A novel L1 retrotransposon marker for HeLa cell line identification

Raheleh Rahbari[1], **Tom Sheahan**, **Vasileios Modes**[2], **Pam Collier**[1], **Catriona Macfarlane**[1], and **Richard M. Badge**[1]

[1] Department of Genetics University of Leicester, Leicester, UK

[2] Department of Cancer Studies and Molecular Medicine Leicester Royal Infirmary, Leicester, UK

Abstract

The HeLa cell line is the oldest, most widely distributed, permanent human cell line. As a nearly ubiquitous inhabitant of laboratories using tissue culture techniques, its aggressive growth characteristics make it a problematic contaminant that can overgrow less robust cell lines. Consequently, HeLa contamination is common in both the research laboratory and cell line repository contexts, and its detection is hampered by the lack of a rapid, sensitive and robust assay. Here we report the development of a HeLa-specific DNA diagnostic test: a single duplex detection PCR assay targeting an L1 retrotransposon insertion. All HeLa clones from a geographically diverse panel were positive by this assay, and the particular L1 insertion we identified appears to be unique to the HeLa cell line. The assay can detect very low levels of HeLa contamination (<1%), and can be performed on un-purified cell pellets, allowing rapid routine screening.

Keywords

cell culture; LINE-1; HeLa; contamination; genotyping

Introduction

The history of the HeLa cell line is long and somewhat controversial (1). Despite its pivotal role as the first continuous human cancer cell line, the identity of the HeLa cell line has been accompanied by uncertainty. Although originally reported in 1952 as a cervical epithelial carcinoma from a female African-American individual, subsequent analysis of tumor pathology and clinical phenotype clearly identified the primary tumor as a rare adenocarcinoma (1). As the first human cell line to be widely cultured, HeLa has undergone strong adaptation to culture conditions. While karyotypically highly abnormal ($4n = 82$) and carrying cytologically distinct marker chromosomes (2), it is phenotypically epithelial in appearance, grows rapidly (doubling time is ~24 h), and shows a lack of contact inhibition (3). These characteristics and its ubiquity in cell culture laboratories have led to frequent cross-contamination. It has been estimated that 18% of cell lines submitted to repositories are contaminated with other cell lines, with HeLa alone being responsible for 25% of these cross-contamination events (2). In many cases, the originally deposited line has been completely replaced, resulting in the loss of potentially unique research resources in the very

facilities designed to ensure their maintenance. There is also little doubt that cell line contamination in general—and HeLa contamination, in particular—has led to the publication of erroneous research data (4), with HeLa sublines masquerading variously as amnion cells (WISH), embryonic lung cells (L132) and liver cells (Chang liver) (5).

The advent of robust and fairly inexpensive short tandem repeat (STR)–based DNA (STR DNA) fingerprinting technologies (6) has enabled the development of international reference standards for cell line identification. These standards, in principle, should eliminate cell line cross-contamination as a source of error in research using established cell lines, and should allow validation of provenance for newly derived human cell lines. The major cell line repositories (Coriell, ATCC, ECACC, DSMZ, JCRB) routinely use STR DNA fingerprinting techniques to monitor the status of their cell line stocks. However, part of the problem with the use of contaminated cell lines is the requirement for the regular and routine assay of cell line identity. It has been suggested that mandatory monitoring of cell line identity, as a condition for funding and/or publication (7), could eliminate cell line contamination. Meanwhile, in the absence of such requirements, the expense of STR-based monitoring to research laboratories is clearly a barrier to self-regulation. Given that the HeLa cell line is disproportionately responsible for cross contamination events, a specific assay for this contamination could be included in cell culture practice, in much the same way as mycoplasma contamination can be routinely monitored by PCR (8). Here we present an assay that may remove some of these barriers.

L1 retrotransposons are active, autonomous mobile elements that are very common in the human genome [comprising 17% of the human genome sequence (9)]. Human-specific L1s (L1Hs) have been inserting into our DNA since the origin of the species and thus are readily employed as population (10) and individual specific (11) genetic markers with a number of useful characteristics. They are large (up to 7 kb) and stable insertions that undergo precise reversion only very rarely (12). During ongoing studies of human-specific, full-length, and potentially active L1 elements, we identified [using genome-wide transposon display methods (11)] a full-length L1 element that appeared to be HeLa-specific. Initial screening experiments showed that this insertion was present in a panel of widely geographically distributed HeLa isolates obtained from cell culture repositories in Europe and the United States, and absent from a large panel of human DNA samples. To fully realize the potential of this insertion as a HeLa-specific molecular diagnostic, we designed a single PCR-based genotyping assay that can be utilized in laboratories with access to standard molecular biology reagents and equipment. The assay is internally controlled, is sensitive to even low levels of HeLa contamination, and can be performed on unpurified cell pellets. We believe it will be of high utility in routine culture hygiene in cell culture laboratories and in tissue culture repositories.

Materials and methods

Tissue culture and DNA sources

The majority of HeLa and non-HeLa cell lines were obtained as gifts from collaborators (see Table 1). Where necessary, lines were revived from frozen storage and cultured at 37°C under 7% CO_2 in low glucose DMEM (Gibco/Invitrogen, Carlsbad, CA, USA) supplemented with 10% fetal calf serum (Gibco/Invitrogen, Carlsbad, CA, USA) and 1× penicillin/streptomycin/glutamate (Gibco/Invitrogen).

DNA samples from anonymous black donors (from Harare, Zimbabwe) were provided by Sir Alec J. Jeffreys FRS (University of Leicester, Leicester, UK). Northern European DNA samples were obtained from the Centre d'Etude du Polymorphisme Humain (CEPH)/ Fondation Jean Dausset. African, African-American, German Caucasian, Asian, and South

American DNA samples were obtained from the Coriell Institute for Medical Research (Camden, NJ, USA).

DNA extraction

DNA was extracted from cultured cells using the Puregene Tissue Culture kit (Qiagen UK, Crawley, West Sussex, UK) following the manufacturer's protocol. DNA integrity was assessed by fractionation on 0.8% (w/v) agarose gels [0.5× Tris-Borate-EDTA (TBE)] alongside intact bacteriophage λ DNA. DNA was visualized by ethidium bromide (0.5 μg/mL) staining under UV illumination.

PCR genotyping

Initial PCR genotyping assays used 1× PCR buffer (45 mM Tris-HCl pH 8.8, 11 mM NH_4SO_4, 0.9 mM $MgCl_2$), 6.7 mM β-mercaptoethanol, 113 μg/mL bovine serum albumin (BSA), 200 μM dNTPs, 0.05 U/μL *Taq* DNA polymerase (Abgene, Epsom, Surrey, UK) and primers [empty site: VM164A (5′-TGCCTCCTAGATCGTATTCCC-3′) and VM-164B (5′-GCACTCTGTGGCATGAAGGT-3′); filled site: VM164A and RB5-PA2 (5′-TGGAAATGCAGAAATCACCG-3′)]. PCR primers were synthesized by Sigma-Genosys/Aldrich (Gillingham, Dorset, UK). Ten to twenty nanograms of genomic DNA were added to 10-μL PCR reactions and cycled in a Tetrad 2 Thermal Cycler (MJ Research/ Bio-Rad, Hercules, CA, USA). Cycle conditions were 30 cycles of 96°C for 30 s, 63°C for 30 s, 72°C for 1 min; then 1 cycle of 72°C for 5 min. Single-tube HeLa-specific PCR used 1× PCR buffer and primers VM164A, VM164B, and RB164K2 (5′-TGGCTCCTCCCTCTATTATCG-3′). Cycle conditions were 30 cycles of 96°C for 30 s, 63°C for 30 s, 72°C for 1 min; then 1 cycle of 72°C for 5 min. PCR reactions were fractionated on 2% (w/v) agarose gels (0.5× TBE) and DNA was visualized by ethidium bromide (0.5 μg/mL) staining.

Dotblotting

PCR-amplified DNA was denatured with 0.5M NaOH, 2M NaCl, 25 mM EDTA and transferred to nylon membranes using a 96-well dotblot vacuum manifold (BRL, Gaithersburg, MD, USA). DNA was covalently linked to the membrane by UV crosslinking (CL-1000 UV Cross-linker, UVP, Cambridge, UK). Radiolabeled empty site–specific probes were generated by labeling 25–50 ng of the empty site amplicon (amplified with primers VM164A/VM164B) with α.^{32}P dCTP using the Amersham Rediprime II DNA labeling system (GE Healthcare, Fairfield, CT, USA) according to the manufacturer's protocol. Hybridization was carried out overnight at 65°C in modified Church buffer (0.5 M NaH_2PO_4/Na_2HPO_4 pH 7.2, 7% SDS, 1 mM EDTA). Hybridized membranes were washed twice in 0.2× sodium chloride-sodium citrate (SSC) and 0.5% sodium dodecyl sulfate (SDS), and once in 0.1× SSC/0.1× SDS, at 65°C. Dotblots were visualized by autoradiography.

Results and discussion

As illustrated in Figure 1A, the AL137164 locus (chr14q12) in HeLa cells harbors a full-length human-specific L1 insertion, located in intron 2 of the *STXBP6* (*amisyn*) gene. This L1 element is inserted in the same transcriptional orientation as the *amisyn* gene, has intact open reading frames (ORFs), and is likely to be active in cell culture retrotransposition assays (V. Modes and R. Badge, unpublished data). HeLa is triploid with respect to this otherwise karyotypically normal chromosome: PCR and direct sequencing of flanking markers showed that HeLa cells carry two empty site chromosomes (i.e., that lack the L1 insertion) and one filled site chromosome (data not shown). All tested HeLa isolates were

heterozygous for the insertion (i.e., they contain at least one empty and one filled site chromosome).

To establish the population frequency of this L1 insertion, and to determine if this marker could be used unequivocally to identify HeLa and HeLa-derived cell lines, we applied locus-specific genotyping assays to diverse human DNA samples. Samples from 72 unrelated black Zimbabwean individuals were genotyped (of DNA we have tested, this sample set is most closely related to the HeLa donor's likely population of origin), using a sensitive dotblot assay. In this assay PCR amplicons specific for the empty site (chromosomes lacking the L1 insertion) site and filled site (chromosomes harboring the L1 insertion) were hybridized with a radiolabeled probe derived from the empty site, as shown in Figure 1B. This probe detects not only the empty site amplicon, but also the filled site, due to the presence of 3′ flanking genomic sequence in the amplicon. All tested Zimbabweans were homozygous for the insertion empty site. Genotyping 364 unrelated individuals from geographically diverse populations (Africans, African-Americans, Northern Europeans, German Caucasians, Asians, and South Americans) failed to identify any other carriers of the AL137164 L1 insertion (data not shown).

The original genotyping assay was modified into a single duplex PCR format (Figure 2A), providing a simple way to discriminate HeLa and non-HeLa cell lines. As shown in Figure 2B, all tested HeLa isolates, including the HeLa S3 subline, were positive using this assay (they all generated a 325-bp insertion–specific amplicon), and 11 tested non-HeLa lines were negative (they lacked this filled site–specific amplicon).

DNA mixing experiments (Figure 3A) using HeLa DNA and Ntera2D1 genomic DNA indicate that the single-tube assay is highly sensitive, and can detect at least 1% of HeLa mixed with non-HeLa DNA. The assay is also robust with respect to the input material: even genomic DNA present in unpurified frozen cell pellets (Figure 3B) can be genotyped using this assay.

It is evident that the HeLa cell line is responsible for a disproportionate number of cell line contamination events, likely in part due to its unusual origin as an aggressive cervical adenocarcinoma, as well as an adaptation to cell culture over its long history (1). Despite the availability of STR-based DNA fingerprinting technology to unequivocally identify cell lines (6), contaminated cell lines are still used erroneously in peer-reviewed publications (4). One reason for this may be ignorance of the status of these lines, but the expense of routine cell line identity monitoring may also play a role in the persistence of these compromised experimental resources. Even vigilant monitoring of cell line identity by cell culture stock centers cannot guard against contamination events in individual laboratories. The widespread sharing of cell line cultures between laboratories and their resultant poor provenance—although pivotal in enabling this particular study—will hinder a solution to this problem. Recent calls to action on this issue (17,18) as well as assessments of the costs of using unauthenticated cell lines in research (7), show that the need to develop inexpensive, rapid and sensitive cell identity assays is topical and pressing.

Having fortuitously discovered what appears to be a HeLa-specific full-length L1 insertion, we screened donated HeLa cell isolates and found that independently acquired and geographically dispersed HeLa isolates all share this marker. To establish whether this marker might occur in other human cell lines, we determined the population frequency of the insertion in a diverse panel of human DNA samples. The insertion could not be located in a panel of 72 DNA samples from Zimbabwean donors, despite this population sample being closely related to the HeLa donor's likely origin (of the DNA we have tested) and the large proportion of worldwide human genetic diversity preserved in African populations.

Given the tiny proportion of human genomes that have been immortalized as cell lines, it therefore seems most likely that this L1 insertion is specific for and unique to the HeLa cell line.

To improve the utility of this marker system, we adapted the genotyping assay to a single PCR. We also showed that the assay is sensitive to small amounts of contaminating cell line DNA, and that it works reliably with frozen cell pellets, a resource that is routinely available in cell culture laboratories.

In conclusion, we present a simple, rapid, and robust assay for detecting HeLa, the most common contaminating cell line. As it is PCR-based, insensitive to DNA condition and purity, and highly specific for HeLa, this assay should be accessible to any lab performing routine molecular biology. As such, it removes at least one barrier (that of cost) to routine cell line monitoring, and may contribute to the elimination of erroneous research based on contaminated or false cell lines.

Acknowledgments

Reference

1. Masters JR. HeLa cells 50 years on: the good, the bad and the ugly. Nat. Rev. Cancer. 2002; 2:315–319. [PubMed: 12001993]

2. MacLeod RA, Dirks WG, Matsuo Y, Kaufmann M, Milch H, Drexler HG. Widespread intraspecies cross-contamination of human tumor cell lines arising at source. Int. J. Cancer. 1999; 83:555–563. [PubMed: 10508494]

3. Stephenson EM. Locomotory invasion of human cervical epithelium and avian fibroblasts by HeLa cells in vitro. J. Cell Sci. 1982; 57:293–314. [PubMed: 7153261]

4. Lacroix M. Persistent use of "false" cell lines. Int. J. Cancer. 2008; 122:1–4. [PubMed: 17960586]

5. Nelson-Rees WA, Flandermeyer RR. HeLa cultures defined. Science. 1976; 191:96–98. [PubMed: 1246601]

6. Masters JR, Thomson JA, Daly-Burns B, Reid YA, Dirks WG, Packer P, Toji LH, Ohno T, et al. Short tandem repeat profiling provides an international reference standard for human cell lines. Proc. Natl. Acad. Sci. USA. 2001; 98:8012–8017. [PubMed: 11416159]

7. Hughes P, Marshall D, Reid Y, Parkes H, Gelber C. The costs of using unauthenticated, over-passaged cell lines: how much more data do we need? BioTechniques. 2007; 43:575–586. [PubMed: 18072586]

8. Uphoff CC, Drexler HG. Detection of mycoplasma contaminations. Methods Mol. Biol. 2005; 290:13–23. [PubMed: 15361652]

9. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, et al. Initial sequencing and analysis of the human genome. Nature. 2001; 409:860–921. [PubMed: 11237011]

10. Witherspoon DJ, Marchani EE, Watkins WS, Ostler CT, Wooding SP, Anders BA, Fowlkes JD, Boissinot S, et al. Human population genetic structure and diversity inferred from polymorphic L1(LINE-1) and Alu insertions. Hum. Hered. 2006; 62:30–46. [PubMed: 17003565]

11. Badge RM, Alisch RS, Moran JV. ATLAS: a system to selectively identify human-specific L1 insertions. Am. J. Hum. Genet. 2003; 72:823–838. [PubMed: 12632328]

12. van de Lagemaat LN, Gagnier L, Medstrand P, Mager DL. Genomic deletions and precise removal of transposable elements mediated by short identical DNA segments in primates. Genome Res. 2005; 15:1243–1249. [PubMed: 16140992]

13. Moore AE, Sabachewsky L, Toolan HW. Culture characteristics of four permanent lines of human cancer cells. Cancer Res. 1955; 15:598–602. [PubMed: 13261081]

14. Puck TT, Marcus PI, Cieciura SJ. Clonal growth of mammalian cells in vitro; growth characteristics of colonies from single HeLa cells with and without a feeder layer. J. Exp. Med. 1956; 103:273–283. [PubMed: 13286432]

15. Chen TR. Re-evaluation of HeLa, HeLa S3, and HEp-2 karyotypes. Cytogenet. Cell Genet. 1988; 48:19–24. [PubMed: 3180844]

16. Varley H, Di S, Scherer SW, Royle NJ. Characterization of terminal deletions at 7q32 and 22q13.3 healed by De novo telomere addition. Am. J. Hum. Genet. 2000; 67:610–622. [PubMed: 10924407]

17. Nardone RM. Eradication of cross-contaminated cell lines: a call for action. Cell Biol. Toxicol. 2007; 23:367–372. [PubMed: 17522957]

18. Nardone RM. Curbing rampant cross-contamination and misidentification of cell lines. Biotechniques. 2008; 45:221–227. [PubMed: 18816888]

19. Kent WJ. BLAT--the BLAST-like alignment tool. Genome Res. 2002; 12:656–664. [PubMed: 11932250]

20. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. The human genome browser at UCSC. Genome Res. 2002; 12:996–1006. [PubMed: 12045153]
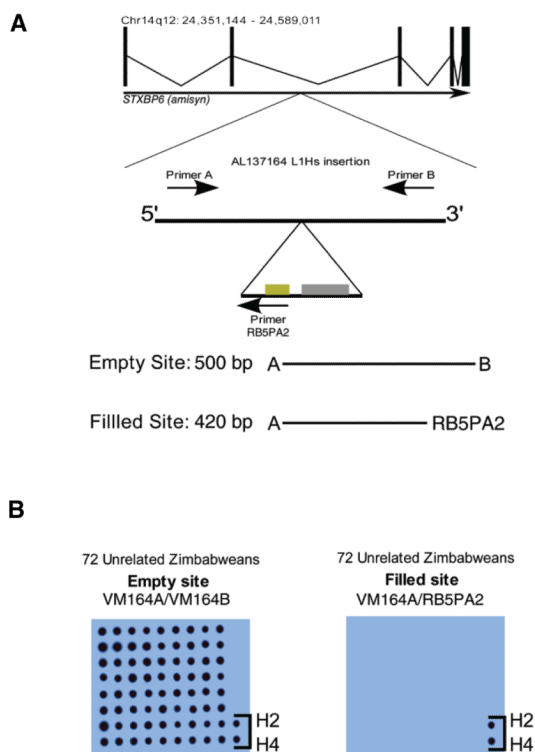
Figure 1. Genomic location and genotyping of the L1 retrotransposon insertion in accession no. AL137164

(A) Illustration of the genomic region containing the AL137164 L1Hs insertion and PCR primer placement. BLAT (19) searches of the genomic sequence flanking the L1 indicate that it occurs in intron 2 of the *STXBP6* (*amisyn*) gene. The genomic coordinates of *STXBP6* are indicated above the intron/exon structure diagram. The insertion does not occur within any of the repeat/indel/structural variation annotation tracks available at the University of California Santa Cruz's genome browser (20) (http://genome.ucsc.edu, data not shown). (B) The AL137164 L1 insertion is absent from a panel of unrelated Zimbabweans. The original two PCR genotyping assays were performed using genomic DNA from 72 unrelated Zimbabwean males. DNA from PCR reactions was immobilized on nylon membranes using a dotblot manifold and UV cross-linking, and hybridized with a radio-labeled probe derived from the empty site. No hybridization signals were detected on the filled site (L1-specific) membrane, with the exception of HeLa DNA–positive controls (H2/H4), while the empty site membrane showed hybridization for all DNA.
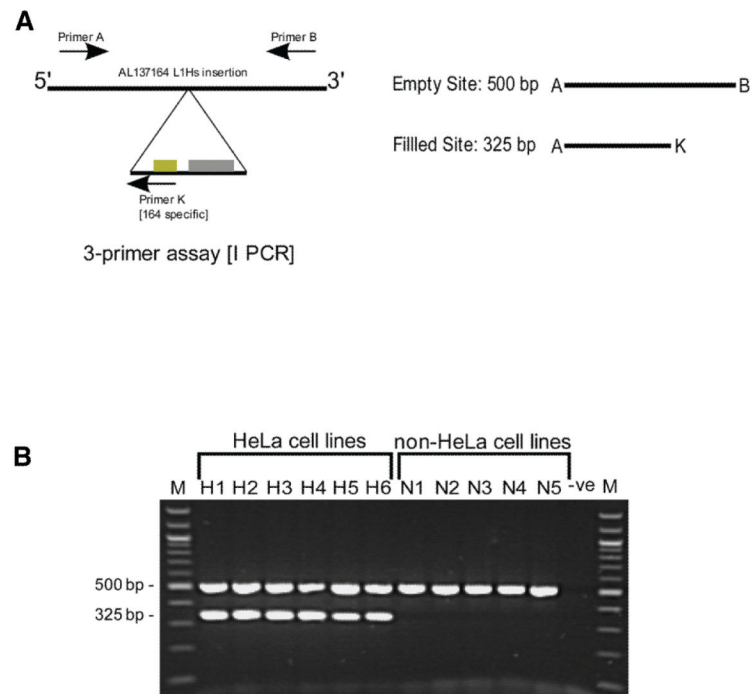
Figure 2. Duplex PCR design and genotyping results

(A) A duplex PCR assay for insertion AL137164. Primers A and B, specific to the AL137164 insertion site, were combined with a primer (K) specific to the junction between the L1 insertion and its 5′ flanking genomic DNA. The primers generate two amplicons of 500 bp and 325 bp from the empty and filled sites, respectively. (B) PCR amplification using the three-primer duplex assay. Genomic DNA from 6 independently sourced HeLa cell lines (H1, H2, H3, H4, H5, and H6) and 5 different non-HeLa cell lines (N1, N2, N3, N4, and N5) was subjected to PCR amplification. M, 100-bp ladder molecular weight marker; Lanes H1–H6, three-primer amplification of HeLa genomic DNA; Lanes N1–N5, three-primer amplification of non-HeLa cell line genomic DNA; -ve, negative control reaction (no input DNA).
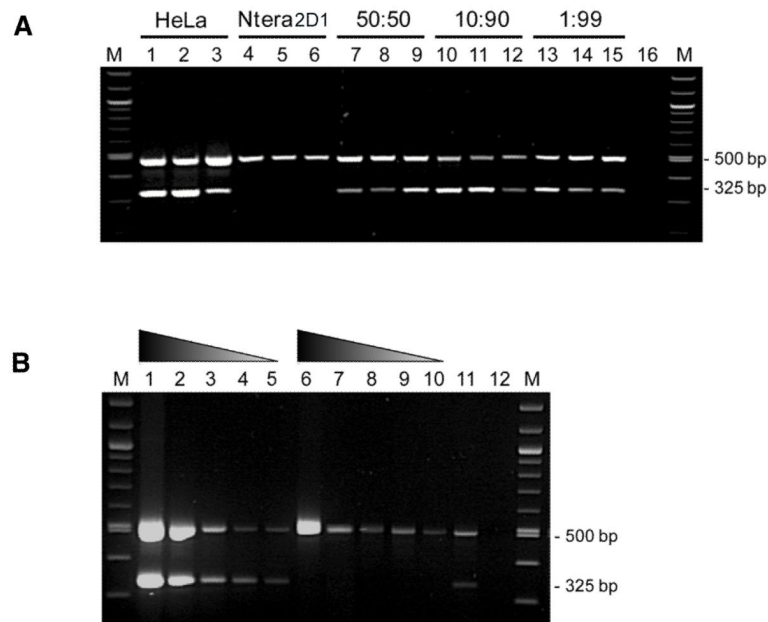
Figure 3. The duplex PCR assay can detect low-level HeLa contamination and can be applied to unpurified cell pellets

(A) The genotyping assay can detect low-level HeLa DNA contamination. Genomic DNA from the HeLa and Ntera2D1 cell lines was mixed in varying ratios and subjected to PCR amplification in triplicate, using the three-primer duplex genotyping assay. The HeLa-specific AL137164 amplicon is detectable at a ratio of 1% HeLa DNA to 99% NTera2D1 DNA. M, 100-bp ladder molecular weight marker; Lanes 1–3, 100% HeLa genomic DNA (20 ng/reaction); Lanes 4–6, 100% Ntera2D1 DNA (20 ng/reaction); Lanes 7–9, 50% HeLa, 50% Ntera2D1; Lanes 10–12, 10% HeLa, 90% Ntera2D1; Lanes 13–15, 1% HeLa, 99% Ntera2D1; Lane 16, PCR negative control (no input DNA). (B) Unpurified frozen cell pellets are suitable genotyping substrates. Frozen cell pellets (~1 × 10^6 cells/mL) from HeLa and non-HeLa cell cultures were defrosted, diluted in 5 mM Tris-HCl, pH 7.0 and boiled for 5 min. The boiled extracts were amplified using the three-primer duplex genotyping assay. M, 100-bp ladder molecular weight marker; Lane 1, input 50% HeLa cell pellet, 50% 5 mM Tris-HCl, pH 7.0; Lane 2, input 10% HeLa cell pellet, 90% 5mM Tris-HCl, pH 7.0; Lane 3, 5% HeLa, 95% 5 mM Tris-HCl, pH 7.0; Lane 4, 2% HeLa, 98% 5 mM Tris-HCl, pH 7.0; Lane 5, 1% HeLa cell pellet, 99% 5mM Tris-HCl, pH 7.0; Lane 6, input 50% non-HeLa cell pellet, 50% 5mM Tris-HCl, pH 7.0; Lane 7, 10% non-HeLa cell pellet, 90% 5mM Tris-HCl, pH 7.0; Lane 8, 5% non-HeLa cell pellet, 95% 5 mM Tris-HCl, pH 7.0; Lane 9, 2% non-HeLa cell pellet, 98% 5mM Tris-HCl, pH 7.0; Lane 10, 1% Non-HeLa cell pellet, 99% 5 mM Tris-HCl, pH 7.0; Lane 11, PCR positive control (10 ng HeLa gDNA); Lane 12, PCR negative control (no input DNA). The gradient triangles indicate the proportion of cell-pellet extract in the input to the genotyping PCR.

Table 1

Cell Lines Used in This Study

Culture	Culture information	Culture source
H1	HeLa	Andrew Fry (University of Leicester, Leicester, UK)
H2	HeLa	Fred Gage (Salk Institute, La Jolla, CA, USA)
H3	HeLa	Raj Patel (University of Leicester)
H4	HeLa	John Moran (University of Michigan, Ann Arbor, MI, USA)
H5	Hep2 [morphologically identical to HeLa (13) and by STR profiling]	Simon Kilvington (University of Leicester)
H6	HeLa S3 [clonal subline derived from the original HeLa culture (14,15)]	ECACC/HPA (Salisbury, UK)
N1	AJ (patient lymphoblastoid cell line) (16)	Nicola Royle (University of Leicester)
N2	BJ (primary foreskin fibroblast cell line)	Nicola Royle; ATCC USA (Manassas, VA, USA)
N3	GM03798 (lymphoblastoid cell line)	Nicola Royle; Coriell Cell Repositories (Camden, NJ, USA)
N4	NTera2D1 teratocarcinoma	Nicola Royle; ATCC USA
N5	HepG2	Fred Tata (University of Leicester); ECACC/HPA