CMLS, Cell. Mol. Life Sci. 61 (2004) 721–726
1420-682X/04/060721-06
DOI 10.1007/s00018-003-3436-6

CMLS Cellular and Molecular Life Sciences

Research Article

Distribution of HIV-1 in the genomes of AIDS patients

L. Tsyba[a, b], A. V. Rynditch[a, b], E. Boeri[c], K. Jabbari[a] and G. Bernardi[a, d, *]

[a] Laboratoire de Génétique Moléculaire, Institut Jacques Monod, 2 place Jussieu, Tour 43, 75005 Paris (France)
[b] Institute of Molecular Biology and Genetics, Ukrainian Academy of Sciences, 150 Zabolotnogo str., 03143 Kiev (Ukraine)
[c] Diagnostica e Ricerca San Raffaele, I.R.C.C.S. Istituto Scientifico San Raffaele, Via Stamira di Ancona 22, 20127 Milan (Italy)
[d] Laboratorio di Evoluzione Molecolare, Stazione Zoologica Anton Dohrn, Villa Comunale, 80121 Naples (Italy), Fax: +39 081 2455807, e-mail: bernardi@szn.it

Received 21 November 2003; received after revision 13 January 2004; accepted 15 January 2004

Abstract. The localization of HIV-1 proviruses in compositional DNA fractions from 27 AIDS patients during the chronic phase of the disease with depletion of CD4+ and different levels of viremia showed the following. (1) At low viremia, proviruses are predominantly localized in the GC-richest isochores, which are characterized by an open chromatin structure; this result mimics findings on HIV-1 integration in early infected cells in culture. (2) At higher viremia, an increased distribution of proviruses in GC-poor isochores (which match the GC poorness of HIV-1) was found; this suggests a selection of cells in which the 'isopycnic' localization leads to a higher expression of proviruses and, in turn, to higher viremia. (3) At the highest viremia, integrations in GC-rich isochores are often predominant again, but generally not at the same level as in (1); this may be the consequence of new integrations from the extremely abundant RNA copies.

Key words. Virus integration; HIV-1 (human immunodeficiency virus-1); isochore; transcription; AIDS (acquired immunodeficiency syndrome).

Investigations of the integration of retroviral cDNA copies into host chromosomal DNA, a crucial step in retrovirus replication, led to divergent proposals concerning the existence of integration site selection [reviewed in refs 1, 2]. Indeed, while integration of retroviral genomes is a site-specific process as far as the viral sequences are concerned since long terminal repeat (LTR) sequences are involved, for the host genome, site selection for integration has remained an open problem, despite investigations of local features of integration loci, as well as the role of DNA-binding proteins and integrase in site selection [reviewed in ref. 1].

Investigations carried out in our laboratory since the late 1970s [see refs 3, 4 for reviews] led, however, to a major conclusion, namely that stable integration and transcription of all retroviruses tested is predominantly found in isochores matching the composition of the retroviral sequences (a situation that was called 'isopycnic integration'). Isochores are long (>300 kb), compositionally fairly homogeneous DNA segments. The human genome is a mosaic of isochores which are distributed in five families – L1, L2, H1, H2, H3 – characterized by an increasing GC level (see footnote to table 1 for the GC levels). We also found that retroviral genomes belong to two compositional classes: a GC-poor class and a GC-rich class [5]. The first class comprises all lentiviruses and spumaviruses, B-type oncoviruses and D-type oncovi-

* Corresponding author.

Table 1. Localization of HIV-1 proviruses in compositional fractions of the human genome.

HIV-1 RNA (copies/ml)	Sample number	Localization of HIV-1 in DNA fractions, GC%		
		L1 + L2	H1	H2 + H3
<80–4100				
	13587	*36.4*		**50.7**
	12298	*40.1*		**51.5**
	13588	*36.5*	*42.7*	**51.2**
	13584			**51**
	7634	*35*	*42*, **46**	**50.2**
	13591			**51.7**
	11719	*37*	*48.5*	**54.8**
	11723	*38*	*46*	**51**
	12263	*38.5*	*46*	**51**
	12273	*37.1*		**50.7**
	12282	*40.8*		**57**
	12274	*35.2, 40.5*	*44.8*	**49, 53.4**
5000–15,800				
	11713	**34, 40**		*47*, **53**
	13589	**34.4**, *41*		**50**
	7620	**34.4**	*45.2*	**50.7**
	13590	**36**	*45.5*	**51.5**
	13585	**34.2**	*42*	**50**
	7638	**35.2**		**48.8**
	13592	**36**	*45.4*	**52**
63,000–280,000				
	6695	**34.5**	*43.5*	**52.5**
	8903	*37.3*		**49.2, 56.6**
	8340	**36**		**50.2**
	7150	**35.8**	*43.6*	**49.8**
	12299	*36.3*	*43*	**51**
	7721	*38.7*		48.2, **54.3**
	7153	**34**	*43*	**50, 56**
	8336	**36**	*41.5*	**48.5**

In each ultracentrifugation analysis, the highest peaks are indicated in bold and underlined; the peak(s) which correspond(s) to more than 50% of the largest peak are in bold; peaks in italics correspond to less than 30% of the largest peak. L1 (<37% GC), L2 (37–41% GC), and H1 (41–46% GC) are the GC-poor, H2 (46–52% GC), H3 (>52% GC), the GC-rich isochore families of human DNA.

ruses which do not contain oncogenes, like mouse mammary tumor virus (MMTV). The second class includes all the oncoviruses that contain oncogenes, such as Rous sarcoma virus (RSV), and some oncoviruses that do not contain oncogenes, such as bovine leukemia virus (BLV) and human T cell leukemia virus type 1 (HTLV-1), except for those of the B type and for some D-type oncoviruses. The recent addition of 13 retroviral genomes to the set of 32 retroviruses originally explored has confirmed the bimodal compositional pattern of retroviral genomes [4].

Data on the experimental localization of retroviruses in the host genome and compositional analysis of genes located in the neighborhood of integrated retroviral sequences [3] were obtained for GC-rich retroviruses such as BLV, RSV, HTLV-1, murine leukemia viruses (MuLVs) and avian leukemia viruses (ALVs), all of which were found in GC-rich regions of the host genome, whereas GC-poor MMTV was mainly found in GC-poor compartments of the mouse genome [6]. The investigations referred to so far concerned stably integrated sequences. Our analyses indicated that the localization of proviruses in compositionally matching chromosomal environments is associated not only with stability of integration, but also, remarkably, with the transcription of proviral sequences [3].

A recent mapping [7] of 48 HIV-1 integration site sequences after a short (14 h) experimental infection of the human T cell line Sup T1 [8] showed, however, that 54% of them were located in genes and 69% in transcribed sequences. As far as the isochore localization is concerned, while the HIV-1 integrations were mainly found in isochores L2, H1 and H2, their highest density (and, therefore, their preference for integration) is in the 'genome core,' the gene-richest and GC-richest isochores H2 and H3 [see ref. 4]. This is a most interesting result in view of the fact that it indicates a non-isopycnic localization of HIV-1, a retrovirus which has a GC-poor genome.

The conclusions of Elleder et al. [7] have been confirmed by a more extensive study of 524 HIV-1 initial integration sites in the human lymphoid cell line [9]. Indeed, the gene-dense regions in which the preferential localization

of these early integration sites was found to coincide (although apparently not noticed by the authors) with the GC-richest isochores, as shown by our previous investigations indicating that a strong gradient of gene density parallels the gradient of isochore GC levels [10].

An interesting aspect of the work of Schröder et al. [9] was that these authors not only investigated HIV-1 integration in vivo, but also in vitro. Using naked genomic DNA from Sup T1 cells as an integration target and HIV-1 preintegration complexes, they concluded that, in this case, the integration pattern did not show the in vivo specificity for genes. Incidentally, HIV-1 preintegration complexes are replication intermediates that can be isolated from infected cells and which contain the viral cDNA, integrase and other viral and cellular proteins.

Additional definition of initial integration site preferences of HIV-1 performed by mapping of 244 integrations of infectious HIV-1 in the human H9 cell line and 135 integrations of recombinant pseudotyped HIV-1 in HeLa cells showed that approximately 60% of them localized in the transcriptional units, but not upstream of the transcriptional start whereas MLV integrations favored the transcription start regions [11].

In the present work, we show that the compositional pattern of proviral integration changes during the course of infection in AIDS patients, and that such changes are correlated with viremia levels. Finally, we propose an explanation for such changes.

Materials and methods

Viral load determination

RNA from plasma samples was extracted by silica adsorption and the number of viral RNA copies was calculated by NASBA isothermal amplification [12], with a sensitivity of 80 copies/ml.

Compositional fractionation

Compositional fractionation of high-molecular-weight DNA from peripheral blood mononuclear cells was performed in preparative shallow CsCl gradients in the presence of buoyant density markers using vertical rotors [13]. The HIV-1 proviral sequences in DNA compositional fractions were detected by slot-blot hybridization with a pBH10-R3 plasmid HIV-1 insert [14] under stringent conditions. The hybridization signals were evaluated with a Phosphor Imager. The modal buoyant density of the DNA molecules hybridizing the probe (as determined by comparison with buoyant density DNA markers) was used to estimate the GC level of DNA in each fraction. Gaussian curves of distribution of proviruses and markers in the compositional fractions of the human genome were obtained using a program based on IgorPro (from WaveMetrics Inc., Lake Oswego, Ore.).

Computational compositional analysis of integration sites

The compositional distribution of HIV-1 integration sites was assessed by the Blast program in a human lymphoid cell line (in vivo) and in naked human DNA (in vitro), based on the sequences determined by Schröder et al. [9] and assessments of the GC level of 100-kb human genome regions around integration sites.

Results and discussion

Localization of HIV-1 proviruses in the genomes of AIDS patients

We localized HIV-1 proviruses in mononuclear cells from peripheral blood isolated from 27 HIV-1-infected individuals in the chronic phase of AIDS. The patients were characterized by CD4+ T lymphocyte depletion (less than 500 lymphocytes/mm^3) and by concentrations of HIV-1 RNA varying from 80 to 280,000 copies/ml of blood. The patients were divided into three groups according to their levels of viremia: low viremia (80–4100 copies/ml), medium viremia (5000–15,800 copies/ml) and high viremia (63,000–280,000 copies/ml).

Hybridization analysis with an HIV-1 probe on DNA compositional fractions from the HIV-1-infected patients in these three groups revealed that proviral copies showed peaks of distribution in different isochore families (fig. 1). The crucial observation made by studying the 27 cases under examination was that a correlation was found between the distribution of the provirus in the genome of patients and their viremia levels.

Indeed, in the first group of patients, characterized by low viremia, the major distribution peaks were located in the H2/H3 isochores (fig. 1A). GC-poor fractions of L1-L2 isochores comprised less than 20% of total provirus. In two cases, proviruses were not even found in L1-L2 isochores. Moreover, the relatively rare intermediate peaks located in several cases in H1 isochores only comprised small percentages of proviruses.

The situation was different in the second group of patients with intermediate levels of viremia (fig. 1B). Here, the proviruses localized in the GC-poor DNA fractions reached levels higher than one-half of those of GC-rich DNA fractions, in a few cases even reaching similar levels.

Finally, in the third group (fig. 1C), the distribution of proviruses was either similar to that of the second group or was characterized by a lower abundance of proviruses in GC-poor compared to GC-rich fractions. As in both previous cases, proviruses localized in intermediate GC level fractions represented a minority of the total integrated sequences.

Figure 1 shows representative situations as found in the first, second and third groups. The figure indicates that the evolution of the integration pattern correlates with

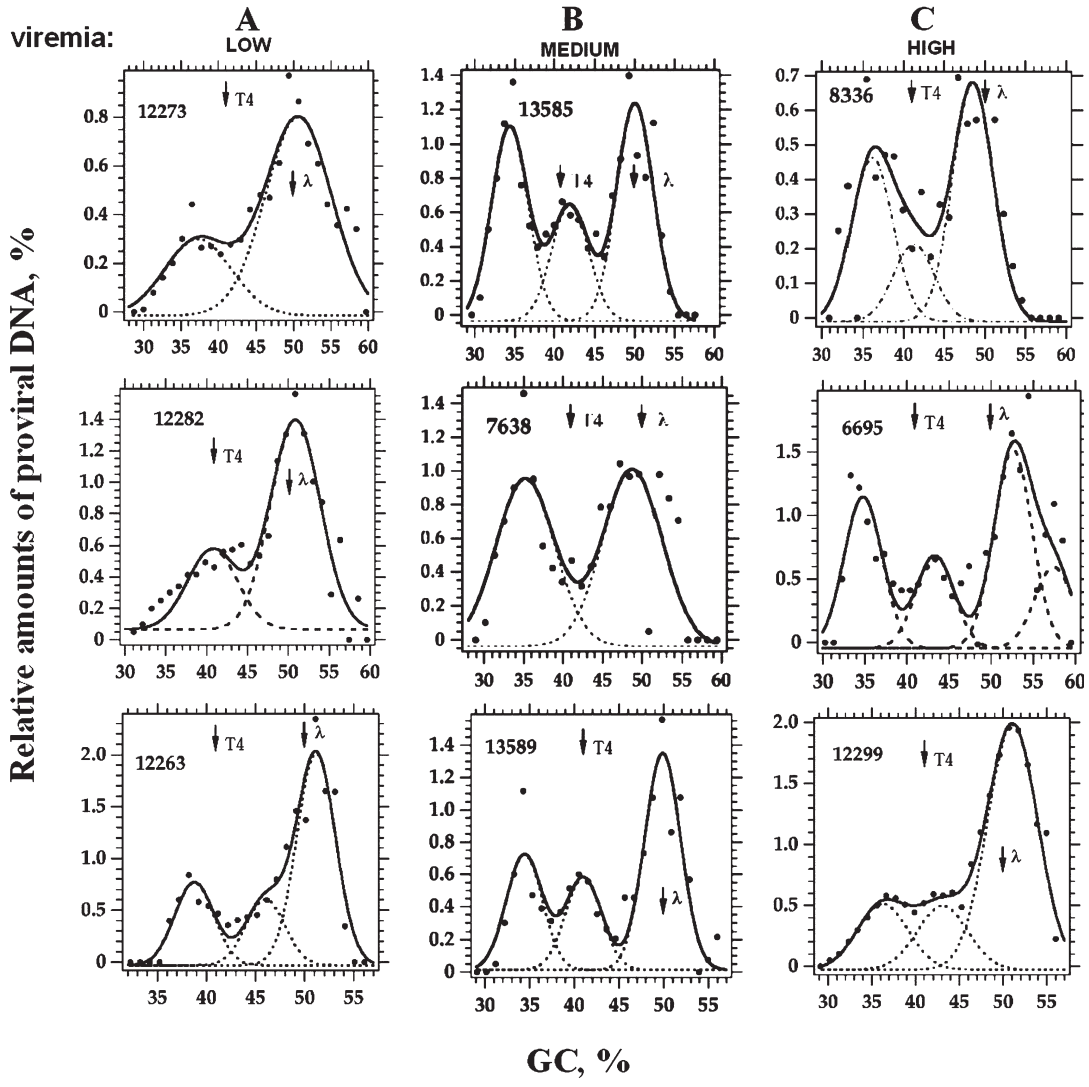

Figure 1. Hybridization of HIV-1 probe on DNAs from HIV-1 infected patients showing an increasing production (from left to right, *A*–*C*) of virus-specific RNA copies (see text). DNAs from the patients listed in table 1 (sample numbers are indicated) were centrifuged in shallow CsCl gradients (see Materials and methods). T4 and *λ* indicate the positions of these phage density markers. Profiles of the other samples listed in table 1 are available upon request. The relative hybridization levels can be judged from table 1.

viremia levels. In the chronic phase of AIDS, a low level of viremia is accompanied by an initial preferential integration in GC-rich regions of the genome, which are characterized by an open chromatin organization [3, 4, 15], whereas higher viremia, accompanied by increased expression, is possibly due to selection of cells with higher integration levels in GC-poor isochores. Finally, the extremely high viremia levels are possibly due to an abundance of novel integrations, that preferentially occur in GC-rich isochores.

Control experiments

Several controls were done to exclude the possibility of artifacts, such as the presence of unintegrated HIV-1 cDNA in the samples of high-molecular-weight genomic DNA, or the influence of HIV-related endogenous viral sequences on the determination of integrated HIV-1 proviruses in compositional fractions of human DNA.

These experiments showed that the existence of HIV-1-specific sequences in the GC-poor compartments of the human genome could not be explained by the presence of unintegrated circular HIV-1 proviral forms. Indeed, such forms were not detected by PCR with primers that cover the circle junction formed by ligation of the U5 and U3 regions. Moreover, unintegrated forms were not detected by Southern blot analysis in high-molecular-weight DNA samples when the DNA load was equal to the amount of DNA used in compositional fractionation.

At the same time, endogenous EHS-1 sequences homologous to the HIV-1 *env* gene [16] and EHS-2 sequences homologous to Rev and gp41 HIV-specific sequences were detected in DNA fractions from uninfected individ-

uals with GC levels equal to 39% and 41%, respectively. This suggested that HIV-specific sequences observed in the GC-rich part of the human genome were not endogenous HIV-like sequences. The results indicated also that GC-poor endogenous sequences homologous to HIV-1 localized in compositionally matching chromosomal environments. This is in agreement with previous findings for endogenous MMTV sequences [6].

An analysis of the data of Schröder et al. [9]

We also assessed the compositional distribution of HIV-1 integration sites in the human lymphoid cell line and in naked human DNA based on the sequences determined by Schröder et al. [9]. The data revealed that the highest density of integration sites are in the GC-richest regions corresponding to H2 (46–52% GC) and H3 (>52% GC) isochore families, in both cases with a higher in vivo integration density compared to in vitro (fig. 2). Since GC-rich regions contain high concentration of genes, these results indicate a preferred integration of HIV-1 into gene-rich regions. The direct determination of integration sites showed that integration is targeting genes in vivo [7, 9] but not in vitro [9]. More recent data indicate that only transcribed regions of genes are favored targets for in vivo HIV-1 integration [11] which could explain this difference. The latter suggested that some local features of GC-rich regions of the human genome are favorable per se for targeting of HIV-integration complexes. The influence of primary sequences on HIV-1 target site selection such as presence of weak consensus sequences in the integration junctions has been reported previously [8, 17].

The higher density of in vivo integration sites in GC-rich regions in comparison with that in vitro revealed that the chromatin environment in H2 and H3 isochore families, which is characterized by open structures [15], is important for primary integration. HIV-1 integration in GC-rich genomic regions has been clearly demonstrated in all cases. However, the profiles of HIV-1 localization in the DNA from the infected individuals with CD4+ T lymphocyte depletion indicated that higher levels of viremia correlated with localization of HIV-1 in GC-poorer regions. This in turn suggested a selection of cells with HIV-1 proviruses integrated in GC-poor compartments in the course of HIV-1-dependent disease. The correlation between localization of integrates in GC-poor isochore families and a high level of viremia suggested that the clonal chromosomal environment in these regions was important to support HIV-1 transcription. Heterogeneity of HIV-1 expression levels in distinct clones as a result of different integration sites was observed [18–20].

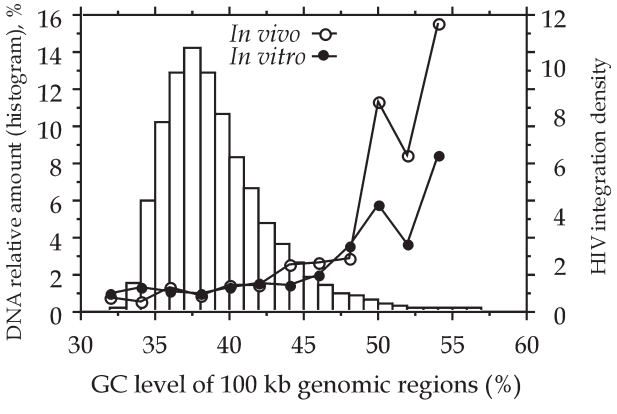

Figure 2. Distribution of 500 in vivo and 110 in vitro HIV integration sites in the human genome [based on the data from ref. 9]. The curves represent the integration density (% integrated viral sequences/% corresponding DNA fractions).

A high level of HIV-1 replication is usually found at 42 h after infection in cell lines and primary CD4+ T cells [21], when the majority of integration sites are located in GC-rich regions. In addition, the preferential integration of HIV detected initially in transcriptionally active genes suggested that such regions are important for efficient expression of the HIV genome [21]. This is not in contradiction with the localization of highly replicated HIV-1 proviruses in the GC-poor part of the genome because (i) active genes are also located in GC-poor isochore families, but at a lower frequency and (ii) primary integration of HIV-1 in GC-poor regions is also observed but less frequently.

These findings are consistent with data showing that highly expressed HIV-1 proviruses could be localized in GC-poor regions of the human genome. In the majority of cases investigated, the localization of integrated proviruses and repeated elements in a matching compositional environment (designated as 'isopycnicity') has been associated with their stability under such conditions [3, 4]. Both stability and chromosomal environment favorable for HIV-1 replication could be factors that contribute to the selection of infected cells in the course of HIV-1-dependent disease.

In conclusion, in the present work we demonstrated that a pattern of integration sites involving GC-rich isochores characterizes not only the initial distribution of HIV-1 proviruses in cell lines, but could also be observed in peripheral blood cells of HIV-infected individuals in a chronic phase of AIDS with a low level of CD4+ T cells and low viremia. This period is followed by an increase in virus production [22].

We also demonstrated the existence of changes in HIV-1 localization in the host genome during the disease and its correlation with viremia levels. Indeed, the increasing production of virus-specific RNA copies is correlated with provirus localization. Representative samples show that low viremia levels are accompanied by a predominance of proviruses in the GC-rich isochores (fig. 1A), that an intermediate viremia level is characterized by a localization in GC-poor isochores which is as important as that in GC-rich isochores (fig. 1B), and that localization

in GC-rich isochores is again predominant at the highest viremia levels (fig. 1 C), although not to the extent of the situation found at low viremia levels. We interpret these results as indicating that production of virus-specific RNA copies increases from the first to the second situation due to selection of the cells with increased transcription for the proviruses located in the compositionally matching GC-poor environment, whereas the last situation is the result of novel integrations from the abundant viral RNA produced.

Since the situation in figure 1 A mimics that of early infected cells, in that proviruses are mainly in GC-rich isochores, and the end phase of the disease is characterized by very high viremia, one is tempted to suggest that the three situations described in this paper may follow each other during the time course of the AIDS disease.

1 Holmes-Son M. L., Appa R. S. and Chow S. A. (2001) Molecular genetics and target site specificity of retroviral integration. Adv. Genet. **43:** 33–69

2 Brown P. O. (1997) Integration. In: Retroviruses, pp. 161–205, Coffin J. M., Hughes S. H. and Varmus H. E. (eds), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

3 Rynditch A. V., Zoubak S., Tsyba L., Tryapitsina-Guley N. and Bernardi G. (1998) The regional integration of retroviral sequences into the mosaic genomes of mammals. Gene **221:** 1–16

4 Bernardi G. (2004). Structural and Evolutionary Genomics: Natural Selection in Genome Evolution. Elsevier, Amsterdam

5 Zoubak S., Rynditch A. and Bernardi G. (1992) Compositional bimodality and evolution of retroviral genomes. Gene **119:** 207–213

6 Salinas J., Zerial M., Filipski J., Crepin M. and Bernardi G. (1987). Non-random distribution of MMTV proviral sequences in the mouse genome. Nucleic Acids Res. **15:** 3009–3022

7 Elleder D., Pavliček A., Pačes I. and Heinar I. (2002) Preferential integration of human immunodeficiency virus type 1 into genes, cytogenetic R bands and GC-rich DNA regions: insight from the human genome sequence. FEBS Lett. **517:** 285–286

8 Carteau S., Hoffmann C. and Bushman F. (1998) Chromosome structure and human immunodeficiency virus type 1 cDNA integration: centromeric alphoid repeats are a disfavored target. J. Virol. **72:** 4005–4014

9 Schröder A. R. W., Shinn P., Chen H., Berry C., Ecker J. R. and Bushman F. (2002) HIV-1 integration in the human genome favors active genes and local hotspots. Cell **110:** 521–529

10 Zoubak S., Clay O. and Bernardi G. (1996) The gene distribution of the human genome. Gene **174:** 95–102

11 Wu X., Li Y., Crise B. and Burgess S. M. (2003) Transcription start regions in the human genome are favored targets for MLV integration. Science **300:** 1749–1751

12 Gemen B. van, Beuningen R. van, Nabbe A., Strijp D. van, Jurriaans S., Lens P. et al. (1994) A one-tube quantitative HIV-1 RNA NASBA nucleic acid amplification assay using electrochemiluminescent (ECL) labelled probes. J. Virol. Methods **49:** 157–168

13 De Sario A., Geigl E.-M. and Bernardi G. (1995) A rapid procedure for the compositional analysis of yeast artificial chromosomes. Nucleic Acids Res. **23:** 4013–4014

14 Fisher A. G., Collalti E., Ratner L., Gallo R. C. and Wong-Staal F. (1985) A molecular clone of HTLV-III with biological activity. Nature **316:** 262–265

15 Saccone S., Federico C., Andreozzi L., D'Antoni S. and Bernardi G. (2002) Localization of the gene-richest and the gene-poorest isochores in the interphase nuclei of mammals and birds. Gene **300:** 169–178

16 Horwitz M. S., Boyce-Jacino M. T. and Faras A. J. (1992) Novel human endogenous sequences related to human immunodeficiency virus type 1. J. Virol. **66:** 2170–2179

17 Stevens S. W. and Griffith J. D. (1996) Sequence analysis of the human DNA flanking sites of human immunodeficiency virus type 1 integration. J. Virol. **70:** 6459–6462

18 Winslow B. J., Pomerantz R. J., Bagasra O. and Trono D. (1993) HIV-1 latency due to the site of proviral integration. Virology **196:** 849–854

19 Jordan A., Defechereux P. and Verdin, E. (2001) The site of HIV-1 integration in the human genome determines basal transcriptional activity and response to Tat transactivation. EMBO J. **20:** 1726–1738

20 Stevenson M., Stanwick T. L., Dempsey M. P. and Lamonica C. A. (1990) HIV-1 replication is controlled at the level of T cell activation and proviral integration. EMBO J. **9:** 1551–1560

21 Corbeil J., Sheeter D., Genini D., Rought S., Leoni L., Du P. et al. (2001) Temporal gene regulation during HIV-1 infection of human CD4+ T cells. Genome Res. **11:** 1198–1204

22 Baltimore D. and Feinberg M. (1989) HIV revealed: toward a natural history of the infection. N. Engl. J. Med. **321:** 1673–1675