SCIENTIFIC REPORTS

Infection with retroviral vectors leads to perturbed DNA replication increasing vector integrations into fragile sites

Assaf C. Bester, Moshe Kafri*, Karin Maoz & Batsheva Kerem

Department of Genetics, The Life Sciences Institute, Edmond J. Safra Campus, The Hebrew University, Jerusalem 91904, Israel.

Genome instability is a hallmark of cancer. Common fragile sites (CFSs) are specific regions in the human genome that are sensitive to replication stress and are prone to genomic instability in different cancer types. Here we molecularly cloned a new CFS, FRA11H, in 11q13. The genomic region of FRA11H harbors a hotspot of chromosomal breakpoints found in different types of cancer, indicating that this region is unstable during cancer development. We further found that FRA11H is a hotspot for integrations of Murine Leukemia Virus (MLV)-based vectors, following CD34+ infections *in vitro* as well as *ex-vivo* during gene therapy trials. Importantly, we found that the MLV-based vector infection *in-vitro* leads to replication perturbation, DNA damage and increased CFS expression. This suggests that infection by MLV-based vectors leads to replication-induced genome instability, raising further concerns regarding the use of retroviral vectors in gene therapy trials.

Common fragile sites (CFSs) are specific regions in the human genome that appear as gaps and constrictions in metaphase chromosomes from cells grown under mild replication stress conditions[1]. These sites are present in all individuals and are considered to be part of the normal chromosomal structure. *In vitro*, most CFSs are induced using low concentrations of aphidicolin, an inhibitor of polymerase α, δ and ε[2,3]. Under these conditions, CFSs are hotspots for sister chromatid exchanges, translocations and deletions[4-6]. At present, 87 common fragile sites have been characterized at low resolution by their cytogenetic appearance on metaphase chromosomes (http://www.ncbi.nlm.nih.gov/gene), but the exact repertoire of common fragile sites depends on the inducer, cell type and method of analysis. Using fluorescence in situ hybridization (FISH) 39 of the CFSs have been molecularly cloned, and their exact genomic localization determined[7].

In vivo, CFSs correlate with chromosomal breakpoints in tumors[8,9]. Recently, several studies have analyzed genomic deletions in cancer cells and demonstrated that CFSs are preferentially unstable in the early stages of cancer development and precede instability in non-fragile regions[10-16]. A large scale analysis of 750 cancer cell lines revealed an instability signature in CFSs that differed significantly from that found in cancer genes, indicating that there is a unique mechanism leading to genomic instability in fragile sites[7]. CFSs were also found to be involved in the occurrence of duplications and amplifications in cancer[12-14]. The breakpoints setting the amplicon boundaries are within the CFSs and they drive breakage-fusion-bridge (BFB) cycles.

Altogether, the results of these studies suggest that DSBs at fragile sites may themselves contribute to oncogenesis by triggering genomic instability that can lead to disruption of tumor suppressor genes or oncogenes harbored within these regions.

Another type of instability at CFSs involves the integration of foreign DNA into the genome. *In vitro* studies have shown that under replication stress conditions, plasmid DNA integrates preferentially into CFSs[17]. *In vivo*, CFSs were also found to be preferential sites for integration of viral DNA. Human papillomavirus (HPV), the most important cancer related virus, was found to integrate preferentially into CFSs in the genome of most cervical cancer cells[18]. Another oncogenic retrovirus found to preferentially integrate into CFSs is the Xenotropic Murine Leukemia related virus (XMRV)[19], a relatively new type of retrovirus associated with a subset of prostate tumors[20].

We have previously found preferential integrations of retroviral-derived vectors, based on Moloney Murine Leukemia virus (MLV), into CFSs[21] both *in vitro* and in gene therapy trials. Unfortunately, in two gene therapy trials on patients with X-linked severe combined immunodeficiency (SCID-X1), five of the 20 treated patients developed monoclonal acute lymphoblastic leukemia-like lymphoproliferation. The leukemia in four of these patients involved activation of the LOM2 proto oncogene[22,23], which resides within the CFS FRA11E[21].
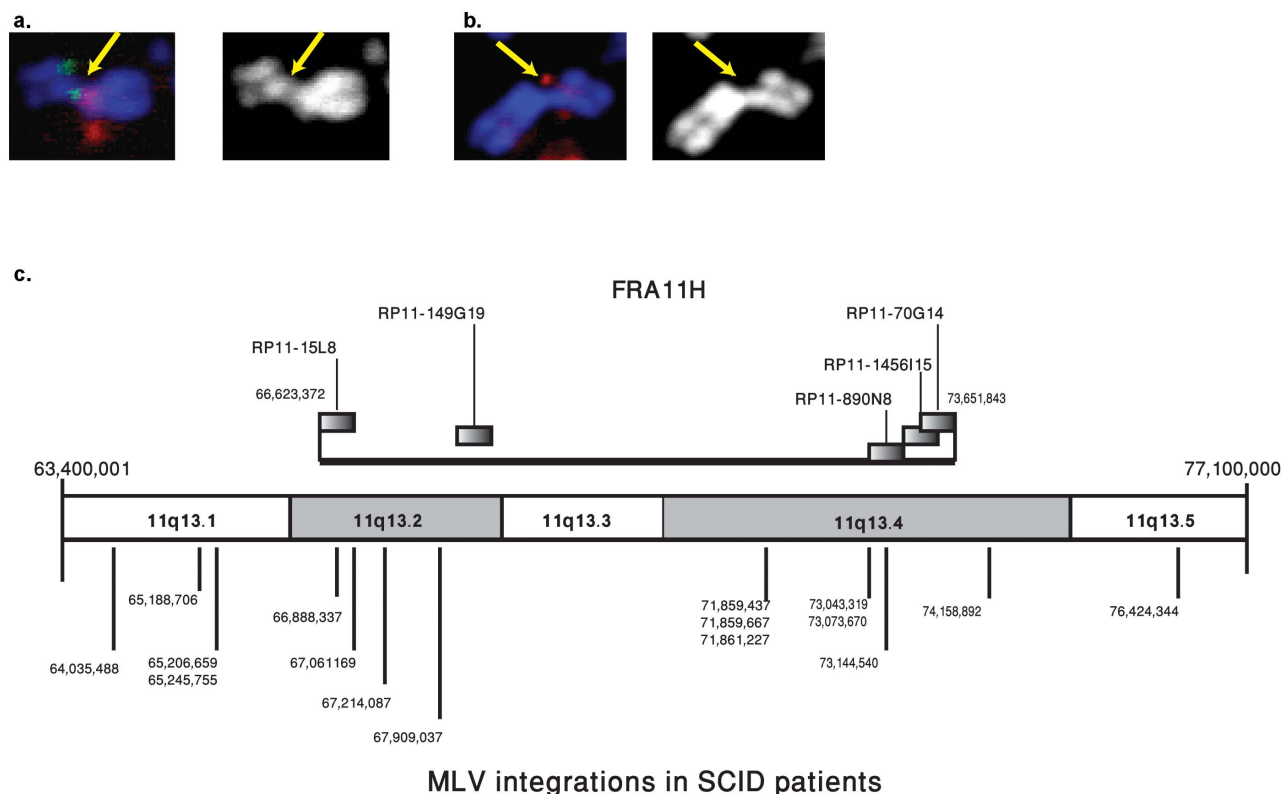
Figure 1 | The positional cloning of FRA11H. (a and b) Examples of FISH signals using clones from 11q13 on chromosome 11 expressing FRA11H. (a) Right, chromosome 11 (stained with DAPI) expressing FRA11H; Left, FISH signals using RP11-809N8 (green) and RP11-15L8 clones as probes (red). Note that the FRA11H gap is in between the clones. (b) Right, chromosome 11 (stained with DAPI) expressing FRA11H; Left, FISH signals using RP11-15L8 as a probe (red). Note that the FISH signals are on the FRA11H uncondensed region. (c) Schematic representation of chromosome 11q13. The horizontal bar above the chromosome represents the CFS, FRA11H region. The gray boxes represent the BACs used for the cloning of the CFS. The vertical lines under the chromosome represent integration sites of MLV-based vectors found in the SCID patients. The numbers indicate the genomic position of each integration.

Here we studied a region in 11q13 known to be a hotspot for chromosomal instability in many cancer types. Recurrent amplifications in this region are found in 15% of breast carcinomas, 13% of lung cancers, 21% of bladder tumors, 50% of esophageal cancers and 45% of oral squamous cell carcinomas[24]. Recently, a large scale analysis of 50 gastric tumors revealed that the 11q13.3 region is among the most common genomic deleted regions[25].

The 11q13 band encompasses 13.7 Mb, in which a CFS FRA11H was characterized by low resolution cytogenetic mapping (http://www.ncbi.nlm.nih.gov/gene/?term=FRA11H). However, whether this CFS contributes to the instability in 11q13 remains unclear.

Here we show that the recurrent instability leading to amplifications in 11q13 is within the common fragile site FRA11H. We further reveal that FRA11H is a preferential site for integration of MLV-based vectors, both *in vitro* and in gene therapy trials. More importantly, using the high resolution DNA combing approach we show that following infection by a MLV-based vector, the replication dynamics of the genomic DNA is significantly perturbed, leading to the formation of DNA double strand breaks (DSBs) and the expression of fragile sites on metaphase chromosomes. These findings shed new light on viral-host interactions, and raise new concerns regarding the use of viral vectors in gene therapy trials.

Results

In order to investigate whether the breakpoints leading to the recurrent amplifications in 11q13 are within the CFS, FRA11H, we analyzed constriction gaps and breaks on metaphase chromosomes from GM00847 cells treated with a low concentration of aphidicolin

(0.4 μM). Fluorescence in situ hybridization (FISH) was performed using BAC clones as probes. These clones flank the region of recurrent genomic instability found in different cancer types in 11q13 (Supplementary Table 1). The results revealed that both BAC clones show FISH signals proximal, distal or on both sides of the chromatin constrictions and gaps in the 11q13 region (Figure 1 and Table 1). These results indicate that a genomic region of ~7 Mb flanked by the analyzed probes is part of the CFS, FRA11H, and that this fragile site is probably involved in chromosome instability in different cancer types.

We further studied whether FRA11H is a preferential site for retroviral integrations. For this we first analyzed 1238 retroviral integration sites (RISs), mapped in cells of patients with X-linked

Table 1 | Summary of FISH results. The position of the FISH probes relative to the gaps and constrictions in 11q13. The results of the adjacent BACs RP11-456I5, RP11-707G14 and RP11-809N8 were pooled

BAC name	Telomeric	On	Split	Centromeric
RP11-15L8	19	3	-	3
RP11-149G19	16	-	1	3
RP11-456I5 RP11-707G14 RP11-809N8	8	3	-	9

SCID (SCID-X1) treated with a retrovirus-based gene therapy vector carrying the γ chain (γc) gene[26,27]. Our analysis revealed 10 retroviral integrations into the ~7 Mb of FRA11H (Figure 1), a significantly higher frequency than the frequency in the entire non-fragile regions in the genome (binomial test p < 0.001). We then analyzed more than 40,000 MLV integrations into human CD34+ cells recently mapped by Cattoglio et al.[28]. Ofthese, 2197 unique integrations were mapped on chromosome 11. This massive integration mapping allowed us to study the distribution of viral integrations into CFSs at a higher resolution and statistical power. First we analyzed viral integrations into FRA11H and compared this to the frequency of integrations along the entire chromosome 11 (excluding centromeric, telomeric sequences and other mapped CFSs: FRA11E, FRA11F and FRA11G (Supplementary Table 2)). Our analysis revealed that the 7 Mb FRA11H region harbors 170/2197 MLV integrations, which is significantly higher than the frequency in the 110 Mb non-fragile regions in chromosome 11 ($p = 1.63*10^{-3}$).

Previously, we analyzed a limited number of MLV integrations into another fragile site on human chromosome 11, FRA11E[21]. Here we analyzed the viral integration into FRA11E in the large integration database mapped by Cottoglio et al.[28]. The analysis revealed an extremely high MLV integration frequency, with 67 integrations into the 2 Mb of FRA11E, significantly different from the integration frequency into the entire chromosome 11 ($p = 2.12*10^{-30}$). These results are consistent with our previous report on a limited number of integration sites mapped in SCID patients, which indicated that FRA11H is a hotspot for MLV integrations in human CD34+ cells[21].

We further aimed to investigate the basis for the preferential integrations of retroviral MLV-based vectors into CFSs. CFSs are sensitive to replication stress conditions and under these conditions their chromatin is uncondensed even in mitosis. Hence, we raised the possibility that MLV retroviral infection leads to replication stress resulting in open chromatin conformation in fragile sites which might predispose these regions to the MLV integration protein complex[29]. To test this possibility we infected HeLa cells with an MLV-based vector (based on pTG-nls-LacZ plasmid), with high infection efficiency (~96%) (Figure 2). Using the high resolution DNA combing method, which enables replication analysis of single DNA molecules (an example is shown in Figure 3a), we analyzed the DNA replication rate 24 hours following infection, and compared it

to the rate in control cells, grown for 24 hours in a medium from packaging cells that does not produce virions. Our results revealed a significant reduction (~35%) in the replication rate following viral infection. The mean replication rate for the control HeLa cells was 1.32 ± 0.04 Kb/min (n = 111), whereas the MLV based-vector infection significantly decreased the rate to 0.86 ± 0.02 Kb/min (n = 237) ($p < 7.8*10^{-8}$) (Figure 3b). Interestingly, following MLV infection the distribution of fork rates changed dramatically. In the control cells, in 25% of the forks the rate was >1.5 Kb/min, whereas only 3.3% of the forks progressed at this high rate following infection with the MLV-based vector. In order to verify that MLV-induced replication stress is a general phenomenon and is not cell-specific, we further analyzed the replication dynamics in normal human fibroblast cells expressing hTERT (BJ). As was found for HeLa cells,
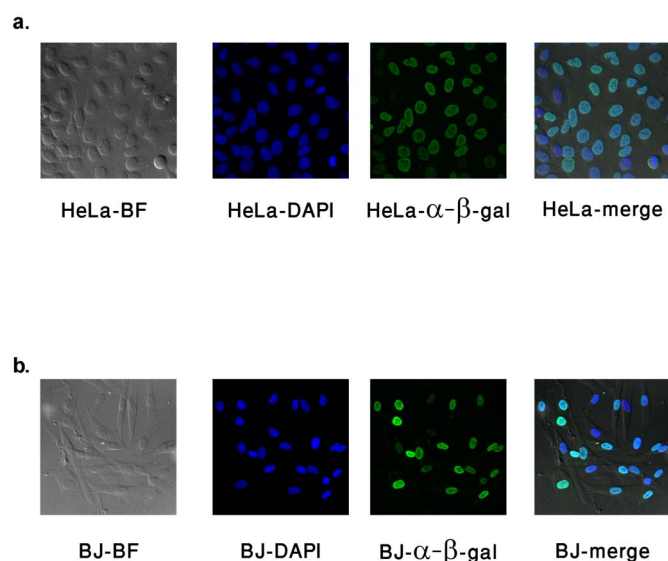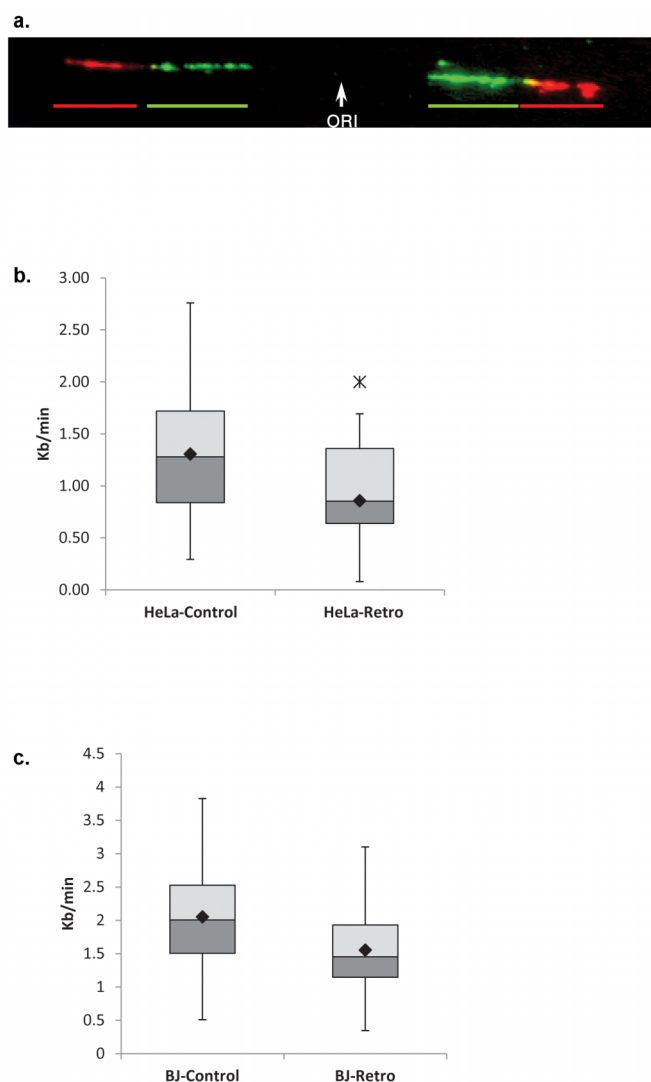

Figure 3 | The replication rate in HeLa and BJ cells 24 hours after infection with a retroviral-based vector. (a) Example of a single combed DNA molecule labeled with IdU (green) and CldU (red). The unstained gap represents the position of the origin (ORI). (b) Box plot of fork rates in HeLa cells infected by a retroviral-based vector (HeLa-Retro) (n = 237) and control cells (HeLa-control) (n = 197) (p-value = $7.8*10^{-8}$) (these results are a summary of two independent infections). (c) Box plot of fork rates in BJ cells infected by a retroviral-based vector (BJ-Retro) (n = 94) and control cells (BJ-control) (n = 108) (p-value = 1.188e-07); (these results are a summary of two independent infections). Lower hinge defines the 25th percentile, upper hinge defines the 75th percentile. The median is shown as a line across the box. Dimond shape indicates the average.

Figure 2 | Examples of infection efficiency by MLV-based vector, expressing nls-Lac Z in (a) HeLa, (b) BJ cells. Immunofluorescence with anti-β-gal (green), nuclei were stained with DAPI (Blue) BF - bright field cells.

the replication rate significantly (p-value = 1.188e-07) decreased following the MLV-based vector infection, from an average of 2.05 ± 0.07 Kb/min (n = 108) in the control non-infected cells, to 1.55 ± 0.06 Kb/min (n = 94) in BJ cells infected by the MLV-based vector (Figure 3c). Furthermore, in BJ infected by the MLV-based vector a dramatic (4- fold) increase in the fraction of slow forks (<1 Kb/min) was found, from 3.7%, in non-infected control cells to 14.8% in the infected cells. The perturbed replication dynamics described in BJ cells is clearly an under-estimation of the phenomenon, since only 65 ± 7% of the cells were infected by the vector (performed in three independent experiments), as revealed by X-gal positive staining. Altogether, our results indicate that a short time after retroviral-based vector infection, the DNA replication was dramatically perturbed. This replication perturbation may lead to the collapse of replication forks, resulting in DNA damage.

In order to study whether the replication perturbation caused by retroviral infection is the result of high viral titer, leading to high multiplicity of infection (MOI), we infected HeLa cells with GFP reporter gene, by a low MOI (1 MOI) of MLV-based vectors. 24 h post infection, cells were sorted by fluorescence-activated cell sorting (FACS), to differentiate GFP positive and negative cells. Replication analysis using DNA combing revealed a significantly slow fork progression rate in the GFP cells (average of 0.6 ± 0.02 Kb/min (n = 127)), compared to the GFP negative cells (average of 1.0 ± 0.04 Kb/min (n = 71)) (p-value = 1.441e-11).

Lentiviral-based vectors are considered as good candidate vectors for gene therapy trials due to their potential to mediate a prolonged therapeutic gene transfer, both in dividing and non-dividing cells. We further analyzed the effect of lentiviral infection on the replication dynamics in HeLa cells infected with a lenti Green Fluorescent Protein (GFP) vector. Our results show that even with a high titer (10 MOI), the replication fork rate remains similar to the observed rate in the non-infected (GFP-negative) cells (average of 1.0 ± 0.03 Kb/min (n = 107)) (p-value = 0.89) (Supplementary Figure 1). This result indicates that replication stress caused by MLV-based vectors, is not a general characteristic of viral infection.

We further explored whether the replication perturbation conferred by retroviral MLV vector infection leads to fragile site expression. For this we analyzed HeLa cells 24 hours after infection with a retroviral-based vector. Our results show that retroviral infection leads to a significant (P < 0.005) level of gaps and constrictions in the infected cells (1.93 CFSs/metaphase) compared to non-infected control cells (1.19 CFSs/metaphase) (Figure 4). Importantly, half of the retroviral-infected cells showed a significant number of fragile sites (≥2). This indicates that infection with a retroviral-based vector leads to the expression of CFSs as a result of replication stress.

In order to investigate the possibility that the replication perturbation induced by the retroviral MLV-based vector infection leads to accumulation of DSBs we analyzed γH2AX foci, which are known to be formed at DSBs (Figure 5a)[30,31] in HeLa cells 24 h post-infection with an MLV-based vector. Immunofluorescence analysis revealed significantly (p < 8.9*10^{-4}) more γH2AX foci in HeLa cells 24 h post infection by an MLV-based vector, with an average of 2.25 ± 0.4 foci/cell (n = 60) compared to 0.87 ± 0.2 foci/cell (n = 61) in the control cells (Figure 5b). Next we analyzed DSB-formation in BJ cells infected with an MLV-based vector. Because the infection in BJ cells was less efficient (only 67% of the cells were positive to the x-gal staining (data not shown), or the $\alpha\beta$−gal antibody (Figure 2b)), we analyzed γH2AX foci only in β−gal positive cells (Figure 5c). Immunofluorescence analysis revealed a significant (p < 5*10^{-14}) increase in γH2AX foci in BJ cells, 24 h post- infection by the MLV-based vector, with an average of 3.5 ± 0.2 foci/cell (n = 105), compared to 0.95 ± 0.2 foci/cell (n = 103) in the control cells (Figure 5d). In order to verify that the observed effects of the MLV-based vector infection resulted from the viral infection and not from the introduction of foreign DNA, we analyzed the formation of DNA damage
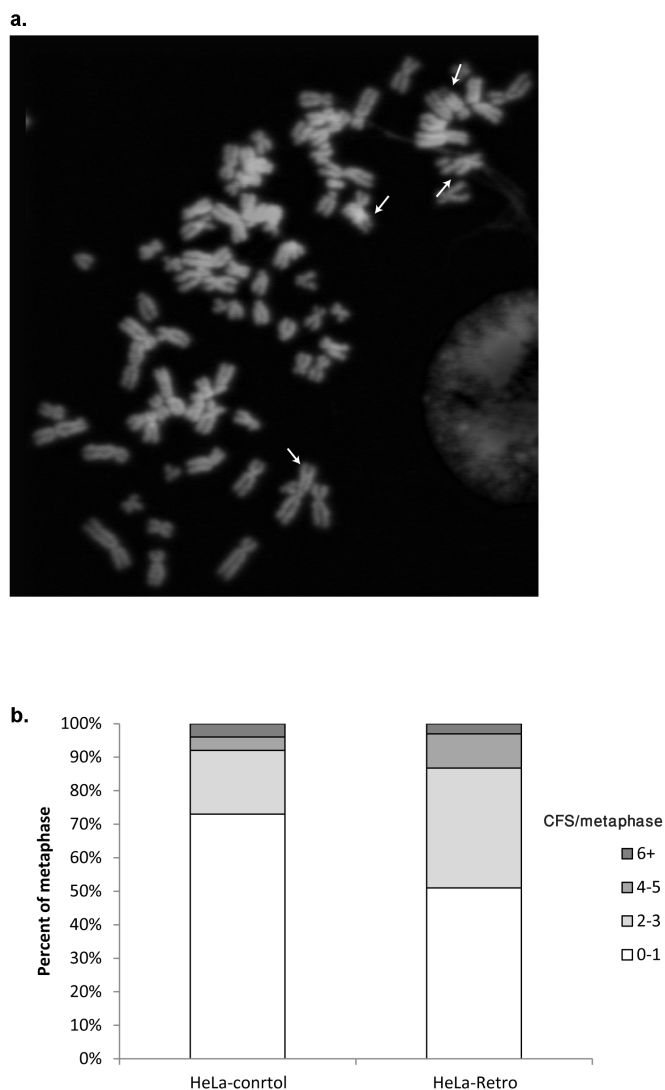

a.

b.

Figure 4 | CFS expression in metaphase chromosomes following retroviral vector infection. (a) Example of metaphase chromosomes from HeLa cells, 24 h after infection with a retroviral-based vector. White arrows indicate CFSs. (b) Distribution of metaphases according to the number of gaps and constrictions per metaphase in HeLa cells, 24 hours after retroviral-based vector infection (HeLa-Retro) (n = 100) or control cells (HeLa-control) (n = 100) (p < 0.005); (these results are a summary of two independent infections).

following tranfection with a plasmid DNA which is the backbone of the retroviral-based vector. As the transfection efficiency of the retroviral plasmid was only >20%, we transfected HeLa cells with the retroviral plasmid encoding GFP (pLVC-GFP). No significant difference (p = 0.67) in the number of γH2AX foci in the GFP positive cells (2.4 ± 0.3, n = 57) was found compared to the GFP negative cells (2.5 ± 0.2, n = 320) (Figure 6), although the transfection itself is toxic, resulting in DNA damage and cell death. These results indicate that transfection of plasmid DNA by itself does not lead to replication-induced DSBs, indicating that such DSBs are specific to the infection by retroviral virons.

In summary, our results reveal that retroviral infection leads to replication perturbation and DSBs suggesting a mechanism for the significantly higher frequency of retroviral integration into CFSs.

Discussion

Genome instability is a hallmark of cancer cells. One type of instability is amplifications that increase the expression of oncogenes. The
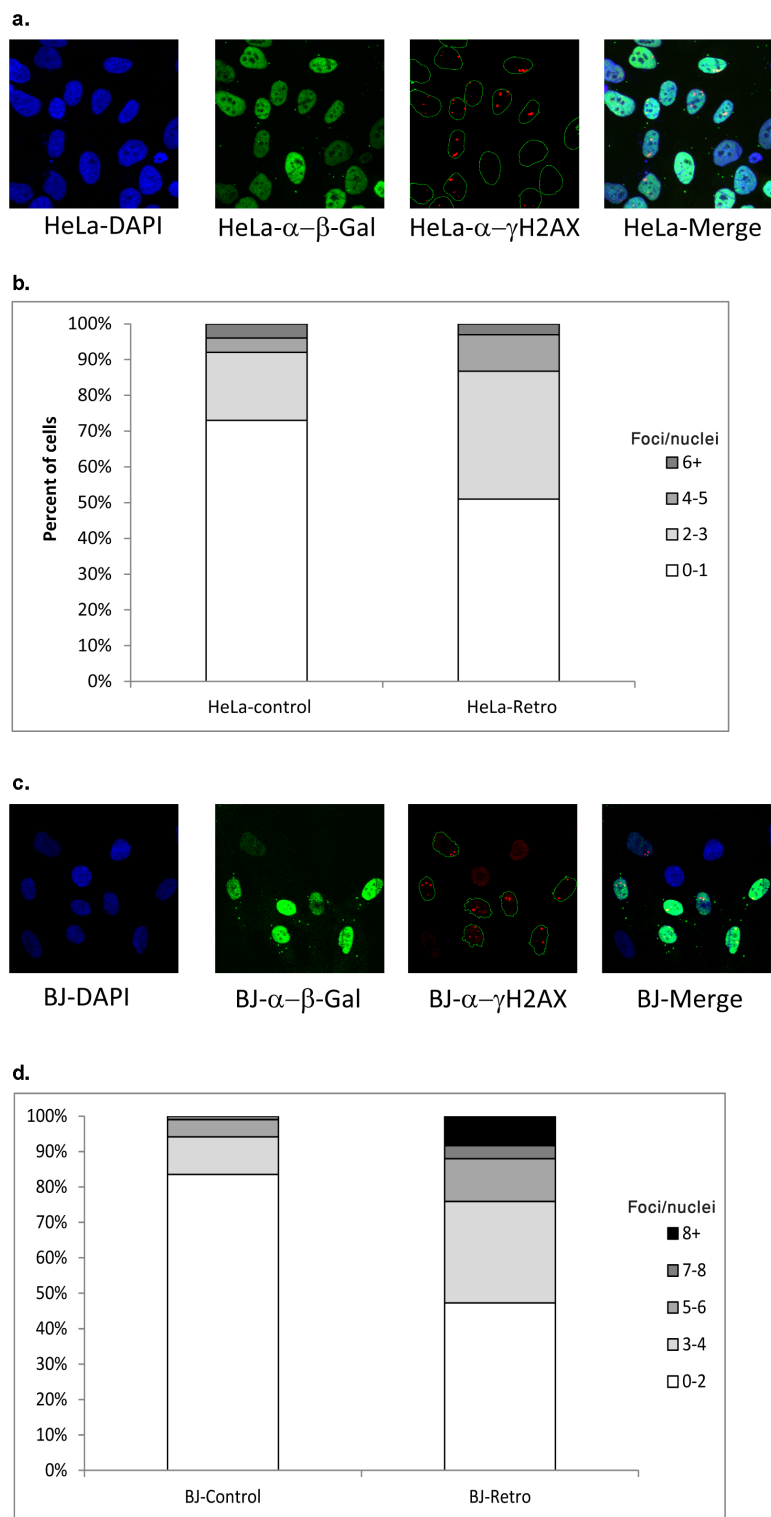
Figure 5 | DSBs accumulated following retroviral-based vector infection. (a) Examples of HeLa cells infected with MLV-based vector following immunofluorescence with anti-β-gal (green) and γH2AX (red). The green outlines mark positive β-gal cells. Nuclei were stained with DAPI (blue). (b) Percent of nuclei with the indicated number of γH2AX foci. HeLa infected by a retroviral based vector (HeLa-Retro, n = 60), control cells (HeLa-control) (control, n = 60), (p < 8.9*10⁻⁴). (c) Examples of BJ cells infected with MLV-based vector following immunofluorescence with anti-β-gal (green) and γH2AX (red). The green outlines mark positive β-gal cells. Nuclei were stained with DAPI (blue). (d) Percent of nuclei with the indicated number of γH2AX foci. BJ cells infected by a retroviral-based vector (BJ-Retro, n = 105), control cells (BJ-control) (control, n = 103), (p < 5*10⁻¹⁴); (these results are a summary of two independent infections).

a.

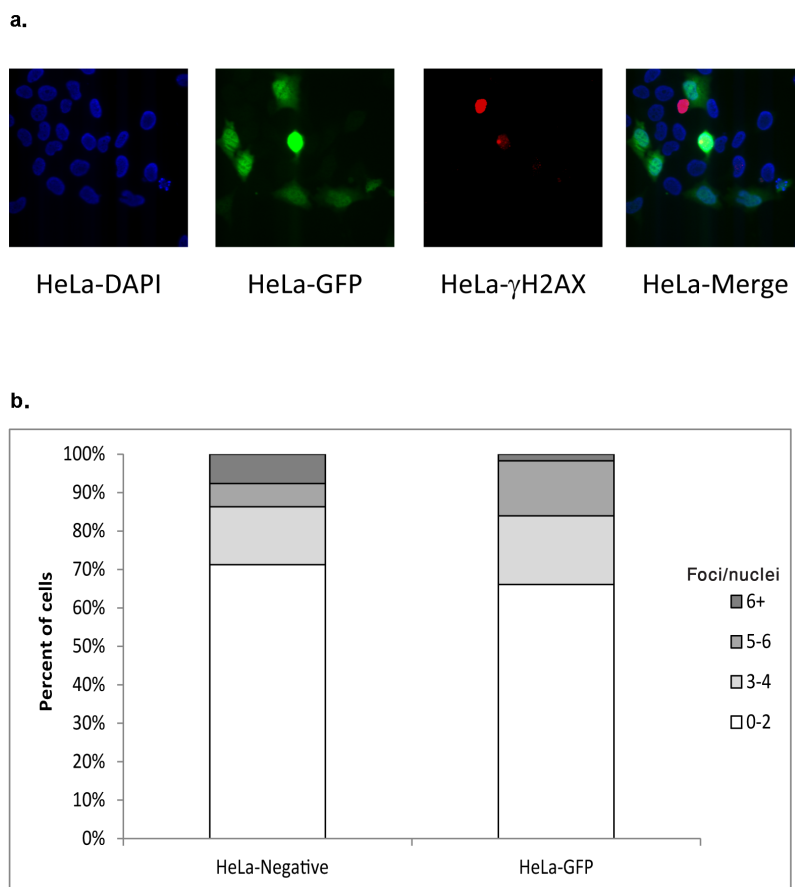HeLa-DAPI   HeLa-GFP   HeLa-γH2AX   HeLa-Merge

b.

Figure 6 | DSBs following transfection with a plasmid DNA encoding for GFP. (a) Examples of HeLa cells infected with MLV-based vector following immunofluorescence with anti-GFP (green) and γH2AX (red). The nuclei were stained with DAPI (blue). (b) Percent of nuclei with the indicated number of γH2AX foci. HeLa cells expressing GFP following transfection (HeLa-GFP, n = 57), tranfected cells that do not express GFP (HeLa-Negative) (control, n = 320), (p = 0.67); (these results are a summary of two independent transfections).

high frequency of amplifications found in 11q13 indicates that this region harbors an oncogene/s which may give a proliferative advantage to the cancer cells. Indeed FRA11H harbors the oncogene *CCND1* that encodes the cyclin D protein, which is involved in different types of cancer[32]. However, the 11q13 amplified region in these cancers is significantly larger than the oncogene, indicating that in addition to the selection force, other mechanisms contribute to 11q13 instability. One such mechanism involves the inherent instability of CFS under replication stress conditions. Instability in CFSs was shown to drive genomic amplifications by setting the boundaries of the amplicons. This mechanism was described for the amplification of different oncogenes including *MET* (FRA7G)[12,14,33], *PIP* (FRA7I)[34], *SMAD5* (FRA5C)[35] among others[36,37]. The preferential breaks promoting genomic amplifications within the replication-sensitive CFSs indicate that the instability in these sites involves replication stress conditions. Several studies have indicated that there is replication stress at early stages of cancer development[16,38] which is a result of aberrant activation of oncogenes[10,15,39]. Our results revealed a large fragile site (at least 7 Mb long), FRA11H, in 11q13, which may explain the large amplifications at this site in many cancers.

A different type of genome instability is the integration of foreign DNA into the human genome. Retroviral integrations are mediated by viral and cellular proteins, and are distributed along the genome with a preference for transcription start sites (TSS) and regulatory elements[28,40,41] and DNase I hypersensitive sites[42]. Large scale analyses of retroviral based vectors have found preferential integration of viral DNA into CFSs[19,21]. Here we showed that the CFS FRA11H is a

preferential site for MLV-based vectors. Many fragile sites are involved in genomic instability in cancer[43]. It is important to note that 11q13, the chromosomal band harboring FRA11H, is a known hotspot for genomic instability in different types of cancer (Supplementary Table 1), although further studies are required to shed light on the importance of FRA11H sequences as a hotspot for genomic instability in cancer.

We further showed that *in vitro* infection by a retroviral-based vector leads to perturbed DNA replication of the host cell, as indicated by the decrease in replication rate and the increase in DNA damage (Figures 2 and 4). The infection also leads to CFS expression (Figure 4), which appears as uncondenced open chromatin at metaphase chromosomes.

The uncondensed chromatin at CFSs compared to the non-fragile DNA might lead to higher accessibility of these regions to the viral integration complex, which may account for the high number of MLV integrations into FRA11H and FRA11E. Integration to the open chromatin of CFSs is consistent with previous results showing preferential MLV integration into other open chromatin regions in the genome, such as TSS[40], regulatory elements[28] and DNase I hypersensitive sites[42].

Altogether, our findings showing that *in vitro* infection by retrovirus vectors leads to replication perturbations raises new concerns regarding the safety of these vectors. This replication perturbation may increase genome instability and mutation rates that could lead to tumorigenicity. This risk is greater especially in patients with challenged immune systems, such as those with SCID. Hence, future studies should explore the molecular basis of this replication stress,

and the ways in which it affects specific risks associated with gene therapy trials and treatments.

Methods

Cells and growth conditions. The simian virus 40-transformed human fibroblast cell line GM00847 (Coriell Cell Repository, Camden, N.J.) was grown in Eagle minimal essential medium supplemented with 10% fetal calf serum 100,000 U/l penicillin and100 mg/l streptomycin. HeLa cells were grown in Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum, 100,000 U/l penicillin and100 mg/l streptomycin.

Retroviral-based vector infection and tranfection. For the MLV-based vector infection, we used packaging cells stably expressing a pseudo-type gibbon ape leukemia virus (GALV) envelope (TGA), and the pTG-nls-LacZ plasmid which is a MLV-based vector (a gift from Prof. Moshe Y. Flugelman, Technion, Israel), and from Prof. Bertrand Boson (Université de Lyon, Lyon, France.)[43].

Evaluation of the LacZ gene transfer to HeLa cells was performed by X-GAL staining.

For non-viral expression of GFP pLVC-GFP plasmid, we transfected cells using the TransIT-HeLaMONSTER Transfection Kit (Mirus Bio LLC).

Preparation of cells for cytogenetic analysis. Cells infected with the MLV-based vector were grown in 2.2 mM caffeine (Sigma-Aldrich) for 24 h before fixation in order to avoid cell cycle arrest conferred by replication-induced DNA damage during the S-phase. On the day of fixation, cells were grown with 0.1 ug/ml Colcemid for 40 minutes, trypsinized and incubated with 0.4% KCl in 37°C for 40 minutes. Cells were fixed with methanol- acetic acid (3 : 1) and spread by dropping on cover slips.

Fluorescent in situ hybridization (FISH). DNA BAC clones were labeled with digoxigenin DY-505-aadUTP (Applied Spectral Imaging) by nick translation. FISH on metaphase chromosomes was performed as previously described[44].

Cytogenetic analysis of hybridization signals and fragile site expression. Green and red fluorescence were visualized using a Nikon B-2A filter cube. For weak signals a modified Chromatech HQ-FITC (Chroma Technology, Brattleboro, Vt.) filter set was used (excitation, 460 to 500 nm; emission, 520 to 600 nm). Images were captured with an intensified charge-coupled device imager (Paultek Imaging, Grass Valley, Calif.) and digitized with a frame grabber (Imascan/MONO-D; Imagraph, Chelmsford, Mass.). The Image-Pro PLUS program (Media Cybernetics, Silver Spring, Md.) was used to measure the fragile site-telomere distance relative to the total length of the p arm of chromosome 11 and compared it to the GDB mapping of the fragile sites, as previously described[45].

Molecular combing. Unsynchronized cells were pulse labeled for 30 min by a medium containing 100 mM of the thymidine analog iododeoxyuridine (IdU). At the end of the first labeling period, the cells were washed twice with a warm medium and pulse labeled once more for 30 min with a medium containing 100 mM of another thymidine analog, chlorodeoxyuridine (CldU). Cells were then harvested, and genomic DNA was extracted, combed, and analyzed as previously described[46,47]. The primary antibody for fluorescence detection of IdU was mouse anti-BrdU (Becton Dickinson), and the secondary antibody was goat anti-mouse Alexa Fluor 488 (Invitrogen). The primary antibody for fluorescence detection of CldU was rat anti-CldU (Novus Biologicals). The secondary antibody was goat anti-rat Alexa Fluor 594 (Invitrogen). The length of the replication signals was measured in micrometers and converted to kilo bases according to a constant and sequence-independent stretching factor (1 mm = 2 Kb), as previously reported[46].

Statistical analysis. A one-tailed t-test was performed for the replication rate analysis, γH2AX foci and chromosomal CFS expression. A binomial test was used to examine the frequency of MLV integrations into CFSs, compared to non-fragile regions. Statistical analysis was performed on at least two biological repeats.

Immunofluorescence for detection of γH2AX foci and βgal. Cells were fixed in 3.7% formaldehyde/PBS for 10 min, permeabilized with 0.5% Triton/PBS, and blocked with 5% BSA/PBS. The primary γH2AX antibodies used were mouse anti-γH2AX (Upstate Biotechnology). Appropriate secondary antibodies were added, Cy3 conjugated (Jackson ImmunoResearch Laboratories) or Alexa Fluor 594 (Invitrogen). For β-gal analysis we used anti- β-Galactosidase (Promega) as a primary antibodies, and Alexa Fluor 488 (Invitrogen) as a secondary antibodies. Images were taken with a Bio-Rad confocal microscope. For focus information analysis at least 50 nuclei for each condition were analyzed.

1. Glover, T. W., Berger, C., Coyle, J. & Echo, B. DNA polymerase alpha inhibition by aphidicolin induces gaps and breaks at common fragile sites in human chromosomes. *Hum Genet* **67**, 136–142 (1984).
2. Cheng, C. H. & Kuchta, R. D. DNA polymerase epsilon: aphidicolin inhibition and the relationship between polymerase and exonuclease activity. *Biochemistry* **32**, 8568–8574 (1993).
3. Ikegami, S. *et al.* Aphidicolin prevents mitotic cell division by interfering with the activity of DNA polymerase-alpha. *Nature* **275**, 458–460 (1978).
4. Glover, T. W. & Stein, C. K. Induction of sister chromatid exchanges at common fragile sites. *Am J Hum Genet* **41**, 882–890 (1987).
5. Glover, T. W. & Stein, C. K. Chromosome breakage and recombination at fragile sites. *Am J Hum Genet* **43**, 265–273 (1988).
6. Wang, L. *et al.* Aphidicolin-induced FRA3B breakpoints cluster in two distinct regions. *Genomics* **41**, 485–488 (1997).
7. Bignell, G. R. *et al.* Signatures of mutation and selection in the cancer genome. *Nature* **463**, 893–8 (2010).
8. Hecht, F., Morgan, R., Hecht, B. K. & Smith, S. D. Common region on chromosome 14 in T-cell leukemia and lymphoma. *Science (New York, N.Y.)* **226**, 1445–7 (1984).
9. Yunis, J. J. J. & Soreng, A. L. Constitutive Fragile Sites and Cancer. *Science* **226**, 1199–1204 (1984).
10. Bester, A. C. *et al.* Nucleotide deficiency promotes genomic instability in early stages of cancer development. *Cell* **145**, 435–46 (2011).
11. Bartkova, J. *et al.* DNA damage response as a candidate anti-cancer barrier in early human tumorigenesis. *Nature* **434**, 864–870 (2005).
12. Hellman, A. *et al.* A role for common fragile site induction in amplification of human oncogenes. *Cancer Cell* **1**, 89–97 (2002).
13. Kotzot, D. *et al.* Parental origin and mechanisms of formation of cytogenetically recognisable de novo direct and inverted duplications. *J Med Genet* **37**, 281–286 (2000).
14. Miller, C. T. *et al.* Genomic amplification of MET with boundaries within fragile site FRA7G and upregulation of MET pathways in esophageal adenocarcinoma. *Oncogene* **25**, 409–418 (2006).
15. Di Micco, R. *et al.* Oncogene-induced senescence is a DNA damage response triggered by DNA hyper-replication. *Nature* **444**, 638–42 (2006).
16. Gorgoulis, V. G. *et al.* Activation of the DNA damage checkpoint and genomic instability in human precancerous lesions. *Nature* **434**, 907–13 (2005).
17. Rassool, F. V. *et al.* Preferential integration of marker DNA into the chromosomal fragile site at 3p14: an approach to cloning fragile sites. *Proceedings of the National Academy of Sciences of the United States of America* **88**, 6657–61 (1991).
18. Lehoux, M., D'Abramo, C. M. & Archambault, J. Molecular mechanisms of human papillomavirus-induced carcinogenesis. *Public health genomics* **12**, 268–80 (2009).
19. Kim, S. *et al.* Integration site preference of xenotropic murine leukemia virus-related virus, a new human retrovirus associated with prostate cancer. *J Virol* **82**, 9964–9977 (2008).
20. Urisman, A. *et al.* Identification of a novel Gammaretrovirus in prostate tumors of patients homozygous for R462Q RNASEL variant. *PLoS pathogens* **2**, e25 (2006).
21. Bester, A. C. *et al.* Fragile sites are preferential targets for integrations of MLV vectors in gene therapy. *Gene therapy* **13**, 1057–9 (2006).
22. Howe, S. J. *et al.* Insertional mutagenesis combined with acquired somatic mutations causes leukemogenesis following gene therapy of SCID-X1 patients. *The Journal of clinical investigation* **118**, 3143–50 (2008).
23. Hacein-Bey-Abina, S. *et al.* Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. *The Journal of clinical investigation* **118**, 3132 (2008).
24. Huang, X., Gollin, S. M., Raja, S. & Godfrey, T. E. High-resolution mapping of the 11q13 amplicon and identification of a gene, TAOS1, that is amplified and overexpressed in oral cancer cells. *Proc Natl Acad Sci U S A* **99**, 11369–11374 (2002).
25. Zhang, D. *et al.* Analysis of DNA copy number aberrations by multiple ligation-dependent probe amplification on 50 intestinal type gastric cancers. *Journal of surgical oncology* **103**, 124–32 (2011).
26. Deichmann, A. *et al.* Vector integration is nonrandom and clustered and influences the fate of lymphopoiesis in SCID-X1 gene therapy. *J Clin Invest* **117**, 2225–2232 (2007).
27. Schwarzwaelder, K. *et al.* Gammaretrovirus-mediated correction of SCID-X1 is associated with skewed vector integration site distribution in vivo. *J Clin Invest* **117**, 2241–2249 (2007).
28. Cattoglio, C. *et al.* High-definition mapping of retroviral integration sites identifies active regulatory elements in human multipotent hematopoietic progenitors. *Blood* **116**, 5507–17 (2010).
29. Panet, A. & Cedar, H. Selective degradation of integrated murine leukemia proviral DNA by deoxyribonucleases. *Cell* **11**, 933–940 (1977).
30. Wang, B., Matsuoka, S., Carpenter, P. B. & Elledge, S. J. 53BP1, a mediator of the DNA damage checkpoint. *Science (New York, N.Y.)* **298**, 1435–8 (2002).
31. Rogakou, E. P., Pilch, D. R., Orr, A. H., Ivanova, V. S. & Bonner, W. M. DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. *J Biol Chem* **273**, 5858–5868 (1998).
32. Musgrove, E. A., Caldon, C. E., Barraclough, J., Stone, A. & Sutherland, R. L. Cyclin D as a therapeutic target in cancer. *Nature reviews. Cancer* **11**, 558–72 (2011).
33. Coquelle, A., Pipiras, E., Toledo, F., Buttin, G. & Debatisse, M. Expression of fragile sites triggers intrachromosomal mammalian gene amplification and sets boundaries to early amplicons. *Cell* **89**, 215–225 (1997).
34. Ciullo, M. *et al.* Initiation of the breakage-fusion-bridge mechanism through common fragile site activation in human breast cancer cells: the model of PIP gene duplication from a break at FRA7I. *Human molecular genetics* **11**, 2887–94 (2002).

35. Zimonjic, D. B. *et al*. SMAD5 gene expression, rearrangements, copy number, and amplification at fragile site FRA5C in human hepatocellular carcinoma. *Neoplasia (New York, N.Y.)* **5**, 390–6 (2003).

36. Pelliccia, F., Bosco, N. & Rocchi, A. Breakages at common fragile sites set boundaries of amplified regions in two leukemia cell lines K562 - Molecular characterization of FRA2H and localization of a new CFS FRA2S. *Cancer letters* **299**, 37–44 (2010).

37. Reshmi, S. C. *et al*. Relationship between FRA11F and 11q13 gene amplification in oral cancer. *Genes Chromosomes Cancer* **46**, 143–154 (2007).

38. Tsantoulis, P. K. *et al*. Oncogene-induced replication stress preferentially targets common fragile sites in preneoplastic lesions. A genome-wide study. *Oncogene* **27**, 3256–64 (2008).

39. Bartkova, J. *et al*. Oncogene-induced senescence is part of the tumorigenesis barrier imposed by DNA damage checkpoints. *Nature* **444**, 633–7 (2006).

40. Wu, X., Li, Y., Crise, B. & Burgess, S. M. Transcription start regions in the human genome are favored targets for MLV integration. *Science (New York, N.Y.)* **300**, 1749–51 (2003).

41. Mitchell, R. S. *et al*. Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. *PLoS Biol* **2**, E234 (2004).

42. Rohdewohld, H., Weiher, H., Reik, W., Jaenisch, R. & Breindl, M. Retrovirus integration and chromatin structure: Moloney murine leukemia proviral integration sites map near DNase I-hypersensitive sites. *Journal of virology* **61**, 336–43 (1987).

43. Koren, B. *et al*. Efficient transduction and seeding of human endothelial cells onto metallic stents using bicistronic pseudo-typed retroviral vectors encoding vascular endothelial growth factor. *Cardiovascular Revascularization Medicine* **7**, 173–178 (2006).

44. Lichter, P., Cremer, T., Borden, J., Manuelidis, L. & Ward, D. C. Delineation of individual human chromosomes in metaphase and interphase cells by in situ suppression hybridization using recombinant DNA libraries. *Hum Genet* **80**, 224–234 (1988).

45. Mishmar, D. *et al*. Molecular characterization of a common fragile site (FRA7H) on human chromosome 7 by the cloning of a simian virus 40 integration site. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 8141–6 (1998).

46. Herrick, J. & Bensimon, A. Single molecule analysis of DNA replication. *Biochimie* **81**, 859–871 (1999).

47. Lebofsky, R., Heilig, R., Sonnleitner, M., Weissenbach, J. & Bensimon, A. DNA replication origin interference increases the spacing between initiation events in human cells. *Molecular biology of the cell* **17**, 5337–45 (2006).

Acknowledgments

Author contributions

A.C.B. and B.K. designed the experiments and wrote the manuscript. A.C.B. performed all the experiments presented in the manuscript. M.K. assited in the cloning of the FRA11H region. K.M. assisted in the DNA combing experiments and analyses.

Additional information

Supplementary information accompanies this paper at http://www.nature.com/scientificreports

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Bester, A.C., Kafri, M., Maoz, K. & Kerem, B. Infection with retroviral vectors leads to perturbed DNA replication increasing vector integrations into fragile sites. *Sci. Rep.* **3**, 2189; DOI:10.1038/srep02189 (2013).