Published online 25 October 2008

Nucleic Acids Research, 2008, Vol. 36, No. 22 ***e148***
doi:10.1093/nar/gkn801

Bacteriophage Mu integration in yeast and mammalian genomes

Anja O. Paatero[1], Hilkka Turakainen[1], Lotta J. Happonen[1], Cia Olsson[2], Tiina Palomäki[2], Maria I. Pajunen[1,3], Xiaojuan Meng[4,†], Timo Otonkoski[2,5], Timo Tuuri[2,6], Charles Berry[7], Nirav Malani[7], Mikko J. Frilander[4], Frederic D. Bushman[7] and Harri Savilahti[1,3,*]

[1]Program in Cellular Biotechnology, Institute of Biotechnology, Viikki Biocenter, [2]Biomedicum Stem Cell Center, Biomedicum Helsinki, University of Helsinki, Helsinki, [3]Division of Genetics and Physiology, Department of Biology, University of Turku, Turku, [4]Program in Developmental Biology, Institute of Biotechnology, Viikki Biocenter, University of Helsinki, [5]Hospital for Children and Adolescents, University of Helsinki, [6]Family Federation of Finland, Helsinki, Finland and [7]Department of Microbiology, University of Pennsylvania School of Medicine, Philadelphia, PA, USA

Received August 26, 2008; Revised October 9, 2008; Accepted October 10, 2008

ABSTRACT

Genomic parasites have evolved distinctive lifestyles to optimize replication in the context of the genomes they inhabit. Here, we introduced new DNA into eukaryotic cells using bacteriophage Mu DNA transposition complexes, termed 'transpososomes'. Following electroporation of transpososomes and selection for marker gene expression, efficient integration was verified in yeast, mouse and human genomes. Although Mu has evolved in prokaryotes, strong biases were seen in the target site distributions in eukaryotic genomes, and these biases differed between yeast and mammals. In *Saccharomyces cerevisiae* transposons accumulated outside of genes, consistent with selection against gene disruption. In mouse and human cells, transposons accumulated within genes, which previous work suggests is a favorable location for efficient expression of selectable markers. Naturally occurring transposons and viruses in yeast and mammals show related, but more extreme, targeting biases, suggesting that they are responding to the same pressures. These data help clarify the constraints exerted by genome structure on genomic parasites, and illustrate the wide utility of the Mu transpososome technology for gene transfer in eukaryotic cells.

INTRODUCTION

The replication cycles of genomic parasites have evolved to optimize replicative success in the context of the host cells they inhabit (1,2). For example, the attachment sites for integrating bacteriophages are commonly located in regions of bacterial genomes that can tolerate insertion without adverse effects on the host cell (1,2). Here we probe the relationship between genome organization and the consequences of integration experimentally. We compared DNA integration in the yeast, murine and human genomes, which differ radically in structure. The *Saccharomyces cerevisiae* genome is composed of ~69% coding regions, and introns are rare (3). In contrast, in the mammalian genomes, transcription units comprise a minority of the sequence. In humans, ~30% of the genome is contained within transcription units, and only 1.5% of the genome is exons, the remainder of genes being intronic (4,5). We used the bacterial transposon phage Mu to integrate new sequences into the three genomes and found that the recovered sites of Mu integration differed radically between yeast and mammals, disclosing differing selective forces that can be understood in terms of differing genome structure.

Bacteriophage Mu replicates its genome using DNA transposition, and it is one of the most thoroughly studied mobile genetic elements (6,7). The catalysis of the DNA-breaking and -joining steps involved in Mu transposition takes place within a nucleoprotein complex called the

*To whom correspondence should be addressed. Tel: +358 2 333 5586; Fax: +358 2 333 6680; Email: harri.savilahti@utu.fi
Present address:
Charles Berry, Department of Family and Preventive Medicine, School of Medicine, University of California San Diego, La Jolla, CA, USA
†Deceased

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

Mu transpososome (8,9), which can be assembled in a minimal *in vitro* system using MuA transposase protein and linear DNA molecules containing a 50-bp Mu R-end segment in each end (8,10,11). The transpososome directs transposon integration into any target DNA with a low stringency of target sequence preference, and this reaction constitutes the basis of Mu *in vitro* transposition technology (11). The technology has been used for DNA sequencing (12), functional analyses of plasmid DNA and virus genomes (11,13–17), protein engineering for structure/function studies (18–21) and generation of gene targeting constructions (22,23). In bacteria, *in vitro*-assembled Mu transpososomes introduced by electroporation are able to direct correct Mu integration, as has been shown with several bacterial species (24,25). Here we show for the first time that Mu transpososome-mediated gene delivery can be used in eukaryotic cells, including yeast, mouse ES cells, human HeLa cells and human ES cells.

For each species, custom Mu transpososomes were generated that directed integration of selectable marker genes for each cell type. Thus, the cells surviving selection in culture have been selected for both efficient expression of the marker gene and proper function of the cellular genome after integration. Integration sites were recovered from each cell type after selection, mapped by sequencing and their locations were compared to previously mapped genomic features. We found extreme differences among cell types in the frequency of integration compared to identifiable genomic landmarks. In yeast, integration sites preferentially accumulated outside of genes compared to the expectation for random integration, while in mammals, integration sites accumulated in introns within genes. These trends disclose the different requirements for genome function and gene expression in yeasts versus mammals that govern the accumulation of newly integrated DNA.

MATERIALS AND METHODS

Strains and media

Eschericia coli DH5α (Life Technologies//Invitrogen, Carlsbad, CA, USA) was used as a standard plasmid host strain. It was grown in Luria Broth with supplementary antibiotics when required: ampicillin (100 μg/ml), kanamycin (10 μg/ml).

Diploid *S. cerevisiae* strain FY1679 (*MATa/MATα ura3-52/ura3-52 his3Δ200/HIS3 leu2Δ1/LEU2 trp1Δ63/TRP1 GAL2/GAL2*; (26) and its haploid derivative FY-3a (*MATa ura3-52*; isolated in this study by random spore analysis) were grown on YPD (1% yeast extract, 2% peptone and 2% glucose) or minimal medium (0.67% yeast nitrogen base and 2% glucose).

HeLa cells were grown in MEM supplemented with 10% fetal calf serum, penicillin–streptomycin and L-glutamine at 37°C in the presence of 5% CO_2. Mouse AB2.2-Prime embryonic stem cells (Lexicon Genetics, Inc., The Woodlands, TX, USA) were grown on gelatinized tissue culture dishes in knockout DMEM supplemented with 15% defined fetal calf serum, penicillin–streptomycin, L-glutamine, non-essential amino acids, sodium pyruvate, mercaptoethanol and 500 U/ml LIF (leukemia inhibitory factor) at 37°C in the presence of 5% CO_2. Human FES29 embryonic stem cells were grown on MEF feeders as described (27). The number of live mammalian cells was determined by trypan blue staining.

Enzymes, reagents and DNA techniques

Commercial proteins and reagents are listed in Supplementary Table 1. MuA transposase proteins were purified in collaboration with Finnzymes (Espoo, Finland) as described (28). Oligonucleotides are listed in Supplementary Table 2. Plasmids and yeast DNA were isolated using appropriate Qiagen kits, and mammalian DNA as described (29). Standard DNA techniques were performed as described (30). DNA sequence determination was done at the sequencing facility of the Institute of Biotechnology. Southern hybridization was carried out using [α-^{32}P]-labeled probes. Autoradiography was done using BAS-1500 or FLA-5000 image readers (Fuji, Stamford, CT, USA).

Plasmids and transposons

The yeast electroporation control plasmid pHTH36 is a G418-selectable version of pYC2/CT (Invitrogen, Carlsbad, CA, USA), and it was made by kanMX4-Mu *in vitro* transposition. *In vitro* transposition assay target plasmid pAPH7 is a derivative of pUC19 (New England Biolabs, Ipswich, MA, USA), containing a cloned 6.6-kb HindIII fragment of phage λ. Transposons (details in Supplementary Figure 1) were constructed in the context of their carrier plasmids, isolated by BglII digestion and purified by anion exchange chromatography as described (11).

Transpososomes

Transpososomes were assembled as described (24). The assembly reaction (80 μl) contained 55 nM transposon DNA fragment, 245 nM MuA, 150 mM Tris–HCl pH 6.0, 50% (v/v) glycerol, 0.025% (w/v) Triton X-100, 150 mM NaCl and 0.1 mM EDTA. The reaction was carried out at 30°C for 4–5 h. Following assembly, several reactions were pooled, and reaction products were concentrated (~10-fold) and desalted using Centricon YM-100 centrifugal cartridges as described (25). Alternatively, the concentration step was done using polyethylene glycol (PEG 6000) precipitation as described (31). The pellet was resuspended in storage buffer (10 mM Tris–HCl pH 6.0, 0.5% glycerol, 0.1 mM DTT). The assembly and concentration of transpososomes were monitored by agarose/BSA/heparin gels as described (24). Preparations were frozen in liquid nitrogen and stored at –80°C. The transpositional activity of these preparations could be measured by an *in vitro* transposition assay using a selectable target plasmid and introduction of the reaction products into competent *E. coli* cells (11). The preparations scored roughly 1×10^6 CFU/μg transposon DNA when measured using 3.44-μg target plasmid (pAPH7) and

cells with a competence status of $\sim 4 \times 10^8$ CFU/μg pUC19 DNA.

Yeast electrocompetent cells and their electroporation

Yeast cells were grown in YPD medium to a stationary phase, diluted 1:10 000 in fresh medium and grown to an OD_{600} of ~2. Cells were collected by centrifugation, resuspended in 1/4 culture volume of LiAc/DTT/TE (0.1 M lithium acetate, 10 mM dithiotreitol, 10 mM Tris–HCl pH 7.5 and 1 mM EDTA), and incubated for 1 h at room temperature. After centrifugation for 10 min (3500 *g*), cells were washed with an equal volume of ice-cold water, re-collected by centrifugation and resuspended in 1/10 of the original culture volume of ice-cold 1 M sorbitol. After re-collection by centrifugation, the pellet was suspended in ice-cold 1 M sorbitol to yield a ~200-fold concentration of the original culture density. One hundred microliters of fresh cell suspension ($\sim 1 \times 10^8$ cells) was used for electroporation. An aliquot (1–2 μl) of transpososome preparation (containing 1 μg transposon DNA) or plasmid DNA (20 ng) was added to the cell suspension. Following incubation on ice (5 min), the mixture was transferred into a chilled Bio-Rad cuvette (0.2-cm electrode spacing), and electroporation was carried out using Genepulser II (Bio-Rad, Hercules, CA, USA) with the following settings: voltage 1.5 kV (diploid strain FY1679) or 2.0 kV (haploid strain FY-3a); capacitance 25 μF; resistance 200 Ω. Following electroporation, 0.9 ml of YPD medium was added, and cells were incubated for 2 h at 30°C. Cells were plated on YPD plates containing G418 (geneticin, 200 μg/ml) for the selection of the transposon marker gene. Under these conditions, ~5–25% of the cells survived electroporation (data not shown). The optimized protocol yielded electroporation efficiencies for diploid and haploid cells that routinely reached the levels of $\sim 1 \times 10^6$ and $\sim 3 \times 10^6$ CFU/μg of introduced plasmid (pYC2/CT) DNA, respectively. The reason for the apparent efficiency difference between the two strains is not known but may be related to differences in cell size or rigidity. We found that when measured with plasmid pHTH36, using G418 selection, the apparent efficiencies were four to six times smaller (Table 3), most probably reflecting the physiological dissimilarity of the selection schemes used.

Electroporation of mammalian cells

HeLa and AB2.2 cells were harvested with trypsin-EDTA, pH 7.4 and washed three times with 1 × PBS (137 mM NaCl, 2.7 mM KCl, 4.3 mM Na_2HPO_4 and 1.47 mM KH_2PO_4). Standard electroporation mixtures contained $\sim 1 \times 10^6$ HeLa cells or $\sim 1 \times 10^7$ AB2.2 cells in 800 μl of 1 × PBS and 2 μg of transposon DNA complexed with MuA transposase. The cells were exposed to a single voltage pulse (250 V; 500 μF) at room temperature, allowed to remain in the cuvette (0.4-cm electrode spacing) for 10 min and then plated onto tissue culture dishes. Selective conditions consisted of 400 μg/ml G418 for HeLa cells and 150 μg/ml G418 for AB2.2 cells. Selection was initiated 2 days after electroporation, and G418-resistant colonies were isolated for clonal propagation after ~10 days of selection. For the enumeration of colonies, cells were fixed with cold methanol and stained with 0.2% methylene blue.

Human ES cells were detached using 1 × Tryple select (Gibco/Invitrogen, Carlsbad, CA, USA) according to the supplier's recommendations and resuspended in hESC medium (27). Cells were then washed three times with 1 × PBS. Transpososomes (5.0 μg) were mixed with $\sim 3 \times 10^6$ cells (800 μl, in 1 × PBS) in a 0.4-cm electrode spacing cuvette and a single voltage pulse (320 V; 200 μF) was given immediately. Medium (800 μl) was then added, and after a 5-min incubation at room temperature, the cells were plated on feeders. Selection with puromycin (0.5 μg/ml) was initiated 2 days after electroporation and continued until visible colonies appeared (~7 days). Propagation and enumeration of colonies was done as described above.

Determination of transposon location

Genomic fragments with a transposon attached to its chromosomal DNA flanks were either self-ligated (ori-containing transposons) to produce plasmids or cloned into pUC19 (other transposons). DNA sequences of both transposon borders were determined from the plasmids using transposon-specific oligonucleotide primers. Some borders from yeast chromosomes were sequenced using a linker ligation method (32). Transposon borders from hES cells were determined using inverse PCR. Two micrograms of genomic DNA was digested with a combination of restriction enzymes (NheI + SpeI + XbaI or DraI + HpaI + SnaBI). Following ligation, the DNA was used as a template in PCR amplification with transposon-specific primers, yielding PCR products that were sequenced. Genomic locations were identified using the BLAST search at SGD (Saccharomyces Genome Database; http://genome-www.stanford.edu/Saccharomyces/), SDSC Biology WorkBench (http://workbench.sdsc.edu/), Ensembl Genome Browser (http://www.ensembl.org/index.html) release v41 (October 2006) or NCBI (http://www.ncbi.nlm.nih.gov/).

For statistical analysis of Mu integration in mouse cells, it was useful to compare integration site distributions to random expectation. For this, matched random controls were generated. A large set of random sites in the mouse genome was drawn computationally. However, recovery of Mu integration sites using restriction enzymes introduces a recovery bias favoring sites near suitable restriction enzyme cleavage sites in the mouse genome. This bias is addressed by the use of matched random controls. Each random site generated *in silico* was annotated for proximity to restriction enzyme recognition sites. For each experimental site of Mu integration, the distance to the restriction site used for recovery was measured, then 10 random sites were drawn that were the same distance from a recognition site for the same enzyme. The statistical analysis (Supplementary Text 1) preserved the pairing between Mu integration sites and matched random controls. This matching procedure 'washes out' recovery biases due to placement of restriction enzyme recognition sites, which otherwise can be severe.

Table 1. Integration of transposon into human HeLa cell genome

Clone	Sequence[a]	Chromosome	Band	Location of duplicated pentamer	Gene(s)		Transposon orientation[c]
RGC16	aggaggaagaACCAG(Kan/Neo-LoxP-Mu[b]) ACCAGgcacatgctg	8	q24.21	128363625-29	FAM84B–MYC	Intergenic	
RGC26	ttaaatgaacTTCAG(Kan/Neo-LoxP-Mu) TTCAGgaaaataatg	12	p12.3	15381980-84	PTPRO_HUMAN	Intron	+
RGC35	ttgttcagttCTGGT(Kan/Neo-LoxP-Mu) CTGGTgactcattgg	2	q31.2	179679743-47	NP_775919.2–SESTD1	Intergenic	
RGC200.1	agggggatccCCGGC(Kan/Neo-p15A-Mu) CCGGCccctgctgcc	5	q35.3	179178676-80	MGAT4B–SQSTM1	Intergenic	
RGC204.1	ttgagtcaagAGGGG(Kan/Neo-p15A-Mu) AGGGGgaagtccggg	1	c21.3	149586575-79	ENSESTG00000020135	Intron	+
RGC205.1	aagcatcaggCTGGG(Kan/Neo-p15A-Mu) CTGGTcaggtggagg	1	p36.13	16855907-11 16949721-25	ENSESTGG00000008139 ENSESTGG00000008135	Intron Intron	− −
RGC209.1	cccagacttcACCAT(Kan/Neo-p15A-Mu) ACCATtgtgtcatac	1	q21.3	152313986-90	Nup210L	Intron	+
RGC210.1	caacaatttcATAGG(Kan/Neo-p15A-Mu) ATAGGgttcagccta	20	q12	38737377-81	RP1-191L6.2-001–MAFB	Intergenic	
RGC214.1	ttgcagtgagCCGAG(Kan/Neo-p15A-Mu) CCGAGatcctgccac	5	q13.3	75118286-90	NP_001013738.1–SV2	Intergenic	

[a]Target site duplications are shown in capitals.
[b]Transposon is the same as Kan/Neo-Mu, except that it contains two loxP sites flanking the Kan/Neo gene.
[c]Transcription from the transposon compared with the direction of local transcription within the specified genomic location; +, same direction; −, opposite direction.

Table 2. Integration of the transposon into the human embryonic stem cell genome

Clone	Sequence[a]	Chromosome	Band	Location of duplicated pentamer	Gene(s)		Transposon orientation[b]
1	ttgcccaggcTGGAG (Puro-eGFP-Mu)TGGAGtacagtggct	1	p34.3	36223437-41	EIF2C3	Intron	−
4	agccaccgcgCCCGG(Puro-eGFP-Mu)CCCGGccaatcctgg	5	q31.1	133903082-86	PHF15	Intron	+
5[c]	tcttcaaataGAGAT(Puro-eGFP-Mu)GAGATggagaatcac	18	p11.1	5408820-24	EPB41L3	Intron	+
8	tgtaactcacCCCTG(Puro-eGFP-Mu)CCCTGgaaggaggct	17	q25.3	72973536-40	SEPT9	Intron	+
9	ggctactgtgGGCAC(Puro-eGFP-Mu)GGCACacacagatac	3	q25.1	152372945-49	MED12L	Intron	+

[a]Target site duplications are shown in capitals.
[b]Transcription from the transposon compared with the direction of local transcription within the specified genomic location; +, same direction; −, opposite direction.
[c]Possibly contains two transposons (Figure 4), but inverse PCR used for sequencing only gives one integration site.

RESULTS

Production of Mu transpososomes and verification of correct integration

Mu transpososomes were assembled by incubating mini-Mu transposon DNA with MuA transposase (Supplementary Figure 1), and correct assembly was verified by agarose gel electrophoresis (Supplementary Figure 2). Transpososome preparations were then concentrated ~10-fold for gene delivery. As negative controls in gene delivery experiments, we used naked transposon DNA without the addition of transposase, and transpososome preparations made using an active site mutant transposase MuA(E392Q) deficient for catalysis (33).

Two features of the integrated transposons could be checked to assess whether DNA integration was mediated by MuA in yeast and mammalian cells. Authentic Mu integration generates a 5-bp duplication at the targeted locus. Integrants were analyzed from yeast, mouse and human cells, and the correct length target sequence duplication was found in 158/168 yeast, 235/299 mouse integrants and 14/14 human integrants analyzed (Tables 1 and 2 and Supplementary Tables 3 and 4). Another test for correct Mu integration involved analyzing the sequences duplicated at the site of insertion. A consensus sequence for the duplicated region, 5′-N-T/C-G/C-G/A-N-3, has been observed both *in vivo* in bacteria and in *in vitro* transposition reactions (34,35). We analyzed the distribution of nucleotides in duplicated pentamers from yeast and mouse ES cells (Supplementary Figures 3 and 4), and evaluated the data from human HeLa and ES cells (Tables 1 and 2). All the results agreed well with those of the previous studies, supporting the idea that the integrations were mediated by MuA transposase.

Below we describe the use of Mu transpososomes to integrate new genes in yeast, mouse and human cells, then the distribution of integration sites relative to genomic features in each cell type.

Table 3. Number of colonies obtained on selection plates following electroporation into diploid and haploid yeast cells[a]

DNA	Transposase	Selection	Colonies generated (CFU/μg DNA)	
			FY1679[b]	FY-3a[c]
KanMX-Mu	MuA (wild-type)	YPD + G418	197 ± 20[d]	1117 ± 401
KanMX-Mu	MuA(E392Q)	YPD + G418	0	1
KanMX-Mu	None	YPD + G418	0	0
KanMX-p15A-Mu	MuA (wild-type)	YPD + G418	53 ± 27	583 ± 247
KanMX-p15A-Mu	MuA(E392Q)	YPD + G418	0	0
KanMX-p15A-Mu	None	YPD + G418	0	0
pHTH36[e]	None	SC-ura	1.6×10^6	2.8×10^6
pHTH36	None	YPD + G418	2.6×10^5	7.2×10^5

[a]The optimal electroporation parameters used are different for the two strains, not allowing a direct comparison between the strains.
[b]Diploid *S. cerevisiae* strain (26).
[c]Haploid derivative of FY1679 (this study).
[d]Reactions with wild-type MuA transposase were electroporated in triplicate. Average number of colonies ± SD is indicated.
[e]Electroporation of plasmid pHTH36 DNA served as a control for competence status.

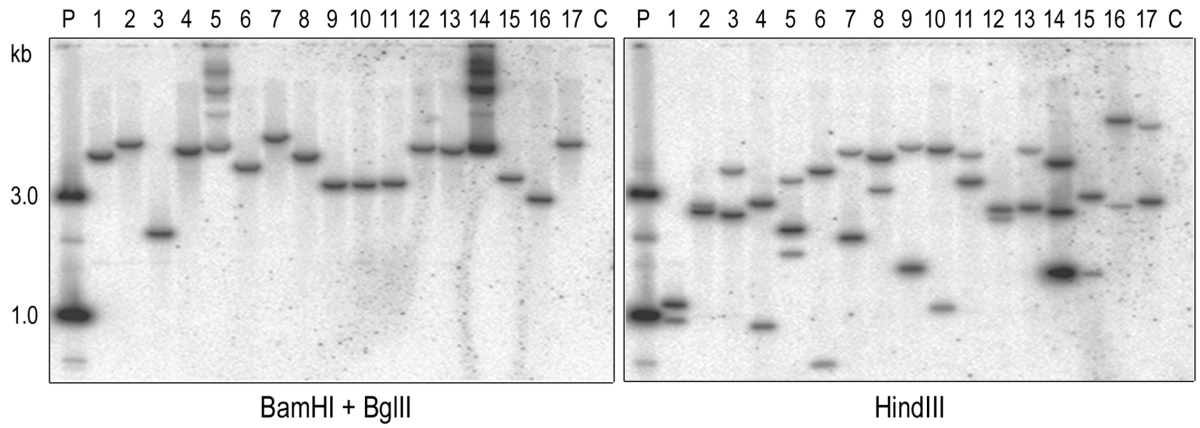

Figure 1. Southern analysis of transposon insertions into *S. cerevisiae* genome. Genomic DNA of 17 G418-resistant yeast clones (diploid strain FY1679) was digested with BamHI + BglII (left) or HindIII (right) and probed with labeled *kanMX4* DNA. (Lanes 1–17) Insertion mutants, (lane C) Genomic DNA of the original FY1679 recipient strain as a negative control, (lane P) Plasmid DNA containing the kanMX4-Mu transposon digested with HindIII as a positive control, fragment sizes on the left.

Mu integration in yeast cells

Transpososome preparations were electroporated separately into diploid (FY1679) and haploid (FY-3a) yeast recipient strains using optimized pulse parameters for each of the strains (see 'Materials and methods' section), then cells were selected for G418 resistance (Table 3). When KanMX4-Mu was used, more than 1000 CFU/μg introduced transposon DNA were obtained with the haploid-recipient strain. The diploid strain yielded ~200 CFU/μg of input KanMX4-Mu transposon DNA. When the longer KanMX4-p15A-Mu transposon was used, the number of G418-resistant colonies was roughly half of that obtained with KanMX4-Mu. Control electroporations yielded no or few resistant colonies in repeated experiments (Table 3 and data not shown).

Southern hybridization was used to detect the presence of transposon DNA and estimate the transposon copy number in 179 G418-resistant isolates (49 haploid, 130 diploid). A representative autoradiograph portraying 17 (diploid) isolates is shown in Figure 1. Of these 17 clones, 15 generated one prominent band when the genomic DNA was cut with enzymes that did not cleave the transposon DNA (BamHI + BglII), and two bands were generated if the transposon was cut once (HindIII). Two isolates produced more complex patterns. These data and additional sequencing data (below) indicated that the 15 clones each contain a single-copy genomic transposon integration, while the more complex patterns arose due to integration into the yeast two micron (2μ) plasmid. Altogether, 49 haploid integrant strains revealed 45 isolates with a single chromosomal integration and for isolates with a single plasmid integration. Of the 130 diploid integrant strains studied, 107 showed a single transposon integration in a chromosomal locus, and 22 isolates displayed a single transposon integrated in the 2μ plasmid. Notably, only one isolate contained two transposons, one in the chromosome and the other in the plasmid.

Mu integration in murine ES cells

Kan/Neo-p15A-Mu transpososomes were introduced into murine ES cells by electroporation, and G418 resistance was selected (Table 4), yielding 2400 CFU/μg DNA.

Table 4. Electroporation of mouse embryonic stem cells (AB2.2) with Kan/Neo-Mu transpososomes

Transposase	G418-resistant colonies generated (CFU/μg DNA)[a]	Survival of cells after electroporation (%)[b]	Survived cells transformed (%)
MuA (wild-type)	2376 ± 324	83	0.20
MuA(E392Q)	38 ± 12	79	0.01
None	98 ± 26	77	0.01

[a]Transpososomes were electroporated in triplicate, average number of colonies ± SD are indicated.
[b]4.6×10^6 viable cells were electroporated, plated in appropriate dilutions and survival was determined in growth medium without selection.

The control inactive mutant complexes yielded ~40 CFU/μg DNA. Transfection of naked transposon DNA yielded ~100 CFU/μg DNA (Figure 2A). Thus, the active transpososomes enhanced the transfection efficiency about 20-fold compared to the linear transposon or about 60-fold compared to the inactive transpososomes. Approximately 0.2% of the cells that survived electroporation were stably transfected.

Southern hybridization was used to verify the presence of transposon DNA and assess the copy number in several G418-resistant isolates. We repeatedly recloned primary isolates to produce cell lines that originated from a single cell. Overall, 47 clonal cell lines were established. An analysis of 16 clones is shown in Figure 2B. Genomic DNA was digested with BamHI and BglII, which did not cut the transposon. Fourteen of these clones produced one band and two clones showed two bands, indicative of integration by one and two transposons, respectively. Altogether, 41 clones contained a single transposon integration, five clones two integrations and one clone three integrations. Thus, a majority of the isolated G418-resistant clones contained a single integrated transposon.

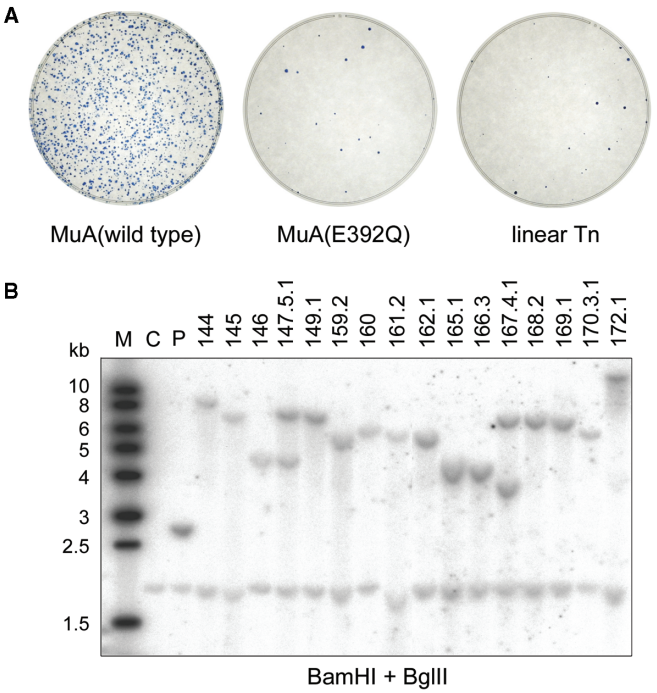

Figure 2. Analysis of transpososome-mediated gene delivery into mouse ES cells. (**A**) Efficiency. Pre-assembled transpososomes made with Kan/Neo-p15A-Mu transposon were introduced into mouse cells by electroporation. Following G418 selection, surviving cell colonies were stained with methylene blue. Gene delivery was analyzed using transpososomes made with wild-type MuA protein (left) and active site mutant MuA(E392Q) (middle). Analysis was done also with linear transposon DNA (right). (**B**) Southern analysis of transposon insertions into the mES cell genome. Genomic DNA of 17 G418-resistant mES cell clones was doubly digested with BamHI + BglII and probed with labeled Kan/Neo-p15A-Mu transposon DNA. (Lanes 1–17) Transposon insertion mutants. (Lane C) Genomic DNA of the original AB2.2 recipient strain as a negative control. (Lane P) AB2.2 genomic DNA spiked with transposon DNA as a positive control. (Lane M) Size marker. The cross-hybridizing band present in all genomic DNA samples served as a loading control.

Mu integration in HeLa cells

As a first application in human cells, a Kan/Neo-loxP-Mu transpososome preparation was used to electroporate HeLa cells. Following selection for G418 resistance, ~640 and ~25 CFU/μg transposon DNA for transpososomes and control naked transposon DNA were obtained. The experiment was then repeated using a Kan/Neo-p15A-Mu transpososome preparation, and the yield of G418-resistant colonies was ~400 CFU/μg transposon DNA. These data indicated that although naked transposon DNA can generate G418-resistant colonies with a low-frequency, evidently reflecting integration by non-homologous recombination, the use of transpososome preparation boosts the transformation efficiency up to 25-fold in HeLa cells.

We used Southern hybridization to analyze transposon DNA in HeLa cells. Figure 3 shows a Southern analysis of 19 HeLa cell clones resistant to G418. Seventeen isolates showed a single band, indicating single-copy insertions. Two isolates produced two bands of similar intensity, evidently reflecting the integration of two transposons. Thus, most of the analyzed HeLa clones contained only one integrated transposon copy in the genome.

Mu integration into the hESC genome

To establish conditions for human ES cell modification, we electroporated Puro-eGFP-Mu transpososomes into the hESC cell line FES29. Following puromycin selection, the yield of resistant colonies was ~2200 and ~65 CFU/μg for transpososomes and control naked DNA, respectively (Figure 4A). Expression of eGFP from the transposon construct was evaluated using FACS analysis (Supplementary Figure 5). Southern hybridization was used to verify the transposon integration in the genomic DNA (Figure 4B). Eight isolates produced only one distinct band hybridizing to the transposon-specific probe. One of the clones gave rise to two bands indicative of possible integration of two transposons. Thus, the Mu-mediated gene delivery technology can be used to modify human stem cells.

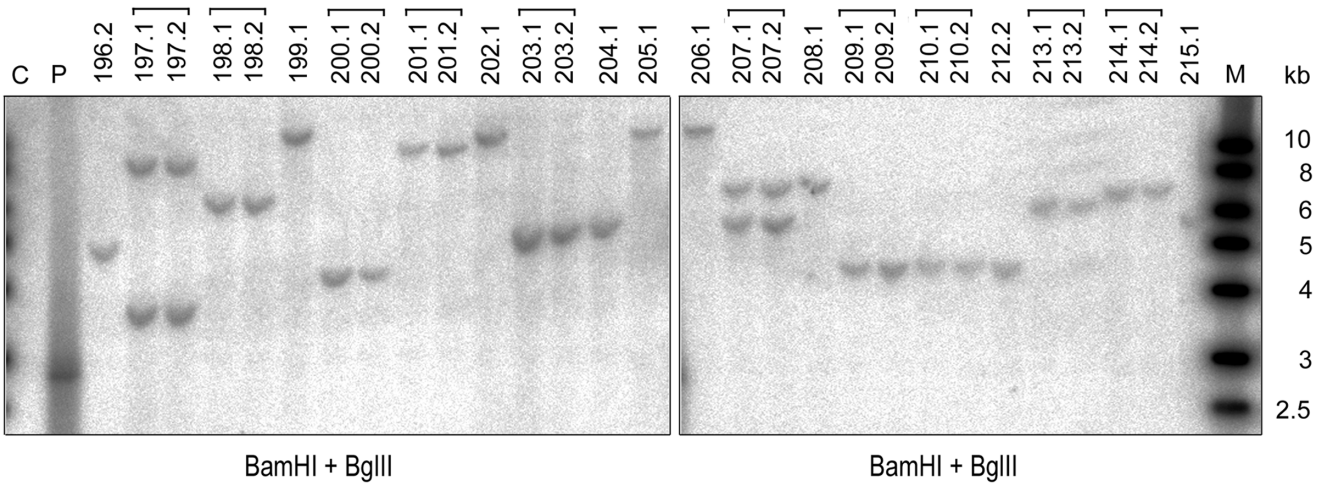

Figure 3. Southern blot analysis of transposon insertions into HeLa cell genome. Chromosomal DNA was doubly digested with BamHI + BglII and probed with labeled Kan/Neo-p15A-Mu transposon DNA. A total of 19 different G418-resistant clones are analyzed, some with their siblings (bracketed) for the verification of clonality. (Lane C) Genomic DNA of the recipient HeLa cell line as a negative control. (Lane P) HeLa cell DNA spiked with transposon DNA as a positive control. (Lane M) Size marker.

Integration target site selection in yeast

The sequences flanking the transposon DNA were determined and mapped to the *S. cerevisiae* genome sequence. Integration sites were distributed over all the yeast chromosomes (Supplementary Figure 6). For diploid cells, 102 sites were available for the statistical analysis of target site selection, and 37 sites for haploid cells (Supplementary Table 3). For comparison, a random set of ~10 000 integration sites was generated *in silico* and analyzed in parallel (Table 5).

The *S. cerevisiae* genome is composed of about ~69% open reading frames (ORFs), and introns are rare. For the random data set, 69% of integration sites were in ORFs (Figure 5A). For the experimental Mu integration site data sets, genes comprised 46% of sites for diploid and 24% for haploid, significantly different from random ($P < 0.001$ for comparison of either to random). Thus, after integration of Mu DNA into the yeast genome and selection for expression from the marker gene, integration sites accumulated outside the yeast genes. This even held true for the diploid strain, where two copies of each gene were present.

Yeast genes that are essential for growth have been cataloged experimentally, allowing the relationship between Mu integration and essential genes to be assessed. The random data set showed 13% integration in essential genes (Figure 5B). For Mu integration in diploids, the figure was 11%, while for haploid the figure was 0%. For haploids, there was a significant reduction in integration sites in essential genes compared to random expectation ($P = 0.036$; Fisher's exact test for comparison over all sites), while for diploids there was no significant difference ($P = 0.66$). Presumably, interruption of essential genes by Mu integration in the haploid strain caused reduced growth, so that cells harboring these integrants did not appear in the selected population, whereas in diploids the other gene copy was sufficient to complement.

Seventeen integrations were found in rDNA. These could not be mapped to a specific location due to the repeated structure of the rDNA region. The proportion of integration sites in rDNA in the two yeast strains were 12% (diploid) and 14% (haploid), compared to 0.3% in the random set, showing a strong favoring of integration in the rDNA cluster ($P < 0.001$ for comparison of either experimental set to random) (Figure 5C).

The *S. cerevisiae* strains studied contain the endogenous 2 μ plasmid, a 6-kb circular DNA present at 60–100 copies per cell. Nineteen Mu transposon integrations were found in the 2 μ plasmid. Because copy number is not known precisely, it is not possible to analyze integration frequency statistically. Nevertheless, it seems probable that Mu integration events also accumulated to a greater extent than expected by chance in the 2 μ plasmid.

Integration target site selection in murine ES cells

To analyze the distribution of Mu integration sites in the mouse genome, we determined the sequence of 299 Mu integration sites in murine ES cells (Supplementary Table 4). Both left and right junctions between Mu DNA and murine DNA were sequenced. A perfect 5-bp duplication of the target site, was found in 235 of the integration sites mapped (Supplementary Table 4A). Some of the clones which lacked the target duplication had deletions, the length of which typically ranged from 1 bp to several kilobases, and a few clones contained some extra DNA (Supplementary Table 4B). These could have occurred by (i) integration by non-homologous recombination into the genome; (ii) integration of two Mu sequences near one another, followed by recombination between the two; or (iii) integration by transposition followed by non-conventional processing of the transposition intermediate, involving a double-strand break [for a plausible model see (17)]. Also nested transposons formed by

integrating into one another and then into the genome were detected, in some cases accompanied by additional rearrangements. We also sequenced some insertion sites from control electroporations, and patterns indicative of non-homologous recombination were revealed (Supplementary Table 4C and D).

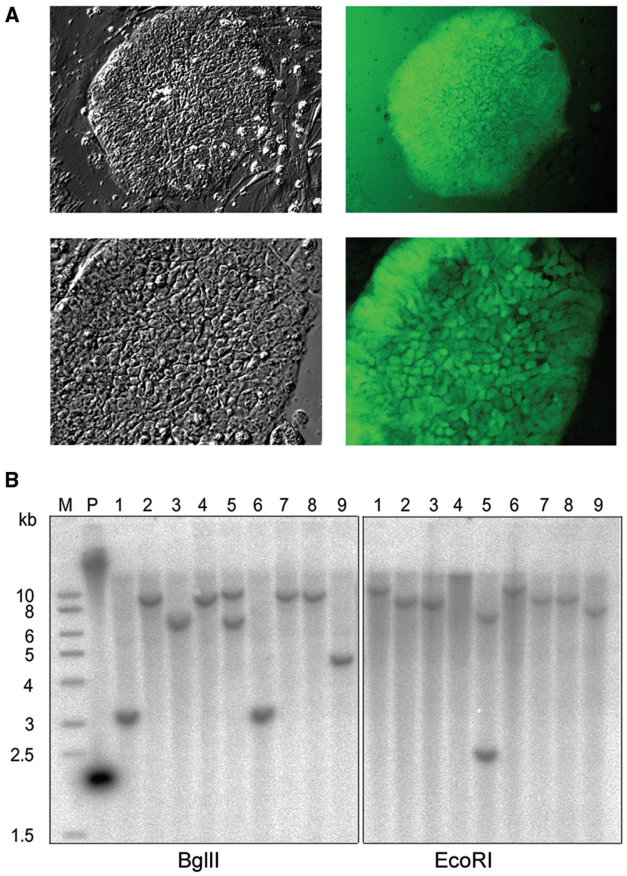

Figure 4. Analysis of transpososome-mediated gene delivery into human ES cells. (**A**) Expression of eGFP. Human ES cell line FES29 was electroporated with Puro-eGFP-Mu transposons and selected for 2 days with puromycin. Surviving fluorescent colonies were isolated and further cultured as clonal cell lines for several passages. Most of the colonies of the clonal isolates showed uniform GFP expression. Two example clones are shown in the phase contrast (left) and fluorescent (right) micrographs. (**B**) Southern analysis of the insertions into the hES cell genome. Genomic DNA of nine puromycin-resistant hES cell clones was digested with BglII (left) or EcoRI (right) and probed with labeled Puro-eGFP-Mu transposon DNA. (Lane P) Undigested genomic DNA of clone 2 spiked with transposon DNA as a positive control. (M) Size marker.

A collection of 214 Mu integration sites in mice was available for statistical analyses (Table 5). This collection was compared to a randomly generated group of control integration sites (Supplementary Text 1). All chromosomes except Y hosted at least four Mu integration events (Supplementary Figure 7). Integration in the mouse genome has also been extensively characterized for murine leukaemia (MLV) and human immunodeficiency virus (HIV), so data sets for these retroviruses (Table 5) were included for comparison in the analysis (Figure 6).

The Mu integration sites were found to be enriched in transcription units compared to random—fully 43% of Mu integration sites were in transcription units, while only 31% of the control sites were in transcription units ($P = 0.0001$; Fisher's exact test). Figure 6A shows a comparison to MLV and HIV, where each data set has been normalized to random expectation. Mu integration resembles MLV in modestly favoring integration in transcription units. In HIV integration, this preference is more pronounced. Both MLV and Mu strongly favored integration near CpG islands (Figure 6B), which are regions enriched in the rare CpG dinucleotide that often correspond to regulatory regions. Mu integration was also strongly favored in gene dense regions (Figure 6C) and regions of high transcriptional intensity (a measure that combines transcription activity with gene density; Figure 6D). Lastly, Mu integration was favored in G/C rich regions (Figure 6E). In general Mu integration resembled MLV integration in these genome features, although the MLV trend was in each case somewhat more pronounced.

Summaries of Mu integration frequency, relative to 54 forms of genomic annotation, can be found in Supplementary Text 1. This analysis includes quantification of the variables in Figure 6, but analyzed over multiple window sizes, and also integration frequency relative to additional types of annotation, such as histone post-translational modification. The single strongest factor positively correlating with Mu integration frequency is the primary sequence of the 20 bp at the site of integration (abbreviated 'score.20'). In addition, many of the features analyzed showed significant correlations with Mu integration frequency, raising the question of which effects are most meaningful. Many types of genomic annotation are

Table 5. Integration site data sets

Host organism	Number of sites	Transposon or virus	Protein coding genes (%)	Reference
Saccharomyces cerevisiae (haploid)	37	Mu transpososome	19.51	This study
Saccharomyces cerevisiae (diploid)	102	Mu transpososome	38.46	This study
Saccharomyces cerevisiae	9930	Random *in silico*	68.89	This study
Mouse ES cell	214	Mu transpososome	43.46	This study
Mouse	2140	Random *in silico*	30.51	This study
Human HeLa cell	9	Mu transpososome	55.56	This study
Human ES cell	5	Mu transpososome	100.00	This study
Human	2178	Random *in silico*	36.00	(58)
Mouse	2309	HIV	60.29	(59)
Mouse	4083	MLV	41.22	(59)

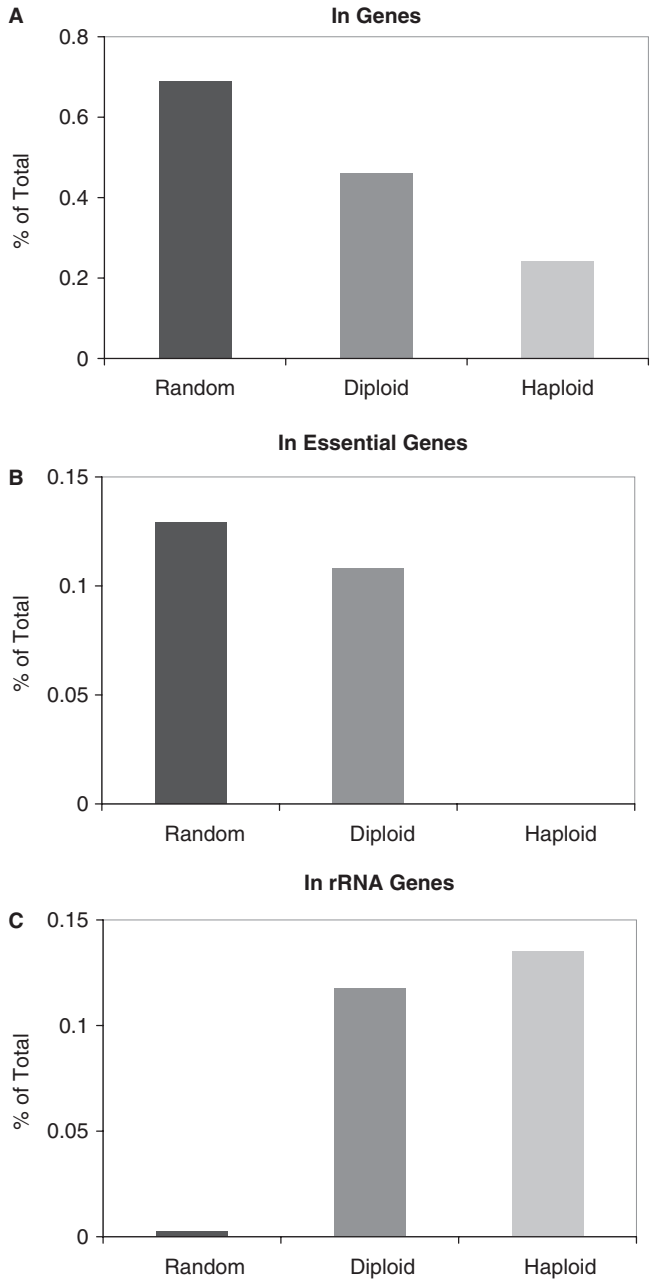

Figure 5. Analysis of integration site distributions in yeast cells. Comparison of integration target site distributions in haploid and diploid strains, and comparison of each to random. (**A**) Percentage of integration events in yeast open reading frames. (**B**) Percentage of integration events in experimentally defined essential genes. (**C**) Percentage of integration events in rRNA genes.

correlated with each other, so it is of interest to investigate which types of features are particularly strong drivers of Mu integration site selection, and which are 'hitchhikers' on other more important determinants.

To explore this, a conditional logit regression model was used to investigate interaction among the 54 variables (types of genomic annotation) and Mu integration frequency (Supplementary Text 1). To accommodate the large number of variables, Bayes Model Averaging was applied as described in (36). This revealed that a simplified model composed of only the local sequence effect (score.20), local G/C content, and histone H3K36 trimethylation (37) captured much of the Mu targeting preference in murine ES cells. H3K36 trimethylation accumulates within transcription units, and its positive association captures the favorable effects of integration within genes. Using the model, it is possible to show that the local sequence effect (score.20) largely accounts for the favored integration near CpG islands, which is consistent with the observation that the favored local target sequence contains the CpG dinucleotide within it. In summary, the dominant effects disclosed using Bayes Model Averaging included local sequence (score.20), the local G/C content, and H3K36 trimethylation.

Integration target site selection in human cells

The number of Mu integration sites analyzed in human cells was low (Tables 1, 2 and 5), but nevertheless a significant effect could be detected for integration frequency in genes. For the Hela cell data, 5/9 integration sites were in transcription units, and for ES cells 5/5 were in transcription units. Pooling the integration sites for all human cells, comparison to random shows a significant enrichment of Mu integration in transcription units ($P = 0.0077$).

DISCUSSION

In this paper we report the use of Mu transposomes to integrate new DNA in *S. cerevisiae*, murine ES cells and human HeLa and ES cells. Donor Mu DNAs were engineered to contain marker genes selectable in each cell type, allowing selection of cells containing at least one copy of integrated Mu transposon. This allowed us to ask where newly integrated Mu DNA accumulated in each of the genomes tested. We found that the combined selection for cell growth together with marker gene expression led to opposite results in the yeast versus mammals—in yeast the newly integrated DNA accumulated in regions between genes, while in mammals Mu DNA accumulated within genes. As is discussed below, this parallels the integration targeting preferences shown by genomic parasites that have evolved to inhabit each cell type.

Mu technology for gene delivery

The Mu transpososome technology described here provides a simple and flexible means of integrating new DNA into many types of cells. Transpososomes can be assembled *in vitro*, bypassing the need of expressing potentially toxic transposases within cells. Host factors are not required, and transpososomes can be conveniently concentrated and stored. In previous studies, we have used Mu transpososomes for the mutagenesis of Gram-negative and Gram-positive bacteria (24,25). Here we show that yeast, murine cells, and human cells could all be efficiently transduced by electroporation. The majority of integration events had the structures expected for MuA-mediated integration. This included the characteristic 5-bp target sequence duplication, and the favored sequence found at the site of integration. In mouse ES cells, 10–20% of the

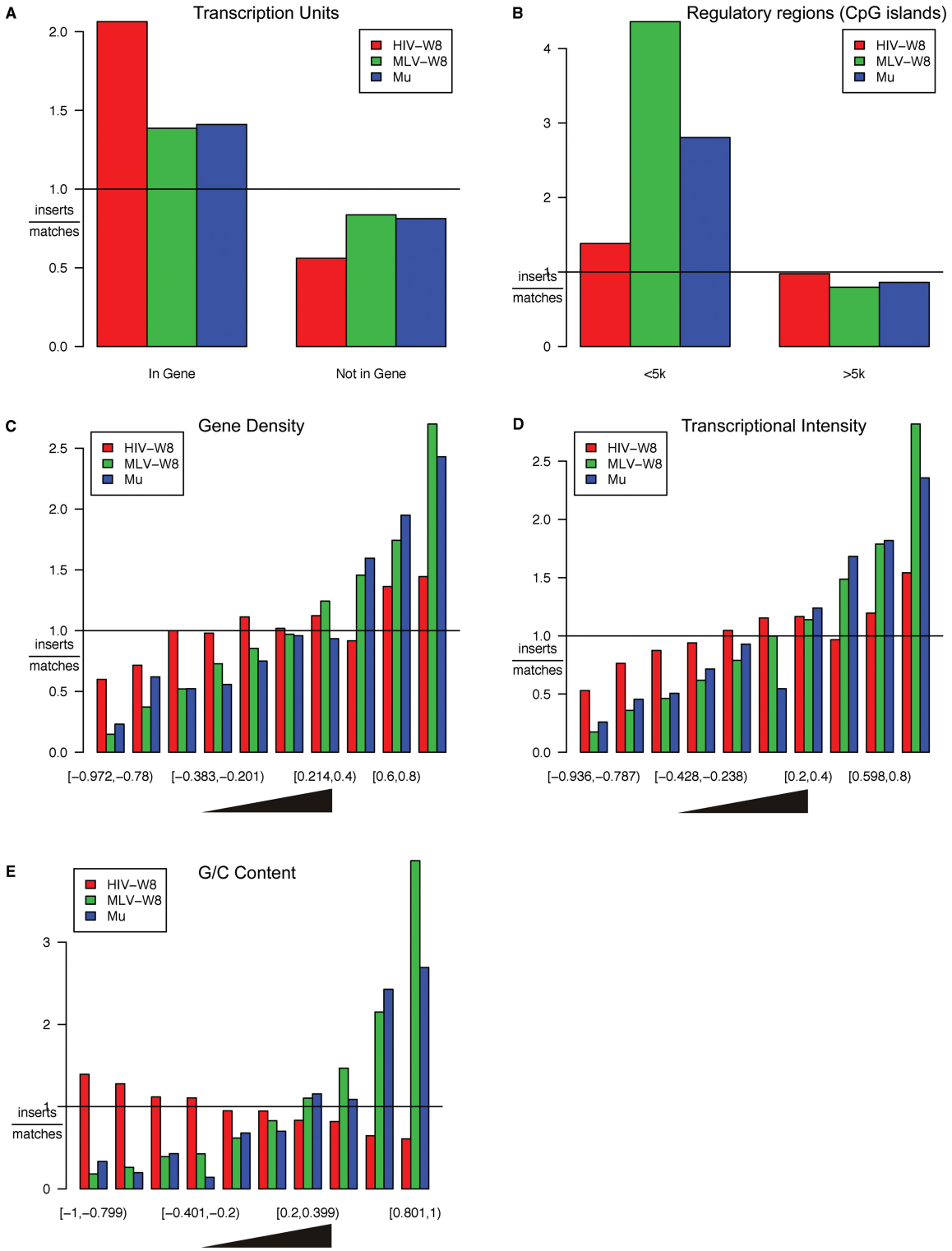

Figure 6. Analysis of integration site distributions in murine cells. The experimentally determined Mu integration sites reported here were compared to previously reported integration sites for MLV and HIV in the murine genome. In each figure, the proportion of Mu integration sites in each category is divided by the proportion in matched random controls—a bar below the line at 1.0 indicates disfavored Mu integration compared to random, while a bar above the line indicates favored integration. (**A**) Mu integration frequency in transcription units (defined as RefGenes). Comparison of Mu to random achieves $P = 5.9e\text{-}4$. (**B**) Mu integration frequency within 5 kb of the center of a CpG island. Comparison of Mu to random achieves $P = 4.22e\text{-}7$. (**C**) Analysis of Mu integration frequency in gene-dense regions. The murine genome was partitioned into 10 bins of increasing gene density (analyzed over four megabase regions), then the proportion of integration quantified in each bin and divided by random. Comparison of Mu to random achieved $P = 3.58e\text{-}10$. (**D**) Analysis of Mu integration as a function of transcriptional intensity. Affymetrix microarray data for murine ES cells was used to quantify transcriptional intensity. Transcriptional intensity was measured exactly as for gene density described above, but only genes in more highly expressed upper half of all genes queried on the microarray were scored. Comparison of Mu to random achieved $P = 6.43e\text{-}8$. (**E**) Mu integration frequency as a function of G/C content. The G/C content was measured over 5-kb intervals. Comparison of Mu to random achieves $P = 5.13e\text{-}14$.

integrants were probably due to non-MuA pathways, as indicated by abnormal DNA structures in the transposon or genomic DNA. Taken together, this establishes Mu as a convenient tool for the generation of stable cell lines and insertional mutagenesis. Mu may also be useful for transgenesis and therapeutic gene transfer. It is notable that the majority of Mu integrant clones contained only one transposon copy per genome, constituting a desirable feature for most applications.

Comparison to other transposon systems in eukaryotes

Mu transpososome technology is a useful addition to transposon-based methods for gene transfer, because the Mu target site preference and integration pattern is different from the others characterized. Sleeping beauty (*SB*) elements favor A/T rich DNA (38) and particular target site sequences (39,40). *SB* integration favors genes and their upstream regulatory regions in some cell types (36,38,39). *SB* elements also frequently target microsatellite DNA during genomic integration (38). Piggyback (*PB*) demonstrated non-random integration site selectivity, including high preference for regions surrounding transcriptional start sites and within long terminal repeat elements (41–43). *Tol2* does not show preference for any specific primary sequence but targets a characteristic local deformation of DNA (44–46). Thus, for insertional mutagenesis applications, it may be attractive to use Mu together with other transposon systems to widen the accessible target sites.

Positions of integrated Mu DNA in yeast and mammalian cells

The most unexpected finding of this study was the opposite biases in the targeting of Mu integration in yeast and mammalian cells. Integration in yeast was disfavored in genes, whereas in mammalian cells integration was actually favored in genes. For mouse ES cells, our deepest data for Mu in a mammalian cell type, a strong positive effect was also seen for H3K36 trimethylation, which marks transcribed regions. Possibly H3K36 trimethylation itself is involved directly in targeting, since integration near sites of H3K36 trimethylation was more strongly favored than integration in the transcription units themselves.

The observed departures from random could be either due to the initial targeting of Mu integration in each genome, or due to differential outgrowth of cells harboring integration events at different positions. We see evidence for both types of effects. Targeting directed by MuA could be seen in the conserved pentamer at the host cell site of integration (score.20). In murine cells, the multivariate modeling revealed that this largely accounts for favored integration near CpG islands. Selection after Mu integration likely took place in the haploid yeast, where no integration events were detected in essential genes, which is readily explained by selection after integration against disruption of essential functions.

In yeast, not only haploid but also diploid cells showed reduced integration in genes. In *S. cerevisiae* 69% of the genome is open reading frames, and introns are rare. Two models may account for the depletion of integration in genes in the diploids—(i) there may be effects of gene dosage at many genes, so that disruption of one copy is selected against during growth; or (ii) genes truncated by Mu integration may sometimes have enough dominant negative activity to confer a growth disadvantage.

In mammalian cells, several studies of retroviral integration have suggested that proviruses located within gene rich regions are well positioned for gene expression (47,48). For mammals, transcription units comprise roughly one-third of the genome, but exons are only about 1.5%, so that most integration events within transcription units are in introns. Mu DNA in introns is well positioned for marker gene expression, but gene disruption is minimized by splicing out of the Mu sequences. Thus we hypothesize that the bias in favor of integration in transcription units and gene-rich regions in mammalian cells is a consequence of selection after integration for efficient marker gene expression.

From the multivariate analysis, G/C richness scored as an important positive correlate of Mu integration frequency. The favored local sequence for Mu integration is G/C rich, potentially explaining the bias. In addition, G/C richness is positively correlated with gene density, which as described above may be favorable for marker gene expression.

Genome structure and integration of new sequences

Given the biases detected for Mu above, one might expect that genomic parasites which inhabit the yeast or mammalian genomes might have evolved to respond to the pressures detected in the Mu study. Indeed, examples can be found for both genomic parasites of yeast and mammals that naturally show the bias in *de novo* integration that were found here for Mu. HIV integration in mammalian cells is strongly favored in transcription units (49–52). A typical HIV infected cell only persists for a day or two after infection before it is destroyed by the cellular immune system or the toxicity of infection. Thus there is strong pressure on HIV for efficient gene expression over a short time, and integration in gene-rich regions promotes efficient HIV gene expression (48), likely explaining in the observed integration targeting preference. For the endogeous Ty and Tf transposons of yeasts an opposite bias is seen. These elements are retrotransposons, which replicate using retrovirus-like reverse transcriptase and integrase enzymes, but all the steps take place in a single cell. Thus, these elements are retrotransposons and not retroviruses. Ty and Tf elements have evolved to integrate in intergenic regions, and extensive experimentation has revealed some of the molecular tethering interactions involved (53–57). In the yeast genome, which is densely packed with genes, and also goes through haploid phases, selective pressures resulted in accumulation of Mu integrants outside of genes. For endogenous Ty and Tf elements, such tendencies have evolved to optimize cohabitation in the yeast genome. For mammalian genomes, which are exon sparse, Mu integration followed by selection for expression results in accumulation within genes. For HIV, *de novo* integration

targeting has evolved to strongly favor genes, probably to allow efficient expression of the viral genome. Thus, the studies with Mu help clarify the connections between eukaryotic genome structure and selection pressures on newly integrated DNA, thereby allowing us to better understand the integration targeting preferences of naturally occurring integrating parasites.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

Sari Tynkkynen, Pirjo Rahkola and Danielle Bansfield are acknowledged for excellent technical assistance.

FUNDING

The Academy of Finland (separate grants to H.S., M.P, X.M., and M.F.); Finnish National Technology Agency, TEKES (to H.S. and T.O.); National Institutes of Health (AI52845 to F.D.). Funding for open access charge: The Academy of Finland.

Conflict of interest statement. None declared.

REFERENCES

1. Bushman,F. (2002) *Lateral DNA Transfer. Mechanisms and Consequences*. Cold Spring Harbor Press, Cold Spring Harbor, NY.
2. Craig,N.L., Craigie,R., Gellert,M. and Lambowitz,A.M. (2002) *Mobile DNA II*., ASM Press, Washington, DC.
3. Dujon,B. (1996) The yeast genome project: what did we learn? *Trends Genet*., **12**, 263–270.
4. Lander,E.S., Linton,L.M., Birren,B., Nusbaum,C., Zody,M.C., Baldwin,J., Devon,K., Dewar,K., Doyle,M., FitzHugh,W. *et al*. (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860–921.
5. Venter,J.C., Adams,M.D., Myers,E.W., Li,P.W., Mural,R.J., Sutton,G.G., Smith,H.O., Yandell,M., Evans,C.A., Holt,R.A. *et al*. (2001) The sequence of the human genome. *Science*, **291**, 1304–1351.
6. Chaconas,G. and Harshey,R.M. (2002) In Craig,N.L., Craigie,R., Gellert,M. and Lambowitz,A.M. (eds), *Mobile DNA II*. ASM Press, Washington, DC, pp. 384–402.
7. Mizuuchi,K. (1992) Transpositional recombination: mechanistic insights from studies of Mu and other elements. *Annu. Rev. Biochem*., **61**, 1011–1051.
8. Craigie,R. and Mizuuchi,K. (1987) Transposition of Mu DNA: joining of Mu to target DNA can be uncoupled from cleavage at the ends of Mu. *Cell*, **51**, 493–501.
9. Surette,M.G., Buch,S.J. and Chaconas,G. (1987) Transpososomes: stable protein-DNA complexes involved in the *in vitro* transposition of bacteriophage Mu DNA. *Cell*, **49**, 253–262.
10. Savilahti,H., Rice,P.A. and Mizuuchi,K. (1995) The phage Mu transpososome core: DNA requirements for assembly and function. *EMBO J*., **14**, 4893–4903.
11. Haapa,S., Taira,S., Heikkinen,E. and Savilahti,H. (1999) An efficient and accurate integration of mini-Mu transposons *in vitro*: a general methodology for functional genetic analysis and molecular biology applications. *Nucleic Acids Res*., **27**, 2777–2784.
12. Haapa,S., Suomalainen,S., Eerikäinen,S., Airaksinen,M., Paulin,L. and Savilahti,H. (1999) An efficient DNA sequencing strategy based on the bacteriophage Mu *in vitro* DNA transposition reaction. *Genome Res*., **9**, 308–315.
13. Laurent,L.C., Olsen,M.N., Crowley,R.A., Savilahti,H. and Brown,P.O. (2000) Functional characterization of the human immunodeficiency virus type 1 genome by genetic footprinting. *J. Virol*., **74**, 2760–2769.
14. Kekarainen,T., Savilahti,H. and Valkonen,J.P.T. (2002) Functional genomics on *Potato virus A*: virus genome-wide map of sites essential for virus propagation. *Genome Res*., **12**, 584–594.
15. Vilen,H., Aalto,J.-M., Kassinen,A., Paulin,L. and Savilahti,H. (2003) A direct transposon insertion tool for modification and functional analysis of viral genomes. *J. Virol*., **77**, 123–134.
16. Kiljunen,S., Vilen,H., Pajunen,M., Savilahti,H. and Skurnik,M. (2005) Nonessential genes of phage ϕYeO3-12 include genes involved in adaptation to growth on *Yersinia enterocolitica* serotype O:3. *J. Bact*., **187**, 1405–1414.
17. Krupovič,M., Vilen,H., Bamford,J.K.H., Kivelä,H.M., Aalto,J.-M., Savilahti,H. and Bamford,D.H. (2006) Genome characterization of lipid-containing marine bacteriophage PM2 by transposon insertion mutagenesis. *J. Virol*., **80**, 9270–9278.
18. Pajunen,M., Turakainen,H., Poussu,E., Peränen,J., Vihinen,M. and Savilahti,H. (2007) High-precision mapping of protein–protein interfaces: an integrated genetic strategy combining *en masse* mutagenesis and DNA-level parallel analysis on a yeast two-hybrid platform. *Nucleic Acids Res*., **35**, e103.
19. Poussu,E., Vihinen,M., Paulin,L. and Savilahti,H. (2004) Probing the α-complementing domain of *E. coli* β-galactosidase with use of an insertional pentapeptide mutagenesis strategy based on Mu *in vitro* DNA transposition. *Proteins*, **54**, 681–692.
20. Poussu,E., Jäntti,J. and Savilahti,H. (2005) A gene truncation strategy generating N- and C-terminal deletion variants of proteins for functional studies: mapping of the Sec1p binding domain in yeast Mso1p by a Mu *in vitro* transposition-based approach. *Nucleic Acids Res*., **33**, e104.
21. Taira,S., Tuimala,J., Roine,E., Nurmiaho-Lassila,E.-L., Savilahti,H. and Romantschuk,M. (1999) Mutational analysis of the *Pseudomonas syringae* pv. *tomato hrpA* gene encoding Hrp pilus subunit. *Mol. Microbiol*., **34**, 737–744.
22. Vilen,H., Eerikäinen,S., Tornberg,J., Airaksinen,M.S. and Savilahti,H. (2001) Construction of gene-targeting vectors: a rapid Mu *in vitro* DNA transposition-based strategy generating null, potentially hypomorphic, and conditional alleles. *Transgenic Res*., **10**, 69–80.
23. Jukkola,T., Trokovic,R., Maj,P., Lamberg,A., Mankoo,B., Pachnis,V., Savilahti,H. and Partanen,J. (2005) *Meox1*^{Cre}: a mouse line expressing Cre recombinase in somitic mesoderm. *Genesis*, **43**, 148–153.
24. Lamberg,A., Nieminen,S., Qiao,M. and Savilahti,H. (2002) Efficient insertion mutagenesis strategy for bacterial genomes involving electroporation of *in vitro*-assembled DNA transposition complexes of bacteriophage Mu. *Appl. Environ. Microbiol*., **68**, 705–712.
25. Pajunen,M.I., Pulliainen,A.T., Finne,J. and Savilahti,H. (2005) Generation of transposon insertion mutant libraries for Gram-positive bacteria by electroporation of phage Mu DNA transposition complexes. *Microbiology*, **151**, 1209–1218.
26. Winston,F., Dollard,C. and Ricupero-Hovasse,S.L. (1995) Construction of a set of convenient *Saccharomyces cerevisiae* strains that are isogenic to S288C. *Yeast*, **11**, 53–55.
27. Mikkola,M., Olsson,C., Palgi,J., Ustinov,J., Palomäki,T., Horelli-Kuitunen,N., Knuutila,S., Lundin,K., Otonkoski,T. and Tuuri,T. (2006) Distinct differentiation characteristics of individual human embryonic stem cell lines. *BMC Dev. Biol*., **6**, 40.
28. Baker,T.A., Mizuuchi,M., Savilahti,H. and Mizuuchi,K. (1993) Division of labor among monomers within the Mu transposase tetramer. *Cell*, **74**, 723–733.
29. Miller,S.A., Dykes,D.D. and Polesky,H.F. (1988) A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res*., **16**, 1215.
30. Sambrook,J. and Russell,D.W. (2001) edn. *Molecular Cloning: A Laboratory Manual,* 3rd edn. Cold Spring Harbor Press, Cold Spring Harbor, NY.
31. Savilahti,H. and Bamford,D.H. (1993) Protein-primed DNA replication: role of inverted terminal repeats in the *Escherichia coli* bacteriophage PRD1 life cycle. *J. Virol*., **67**, 4696–4703.

32. Reddy,M.K., Nair,S. and Sopory,S.K. (2002) A new approach for efficient directional genome walking using polymerase chain reaction. *Anal. Biochem.*, **306**, 154–158.
33. Baker,T.A. and Luo,L. (1994) Identification of residues in the Mu transposase essential for catalysis. *Proc. Natl Acad. Sci. USA*, **91**, 6654–6658.
34. Mizuuchi,M. and Mizuuchi,K. (1993) Target site selection in transposition of phage Mu. *Cold Spring Harb. Symp. Quant. Biol.*, **58**, 515–523.
35. Haapa-Paananen,S., Rita,H. and Savilahti,H. (2002) DNA transposition of bacteriophage Mu: a quantitative analysis of target site selection *in vitro*. *J. Biol. Chem.*, **277**, 2843–2851.
36. Berry,C., Hannenhalli,S., Leipzig,J. and Bushman,F.D. (2006) Selection of target sites for mobile DNA integration in the human genome. *PLoS Comput. Biol.*, **2**, e157.
37. Mikkelsen,T.S., Ku,M., Jaffe,D.B., Issac,B., Lieberman,E., Giannoukos,G., Alvarez,P., Brockman,W., Kim,T.-K., Koche,R.P. *et al.* (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*, **448**, 553–560.
38. Yant,S.R., Wu,X., Huang,Y., Garrison,B., Burgess,S.M. and Kay,M.A. (2005) High-resolution genome-wide mapping of transposon integration in mammals. *Mol. Cell. Biol.*, **25**, 2085–2094.
39. Geurts,A.M., Hackett,C.S., Bell,J.B., Bergemann,T.L., Collier,L.S., Carlson,C.M., Largaespada,D.A. and Hackett,P.B. (2006) Structure-based prediction of insertion-site preferences of transposons into chromosomes. *Nucleic Acids Res.*, **34**, 2803–2811.
40. Liu,G., Geurts,A.M., Yae,K., Srinivasan,A.R., Fahrenkrug,S.C., Largaespada,D.A., Takeda,J., Horie,K., Olson,W.K. and Hackett,P.B. (2005) Target-site preferences of *Sleeping Beauty* transposons. *J. Mol. Biol.*, **346**, 161–173.
41. Ding,S., Wu,X., Li,G., Han,M., Zhuang,Y. and Xu,T. (2005) Efficient transposition of the *piggyBac* (*PB*) transposon in mammalian cells and mice. *Cell*, **122**, 473–483.
42. Wilson,M.H., Coates,C.J. and George,A.L. Jr. (2007) *PiggyBac* transposon-mediated gene transfer in human cells. *Mol. Ther.*, **15**, 139–145.
43. Wu,S.C.-Y., Meir,Y.-J.J., Coates,C.J., Handler,A.M., Pelczar,P., Moisyadi,S. and Kaminski,J.M. (2006) *piggyBac* is a flexible and highly active transposon as compared to *Sleeping Beauty*, *Tol2*, and *Mos1* in mammalian cells. *Proc. Natl Acad. Sci. USA*, **103**, 15008–15013.
44. Clark,K.J., Carlson,D.F., Foster,L.K., Kong,B.-W., Foster,D.N. and Fahrenkrug,S.C. (2007) Enzymatic engineering of the porcine genome with transposons and recombinases. *BMC Biotechnol.*, **7**, 42.
45. Hackett,C.S., Geurts,A.M. and Hackett,P.B. (2007) Predicting preferential DNA vector insertion sites: implications for functional genomics and gene therapy. *Genome Biol.*, **8**, S12.
46. Koga,A. (2004) Transposition mechanisms and biotechnology applications of the medaka fish *Tol2* transposable element. *Adv. Biophys.*, **38**, 161–180.
47. Lewinski,M.K., Yamashita,M., Emerman,M., Ciuffi,A., Marshall,H., Crawford,G., Collins,F., Shinn,P., Leipzig,J., Hannenhalli,S. *et al.* (2006) Retroviral DNA integration: viral and cellular determinants of target-site selection. *PLoS Pathog.*, **2**, e60.
48. Lewinski,M.K., Bisgrove,D., Shinn,P., Chen,H., Hoffmann,C., Hannenhalli,S., Verdin,E., Berry,C.C., Ecker,J.R. and Bushman,F.D. (2005) Genome-wide analysis of chromosomal features repressing human immunodeficiency virus transcription. *J. Virol.*, **79**, 6610–6619.
49. Hematti,P., Hong,B.-K., Ferguson,C., Adler,R., Hanawa,H., Sellers,S., Holt,I.E., Eckfeldt,C.E., Sharma,Y., Schmidt,M. *et al.* (2004) Distinct genomic integration of MLV and SIV vectors in primate hematopoietic stem and progenitor cells. *PLoS Biol.*, **2**, e423.
50. Mitchell,R.S., Beitzel,B.F., Schröder,A.R.W., Shinn,P., Chen,H., Berry,C.C., Ecker,J.R. and Bushman,F.D. (2004) Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. *PLoS Biol.*, **2**, e234.
51. Schröder,A.R.W., Shinn,P., Chen,H., Berry,C., Ecker,J.R. and Bushman,F. (2002) HIV-1 integration in the human genome favors active genes and local hotspots. *Cell*, **110**, 521–529.
52. Wu,X., Li,Y., Crise,B. and Burgess,S.M. (2003) Transcription start regions in the human genome are favored targets for MLV integration. *Science*, **300**, 1749–1751.
53. Ji,H., Moore,D.P., Blomberg,M.A., Braiterman,L.T., Voytas,D.F., Natsoulis,G. and Boeke,J.D. (1993) Hotspots for unselected Ty*1* transposition events on yeast chromosome III are near tRNA genes and LTR sequences. *Cell*, **73**, 1007–1018.
54. Kirchner,J., Connolly,C.M. and Sandmeyer,S.B. (1995) Requirement of RNA polymerase III transcription factors for *in vitro* position-specific integration of a retroviruslike element. *Science*, **267**, 1488–1491.
55. Yieh,L., Kassavetis,G., Geiduschek,E.P. and Sandmeyer,S.B. (2000) The Brf and TATA-binding protein subunits of the RNA polymerase III transcription factor IIIB mediate position-specific integration of the *gypsy*-like element, Ty*3*. *J. Biol. Chem.*, **275**, 29800–29807.
56. Brady,T.L., Schmidt,C.L. and Voytas,D.F. (2008) Targeting integration of the *Saccharomyces* Ty*5* retrotransposon. *Methods Mol. Biol.*, **435**, 153–163.
57. Leem,Y.-E., Ripmaster,T.L., Kelly,F.D., Ebina,H., Heincelman,M.E., Zhang,K., Grewal,S.I., Hoffman,C.S. and Levin,H.L. (2008) Retrotransposon Tf*1* is targeted to Pol II promoters by transcription activators. *Mol. Cell*, **30**, 98–107.
58. Ciuffi,A., Llano,M., Poeschla,E., Hoffmann,C., Leipzig,J., Shinn,P., Ecker,J.R. and Bushman,F. (2005) A role for LEDGF/p75 in targeting HIV DNA integration. *Nat. Med.*, **11**, 1287–1289.
59. Wang,G.P., Garrigue,A., Ciuffi,A., Ronen,K., Leipzig,J., Berry,C., Lagresle-Peyrou,C., Benjelloun,F., Hacein-Bey-Abina,S., Fischer,A. *et al.* (2008) DNA bar coding and pyrosequencing to analyze adverse events in therapeutic gene transfer. *Nucleic Acids Res.*, **36**, e49.