

Integration site preferences of endogenous retroviruses

Domenica Taruscio and Laura Manuelidis

Yale Medical School, 333 Cedar Street, New Haven, CT 06510, USA

Received April 23, 1991 / in revised form May 21, 1991

Accepted June 3, 1991 by P.B. Moens

Abstract. Retroviruses have the ability to integrate into the genome of their host, in many cases with little apparent sequence or site specificity. However, relatively few studies have addressed more general features of chromosomal integration. In this study we directly visualized the chromosomal organization of three representative endogenous retroviruses by *in situ* hybridization. Because there are 50–1000 copies of each of these retroviruses in the genome, it was possible to evaluate repeated integration events. Each retroviral sequence exhibited a unique and markedly different integration pattern. In order to characterize more precisely the chromosomal domains targeted by each retrovirus, later replicating domains were differentially labeled. Additionally, prototypic SINES and LINES (short and long interspersed reiterated sequences), which are inhomogeneously distributed on chromosome arms, were simultaneously detected. Retroviral copies of ≥ 2 kb were found (i) exclusively in a discrete set of later replicating domains, most of which have the staining characteristics of constitutive heterochromatin, (ii) widely represented in disparate types of chromosome domains, or (iii) almost completely confined to CpG Alu-rich regions that are known to be early replicating. Retroviral elements in Alu-rich domains would be expected to be actively transcribed in all cells. Surprisingly, hybridization to blots of brain RNA showed an ~ 25 fold lower level of transcripts from these Alu associated elements than from retroviral sequences restricted to later replicating, heterochromatic domains. Retroviral insertions may subvert more typical transcriptional characteristics of a domain. The present results indicate that there are highly specific integration patterns for each endogenous retrovirus that do not readily relate to their sequence or particle classification. Each host genome may utilize these elements for contrary, and possibly beneficial functions.

Introduction

Retroviruses have the capacity to reverse transcribe their RNA genome into DNA. This retroviral DNA can integrate into host chromosomes and act as a stable genetic element. In acute retroviral infections, susceptible somatic cells often produce viral particles and exhibit pathological sequelae. The demise of an infected cell can be relatively rapid. On the other hand, retroviral integration itself can be deleterious, even when the production of infectious particles is inefficient. There are numerous examples of retroviral pathology associated with insertional mutagenesis, chromosomal breakage and oncogene activation (reviewed in Weiss et al. 1985). Often insertional changes initiate a series of progressive events with consequences as diverse as tumor formation and slow degenerative changes. Infection with HTLV-1 for example, can lead to leukemia as well as to the degenerative CNS myelopathy of tropical spastic paraparesis (Vernant et al. 1987). Such sequelae may not be apparent for many years. During this time an infectious retrovirus has the opportunity to integrate into germ line cells. When this occurs, the provirus can be propagated as a stable Mendelian gene in progeny. These endogenous retroviruses have been described in many vertebrate species, and at least some of them appear to have initially integrated into the germ line cells of their host over 1 million years ago.

In recent years it has become apparent that many different classes of endogenous retroviruses are present in each species. Endogenous retroviruses may be classified according to their nucleic acid sequence, or by the type of particle with which they are associated. It is intriguing that in the genome of a given species, some of these proviruses are present only as a single copy, whereas others are represented at intermediate copy levels (of 30–300 members), or at very high copy levels of 1000 or more homologous elements. Multiplicity of genomic copies is not readily correlated with either a

particular sequence or particle classification. For example, endogenous retroviruses of similar sequence, present at ~1000 copies per haploid genome in mouse and Syrian hamster cells appear to relate to intracisternal particles (IAP) of an A type or R type morphology respectively (Kuff and Lueders 1988). In humans, a C-type retroviral sequence is represented ~1000 times in the haploid genome (Fraser et al. 1988), whereas SHIA-like elements are represented only 50 times in the human genome (Ono et al. 1986). Furthermore, at least three other human C-type endogenous proviruses are present at only one copy per haploid genome (Leib-Mosch et al. 1990). At issue is why certain retroviruses attain high copy numbers, whereas others are restricted to low or single copy numbers. Although the deleterious effects of retroviral integration have been the focus of many investigations, the amplification of selected proviruses may indicate a beneficial function for the host species. This view has rarely been considered.

If an endogenous retrovirus assumes some beneficial function, it might be organized in the genome in a unique way. One attribute of proviruses that is not entirely understood is the potential to integrate selectively into preferred chromosomal sites. Initial restriction mapping and limited sequence data had failed to reveal proviral integration preferences in acute infections (e.g. Dhar et al. 1980; Panganiban 1985). However, more recent experiments detailing DNase I-hypersensitive sites (Rohdewohld et al. 1987), the transcriptional activity of flanking DNA (Mooslehner et al. 1990) and sequence analysis of independent integration events (Shih et al. 1988) have verified proviral insertions that are non-random. These investigators have suggested that higher order structural features, or an active transcriptional capacity of chromatin may be fundamental for preferred integration events. Nonetheless, relatively few studies directly address more general macromolecular and chromosomal levels of retroviral integration. We sought to visualize directly these higher levels of organization. Such studies can complement more detailed sequence analyses of preferred integration sites.

Most proviral experiments by necessity assay cell systems where viral integration is reasonably rapid, and often limited to a single or very few insertion sites. In contrast, the evolution of endogenous retroviruses often entails multiple insertions and amplifications. This multiplicity allows one to assess features of organization that are more global in nature. One can reasonably determine if a particular set of chromosomal features are targeted by all retroviruses. A more detailed knowledge of the preferred chromosomal sites targeted by different endogenous retroviruses could additionally clarify the functional capacity, as well as the dynamic role of these viruses in shaping the genome. In this study we reexamined the chromosomal position of two homologous high copy endogenous rodent retroviruses (associated with intracisternal A and R particles; Kuff and Lueders 1988) using high resolution in situ hybridization. Additionally we mapped the chromosomal positions for a moderate copy number C-type human full length element (Repaske et al. 1985) that has not been previously

studied by in situ hybridization. A 2.5 kb region of the single copy, tissue specific gene for glial fibrillary acidic protein (GFAP; Lewis et al. 1984) was also mapped. This control confirmed our ability to reproducibly detect elements of >2 kb, a size that would include many truncated proviral elements. Finally, in order to characterize further the chromosomal domains targeted by each provirus we (i) labeled later replicating domains with bromodeoxyuridine (BrdU; Manuelidis and Borden 1988) for simultaneous detection with each retroviral probe, and/or (ii) simultaneously evaluated the banding pattern of prototypic SINES and LINES (short and long interspersed reiterated sequences). These repeated sequences are preferentially concentrated in band size domains with different functional attributes (reviewed in Manuelidis 1990). Remarkably, each of the three retroviral elements displayed very different patterns of chromosomal integration. Each of these patterns may be the consequence of evolving host control mechanisms.

Materials and methods

Cell cultures, BrdU pulse labeling and preparation of metaphase spreads. A Syrian hamster cell line (TC774) derived from normal adult brain that spontaneously transformed in vitro was propagated as described (Manuelidis EE et al. 1987). A methylcholanthrene induced glioblastoma cell line from an NIH Swiss mouse producing numerous C-type particles (TC509; Manuelidis and Manuelidis 1976) and normal *Mus domesticus* (B6C3F1/J) spleen cells in short term culture (Boyle et al. 1990) were compared. Short term cultures of normal phytohemagglutinin stimulated human lymphocytes were grown as described (Cremer et al. 1988). In experiments using replication banding, log phase cells were treated with 10 µg/ml BrdU and 0.5 µg/ml fluorodeoxyuridine (Manuelidis and Borden 1988). The optimal BrdU exposure time to produce positive late replication bands was determined empirically for each cell line (range 6–8 h). Cells were treated with 4×10^{-7} M colchicine for 30 min prior to collection, swelling and acid-methanol fixation of metaphase spreads.

Preparation and source of probes. Plasmid DNAs were purified as described (Maniatis et al. 1982), and DNAs were labeled with 11-dUTP ligands by nick translation (Manuelidis and Ward 1984) using DNAase I concentrations sufficient to result in an average probe size of ~250 bp for efficient penetration (Manuelidis 1985) as determined by agarose gel electrophoresis. Either biotin 11-dUTP (Sigma), or digoxigenin 11-dUTP (diluted 1:3 with dTTP as recommended by Boehringer) were incorporated during a standard 2 h reaction at 15° C. Labeled DNA was separated from unincorporated nucleotides in G-50 spun columns equilibrated in 50 mM Tris-Cl, 2 mM EDTA, 0.1% SDS, pH 8.0. Incorporation was considered adequate if 0.5 pg of DNA was visualized after development of nitrocellulose dot blots using alkaline phosphatase digoxigenin antibodies (Boehringer) or streptavidin-alkaline phosphatase (BRL) to detect the labeled DNAs.

Two plasmids containing the Syrian hamster endogenous retroviral sequence from the 5' and 3' halves of the gene (pSHIA 12-1 and 12-2) were studied alone or in combination and were provided by E. Kuff (Lueders and Kuff 1983). A plasmid containing a comparable mouse 7.1 kb endogenous retroviral sequence (pMIA14; Mietz et al. 1987) was also a gift of E. Kuff. The full-length cloned 8.8 kb human endogenous retroviral sequence (HERV 4.1; Rabson et al. 1985) was donated by A. Rabson. A plasmid containing 2.5 kb of mouse GFAP (provided by N. Cowan; Lewis et al. 1984) was used as a single copy control to demonstrate detection limits under standard conditions. Additionally, a cloned mouse L1 repeat

(clone KS13A from T. Fanning, NIH) and Alu-PCR (polymerase chain reaction) products from human DNA were used in two-color hybridization studies to assess more precisely the position of retroviral sequences relative to a reproducible chromosomal banding pattern.

Alu-PCR products of 0.3–3-kb were generated from high molecular weight human DNA using 1 μ M of primer no. 517 (CGACCTCGAGATCT(C/T)(G/A)GCTCACTGCAA; Nelson et al. 1989) and an annealing temperature of 55° C in a 35 cycle reaction with Taq polymerase and buffer (Cetus kit). The PCR reaction was modified to include 2.5 mM MgCl₂ (Baldini and Ward 1991). Prior to nick translation, PCR products were ethanol precipitated in the presence of 2 M ammonium acetate and dissolved at appropriate concentrations for nick translation.

In situ hybridization. Chromosome spreads were denatured in 70% formamide, 2 \times SSC at 70° C, dehydrated in increasing concentrations of ethanol (70%–100%), air-dried and prewarmed to 42° C just prior to hybridization. (1 \times SSC is 0.15 M NaCl, 0.015 M sodium citrate.) Biotinylated rodent retroviral probes (SHIA and MIA) were used at a final concentration of 5 μ g/ml; GFAP was used at 2.5 and 5 μ g/ml (with equivalent results), and HERV 4.1 was used at 10 μ g/ml hybridization solution. Carrier salmon sperm DNA was included to yield a final total DNA concentration of 700 μ g/ml. Appropriate quantities of probe and carrier DNA were ethanol precipitated, vacuum dried and resolubilized in hybridization solution [final concentration of 50% formamide (v/v), 2 \times SSC and 10% dextran sulfate]. This mixture was denatured at 75° C for 5 min, applied to slides (2.5 μ l/cm²), and sealed with coverslips/rubber cement for hybridization at 37°–38° C for 10–16 h.

In some experiments mouse genomic competitor DNA (200 μ g/ml) was included to reduce non-specific background in GFAP hybridizations. In this case, the heated hybridization mixture was partially reannealed (for 10 min at 37° C) before application to slides (Lichter et al. 1988). In double labeling experiments with biotinylated GFAP and digoxigenin labeled L1 sequences, it was most convenient to mix equal parts of two separate hybridization mixtures; partially renatured mouse genomic and GFAP DNA were mixed with freshly denatured L1 on ice to yield a final concentration of 2.5 μ g/ml GFAP and 5 μ g/ml of the L1 repeat. In double hybridizations of MIA with L1, or of HERV 4.1 with Alu-PCR products, it was not necessary to include competitor genomic DNA; a single hybridization mixture containing both probes was completely denatured. Digoxigenin labeled Alu-PCR products were used at 6 μ g/ml. In experiments for the simultaneous detection of a biotinylated probe and metabolically incorporated BrdU, dextran sulfate was sometimes omitted to improve the morphology of BrdU banding. In this case, chromosomes were denatured together with probe(s) in formamide-SSC as previously described (Manuelidis 1985; Manuelidis and Borden 1988).

Detection of labeled probes. Post-hybridization washes were done in 50% formamide, 2 \times SSC at 42° C (3 \times 5 min each) and in 0.1 \times SSC at 60° C (3 \times 5 min each). Slides were then blocked in 3% bovine serum albumin (BSA) in 4 \times SSC. Biotinylated probes were detected with either 5 μ g/ml Avidin DCS-FITC (fluorescein isothiocyanate) or with 20 μ g/ml Avidin DCS-rhodamine (Vector Labs.) in detection buffer (4 \times SSC, 0.1% Tween-20 and 1% BSA) for 30 min at 37° C. Each double labeling experiment was designed to discriminate each of two DNA labels with either green (FITC) or red (rhodamine) fluorescence in chromosomes counterstained with the blue fluorescent dye 4',6-diamidino-2-phenylindole (DAPI) (e.g. Fig. 2c–e). For simplicity, we detail a representative combination (green for probe, red for banding). However, these results were confirmed in reverse color combination experiments with comparable secondary detector reagents (e.g. Fig. 3a–c). Additionally it should be noted that very late replicating centromeric regions are often not visualized in BrdU/hybridization studies. These satellite DNA regions probably renature during hybridization because they bind BrdU antibodies just after denaturation, but lose their ability to bind these single strand requiring antibodies after *in situ* hybridization.

In double-probe labeling experiments for biotinylated probe and metabolically labeled BrdU substituted DNA, a primary antibody incubation in 1:700 dilution of mouse monoclonal anti-BrdU (Partec) in detection buffer was followed by secondary detection with a 1:700 dilution of goat anti-mouse IgG conjugated to rhodamine (Jackson Labs.) combined with 5 μ g/ml Avidin DCS-FITC to detect the biotin residues. For simultaneous detection of biotin and digoxigenin labeled probes we generally used a single step detection procedure with a combination of rhodamine labeled anti-digoxigenin Fab fragments (1:100, Boehringer) and avidin-FITC in detector buffer. Detector reagents were allowed to bind for 30–60 min at 37° C. Slides were washed in detector buffer without BSA (3 \times 5 min each) after each detector step. DAPI staining and mounting in antifade solution were done as described (Cremer et al. 1988).

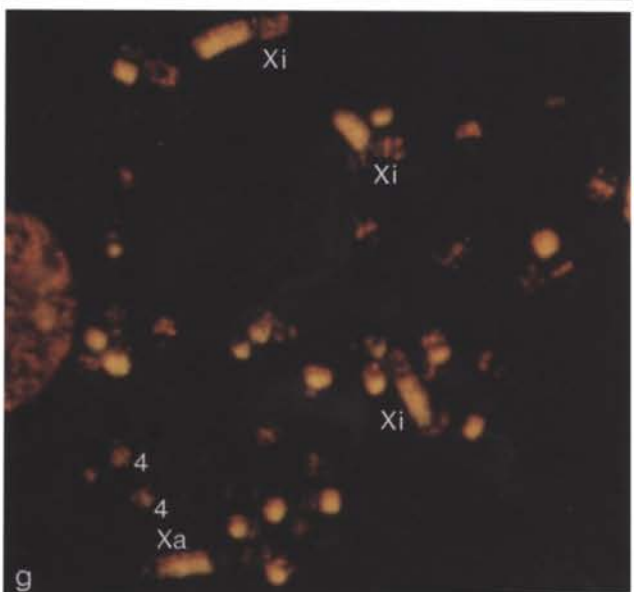
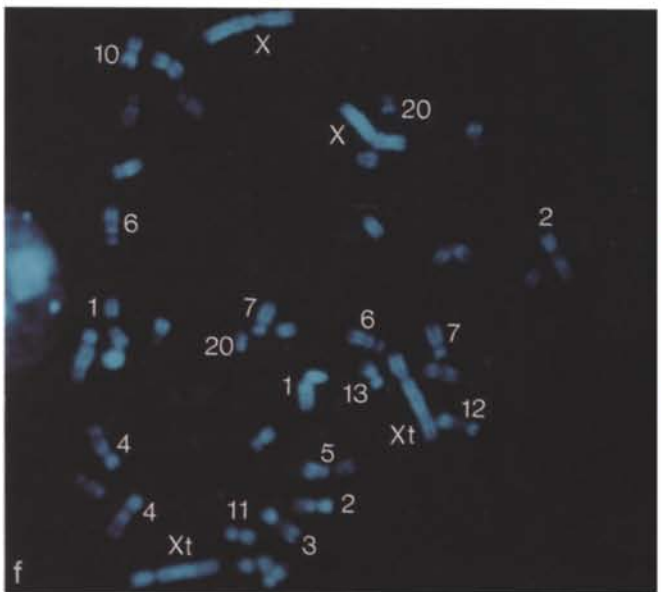
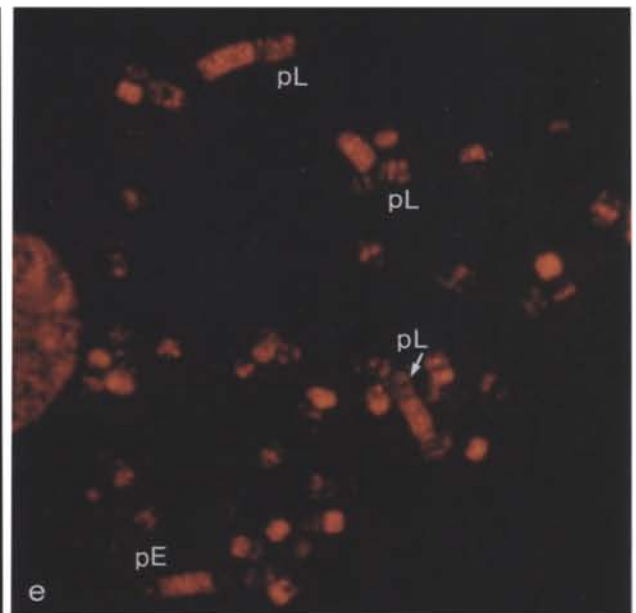
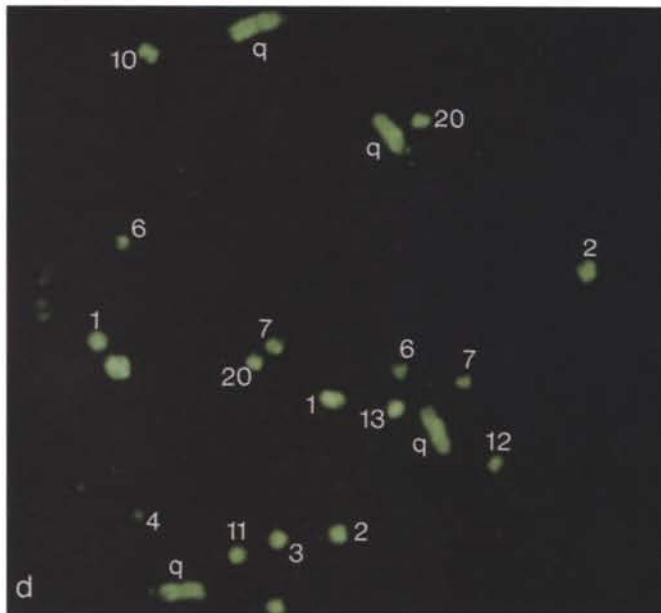
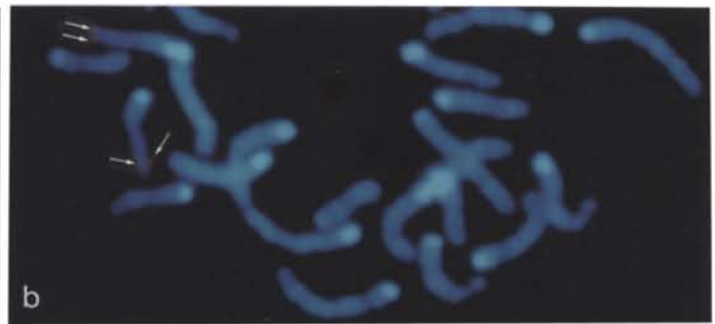
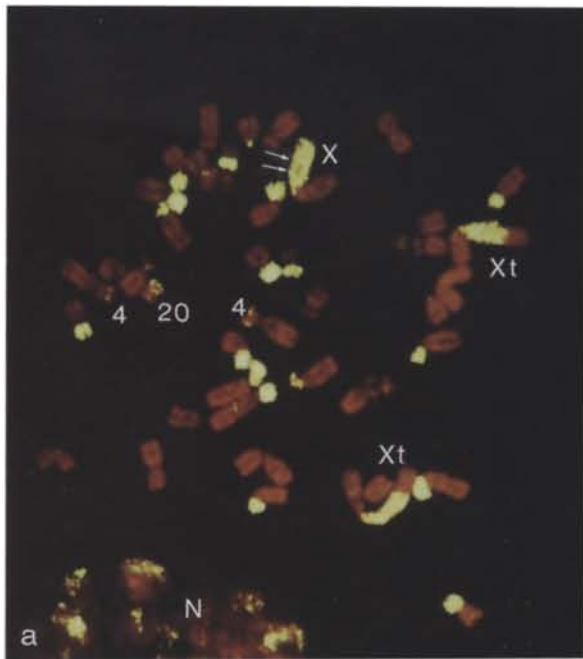
Standard filter set combinations (from Zeiss or Nikon) for FITC, rhodamine and DAPI fluorescence were used to view spreads by conventional epifluorescence microscopy. Sets were chosen to minimize overlap of different color fluorescent signals, and data was verified in reverse color experiments. Additionally, some preparations were examined by confocal microscopy (BioRad MRC-500 scanner with a filter system optimized to obtain separate images for each fluorochrome). In these studies the z-plane with the highest density of fluorescent signals was recorded, but it should be noted that slightly out of focus signals (at adjacent z-planes) were lost, and it was not possible to image DAPI fluorescence in this set-up. A cooled Photonics CCD camera mounted on a Nikon microscope and attached to a Macintosh IIX computer was used to digitize blue, green and red emitting fluorochrome separately. Each of these grey scale images was then pseudocolored and merged as described (Boyle et al. 1990). Most merged images overlapped quite well, and no corrections were made for occasional slightly shifted images.

Isolation of total brain RNA and RNA blot hybridization. Fresh normal adult hamster cerebrum and human cortex were lysed in guanidinium isothiocyanate solution, and RNA was isolated by CsCl gradient centrifugation (Chirgwin et al. 1979). Normal human cortex (Borden and Manuelidis 1988) was submerged in lysing solution in the operating room to minimize RNA degradation. Each purified RNA sample (10 μ g) was resolved on a 0.8% glyoxal denaturing agarose gel (Maniatis et al. 1982), and electrophoretically eluted on to Zeta Probe membranes as described (Manuelidis et al. 1987). Plasmid probes were linearized with appropriate restriction enzymes, labeled with ³²P by random priming (Feinberg and Vogelstein 1983) and hybridized and washed as described (Manuelidis et al. 1987). The stringency was comparable to that used for *in situ* hybridization. Autoradiography was done at –70° C with an intensifying screen. Both RNA preparations, assessed with the GFAP probe, showed a single band, negligible breakdown of RNA and confirmed equivalent gel loads of each RNA sample. λ HindIII marker DNA fragment labeled with ³²P using kinase were used to check elution efficiencies over a broad size range on membranes. Autoradiographs were digitized using a Dage-MTI camera attached to a Raster-Ops digitizing board in a Macintosh II computer, and were evaluated (along with some CCD acquired fluorescent chromosome signals) using the NIH image program of Wayne Rastband (public domain software).

Results

Syrian hamster proviral sequences are confined to later-replicating heterochromatin domains

Studies with either the 5' or 3' halves of SHIA displayed the same chromosomal positions, and thus only hybridizations with both halves of the probe are demonstrated here. Figure 1a shows the typical hybridization pattern



of the SHIA probe to chromosomes from a transformed hamster brain cell line, TC774. The 950 SHIA copies (Kuff and Lueders 1988) were confined to only 13 of the 44 hamster chromosomes in this cell line. Furthermore, SHIA sites were highly clustered on only specific portions of these chromosomes. In most cases the p arms were selectively decorated. Evaluation of multiple spreads, counterstained with DAPI for chromosome identification (as in Fig. 1d and f) revealed a high concentration of SHIA sequences in the p arms of chromosomes 1, 2, 3, 6, 7, 8, 10, 11, 12 and 13. A more discrete and less intense hybridization signal was seen on the distal third of 4p. The q arms of chromosomes 20 and X were also heavily labeled by the SHIA probe. In most of the above regions the SHIA signal was reasonably uniform, although the Xq and 20q arms showed some focal variations in signal intensity that were suggestive of banding (Fig. 1a). Most TC774 cells contained three X chromosomes, two of which displayed translocated chromosome regions that were negative for SHIA sequences (Fig. 1a). These translocated regions appeared to represent the entire 19 chromosome as judged by both DAPI staining, and the corresponding absence of either 19 homolog in incomplete metaphase spreads with 44 chromosomes.

In an earlier study using these same probes and autoradiographic detection of in situ hybridization signals, a high concentration of SHIA sequences on the p arms of chromosomes 6, 11 and 12 was not demonstrable

(Kuff et al. 1986). Furthermore, 46% of the total SHIA autoradiographic grains were diffusely dispersed, in some cases with high intensity of many chromosome q arms. In contrast, we were not able to detect any minor signals on any chromosome arm beside Xq and 20q. Occasional weak signal seen on only a single metaphase chromatid (Fig. 1a) were considered to be spurious. Because a relative lack of sensitivity in our fluorescent in situ procedure could preclude the detection of widely dispersed SHIA sequences, we hybridized a single copy 2.5 kb GFAP probe, not previously mapped by in situ hybridization, under identical conditions. Several different chromosome preparations were qualitatively evaluated (over 100 metaphase spreads). The 2.5 kb GFAP sequence was reproducibly detected on both sister chromatids of both 11 homologs (as in Fig. 1b) in 30%–50% of all metaphase spreads. Thus we find no evidence for more diffusely dispersed copies of SHIA >2 kb in length. Very short members of SHIA, such as solo LTR (long terminal repeat) elements are probably undetectable by any current in situ procedure. For reference, the SHIA LTR (encoding enhancer, promoter and polyadenylation signals) is only 375 bp, whereas the 5' LTR and coding portion of p27gag is 3010 bp in length (Ono et al. 1985). Our controls clearly indicate that a truncated copy of SHIA capable of coding for only p27 would be detectable in the present experiments.

Almost all of the chromosome arms labeled by SHIA are uniformly heterochromatic when stained by both denaturing and Giemsa-trypsin banding methods. The vast majority of SHIA labeled sites correspond to regions that have been considered to be constitutively heterochromatic (Kuff et al. 1986; Kuff and Lueders 1988). With very rare exceptions, constitutive heterochromatin is transcriptionally inactive in all cell types. In order to understand more fully the functional correlates of SHIA-rich domains, we labeled later replicating regions with BrdU. Simultaneous detection of SHIA and BrdU positive sites in TC774 showed all the SHIA positive regions were later replicating (Fig. 1d–g). We were not able to resolve unambiguously any early replicating domains in SHIA positive regions. Because very long prometaphase chromosomes were not studied in detail, the presence of a very few SHIA sequences within unresolved early replicating domains cannot be ruled out. Given this limitation, it can reasonably be concluded that all SHIA sequences are embedded in later replicating domains in TC774, which on the basis of DAPI counterstaining characteristics are also relatively rich in the bases A and T. Although AT richness and later replication are often considered to indicate transcriptional dormancy of a domain, abundant SHIA transcripts were identified in this cell line in extension reactions with a 55 base TaqI-PstI fragment 3' to the primer binding site (Murdoch et al. 1990). This abundant SHIA transcription was therefore almost certainly derived from later replicating regions in this cell line, but did not result in the formation of characteristic intracisternal R particles as determined by ultrastructural analysis (Manuelidis et al. 1987).

Some later replicating domains on q arms that corresponded to G-dark bands showed no integration of

Fig. 1a–g. Detection of retroviral SHIA and mouse single copy glial fibrillary acidic protein (GFAP) sequences. **a** Syrian hamster chromosomes (TC774) hybridized to biotinylated pSHIA 12.1 and 12.2, detected with Avidin-FITC (fluorescein isothiocyanate) and counterstained with 0.1 µg/ml propidium iodide. Several p arms show a high density of SHIA hybridization signals along their entire length. The 4p arm has only a small discrete signal, and the labeled 20q arm contains an unlabeled central region. Some labeled Xq chromosome arms show band-like variations in signal intensity (e.g. arrows). This cell contains three X chromosomes, two of which have a 19 chromosome translocation (Xt). N nucleus. Confocal microscopy. **b, c** A region of the same normal mouse metaphase spread simultaneously probed with the single copy 2.5 kb GFAP sequence and the L1 repeat. In **b** the red GFAP signal on both sister chromatids (arrows in each of the two 11 homologs) is superimposed on the DAPI (4',6-diamidino-2-phenylindole) blue image. In **c** GFAP is simultaneously localized to a small L1-positive green band at 11E1 by double-probe hybridization. CCD acquisition of conventional epifluorescence signals, with digital superimposition of each of two colors. **d–g** A single TC774 metaphase spread with **d** biotinylated SHIA signals in green (FITC detection), **e** late replicating BrdU (5-bromodeoxyuridine) positive bands in red (rhodamine detection) and **f** DAPI counterstaining (blue). In **g** (merged image of **d** and **e**), all green SHIA signals appear yellow since they overlie later replicating chromosome domains. Some later replicating regions contain no SHIA sequences. Four X chromosomes are identified; in the DAPI image, two X chromosomes bear the 19 translocation (Xt). All four X chromosomes have later replicating SHIA positive q arms whereas three have later replicating p arms (pL) that distinguish the inactive (Xt) from the active X (Xa). Xa contains a 19 translocation. DAPI identification of chromosomes 2 and 4 was confirmed by p/q length measurements in photographic enlargements of more than five metaphases hybridized to SHIA

SHIA sequences (as on chromosome 5, Fig. 1f and g). Therefore SHIA integration is not selective for later replication per se. Interestingly, all SHIA positive Xq arms were later replicating, a feature that has previously been noted in female Syrian hamster fibroblasts (Savage and Bhunya 1980). Nevertheless, the inactive X chromosomes here were readily identified by their later replicating p arms (Fig. 1e and g). Only one active X was seen, even in cells with four X chromosomes (Fig. 1g). Surprisingly, the active X, as in the cell shown, is not the normal X but one that harbors a 19 translocation.

The late replication of all Xq arms, as well as unusual Giemsa staining characteristics of Xq and other uniformly heterochromatic arms in Syrian hamster chromosomes (Popescu and DiPaolo 1979; Li et al. 1982) underscore special features of these domains in this species. It is remarkable that such a large proportion of SHIA sequences are concentrated in these regions. Most other resolved sites of SHIA (as on 4p and 20q) corresponded well with the position of G-dark and DAPI bright chromosome bands. However, a few domains labeled by SHIA included G-light bands, as in 10p and 13p, that were not well resolved in our spreads. The proportionally small size of these G-light bands, and the overall intensity of SHIA hybridization at other heterochromatic sites lead us to estimate that there are maximally 50, and possibly even no SHIA elements residing in G-light domains.

Molecular strategies for characterizing domains allow a more precise localization of single copy GFAP

Studies with human chromosomes have shown that SINES, exemplified by Alu repeats, and LINES, such as L1 repeats, are inhomogeneously distributed on chromosome arms (cf. Manuelidis 1990). In general, interspersed Alu repeats preferentially concentrate in regions that correspond to G-light domains. These regions are early replicating, and are thought to contain the majority of housekeeping genes which are transcriptionally active in all cells. In contrast, interspersed L1 repeats preferentially collect in G-dark regions which contain many tissue-specific genes. Studies of normal lymphocytes have directly shown that L1 sequences collect in later replicating chromosome domains (Manuelidis and Chen 1990). Unique chromosomal fragments with a disproportionately high representation of LINES can be as long as 3 Mb as determined by pulsed field gel electrophoresis (PFGE; Chen and Manuelidis 1989) indicating that L1 repeats can span the entire length of a microscopically visible G-dark band. Recently, a roughly parallel situation has been observed in mice, as L1 sequences also selectively concentrate, with some exceptions, in G-dark bands of *M. domesticus* chromosomes. These mouse L1 hybridization patterns are useful for identifying and subdividing each chromosome (Boyle et al. 1990). A typical mouse L1 hybridization pattern is shown in Fig. 2b.

Because GFAP is a tissue specific gene normally expressed only in astrocytes of the brain, according to the

scheme above it should be found in a G-dark band, and possibly cosegregate with L1 repeats. In order to assess more precisely the position and residence characteristics of GFAP, we simultaneously hybridized the mouse L1 and GFAP sequences. Each of these sequences was detected with a different fluorochrome. As in single probe studies, ~40% of chromosome spreads showed two GFAP signals on each metaphase homolog. GFAP localized within a small L1 positive band at 11E1 (Fig. 1b and c). This L1 band corresponds to a G-dark region in chromosome 11 (Lyon and Searle 1989). The current data are in accord with, but more precise than, the GFAP position previously mapped by hybrid cell line analysis (Bernier et al. 1988). The determined GFAP position is consistent with the preferential localization of other tissue specific genes in G-dark, LINE-rich chromosome bands.

Homologous mouse proviral sequences are widely dispersed in the genome

The mouse intracisternal A particle sequence (MIA) is highly homologous to SHIA, and is present in comparable copy numbers in the genome (Kuff and Lueders 1988). However, unlike SHIA, the MIA sequence hybridized to all normal mouse chromosomes, and appeared to be quite dispersed, even by confocal microscopy (Fig. 2a). MIA hybridization studies were done in both an aneuploid cell line that actively produces intracisternal and budding C-type particles (Manuelidis and Manuelidis 1976) as well as in normal mouse spleen cells. The patterns in both cell types were indistinguishable, but the diffuse nature of the MIA hybridization pattern


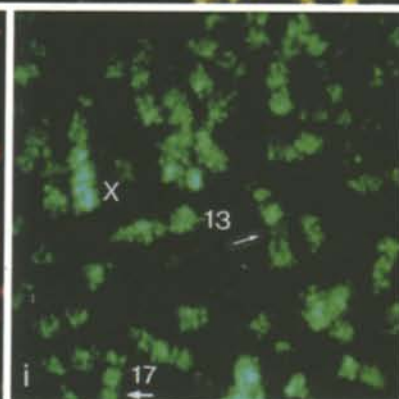
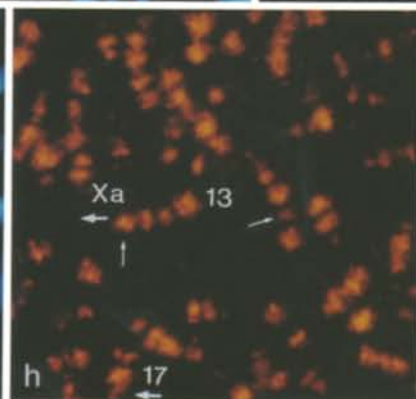
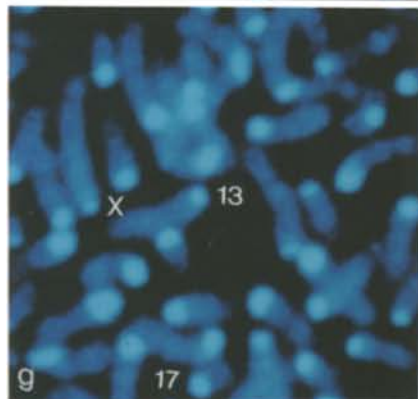
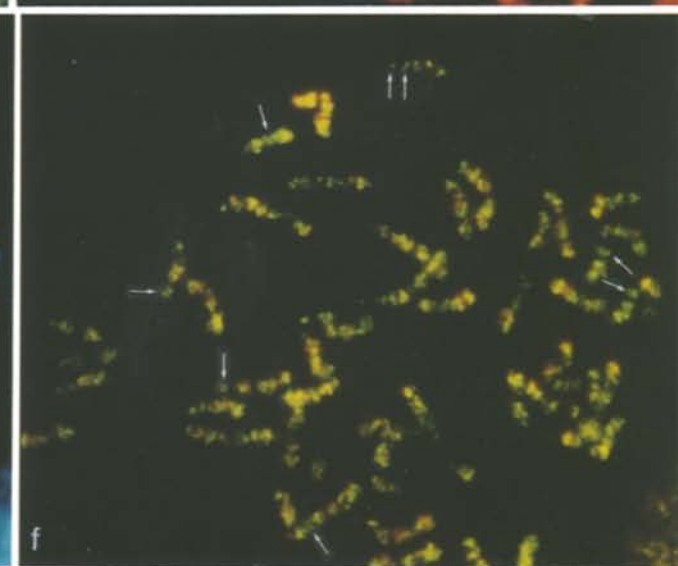
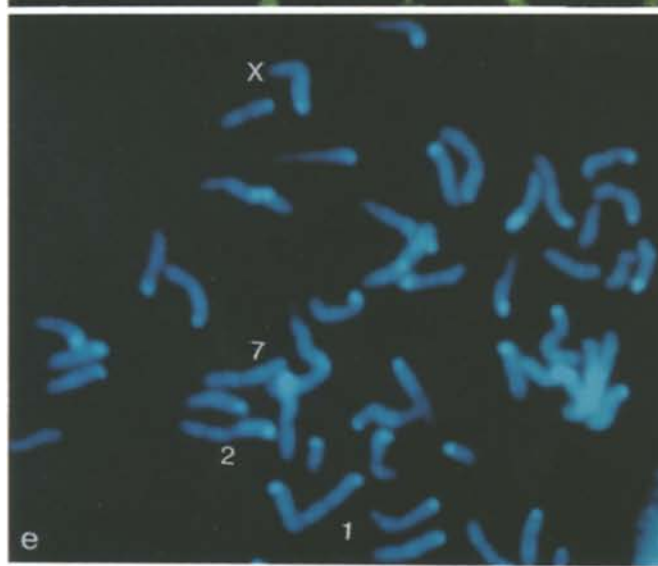
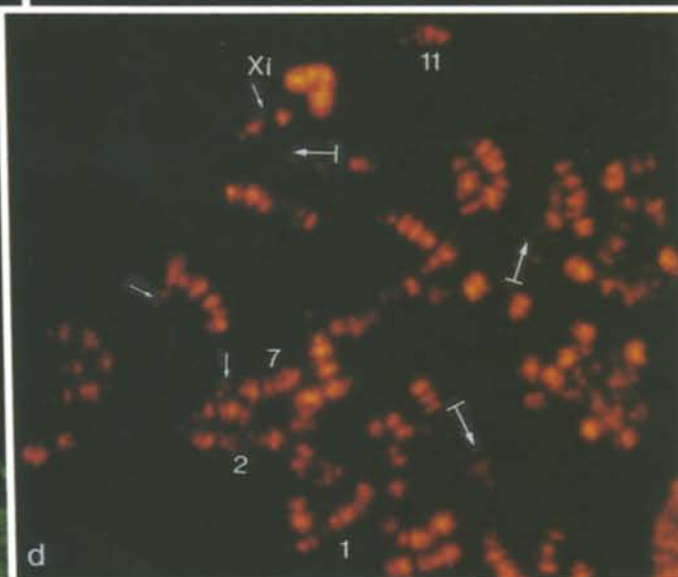
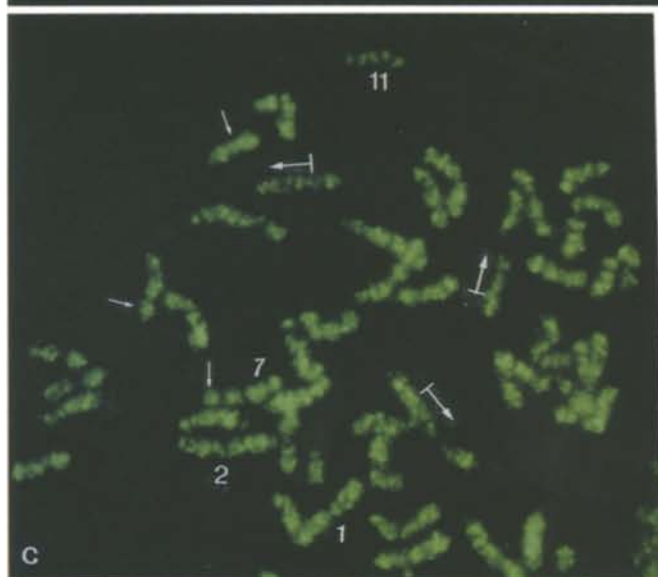
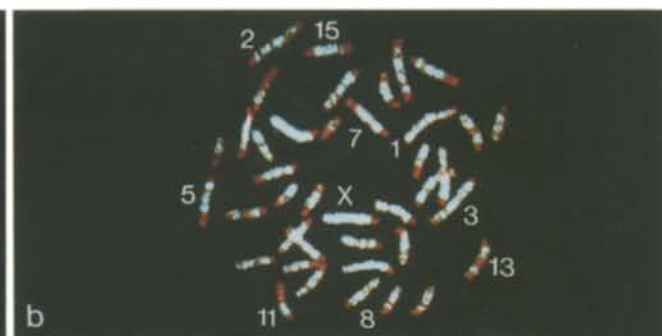
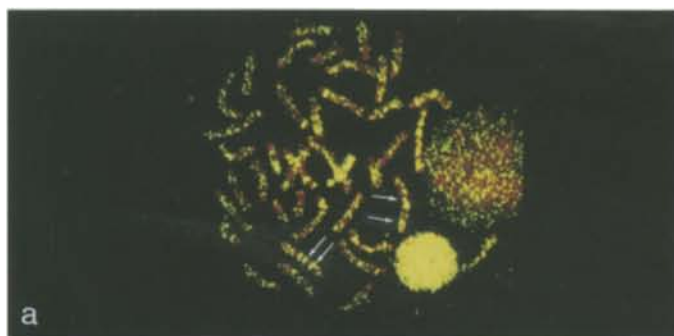


Fig. 2a-j. Repeated sequences detected in *Mus domesticus* chromosomes. **a** Normal mouse chromosomes hybridized to biotinylated MIA, counterstained with propidium iodide. Confocal microscopy emphasizes band-like concentrations of MIA sequences (arrows), but a more dispersed pattern of integration is evident in most chromosome arms at this single z-plane. **b** Characteristic L1 hybridization pattern in normal mouse chromosomes counterstained with propidium iodide. L1 signals are clustered, with less dispersion than MIA sequences. This reproducible L1 pattern was subsequently used to identify each chromosome in double hybridization studies (e.g. Fig. 3). Confocal microscopy. **c-f** Demonstration of MIA sequences in both early and later replicating chromosome domains of TC509. A single metaphase spread is shown with **c** MIA signals **d** BrdU positive bands and **e** DAPI fluorescence. In **f** (the merged image of **c** and **d**), MIA sites within late replicating domains become yellow, whereas MIA sequences in earlier replicating (BrdU negative) domains remain green. Note several band-like sites of MIA integration (arrows) as well as regions with more dispersed MIA elements (terminated arrows) that map to early replicating domains. A late replicating inactive X chromosome is at Xi. CCD acquisition. **g-i** Incongruence of L1 sequences and later replicating bands in TC509. **g** DAPI staining, **h** later replicating bands and **i** L1 hybridization in same field. Although many L1-positive regions correspond to later replicating domains, some later replicating bands have a paucity of L1 sequences (thin arrows). Conversely, some L1-rich regions are early replicating (thicker arrows), such as on autosome 17 and the active X (Xa). The inset shows early replicating L1 regions in a merged image. CCD acquisition



precluded detection of very subtle or minor differences in chromosomal localization. Although these high resolution in situ patterns were comparable to those previously described in autoradiographic studies (Kuff et al. 1986), we further characterized the nature of MIA domains. Double labeling experiments with BrdU and the MIA probe showed that MIA integrated into both early and later replicating domains (Fig. 2c–f). The BrdU positive signals generally defined well demarcated bands, in contrast to the more dispersed MIA signals. In some chromosome arms, such as the distal region of chromosomes 7 and 11, band like concentrations of MIA corresponded to early replicating domains (Fig. 2c and d). Thus the presence of relatively high concentrations of MIA in these band like domains appear to be insufficient for designating later replication. Other more dispersed MIA integration sites were also apparent in early replicating domains. Although it is difficult to estimate the proportion of MIA sequences in early and later replicating domains, the hybridization pattern would suggest that as much as ~25% of MIA sequences (~250 copies) may be stably integrated in early replicating domains.

The correspondance of L1-rich regions with later replication was also evaluated. Although L1 hybridization patterns were useful for identifying each chromosome, and many of the L1-rich regions were later replicating, there were clearly regions that deviated from this pattern (Fig. 2g–j). For example chromosome 17 displayed a distal L1-rich band that was early replicating, as did other chromosomes. The active X chromosomes which are very rich in L1 repeats were also early replicating (Fig. 2h and i). In addition, several later replicating bands did not contain high concentrations of L1 (Fig. 2j). These studies would suggest that L1 segregation into later replicating domains may be less complete in *M. domesticus* than it is in *Homo sapiens*. These studies in mouse also indicate that L1 repeats, as SHIA sequences are not the dominant determinants of later replication in this species.

Because both the MIA and L1 hybridization patterns could be distinguished from later replicating patterns, we sought to determine whether MIA and L1 were more convergent with each other than with BrdU banding patterns. Figure 3a–d shows simultaneous hybridization with L1 and MIA probes in normal mouse chromosomes. The patterns were only partially convergent. For example, although higher concentrations of MIA roughly corresponded to L1 bands in chromosome 1, MIA was found in regions with little L1 (such in the more distal regions of chromosomes 2, 7 and 11). Regions with high concentrations of L1, such as portions of chromosomes 2, 3 and X also did not show similarly high concentrations of MIA. The above results indicate that MIA sequences do not selectively integrate into L1-rich domains. Clearly, the MIA integration pattern is more diffuse than that of L1. Furthermore, some MIA sequences are reasonably concentrated in regions that correspond to G-light domains, such as those on the distal portion of chromosomes 7 and 11. Although future studies are likely to define chromosome domains and their boundaries with more precise monospecific and combin-

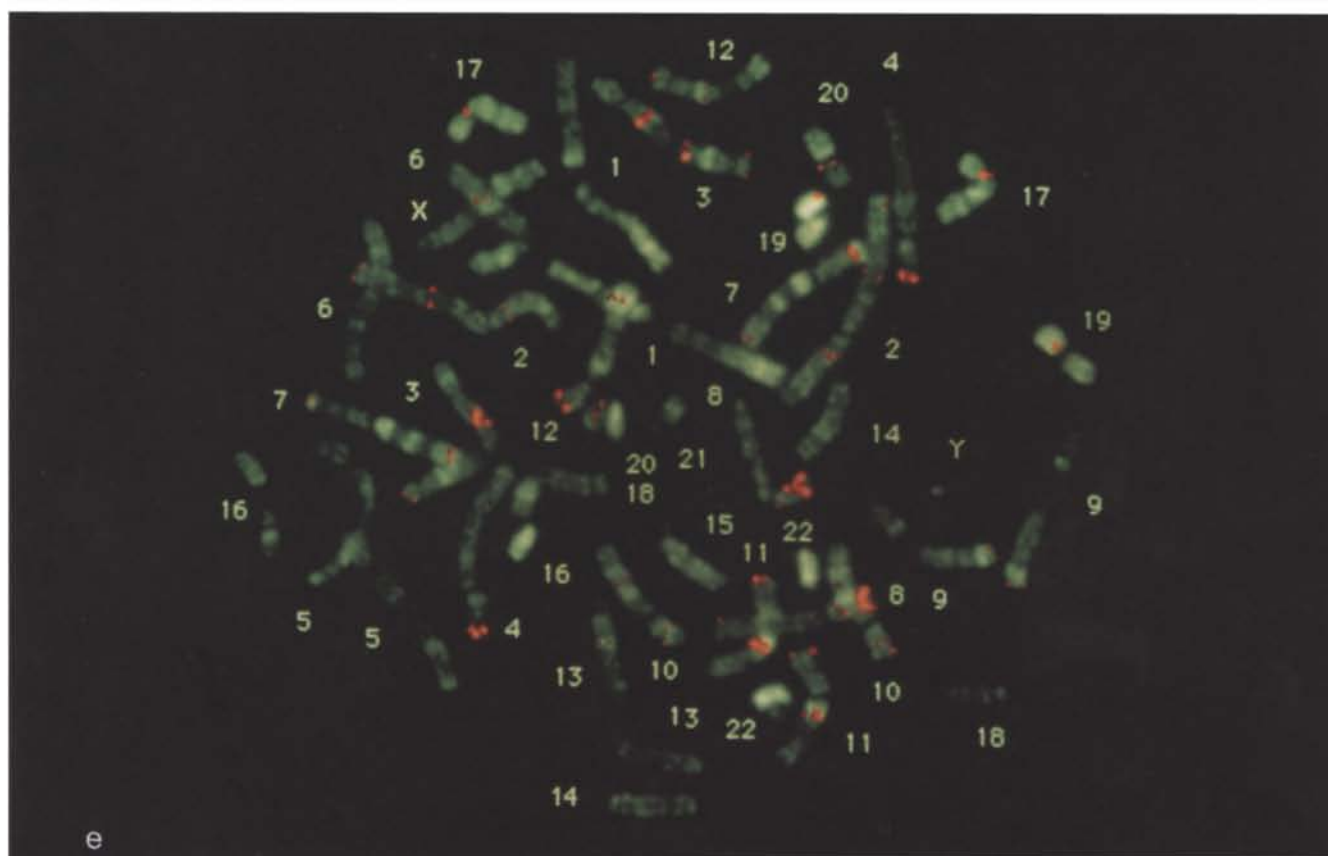
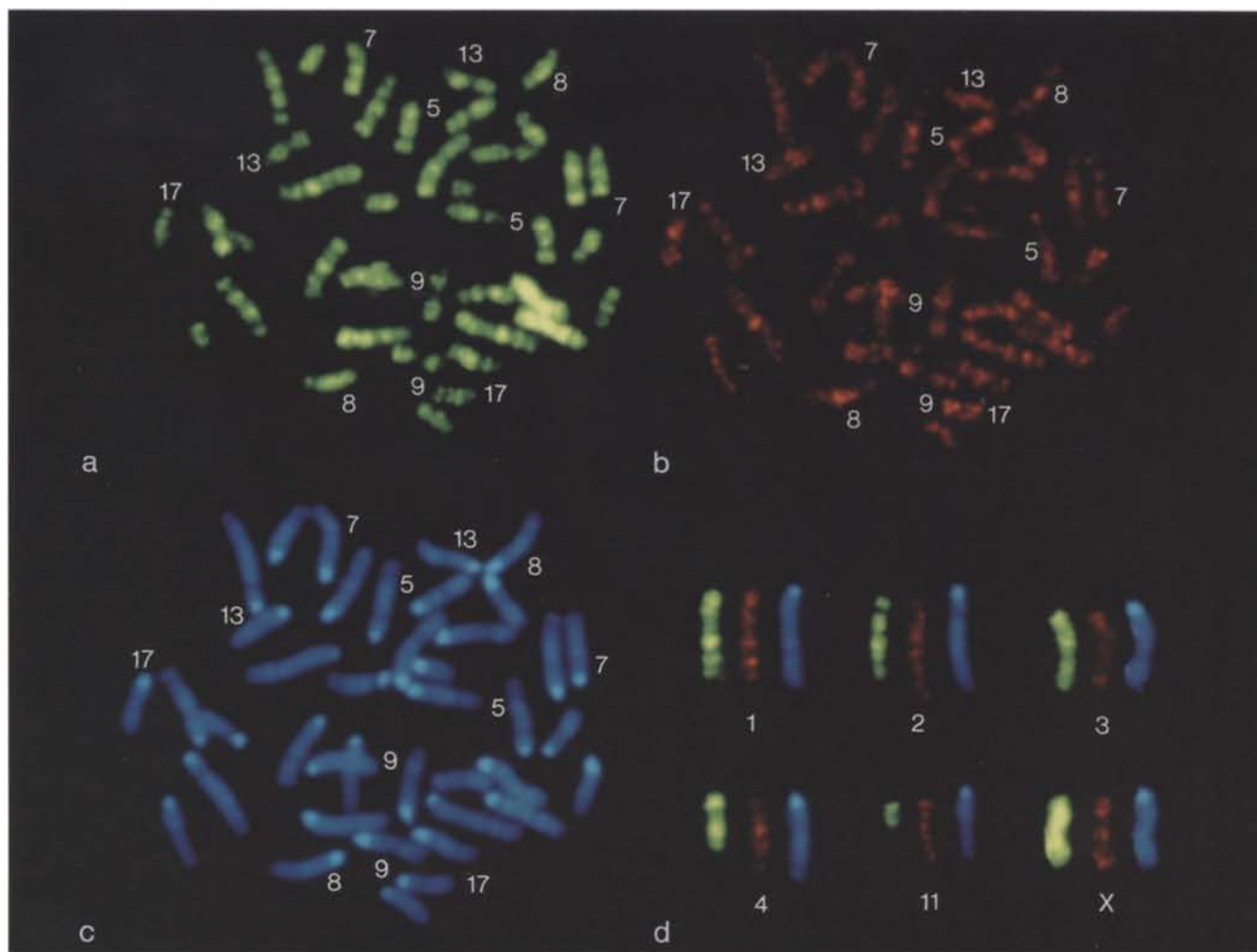
atorial sequence attributes, the present studies emphasize an evolving organizational diversity of like elements in different host genomes.

Human retroviral sequences of moderate copy number prefer selected Alu-rich and G-light domains

Previous Southern hybridization studies with the HERV 4.1 clone in rodent × human hybrid cell lines showed wide dispersion of these proviruses as both full length and truncated *pol-env* copies (Steele et al. 1986). Cloning and Southern blot studies have also failed to reveal tandem repeats of HERV 4.1. On the other hand, recent PFGE studies suggested that the ~50 HERV copies in the genome might be clustered together in selected loci, as the HERV probe hybridized to relatively few fragments of DNA between 50–400 kb (Chen and Manueldis 1989). Only one of these fragments clearly corresponded to an L1-rich fragment, whereas the others were in DNA fragments harboring many SfiI sites and showing intense Alu hybridization. The latter features suggested that HERV 4.1 elements might selectively associate with CpG islands of active chromatin domains. We used in situ hybridization in order (i) to test for non-random HERV integration, (ii) to confirm independently the presence of closely spaced (clustered) HERV elements in the genome, and (iii) to determine whether HERV 4.1 had an integration preference for G-light and Alu-rich domains. Experiments were done with HERV 4.1 alone, or in combination with Alu-PCR products. The localization of HERV was comparable in both single and two probe hybridization studies. The simultaneous decoration of Alu-rich regions gave a highly refined in situ map of HERV 4.1 positions.

Figure 3e shows a representative HERV 4.1 hybridization pattern in normal human chromosomes. Some of the HERV signals are large and very bright (e.g. on chromosomes 8 and 11), whereas other signals on both homologs are considerably smaller and fainter (e.g. on chromosomes 2 and 7). Simultaneous hybridization with Alu-PCR products shows almost all HERV members are positioned within Alu-rich chromosome domains. The Alu-PCR banding pattern corresponds well with known G-light and early replicating domains. Interestingly, HERV signals were disproportionately repre-

Fig. 3a–e. Top panel a–d simultaneous hybridization of biotinylated L1 and digoxigenin labeled MIA sequences to normal mouse spleen cells. a Typical L1 banding pattern, b MIA sequences detected in the same spread, c DAPI fluorescence. Sets in d show high concentrations of MIA can correspond to L1-rich regions (chromosome 1), but are clearly present in L1-poor domains, such as more distal regions of chromosomes 2, 4, and 11. Other chromosomes with disparate L1 and MIA cluster patterns are also numerically identified in both homologs. Bottom panel e simultaneous hybridization of human HERV 4.1 (red signals) and Alu-PCR (polymerase chain reaction) products (green signals) in normal human lymphocytes. This representative merged image shows that most signals overlie Alu-positive domains. Note prominent large signals on chromosomes 8 and 11 and the preponderance of telomeric sites



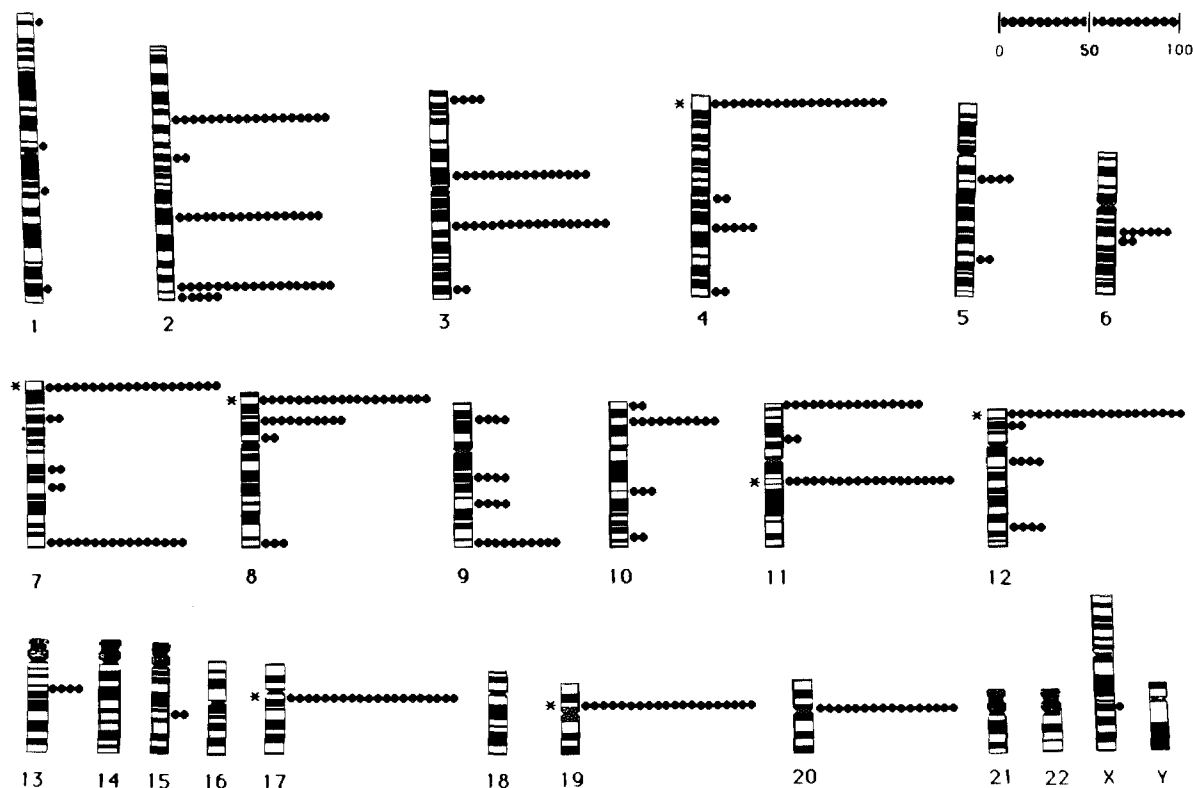


Fig. 4. The location of HERV 4.1 sequences with respect to G-band landmarks. Positive signals (where both chromatids of each of the two homologs were labelled) counted in 100 metaphase spreads.

Each dot on the % scale represents five such observations. Sites with an *asterisk* indicate large, prominent signals consistent with >1 full length (8.8 kb) equivalent of HERV

sented in telomeric regions, and were often seen at the telomeric tip of the chromosome. Merged HERV and Alu hybridization images were used to analyze all the sites giving reproducible signals on both chromatids of both homologous chromosomes. Quantitative counts at each site for 100 metaphases from a normal Caucasian male are shown in Fig. 4. Quantitation of an additional 30 metaphases from the blood of a normal Asiatic male showed no discernible differences in the chromosomal position of HERV (data not shown). Previous restriction length polymorphism studies have also failed to detect variations in HERV in different individuals (Repaske et al. 1985), and in long gels of HindIII, BamHI and EcoRI digests from multiple individual DNA samples, we have only been able to identify a single minor band compatible with a restriction site polymorphism (T. Chen and L. Manuelidis, unpublished data). Thus both in terms of position and fine DNA structure, HERV 4.1 is a very stable genetic element exhibiting little if any mobility.

Quantitative analysis yielded 30 reproducible integration sites for HERV 4.1 that were ≥ 2 kb in length. Twelve of these sites contained truncated copies of 2–4 kb, as judged by counts in comparison with the 2.5 kb single copy GFAP probe (Fig. 4 and Table 1). When both chromatids of both homologs gave a positive hybridization signal in 80%–90% of all metaphase spreads, the site was considered to contain the equivalent of 1–2 full length (8.8 kb) copies of HERV. Although multiple truncated and full length copies within a single locus

cannot be distinguished by in situ hybridization, several large, very bright hybridization signals were likely to contain clusters of HERV, equivalent to 3–6 full length copies. These larger strong signals were detected at 7 of the 30 positive sites (noted by an asterisk). One of these sites, on 11q13.3, contained two closely spaced signals that could be resolved from each other in longer metaphase preparations. From semiquantitative in situ hybridization there are a total of 32–55 copies of HERV (equivalent to ~ 400 kb of this provirus) in the haploid human genome (Table 1). These results are in good accord with the estimate of 35–50 copies of HERV from Southern blot studies (Steele et al. 1984).

Previous studies utilizing rodent \times human hybrid cell lines and Southern blotting had indicated full length HERV 4.1 copies reside somewhere on chromosomes 8, 11 and 12, with additional truncated copies on several other chromosomes (Steele et al. 1986). We have detected definitive HERV 4.1 chromosome integration sites in normal human chromosomes that were not previously detected in the hybrid cell studies. Some of these sites are quite prominent by in situ hybridization (e.g. strong signals on chromosomes 2, 3, 4, 7, 17 and 19). The absence of any in situ hybridization signals on the Y chromosome indicated that the stringency of hybridization used here was sufficient to distinguish HERV 4.1 from another endogenous retroviral sequence with 73% homology that is restricted to the Y chromosome (Silver et al. 1987). We have also not been able to detect a reasonably long truncated HERV member on chromosome

Table 1. Semiquantitative estimate of HERV 4.1 sequences at different sites based on in situ hybridization

Site	Chromosome	Counts ^a	%	Copy no. ^b	G-Light and Alu+	T-bands ^c and very high GC	Break & [hotspots] ^c
1	2 p13	90	3	1-2	Yes		+
2	q24	85	3	1-2	G-dark		
3	p37.1	90	3	1-2	Yes	++	++
4	q37.3	25		Truncated	Yes	++	++
5	3 p25	20		Truncated	Yes		+
6	p12	80	3	1-2	Yes		+
7	q21	90	3	1-2	Yes	++	+++&[+++]
8	4 p16	*100	9	3-5	Yes	++	++
9	q27	30		Truncated	Yes		
10	5 q13.1	20		Truncated	Yes		++
11	6 q16	30		Truncated	G-dark	++	+
12	7 p22.2	*100	9	3-5	Yes	++	+++&[+++]
13	q22	30		Truncated	Yes	++	++&[+++]
14	q36.3	80	3	1-2	Yes		
15	8 p23	*100	9	3-5	Yes		
16	p21.1	50	2	≤1	Yes		+
17	9 p22	20		Truncated	Yes		+
18	q21.1	20		Truncated	G-dark		
19	q22.3	20		Truncated	Yes		++
20	q34.3	50	2	≤1	Yes	++	
21	10 p13	50	2	≤1	Yes		
22	11 p15.3	80	3	1-2	Yes	±	++++&[+++]
23	q13.3	*100 (2)	12	4-6	Yes	++	++++&[+++]
24	12 p13.3	*100	9	3-5	Yes		++&[+]
25	q13.13	20		Truncated	Yes	++	+&[+++]
26	q24.33	20		Truncated	Yes	++	++++&[+++]
27	13 q14.3	20		Truncated	Yes		
28	17 q11.2	*100	9	3-5	Yes		
29	19 p13.1	*100	9	3-5	Yes	++	
30	20 p11.1-11.2	80	3	1-2	Junction		±
Sum ^d			96%	32-55	90%	45%	60%

^a Counts represent two different blood samples as assessed in Fig. 4. The seven sites marked with an asterisk represent >8.8 kb of HERV as judged by the intensity and/or size of the hybridization signal. See text for details of 11q13.3

^b Copy number estimated in terms of full length (8.8 kb) HERV equivalents at different sites. At 11q13.3 two closely spaced hybridization signals were detected. Truncated copies correspond to HERV members estimated to be 2-4 kb in length based on detection frequency of the 2.5 kb GFAP clone

^c Data for T-bands, very high GC content, breakpoints and hotspots are based on mapping data compiled by Holmquist (1991); ± indicates ambiguity

^d The sums exclude the 11 truncated members, estimated to represent a total of 25-30 kb of HERV, or ~4% of the total genomic HERV DNA

14, and this site may represent an unrelated genomic repeat within the 3' probe previously used to detect HERV 4.1 members. Previously we were also unable to detect HERV signals in PFGE studies of a hybrid cell line that initially contained human chromosomes 3 and 4 (Chen and Manuelidis 1989). However, we sub-

sequently found that our passaged hybrid cells, not maintained under selective conditions, lost human chromosomes and especially chromosome ends with high frequency during serial passage. Little is known about potential deletion mechanisms that may be operative in hybrid cells, but it is conceivable that mixing chromo-

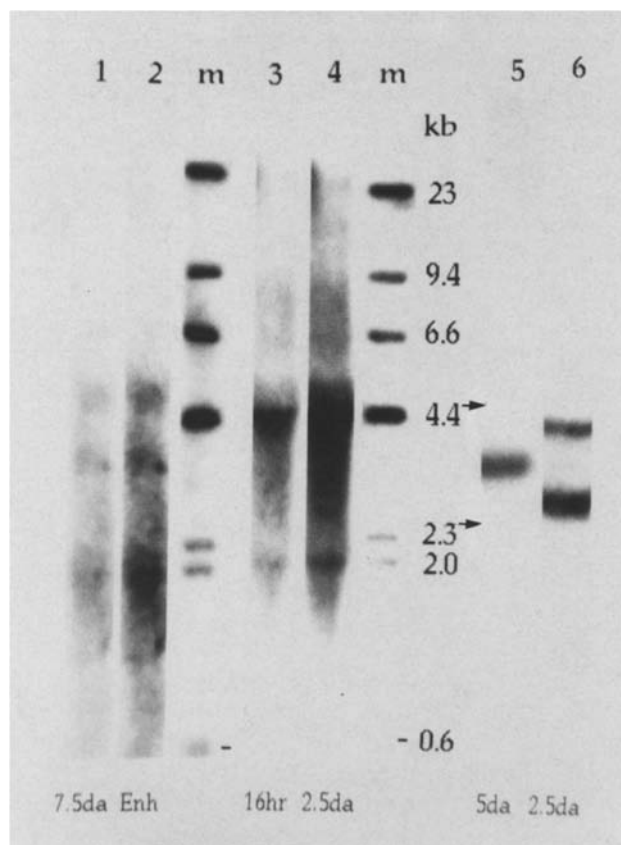


Fig. 5. Blot loaded with 10 μ g per lane of total RNA from normal brain treated with glyoxal. *Lane 1*, human brain RNA hybridized to HERV 4.1 and autoradiographed for 7.5 days. *Lane 2*, a three-fold digital enhancement of lane 1. *Lanes 3 and 4*, Syrian hamster brain RNA hybridized to SHIA, exposed for 16 h and 2.5 days respectively. Note 6–9 kb smear of SHIA RNA. Densitometric scans showed intensities in *lanes 2 and 3* were roughly equivalent, indicating an ~ 25 -fold difference in the levels of HERV and SHIA RNA. *Lanes 5 and 6*, parallel blots of the same hamster brain RNA preparation hybridized to GFAP and neurofilament single copy probes show the expected 3.4 kb GFAP and 2.5/4 kb neurofilament transcripts (Manuelidis et al. 1987) with negligible breakdown. Specific activities of all probes were equivalent, and the human RNA preparation was comparably intact (data not shown). Markers are 32 P labeled and glyoxylated λ HindIII fragments

somes from two different species, each with their own particular retroviral complements, may lead to unusual or subtle recombinations and deletions at retroviral sites. Additionally, since many of the HERV 4.1 sites are within Alu-rich domains associated with breakpoints and recombination (Table 1), some HERV sequences may be especially prone to deletion.

Because semiquantitative in situ hybridization analysis and independent estimates of copy number by Southern blotting were in close agreement, it was possible to analyze further the characteristics of HERV integration. Ninety percent of the copies were present in Alu-rich domains that were G-light. Only two sites with the equivalent of 1–2 copies of HERV were detectable in a G-dark and L1-rich domain. These were at 2q24, and

in the unresolved junction region at 20p11. One of these two chromosomal sites probably corresponds to an ~ 370 kb SfiI PFGE fragment with a high concentration of L1 repeats that was detectable with relatively short autoradiographic exposures for HERV 4.1 (Chen and Manuelidis 1989). L1 sequences have also been identified adjacent to one truncated 4.1 kb HERV clone (Steele et al. 1984). Although interspersed repeat characteristics of HERV clones have not been extensively documented, some HERV clones show Alu repeats adjacent to the LTRs (Repaske et al. 1983). The present in situ hybridization studies show in fact that most HERV members are positioned within Alu-rich domains, with 66% of the total genomic HERV sequences clustered at only seven sites by in situ hybridization. In accord with this, six heavily labeled SfiI fragments from 50–230 kb, with one fragment likely to contain a poorly resolved doublet, were detected by PFGE. Thus the clustering of many HERV members in submicroscopic domains is now reasonably established. Comparatively more minor HERV sites, possibly representing single copy or truncated HERV members were also detectable in other SfiI fragments of similar size with longer autoradiographic exposures. These findings are also compatible with the in situ localization of most truncated members to Alu-rich domains with CpG islands.

Most remarkably, $\sim 45\%$ of HERV 4.1 was found in very GC-rich T-bands (Table 1), and these sites account for $<15\%$ of the human genome (Ambros and Sumner 1987; Holmquist 1991). Thus the majority of HERV shows quite a restricted integration pattern. The selectivity of HERV is even more apparent when one considers that $\sim 50\%$ of the copies are localized in distal telomeric regions. Interestingly, $\sim 60\%$ of the HERV integration sites were also associated with known chromosomal bands exhibiting breakpoints and/or radiation hot spots (Table 1). This association might suggest that HERV either selectively utilizes such sites in its amplification/reintegration cycle or, that HERV integration contributes to the structural fragility of these chromosome domains. The general pattern of integration of HERV 4.1 is almost the exact opposite of that seen with SHIA with respect to chromosomal band characteristics.

Relative abundance of SHIA and HERV transcripts in adult brain

Because virtually all copies of SHIA reside in later replicating domains that are thought to be transcriptionally inactive, one might predict that SHIA transcripts would be very limited in normal cells in vivo. On the other hand, a high level of transcription might be expected from the ~ 45 HERV 4.1 copies in Alu-rich domains with early replication characteristics. We therefore compared the steady state levels of RNA using these two probes. Brain was chosen because it contains a diversity of cell types that reflect a spectrum of chromatin configurations, from the very heterochromatic oligodendrocyte to the extremely euchromatic large neuron. Neither normal hamster nor human adult brain display charac-

teristic retroviral particles in any cell type by electron microscopy.

Figure 5 shows equal amounts of total RNA from normal adult human and Syrian hamster brain hybridized with HERV and SHIA radiolabeled probes of comparable specific activity. These profiles are representative of more than ten independent brain RNA preparations, and emphasize a disproportionate representation of less than full length copies. Transcripts from internally deleted or truncated elements are also typical for other RNAs probed with IAP (intracisternal A peptide sequence) and HERV 4.1 (Kuff and Lueders 1988; Rabson et al. 1985). Visual inspection shows that the human transcripts are very faint as compared with the hamster transcripts, despite the 11-fold longer autoradiographic exposure for HERV (lanes 1 and 3). Densitometric analysis indicated an ~ 25 -fold difference in HERV and SHIA steady state RNA levels in normal brain. Additionally, hamster RNA full length members are readily visible. However, only in polyA selected human brain RNA have we unambiguously been able to detect full length HERV RNAs of ~ 9 kb (G. Murdoch and L. Manuelidis, in preparation), indicating such transcripts are of very low abundance. The ~ 1.7 and 3.6 kb human brain RNAs correspond to previously identified spliced *pol-env* transcripts in polyadenylated placental and carcinoma RNAs (Rabson et al. 1985) but the 3.0 kb *pol-env* placental transcript is not detected in brain. Notably, the RNA blot signal intensity of the 1.7 and 3.6 kb HERV transcripts is less than that produced by unstimulated single copy GFAP in normal adult brain (Manuelidis et al. 1987). Figure 5, lanes 5 and 6 show the intact single GFAP, and two established neurofilament specific gene products, for comparison in equivalent amounts of brain RNA. Thus most HERV members, despite their chromosomal location, appear to be transcriptionally less active than quiescent GFAP. Although unique post-transcriptional processing features could account for the relatively low levels of HERV transcripts, specific HERV sequence motifs, possibly in combination with CpG gene methylation, may selectively inhibit transcription from the ~ 45 HERV genomic members identified here in active chromosome domains. In summary, the transcriptional capacity of HERV 4.1 is very limited, and contrary to its expected behavior from its chromosomal position.

In contrast, multiple SHIA transcripts were visible in total RNA in relatively short autoradiographic exposures of hamster brain RNA. The very abundant smaller transcripts of 4 – 5 kb probably derive from several truncated SHIA genomic elements at different chromosomal loci, as SHIA elements with large internal deletions have been isolated, and transcripts of similar size are observed in numerous mouse preparations where there is no evidence for splicing of IAP transcripts (Kuff and Lueders 1988). The smear of hamster products was not due to RNA breakdown as deduced by the intactness of single copy transcripts (lanes 5 and 6). A similar SHIA profile is also reproducibly seen in other hamster tissue RNA preparations (Murdoch et al. 1990). Moreover, the presence of a series of transcripts of between 6 – 9 kb indicates

transcription of multiple SHIA genomic members that are likely to derive from close to full length genomic elements. The predicted full length SHIA transcript is 8.1 kb. There was a distinct autoradiographic band of the size within the higher molecular weight RNA smear. The longer SHIA RNAs (> 8.1 kb) almost certainly represent transcription of variably sized genomic SHIA members that are read through the $3'$ LTR into flanking cellular sequences, as has been described for other endogenous retroviral elements (Wilkinson et al. 1990). This type of readthrough could denote a select set of genes under the control of the SHIA $5'$ LTR where "constitutively" inactive heterochromatic features are subverted. In this case one might expect a tissue specific pattern of SHIA expression. In fact, spleen and thymus show a high abundance of SHIA transcripts (two- to fourfold greater than brain), whereas liver has almost undetectable levels of SHIA RNA (Murdoch et al. 1990). Furthermore, the higher molecular weight RNA species are very prominent in spleen and thymus.

Given the in situ hybridization results detailed above, it is likely that multiple SHIA genes within heterochromatic domains are transcriptionally competent in brain, as they are in thymus, spleen and cultured cells. Although the SHIA data provide an exceptional example of transcriptional gene activity from special heterochromatic domains, the tissue specific pattern of SHIA RNA is in keeping with attributes generally ascribed to other G-dark regions that do not have long tandem repeats. In fact, the term constitutive should probably be reserved for regions with long stretches of satellite DNA which are very tightly folded in interphase and transcriptionally inactive in almost all cells (Manuelidis 1990, 1991). The finding of numerous SHIA genes, as well as single copy GFAP within G-dark chromatin should also dispel the simplified notion that "few genes at all" reside in these regions (Bickmore and Sumner 1989).

Discussion

Although retroviruses utilize a similar strategy for invading the host genome (see Varmus 1988 for a review), the present studies show that retroviral integration can be fastidiously selective in a natural setting. We have directly visualized more general macromolecular features of retroviral chromosome integration that are not readily appreciated in detailed molecular studies. The present studies emphasize both the diversity and specificity of retroviral integration. The three endogenous retroviruses studied here have very different patterns of genomic integration despite their overall structural similarity ($5'$ LTR-*gag-pol-env*-LTR reading frames), and relative simplicity by comparison with HIV or HTLV-1. Overall chromatin attributes, as defined by Giemsa staining characteristics, time of replication in the cell cycle, and currently defined DNA sequence signatures (such as the preponderance of Alu or L1 repeats) appear to be poor predictors of retroviral insertion sites. This may be due to complex and dynamic host-retroviral interactions involving repeated insertions and deletions over a long period of time.

Hamster IAP endogenous retroviruses are integrated selectively into heterochromatic portions of the genome, most of which show unusual Giemsa staining characteristics. Although the basis for the unusual nature of the Syrian hamster heterochromatin is not known, it is notable that numerous SHIA elements are found in all of these regions. All domains occupied by SHIA were also later replicating, including the q arm of the active X chromosome, a region that would be expected to be early replicating in most other species. Although the presence of many SHIA sequences in the Xq arm could be construed as determinants of later replication, the above studies show that SHIA sequences are not required for later replication, because some large, later replicating domains are free of SHIA sequences. However, SHIA elements may delineate, or contribute attributes to a select set of later replicating domains. They may also define some special protein binding domains within hamster heterochromatin.

By comparison, the very closely related mouse IAP sequence, present at comparable genomic copy numbers, was distributed on all chromosomes with a very dispersed integration pattern. As many as 25% of the total MIA sequences were found in early replicating domains with a paucity of L1 sequences. Most notably, these different IAP integration patterns in hamster and mouse do not denote comparable differences in transcriptional properties. In both instances, a similar tissue specific pattern of expression is seen (Kuff and Lueders 1988; Murdoch et al. 1990). These findings may imply a similar type of coevolution of IAP elements with tissue specific genes in rodents. In this context, a demonstrable role for defective Class II endogenous MMTV virus in shaping the immunological repertoire (Frankel et al. 1991) poses the question of potential class II IAP functions in normal immunocompetence. In the unique heterochromatin of Syrian hamster, retroviral LTRs with their promoter, enhancer and polyadenylation signals, might drive the expression of adjacent essential genes that otherwise would be completely silenced. The RNA blotting results for SHIA were consistent with the recruitment of adjacent cellular genes in normal tissue.

The position of the human endogenous HERV 4.1 elements within GC-rich, actively transcribed and early replicating chromatin regions is more in keeping with current concepts of preferred retroviral integration sites (Rohdewald et al. 1987). However, as in the case of SHIA, the surrounding chromatin features did not correctly indicate the transcriptional capacity of HERV. Steady state HERV RNA was quite low, contrary to expectations for these Alu-rich, early replicating domains. Thus unlike most other genes studied to date, including GFAP, the chromosomal setting of each endogenous retrovirus inadequately forecasts its transcriptional expression. Although the site of chromosomal integration can determine the activity of a newly inserted provirus (Jahner and Jaenisch 1985), the present results conversely suggest that a retroviral sequence may also be capable of modulating or overriding local chromosomal characteristics.

The more detailed preferences of HERV for a highly

selected subset of GC-rich domains, i.e. the disproportionate HERV copy numbers in telomeric regions, in very GC-rich T-bands, and in relatively close proximity to each other at six to seven loci by both *in situ* hybridization and PFGE analysis (Chen and Manuelidis 1989) has not previously been appreciated. The presence of clusters of HERV in close proximity to each other, as well as the finding of several HERV loci on a single chromosome (e.g. chromosome 2), could suggest that regional, chromosome specific, and possibly even spatial parameters are involved in the selection of integration sites. The predilection of HERV 4.1 for telomeric regions, and bands with breakpoints and hotspots may also indicate chromosomal repair mechanisms are sometimes required for integration.

The integration site preference of this C-type virus does not simply relate to its classification, because other human C-type (Class I) proviruses can have a very dispersed chromosomal distribution, similar to that seen with the MIA (Class II) sequence (Fraser et al. 1988). Because LTR sequences are unique for each retrovirus, these elements may have a role in determining both integration site preferences and functional attributes (such as transcription or higher order folding), in concert with controls specified by the host. The importance of LTR sequences in this context is underscored by the vastly different integration patterns of MIA and SHIA sequences. MIA and SHIA sequences are highly homologous overall, but have considerable divergence in their LTR regions. There also seems to be some mechanism whereby a host can choose a particular retroviral sequence for amplification to high copy numbers. SHIA related members are less numerous in other rodents (Anderson et al. 1990), and are present at only 50 copies in human DNA (Ono et al. 1986). Whether LTR sequences have a role in this selection for further amplification is entirely unknown.

Because endogenous retroviral elements can reintegrate into the genome via conventional reverse transcriptional mechanisms (Heidmann and Heidmann 1991), new genomic members can be continuously inserted into the host genome. However, transposition can be highly mutagenic for the retroviral sequence. Thus the host genome is capable of tailoring these elements. Presumably, truncated retroviral elements with only a single LTR would be incapable of reintegration (Weiss et al. 1985; Varmus 1988). The truncation and internal deletions of IAP sequences (Kuff and Lueders 1988) as well as the observation of a fair proportion of truncated HERV members here, could indicate these changes are part of a stabilizing mechanism in the genome. Indeed, the positional stability of HERV 4.1 in genetically divergent individuals argues for a fastidious maintenance of these restricted integration sites.

In view of the apparent stability of these elements, their selected amplification, and their extraordinarily different patterns of integration, it is worth reconsidering potential beneficial functions of endogenous retroviruses that might yield selective advantages for the host. The present results indicate contrary transcriptional capacities of SHIA and HERV sequences that may be utilized

by the host to override more general characteristics of a chromosome domain. In addition, comparable endogenous retroviruses are known to be involved in pseudotyping, the formation of mosaic viral forms (Hu and Temin 1990), the production of regulatory cytoplasmic factors (Kuff and Leuders 1988) and the occupation of cytoplasmic sites, all of which may protect the host from extrinsic retroviral infections. Notably, HIV pseudotypes can be blocked 1000-fold by prior infection of cells with an amphotropic MLV retrovirus (Spector et al. 1990). The finding of inverted genomic orientations, and both antisense and double-stranded RNA intermediates (Kuff and Leuders 1988), may also indicate a protective function for high copy retroviral elements. In an organism, at least some of these pathways might be utilized to attenuate exogenous infections.

The long affair of complex organisms with retroviral elements, which predates the emergence of eukaryotic organisms (Temin 1989), suggests a symbiotic rather than purely parasitic relationship. One macabre aspect of endogenous retroviruses deserves special mention in the context of the potential primate origin of HIV. Some species may infect their competitors by shedding virus that is detrimental to others but not to themselves. Several endogenous MLV retroviruses that are xenotropic (i.e. infectious for other strains or species) produce an apparently innocuous viremia in their host of origin (Weiss et al. 1985). A malicious phenotype of this sort might escape detection in a laboratory setting, but could be advantageous for the host in a natural environment.

Acknowledgements. We thank Gwyn Ballard for help with the CCD equipment, and A. Boyle for advice concerning mouse L1 hybridizations. E. Zelazny provided invaluable technical assistance with cell preparations and color printing. This work was supported by NIH grants CA15044 and AG03106. The salary of D. Taruscio was generously supported by the Istituto Superiore di Sanita', Rome, Italy.

References

- Ambros PF, Summer AT (1987) Correlation of pachytene chromomeres and metaphase bands of human chromosomes, and distinctive properties of telomeric regions. *Cytogenet Cell Genet* 44:223–228
- Anderson KP, Lie YS, Low ML, Williams SR, Fennie EH, Nguyen TP, Wurm FM (1990) Presence and transcription of intracisternal A-particle-related sequences in CHO cells. *J Virol* 64:2012–2032
- Baldini A, Ward DC (1991) In situ hybridization banding of human chromosomes with Alu-PCR products: a simultaneous karyotype for gene mapping studies. *Genomics* 9:110–114
- Bernier L, Coleman DR, D'Eustachio P (1988) Chromosomal locations of genes encoding 2', 3' cyclic nucleotide 3'-phosphodiesterase and glial fibrillary acidic protein in the mouse. *J Neurosci Res* 20:497–504
- Bickmore WA, Sumner AT (1989) Mammalian chromosome banding – an expression of genome organization. *Trends Genet* 5:144–148
- Boyle A, Ballard SG, Ward DC (1990) Differential distribution of long and short interspersed element sequences in the mouse genome: chromosome karyotyping by fluorescence in situ hybridization. *Proc Natl Acad Sci USA* 87:7757–7761
- Borden J, Manuelidis L (1988) Movement of the X chromosome in epilepsy. 242:1687–1691
- Chen TL, Manuelidis L (1989) SINEs and LINEs cluster in distinct DNA fragments of Giemsa band size. *Chromosoma* 98:309–316
- Chirgwin JM, Przybyla AE, MacDonald RJ, Rutter WJ (1979) Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry* 18:5294–5299
- Cremer T, Tesin D, Hopman AHN, Manuelidis L (1988) Rapid interphase and metaphase assessment of specific chromosomal changes in neuroectodermal tumor cells by in situ hybridization with chemically modified DNA probes. *Exp Cell Res* 176:199–220
- Dhar R, McClements W, Enquist LM, Vande Woude G (1980) Nucleotide sequences of integrated Moloney sarcoma provirus long terminal repeats and their host and viral junctions. *Proc Natl Acad Sci USA* 77:3937–3941
- Feinberg AP, Vogelstein B (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal Biochem* 132:6–13
- Frankel WN, Rudy C, Coffin JM, Huber BT (1991) Linkage of MIs genes to endogenous mammary tumor viruses of inbred mice. *Nature* 349:526–528
- Fraser C, Humphries RK, Mager DL (1988) Chromosomal distribution of the RLV-H family of human retrovirus-like sequences. *Genomics* 2:280–287
- Heidmann O, Heidmann T (1991) Retrotransposition of a mouse IAP sequence tagged with an indicator gene. *Cell* 64:159–170
- Holmquist GP (1990) Mutational bias, molecular ecology and chromosome evolution. In: Obe G, Evans JH, Natarajan HS (eds) *Advances in mutagenesis research*, vol 2. Springer, Berlin Heidelberg New York, pp 95–126
- Hu W-S, Temin HM (1990) Retroviral recombination and reverse transcription. *Science* 250:1227–1233
- Jahner D, Jaenisch R (1985) Chromosome position and specific demethylation in enhancer sequences of germ line-transmitted retroviral genomes during mouse development. *Mol Cell Biol* 5:2212–2220
- Kuff EL, Lueders KK (1988) The intracisternal A-particle gene family: structure and functional aspects. *Adv Cancer Res* 51:183–275
- Kuff EL, Fewell JE, Lueders KK, DiPaolo JA, Amsbaugh SC, Popescu NC (1986) Chromosome distribution of intracisternal A-particle sequences in the Syrian hamster and mouse. *Chromosoma* 93:213–219
- Leib-Mosch C, Brack-Werner R, Werner T, Bachmann M, Faff O, Erfle V, Hehlmann R (1990) Endogenous retroviral elements in human DNA. *Cancer Res [Suppl]* 50:5636–5642
- Lewis SA, Balcarek JM, Krek V, Shelanski M, Cowan NJ (1984) Structure of a cDNA clone encoding mouse glial fibrillary acidic protein: structural conservation of intermediate filaments. *Proc Natl Acad Sci USA* 81:2743–2746
- Li S, Pathak S, Hsu TC (1982) High resolution G-banding patterns of Syrian hamster chromosomes. *Cytogenet Cell Genet* 33:295–302
- Lichter P, Cremer T, Borden J, Manuelidis L, Ward CD (1988) Delineation of individual human chromosomes in metaphase and interphase cells by in situ suppression hybridization using recombinant DNA libraries. *Hum Genet* 80:224–234
- Lueders KK, Kuff EL (1983) Comparison of the sequence organization of related retrovirus-like multigene families in three evolutionarily distant rodent genomes. *Nucleic Acids Res* 11:4391–4408
- Lyon MF, Searle AG (1989) *Genetic variants and strains of the laboratory mouse*, 2nd edn. Oxford University Press, Oxford
- Maniatis T, Fritsch EF, Sambrook J (1982) *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York
- Manuelidis EE, Fritch WW, Kim JH, Manuelidis L (1987) Immortality of cell cultures derived from brains of mice and hamsters infected with Creutzfeldt-Jakob disease agent. *Proc Natl Acad Sci USA* 84:871–875

- Manuelidis L, Manuelidis EE (1976) Amount of satellite DNA in four experimentally induced tumors of the central nervous system. Quantitative changes in glioblastoma producing C-type particles. *J Natl Cancer Inst* 56:43–50
- Manuelidis L, Ward D (1984) Chromosomal and nuclear distribution of the HindIII 1.9 kb human DNA repeat segment. *Chromosoma* 91:28–38
- Manuelidis L (1985) In-situ detection of DNA sequences using biotinylated probes. *Focus* 7:4–8
- Manuelidis L, Tesin D, Sklaviadis T, Manuelidis EE (1987) Astrocyte gene expression in Creutzfeldt-Jakob disease. *Proc Natl Acad Sci USA* 84:5937–5941
- Manuelidis L, Borden J (1988) Reproducible compartmentalization of individual chromosome domains in human CNS cells revealed by in situ hybridization and three dimensional reconstruction. *Chromosoma* 96:397–410
- Manuelidis L (1990) A view of interphase chromosomes. *Science* 250:1533–1540
- Manuelidis L, Chen TL (1990) A unified model of eukaryotic chromosomes. *Cytometry* 18:8–25
- Manuelidis L (1991) Heterochromatic features of an 11 Mb transgene in brain cells. *Proc Natl Acad Sci USA* 88:1049–1053
- Mietz JA, Grossman Z, Lueders KK, Kuff EL (1987) Nucleotide sequence of a complete mouse intracisternal A-particle genome: relationship to known aspects of particle assembly and function. *J Virol* 61:3020–3029
- Mooslehner K, Karls U, Harbers K (1990) Retroviral integration sites in transgenic Mov mice frequently map in the vicinity of transcribed DNA regions. *J Virol* 64:3056–3058
- Murdoch G, Sklaviadis T, Manuelidis EE, Manuelidis L (1990) Potential retroviral RNAs in Creutzfeldt-Jakob Disease. *J Virol* 64:1477–1486
- Nelson DL, Ledbetter SA, Corbo L, Victoria MF, Ramirez-Solis R, Webster TD, Ledbetter DH, Caskey CT (1989) Alu polymerase chain reaction: a method for rapid isolation of human-specific sequences from complex DNA sources. *Proc Natl Acad Sci USA* 86:6686–6690
- Ono M, Toh H, Miyata T, Awaya T (1985) Nucleotide sequence of the Syrian Hamster intracisternal A-particle gene: close evolutionary relationship of type A particle to types B and D oncovirus genes. *J Virol* 55:387–394
- Ono M, Yasunaga T, Miyata T, Ushikubo H (1986) Nucleotide sequence of human endogenous retrovirus genome related to the mouse mammary tumor virus genome. *J Virol* 60:589–598
- Panganiban AT (1985) Retroviral DNA integration. *Cell* 42:5–6
- Popescu NC, DiPaolo JA (1979) Heterogeneity of constitutive heterochromatin in somatic Syrian hamster chromosomes. *Cytogenet Cell Genet* 24:53–60
- Rabson AB, Hamagishi Y, Steele PE, Tykocinski M, Martin MA (1985) Characterization of human endogenous retroviral envelope RNA transcripts. *J Virol* 56:176–182
- Repaske R, O'Neil RR, Steele EE, Martin MA (1983) Characterization and partial nucleotide sequence of endogenous C retrovirus segments in human chromosomal DNA. *Proc Natl Acad Sci USA* 80:678–682
- Repaske R, Steele PE, O'Neil RR, Rabson AB, Martin MA (1985) Nucleotide sequences of a full-length human endogenous retroviral segment. *J Virol* 54:764–772
- Rohdewohld H, Weiher H, Reik W, Jaenisch R, Breindl M (1987) Retrovirus integration and chromatin structure: Moloney murine leukemia proviral integration sites map near DNase I-hypersensitive sites. *J Virol* 61:336–343
- Savage JRK, Bhunya SP (1980) Cytological sub-division of S-phase in the Syrian hamster (*Mesocricetus auratus*). *Chromosoma* 77:169–180
- Shih CC, Stoye JP, Coffin JM (1988) Highly preferred targets for retrovirus integration. *Cell* 53:531–537
- Silver J, Rabson A, Bryan T, Willey R, Martin M (1987) Human retroviral sequences on the Y chromosome. *Mol Cell Biol* 7:1559–1562
- Spector DH, Wade E, Wright DA, Koval V, Clark C, Jaquish D, Spector SA (1990) Human immunodeficiency virus pseudotypes with expanded cellular and species tropism. *J Virol* 64:2298–2308
- Steele PE, Rabson AB, Bryan T, Martin MA (1984) Distinctive termini characterize two families of human endogenous retroviral sequences. *Science* 225:943–947
- Steele PE, Martin MA, Rabson AB, Bryan T, O'Brian SJ (1986) Amplification and chromosomal dispersion of human endogenous retroviral sequences. *J Virol* 59:545–550
- Temin HM (1989) Retrons in bacteria. *Nature* 339:254–255
- Varmus H (1988) Retroviruses. *Science* 240:1427–1435
- Vernant JC, Maurs L, Gessain A, Barin F, Gout O, Delaporte JM, Sanhadji K, Buisson G, de-The G (1987) Endemic tropical spastic paraparesis associated with human T-lymphotropic virus Type I: a clinical and seroepidemiological study of 25 cases. *Ann Neurol* 21:123–130
- Weiss R, Teich N, Varmus H, Coffin J (1985) RNA tumor viruses, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York
- Wilkinson DA, Freeman D, Goodchild NL, Kelleher CA, Mager DL (1990) Autonomous expression of RTVL-H endogenous retroviruslike elements in human cells. *J Virol* 64:2157–2167