

Original Article

Open Access



Innovative use of forensic STRs for authentication of human cell lines stored over 34 years

Anqi Chen¹, Ji Chen¹, Yujia Xuan¹, Jinyuan Zhao¹, Chaoneng Ji², Suhua Zhang¹

¹Institute of Forensic Science, Fudan University, Shanghai 200032, China.

²State Key Laboratory of Genetic Engineering, School of Life Sciences, Fudan University, Shanghai 200438, China.

Correspondence to: Prof. Suhua Zhang, Institute of Forensic Science, Fudan University, No. 131 Dongan Road, Xuhui district, Shanghai 200032, China. E-mail: zhang_suhua@fudan.edu.cn

How to cite this article: Chen A, Chen J, Xuan Y, Zhao J, Ji C, Zhang S. Innovative use of forensic STRs for authentication of human cell lines stored over 34 years. *J Transl Genet Genom.* 2025;9:149-166. <https://dx.doi.org/10.20517/jtgg.2025.08>

Received: 13 Jan 2025 **First Decision:** 4 Jun 2025 **Revised:** 16 Jun 2025 **Accepted:** 26 Jun 2025 **Published:** 30 Jun 2025

Academic Editor: Taosheng Huang **Copy Editor:** Ping Zhang **Production Editor:** Ping Zhang

Abstract

Aim: In forensic science, the authentication of biological samples is critical for accurate analysis and evidence validation. To evaluate the authenticity and genetic stability of 91 long-term preserved human cell line samples under cryogenic conditions over 34 years, and to explore the applicability of forensic-grade short tandem repeat (STR) profiling for long-term cell line authentication.

Methods: We conducted STR-based authentication using 23 forensic STR markers. STR profiles were analyzed using the Tanabe and Masters algorithms for authentication. The study also examined the effects of prolonged passaging and genetic modification on STR stability, and investigated potential contamination events by comparing profiles with public databases.

Results: All uniquely labeled human cell lines were successfully revived and yielded complete STR profiles, confirming the efficacy of long-term cryopreservation. One male cell line showed the Y-indel in the absence of a Y-allele at Amelogenin, indicating that the Y-indel alone is insufficient for accurate sex determination. Authentication using the Tanabe and Masters algorithms revealed that stringent criteria provided more definitive results than scoring-based methods. Instances of relatedness among unauthenticated cell lines to problematic references in existing databases suggest early-stage contamination rather than recent laboratory errors. STR profile stability persisted through routine cell line passages, and genetically edited cell lines retained profiles identical to their parental counterparts, indicating a need for additional differentiation markers for accurate derivation authentication. No significant differences in the number of altered alleles between normal and tumor cell lines,



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.



implying cell type alone did not dictate STR stability.

Conclusion: This study represented one of the most extensive single-laboratory investigations into cell line preservation using forensic-grade tools, providing new reference alleles and valuable insights into cell line authentication. The findings supported the application of forensic STR kits beyond traditional forensic samples, offering a new perspective on their utility in genetic research and laboratory management.

Keywords: Cell line authentication, human cell lines, HeLa contamination, STR profiling, STR status

INTRODUCTION

Cell lines serve as essential experimental models, which have been widely utilized in biomedical research and drug development^[1]. However, frequent exchange of cell lines between laboratories increases the risk of mislabeling, misidentification, and contamination, leading to incorrect and irreproducible research outcomes^[2]. In 1967, Walter Nelson-Rees first reported cross-contamination in human cell lines, but his findings did not gain much immediate attention, allowing problematic cell lines to persist in published studies^[3]. It was not until the 1980s that the scientific community began to recognize this issue^[4]. Since then, increasing attention has been paid to cell line contamination, with many journals now requiring authentication as a prerequisite for publication.

Among various authentication methods^[5,6], short tandem repeat (STR) profiling stands as the gold standard^[7], especially in forensic science due to its precision in identifying genetic variation. Initially, the American Type Culture Collection (ATCC) Standards Development Organization Workgroup recommended eight STR markers for human cell line authentication, later expanding to 13 STRs to improve accuracy^[8]. Forensic STR markers are highly stringent and exhibit greater polymorphism, offering strong discriminatory power^[9]. The inclusion of additional STR loci in authentication can further enhance accuracy, similar to forensic science applications where testing more markers improves individual identification reliability^[10]. As the diversity of cell lines increases, expanding STR loci will likely become necessary to meet evolving authentication needs, reflecting a forensic-like emphasis on precision and reliability. Therefore, adopting forensic-grade STR markers for cell line authentication represents an innovative approach ensuring more robust and reliable analysis.

Researchers who conduct cell-based experiments seek to confirm that their cell lines are correctly labeled and free from aberrations. In most cases, cell line authentication aims to prevent cross-contamination, misidentification, and genetic drift. Furthermore, assessing the impact of long-term storage on cell line integrity is essential, yet comprehensive studies in this area remain limited. Our laboratory began using cell lines in 1990 and has documented 91 human cell lines by a limited, experienced team to ensure reliability.

Given these circumstances, it was imperative to conduct a thorough investigation of our cell line inventory. The objectives of this study were threefold: (1) to assess the preservation status of cell lines stored over the past 34 years and identify any potential contamination events that may compromise research outcomes; (2) to explore the genetic changes that occurred in cell lines during long-term passaging, thereby evaluating the stability and variation of STR loci over time; and (3) to demonstrate the application of forensic-standard STR profiling combined with comparative algorithms as a reliable approach for cell line authentication. These questions are essential for understanding the integrity of stored biological resources and for ensuring the validity of experimental results in biomedical research.

MATERIALS AND METHODS

Sample selection

According to our laboratory regulations, all cell line samples stored in the liquid nitrogen tank were documented on paper. To ensure comprehensive backup, samples with different names were treated as distinct cell line strains for subsequent STR genotype analysis. For instance, samples labeled as 293FT, HEK293 LC, HEK293, 293y99, K562, k562, HUH7, and HUH7-2 were considered as eight different samples, rather than three different cell line strains (HEK293, K562, and HUH7). This approach resulted in the selection of 91 different cell line samples for the subsequent cell line backup [[Supplementary Table 1](#)].

Cell line culture and DNA extraction

The 91 cell line samples were selected for subsequent cell line cultivation. The cell lines referenced in publications were cultured following the experimental conditions described in the respective articles. For the cell lines not involved in any publications, cultivations were performed according to the vendor's instructions.

Genomic DNA was extracted from 5×10^6 cells using the QIAamp DNA Blood Mini Kit (Qiagen, Venlo, The Netherlands), according to the manufacturer's instructions. DNA quantification was performed using a Qubit fluorometer (Life Technologies, CA, USA), and all DNA samples were stored at -80°C until use.

STR genotyping and alternation status evaluation

STRs were analyzed using the SiFaSTR™ 23-plex system (Academy of forensic sciences, Shanghai, China), which includes 21 autosomal STRs (D3S1358, D5S818, D2S1338, TPOX, CSF1PO, Penta D, TH01, vWA, D7S820, D21S11, Penta E, D10S1248, D8S1179, D1S1656, D18S51, D12S391, D6S1043, D19S433, D16S539, D13S317, and FGA) and two sex-related polymorphisms (Amelogenin and Y indel). The PCR reactions were conducted according to the manufacturer's protocol, and DNA genotyping was performed in a Classic 116 Genetic Analyzer (SUPERYEARS, Nanjing, China) using GeneManager Software (SUPERYEARS, Nanjing, China).

The alteration status evaluation focused solely on the 21 autosomal STRs. By comparing the query genotype with the reference genotypes of the cell lines, five types of statuses were determined^[11,12] as follows: (1) stable (S): no alteration occurred; (2) loss of heterozygosity (L): an allele was lost in the query cell line sample compared to the reference alleles; (3) occurrence of an additional allele (Aadd): an additional allele appeared in the query cell line sample (e.g., allele 16, 17 > allele 16, 17, 19); 4. occurrence of a new allele (Anew): allele replacement occurred in the query cell line sample (e.g., allele 16, 17 > allele 16, 19).

Human cell line authentication

Conventionally, cell line authentication is performed by comparing the obtained STR profiles to those registered in the cell line databases^[13]. In this study, two matching algorithms were applied to evaluate the relatedness between the query sample and the reference cell lines. The scores were computed using the following algorithms: (1) Tanabe algorithm: percent match = $\frac{2 \times \text{number shared alleles}}{\text{total number of alleles in the query profile} + \text{total number of alleles in the reference profile}} \times 100\%$ (2) Masters algorithm: percent match = $\frac{\text{number shared alleles}}{\text{total number of alleles in the query profile}} \times 100\%$. Different algorithms applied distinct similarity thresholds to classify the relationship between two cell lines^[14-16]. According to the Tanabe algorithm, a similarity score ranging from 90% to 100% indicated that the two profiles were related, likely originating from the same donor. Scores between 80% and 90% were considered ambiguous or mixed results, suggesting possible relatedness but requiring further investigation. Scores below 80% were regarded as unrelated. In contrast, the Masters algorithm was slightly more lenient. A similarity score of 80% or higher indicated relatedness, while scores from 60% to 80% suggested mixed or uncertain results. Scores below 60% were interpreted as

unrelated, reflecting a likely distinct origin. Tanabe's "Related" threshold ($\geq 90\%$) was stricter than that of Masters ($\geq 80\%$) due to differences in the underlying allele-matching algorithms. In Tanabe's method, the number of shared alleles was multiplied by 2 and divided by the total number of alleles in both profiles. This formula placed greater emphasis on exact matches and penalized allele imbalances more heavily, especially in polyploid or contaminated lines.

To search for the correct reference cell line, we utilized the obtained STR results and conducted a search using the online STR similarity search tool CLASTR (Cell Line Authentication using STR, version 1.4.4) (<https://www.cellosaurus.org/str-search/>). The matching scores were recalculated using the two algorithms based on the highest-scoring results obtained from CLASTR 1.4.4.

Statistics analysis

Statistical analyses were performed using Prism 9.0 software (GraphPad, San Diego, USA) and Microsoft Excel 2019 (Microsoft Corporation, Redmond, USA). An unpaired t-test was conducted to compare the two groups, and a P -value < 0.05 was considered statistically significant.

RESULTS

Overview of the human cell lines stored in the laboratory

During the cell line backup process, we revived a total of 91 different cell samples based on the labels on the cell cryovials, which were later confirmed to represent 75 unique cell line strains [Supplementary Table 1]. According to the publications of our laboratory, a total of 80 human cell lines should have been documented, with nearly half of these cell lines and their research outcomes published. A total of 93.75% (75 out of 80) of the cell lines were successfully recovered during the cell line backup process. Although BEAS-2BS, FDIX, FL, HT-1080 and Raji had related outcomes published, we did not find them during the backup process and were considered lost [Table 1].

Most human cell lines in our laboratory were purchased from ATCC and the Cell Bank of Type Culture Collection of Chinese Academy of Sciences. Approximately 12.5% (10 out of 80) of the cell lines were kindly provided by the collaborators [Figure 1A]. Among them, about 18.75% (15/80) of the cell lines were derived from normal tissues, while 77.5% (62/80) were derived from tumor tissues. Lung cancer cell lines were the most numerous ($n = 16$), followed by liver cancer cell lines ($n = 8$), colon cancer cell lines ($n = 6$), and breast cancer cell lines ($n = 5$) [Figure 1B].

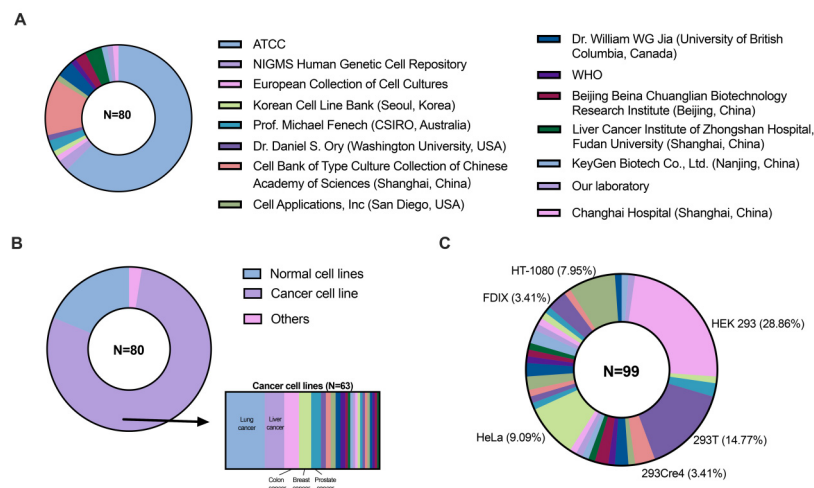
From 1990 to 2024, our laboratory published 60 articles involving 33 human cell lines. HEK 293 was the most frequently used human cell line, accounting for 28.86% of our publications. Publications using 293T, HeLa, and HT-1080 cells ranked second to fourth, accounting for 14.77%, 9.09%, and 7.95%, respectively. FDIX and HT-1080 tied for fifth place, each being used for four articles [Figure 1C].

STR profiling of cell lines stored in the lab over 34 years

A total of 23 STRs were tested, but only 21 autosomal loci were used for cell authorization. For the sex-determining genotypes, excluding the HFLS-OA and RKO cell lines, which lacked sex-related polymorphism data, about 24.72% (22/89) of the cell lines exhibited changes in genetic sex. According to the reference genomes, the cell lines should be derived from 46 male samples and 43 female samples. The results for the female cell lines showed 100% concordance with expectations. However, among the male-origin cell lines, only 24 showed the expected sex-determining profiles, while 20 samples completely lost male characteristics at these loci [Figure 2A].

Table 1. Overview of the human cell lines stored in the laboratory

Cell line	Publication	Cell line	Publication	Cell line	Publication	Cell line	Publication
ALK	/	NCI-H520	/	BPH-1	[17]	SW620 (CCL-227)	/
GM12878	/	95-D	/	HeLa	[18-25]	U-251MG	/
KELLY	/	HCC4006 (CRL-2871)	/	WI-38	/	HEP2	[26]
SNU-2535	/	NCI-H1975	/	M059J	/	CNE-1 (CNE)	/
NCI-H596	/	U-2 OS	[27,28]	M059K	/	A-431	/
GM13705	[29]	BT-483 (ATCC=HTB-121)	/	NCI-H460	/	SK-OV-3	/
GM12593	[30]	22Rv1	[17]	ECV-304	/	7721	[31]
HTB-178 (CRL-5928)	/	NCI-H226	/	MCF7	[32]	L-02	[31,33]
HTB19 (BT-20)	/	Ca Ski	/	MRC-5	/	SK-N-AS	/
GM24385	/	SW 1271 (CRL-2177)	/	RKO (CRL-2577)	/	CCL-187	/
HEK 293	[18,20,22,23,25,27,34-45]	SW48	/	Huh-7	[23]	5637	[46]
AD-293	[43]	A549	[23,47]	MHCC97-L	/	NCCIT	/
293GPG	[48,49]	SaOS-2	[27]	MHCC97-H	[31]	PC-3	[17]
293T	[26,27,33,37,50-58]	HFLS-OA	/	QGY-7703	[31,59]	DU145	[17]
293Cre4	[60-62]	CRL-7585	/	HCCLM3	/	SW480	/
NCI-H1703	/	ZR-75-30	/	Hep G2	[63,64]	BEAS-2BS	[23]
NCI-H1299	[28]	Jurkat, Clone E6-1	/	MDA-MB-231	/	FDIX	[65-67]
EJ	/	SK-HEP-1	/	KB	/	FL	[68]
NCI-H1563	/	K-562	[69]	LNCaP	[17]	HT-1080	[65,66,70-74]
NCI-H2170	/	PNT1B	[17]	SW1417 (CCL-238)	/	Raji	[21]

**Figure 1.** Summary of the human cell lines recorded in the laboratory from 1990 to 2024. (A) Origin of the cell lines; (B) Types of the cell lines; (C) Distribution of the cell lines used in published studies.

Among the 17 normal cell lines, the distribution of genotypes was as follows: homozygous genotypes accounted for 33.89% (121/357), heterozygous genotypes for 57.42% (205/357), and polyploid genotypes for 8.68% (31/357). In cancer cell lines, the proportions of homozygous, heterozygous, and polyploid genotypes

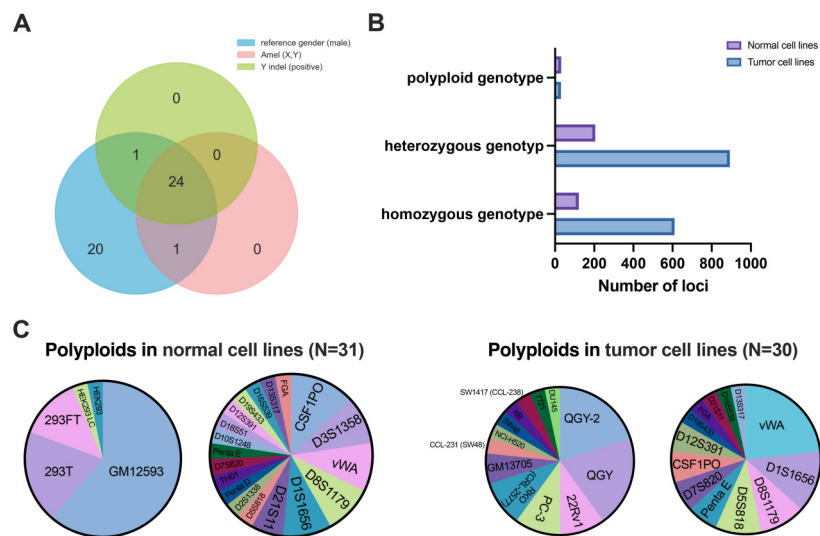


Figure 2. STR profiling of the human cell line samples revived in this study ($n = 91$). (A) Sex-related polymorphisms in the human cell lines; (B) Distribution of homozygous, heterozygous, and polyloid genotype loci in normal cell lines and cancer cell lines; (C) Distribution of the polyloid genotypes in normal cell lines and cancer cell lines.

were 39.79% (610/1533), 58.25% (893/1533), and 1.96% (30/1533), respectively [Figure 2B]. The polyloid genotypes were distributed across the 73 tumor cell lines, with the QGY cell lines having the most ($n = 6$). Interestingly, the polyloid genotype loci in the normal cell lines were predominantly found in GM12593 ($n = 19$, 61.29%). The vWA locus was the most frequent site for polyloid genotype in tumor cell types, whereas the CSF1PO locus was the most altered in normal cell lines. Additionally, alternations in D1S1656 and D8S1179 were commonly observed across different cell line types [Figure 2C].

Most cell line identities successfully verified using two algorithms

In this study, we successfully revived and typed 91 cell samples. Except for GM13705, GM12593, GM24385, HFLSOA, and MHCC97-L, all samples had registered STR profiling data in the ATCC or Cellosaurus databases. Two algorithms were employed for cell line authentication, and no statistically significant differences were observed between the scoring results [Figure 3A]. Approximately 94.51% (86/91) of the authentication results were identical using both algorithms, with discrepancies found in the conclusions for five samples [Figure 3B].

The inconsistent outcomes primarily fell into two scenarios: first, where the Masters algorithm classified the result as “Related”, while the Tanabe algorithm categorized it as “Unrelated” ($n = 1$, $\text{score}_{\text{Masters algorithm}} = 0.8236$; $\text{score}_{\text{Tanabe algorithm}} = 0.8889$); and second, where the Masters algorithm categorized the result as “Unrelated”, while the Tanabe algorithm resulted in “Mixed results” ($n = 4$, $\text{score}_{\text{Masters algorithm}} = 0.6129 \sim 0.7353$; $\text{score}_{\text{Tanabe algorithm}} = 0.6897 \sim 0.7692$). Since there were no significant score differences between the two algorithms, the scores for inconsistent conclusions were very close. We validated the outcomes of different scoring methods under the same criterion using cross-validation, revealing that identical authentication results were obtained regardless of the scoring method used [Figure 3C].

A total of eight samples failed to match their proclaimed cell line identities. To further investigate their origins, CLASTR 1.4.4 was applied to search for STR similarities in the Cellosaurus database. Results showed that seven out of eight samples had related matches. Although the relationship between Sample 293FT and HEK293 was questionable, it showed higher similarity to two HEK293 derivatives. Samples EJ,

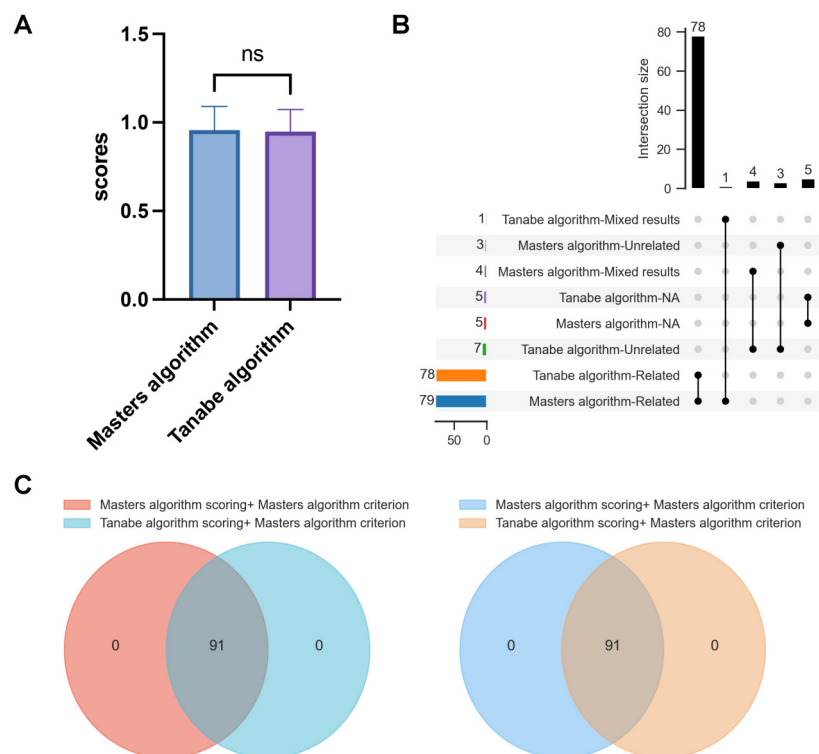


Figure 3. Comparative analysis of authentication results using different algorithms. (A) Scores obtained from the different algorithms; (B) Authentication outcomes produced by different algorithms; (C) Intersection of authentication results derived from different scoring methods under the same criterion. ns indicates no significant difference.

Sample QGY, Sample QGY-2, and Sample CNE successfully matched several human cell lines, but all the matched cell lines were considered contaminated. The typing results of Sample PC-3 differed from its reference genotype, resulting in “Mixed results” with both algorithms. However, the STR profiling showed a higher similarity to LNCaP C4-2B ($\text{score}_{\text{Tanabe algorithm}} = 90.91$, $\text{score}_{\text{Masters algorithm}} = 93.75$).

Fortunately, our laboratory possessed the parent cell line of LNCaP C4-2B, LNCaP. By comparing the STR profiling results of LNCaP with Sample PC-3, we found the scores using both algorithms were above 0.9. This further confirmed that Sample PC-3 was mislabeled and should be identified as the LNCaP cell line. Using the same method, we found that Sample WI-38 was also mislabeled, as its STR results were identical to those of 2BS [Table 2].

Comparative analysis of the multiple-backup cell lines and their derivatives

In this study, a total of 31 cell samples had multiple backups recorded in the laboratory. Due to its known issue, QGY-7703 was excluded from the analysis of this section. We compared the genotypes of cell samples within the same strains and found that only 12 cell lines had identical results. Among them, only eight samples showed exact matches to their reference genotypes [Supplementary Table 2]. Although the STR profiles of Sample H1299 vs Sample H1299-1, and Sample 252 vs. Sample CCL-231 differed from their references, the typing results within each pair were completely identical. For the remaining samples, one to five loci differed from the reference.

HEK293 and its derivatives had the highest number of backups in this study. Among the six backups, the typing results differed from the reference except for Sample 293y99 (the oldest recorded cell sample in our

Table 2. Cell line authentication results for eight “unrelated” samples

Label on cryovials	Reference cell line		Masters algorithm		Tanabe algorithm	
	Cell line	Reference	Score	Outcome	Score	Outcome
293FT	HEK 293	https://www.cellosaurus.org/CVCL_9804	0.8235	Related	0.8889	Unrelated
	HKb20 (HEK293 Hierarchy)	https://www.cellosaurus.org/CVCL_E065	0.9286	Related	0.9630	Related
	tsA201(HEK293 Hierarchy)	https://www.cellosaurus.org/CVCL_2737	0.9286	Related	0.9630	Related
EJ	EJ	https://www.cellosaurus.org/CVCL_7039	0.5000	Mixed results	0.4444	Mixed results
	JCA-1 (Problematic cell line: Shown to be a T24 derivative)	https://www.cellosaurus.org/CVCL_4015	1.0000	Related	1.0000	Related
	SLR20 (Problematic cell line: Shown to be a T24 derivative)	https://www.cellosaurus.org/CVCL_V606	1.0000	Related	1.0000	Related
TSU-Pr1 (Problematic cell line: Contaminated. Shown to be a T24 derivative)		https://www.cellosaurus.org/CVCL_4014	1.0000	Related	1.0000	Related
WI-38	WI-38	https://www.atcc.org/products/ccl-75	0.4194	Mixed results	0.4333	Mixed results
	2BS	https://www.cellosaurus.org/CVCL_BT02	1.0000	Related	1.0000	Related
QGY	QGY-7703 (Problematic cell line: Contaminated. Shown to be a HeLa derivative)	https://www.cellosaurus.org/CVCL_6715	0.6129	Unrelated	0.7308	Mixed results
	MGC-803 (Problematic cell line: Contaminated. Seems to be a hybrid of HeLa and probably of a gastric cancer cell from an Asian individual)	https://www.cellosaurus.org/CVCL_5334	0.9130	Related	0.9231	Related
QGY-2	QGY-7703 (Problematic cell line: Contaminated. Shown to be a HeLa derivative)	https://www.cellosaurus.org/CVCL_6715	0.6129	Unrelated	0.7308	Mixed results
	MGC-803 (Problematic cell line: Contaminated. Seems to be a hybrid of HeLa and probably of a gastric cancer cell from an Asian individual)	https://www.cellosaurus.org/CVCL_5334	0.9149	Related	0.9348	Related
293T	293T	https://www.cellosaurus.org/CVCL_0063	0.7353	Unrelated	0.7692	Mixed results
	STAR-A (HEK293T derivative)	https://www.cellosaurus.org/CVCL_AR74	0.8000	Related	0.8889	Unrelated
CNE	CNE-1 (CNE)	https://www.cellosaurus.org/CVCL_6888	0.7407	Unrelated	0.6897	Mixed results
	HeLa/Blvra KO (Problematic cell line: (Contaminated. Seems to be a hybrid of HeLa and of cell line of unknown origin.)	https://www.cellosaurus.org/CVCL_D519	1.0000	Related	1.0000	Related
PC-3	PC-3	https://www.atcc.org/products/crl-1435	0.2188	Mixed results	0.2500	Mixed results
	LNCaP C4-2B	https://www.cellosaurus.org/CVCL_4784	0.9375	Related	0.9091	Related

lab, a backup of HEK293 stored in 1999), Sample 293gpg, and Sample 293CRE. Nevertheless, they were classified as HEK293 derivatives by both cell line authentication algorithms.

Potential cross-contamination of the human cell lines with HeLa cells

As mentioned previously, several cell samples did not show a “related” outcome to any cell lines because they were registered as contaminated problematic cell lines. To investigate potential cell line contamination in our laboratory, we compared all 91 successfully revived cell samples with HeLa cells. Nine cell samples showed the outcomes of “Unrelated” or “Mixed results”. Sample Hela NFkB and Sample HELA were both HeLa derivatives, confirming their results as correct. The remaining seven cell samples were documented as contaminated by HeLa [Figure 4].

Analysis of STR alternations in tumor and normal cell lines

To investigate STR alternations in different cell lines, we analyzed the STR status in tumor and normal cell lines. Due to the difficulty in determining specific cellular attributes for contaminated cell lines, all problematic cells and those without reference STR results were excluded from this analysis.

An average of 1.43 ± 1.99 altered loci were detected in the normal cell lines, whereas the tumor cell lines contained approximately 0.75 ± 1.09 variant loci per sample. There was no significant statistical difference between the two groups ($P > 0.05$) [Figure 5A]. No STR alternations were observed in TH01, D12S391, D10S1048, and D6S1043 in any cell lines. D1S1656 and D3S1358 showed a low tendency for mutation, while vWA, D7S820, and D8S1179 were the most frequently altered STRs in this study [Figure 5B]. Alternations in normal cell lines were mainly confined to eight STRs, whereas the alternations in tumor cells were widely distributed. Regardless of cell line type, status of L was the most detected alternation, followed by the occurrence of Aadd and Anew [Figure 5C].

For different tumor cell lines, status of L was the most frequently observed type of STR alternation, constituting all alternations in lung cancer, liver and breast cancer cell lines. Anew was observed in colon and prostate cancer cell lines, while Aadd was detected only in prostate cancer cell lines. The distribution of STR alterations varied among different tumor types.

In the lung cancer cell lines, vWA and D8S1179 were the most common altered loci. In the liver cancer cell lines, the most common variated loci were D8S1179 and D13S317. In the colon cancer cell lines, vWA was the most variable STR locus, while in the breast cancer cell lines, it was TPOX. The distribution of STR variant loci was more balanced in the prostate cancer cell lines, with FGA, D21S11, and D13S317 each occurring twice, and Penta D and D16S539 each occurring once [Figure 6].

DISCUSSION

Long-term storage reveals mislabeling and cell line loss

From 1990 to 2024, our laboratory documented 80 human cell lines. Given the importance of cell line authentication for reliable and accurate research results^[75], we authenticated the cell lines during the backup process using the established STR identification methods.

To ensure the reliability and traceability of the cell lines, the liquid nitrogen tank was managed by only three authorized personnel over the past 34 years. However, instances where the same cell line strains were labeled and documented with different names still occurred. For example, the sample labeled 293y99 was later confirmed to represent “HEK293 stored in 1999”. Subsequent STR verification revealed many duplicates (as detailed later). Therefore, although 91 “different” cell line samples were selected for identification, they represented 75 unique cell line strains. The observation underscored the significant inconvenience improper labeling can cause in future experiments. Cell line naming has been a long-standing problem in life sciences^[76]. Despite restricted access and management by a limited number of

Label on cryovials	Outcome		Comments
	Masters algorithm	Tanabe algorithm	
Hela NFkB	Related	Related	Hela derivate
HELA	Related	Related	HeLa
QGY	Mixed results	Unrelated	Problematic cell line: Contaminated. Shown to be a HeLa derivative (PubMed=28807831).
KB	Related	Related	Problematic cell line: Contaminated. Shown to be a HeLa derivative (PubMed=4864103; PubMed=5641128; PubMed=1246601; PubMed=20143388).
QGY-2	mixed results	Unrelated	Problematic cell line: Contaminated. Shown to be a HeLa derivative (PubMed=28807831).
HEP2	related	related	Problematic cell line: Contaminated. Shown to be a HeLa derivative (PubMed=4864103; PubMed=5641128; PubMed=566722; PubMed=1246601; PubMed=20143388).
CNE	related	related	Problematic cell line: Contaminated. Seems to be a hybrid of HeLa and of cell line of unknown origin.
7721	related	mixed results	Problematic cell line: Contaminated. Shown to be a HeLa derivative (PubMed=26116706; PubMed=28807831).
L02	related	mixed results	Problematic cell line: Contaminated. Shown to be a HeLa derivative (PubMed=26116706).

■ "Unrelated" to HeLa
■ Other results

N=91

Figure 4. Description of cell samples with outcomes of "Unrelated" or "Mixed results" (n = 9).

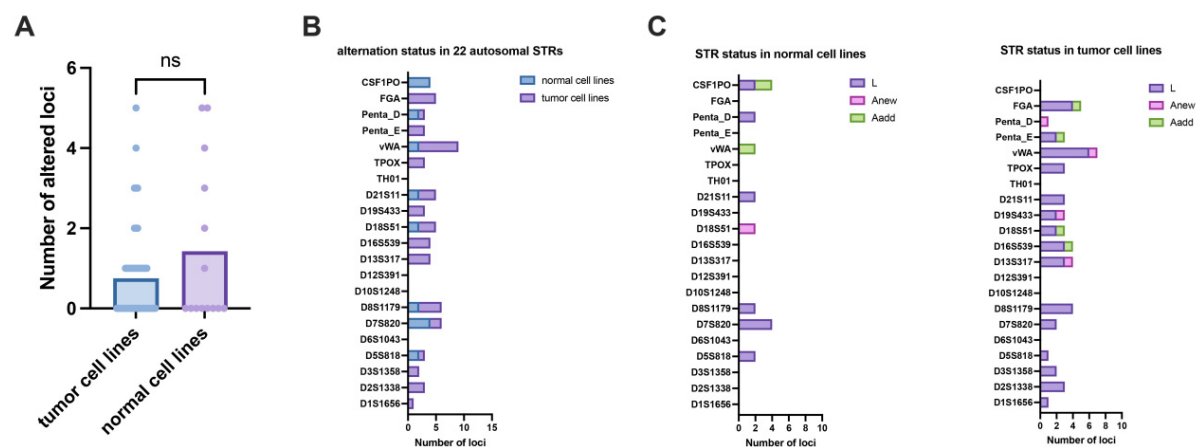


Figure 5. STR status in normal cell lines and tumor cell lines. (A) Number of altered loci in normal cell lines and tumor cell lines; (B) Number of altered STRs in two types of cell lines; (C) Genetic status at each STR locus in normal and tumor cell lines. STR: Short tandem repeat.

authorized personnel, inconsistent labeling practices can still lead to misunderstandings for future users over time. Additionally, five cell lines (BEAS-2BS, FDIX, FL, HT-1080, and Raji) were considered lost as they were not found in the liquid nitrogen tank. Among these, FDIX, a skin fibroblast from an adult hemophilia B patient (FDIX)^[65-67], likely failed to be preserved due to its status as a non-finite cell line, making it unsuitable for long-term use. The loss of the other four cell lines likely resulted from irregular handling and transfer practices that led to sample loss. Most cell lines in our laboratory were tumor cell lines, with lung cancer cell lines being the most common, followed by liver and colon cancer cell lines. However, HEK 293 and its derivatives were the most extensively used in our publications, accounting for a quarter of all cell lines.

STR-based sex identification shows male-specific instability

Amelogenin and Y-indel play crucial roles in sex determination in forensic identification^[77]. Similarly, Amelogenin plays an essential role in cell identification, although it is not included in the match score calculation. In this study, the expected number of cell lines from males and females was 46 and 43, respectively. However, the observed concordance was only 75.28% (67/89). All inconsistencies occurred in cell lines of male origin, primarily manifesting as a loss of heterozygosity in Amelogenin. Although the

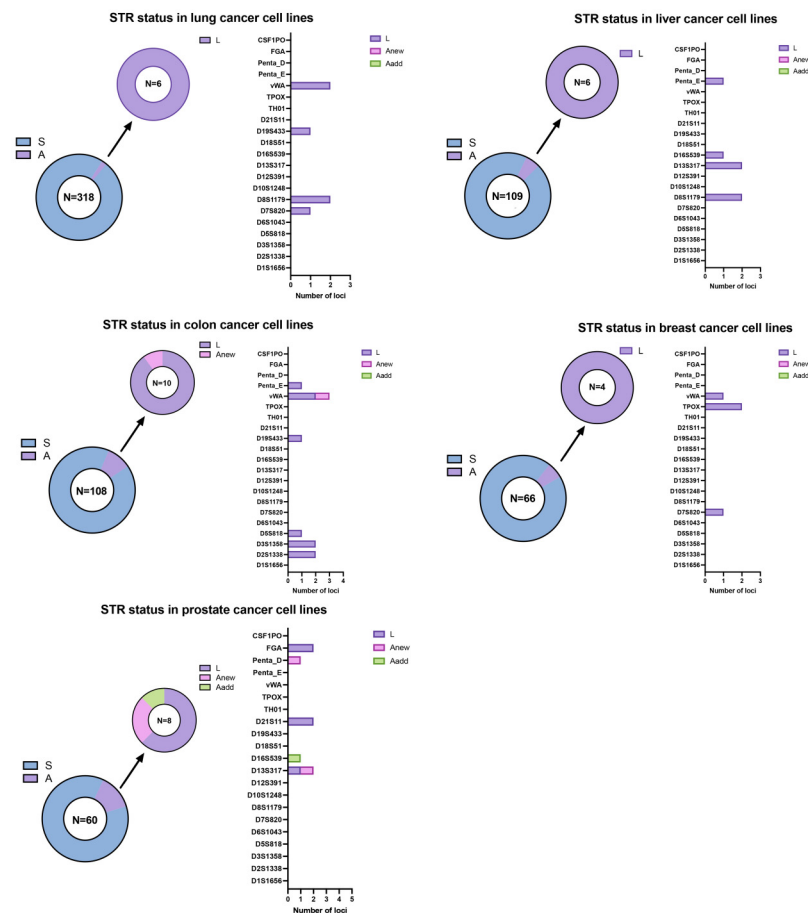


Figure 6. STR status in five different tumor cell lines. STR: Short tandem repeat.

Y-indel was included as an additional marker for sex determination, it did not improve the accuracy. In most cases, the loss of the Y allele at Amelogenin occurred simultaneously with the loss of the Y-indel.

Unexpected polyploidy and cross-contamination in normal cell lines

Normal human cells are diploid, resulting in one to two alleles at an STR locus. In this study, three normal cell lines (GM12593, 293T, and 293FT) exhibited a polyploid genotype. Notably, the GM12593 cell line showed an extremely high number of polyploid genotypes, with 19 out of 21 autosomal STRs (90.48%). GM12593 cell line is derived from a healthy female's B-lymphocytes and is expected to exhibit the characteristics of a diploid cell. Although genomic alternations can occur over successive passages, such extensive mutations are rare. Given that GM12593 and GM13705 originated from the same donor, we compared their genotypes and found that all alleles in GM13705 can be traced back to GM12593. This suggested that GM12593 was likely contaminated by GM13705, but the extract time of contamination was currently difficult to ascertain^[29,30].

Algorithm evaluation highlighted STR criteria over scoring methods

To investigate the origins of the cell samples, we analyzed them using the gold standard STR profiling technique for cell line authentication^[8]. The Tanabe algorithm is recommended for STR profiling, while the Masters algorithm is useful for identifying contaminating cell lines. In this study, we applied both methods and found that the vast majority of samples yielded consistent results, demonstrating the reliability of both

methods. Additionally, all samples identified as “related” by the Tanabe algorithm were also confirmed by the Masters algorithm. The outcomes were not true in the opposite scenario, suggesting that the Tanabe algorithm may be more stringent. Since both algorithms rely on the ratio of query alleles to reference alleles and showed no statistical difference in score values, this suggested that the choice of scoring method has a limited impact on the authentication outcomes. Cross-validation of scoring and criteria revealed complete consistency in results under the same criterion, suggesting that the criterion was more crucial than the scoring method. Sample 293FT (recorded as HEK293) was the only case that yielded conflicting results between the two methods. The Masters algorithm identified it as “related” to HEK293 (score: 0.8236), while the Tanabe algorithm classified it as “Unrelated” to HEK293 (score: 0.8889). HEK293, established in 1973, is one of the most widely used cell lines due to its robust growth in culture and high transfection efficiency^[78]. After prolonged passaging, cell lines may develop subtypes. Since the Tanabe algorithm score for this sample was close to the “related” threshold, we speculated that it may represent another subtype of HEK293. Using CLASTR 1.4.4 software to search for similar cell lines confirmed that Sample 293FT and Sample 293T showed high similarity to HEK derivatives, suggesting they may represent new HEK293 subtypes.

To enhance the accuracy and reliability of cell line identification, ATCC increased the number of STR markers used for authentication from 8 to 13 in 2019^[13]. The cell lines showing high similarity to Sample 293FT and Sample 293T were analyzed using only 8 STR loci, which may provide insufficient evidence for definitive identification. Moreover, the reference cell lines used for comparison were all problematic, suggesting that contamination likely occurred prior to their acquisition by our laboratory. Therefore, it was not difficult to understand why these eight cell lines initially failed to match any related reference cell lines.

STR profiling traced mislabeling events and confirmed minimal research impact

According to CLASTR search results, Sample PC-3 was related to LNCaP C4-2B. Although the comparison initially used only eight loci, our laboratory possessed a validated LNCaP cell line. We subsequently compared the profiles across 21 autosomal loci, confirming their close relationship (data not shown). This observation confirmed that PC-3 was a mislabeled cell line rather than one of unknown origin. Interestingly, both Sample WI-38 and Sample PC-3 were mislabeled, yet were classified as “Mixed results”, indicating that neither authentication method could clearly identify these problematic cell lines. In our laboratory, the PC-3 cell line was used in one publication. In that study, Tian *et al.*^[17] investigated the relationship between four microRNAs (miR-19b, miR-23b, miR-26a, and miR-92a) and prostate cell proliferation in vitro, using four different cell lines, including PC-3 and LNCaP. Therefore, we suspected that the mislabeling likely occurred during this period. However, the expression trend of miR-23b was downregulated in PC-3 but upregulated in LNCaP. This indicated that the expression trends of miR-23b in LNCaP and PC-3 were opposite, suggesting that LNCaP and PC-3 were still independent cell lines at that time. The labeling error for PC-3 may have occurred after the publication. Fortunately, since the PC-3 cell line was not used in any published studies afterward, it did not result in erroneous findings for the scientific community.

Another mislabeled sample was WI-38, which should be 2BS based on the comparison results. WI-38 is a normal diploid cell line derived from normal embryonic lung tissue^[79], and the cell line in our laboratory was sourced from the World Health Organization (WHO) as a gift to Wuxi Co., Ltd. for vaccine development. Although 2BS was also a fibroblast cell line derived from fetal lung tissue, it has already undergone STR alternations (https://www.cellosaurus.org/CVCL_BT02). Our observations were completely consistent with the subtype, as all loci exhibited diploid characteristics. Therefore, we asserted that the 2BS cell line stored in our laboratory represented a very early passage. These findings underscored the challenges in managing cell lines, which are susceptible to being mixed up over long periods of use. Even large professional organizations like the WHO can make mislabeling mistakes, highlighting the necessity of

authenticating cell lines before conducting experiments.

STR-based authentication confirmed derivatives but lacked subtype resolution

In the records of our laboratory, 31 cell samples had multiple backups or derived lines. We compared these samples with their corresponding reference typing profiles and found that the results varied from one another. Even the reference STRs have shown several different genotyping results, indicating that the cell lines underwent multiple variations during passaging. Three cell line strains (A549, K562, and NCI-H460) showed the same STR profiles as the reference, and two cell samples (NCI-H1299 and SW48) showed identical profiles within the backups, suggesting that their backup times were likely very close, resulting in no changes. Six cell samples were recorded as HEK293 or its derivative. Although their STR profiles varied, they were all identified as HEK293 derivatives. Only 293y99, the oldest recorded HEK293 backup in our laboratory, showed results that were completely consistent with the reference cell line. Sample 293gpg was a human 293-derived retroviral packaging cell line^[48,49], and Sample 293CRE was a Cre-loxP recombination cell line^[60-62]. Interestingly, their STR profiling remained the same with the parental cell line. Although STR polymorphisms are hotspots for homologous recombination, these genetic modifications did not alter the number of repeating units. Therefore, the current STR typing method could not differentiate the two cell lines, indicating that additional methods were needed to distinguish their differences. Castro *et al.*^[5] developed a 24-plex single nucleotide polymorphism profiling assay for human cell line authentication, and found it was a more robust approach for analyzing mismatch repair deficient cell lines. Lung *et al.*^[6] conducted metagenomic high-throughput sequencing for cell line authentication and contamination, providing an alternative approach for species identification and contamination detection. To date, high-throughput sequencing-based cell line authentication research has primarily focused on identifying more stable genetic markers rather than expanding intercellular differences. Therefore, future research should concentrate on developing new technologies and methods to identify and distinguish cell lines, ensuring the recognition of cell lines within the same hierarchy more accurately.

No evidence of HeLa contamination detected

Cross-contamination is considered the most common source of cell line contamination^[7], often occurring during the establishment of the cell line^[80]. Several studies^[2,81,82] suggested that HeLa cells played a central role in cross-contamination due to their rapid growth. In Chinese laboratories, more than 90% of cross-contaminated foreign cells were HeLa cells^[83]. Since our laboratory also maintained a HeLa cell line, the possibility of contamination with HeLa cells could not be ruled out. To investigate potential contamination, all revived cell samples were compared with HeLa cells. We found seven non-HeLa derivative samples were classified as “related” results. The QGY cell line is registered as a HeLa derivative; hence, the Masters algorithm results were more accurate. Therefore, we concluded that the Tanabe algorithm was more stringent in determination, and the Masters algorithm was better at exclusion. Scanning the cell line database trace revealed that all HeLa-induced contamination has been reported previously, suggesting early contaminations, possibly before these cells were introduced into our laboratory. Over the past 34 years, the cell line preservation status in our laboratory has been good, with no instances of HeLa cell contamination in our stock.

Characterizing STR instability and authentication markers

STR polymorphisms are hotspots for homologous recombination, which may display alternations in the number of repeating units^[84]. Additionally, a somatic mutation from microsatellite instability may complicate the STR typing^[11]. In this study, we investigated the STR profiles of the stocked cell lines at 23 loci. STR alternations were found in both normal diploid cells and tumor cells, with the number of alternations in normal cell lines not being less than in tumor cells. This may be attributed to the accumulation of alternations occurring during the passaging. TH01, D12S391, D10S1048, and D6S1043

showed relatively low alternation frequencies in cell lines, suggesting they may be ideal markers for authentication. In contrast, loci such as vWA, D7S820, and D8S1179 exhibited higher instability. Based on the principles of cell passage and STR replication, factors like high GC content or motif complexity could influence stability^[85-87]. However, this study found that both stable and unstable loci have comparable GC content and motif complexity (data not shown). We further examined the genomic context and found that most of them are located in intergenic regions or introns. As the mechanisms governing STR stability in these non-coding regions remain poorly understood, it was difficult to draw definitive conclusions at this stage. Further research was needed to clarify the biological significance of these observations. pLOH, L, Aadd, and Anew are the four types of alterations that may occur at STR loci^[12]. Since reference STR in public databases does not provide peak height information, pLOH-type alternation cannot be assessed. The status of L was the most frequent alternation, consistent with the observations reported by Rubocki *et al.*^[88]. Moreover, it was the only type of alternation detected in the lung cancer, liver and breast cancer cell lines. Although earlier genomic instability studies revealed a preference for STR alternations in different tumor types and predicted the potential biological significance behind these changes^[12], the preferred loci observed in this study differed from previous research, possibly due to the limited sample size. Compared to the number of cases reported previously ($n = 407$), the cell line number involved in this study was relatively small ($n = 65$). Therefore, the current findings should be considered exploratory, and larger cohorts will be required to establish tumor-specific STR alteration signatures. On the other hand, tumor cells experienced different environmental conditions between in vivo and in vitro cultures. The varied selection pressures may lead to different mutation outcomes ultimately. Although differences exist between tumor cell lines and tumor tissues, research using tumor cell lines remained representative as they retain specific characteristics of the primary tumor, such as gene mutations and phenotypic features. As noted by Freedman *et al.*^[76], DNA profiling provided limited insight into the complex molecular and phenotypic characteristics of a cell line. At the same time, cell lines serve as a foundational tool for further research, making it crucial to guarantee their authenticity before conducting experiments.

CONCLUSION

In this study, we used 23 forensic STR markers to authenticate 91 human cell line samples stored in our laboratory over the past 34 years. First, we confirmed that all revived cell lines were correctly authenticated and free from HeLa contamination, demonstrating the effectiveness of long-term preservation when coupled with periodic verification. Second, we observed STR alterations across both normal and tumor cell lines, indicating that genomic changes can accumulate with extended passages and should be considered in experimental interpretation. To support future research, we documented these genotypic variations and provided updated reference STR profiles for multiple widely used cell lines. Third, both the Master and Tanabe authentication algorithms were reliable, although they differed in stringency. However, our findings indicated that STR profiling alone was inadequate for distinguishing cell line derivatives within the same lineage. These findings underscore the importance of obtaining cell lines from reputable repositories and highlight the need for comprehensive multi-omics authentication frameworks to improve reproducibility in biomedical research.

DECLARATIONS

Author's contributions

Proposed the initial idea. Zhang S

Developed the methodology, supervised and administrated the project: Ji C, Zhang S

Accomplished the conceptualization and mainly wrote the manuscript, composed and drew figures, made a formal analysis: Chen A

Carried out experiments, gathered and curated data, investigated the resulting data and validated the results:

Chen A, Chen J, Xuan Y, Zhao J

Acquired the funding: Chen A, Ji C

All authors read and approved the final manuscript.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Financial support and sponsorship

This study was supported by grants from National Natural Science Foundation of China (82302124) and National Key Research and Development Project of China (2021YFC2301500).

Conflicts of interest

Zhang S, Guest Editor of the Special Issue “Molecular Innovation in Forensic Genetics” in *Journal of Translational Genetics and Genomics*, also serves as a Junior Editorial Board member of the journal. He was not involved in any steps of the editorial process for this manuscript, including reviewer selection, manuscript handling, or decision making. The other authors declared that there are no conflicts of interest.

Ethics approval and consent to participate

No human/animal experiments were performed; all cell lines were commercially sourced or donated.

Consent for publication

Not applicable.

Copyright

© The Author(s) 2025.

REFERENCES

1. Kaplan J, Hukku B. Cell line characterization and authentication. animal cell culture methods. *Methods Cell Biol.* 1998;57:203-16. DOI
2. Corral-Vázquez C, Aguilar-Quesada R, Catalina P, et al. Cell lines authentication and mycoplasma detection as minimum quality control of cell lines in biobanking. *Cell Tissue Bank.* 2017;18:271-80. DOI
3. Masters JR. Cell-line authentication: End the scandal of false cell lines. *Nature.* 2012;492:186. DOI PubMed
4. Nelson-Rees WA, Daniels DW, Flandermeyer RR. Cross-contamination of cells in culture. *Science.* 1981;212:446-52. DOI PubMed
5. Castro F, Dirks WG, Fährnich S, Hotz-Wagenblatt A, Pawlita M, Schmitt M. High-throughput SNP-based authentication of human cell lines. *Int J Cancer.* 2013;132:308-14. DOI PubMed PMC
6. Lung O, Candlish R, Nebroski M, Kruckiewicz P, Buchanan C, Moniwa M. High-throughput sequencing for species authentication and contamination detection of 63 cell lines. *Sci Rep.* 2021;11:21657. DOI PubMed PMC
7. Capes-Davis A, Reid YA, Kline MC, et al. Match criteria for human cell line authentication: where do we draw the line? *Int J Cancer.* 2013;132:2510-9. DOI
8. Almeida JL, Cole KD, Plant AL. Standards for Cell Line Authentication and Beyond. *PLoS Biol.* 2016;14:e1002476. DOI PubMed PMC
9. Zheng H, Tao R, Zhang J, et al. Development and validation of a novel SiFaSTRTM 23-plex system. *Electrophoresis.* 2019;40:2644-54. DOI
10. Huang Y, Wang M, Liu C, He G. Comprehensive landscape of non-CODIS STRs in global populations provides new insights into challenging DNA profiles. *Forensic Sci Int Genet.* 2024;70:103010. DOI
11. Chen A, Zhang S, Li J, Ji C, Chen J, Li C. Detecting genetic hypermutability of gastrointestinal tumor by using a forensic STR kit. *Front Med.* 2020;14:101-11. DOI
12. Chen A, Xiong L, Qu Y, et al. Opportunity of Next-Generation Sequencing-Based Short Tandem Repeat System for Tumor Source Identification. *Front Oncol.* 2022;12:800028. DOI PubMed PMC
13. Almeida JL, Korch CT. Authentication of human and mouse cell lines by short tandem repeat (STR) DNA genotype analysis. In:

- Markossian S, Grossman A, Baskir H, et al., editors. *Assay Guidance Manual* [Internet]. Bethesda (MD): Eli Lilly & Company and the National Center for Advancing Translational Sciences;2004. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK144066/> [Last accessed on 30 June 2025].
14. Masters JR, Thomson JA, Daly-Burns B, et al. Short tandem repeat profiling provides an international reference standard for human cell lines. *Proc Natl Acad Sci U S A.* 2001;98:8012-7. DOI PubMed PMC
 15. Korch CT, Hall EM, Dirks WG, et al. Human cell line authentication: standardization of short tandem repeat (STR) profiling. ASN-0002 Revised 2021, by American National Standards Institute - American Type Culture Collection Standards Development Organization. Available from: <https://www.atcc.org/products/135-ansi-str>. [Last accessed on 30 June 2025].
 16. CLASTR. CSSS, Interpretation of results. Available from: <https://www.cellosaurus.org/str-search/help.html>. [Last accessed on 20 Jun 2025].
 17. Tian L, Fang YX, Xue JL, Chen JZ. Four microRNAs promote prostate cell proliferation with regulation of PTEN and its downstream signals in vitro. *PLoS One.* 2013;8:e75885. DOI PubMed PMC
 18. Wang BY, Xu GL, Zhou CH, et al. ΦC31 integrase interacts with TTRAP and inhibits NFκB activation. *Mol Biol Rep.* 2010;37:2809-16. DOI
 19. Yang X, Wang F, Wang Y, Lu D, Qiu X, Xue J. Constitutive expression of human coagulating factor IX in HeLa cells by homologous recombination of the promoter. *Sci China C Life Sci.* 2001;44:18-24. DOI
 20. Lai L, Chen L, Wang J, et al. Skeletal muscle-specific expression of human blood coagulation factor IX rescues factor IX deficiency mouse by AAV-mediated gene transfer. *Sci China C Life Sci.* 1999;42:628-34. DOI
 21. Zhou C, Lu D, Zhu Q, Qiu X, Xue J. Effects of CIITA antisense RNA on the expression of HLA class II molecules. *Chin Sci Bull.* 2000;45:2068-71. DOI
 22. Chen L, Chen H, Zou B, et al. Preparation of rAAV/hFIX by HSV/AAV hybrid helper virus and evaluation of its safety. *Chin Sci Bull.* 2003;48:1369-1374. DOI
 23. Dong X, Tian W, Wang G, et al. Establishment of an AAV reverse infection-based array. *PLoS One.* 2010;5:e13479. DOI PubMed PMC
 24. Lu H, Chen L, Wang X, Lu D, Qiu X, Xue J. Preparation of a recombinant adeno-associated viral vector with a mutation of human factor IX in large scale and its expression in vitro and in vivo. *Chin Sci Bull.* 2001;46:1367-71. DOI
 25. Lu DR, Wang Q, Cong S, Wang ZM, Qiu XF, Xue JL. Detection and safety of wild-type adenovirus in HSV-tk/NAS system gene therapy for brain tumors. *Chin J Virol.* 1998;11-16. (in Chinese) Available from: https://kns.cnki.net/kcms2/article/abstract?v=UThtwiquHbf182koZgsr3yGgO4S9-Rl5duqPi3UBJxok1tja4GZNFsUrqu0uc8-fmk1-7EYknfP1JmJmx18wn0sdRh1TYtU64sH-tRGPrGfIXVE7UYJ_DehY4EpS_5-Bfkq5JyNmJ1Ifs-0dwViKScIRG33UI7ki0LwOqDyzxI=&uniplatform=NZKPT&language=CHS. [Last accessed on 30 June 2025].
 26. Wang R, Li KM, Zhou CH, Xue JL, Ji CN, Chen JZ. Cdc20 mediates D-box-dependent degradation of Sp100. *Biochem Biophys Res Commun.* 2011;415:702-6. DOI
 27. Zhou C, Shen Q, Xue J, Ji C, Chen J. Overexpression of TTRAP inhibits cell growth and induces apoptosis in osteosarcoma cells. *BMB Rep.* 2013;46:113-118. DOI PubMed PMC
 28. Wang J, Li W, Wang R, Xue J, Chen J. Adeno-associated virus Rep78 restricts adenovirus E1B55K-mediated p53 nuclear exportation. *Acta Biochim Biophys Sin.* 2013;45:135-40. DOI
 29. Ni J, Li Z, Xue JL, Wang X. Effects of -epigallocatechin gallate on expressions of hMLH1 and hMSH2. *Carcinog Teratog Mutagen.* 2013;25:276-79. DOI
 30. Ni J, Chen HY, Cao N, Zhou T, Xue JL, Wang X. Effects of folate with different oxidative states on genomic DNA methylation in human lymphoblast cell line. *Cancer Res Prev Trea.* 2014;41:256-9. Available from: https://kns.cnki.net/kcms2/article/abstract?v=UThtwiquHbf182koZgsr3yGgO4S9-Rl5duqPi3UBJxok1tja4GZNFsUrqu0uc8-fmk1-7EYknfP1JmJmx18wn0sdRh1TYtU64sH-tRGPrGfIXVE7UYJ_DehY4EpS_5-Bfkq5JyNmJ1Ifs-0dwViKScIRG33UI7ki0LwOqDyzxI=&uniplatform=NZKPT&language=CHS [Last accessed on 30 June 2025].
 31. Fang YX, Zhang XB, Wei W, et al. Development of chimeric gene regulators for cancer-specific gene therapy with both transcriptional and translational targeting. *Mol Biotechnol.* 2010;45:71-81. DOI
 32. Li Z, Qian Z, Chen F, Jiang S, Meng L, Chen J. Identification of Key lncRNA-mRNA Pairs and Functional lncRNAs in Breast Cancer by Integrative Analysis of TCGA Data. *Front Genet.* 2021;12:709514. DOI
 33. Zhu H, Chen X, Li F, Gong J, Xue J. Expression of human clotting factor IX mediated by recombinant lentiviral vector in cultured cells and hemophilia B mice. *Chin Sci Bull.* 2003;48:2196-200. DOI
 34. Lin Y, Li ZH, Wang JJ, Xu GL, Shen Q, Tian L, Xue JL, Chen JZ. Sp100 interacts with phage ΦC31 integrase to inhibit its recombination activity. *Acta Biochim Pol.* 2011;58:67-73. PubMed
 35. Zhou GY, Ji CN, Shen Q, Xue JL, Chen JZ. FHC interacts with bim and protects cell from apoptosis. *Prog Biochem Biophys.* 2011;38:859-65. Available from: https://kns.cnki.net/kcms2/article/abstract?v=UThtwiquHbdyQhXhTYrDqwl9EC4vM4CcXNiG7g_dJucj4hsvk_TNQlVD6FEgJGthmt4qzBdJGdi2Ee0ki_Vsg6KBQA31P2-KcQxM_iCJ7F5bzBZIntu-gT7n_Y7HrBKUM8H7gZv7rwlr2pLo48ulqQjg2Y45zZILoCL4ZmyTKnmv0Oz7uMoA=&uniplatform=NZKPT&language=CHS. [Last accessed on 30 June 2025].
 36. Yue YB, Xue YY, Tian L, Xue JL, Chen JZ, Jia W. Functional differentiation between Rep-mediated site-specific integration and transcriptional repression of the adeno-associated viral p5 promoter. *Hum Gene Ther.* 2010;21:728-38. DOI PubMed

37. Zhang JQ, Wang JJ, Li WJ, et al. Cellular protein TTRAP interacts with HIV-1 integrase to facilitate viral integration. *Biochem Biophys Res Commun.* 2009;387:256-60. DOI
38. Xu GL, Pan YK, Wang BY, et al. TTRAP is a novel PML nuclear bodies-associated protein. *Biochem Biophys Res Commun.* 2008;375:395-8. DOI
39. Chen JZ, Ji CN, Xu GL, et al. DAXX interacts with phage ΦC31 integrase and inhibits recombination. *Nucleic Acids Res.* 2006;34:6298-304. DOI
40. Feng D, Chen J, Yue Y, Zhu H, Xue J, Jia WW. A 16bp Rep binding element is sufficient for mediating Rep-dependent integration into AAVS1. *J Mol Biol.* 2006;358:38-45. DOI
41. Chen JZ, Huang SD, Ji CN, Pang RY, Xie Y, Xue JL. Identification, expression pattern, and subcellular location of human RIP isoforms. *DNA Cell Biol.* 2005;24:464-9. DOI
42. Chen JZ, Ji CN, Gu SH, et al. Over-expression of Bim alpha3, a novel isoform of human Bim, result in cell apoptosis. *Int J Biochem Cell Biol.* 2004;36:1554-61. DOI
43. Zhang J, Chen J, Liu L, et al. Different gene expression profiles of AD293 and HEK293 cell lines that show contrasting susceptibility to apoptosis induced by overexpression of Bim L. *Acta Biochim Pol.* 2006;53:525-30. DOI
44. Xue YY, Wang R, Yue YB, Xue JL, Chen JZ. Role and fate of SP100 protein in response to Rep-dependent nonviral integration system. *Appl Microbiol Biotechnol.* 2013;97:1141-7. DOI PubMed
45. Wang Q, Lu D, Xing Y, Xue J, Qiu X. Suicide effect on rat gliomas mediated by recombinant adenovirus thymidine kinase/acyclovir system. *Chin Med J.* 1998;111:483-7. PubMed
46. Zhang S, Wang Q, Li W, Chen J. MIR100HG regulates CALD1 gene expression by targeting miR-142-5p to affect the progression of bladder cancer cells in vitro, as revealed by transcriptome sequencing. *Front Mol Biosci.* 2021;8:793493. DOI
47. Wei W, Guo DW, Fang YX, Xue JL, Guo SR, Tian L. Biological evaluation on oncolytic herpes simplex virus with PEGylated folate modification. *Biotechnol Bull.* 2010;201-7. Available from: https://kns.cnki.net/kcms2/article/abstract?v=UThtwiquHbeox2S1Qvpynyi47ShVMOOIks5U4vaK4fXjoTybTHvCQ0jG1NsQARqfxJiseMCxqr2HnrDkeRsb-LYNKkV6dEXgsU6h0KAzDX63IO7aHHFS3NdbeY6KIReKUz1SY5516mRXx_T85uDFGMgELa2WICKK8hDU8c6px150_30N-Y7lpQ=&uniplatform=NZKPT&language=CHS. [Last accessed on 30 June 2025].
48. Wang H, Chen L, Wang X et al. Efficient transfer and expression of human clotting factor IX cDNA in neonatal hemophilia B mice mediated by VSV-G pseudotyped retrovi-rus. *Chin Sci Bull.* 2001;1534-8. Available from: https://kns.cnki.net/kcms2/article/abstract?v=UThtwiquHbctf_3kRHESrWU_41LerqhbRNaEkLYHv2AKYryJ1d3UO7dqERobgw2z2VeEegYDC7-QEZpRQaIHtczHmV2IAp_8BwT8jtmT4d7VHMBesQ6_TGmVP0KZ0Db3IAiLQ-Y4839zg8WMaLkQIZFPmUmpxWtML5CyIlsFzUGMyLrG27Org=&uniplatform=NZKPT&language=CHS. [Last accessed on 30 June 2025].
49. Wang F, Wang HW, Lu DR, Xue JL, Zhang X. Vesicular stomatitis virus G-protein retrovector mediated a herpes simplex virus thymidine kinase gene transduction and expression in the human retinal pigment epithelial cells. *Chin J Ophthalmol.* 2003;39:201-5. PubMed
50. Li K, Wang R, Zhou CH, Xue JL, Ji CN, Chen JZ. Cdh1 mediate D-box-dependent degradation of Sp100. *J Fudan Univ. (Nat Sci.)* 2012;51:315-21. Available from: https://kns.cnki.net/kcms2/article/abstract?v=UThtwiquHbflnEqKIsr6SXzUhyEZTVmq63kajmFNNIfgBV2dlAEbCrOHFgGTGKorFH3Bty5aPjZhnjnhdyMRhQPdoa80fZdXFKG1940IoB7h01QYkOeDC_ojt8IZ7LoD62UYZSXDDNoALop23Fr1buh_Gbnto69yKT8D5GIBMt5iaarYQKyQ=&uniplatform=NZKPT&language=CHS. [Last accessed on 30 June 2025].
51. Li WJ, Huang L, Zhang JQ, et al. The 156KELK159 tetrapeptide of HIV-1 integrase is critical for lentiviral gene integration. *Mol Biol Rep.* 2012;39:343-9. DOI
52. Huang L, Xu GL, Zhang JQ, et al. Daxx interacts with HIV-1 integrase and inhibits lentiviral gene expression. *Biochem Biophys Res Commun.* 2008;373:241-5. DOI
53. Chen H, Yao H, Huang L, Shen Q, Jia W, Xue J. Expression of human factor IX gene in murine plasma through lentiviral vector-infected haematopoietic stem cells. *Clin Exp Pharmacol Physiol.* 2006;33:1196-201. DOI
54. Yao H, Chen H, Huang L, Shen Q, Jia W, Xue J. Isolation and cultivation of murine hematopoietic stem cells and expression of hFIX mediated by recombinant lentiviral vectors in vitro. *Front Biol China.* 2006;1:259-62. DOI
55. Chen XG, Zhu HZ, Gong JL, Li F, Xue JL. Efficient delivery of human clotting factor IX after injection of lentiviral vectors in utero. *Acta Pharmacol Sin.* 2004;25:789-93. Available from: https://kns.cnki.net/kcms2/article/abstract?v=UThtwiquHbdNfbJg6UjHCpmqw4Yzcg-Ny2wcqBTxbFmi6ZlWSJTFSQtFKFa-b8xLI11xL70E_-SHJS5z4Xjh1XotVxUzis2lbDpBMXVjw9mfSmgPsiJqGmA1PtKWDzT01WudquZzpn9tjUz-KLzS6xpcVFlosjbrpVTk4V0KJENCxf77oKefA=&uniplatform=NZKPT&language=CHS. [Last accessed on 30 June 2025].
56. Chen XG, Wang N, Yao JH, et al. Gene delivery to mice spermatogenic stem cells by ffectene. *Chin Med J.* 2004;117:630-2. Available from: https://kns.cnki.net/kcms2/article/abstract?v=UThtwiquHbfpD3gcsppvIZK2xu2Tag3j0ZPJGqb-YWMso3G7VS2F0km21dXUQgUZhVzpqjKY53rCtoTafqQolmrZlnrMJ5x2WARVvaAidKd-3WLJgFp9zjXvEXSL3eVZb-Lssjq_lqWmTNCM5EX1wE05t7U4WDo2h1WRMPKekpwhK-f65mCxx=&uniplatform=NZKPT&language=CHS. [Last accessed on 30 June 2025].
57. Li Z, Wu S, Wang J, et al. Evaluation of the interactions of HIV-1 integrase with small ubiquitin-like modifiers and their conjugation enzyme Ubc9. *Int J Mol Med.* 2012;30:1053-60. DOI

58. Wang JJ, Wang R, Li W, Xue JL, Chen JZ. Daxx interacts with phage ΦBT1 integrase and inhibits its recombination. *Prog Biochem Biophys.* 2013;40:37-42. [PubMed](#)
59. Fang Y, Xue JL, Shen Q, Chen J, Tian L. MicroRNA-7 inhibits tumor growth and metastasis by targeting the phosphoinositide 3-kinase/Akt pathway in hepatocellular carcinoma. *Hepatology.* 2012;55:1852-62. [DOI](#) [PubMed](#)
60. Chen L, Chen H, Li G, Lu D, Xue J. Novel rAAV production system with low contamination of helper virus. *Chin Sci Bull.* 2003;48:472-5. [DOI](#)
61. Gao X, Ye C, Shi D, Chen L, Qiu X, Xue J. A study of gene transfer and expression of human clotting factor IX in hemophilia B mice mediated by mini-adenoviral vector. *Sci China C Life Sci.* 2003;46:631-40. [DOI](#)
62. Gao XB, Shi D, Wang F, Lu DR, Qiu XF, Xue JL. In vitro expression of human clotting factor IX minigene mediated by mini adenoviral vector. *Chin J Virol.* 2000;294-8. Available from: <https://europepmc.org/article/cba/341696>. [Last accessed on 30 June 2025]
63. Zhao Y, Li M, Miao N, et al. Use of miRNA sequencing to reveal Hub miRNAs and the effect of miR-582-3p/SMAD2 in the progression of hepatocellular carcinoma. *Front Genet.* 2022;13:819553. [DOI](#) [PubMed](#) [PMC](#)
64. Zhu P, Pei Y, Yu J, et al. High-throughput sequencing approach for the identification of lncRNA biomarkers in hepatocellular carcinoma and revealing the effect of ZFAS1/miR-150-5p on hepatocellular carcinoma progression. *PeerJ.* 2023;11:e14891. [DOI](#)
65. Hsueh JL. Clinical protocol of human gene transfer for hemophilia B. *Hum Gene Ther.* 1992;3:543-52. [DOI](#) [PubMed](#)
66. Dai YF, Qiu XF, Xue JL, Liu ZD. High efficient transfer and expression of human clotting factor IX cDNA in cultured human primary skin fibroblasts from hemophilia B patient by retroviral vectors. *Sci China B.* 1992;35:183-93. Available from: https://kns.cnki.net/kcms2/article/abstract?v=UThwtwiquHbfc49TnNHmygfQ-0kMSwAw-Vm817rPzp9EzH0k8-VE9RCt9CgWThZTiTsFNXn4XqcrF5GF04PAE0ARlkxPRu_WOb50JXD101vtK3FZ9N4r2S7JmGiHHXL1CdJWzgArVhdu4Jd4FPs_X6jyFmBGn1QaK-G1ldk9KPlr0SNpFu_G2sQ==&uniplatform=NZKPT&language=CHS. [Last accessed on 30 June 2025].
67. Zhou JM, Dai YF, Qiu XF, Hou GY, Akira Y, Xue JL. Expression of human factor IX cDNA in mice by implants of genetically modified skin fibroblasts from a hemophilia B patient. *Sci China B.* 1993;36:1082-92. [PubMed](#)
68. Wang HW, Zheng B, Xue JL, Xiu XF. Expression of human clotting factor IX with EBV shuttle vector. *Chin Sci Bull.* 1995;1829-33. Available from: https://kns.cnki.net/kcms2/article/abstract?v=UThwtwiquHbfcNi9W5FuF6GkEv1S8C3wWV7IA6mRubAL67_i1eub43knocA2yJfoD6KJtGOQctxuOIRuTPNwUuJ8MogmBuEKyi145wzEQgUN1PmGpNqp_L1-zGGzbjcyXjOyx5038a5PxilBAU6PQexb7bcVSouijV9KHNORWLvH5Iy5kKbdWA_g==&uniplatform=NZKPT&language=CHS. [Last accessed on 30 June 2025].
69. Wang JS, Xia XM, Chen ZX, Lu DR, Xue JL, Ruan CQ. Construction of a polycistronic retroviral vector for multidrug resistance genes and transfection of human umbilical cord blood CD34+ cells. *Natl Med J China.* 2000;77-79. (in Chinese) Available from: https://kns.cnki.net/kcms2/article/abstract?v=UThwtwiquHbfcMMmCA3PFNLmuvfhcFD0vcblvBa_sTbbSzR-ehsGVVWPxKfJxdW2fTPl-F_FbmixyhoVn_CIrRKYt4wnSrMc1ad16MOF-oY2NZapMfY-9VUnOCjlmULfVpNxYzD5rpf2tOuPvkJoHsJTyMBHH10D4bAlIRVx80C_TOVmQFLQ==&uniplatform=NZKPT&language=CHS. [Last accessed on 30 June 2025].
70. Lu DR, Qiu XY, Zheng B, Wang HW, Qiu XF, Xue JL. Regulation of in vitro expression of human coagulation factor IX cDNA. *Prog Nat Sci.* 1998;43-48. Available from: https://kns.cnki.net/kcms2/article/abstract?v=UThwtwiquHbfc0h0njtXqb4MdnUMdbaeYKBmW07gZ0398eL9P6t5a_Ee710ndFwpH0WA847E1hYtOThwInZWfjxWSaHU0b5190nb1e5Y7gFHg_iHZStgealAgEhUUGx5Eg9fpQsXfry7tPtbgCPkNDlqBby3IT-5eocC26RCrmupBrRAuKCsQ==&uniplatform=NZKPT&language=CHS. [Last accessed on 30 June 2025].
71. Lu DR, Zhou JM, Zheng B, et al. Stage I clinical trial of gene therapy for hemophilia B. *Sci China B.* 1993;36:1342-51. [PubMed](#)
72. Wang P, Dai YP, Qiu XF, Xue JL. Efficient expression of human factor IX cDNA in human cells under the control of the human cytomegalovirus promoter. *Chin Sci Bull.* 1991;36:1021-3. (in Chinese) Available from: https://kns.cnki.net/kcms2/article/abstract?v=UThwtwiquHbfc4eBgXnVpQMKWKZoQT4DLbRwJp_5k91DOY1ZSjU3fp0VVqzMTY1bF7amGWgBHj-Uf2MWPxnrWvtduKnaRCHfJ5iPEoaD01eSE9bSBm36bCWoj_SFUVuRXMHmPaot0W7-Qasol9135yqHeNbH86IP1Aj4Lj5nDinibd3pxjvvdhA==&uniplatform=NZKPT&language=CHS. [Last accessed on 30 June 2025].
73. Zhou JM, Hou GY, Hu YP, Qiu XF, Xue JL. Construction of recombinant plasmid containing human factor IX cDNA and its expression by transfected cells in mice. *Chin Sci Bull.* 1993;956-60. Available from: https://kns.cnki.net/kcms2/article/abstract?v=UThwtwiquHbfcE5zJ5jATPklJl-HWtUPg0jWykVU0Ig5AqMjWC6UkfMyxQX8tQRIJm7Nnf70H42OiQCzySbiBfU6YsWGuKEknRC4rQhd2kj2-tsylhDLmyiZQyU2Luo--ONpNheubkXysptvey1THO_zwJryE_VjEmZNXGpSvvcOU2Gv8buYt4_MA==&uniplatform=NZKPT&language=CHS. [Last accessed on 30 June 2025].
74. Hu Y, Qiu X, Xue J, Liu Z. Regulation of human clotting factor IX cDNA expression in transgenic mice. *Sci China B.* 1995;38:825-33. [PubMed](#)
75. Souren NY, Fusenig NE, Heck S, et al. Cell line authentication: a necessity for reproducible biomedical research. *EMBO J.* 2022;41:e111307. [DOI](#)
76. Freedman LP, Gibson MC, Ethier SP, Soule HR, Neve RM, Reid YA. Reproducibility: changing the policies and culture of cell line authentication. *Nat Methods.* 2015;12:493-7. [DOI](#) [PubMed](#)
77. Takayama T. A novel mutation at the AMEL primer binding region on the Y chromosome in AMELY negative male. *Int J Legal Med.*

- 2022;136:519-26. [DOI](#)
78. and Let Live: The remarkable story of HEK293 Cells. *Hum Gene Ther.* 2020;31:485-7. [PubMed](#)
79. Hayflick L. The limited in vitro lifetime of human diploid cell strains. *Exp Cell Res.* 1965;37:614-36. [DOI](#) [PubMed](#)
80. 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;83:555-63. [DOI](#) [PubMed](#)
81. Lucey BP, Nelson-Rees WA, Hutchins GM. Henrietta lacks, hela cells, and cell culture contamination. *Arch Pathol Lab Med.* 2009;133:1463-7. [DOI](#) [PubMed](#)
82. Lin LC, Elkashty O, Ramamoorthi M, et al. Cross-contamination of the human salivary gland HSG cell line with HeLa cells: a STR analysis study. *Oral Dis.* 2018;24:1477-83. [DOI](#)
83. Huang Y, Liu Y, Zheng C, Shen C. Investigation of Cross-contamination and misidentification of 278 widely used tumor cell lines. *PLoS One.* 2017;12:e0170384. [DOI](#) [PubMed](#) [PMC](#)
84. Wahls WP, Wallace LJ, Moore PD. Hypervariable minisatellite DNA is a hotspot for homologous recombination in human cells. *Cell.* 1990;60:95-103. [DOI](#)
85. Liu Y, Li J, Wu Q. Short tandem repeats of human genome are intrinsically unstable in cultured cells in vivo. *Gene.* 2023;877:147539. [DOI](#)
86. Press MO, Carlson KD, Queitsch C. The overdue promise of short tandem repeat variation for heritability. *Trends Genet.* 2014;30:504-12. [DOI](#) [PubMed](#) [PMC](#)
87. Gymrek M, Willems T, Guilmatre A, et al. Abundant contribution of short tandem repeats to gene expression variation in humans. *Nat Genet.* 2016;48:22-9. [DOI](#)
88. Rubocki R, Duffy K, Shepard K, Mccue B, Shepherd S, Wisecarver J. Loss of heterozygosity detected in a short tandem repeat (STR) locus commonly used for human DNA identification. *J Forensic Sci.* 2000;45:1087-9. [DOI](#)