Cell line fingerprinting using retroelement insertion polymorphism

Svetlana V. Ustyugova, Anna L. Amosova, Yuri B. Lebedev, and Eugene D. Sverdlov

Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Moscow, Russia

BioTechniques 38:561-565 (April 2005)

Human cell lines are an indispensable tool for functional studies of living entities in their numerous manifestations starting with integral complex systems such as signal pathways and networks, regulation of gene ensembles, epigenetic factors, and finishing with pathological changes and impact of artificially introduced elements, such as various transgenes, on the behavior of the cell. Therefore, it is highly desirable to have reliable cell line identification techniques to make sure that the cell lines to be used in experiments are exactly what is expected. To this end, we developed a set of informative markers based on insertion polymorphism of human retroelements (REs). The set includes 47 pairs of PCR primers corresponding to introns of the human genes with dimorphic LINE1 (L1) and Alu insertions. Using locus-specific PCR assays, we have genotyped 10 human cell lines of various origins. For each of these cell lines, characteristic fingerprints were obtained. An estimated probability that two different cell lines possess the same marker genotype is about 10^{-18} . Therefore, the proposed set of markers provides a reliable tool for cell line identification.

INTRODUCTION

Cell lines are an indispensable, renewable resource for numerous studies in various fields of modern biology and medicine (1–5). It is commonly assumed that the cell features in a particular cell line and those in the human tissue from which this line originated are similar. However, most human cell lines are prone to various sorts of genetic rearrangements that affect biochemical, regulatory, and other phenotypic features of cells during their cultivation (6). Various chromosomal abnormalities, including aneuploidy, numerous rearrangements, and loss of chromosome regions, are characteristic alterations of cell genomes, especially if the cells are of tumor origin (7,8). Therefore, it would be very useful to permanently monitor the authenticity of the cell lines and/or to have a comprehensive set of standard tests for confirming the cellular identity. Routinely used descriptions of established cell lines include their origin and cultivation conditions, but this can be insufficient. Appropriate tests should be performed to authenticate the cell line identity. To test the identity, pheno-

typic (e.g., morphological analysis) or genotypic characteristics can be used. In many cases, isoenzyme analysis may be applied to identify the cell line species of origin. An alternative strategy would be to demonstrate the presence of unique markers, for example, by using banding cytogenetics to detect a unique marker chromosome or DNA analysis to detect a genomic polymorphism pattern (e.g., restriction fragment-length polymorphism, variable number of tandem repeats, or genomic dinucleotide repeats). Expression of a particular product may represent a complementary approach for confirmation of identity. At present, several types of tests, including those for tissue-specific marker expression, detection of a karyotype, and chromosome aberrations by G-banding or by fluorescence in situ hybridization (FISH), are in rather common use for characterization of distinct cell lines (9). Modern approaches for more scrupulous examination of cell lines also include multiplex-FISH (M-FISH), spectral karyotyping (SKY), comparative genomic hybridization (CGH), fluorescence immunophenotyping and interphase cytogenetics as

a tool for investigation of neoplasms (FICTION), and oligonucleotide-primed in situ DNA synthesis (PRINS) (9–12). However, these methods are time-consuming and require high-technology instrumentation. It is highly desirable, therefore, to have a standard set of genomic markers that would allow a fast, reliable, and efficient authentication of the cell line identity.

One type of informative markers could be polymorphic insertions of various retroposons. Insertion polymorphism of retroelements (REs) is now attracting considerable attention. Due to some features, this type of polymorphism seems to be advantageous over the others [single nucleotide polymorphism (SNP) or microsatellites]. These features include a known ancestral state (absence of RE), stability of insertion (there is no special mechanism of removing inserted REs), and relatively easy detection by standard locus-specific PCR assay. Moreover, an event of two independent integrations into one and the same locus is highly improbable (13,14).

Here we report a set of polymorphic LINE1 (L1) and Alu markers and its application for genotyping of human cell lines with different phenotypes and of various tissue origins. We demonstrate that each cell line is characterized by a unique pattern of distribution of the RE-containing and not containing alleles that can serve as a fingerprint for the cell line. We also believe that the data obtained will be of use for hybrid cell production and characterization.

MATERIALS AND METHODS

Cell Lines and DNA Samples

The human cell lines used in this work were as follows: Tera-1 (human testicular malignant embryonal carcinoma; ATCC, Manassas, VA, USA), Tera-2 (pluripotent human testicular embryonal carcinoma; ATCC), HL-60 (acute promyelocytic leukemia; ATCC), RMS-13 (human fibroblast rhabdomyosarcoma; ATCC), HEK293 (primary human embryonal kidney transformed by human adenovirus type 5 DNA; ATCC), HeLa (human epithelial adenocarcinoma;

ATCC), HT1080 (human epithelial fibrosarcoma; ATCC), and Jurkat (acute T cell leukemia; ATCC). These cell lines were obtained from the collection of the Institute of Cytology, St. Petersburg, Russia and maintained as recommended. Cell lines NGP-127 (neuroblastoma) and OsA-CL (osteocarcinoma) were kindly provided by Paul S. Meltzer (NHGRI, NIH, Bethesda, MD, USA). Additional human DNA samples were isolated from individual human placentas, and nonhuman DNA samples were isolated from chimpanzee (*Pan troglodytes*) peripheral blood lymphocytes. All DNA samples were obtained using proteinase K digestion and phenol extraction as described elsewhere (15).

Oligonucleotides

Primers for genomic PCR and L1-specific primers were designed using the Primer 3.0 (www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi) software. The primer pairs for genomic PCR are shown in Supplementary Table S1, which can be seen at the *BioTechniques*' web site at www.BioTechniques.com. Oligonucleotide primers for PCR amplification were synthesized using an ASM-102U DNA synthesizer (Biosset Ltd., Novosibirsk, Russia).

PCR Amplification

PCR amplification of 26 individual L1 and 21 Alu elements was performed in 25- μ L reaction volumes that contained 20–50 ng genomic DNA, 10 pmol each of oligonucleotide primers, 125 μ M each of dNTPs, 16.6 mM $(\text{NH}_4)_2\text{SO}_4$, 67 mM Tris-HCl buffer, pH 8.8, 1.5 mM MgCl_2 , and 0.5 U *Taq* DNA polymerase (SibEnzyme, Novosibirsk, Russia). The strategy of locus-specific PCR is illustrated in Figure 1 and described in the Results and Discussion section. Apart from primers corresponding to 5' (G-forward) and 3' flanks (G-reverse) of RE integration sites, Ta L1-specific primers were used in several cases. L1-forward primer (5'-GGAGATATACCTAATGCTAG-ATGACACA-3') corresponded to the 3' untranslated region (UTR) of the Ta L1Hs consensus sequence, whereas L1-reverse primer (5'-GCTGTTCTATTC-

GGCCATCT-3') was complementary to the outermost 5' end of full-size L1Hs elements. DNA in each reaction mixture was predenatured for 2 min at 95°C and then PCR-amplified for 30 cycles at 95°C for 20 s, at the annealing temperature (specified for each locus in Supplementary Table S1) for 20 s, and at 72°C for 1 min elongation. PCRs were performed in either an MJ Research PTC-200 thermal cycler (Bio-Rad Laboratories, Hercules, CA, USA) or an OmniGene HB-TR3-CM thermal cycler (Hybaid Ltd, Teddington, Middlesex, UK). Samples (8 μ L) of each resulting reaction mixture were taken and fractionated in 1.5% agarose gels supplemented with 0.05 μ g/mL ethidium bromide. PCR products were visualized using a UV illuminator.

RESULTS AND DISCUSSION

Insertions of L1 and Alu REs were used to choose informative genetic markers for cell line genotyping for several reasons. A subset of the AluY family and L1s of the Ta subfamily retained their retroposition activity up to the present time, so one in every 50 or 200 humans has a new genomic L1 or Alu insertion that occurred in parental germ cells (16–18). At least 45% of 500 known Ta L1 insertions and roughly 1200 AluY insertions are thought to be dimorphic in humans and could serve as "identical-by-descent" genetic markers (17–22). Several REs polymorphic with respect to the presence or absence of their insertions in individual genomes were located within intronic parts of human genes (23–27).

To develop a set of markers, we have selected 47 human genes known to contain dimorphic L1 or Alu insertions within their introns (19,25–27) and used these insertions to genotype 10 human cell lines. To detect the presence or absence of RE insertions in each of the loci, a locus-specific PCR assay was used, and RE(+/-) alleles were identified by the length of the corresponding PCR products. The strategy of the PCR assay and various types of PCR products are schematized in Figure 1A. Alu insertions were detected by PCR with unique genomic primers that corresponded to genomic DNA sequences flanking each

of the RE integration sites (primers G-forward and G-reverse in Figure 1, A and B, and in Supplementary Table S1).

Since most of Ta L1 elements in the human genome are 5' truncated, both L1-containing and L1-lacking alleles are amplified during single PCR with a pair of unique primers flanking the insert (Figure 1A and Supplementary Table S1). The formation of a longer PCR product (P1 in Figure 1A) indicates the presence of an RE insertion in the corresponding DNA locus, while the amplification of an RE-lacking allele results in a shorter PCR fragment (P2 in Figure 1A). A simultaneous formation of both types of PCR fragments suggests a genotype heterozygous for the insertion. For example, the results presented in Figure 1C indicate (L1+/L1-) heterozygosity at *AUTS2* locus in cell lines Tera-2, HeLa, and in one of the placenta samples. The remaining genomes were shown to be (L1-/L1-) homozygous at *AUTS2*. In the same way, 11 loci containing known and predicted human genes with dimorphic intronic insertions of relatively short truncated Ta L1Hs elements were analyzed (Supplementary Table S1).

Fifteen other investigated L1 loci were found to be dimorphic for insertions of 1.7–6 kb long L1 elements (for detail, see Supplementary Table S1). Due to the large size of these insertions, successful simultaneous PCR amplification of L1-containing and L1-lacking fragments was rather problematic, which required us to carry out additional parallel amplifications using L1-specific primer(s) and one of the flanking primers. The first of the L1-specific primers corresponded to part of the 3' UTR L1 sequence and carried at its 3' end the three nucleotides ACA, characteristic of the Ta subfamily. Loci presumably containing full-size L1 elements were amplified with G-forward flanking specific primer and L1-reverse universal primer against the 5' UTR of the Ta L1 sequence. PCR amplifications resulted in the formation of P2-, P3-, and P4-type PCR fragments as shown in Figure 1D for the *RPIA* locus analysis. All 21 AluYs tested in this study demonstrated insertional dimorphism in at least one of 10 investigated genomes.

Supplementary Table S2 shows that, in total, 38 (17 L1 and 21 Alu) RE insertions were found to be polymorphic in the human cell lines. Supplementary

Table S2 also presents genotyping patterns characteristic of each tested cell line. Pairwise comparisons of these patterns revealed from 5 (HeLa/Tera-2) to 26 (HL60/NGP-127), with an average of 18, loci distinguishing one cell line from another. For example, the genomes of Tera-1 and Tera-2 differed in 18 of 47 investigated loci. Taking into account that a diploid cell

line may have three genotypes for each locus (+/+, +/-, or -/-) and assuming for simplicity that all these genotypes are equally probable, one can calculate that there can be over 10^{18} ($1/3^{38}$) unique genotypes differing in combinations of 38 dimorphic loci. Thus, the probability that two different cell lines have identical genotyping patterns is extremely low, and the developed set

of polymorphic markers is sufficient to provide an individual fingerprint for any human cell line.

However, as can be seen from Supplementary Table S2, distribution of three genotypes (+/+, +/-, or -/-) is asymmetric for most loci. Moreover, for routine cell line identification, the number of tested loci can be reduced and restricted to the most informative loci. Therefore, we have developed the following protocol for routine cell line identification using a subset of locus-specific markers for the *ABLIM2*, *ACTG2*, *AK127695*, *AMPH*, *BHC80*, *BPGM*, *DSCAM*, *FLJ21269*, *GRID2*, *LAMA2*, *MYRIP*, *NF1*, *NRP1*, *SEMA3A*, *SLC25A18*, and *ZFPM2* genes.

First, 18 tubes are needed for PCR amplification. To 16 of the tubes, add 10 pmol each of G-forward and G-reverse specific primers to amplify fragments of the *ABLIM2*, *ACTG2*, *AK127695*, *AMPH*, *BPGM*, *DSCAM*, *FLJ21269*, *GRID2*, *LAMA2*, *MYRIP*, *NF1*, *NRP1*, *SEMA3A*, *BHC80*, *SLC25A18*, and *ZFPM2* genes. To the remaining 2 tubes, add 10 pmol of the L1-forward primer and 10 pmol of the G-reverse primer that corresponds to the *FLJ21269* and *GRID2* gene loci, respectively. Adjust the volume of the PCR mixture in each tube to 25 μ L with 1 $\times$ PCR buffer containing 125 μ M each of four dNTPs, 20 U/mL *Taq* DNA polymerase, and 0.8–1 μ g/mL of the cell line genomic DNA as template. The amplification reaction is performed as follows: predenaturation at 95°C for 2 min, then 30 cycles at 95°C for 20 s (denaturation), at annealing temperatures (T_{ann} ; see Supplementary Table S3) for 20 s (annealing), and at 72°C for 1 min (elongation). PCR products are analyzed by electrophoresis in 1.2% agarose gels. Genotyping patterns characterizing cell lines of interest are constructed and compared with those of reference genotypes.

An alternative and probably less time-consuming approach to the practical use of the developed cell line fingerprinting might be multiplex locus-specific PCR. In this case, the markers can be subdivided into groups in which the annealing temperature of specific primers should be identical, while the corresponding PCR products

Figure 1. Detection of retroelement (RE) insertions in the cell genome. (A) Locus-specific PCR strategy. Non-RE genomic DNA is represented by lines, and RE sequences is represented by boxes. Positions and orientations of G-forward (G-for), G-reverse (G-rev), L1-reverse (L1-rev), and L1-forward (L1-for) primers are shown by arrows. Open triangles denote target site duplications (TSDs). AAA represents the poly(A) tail of the RE. Different types of PCR products are denoted by framed letters P1–P4. (B) An electrophoretic pattern of PCR products for Alu containing *TMC5* locus. Framed letters P1 and P2 indicate expected positions of P1- and P2-type PCR products, respectively. The sources of genomic DNAs used as templates for the PCR are shown above the lanes. Lane M, a 100-bp ladder length marker (SibEnzyme). (C) An electrophoretic pattern of PCR products for L1 containing *AUTS2* locus. (D) An electrophoretic pattern of PCR products for L1 containing *RPIA* locus.

should be different in length.

A number of such groups can be formed for the gene loci used here for the fingerprinting. One of them, with a T_{ann} of 60°C, includes the markers of L1-containing genes *BPGM*, *FLJ21269*, and *GRID2*, the amplification products of which can be easily identified among total multi-PCR products, because their lengths are markedly different (592/169/276, 122/224, and 330/360 bp, respectively).

We hope that the markers proposed in this report, the number of which can certainly be increased, will be useful for developing standard sets of genetic signatures to reliably identify human cell lines in laboratory practice. They also can be successfully used for construction of hybrid cell lines and characterization of their genetic content.

ACKNOWLEDGMENTS

We thank Boris Glotov for the critical reading of this manuscript and for valuable comments and Victor Potapov and Nadezhda Skaptsova (Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Moscow, Russia) for synthesis of the oligonucleotides. This work was supported by INTAS-01-0759, Russian Foundation for Basic Research 02-04-48614 grants, and the Molecular and Cellular Biology Program of the Presidium of the Russian Academy of Sciences.

COMPETING INTERESTS STATEMENT

The authors declare no competing interests.

REFERENCES

- Carpenter, A.E. and D.M. Sabatini. 2004. Systematic genome-wide screens of gene function. *Nat. Rev. Genet.* 5:11-22.
- Nishizuka, S., L. Charboneau, L. Young, S. Major, W.C. Reinhold, M. Waltham, H. Kouros-Mehr, K.J. Bussey, et al. 2003. Proteomic profiling of the NCI-60 cancer cell lines using new high-density reverse-phase lysate microarrays. *Proc. Natl. Acad. Sci. USA* 100:14229-14234.
- Nishizuka, S., S.-T. Chen, F.G. Gwadry, J. Alexander, S.M. Major, U. Scherf, W.C. Reinhold, M. Waltham, et al. 2003. Diagnostic markers that distinguish colon and ovarian adenocarcinomas: identification by genomic, proteomic, and tissue array profiling. *Cancer Res.* 63:5243-5250.
- Zhumabayeva, B., A. Chenchik, P.D. Siebert, and M. Herrler. 2004. Disease profiling arrays: reverse format cDNA arrays complementary to microarrays. *Adv. Biochem. Eng. Biotechnol.* 86:191-213.
- Ross, D.T., U. Scherf, M.B. Eisen, C.M. Perou, C. Rees, P. Spellman, V. Iyer, S.S. Jeffrey, et al. 2000. Systematic variation in gene expression patterns in human cancer cell lines. *Nat. Genet.* 24:227-235.
- Hiorns, L.R., T.D. Bradshaw, L.A. Skelton, Q. Yu, L.R. Kelland, and B. Leyland-Jones. 2004. Variation in RNA expression and genomic DNA content acquired during cell culture. *Br. J. Cancer* 90:476-482.
- Bonassi, S., A. Znaor, H. Norppa, and L. Hagmar. 2004. Chromosomal aberrations and risk of cancer in humans: an epidemiologic perspective. *Cytogenet. Genome Res.* 104:376-382.
- Zhao, X., C. Li, J.G. Paez, K. Chin, P.A. Janne, T.-H. Chen, L. Girard, J. Minna, et al. 2004. An integrated view of copy number and allelic alterations in the cancer genome using single nucleotide polymorphism arrays. *Cancer Res.* 64:3060-3071.
- Naumann, S., D. Reutzel, M. Speicher, and H.J. Decker. 2001. Complete karyotype characterization of the K562 cell line by combined application of G-banding, multiplex-fluorescence in situ hybridization, fluorescence in situ hybridization, and comparative genomic hybridization. *Leuk. Res.* 25:313-322.
- Lindbjerg, A.C., M. Ostergaard, B. Nielsen, B. Pedersen, and J. Koch. 2000. Characterization of three hairy cell leukemia-derived cell lines (ESKOL, JOK-1, and hair-M) by multiplex-FISH, comparative genomic hybridization, FISH, PRINS, and dideoxyPRINS. *Cytogenet. Cell Genet.* 90:30-39.
- Harris, C.P., X.Y. Lu, G. Narayan, B. Singh, V.V. Murty, and P.H. Rao. 2003. Comprehensive molecular cytogenetic characterization of cervical cancer cell lines. *Genes Chromosomes Cancer* 36:233-241.
- Roschke, A.V., T. Giovanni, S.G. Kristen, N. McTyre, K.J. Bussey, S. Lababidi, D.A. Scudiero, N. John, et al. 2003. Karyotypic complexity of the NCI-60 Drug-Screening Panel. *Cancer Res.* 63:8634-8647.
- Roy-Engel, A.M., M.L. Carroll, M. El-Sawy, A.H. Salem, R.K. Garber, S.V. Nguyen, P.L. Deininger, and M.F. Batzer. 2002. Non-traditional Alu evolution and primate genomic diversity. *J. Mol. Biol.* 316:1033-1040.
- Vincent, B.J., J.S. Myers, H.J. Ho, G.E. Kilroy, J.A. Walker, W.S. Watkins, L.B. Jorde, and M.A. Batzer. 2003. Following the LINES: an analysis of primate genomic variation at human-specific LINE-1 insertion sites. *Mol. Biol. Evol.* 20:1338-1348.
- Sambrook, J. and D.W. Russell. 2001. *Molecular Cloning: A Laboratory Manual*, 3rd ed. p. 6.6-6.7. CSH Laboratory Press, Cold Spring Harbor, NY.
- Ostertag, E.M., R.J. DeBerardinis, J.L. Goodier, Y. Zhang, N. Yang, G.L. Gerton, and H.H. Kazazian. 2002. A mouse model of human L1 retrotransposition. *Nat. Genet.* 32:655-660.
- Batzer, M.A. and P.L. Deininger. 2002. Alu repeats and human genomic diversity. *Nat. Rev. Genet.* 3:370-379.
- Kazazian, H.H. 2004. Mobile elements: drivers of genome evolution. *Science* 303:1626-1632.
- Myers, J.S., B.J. Vincent, H. Udall, W.S. Watkins, T.A. Morrish, G.E. Kilroy, G.D. Swergold, J. Henke, et al. 2002. A comprehensive analysis of recently integrated human Ta L1 elements. *Am. J. Hum. Genet.* 71:312-326.
- Watkins, W.S., A.R. Rogers, C.T. Ostler, M.J. Bamshad, A.-M.E. Brassington, M.L. Carroll, S.V. Nguyen, J.A. Walker, et al. 2003. Genetic variation among world populations: inference from 100 Alu insertion polymorphisms. *Genome Res.* 13:1607-1618.
- Boissinot, S., A. Entezam, L. Young, P.J. Munson, and A.V. Furano. 2004. The insertional history of an active family of L1 retrotransposons in humans. *Genome Res.* 14:1221-1231.
- Carter, A.B., A.-H. Salem, D.J. Hedges, C.N. Keegan, B. Kimball, J.A. Walker, W.S. Watkins, L.B. Jorde, and M.A. Batzer. 2004. Genome-wide analysis of the human Alu Yb-lineage. *Hum. Genomics* 1:1-13.
- Boissinot, S., P. Chevret, and A.V. Furano. 2000. L1 (LINE-1) retrotransposon evolution and amplification in recent human history. *Mol. Biol. Evol.* 17:915-928.
- Sheen, F.M., S.T. Sherry, G.M. Risch, M. Robichaux, and I. Nasidze. 2000. Reading between the LINES: human genomic variation induced by LINE-1 retrotransposition. *Genome Res.* 10:1496-1508.
- Carroll, M.L., A.M. Roy-Engel, S.V. Nguyen, A.-H. Salem, E. Vogel, B. Vincent, J. Myers, Z. Ahmad, et al. 2001. Large-scale analysis of the Alu Ya5 and Yb8 subfamilies and their contribution to human genomic diversity. *J. Mol. Biol.* 311:17-40.
- Buzdin, A., S. Ustyugova, E. Gogvadze, Y. Lebedev, G. Hunsmann, and E. Sverdlov. 2003. Genome-wide targeted search for human specific and polymorphic L1 integrations. *Hum. Genet.* 112:527-533.
- Mamedov, I.Z., E.S. Arzumanyan, A.L. Amosova, Y.B. Lebedev, and E.D. Sverdlov. 2005. Whole-genome experimental identification of insertion/deletion polymorphisms of interspersed repeats by a new general approach. *Nucleic Acids Res.* 33:e16.

Received 3 September 2004; accepted 19 October 2004.

Address correspondence to Yuri B. Lebedev, Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, 16/10 Miklukho-Maklaya, 117997 Moscow, Russia. e-mail: yuri@humgen.siobc.ras.ru