ARTICLE

doi:10.1016/S1525-0016(03)00215-6

Global Analysis of Cellular Transcription Following Infection with an HIV-Based Vector

Richard Mitchell,[1] Chih-Yuan Chiang,[1] Charles Berry,[2] and Frederic Bushman[1,*]

[1]*Infectious Disease Laboratory, The Salk Institute, 10010 North Torrey Pines Road, La Jolla, California 92037, USA*
[2]*Department of Family/Preventive Medicine, University of California at San Diego School of Medicine, La Jolla, California 92093, USA*

**To whom correspondence and reprint requests should be addressed. Fax: (858) 554-0341. E-mail: bushman@salk.edu.*

We have examined the changes in cellular transcription resulting from infection with HIV-based vectors. Previous work suggested that the incoming viral genome may under some circumstances be detected as DNA damage, so to explore this possibility, we compared the transcriptional response to infection with an HIV-based vector to the response to treatment with the DNA-damaging agent etoposide. Expression levels of about 12,000 cellular RNA transcripts were determined in a human B-cell line at different times after either treatment. Statistical analysis revealed that the infection with the lentivirus vector resulted in quite modest changes in gene expression. Treatment with etoposide, in contrast, caused drastic changes in expression of genes known or inferred to be involved in apoptosis. Statistically significant though subtle parallels in the cellular transcriptional responses to etoposide treatment and HIV-vector infection could be detected. Several further data sets analyzing infections with HIV-based vectors or wild-type HIV-1 showed similar modest effects on cellular transcription and very modest parallels among different data sets. These findings establish that HIV-vector or HIV-1 infection has remarkably little effect on cellular transcription. The statistical methods described here may be of wide use in mining microarray data sets. Our observations support the idea that gene therapy with HIV-based vectors should not be particularly toxic to cells due to disruption of cellular transcription.

INTRODUCTION

To gain insight into the cellular response to infection by HIV-1 [1,2], we have analyzed changes in host cell transcription following infection with an HIV-based vector [3,4]. The HIV-based vector used is able to carry out reverse transcription and integration, but not later stages of the viral life cycle. Reverse transcription and integration each take place in large nucleoprotein complexes derived from the viral core, the first of which is the reverse transcription complex, which gives rise to the subsequent preintegration complex (PIC). PICs then carry out DNA integration into the host cell chromosome [5–8]. Analysis of infection with the HIV vector allows the cellular response to this subset of viral replication intermediates to be analyzed without the complication of later viral replication steps.

Several studies have suggested that unintegrated retroviral DNA can be toxic to cells [9–13]. One study suggested that the free DNA ends in the unintegrated linear viral DNA may be detected as a DNA double-strand break, leading to induction of a cellular DNA damage response [10]. Other mechanisms for toxic signaling following infection have also been proposed [14,15]. According to these ideas, there should be similarities in the cellular responses to retroviral infection and cellular DNA damage.

To investigate this issue, we compared the transcriptional responses of cells infected with an HIV-based vector to the responses of cells that had experienced DNA damage by treatment with etoposide. As a target cell, we used the previously characterized Nalm6 ligase4$^{-/-}$ human B cell line [16]. This line is impaired for repair of double-strand breaks due to a deletion of the *ligase IV* gene and is known to be particularly sensitive to the toxic effects of infection by HIV-based vectors [10].

Several previous studies have investigated the cellular transcriptional response to infection by replication-competent HIV [17–20]. Corbeil *et al*. studied infection of cultured human CEM cells using Affymetrix arrays that queried the activity of 6800 genes and found that cellular genes involved in the interferon response, DNA-damage response, and apoptosis changed in activity [17]. More than 30% of genes showed reduced expression 72 h after infection, coinciding with the onset of high-level viral transcription. A mostly different set of genes was found to be induced by infection of human primary cells (peripheral blood mononuclear cells) using Affymetrix arrays querying about 12,000 genes. This study was complicated by the observation of significant differences in the response of different human donors [18]. More recently,

1525-0016/03 $30.00

van't Wout *et al.,* using spotted cDNA arrays to query the activity of 4600 genes in several cultured T cell lines, did not observe preferential activation of stress-response or interferon-induced pathways [19]. Instead, transcriptional changes were generally modest and mostly specific to HIV-1 infection. Thus transcriptional profiling studies of HIV infection published to date have yielded rather divergent pictures, emphasizing the need for statistically rigorous analysis of multiple replicates and careful comparisons between conditions.

To assess possible similarities between HIV infection and DNA damage, we have compared cellular transcriptional profiles of time courses after (1) infection with an HIV-based vector or (2) treatment with etoposide. Etoposide traps covalent protein–DNA complexes formed by type II topoisomerases, thereby stabilizing DNA double-strand breaks. Statistical analysis showed some similarities could be seen between the two treatments in genes encoding proteins involved in DNA metabolism and repair, cell cycle progression, and apoptosis. However, the most prominent finding was that infection with the HIV-based vector had remarkably little effect on cellular transcription. In contrast, strong effects were seen in the etoposide-treated cells, allowing identification of new candidate genes potentially involved in proapoptotic pathways. Perhaps surprisingly, a statistically rigorous comparison of several transcriptional profiling studies of cells infected with HIV vectors or replication-competent HIV-1 showed similar modest effects on host cell transcription and quite modest parallels between data sets.

RESULTS

Experimental Treatments

To compare the cellular transcriptional responses to (1) infection with the HIV-based vector and (2) treatment with etoposide, we analyzed sequential time points by transcriptional profiling as outlined in Fig. 1A. We infected cells in duplicate with the HIV-based vector (60 ng p24 antigen per 5×10^5 cells) or treated them in duplicate with etoposide (25 μM). We also carried out mock treatments in duplicate in parallel. Care was taken to match the mocks as closely as possible to the experimental treatment. For the HIV infection, we added Polybrene (4 μg/ml) and "spinoculated" the virus onto cells [22] to increase the efficiency of infection. For the mock samples, we also treated the cells with 4 μg/ml Polybrene and mock spinoculated them. For the etoposide treatment, we applied etoposide in DMSO, so we exposed the mock-treated cells to the same DMSO concentrations.

Following infection with the HIV-based vector, we removed cells at different times and analyzed them by FACS to detect GFP expression (Fig. 1B). GFP expression became detectable after 8 h, and after 48 h 90% of cells were GFP positive. No significant signal was seen in mock-infected cells.

We monitored the effect of etoposide treatment by assaying cell viability, since etoposide is known to be a potent inducer of apoptosis. A previous study of the response of 6591 transcripts revealed that genes for the cellular stress response and cell death were among those affected by etoposide treatment [26]. As expected, cell numbers declined steadily after treatment with etoposide, so that the majority of cells had died within 48 h (Fig. 1C). Control cells, in contrast, continued to double normally. For the HIV infection, under the conditions used here, the cell numbers did not decline significantly compared to the mock-treated cells (data not shown).

We removed aliquots of cells at time zero and at 8, 16, 24, and 48 h from each culture and analyzed them. We extracted the RNA, converted polyadenylated RNA to biotinylated cRNA, and applied it to the Affymetrix HU95Av2 Gene Chips, each of which probes the expression of about 12,000 human genes. We detected the annealed cRNA using avidin–phycoerythrin and quantified the fluorescence intensity. We processed the raw fluorescence intensity data using the Affymetrix MicroArray Suite 3.1 and converted it to expression intensity values using the rma algorithm [24].

Statistical Analysis of Transcription in Cells Infected with the HIV Vector or Treated with Etoposide

We wished to analyze the transcriptional profiling data for statistically significant trends within the HIV and etoposide data sets and for the parallels between data sets. First we tested differences over time between mock-infected and vector-infected cells using the *F* statistic (for time-by-treatment interaction). This test can detect general patterns of differences in the time courses of expression level measurements, so further study of significant findings is needed to indicate the nature of those differences. The linear model included contrasts for replicate, for active versus mock treatment, for time of measurement (4 degrees of freedom), and for differential trend according to treatment (4 degrees of freedom). We used permutation methods to estimate the false discovery rate when claiming that genes above a fixed threshold for the *F* statistic are differentially expressed [27–29]. For the *F* statistic, the rows of the submatrix of contrasts for the treatment-by-time interaction corresponding to one of the replicates are permuted with respect to the other contrasts and the expression level data. This leads to 120 unique permutation statistics and is invariant with respect to choice of which replicate is permuted; no other permutation scheme preserves orthogonality of the contrasts.

The expected number of false positive discoveries among the 200 genes with the largest *F* statistics for the etoposide data is 13.25, and empirical Bayes estimates [29] suggest that 15% of genes on the array were differentially expressed. Thus, etoposide induces differential expression

doi:10.1016/S1525-0016(03)00215-6

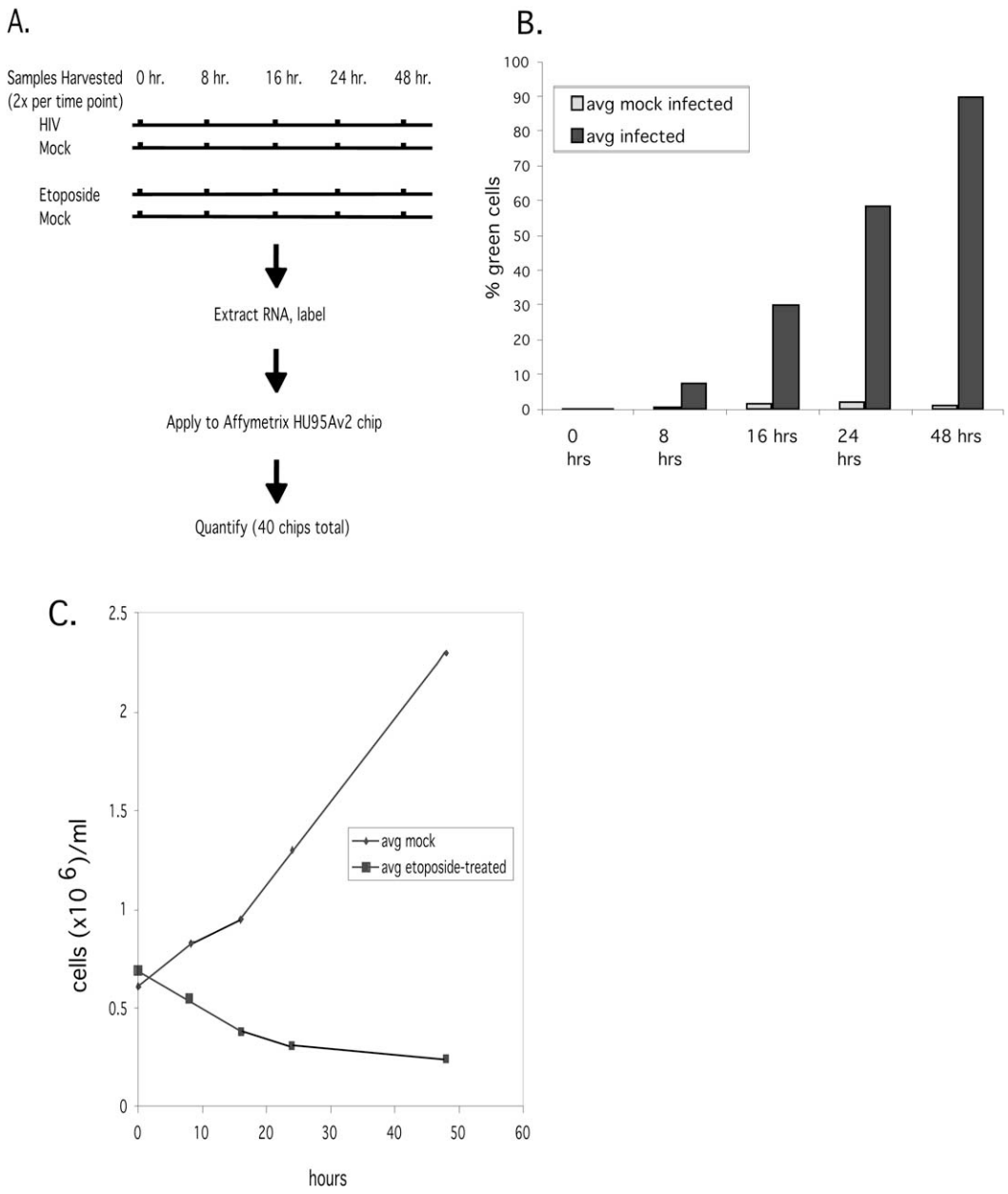


FIG. 1. Experimental plan and characterization of the treatments studied. (A) Diagram of the experimental strategy. (B) Transduction by the HIV-based vector measured by FACS. (C) Toxicity of treatment with etoposide.

in many genes and several hundred are identified with few false positive errors.

We subjected the 200 genes with the largest *F* statistics to hierarchical clustering (Fig. 2A). Hierarchical clustering used etoposide expression values normalized by subtracting gene-specific means to favor clustering of genes that had similar expression patterns regardless of the overall level of expression. We performed average distance agglomerative clustering [30] using the hclust function from the mva package for R [25].

As is evident in Fig. 2A, the majority of the trends shows a strong linear component (by the time scale depicted). However, genes in the clusters displayed at the top and bottom of the figure show strong departures from linear trend. Also evident in Fig. 2A (right column) is an apparent absence of effect of the HIV vector over the same

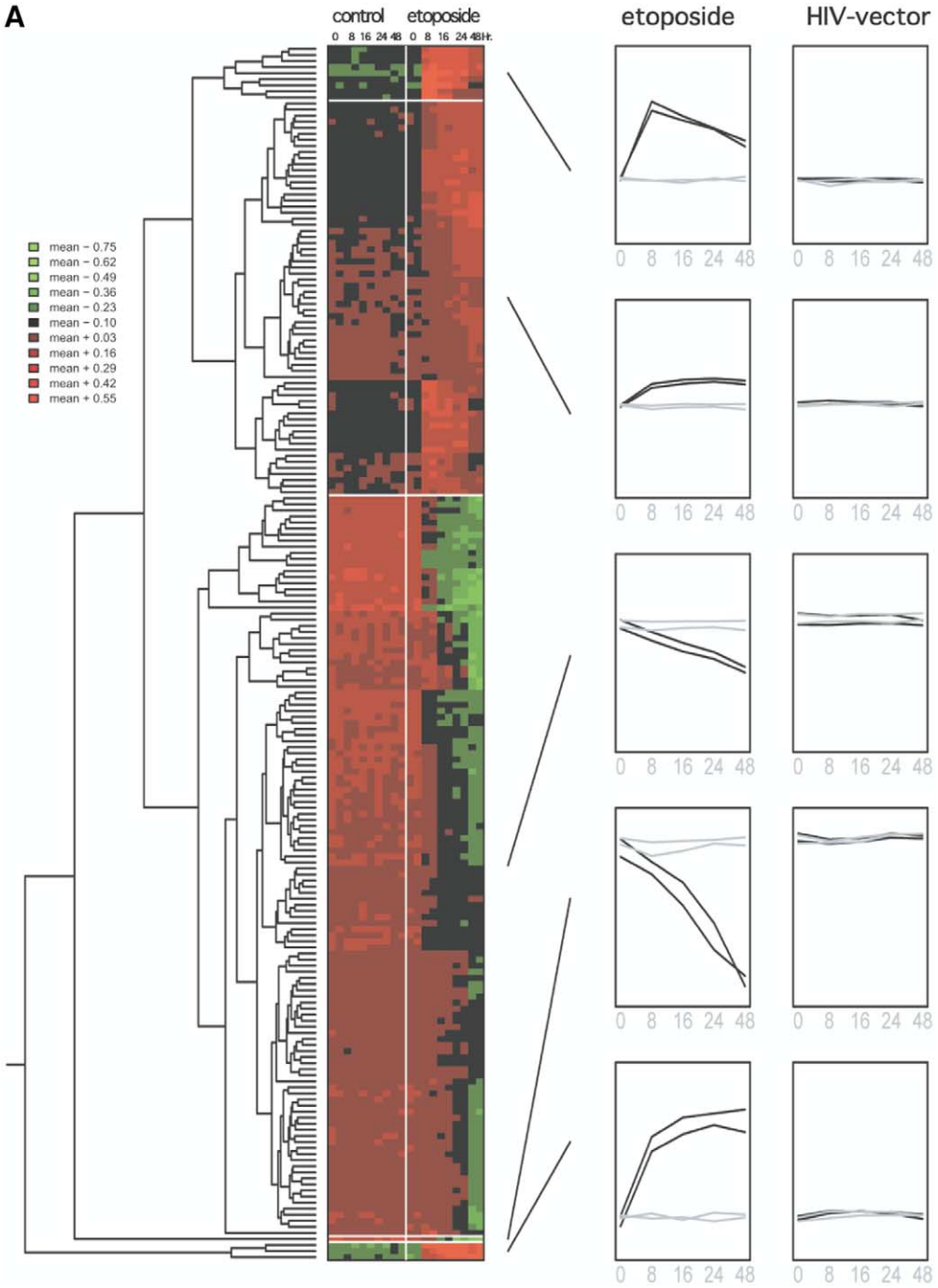


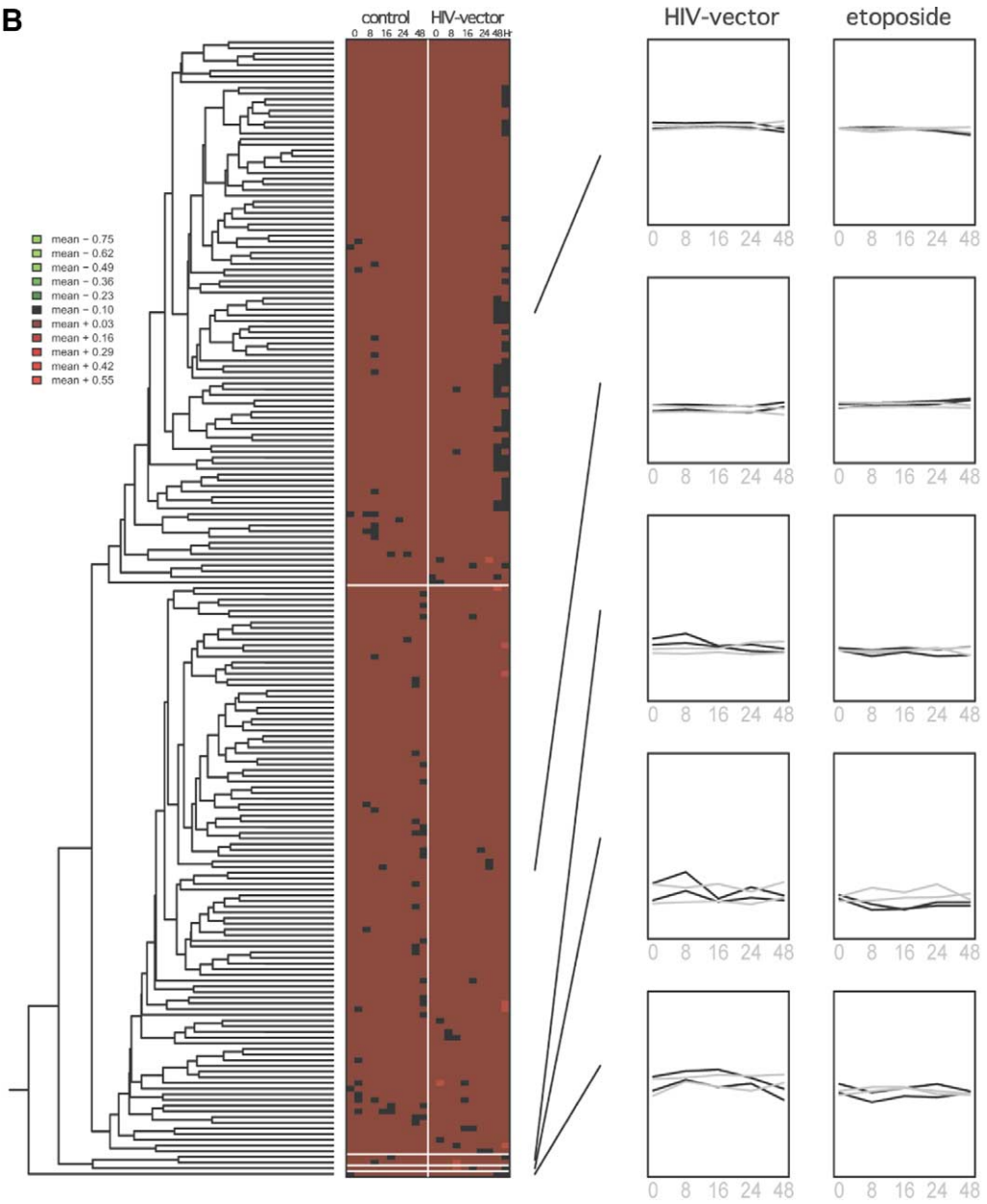
FIG. 2. Analysis of microarray data by hierarchical clustering. (A) Hierarchical clustering of 200 genes with largest *F* statistic for differential expression after etoposide treatment. The dendrogram at left shows the order in which genes were combined to form clusters. The difference between the expression level of a gene and its average over all conditions is shown in false color (green for lower values, red for higher values) for each gene. Each replicate is shown as a separate tile under the indicated times. The relative expression levels over time for each of the five clusters indicated are plotted to the right. Each of the two replicates is shown by a separate line. Dark lines indicate the results for etoposide treatment; gray lines indicate results for mock treatment. At the far right, the cluster averages for HIV vector-treated cells are shown for comparison. (B) Hierarchical clustering of 200 genes with largest *F* statistic for differential expression following infection with the HIV-based vector. Markings parallel those in A.

sets of genes. In fact, the lines for the untreated and HIV-treated are very similar to their baselines and to the etoposide controls. This impression is strengthened by noting that the *F* statistics for differential trends due to HIV-vector infection do not stand out from the permutation distribution and by the unimpressive profiles of the time courses for the genes with the 200 largest *F* statistics (Fig. 2B). Among the 50 largest *F* statistics for HIV-vector infection, 33.65 false discoveries are expected, illustrating the weakness of the signal for changes in transcriptional activity due to HIV-vector infection.

While these results suggest little if any differential in gene expression due to the HIV-vector infection, a penalty is paid for using the *F* statistic rather than a statistic

doi:10.1016/S1525-0016(03)00215-6

FIG. 2 (*continued*)



that searches for just a linear (or quadratic) trend when the dominant component is linear (or quadratic). Therefore we also applied another statistical test. We used the *t* statistic for linear trend in the time codes (0 = 0 h, 1 = 8 h, 2 = 16 h, 3 = 24 h, 4 = 48 h) and that for quadratic trend to examine further the data for etoposide- and HIV-vector-treated cells (compared to their respective controls). The permutation distribution for the *t* statistics for treatment by linear trend is obtained by permuting the contrast for the linear trend by treatment under a model that includes all seven contrasts for lower order terms. Each replicate is independently permuted, leading to 120 × 120 unique permutation statistics. A random sample of 120 of these is used. For quadratic trend, there is no permutation scheme that preserves orthogonality of the contrast for quadratic trend by treatment with respect to the lower order terms, so an asymptotic procedure is used. Residuals from the model fitting the lower order terms are permuted within replicate and treatment, and the full model is then fitted. This procedure is repeated 120 times.

In addition, we performed tests of the rank correlation between the HIV *t* statistic and that for etoposide to assess concordance in expression level changes. Even when few *t* statistics of large magnitude are found for HIV, a signif-

icant correlation with the *t* statistics for etoposide would suggest the presence of changes due to HIV. These tests use the permuted test statistics for the linear and quadratic trends as the reference distribution.

As expected, many genes showed linear trends after etoposide treatment; 1577 genes had *t* statistics with magnitudes of 3.0 or more, while 126.9 false discoveries were expected using the 3.0 threshold. For HIV, 298 genes had *t* statistics with magnitudes of 3.0 or greater, while 79.4 are expected under the permutation distribution. Thus analysis using the *t* statistic did reveal a set of genes affected by HIV-vector infection, though some are expected to be "false" discoveries. However, for the HIV-vector infection the magnitude of the changes detected by the linear *t* test are still quite modest, and analysis by hierarchical clustering yielded unimpressive plots resembling Fig. 2B (data not shown).

Of the 1577 genes that showed linear trends under etoposide, 48 showed similar trends under HIV and 18 showed trends ($|t| \geq 3$) in the opposite direction. When only the sign of the HIV *t* statistic is considered, 1228 showed similar trends and 349 showed linear trends in the opposite direction. Further, the rank correlation between the two sets of 12,625 *t* statistics was 0.42 ($P < 0.01$). These results suggest that HIV infection does induce changes in expression level (albeit weak ones) that parallel some of those seen for etoposide.

The *t* statistic for quadratic trend under etoposide showed that 899 genes had *t* statistics with magnitudes greater than 3.0, while 385 false discoveries would be expected. For HIV 112 genes had *t* statistics with magnitudes greater than 3.0, while 227.2 false discoveries were expected. The rank correlation coefficient between the two sets of *t* statistics is 0.127 ($P = 0.22$). The failure of these results to suggest HIV-induced changes in expression level reflects the absence or small size of quadratic components of change, but it may also reflect the lower power for detecting the quadratic component.

In summary, etoposide treatment induced quite strong changes in gene expression while those induced by infection with the HIV vector were very weak. Analysis using the *F* statistic showed little change in cellular transcription following HIV-vector infection, and so little similarity between etoposide and HIV. Analysis using instead the more sensitive *t* statistic for linear trends showed some parallel changes in cellular transcription due to both HIV-vector infection and etoposide treatment, and in this case subtle but statistically significant parallels could be discerned ($P < 0.01$). The specific genes affected are discussed in more detail below, but first we describe a global comparison of the etoposide and HIV-vector data sets to data monitoring the transcriptional response to infection by replication-competent HIV-1.

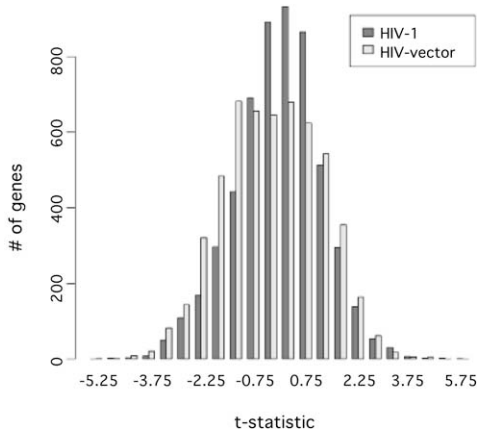


FIG. 3. Comparison of the HIV-vector and HIV-1 data sets using the linear *t* statistic. See text for details.

Statistical Comparison of Infection with Replication-Competent HIV-1, Infection with the HIV-Based Vector, or Treatment with Etoposide

To assess the generality of our conclusions, we next compared the above data set monitoring infection of B cells with an HIV-based vector ("HIV vector data set" below) to a published data set on the transcriptional response of CEM cells, a T cell line, to infection with wild-type HIV [17] (termed "HIV-1 data set" below). The analysis method used *t* statistics for the difference in trend under treated and control conditions. (The regressions from which these were computed also included individual terms for each time of measurement, for HIV versus control, and for replicate in the HIV-vector experiment.) The timing of the measurements is different in the two studies. In the HIV-vector study, the last measurement is at 48 h and there is a baseline measurement. The actual values used in data analysis are just 1, 2, . . . , 5. In the wild-type HIV study, the earliest measurement is at 30 min and the last is at 72 h. The values used are 1.0625, 1.25, 1.5, 2, 3, 4, 5, 6, which are linear interpolations or extrapolation of the values used in the data in the HIV-vector experiment. The distributions of values for the *t* statistics for difference in trend (comparing the HIV-1 data to control) are depicted in Fig. 3. The values for the *t* statistics from the HIV-1 data have only 6 degrees of freedom associated with them while those in the HIV-based vector data have 12. Since the *t* distribution with 6 degrees of freedom is slightly wider than that for 12 degrees of freedom, the *t* statistics for the HIV-1 data set were rescaled as

$$y = t_{12}(t_6^{-1}(x)),$$

where $t_{12}(n)$ refers to the quantile of the *t* distribution for 12 degrees of freedom and $t_6^{-1}(n)$ refers to the area in the left "tail" of the *t* distribution with 6 degrees of freedom.

doi:10.1016/S1525-0016(03)00215-6

TABLE 1: Genes found to be most strongly affected by etoposide treatment using the linear *t* statistic and comparison to the response of these genes after HIV-1 and HIV vector infection

LocusLinkID	Gene	HIV vector	HIV-1	Etoposide	Function
1351	*COXB*	−1.80	−2.14	15.31	Cytochrome *c* oxidase subunit VIII
6631	*SNRPC*	−0.30	−0.68	10.50	Small nuclear ribonucleoprotein polypeptide C
5439	*POLR2J*	−3.78	−0.86	10.35	Polymerase (RNA) II (DNA-directed) polypeptide J
6223	*RPS19*	0.24	−0.34	10.28	Ribosomal protein S19
941	*CD80*	0.04	0.70	10.17	CD80 antigen (CD28 antigen ligand 1, B7-1 antigen)
1347	*COX7A2*	−1.31	−0.10	10.14	Cytochrome *c* oxidase subunit VIIa polypeptide 2
1212	*CLTB*	−3.71	2.06	10.10	Clathrin, light polypeptide (Lcb)
3265	*HRAS*	−2.68	4.29	9.88	v-Ha-ras Harvey rat sarcoma viral oncogene homolog
10437	*IFI30*	−1.52	1.02	9.81	Interferon, γ-inducible protein 30
5714	*PSMD8*	−2.19	−2.24	9.37	Proteasome 26S subunit, non-ATPase, 8
1801	*DPH2L1*	−1.60	2.09	9.24	Diptheria toxin-resistance protein
8337	*HIST2H2AA*	0.06	−0.81	8.98	Histone 2, H2aa
5510	*PPP1R7*	−2.17	−1.54	8.93	Protein phosphatase 1, regulatory subunit 7
3956	*LGALS1*	−2.65	1.60	8.91	Lectin, galactoside-binding, soluble, 1 (galectin 1)
1327	*COX4I1*	−0.91	−0.27	8.88	Cytochrome *c* oxidase subunit IV isoform 1
5928	*RBBP4*	−1.53	−1.28	−10.56	Retinoblastoma binding protein 4
6418	*SET*	−0.86	−3.52	−10.57	SET translocation (myeloid leukemia-associated)
10151	*HNRPA3*	−1.74	−0.85	−10.58	Heterogeneous nuclear ribonucleoprotein A3
7852	*CXCR4*	−1.00	−0.98	−10.79	Chemokine (C-X-C motif) receptor 4
387	*ARHA*	−0.88	−3.37	−11.43	Ras homolog gene family, member A
16	*AARS*	−1.75	−1.82	−11.59	Alanyl-tRNA synthetase
1938	*EEF2*	−1.65	−1.75	−11.60	Eukaryotic translation elongation factor 2
3543	*IGLL1*	−2.37	−1.04	−11.73	Immunoglobulin λ-like polypeptide 1
23181	*DIP2*	−0.99	0.53	−12.83	Disco-Interacting protein 2 (*Drosophila*) homolog
3646	*EIF3S6*	−1.00	−1.63	−12.91	Eukaryotic translation initiation factor 3, subunit 6
4691	*NCL*	−0.73	−3.06	−13.20	Nucleolin
1299	*COL9A3*	−1.67	0.69	−14.22	Collagen, type IX, α3
4500	*MT1L*	−1.30	−1.85	−15.58	Metallothionein 1L
4494	*MT1F*	−0.83	0.79	−17.37	Metallothionein 1F (functional)
4496	*MT1H*	−1.62	−2.09	−17.65	Metallothionein 1H

The values in the column are linear *t* statistics.

As is evident, the tails of the two distributions show no appreciable differences. Thus there are relatively few genes showing significant changes due to infection with either the HIV-based vector or HIV-1 in these data sets.

The Pearson correlation between the *t* statistics for the HIV-1 and HIV vector data sets shows a modest correlation of 0.209, while similar *t* statistics from each of the two replicates of the HIV-vector data show a correlation of 0.202 with each other. The correlations of a similar collection of *t* statistics from the etoposide data are 0.239 and 0.376 for HIV-1 and HIV-vector data, respectively. These low correlations are as expected if there are relatively few parallels among the data sets due to low signal intensity.

Genes Most Strongly Affected by Treatment with Etoposide

We begin the discussion of transcriptional changes in specific genes by describing the robust effects seen after treatment with etoposide. Table 1 compares the genes most strongly affected by etoposide treatment as judged by the linear *t* statistic to the linear *t* statistics for the same genes in the HIV-vector and HIV-1 data sets. The 30 most

strongly affected genes (15 most positive and 15 most negative *t* statistic values) showed relatively strong trends ($|t$ statistic$| > 8.88$). The sign of the *t* statistic indicates whether the gene was activated (positive value) or repressed (negative value) by the treatment. Three upregulated genes encoded subunits of cytochrome *c* oxidase, a mitochondrial respiratory protein previously implicated in proapoptotic signaling [31]. The most strongly downregulated genes were three metallothionein genes—metallothioneins have been previously implicated as antiapoptotic, so their downregulation is expected to promote apoptosis [32]. The nucleolin gene was also strongly downregulated, a change that has been seen previously in cells undergoing apoptosis in response to several stimuli [33]. Some of the genes found to be strongly modulated in this data set have not been previously associated with apoptosis (viz. histone H2AA, RNA polymerase II subunit J, translation factor S19, and EIF3S6) and are of interest for further investigation of the apoptotic mechanism.

Additional significantly affected genes encoded DNA damage response factors and apoptotic regulators, as suggested from previous work [26]. The genes for the proapoptotic factors BAK and BAX-δ were upregulated, while the genes for the antiapoptotic proteins BCL-2 and BCL-$_{xL}$ were downregulated. Affected genes encoding further apoptotic regulators included PDCD8, DAP3, caspase 10, BAX, SIVA, and BNIP3L. Genes encoding products involved in the DNA damage response that were altered in activity included MRE11A; NBS1; DNAPK; Ku86; LIG1; BLM; ERCC2, 4, 5, and 6; RAD21, 23, and 54L; MSH6; AP lyase; MGMT; OGG1; ADPRTL2; and MPG. Genes involved in cell cycle control and response to stress were also affected.

None of the 30 most strongly affected genes (Table 1) were significantly altered in the same direction in the HIV-vector data set, and only 3 were affected in the same direction in the HIV-1 data set. Thus in the genes most strongly affected by etoposide there are relatively few parallels to the HIV-vector or HIV-1 data set. We return to those genes similarly affected in multiple data sets under Discussion.

Genes Most Strongly Affected by Infection with the HIV-Based Vector

Table 2 compares the genes most strongly affected by infection with the HIV-based vector to the effects of infection with HIV-1 or treatment with etoposide on those same genes. Comparison with Table 1 shows that the *t* statistics for the genes in the HIV-vector data set are much lower than those for genes most affected by etoposide, indicating weaker signal. This set of genes is expected to be composed of some significantly affected by infection and others that are incorrectly identified.

Two histocompatibility genes are downregulated, as are a number of genes for signaling proteins. Upregulated genes include an integrin, the Kell blood group factor, and several signaling molecules. Again there were few parallels with the HIV-1 data set or the etoposide data set (two and five genes affected in the same direction, respectively).

Genes Affected by Replication-Competent HIV-1

Table 3 compares genes most strongly affected by infection with replication-competent HIV-1 [17] to genes affected by infection with the HIV-based vector or etoposide treatment. As has been reported previously, in the HIV-1 data set there were apparent effects on genes encoding transcription factors, signal transduction molecules, and proteases. However, the fraction of genes detected as significantly affected is low and the magnitude of the changes is not large. None of the genes in Table 3 showed significant change after infection with the HIV-based vector, and 6 of the 30 genes showed potentially significant changes after treatment with etoposide.

Verification of Expression Changes Using Quantitative PCR

The changes in gene activity observed by transcriptional profiling were verified for seven representative genes by fluorescence-monitored PCR. Aliquots of RNA from each time point were reverse transcribed, then cDNA was used as a template for amplification. Accumulation of DNA products was monitored during amplification by the increase in fluorescence of added SYBR green dye, which fluoresces upon binding to double-stranded DNA.

Figure 4 compares the activity of the seven genes measured by fluorescence-monitored PCR (left column) and transcriptional profiling (right column). The PCR values were normalized to β-actin to account for possible differences arising during cDNA preparation. Genes encoding cellular c-myc, metallothionein, and nucleolin were significantly decreased in activity after the etoposide treatment in the transcriptional profiling data set (t statistic < -3), and this was reproduced in the quantitative PCR assay (Fig. 4). Genes for ATP synthetase and SNRPC were induced by etoposide treatment (t statistic > 3), and this was reproduced by PCR analysis. No significant changes were seen after etoposide treatment in the genes for acidic ribosomal protein P0 or PKR in the transcriptional profiling data, and no significant changes were seen by fluorescence-monitored PCR. Regarding the response of these genes after infection with the HIV-based vector, none showed significant changes in the transcriptional profiling data, and no significant changes were detected by fluorescence-monitored PCR (Fig. 4, compare lines marked I1 and IM1). Thus we were able to confirm the general trends seen for seven genes in the transcriptional profiling data using quantitative PCR.

doi:10.1016/S1525-0016(03)00215-6

TABLE 2: Genes found to be most strongly affected by HIV-vector infection using the linear *t* statistic and comparison of the response of these genes to the response after HIV-1 infection and etoposide treatment

LocusLinkID	Gene	HIV vector	HIV-1	Etoposide	Function
8516	*ITGA8*	4.31	0.60	1.10	Integrin, α8
3792	*KEL*	4.20	1.42	7.67	Kell blood group
5026	*P2RX5*	4.17	−0.16	3.12	Purinergic receptor P2X, ligand-gated ion channel, 5
8613	*PPAP2B*	4.14	0.11	0.24	Phosphatidic acid phosphatase type 2B
6505	*SLC1A1*	4.14	0.75	0.29	Solute carrier family 1, member 1
5653	*KLK6*	3.89	−0.67	3.51	Kallikrein 6 (neurosin, zyme)
10234	*P37NB*	3.88	0.58	0.43	37-kDa leucine-rich repeat protein
3270	*HRC*	3.74	−0.58	1.89	Histidine-rich calcium-binding protein
7123	*TNA*	3.71	−1.87	2.97	Tetranectin (plasminogen-binding protein)
438	*ASMT*	3.68	−1.06	1.74	Acetylserotonin O-methyltransferase
4796	*NFKBIL2*	3.58	0.33	−0.02	Nuclear factor of κ light chain gene enhancer inhibitor-like 2
1961	*EGR4*	3.44	2.60	0.18	Early growth response 4
1585	*CYP11B2*	3.43	2.53	2.08	Cytochrome P450, family 11, subfamily B, polypeptide 2
22915	*MMRN*	3.42	2.43	2.22	Multimerin
2039	*EPB49*	3.42	1.55	6.40	Erythrocyte membrane protein band 4.9 (dematin)
23526	*HA-1*	−3.89	0.10	−1.77	Minor histocompatibility antigen HA-1
3134	*HLA-F*	−3.90	0.54	0.81	Major histocompatibility complex, class 1, F
381	*ARF5*	−3.92	−0.81	0.73	ADP-ribosylation factor 5
682	*BSG*	−3.93	−3.35	3.06	Basigin (OK blood group)
6204	*RPS10*	−4.00	−0.19	−0.67	Ribosomal protein S10
8189	*SPK*	−4.03	−1.09	−0.35	Symplekin
578	*BAK1*	−4.13	1.27	3.49	BCL2-antagonist/killer 1
4682	*NUBP1*	−4.16	−0.33	−3.02	Nucleotide binding protein 1 (MlnD homolog, *Escherichia coli*)
25828	*TXN2*	−4.26	−3.76	−0.34	Thioredoxin 2
9522	*SCAMP1*	−4.27	0.68	0.65	Secretory carrier membrane protein 1
27238	*T54*	−4.27	−0.68	−0.84	T54 protein
5438	*POLR2I*	−4.30	−2.24	0.18	Polymerase (RNA) II (DNA directed) polypeptide I
9159	*PCSK7*	−4.32	−1.61	0.71	Proprotein convertase subtilisin/kexin type 7
4125	*MAN2B1*	−4.75	−0.41	0.65	Mannosidase, α, class 2B, member 1
3135	*HLA-G*	−5.41	−0.38	4.99	HLA-G histocompatibility antigen, class I, G

DISCUSSION

Which human genes are most strongly affected by HIV infection across many data sets? Even when signal intensity is low, comparison of enough independent measurements has the potential to reveal subtle but biologically relevant changes in transcription. Figure 5 shows a summary of genes affected in more than two of five transcriptional profiling data sets monitoring the response to HIV infection or DNA damage. Each row in the figure indicates a single gene, and each numbered column indicates a transcriptional profiling data set. Red indicates genes significantly upregulated, green downregulated, and gray not regulated. Columns 1 and 2 show genes affected by infection with HIV-based vectors, columns 3 and 4 show genes affected by HIV-1, and column 5 shows the etoposide data set. The *t* statistics are shown where available (columns 1, 3, and 5). The genes found to be affected by HIV-1 infection in column 4 are of particular interest, since these are derived from 71 genes previously found to be reproducibly affected by HIV-1 infection across multiple data sets [19]. Other methods were used to identify potentially regulated genes in column 2 (Fig. 5 legend).

TABLE 3: Genes found to be most strongly affected by HIV-1 infection using the linear *t* statistic and comparison to the response of these genes after HIV-vector infection and etoposide treatment

LocusLinkID	Gene	HIV vector	HIV-1	Etoposide	Function
6925	*TCF4*	−1.72	9.57	−1.00	Transcription factor 4; binds E box
4162	*MCAM*	2.05	7.13	−0.84	Melanoma cell adhesion molecule
3142	*HLX1*	2.01	6.50	0.83	H2.0-like homeobox 1 (*Drosophila*)
4149	*MAX*	−1.14	6.13	−2.17	MAX protein
11054	*OGFR*	−2.18	5.63	−0.34	Opioid growth factor receptor
10628	*TXNIP*	−0.34	5.06	1.46	Thioredoxin interacting protein
10136	*ELA3A*	1.20	4.99	−0.01	Elastase 3A, pancreatic (protease E)
433	*ASGR2*	0.73	4.98	0.48	Asialoglycoprotein receptor 2
7067	*THRA*	1.93	4.81	1.81	Thyroid hormone receptor, α (v-erb-a homolog)
2224	*FDPS*	−1.97	4.67	−3.42	Farnesyl diphosphate synthase
5328	*PLAU*	−1.00	4.67	−0.52	Plasminogen activator, urokinase; inhibits HIV
616	*BCRL4*	0.00	4.60	4.58	Breakpoint cluster region-like 4
1604	*DAF*	−2.33	4.33	−0.99	Decay accelerating factor for complement (CD55)
10794	*ZNF272*	1.62	4.33	1.58	Zinc finger protein 272
3265	*HRAS*	−2.68	4.29	9.88	v-Ha-ras Harvey rat sarcoma viral oncogene homolog
498	*ATP5A1*	−0.77	−4.38	−3.15	ATP synthase, mitochondrial F1 complex, α subunit
6905	*TBCE*	−1.03	−4.57	0.86	Tubulin-specific chaperone e
130	*ADH6*	−0.05	−4.58	−0.23	Alcohol dehydrogenase 6 (class V)
1457	*CSNK2A1*	−1.02	−4.60	−0.77	Casein kinase 2, α1 polypeptide
25805	*NMA*	−0.65	−4.70	1.94	Putative transmembrane protein
805	*CALM2*	−0.58	−4.85	−6.55	Calmodulin 2 (phosphorylase kinase, δ)
7137	*TNNI3*	0.51	−5.10	0.30	Troponin I, cardiac
3312	*HSPA8*	−0.03	−5.22	−5.71	Heat shock 70-kDa protein 8
5550	*PREP*	−0.55	−5.25	0.12	Prolyl endopeptidase
5514	*PPP1R10*	0.20	−5.38	0.19	Protein phosphatase 1, regulatory subunit 10
57007	*RDC1*	−0.50	−5.52	−0.09	G protein-coupled receptor
2303	*FOXC2*	0.87	−6.01	0.23	Forkhead box C2 (MFH-1, mesenchyme forkhead 1)
10239	*AP3S2*	−1.85	−6.02	−0.59	Adaptor-related protein complex 3, σ2 subunit
4319	*MMP10*	0.09	−7.26	2.06	Matrix metalloproteinase 10 (stromelysin 2)
3678	*ITGA5*	0.15	−7.33	−3.02	Integrin, α5 (fibronectin receptor, α polypeptide)

Relatively few genes were significantly affected in each infection experiment, and of those only 119 were common among two or more experiments. Only 3 genes were affected in three data sets, and none were affected in more than three. Most of the genes on the list are common to the etoposide data set and one infection data set. These genes should be interpreted with caution, since many more genes were significantly affected in the etoposide data set than in the others, so there is greater potential for chance matches. Thus parallels among the data sets appear to be quite modest, consistent with the view that infection has only slight effects on cellular transcription.

The commonly affected genes in Fig. 5 were analyzed for function using the Amigo gene ontology resource through the Pogo browser (http://gene.ucsd.edu/pogo.jsp) and NCBI LocusLink. Genes in the DNA metabolism/repair, apoptosis, and immune response categories are of particular interest and are marked (yellow, orange, and blue, respectively). The finding of genes in the first two categories provides some support for the hypothesis that unintegrated retroviral DNA can induce a DNA damage response in cells [10]. Many of the remaining genes are involved in signal transduction or transcriptional control, again consistent with a role in signaling after infection.

doi:10.1016/S1525-0016(03)00215-6

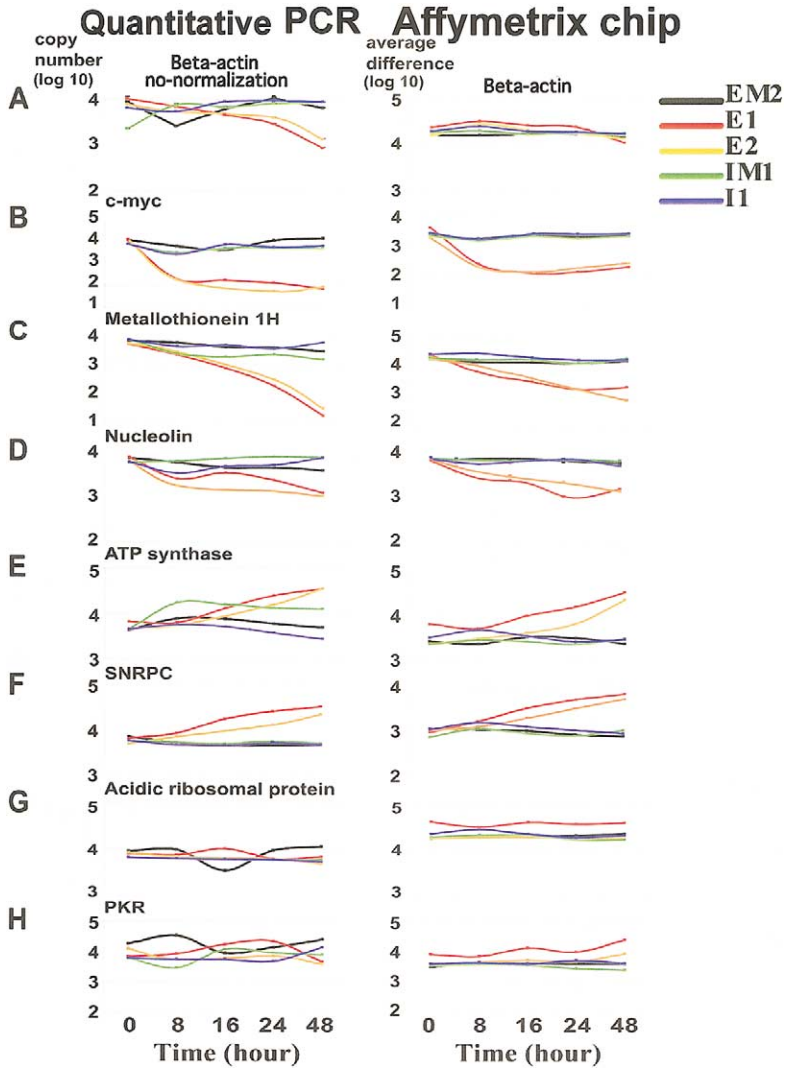
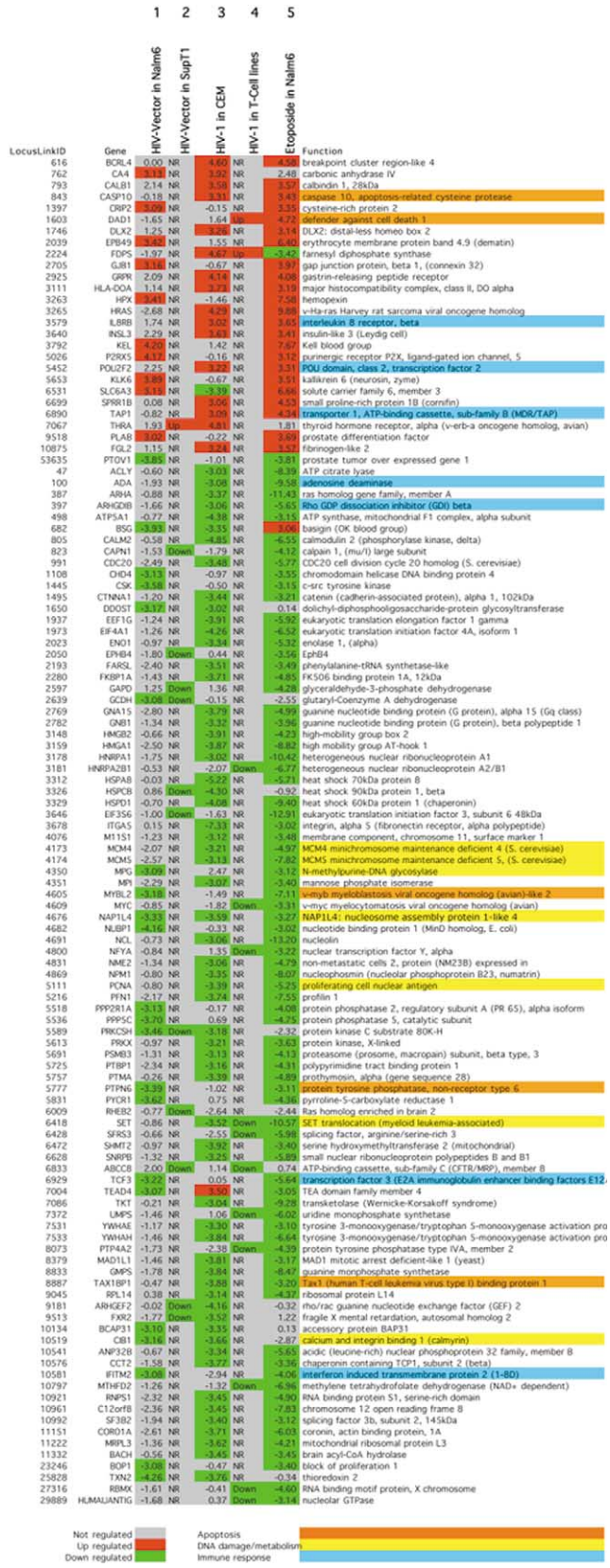


FIG. 4. Comparison of results of quantitative PCR (left column) and transcriptional profiling (right column) analysis for eight genes. The color code for RNA samples analyzed is shown at the top right: EM2 indicates etoposide mock (carried out in parallel with etoposide experiment E2), E1 and E2 indicate two separate etoposide treatments, IM1 indicates the infected cell mock, and I1 indicates the infected cell samples. LocusLink ID numbers for the genes studied are as follows: β-actin (60), c-myc (4609), metallothionein (4496), nucleolin (4691), ATP synthase (521), SNRPC (6631), acidic ribosomal protein P0 (6175), and PKR (5610).

The observation that genes in the immune response category were affected suggests that infected cells may be sensing infection and responding to control viral spread [34].

Of the three genes regulated in three data sets, two were downregulated and members of the Pogo DNA metabolism/repair category. These were the nucleosome assembly protein NAP1L and the SET protein. The third protein, protein kinase C substrate 80K-H, was also downregulated. Regulation of these three genes warrants further study as potentially biologically relevant cellular responses to infection.

Pogo analysis of the HIV-vector data alone revealed that several further genes that encoded proteins involved in DNA metabolism/repair, cell cycle, and apoptosis were affected ($|t$ statistic$| > 3$), further supporting the idea that there are parallels in the cellular response to HIV infection and induction of DNA damage. Seven genes encoding proteins important in DNA metabolism and repair were repressed by infection (CIB1, UBE1, DFFA, NAP1L4, ERCC1, MPG, and POLE) and none were activated. This may reflect induction of a proapoptotic pathway, in which genes for DNA repair proteins are downregulated to allow chromosomal degradation. Nine genes involved in cell cycle progression were repressed (CSK, MYBl2, NOL1, SSSCA1, AKAP8, NUBP1, PPP5C, DCTN3, and PPM1G) and none were induced. Several of these genes encode functions important in signal transduction and transcriptional control, providing candidates for the intermediate signaling molecules that result in growth arrest and apoptosis following infection [10]. Six genes encoding apoptotic regulators were also repressed (DAP3, STK3, BAG1, SCARB1, PTPN6, and BAK1) and none were activated, indicating effects on the pathways directing



apoptosis [10,17]. While these data suggest a possible hypothesis regarding the role of host factors in viral replication, the magnitudes of the changes in gene activity are modest and further testing is needed to substantiate a possible role in the viral life cycle.

Why did the four transcriptional profiling data sets in Fig. 5, columns 1–4, yield mostly different gene sets? As shown in Figs. 2–4, the overall extent of changes induced by infection was quite modest—thus affected genes may be few in number, the magnitude of transcriptional changes modest, and their detection complicated by noise in the measurements. Extreme possibilities for the differences between infection data sets (Fig. 5, columns 1–4) could be (1) lack of signal and poor discrimination between signal and noise or (2) genuine differences between the cellular response to infection in the different transcriptional profiling experiments. Many parameters differed among the experiments in Fig. 5, columns 1–4, including the HIV-vector or HIV-1 genome used, the cell-type infected, the gene chip used, the handling of cells during workup, and the subsequent data processing. Accumulation of further transcriptional profiling data sets should allow resolution of some of these issues, particularly with multiple experimental replicates and careful statistical analysis of each condition analyzed.

In summary, infection of cells with the HIV vector had quite modest effects on transcription, particularly compared to etoposide treatment. Only a small number of genes in the infected cells and etoposide-treated cells showed common changes in transcription because so few genes responded to infection. Comparison of genes affected by infection with HIV vectors or HIV-1 across four data sets showed quite modest parallels, though a few commonly affected genes could be identified and these are of interest for further study (Fig. 5). Some of the commonly affected genes are involved in the response to DNA damage and apoptosis, as are further genes uniquely affected in each data set, providing some support for the idea that unintegrated HIV DNA genomes can be detected as DNA damage. The statistical methods described in this study may be useful in identifying subtle parallels between transcriptional profiling data sets in other studies.

FIG. 5. Genes affected in multiple data sets monitoring global transcriptional responses to treatment with HIV-based vector, HIV-1, or etoposide. Any gene significantly affected in two or more data sets is shown. Color codes are indicated at the bottom. The *t* statistics are plotted where available ($|t$ statistic$| > 3$ was judged significant in columns 1, 3, and 5). For the "HIV-1 in T-cell line" data set, the 71 genes called significant by [19] were used as the initial set for comparison. For the "HIV-vector in SupT1" data set, the most strongly affected 200 genes were identified by: (i) ranking genes by expression level before and after infection, (ii) selecting those genes with the greatest changes in rank order following infection, and (iii) selecting those genes called present by the Affymetrix software prior to infection (downregulated genes) or after infection (upregulated genes). Origins of data sets are as follows: 1, HIV-vector in Nalm6, this work; 2, HIV-vector in SupT1 [35]; 3, HIV in CEM [17]; 4, HIV in T-cell lines [19]; 5, Etoposide in Nalm6, this work.

doi:10.1016/S1525-0016(03)00215-6

TABLE 4: Correlating LocusLink IDs with Affymetrix probe sets

		hu6800									
		1	**2**	**3**	**4**	**5**	**6**	**7**	**9**	**10**	**13**
Hu95Av2	**1**	3622	270	37	5	1	0	0	0	0	0
	2	825	215	37	12	0	1	0	0	0	0
	3	196	105	27	6	1	0	0	0	0	0
	4	44	27	16	8	0	0	0	0	0	0
	5	5	9	5	1	0	0	0	0	0	0
	6	3	3	3	2	1	2	0	0	0	0
	7	1	0	3	1	2	0	1	0	1	0
	8	1	1	0	0	0	1	0	1	0	0
	11	0	0	0	0	0	0	0	0	0	1

The number of LocusLinkIDs for each Affymetrix probe set is shown for the HU95Av2 and HU6800 chips.

Our findings indicate that transduction of cells with lentiviral vectors causes relatively modest changes in cellular transcription, suggesting that in this respect these vectors may be relatively benign when used for gene therapy.

METHODS

Cell culture and viral-vector infection. The Nalm6 ligase4$^{-/-}$ cell line, a human B cell line with both copies of the ligase 4 gene inactivated by deletion [16], was cultured in RPMI with 10% FBS, 1% penicillin/streptomycin/glutamine, 1 mM Hepes, 1 mM Na pyruvate, 100 μM nonessential amino acids, and 55 μM β-mercaptoethanol. The HIV-based vector used was produced by transfection of 293T cells with three plasmids [3]. One plasmid contained the genome that becomes integrated (the self-inactivating HIV vector (p156RRLsinPPTCMVGFPPRE) described in [4,21]). The genome has the HIV LTRs bracketing the HIV central polypurine tract and the gene for the green fluorescent protein (GFP). The packaging plasmid (pdeltaR9 [3]) encodes HIV *gag–pol* genes under control of the CMV promoter. This construct also encodes all of the HIV auxiliary genes (*tat, rev, nef, vif, vpr, vpu*)—thus the transcriptional changes seen after infection may potentially be affected by action of any of these proteins imported in viral particles. The envelope plasmid encodes the gene for VSV-G, the envelope protein for VSV, which allows the vector particles to gain entry into the B cells. Supernatants were collected 48 h after transfection, centrifuged to pellet cellular debris, and then filtered through 0.45-μm filters. Viral particles were then pelleted by centrifugation at 23,000*g* and resuspended in 1/30 volume of fresh medium. Recovery of viral particles was quantified using the Dupont p24 ELISA.

Nalm6 ligase4$^{-/-}$ cells were infected at a multiplicity of infection of 10 (as determined by titration on the permissive 293T cell line) in the presence of Polybrene (4 μg/ml) added to both mock- and vector-infected cells. Cells were centrifuged at 1200*g* at 20°C for 1 h in the presence of virus (spinoculation, as described in [22]) to increase the efficiency of infection. Half of the culture medium was then removed and replaced with fresh medium. The cells were incubated at 37°C. Aliquots of 5×10^6 cells were taken at time 0 (before addition of the vector), 8, 16, 24, and 48 h after vector addition or mock treatment. Cells were resuspended in 250 μl of PBS to which 750 μl of Trizol was added and stored at −80°C. RNA was extracted by treatment with chloroform and alcohol precipitation.

Treatment to induce DNA damage. Etoposide dissolved in DMSO was added to the cells at a final concentration of 25 μM. DMSO was added at the same concentration to the control cells. Aliquots of 5×10^6 cells were taken at time 0 (before addition of the etoposide), 8, 16, 24, and 48 h after etoposide addition or mock treatment.

Transcriptional profiling. Labeled RNA (cRNA) was prepared as per the Affymetrix protocol [23]. The RNA was hybridized to Affymetrix Gene Chips HU95Av2. The data were processed using Affymetrix GeneChip MicroArray Suite software 4.0. Normalization of expression level was carried out using the rma function [24] in the affy package (http://www.bioconductor.org/) for R [25]. Logarithms of the expression scores were used in all analyses. Lists of genes affected by infection with the HIV-based vector, treatment with etoposide, or both are available upon request.

The data produced here used the Hu95Av2 Affymetrix chip, while those produced by [17] used the Hu6800 chip. The probe sets on these two chips overlap, but are not identical, requiring development of methods to map between them. To match the probe sets on these two chips we used the LocusLink IDs provided in the packages hgu95a and hu6800 available on the Bioconductor Web site (http://www.bioconductor.org/), kindly provided by Jianhua Zhang (jzhang@jimmy.harvard.edu). Using these we listed the Affymetrix Probe IDs for each LocusLink ID (eliminating LocusLink IDs for which there was no probe set for at least one chip). In all, there are 5503 LocusLink IDs for which there is at least one matching Affymetrix probe set. In some cases, there are multiple Affymetrix probe set IDs corresponding to a single LocusLink ID. The number of probe sets of each is presented in Table 4. To compare expression values for the two

TABLE 5: Oligonucleotides used for SYBR green quantification of gene activity

Actin	
Forward primer	CGAGAAGATGACCCAGATCATGTT
Reverse primer	CCTCGTAGATGGGCACAGTGT
c-myc	
Forward primer	TCAAGAGGTGCCACGTCTCC
Reverse primer	TCTTGGCAGCAGGATAGTCCTT
Metallothionein 1H	
Forward primer	GCCCAGGGCTGCATCTG
Reverse primer	TTACGTGTCATTCTGTTTTCATCTGA
Nucleolin	
Forward primer	CGCTAAAGAAGCTTTAAATTCCTGTAA
Reverse primer	GGTGATCCCCTGGGTCCTT
ATP synthase	
Forward primer	CCGCTGCTGGTCCAAAGA
Reverse primer	GAAACGGATTGCCAGAGAATTG
SNRPC	
Forward primer	TGCAGTGGAAGGAAACACAAAG
Reverse primer	TTGTTTTGTCAATCAGGCTCTGA
Acidic ribosomal protein	
Forward primer	AGATGCAGCAGATCCGCAT
Reverse primer	GTTCTTGCCCATCAGCACC
PKR	
Forward primer	AACCTGAGGATCGACCTAACACA
Reverse primer	ACACAGTCAAGGTCCTTAGTATTTCAGAT

chips, we first calculated the expression values from the respective CEL files using the rma function of the affy library (authors: Rafael A. Irizarry, Laurent Gautier, Leslie M. Cope, and Benjamin Milo Bolstad with contributions from Magnus Astrand and Christopher Workman) available from the Bioconductor Web site. Probe sets were normalized for each experiment, then the average value for the collection of probe sets corresponding to a single LocusLink ID was found for each chip and for every ID. Thus, 5503 values are calculated for each chip.

Quantitative PCR. Changes in gene activity were confirmed for seven genes using quantitative PCR. One microgram of each RNA sample was reverse transcribed using the Omniscript kit from Qiagen. PCR products were detected using the fluorescent dye SYBR green (Applied Biosystems). The primers used for this analysis are shown in Table 5. Formation of a unique DNA product was confirmed after each PCR by verifying that products had a single melting temperature. Fluorescence-monitored PCR values were normalized to β-actin values to account for any differences in cDNA recovery between samples. Time zero values were also normalized to a common value in each panel to parallel the data processing of gene chip data.

ACKNOWLEDGMENTS

We thank members of the Salk Institute Infectious Disease Laboratory for helpful discussions. This work was supported by Grants AI34786 and AI53820 from the National Institutes of Health, the James B. Pendleton Charitable Trust, the Berger Foundation, the Margaret Morris Foundation, and Frederic and Robin Withington.

RECEIVED FOR PUBLICATION FEBRUARY 25, 2003; ACCEPTED JUNE 16, 2003.

REFERENCES

1. Coffin, J. M., Hughes, S. H., and Varmus, H. E. (1997). Retroviruses. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
2. Bushman, F. D. (2001). Lateral DNA Transfer: Mechanisms and Consequences. Cold Spring Harbor Laboratory Press, New York.
3. Naldini, L., *et al.* (1996). In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. *Science* **272:** 263–267.
4. Zufferey, R., *et al.* (1998). Self-inactivating lentivirus vector for safe and efficient in vivo gene delivery. *J. Virol.* **72:** 9873–9880.
5. Brown, P. O., Bowerman, B., Varmus, H. E., and Bishop, J. M. (1987). Correct integration of retroviral DNA in vitro. *Cell* **49:** 347–356.
6. Farnet, C. M., and Haseltine, W. A. (1990). Integration of human immunodeficiency virus type 1 DNA in vitro. *Proc. Natl. Acad. Sci. USA* **87:** 4164–4168.
7. Fassati, A., and Goff, S. P. (1999). Characterization of intracellular reverse transcription complexes of Moloney murine leukemia virus. *J. Virol.* **73:** 8919–8925.
8. Miller, M. D., Farnet, C. M., and Bushman, F. D. (1997). Human immunodeficiency virus type 1 preintegration complexes: studies of organization and composition. *J. Virol.* **71:** 5382–5390.
9. Temin, H. M., Keshet, E., and Weller, S. K. (1980). Correlation of transient accumulation of linear unintegrated viral DNA and transient cell killing by avian leukosis and reticuloendotheliosis viruses. *Cold Spring Harbor Symp. Quant. Biol.* **44:** 773–778.
10. Li, L., *et al.* (2001). Role of the non-homologous DNA end joining pathway in retroviral infection. *EMBO J.* **20:** 3272–3281.
11. Butler, S., Hansen, M., and Bushman, F. D. (2001). A quantitative assay for HIV cDNA integration in vivo. *Nat. Med.* **7:** 631–634.
12. Zennou, V., *et al.* (2000). HIV-1 genome nuclear import is mediated by a central DNA flap. *Cell* **101:** 173–185.
13. Bell, P., Montaner, L. J., and Maul, G. G. (2001). Accumulation and intranuclear distribution of unintegrated human immunodeficiency virus type 1 DNA. *J. Virol.* **75:** 7683–7691.
14. Daniel, R., Katz, R. A., and Skalka, A. M. (1999). A role for DNA-PK in retroviral DNA integration. *Science* **284:** 644–647.
15. Daniel, R., *et al.* (2003). Evidence that the retroviral DNA integration process triggers an ATR-dependent DNA damage response. *Proc. Natl. Acad. Sci. USA* **100:** 4778–4783.
16. Lieber, M. R. (1998). Pathological and physiological double-strand breaks. *Am. J. Pathol.* **153:** 1323–1332.
17. Corbeil, J., *et al.* (2001). Temporal gene regulation during HIV-1 infection of human $CD4^+$ T cells. *Genome Res.* **11:** 1198–204.
18. Vahey, M. T., *et al.* (2002). Impact of viral infection on the gene expression profiles of proliferating normal human peripheral blood mononuclear cells infected with HIV type 1 RF. *AIDS Res. Hum. Retroviruses* **18:** 179–192.
19. van't Wout, A. B., *et al.* (2003). Cellular gene expression upon human immunodeficiency virus type 1 infection of $CD4^+$-T-cell lines. *J. Virol.* **77:** 1392–1402.
20. Ryo, A., *et al.* (1999). Serial analysis of gene expression in HIV-1 infected T cell lines. *FEBS Lett.* **462:** 182–186.
21. Follenzi, A., Ailes, L. E., Bakovic, S., Gueuna, M., and Naldini, L. (2000). Gene transfer by lentiviral vectors is limited by nuclear translocation and rescued by HIV-1 pol sequences. *Nat. Genet.* **25:** 217–222.
22. O'Doherty, U., Swiggard, W. J., and Malim, M. H. (2000). Human immunodeficiency virus type 1 spinoculation enhances infection through virus binding. *J. Virol.* **74:** 10074–10080.
23. Affymetrix (2000). *Affymetrix Microarray Suite User Guide ver. 4*.
24. Irizarry, R. A., et al. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. In preparation.
25. Ihaka, R., and Gentleman, R. (1996). A language for data analysis and graphics. *J. Comput. Graph. Stat.* **5:** 299–314.
26. Wang, Y., Rea, T., Bian, J. G. S., and Sun, Y. (1999). Identification of the genes responsive to etoposide-induced apoptosis: application of DNA chip technology. *FEBS Lett.* **445:** 269–273.
27. Tusher, V. G., Tibshirani, R., and Chu, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. USA* **98:** 5116–5121.
28. Efron, B., Tibshirani, R., Storey, J. D., and Tusher, V. (2001). Empirical Bayes analysis of a microarray experiment. *JASA* **96:** 1151–1160.
29. Efron, B., and Tibshirani, R. (2002). Empirical Bayes methods and false discovery rates for microarrays. *Gen. Epidemiol.* **23:** 70–86.
30. Hartigan, J. A. (1995). Clustering Algorithms. Wiley, New York.
31. Chandra, D., Liu, J. W., and Tang, D. G. (2002). Early mitochondrial activation and cytochrome c up-regulation during apoptosis. *J. Biol. Chem.* **277:** 50842–50854.
32. Lu, H., *et al.* (2002). Metallothionein protects retinal pigment epithelial cells against apoptosis and oxidative stress. *Exp. Eye Res.* **74:** 83–92.
33. Mi, Y., *et al.* (2003). Apoptosis in leukemia cells is accompanied by alterations in the levels and localization of nucleolin. *J. Biol. Chem.* **278:** 8572–8579.
34. Arendt, C. W., and Littman, D. R. (2001). HIV: master of the host cell. *Genome Biol.* **2:** 1030.1–1030.4.
35. Schroder, A., *et al.* (2002). HIV-1 integration in the human genome favors active genes and local hotspots. *Cell* **110:** 521–529.