Institutional Best Laboratory Practices for Cell Line and Tissue Sample Authentication to Ensure Valid, Reproducible and Robust Research

Christopher T Korch[1] and Rebecca E Schweppe[2]

[1]Division of Medical Oncology
[2]Division of Endocrinology, Metabolism, and Diabetes
School of Medicine
University of Colorado - Anschutz Medical Campus
12801 East 17th Avenue, RC-1 South
Aurora, Colorado 80045, USA

Table of Contents

Summary

The biomedical research community has a major problem with using imposter cell lines and incorrect tissue samples. The world-wide average incidence of using misidentified cell lines is 22%. This problem is partially addressed by granting agencies and journals requiring some type of cell line authentication. But to tackle this problem at its source, specifically in the research laboratories, we have designed a set of Best Laboratory Practices (BLPs) for the authentication of cell lines and tissue samples by four methods. Briefly, the four methods for authentication of cell lines and tissue samples include:

1) verification of cell line identity via the Cellosaurus and ICLAC websites,
2) genotyping by STR genotyping,
3) identification of any non-human species present in human cultures, and
4) mycoplasma testing of all cell cultures.

To utilize these BLPs, we have developed detailed recommendations of how they can be implemented in the laboratories of members of a research community. These include recommendations for evaluating cell cultures and tissue samples prior to their use, developing appropriate laboratory protocols to minimize the risk of working with misidentified cell samples, and confirming the authenticity of the cell lines and tissues used in the experiments before submitting grants, manuscripts, and sharing samples with fellow researchers. A supplement to this document includes additional information and guidelines. The goal of these Best Laboratory Practices is to protect the integrity, validity, robustness, and reproducibility of biomedical research produced by members of a biomedical research organization.

Introduction

Begley [14] reported that a majority (50%-90%) of published biomedical research is not reproducible and Freedman et al. [31] estimated that up to US $28 billion is spent each year on irreproducible biomedical research. A major contributor to the irreproducibility of biomedical research is the infrequent authentication of established cell lines and tissue samples and the frequent use of misidentified cell lines [32, 33, 39, 46, 47, 63, 64, 85]. Studies of data from labs around the world show the incidence of misidentified cell lines ranges from 10% to 100%, with an average of 22% or 2 of 9 cell lines being incorrect [46]. Many cell lines that were shown by Stanley Gartler to be misidentified as early as 1967 [36] are still being used under their false identities [38, 39, 47] to model the incorrectly identified tissues. Vaughan et al. reported that out of 574 articles between 2000 and 2014 using the cell line KB established in 1955, only 57 described the cell line correctly as actually being an imposter derived from HeLa cells [84]. Korch and Capes-Davis reported that the HeLa-derived cell lines Hep-2 and Intestine 407 (originally claimed to be from liver and normal intestinal cells) were used in 8,497 and 1,397 articles, respectively, under their false identities and this usage continues to this day [47]. Horbach and Halffman [39] found that 251 of over 500 known false cell lines with no known authentic samples were used in 32,755 publications and these articles were cited conservatively one-half million times in the scientific literature. Additionally, 7% or more of human genomic

data are sequences derived from mycoplasma, a bacterial genus that frequently contaminates cell line cultures [49] and which can alter metabolic responses and impact the analysis of contaminated cell lines [1, 12, 18, 27-29].

In 2018, the Biocompare company [16, 67] produced an excellent documentary that is worth-while viewing, in which we and eleven other prominent scientists, who have been combating this cell line authenticity problem, discussed many aspects of this issue. In 2022, the *International Journal of Cancer* (IJC) reported that 22.9% of submitted manuscripts which used human cell lines that had different levels of problems with the identifies of the cell lines used for research (i.e., 5.4% of the manuscripts required minor edits of incomplete cell line identity information; 9.8% had moderate problem due to using, e.g., a false cell line which had to be removed because it was misidentified but which was not a major basis of the study; and 7.6% which used two or more false cell lines and/or falsified data and documentation) [78]. Of all the manuscripts, 4.7% were rejected because of "severe, unaddressed cell line problems," and yet were found subsequently to have been published in other journals without addressing the cell line issues.

Clearly, this major problem continues to contaminate the biomedical literature with unreliable research [38-40, 44-47, 50, 78, 84, 87, 88]. Many have proposed that journals and funding agencies, as gate keepers of the biomedical literature, should require authentication of cell lines for acceptance of manuscripts and grant applications [35, 51, 57, 58, 62, 85]. Some journals and some funders, such as the NIH, have begun to require some level of sample authentication, but most requirements imposed by these gate keepers are not very stringent [78] allowing for continued publication of research based on false cell lines [31, 47, 84].

One approach for dealing with published reports based on using misidentified cell lines is that the articles be retracted. This appears to be impractical and an avoided option. Thousands of publications have used false cell lines [38, 39, 47, 84], but only about 100 articles have been retracted or issued notices of concern or corrigenda (see Appendices 2 and 3 in [46] and search Retraction Watch http://retractiondatabase.org/RetractionSearch.aspx for articles retracted because of contaminated cell line and tissue samples). Retractions of research articles are costly, as illustrated by the work of Linger et al. [52-55], infrequent [46], and can be detrimental for the reputations of researchers and their institutions [79]. Stern et al. estimated the average financial cost of retracting an article is between US$300,000 and US$400,000, not to mention its costs in decreased reputation of the research group and institution and their ability to obtain grants in the future [79]. Potentially, many billions of dollars have been used to support research based on false cell lines [31, 47].

The cross-contamination or misidentification of cell cultures, cell lines, stem cells, and xenografts (both of tissue samples and cell suspensions) can have seriously compromised research using these cells to model diseases, biological phenomena, drug testing in pre-clinical trials, and development and improvement of therapies for metabolic diseases and cancers of specific organs. This ultimately results in the waste of limited research resources [39, 46, 47, 75, 78]. Clearly, researchers need to beware of basing their research on publications using false/imposter and misidentified cell lines and citing publications based on such cell lines.

Six valuable resources have been developed to address this enormous problem: The International Cell Line Authentication Committee's website (ICLAC) [41], Cellosaurus – a Cell Line Knowledge Resource website [10]; and four guidance manuals for the authentication of human and mouse cell lines by analysis of STR genotypes, namely two ANSI-ATCC standards [6, 48] and two Assay Guidance manuals available on the NCBI website [4, 72]. To address to this alarming problem, the NIH issued three notices that "key biological and/or chemical resources" need to be regularly authenticated (NOT-OD-15-103, NOT-OD-16-011, NOT-OD-16-012, effective January 25, 2016). Several granting agencies now require some degree of authentication of key biologic reagents, including cell lines and tissue samples. Increasingly, a number of journals are starting to require that authors provide some kind of evidence that the cell lines that they used were authenticated prior to manuscript acceptance. The IJC has an exemplary cell line policy when considering manuscripts for publication (see [78] and Supplementary Information).

It is imperative that members of the biomedical research community alter their research practices in a manner that recognizes the need for these BLPs so that this important problem can be addressed. Several publications describe "best practices" for the culturing, handling, and authentication of cell lines and tissue samples [5, 6, 13, 34, 37, 60, 71-73, 75, 78]. Building on the *Cell Line Authentication Policy of the MD Anderson Cancer Center* [83] (which is the only publicly available cell line policy we could find), the information on the websites of the International Cell Line Authentication Committee (ICLAC) and Cellosaurus, and our own experience (in both our own research and running the University of Colorado Cancer Center DNA Sequencing & Analysis core facility), we propose the following cell line and tissue sample authentication Best Laboratory Practices (BLPs) for biomedical research institutions.

Four Best Practices for the Authentication Cell Line and Tissue Samples used in Biomedical Research

The Best Laboratory Practices for Cell Line Authentication recommend that all cell cultures and tissue samples used by any member of a research team should be validated by all four of the following authentication methods. These methods should be applied at the following stages of a project: a) at the beginning of a project before using samples for experiments, b) regularly during handling of samples, c) whenever a novel phenotype is noted, d) after any phenotypic selection, e) at the end of a project, f) before submitting grants, and manuscripts, and g) prior to sharing samples with fellow researchers. In addition, Microsatellite Stable (MSS) cell cultures should not be passaged more than 10-20 times and not used when unusual phenotypes are noticed. In such cases, the experiments should be started anew with a fresh aliquot of the authenticated batch of cells. MSI-Unstable cell cultures should not be passaged more than 5-10 times because of their inherently rapid genetic drift.

Method 1 - Confirming that the cell lines proposed to be used for research are known to be authentic and not misidentified/cross-contaminated lines by verifying their identity

on the Cellosaurus (https://www.cellosaurus.org/) and ICLAC (https://iclac.org/) websites;

Method 2 - Genotyping by STR / DNA profile analysis of human cell samples (i.e., cell lines, patient-derived explants (PDX), xenograft tissue) and by STR (mouse, dog, rat) or SNP genotyping of other non-human cell lines to authenticate their genetic identity;

Method 3 - Identification of any non-human species present in human culture samples by species-specific PCR analyses (e.g. [24]) or comparable methods to ensure the samples are authentic and not contaminated with cells from inappropriate animal species [7, 24, 40]; and

Method 4 - Analysis of all cell culture samples to ensure they are not contaminated with mycoplasma or other microbes, which can affect cell line phenotypes.

All laboratory research team members working with cell lines, xenografts, and tissue samples are encouraged to implement these four methods in their laboratories to ensure that they publish valid and reproducible research. The laboratories may need to identify a service facility that can perform at least Method 2, while the other three methods can be performed in most laboratories. The researchers will need learn how to interpret the data from the four methods as is described in the ANSI-ATCC ASN-00002-2022 Standard [48], the NCBI Assay Guidance Manual [4], and the Match Criteria explanation on the ICLAC website (https://iclac.org/resources/match-criteria-worksheet/).

Implementation of Guidelines and Methods to Ensure the Authenticity of Cell Lines and Tissue Samples and Adherence to these Best Laboratory Practices

We have developed the following guidelines and methods to encourage implementation and adherence with the BLPs based on the above four authentication methods, These suggestions are derived, in part, from those presented in Appendix G of the 2021 - 2022 revision of the ANSI-ATCC Standard ASN-0002 describing Human Cell Line Authentication by STR DNA profiling [48] and from other resources. These steps consist of (I) determining which cell lines and tissues are appropriate for the planned experiments, (II) implementing appropriate laboratory protocols to minimize the risk of working with misidentified cell samples, (III) being aware of any warning signs that the cell line and tissue sample are no longer correct, (IV) confirming the authenticity of the cell lines and tissues used in the experiments before submitting grants, manuscripts, and sharing samples with fellow researchers, and (V) considering patient confidentiality when publishing human STR data.

Disclaimer

Mention of specific commercial equipment, instruments, or materials (or suppliers, software, etc.) in this document is only to foster understanding. It does not imply that we recommend or endorse any of the materials or equipment identified or imply that these are necessarily the best available for the purpose.

I. Eight Steps for Evaluating Cell Lines and Tissue Samples Prior to their Use in Research

1. All laboratories should establish mandatory training in the Best Laboratory Practices and safe tissue culturing techniques referenced above to prevent contaminating the established cell lines with cells from foreign cell lines or microbes including mycoplasma. These precautions are described on the ICLAC website and by Capes-Davis and Freshney [19], Freshney [34], Geraghty et al. [37], Korch et al. [48], the 2018 OECD Guidance Document on Good In Vitro Method Practices (GIVIMP) and others [13, 60, 71-73].

   To implement these techniques, each laboratory should develop a set of documents detailing these standard best laboratory practices for the consistent and required uniform training of all laboratory personnel as described in the ANSI-ATCC Standard ASN-0002-2022 [48]. The goal is to minimize inconsistent and irreproducible results arising through variable verbal transmission of laboratory practices. See the attached Supplementary Information document for additional details.

   Laboratories are encouraged to (a) provide new lab personnel with detailed mandatory training on the handling of cell lines and tissue samples and (b) offer all lab personnel annual refresher training on these techniques to reduce the likelihood of sample mix-ups, etc.

2. Before obtaining cell lines for a research project, the names of these cell lines should be checked for what is known about them on the websites of ICLAC [21], Cellosaurus [10], the Cell Line Project at the COSMIC Catalogue Of Somatic Mutations In Cancer [82], and the Biosample database of the NCBI (https://www.ncbi.nlm.nih.gov/biosample, which has STR electropherographic data for some of the included cell line data).

   - Are any of the cell lines known to be misidentified?
   - Is there a reference STR or SNP genotype or other information about the cell lines available for comparison?

   Compile the collected information in a searchable database for future use, including in a spreadsheet (e.g., Excel or .CSV file available from the CLASTR-Cellosaurus website) which would allow searching for potentially matching STR profiles.

3. It is best that established cell lines are obtained only from sources that can demonstrate the authenticity of the material. Commercial sources, such as the ATCC, RIKEN, JCRB, or DSMZ, do this routinely. Most colleagues are not likely or able to provide authenticated cell lines, which were not developed in their labs.

4. Prior to their use, all cell cultures brought into a lab should be incubated in a separate Quarantine Incubator until the cultures are shown to be both free of microbial contamination and not cross-contaminated with other cell lines, especially rapid growers

(for example as, but not limited to, HeLa, M14, MCF-7). An exception can be if the cell line was received directly from a reliable repository, such as one of those described above (item 3), with documentation confirming its identity and purity.

5. When establishing new cell lines or xenografts, it is critical to determine the STR genotype of the original patient tissue sample to serve as a reference genotype for all future work. Preserve original tissue samples and, if possible, a normal sample of tissue or blood for any other future molecular characterization of the total (genomic and mitochondrial) DNA.
6. These BLP guidelines should apply also for the use of cell lines from other organisms, i.e., check for interspecies contamination when working with cells from different species, such as human xenografts in mice. When possible, authenticate cell lines and tissues from mouse, rat, dog, and other species using STR or SNP genotyping [3, 7, 10, 17, 24, 26, 61, 68, 80]. New databases and technologies are under continuous development and should be incorporated into the laboratory's best research practices.
7. Cell lines should also be considered potential sources of infectious agents such as pathogenic viruses. This could be due to the original patient sample being infected or could have occurred sometime between the time that the cells were first established and when received by the researcher's laboratory. Therefore, these cultures should be handled with appropriate *Universal Precautions*. The CDC and WHO have published online protocols for the safe handling of human cell cultures [22, 86].
8. Newly established human and mouse cell lines and preclinical models (xenografts, 3D cultures, stem cells, etc.) generated by a laboratory should be subjected to short tandem repeat (STR) genotype analysis and other appropriate molecular techniques, including species confirmation [7, 24] and phenotype, to establish baselines for future authentication efforts.

II. Most Common Causes of Cell Line Contamination/Misidentification and Recommended Cell Culture Handling Protocols to Avoid these Pitfalls

A. Most Common Causes of Cell Line Contamination/Misidentification:

- Handling multiple lines at the same time in a tissue culture hood since some cells, like those of HeLa, have been shown to "fly" by survival and dispersal in aerosols [25, 67].
- Having two persons working together in the same hood at the same time.
- Using the same bottle of growth medium for multiple cell lines.
- Using the same pipette to apply medium or wash solutions to several different cultures.
- Mislabeling a tissue culture flask or putting the wrong cell suspension in a flask.
- Working with a more aggressive, faster growing cell culture before working with a less aggressive slower growing culture instead of vice versa.
- Not properly cleaning the working surface in the hood between cell lines.

- Storing flasks and equipment in the tissue culture hood and thus preventing complete sanitation of the hood's surface.
  The hood should be cleaned by wiping down with a 10% bleach solution followed by 70% ethanol, and then irradiating it with UV light both before and after a day's work. Note that the UV lights in hoods may not be effective as the light tubes lose their intensity with use and most labs never monitor their intensity over time.
- Not cleaning the tissue culture incubator regularly and not using clean water in the heater jacket of the incubator.

These potential pitfalls should be avoided stringently in order to minimize the chances of cell line mix-ups and cross-contaminations and to preserve the authenticity of the culture samples in one's research. Below are ten recommendations for the handling of cell lines and tissue cultures that should be implemented in all laboratories to avoid using the wrong samples.

B. **Ten Protocols Recommended for Safe Handling Cell and Tissue Cultures**

1. Dedicate individual bottles of media to each cell lines to avoid accidental cross-contamination of cell lines.
2. Label media bottles, culture flasks, and storage vials with clear labels or with printed labels before starting to use them to avoid misidentification of samples.
3. Work with only one cell line at any time to avoid mix-ups and cross-contaminations.
4. When handling cell lines, work with the faster growing cultures and new, untested (quarantined) cells last to minimize the likelihood of rapid growers and potentially contaminated cell lines taking over slower growing cultures.
5. Initially, a large batch of a chosen cell line should be generated from which multiple aliquots are preserved in liquid nitrogen. Test an aliquot of this batch to determine its STR or SNP genotype and whether it is contaminated with mycoplasma before the cells are used for research.
   - Does the STR genotype match the published data or the STR genotype of the original donor?
6. Parson et al. [69] and Korch et al. ([45, 46], unpublished observations) have described how the STR genotypes of MSI-Unstable cell lines can change more rapidly than those of MSI-Stable cell lines with passaging. Cell lines can be checked for microsatellite instability (MSI) using tests such as described by Bacher et al. [9], the ANSI-ATCC ASN-0002-2022 Standard [48], or the Promega kit (MD1641, see https://www.promega.com/products/molecular-diagnostics/amplification/microsatellite-instability-msi-analysis/?catNum=MD1641). MSI and genetic drift due to excessive passaging may provide an explanation for experimental STR genotypes not matching exactly those expected for the cell line(s) and their phenotypes changing.

7. Cell lines routinely used by a laboratory should be authenticated at regular intervals to confirm their identity; a minimum of semi-annual testing is recommended. Minimally. cell lines should be authenticated at the beginning and conclusion of a study so as not to jeopardize the submission of manuscripts and grants and the reproducibility of the findings. The best testing regime would be at the following time points:

    a) The start of the project before using cell lines for experiments,
    b) Regularly during a project when handling samples / culturing cells,
    c) Whenever a novel phenotypic behavior is noticed (see item 1 in section III),
    d) After phenotypic selection (e.g., drug resistance, growth as xenografts; see item 2 in section III),
    e) At the end of a project,
    f) Before submission of grant applications,
    g) Before manuscript submissions, and
    h) Prior to sharing of samples with fellow researchers.

    It can be useful to frequently save aliquots of the residual cell suspensions after passaging cells and store them frozen at -20°C (viability is not necessary). If future analysis detects cross-contamination or sample mix-ups at one of the above steps a through g, DNA from these frozen samples can be extracted and analyzed to ascertain when cross-contaminations or mix-ups may have occurred. This could indicate from which stage the experiments should be repeated.

8. Generally, MSI-Stable cell lines should not be passaged more than 10-20 times and MSI-Unstable cell lines should not be passaged more than 5-10 times to minimize the effects of genetic drift of these cultures.

9. The level of confluence of cells can affect gene expression, protein expression and phosphorylation, and the localization of proteins in cells. Therefore, one should harvest cells at consistent levels of confluence to maintain reproducibility.

10. Media components are variable between suppliers and between batches from the same supplier [76]. Therefore, each batch of culture media and their components (e.g., serum) should be recorded and checked, if necessary, for suitability before being used so that results are reproducible.

III. Three Possible Warning Signs of Undesirable Changes in Cell Line Cultures

Below are three common signs that a cell line culture has been infected or replaced by a population of undesired intra- or inter-species cells.

1. Sudden changes in cell line behavior (e.g., grows faster, change of morphology, or other phenotype) often indicates contamination, an outgrowth of a variant subline, or even the presence of another species. Authenticate such cell lines to verify their identity and purity.

2. The identity of cell lines subjected to selection after modification (e.g., stable transfection and drug selection) should be verified as this process can lead to outgrowth of variants or cross-contaminating "imposter" or hardier cells.
3. Langdon [49] showed that about 7% of DNA sequences from the Human 1000-Genomes Project are from mycoplasma, common human parasitic or commensal bacteria, which commonly contaminate cell cultures. Mycoplasma are known to alter metabolic responses and impact DNA and RNA analysis of contaminated cell line cultures [1, 12, 18, 27-29]. Therefore, all cell lines should be routinely checked for mycoplasma infection and treated accordingly when necessary.

   Cultures can be tested for the presence of mycoplasma with a PCR assay (e.g., Bulldog-Bio eMyco Plus kit) or a luciferase luminescence-based assay (e.g., MycoAlert™ PLUS Mycoplasma Detection Kit Catalog # LT07-703).

IV. Analytical Steps in the Cell Line and Tissue Authentication Process

During the use of cell lines and prior to publishing research results, the frequency of STR genotyping and percent match to the reference sample should be reported using the "Tanabe" algorithm [20] as explained elsewhere [4, 48]. Investigators should understand that cell lines or tissue samples having identical or nearly identical STR- or SNP-based genotypes does not prove that the samples are genetically identical. It only shows that they were originally derived from the same donor.

Any investigator sharing cell lines with fellow researchers, has the ethical responsibility to provide only authenticated cell lines. Results of STR genomic profiling and mycoplasma testing should accompany the cell line transferred to the new investigator.

All laboratories should ensure they comply with the cell line authentication policies of the granting agencies that support their research. If policies from multiple agencies apply, the most stringent regulations should be followed.

V. Considerations of Patient Confidentiality when Publishing Human STR Data

When establishing institutional cell line and tissue authentication guidelines as described herein and deciding to publish human STR genotyping data, patient confidentiality and genetic privacy may need to be considered. The human STR genotyping assays were developed for identification of forensic samples, with newer kits detecting the alleles in 25 or more STR loci. These loci are in noncoding chromosomal regions that are of minimal prognostic or diagnostic value. This expansion of available genetic data increases the potential for linking cell lines with deleterious germline mutations to specific patients and familial relatives, but only if the donor samples are not anonymous or as in forensic cases where individuals have their STR genotypes in a forensic database.

To address this concern, the M.D. Anderson Cancer Center **(**see Supplemental Information**)** and the Japanese Protection of Personal Information Act require that the alleles at only eight STR loci be published as opposed to alleles at thirteen STR loci recommended by the 2022 ANSI-ATCC standard. Genetic privacy and patient confidentiality regulations in the USA, Europe, Australia seem not to specify whether STR genotypes may be published. Journals do not specify a maximum limit of the number STR loci data. The International Journal of Cancer (IJC) requests data for a minimum of 8 STR loci. However, there are already 8,327 human STR genotypes in Cellosaurus (as of November 2022), which were obtained from published and unpublished sources of cell lines from anonymous donors and can include data for up to 31 STR loci, with most profiles containing data for between 8 and 17 loci. Souren et al. [78] argue for one limitation on STR data and that is that the actual electropherograms should not be published, but only be provided to the editor and reviewers for their use. This is because the IJC received several manuscripts in which the authors had copy-and-pasted electropherograms from articles published by other authors. Greater details of this issue are presented in the Supplementary Information.

Therefore, we recommend all scientists and research entities consider the above when establishing their own guidelines regulating the release of human STR results and comply with their local and national regulations that govern the publication of these data.

VI. Supplementary Guidelines, Recommendations, and Useful Information

A printout of the ***Cell Line Authentication Flow Chart*** below can be laminated and posted in tissue culture laboratories as a reminder to all of the critical steps for ensuring authenticated, valid, and reproducible results. Additional sets of recommendations and useful information are described in the attached Supplementary Information file.

Definitions

The following definitions of three frequently used terms when discussing the authenticity cell lines are taken from the ICLAC website.

Authentication

The aim of authentication is to confirm or verify the identity of a cell line, demonstrating that it is derived from the correct species and donor. Testing involves comparison of a test sample to other reference samples from that donor, or to a database of reference samples if donor material is not available, to see whether their genotypes correspond. Ideally, the test method should distinguish between different species and different individuals within that species, although this will depend on the technology available to the field of authentication testing.

Not all currently used test methods have the power of discrimination of STR profiling or SNP testing; therefore, authentication may not in all cases lead to unambiguous identification of

cells to a specific donor. Where unambiguous identification is not possible, species verification using methods such as mitochondrial CO1 barcoding [7] or species-specific PCR of mitochondrial cytochrome B [24], is used as the best alternative currently available.

Cross-contamination

The term contamination refers to introduction of foreign material into a cell culture. Cross-contamination occurs when that foreign material consists of cells from another culture, either human cells or non-human cells arising from species such as mouse or rat. Cross-contamination initially results in a mixed culture, containing cells from the authentic culture and the contaminant. If the contaminant has a survival advantage – for example, if it proliferates more rapidly – it will overgrow and replace the authentic cells within the culture. A contaminant usually comes from a different donor or species and so can be detected by authentication testing.

Misidentification

A misidentified cell line no longer corresponds to the donor or species from which it was originally established. Misidentification may arise due to cross-contamination. It may also arise from a variety of errors, including mislabeling of samples. If it happens early – for example, during cell line establishment – there will be no authentic material retained, and the cell line is considered to be a falsely identified or misidentified cell line. If misidentification happens late – for example, after the cell line is established and distributed to other locations – then authentic material may still exist and only some stocks may be false.

Misidentification does not refer to problems with the technical procedure of authenticating cell lines. It also does not typically extend to other characteristics such as tissue type, cell type, or disease state. If the tissue type, cell type, or disease state of a cell line is incorrectly attributed, the cell line is considered to be misclassified.

Contact Information

The authors can be contacted at the following addresses:
Christopher.Korch@CUAnschutz.edu and Rebecca.Schweppe@CUAnschutz.edu.

APPENDIX

Best Laboratory Practices for Cell Line and Tissue Sample Authentication to Ensure Valid, Reproducible and Robust Research

Supplementary Technical Guidelines and Information

Christopher T Korch[1] and Rebecca E Schweppe[2]

[1]Division of Medical Oncology
[2]Division of Endocrinology, Metabolism, and Diabetes
School of Medicine
University of Colorado - Anschutz Medical Campus
12801 East 17th Avenue, RC-1 South
Aurora, Colorado 80045, USA

Table of Contents

Purpose of Supplemental Guidelines and Information

Below are additional descriptions of different aspects of cell line and tissue authentication and suggestions of how to prepare manuscripts for submission of grant applications and for publication of results in journals.

Section 1 - Human Genotyping by the STR / DNA Profiling Method

Human cell lines and tissue samples must be authenticated before they are used (including for each new lot of cryopreserved cells), during their usage, and prior to publication of the research results. Two seminal references describe how to perform authentication by STR genotyping of human and cell lines. The ANSI-ATCC ASN-0002 Standard by Korch et al. [48] published in 2021 with minor revisions in 2022 [48] is an update of the 2011 ANSI-ATCC Standard [6]. The update describes in great detail how to perform STR genotyping of human samples by laboratories offering this service and how to interpret STR data appropriately by both the service personnel and the researcher. The Assay Guidance Manual (AGM) written by Almeida and Korch in 2022 [4], published by the National Center for Biotechnology Information, describes how to perform the STR analysis of both human and mouse cell line and tissue samples and shows how researchers can understand and interpret the resulting data.

Human STR data can be obtained by using one of several commercial kits (e.g., ABI Identifiler Plus or GlobalFiler, or Promega Powerplex -16, -18D, 21, Fusion, or Fusion 6C), which test 13 or more STR loci on human autosomal chromosomes and of the allele(s) at the Amelogenin locus on the human gonosomal chromosomes by PCR amplification. Numerous STR kits from different suppliers are listed in the 2022 ANSI-ATCC Standard for authentication of human cell lines [48]. The mention of specific products is only for illustrative purposes and should not be construed as recommendation of any specific product. At a minimum, the STR alleles present, at least, at the following 13 loci must be determined: CSF1PO, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, FGA, TH01, TPOX, and vWA. Some commercial kits test for additional loci which can be helpful in identifying culture samples that have very few (≤4) diallelic STR loci.

Historically, allelic data for the amelogenin locus on the X and Y chromosomes has been included in the calculations of the STR genotype match scores. This is not valid as discussed elsewhere [4, 48] since it is not an STR locus and as reported by others approximately 40% of male-derived cell lines lack the Y-linked amelogenin locus [50]. Therefore, this locus should not be included in calculating the genotype match scores.

Calculations of match scores by three different algorithms are available by using the CLASTR tool and associated data entry template on the Cellosaurus website (https://www.cellosaurus.org). Explanations for using the different algorithms are explained in the help menu and elsewhere [4, 48]. The resulting STR genotypes should then be used to search different databases to confirm the identity of the culture samples being used. Recommended guidelines and other resources, including links to STR databases, are available at the ICLAC website. The most useful STR genotyping search engine is the CLASTR tool at the Cellosaurus knowledge resource on cell lines website [11]. From the literature, as of September 2022, this database has compiled information about 141,885 cell lines, of which 106,170 are human-

derived, 21,928 are from mouse, and 2,637 from rat. The CLASTR search tool contains the STR profiles for 8,159 human cell lines, 78 mouse cell lines, and 36 dog cell lines.

SNP analysis is also becoming an accepted method for culture identification using a minimum of 48 SNPs [50]. However, a search tool is not currently available for identifying cell lines based on SNPs.

Section 2 - Additional Authentication Considerations

Section 2.1 - Growth Media and Cultivation Conditions

- Media components are variable between suppliers and between batches from the same supplier [76]. Therefore, each batch of culture media and their components should be recorded and validated, if necessary, for suitability before being used so that results are reproducible. The media will not affect the STR genotype, but can alter the phenotype of the cells; for example, changing the expression of specific genes being used to characterize the cell line.
- The level of confluence of cells can affect the expression of many genes in cell lines [65]. Therefore, one should harvest cells at consistent levels of confluence to maintain reproducibility between assays.

Section 2.2 - Cell Line Heterogeneity and Genetic Drift

Researchers need to be aware that cell line cultures are heterogeneous mixtures of cells with slightly different genotypes and phenotypes that have arisen in the original tumor or during culturing due to genetic drift and selection. This may or may not give rise to variant STR genotypes.

Kleensang et al. described how two aliquots of the same batch of MCF-7 (RRID:CVCL_0031)obtained at the same time from the ATCC could diverge phenotypically within 7 passages even though their STR genotypes at 8 loci were identical [43]. It might seem that establishing cell lines from single cells would yield genotypically pure clones that were derived from single cells. However, these can also evolve at the genetic and physiological levels anew and develop into genetically heterogeneous cultures as shown for MCF-7 and other cell lines by Ben-David et al. [15] and for HeLa cells (RRID:CVCL_0030) by Liu et al. [56].

Genetic instability can be due to microsatellite instability (MSI) caused by defects in either the cell's DNA mismatch repair or its DNA replicative machinery; however, MCF-7 and HeLa cells are MSI-Stable. Researchers should be aware that microsatellite instability can, with culturing and selection, give rise to variant STR profiles, which may complicate the interpretation of the STR data. Assaying for shared SNPs may be used to confirm relatedness of MSI-Unstable cell lines that show variable STR profiles as shown by Korch et al. [45].

Section 2.3 - Identification of Cell Line Species by PCR Assays

The presence of interspecies contaminants (e.g., non-human cells in human-derived cell lines) can further confuse research endeavors and conclusions. The STR profiling kits used for

human-derived samples cannot detect the presence of non-human cells in tissue samples and cultures. It has been shown that interspecies contamination of cell lines (~ 9%) can be as frequent as intra-species contaminations [21]. This is especially important when working with cell lines from different species and with xenograft models.

Cooper et al. described a multiplex species-specific PCR-based assay, which tests for the presence of two mitochondrial genes, to rapidly identify the most common cell culture species and quickly detect inter-species contaminations (>2-5%) in a cell line culture/tissue sample [24]. The so-called universal DNA barcoding method of the International Barcode of Life organization based on determining the sequence of the standard barcode for almost all metazoan animal groups (mitochondrial cytochrome c oxidase I gene, "COI") is a highly effective method for identifying many animal groups [24]; however, this method cannot detect contaminations < 20%.

If rodent tissues and cell cultures need to be authenticated, one should consider using STR genotyping for mouse lines [3, 4] or rat cell lines. SNP genotyping assays for either mouse [26, 61] or rat [70, 77] have been published, but we do not know whether they are commercially available. If dog cell lines need to be authenticated, one can use the STR genotyping method described by O'Donoghue et al. [66], which is compatible with the CLASTR search tool. Authentication by STR and microsatellite genotyping methods of canine, mouse, and rat cell lines are commercially available, for example, using LabCorp's Cell Line Authentication and Research Services or the CellCheck service at Idexx BioAnalytics. The CLASTR search tool in the Cellosaurus can search for genotype matches in the dog and mouse STR databases [10, 11, 74].

Detection of interspecies contaminations using COI, cannot detect interspecies hybrid cells. To partially address this possibility, Huang et al. described a method that screens for both mitochondrial COI regions and some autosomal regions of 10 species [40]. Almeida et al. have included two human STR loci in their mouse-specific STR genotyping assay, which would allow the detection of two human chromosomes among mouse chromosomes in a sample [2]. Another approach to detect mouse-human cell hybrids ("hybridomas") could be to test a DNA sample for chromosomal-linked STR alleles with both a mouse STR kit and a human STR kit.

Section 3 - Templates for Grant and Manuscript Descriptions of Guidelines and Cell Line Authentication Procedures used in Report

Assuming that the policy guidelines are fully implemented in a researcher's laboratory, below are a few generic examples of short descriptions which should be modified appropriately for the methods section of grant applications and manuscripts. The highlighted underlined spaces are intended for specifying the indicated information.

Section 3.1 - Established Human Cell Lines, Patient-derived Explants and Xenografts of Human Tissue, and Development of New Human Cell Lines

- The authenticity of cell lines used these experiments was checked by verifying that the cell line's name, or variants of the name, was not present in the list of known misidentified cell lines on the International Cell Line Authentication Committee's website or the Cellosaurus cell line database [10].

- Prior to their use, all established human cell lines used in these experiments were genotyped by STR analysis using the _____ kit (specify kit used, e.g., ThermoFisher Applied Biosystems AmpFLSTR® Identifiler® PCR Amplification Kit, catalog number 4322288; or Promega Powerplex 16HS System. catalog number DC2101) or by the _____ laboratory (specify resource or commercial entity which performed the assays).
- The identity of the cell line was confirmed by comparing the resulting STR genotype to reference profiles from the originator or source of the cell line or through searching the consolidated STR data from four cell line repositories (ATCC, DSMZ, JCRB, RIKEN) base available at the DSMZ website [30] and/or the Cellosaurus STR database. The percent match with the reference genotype was ___ % (insert calculated value) using the Tanabe algorithm (and/or one of the other algorithms available on CLASTR without including the data for the amelogenin result in the search.
- In the case of newly established cell lines or if the reference STR genotype of an established cell line is not available in the literature or an STR database, describe whether:
    - Attempts were made to obtain a sample of the original tissue from which the cell line was established; and
    - Its STR genotype was determined and compared to the cell line to verify that the cell line was established from the specified tissue/patient and that it was not a contaminant.

  If the STR genotype was determined, provide the data for publication in a table format in either the main body or the supplemental information of the report and submit it to Cellosaurus for inclusion in this essential resource.

  NOTE - Actual electropherogram(s) should be only provided for the editors and reviewers and should not be published to avoid being copied by others.
- All cultures were STR genotyped at the end of each series of experiments and prior to submission of grant applications and manuscripts.
- All cultures were confirmed not to be contaminated with mycoplasma by a PCR assay, e.g., Bulldog- Bio eMyco Plus kit (Catalog # 2523448) or the luciferase luminescence-based Lonza MycoAlert™ PLUS Mycoplasma Detection Kit (Catalog # LT07-118).
- If SNP genotyping was used, modify the above descriptions appropriately. Note that as of September 2022 there is no searchable public database for comparing SNP genotyping data between samples.

Section 3.2 - Additional Descriptions for Xenografts, 3D cultures, and Stem cells

- The STR genotype of the source tissue (e.g., original tumor tissue, lymphocytes from blood) was determined and used as a reference genotype for any xenografts that were propagated through, e.g., mice.
- Each xenograft tumor passage was checked by STR genotyping to confirm that there had not been any inadvertent mix-ups of samples or contamination by a cell line being used concurrently; and

- The presence of inter-species contaminating cells (e.g., mouse cells in a human xenograft) was checked by the method of Cooper et al. [24].

  NOTE – The Cooper et al. method checks for the presence of mitochondrial DNA and not for the presence of autosomal DNA from other species. Therefore, it cannot detect mouse – human cell hybrids which have a mixture of mouse and human chromosomes. Such hybrids need to be screened for by karyotyping of individual cells as was done by Jacobsen et al. [42] or by multiplex PCR analysis that screens for the presence of one or more SNPs or mouse chromosomal STR alleles on multiple mouse chromosomes. This PCR assay should be considerably more sensitive than karyotyping. Alternatively, as noted above a DNA sample may be tested with multiple STR kits.

Section 4 - The Cell Line Authentication Policy implemented at the *International Journal of Cancer* (IJC) with Modifications.

This journal's cell line authentication policy is the clearest and most definitive of any such policy. Below, are the seven main stipulations for manuscript submissions to the IJC as outlined by Souren et al. [78], which are based on the requirements presented on the IJC website, with a few clarifications we recommend:
https://onlinelibrary.wiley.com/page/journal/10970215/homepage/ForAuthors.html#AUCEL

1. Authors must provide cell line authentication documents that are not older than three years of all continuous human cell lines used in their manuscript.
   - Note that IJC requires electropherograms of authentication data and they will check the data against Cellosaurus. Also, Souren et al.[78] suggested that these data be checked by a person qualified to do so.
   - Souren et al. [78] note that IJC recommends that the submitted high quality legible electropherograms not be published (i.e., only be included for the reviewers and editors to examine). In the main text or supplementary information, which is to be published, include a table summarizing the STR results. We suggest placing the STR data in an Excel/CSV table using the Cellosaurus / CLASTR format (see CLASTR website) so that it is accessible and easily searchable by the readers.

     The reason for not including the electropherograms is that they may be copied by others and claimed as their own as the IJC has reported [78].
2. STR profiling is the preferred method for cell line authentication of human, mouse, rat, and dog cell lines. Authors can perform STR profiling in their own laboratory or use the service provided by a laboratory or cell bank with certified quality control. Either way, the cell line authentication documents submitted with a manuscript should include high-quality legible electropherograms.
3. For continuous human cell lines obtained within the previous three years from a commercial source that guarantees cell line authenticity through in-house quality control measures (e.g., ATCC, DSMZ), the corresponding purchase orders or invoices are acceptable evidence of cell line authentication.

- Note that the ANSI-ATCC Standard for Authentication of Cell Lines by STR Profiling [48] and several cell line authentication policies recommend that profiling be performed more frequently than every three years (e.g., once per year) and when phenotypic changes are noted in the culture.
- Note that the IJC has received falsified invoices for some manuscripts [78]. Also, in a recent PubMed literature search for the sources of the false cell line L-02 (a.k.a. HL-7702), some reports claimed it was obtained from the ATCC. However, the ATCC never had this imposter cell line among its offerings (Korch, data not show, 2022).

4. The IJC also requests authentication of human cell lines for which no reference STR profile is available. Prior to submission, the obtained STR profile should be compared to a public database (e.g., Cellosaurus), and should show that the cell line is unique and not cross-contaminated or misidentified.
5. The IJC also accepts single nucleotide polymorphism (SNP) based cell line authentication reports from service providers with certified quality control, but only for cell lines for which a SNP-based reference profile is publicly available.
6. Authors of studies describing the establishment of new human cell lines are strongly encouraged to include the summarized STR results in the manuscript for future reference, such as inclusion of data in Cellosaurus, preferably in the .csv, .txt, or .xlsx format of the Cellosaurus template for CLASTR searches. See screenshot below from https://www.cellosaurus.org/str-search/help.html.

3. Input File

Using the Load File button from the user interface, it is possible to directly import STR profile data from a table file. Both mono and multi-samples files are supported. The functionality can be used to perform a similarity search on several samples at a time or to load quickly and reliably the marker data of a sample into the user interface.

3.1 Formats

The table file can be formatted either as an Excel file (.xls or .xlsx extension) or as a plain text file (.csv, .tsv or .txt extension). By default, the tool will assume that each row (except from the header) is a distinct sample. Note that a column named "Sample Name", "Name" or "Sample" is required and each submitted sample needs to have a corresponding value. The ordering of the marker columns is not important. The name of the markers need to be indicated correctly. For Amelogenin, the program recognizes "Amel" and "AM" as valid names.

Sample Name	Amelogenin	CSF1PO	D5S818	D7S820	D13S317	D16S539	TH01	TPOX	vWA
BICR 16	X	12	13	10	11	12	6,9	8,11	17,19
ND31618	X	12,13	10	8,10	11,12	11	7,9	8,9	16
Lu-138	X	11,12	10,12	10,11	8,11	11	7,9	8,11	17,18
NCI-H125	X	7	10	10	11	5,12	7	8,9	17
GK-5	X,Y	10,11	10,12	10	12	11	9,9.3	8	16,19

Example of a properly formatted table

7. The following information must be included in the Materials and Methods section:
   - All cell lines used must be listed using the official cell line name and its Research Resource Identifier (RRID) as available in the ExPASy Cellosaurus database (e.g., HeLa

(RRID: CVCL_0030). Note that Babic et al. [8] showed that there are fewer problematic cell lines in manuscripts citing the RRIDs.

- The source/supplier of all cell lines used must be provided.
- A statement confirming that all human cell lines have been authenticated using STR (or SNP) genotyping within the last 3 years.
- A statement confirming that all experiments were performed with mycoplasma-free cells.

Section 5 - Considerations of Patient Confidentiality when Publishing Human STR Data

When establishing institutional cell line and tissue authentication guidelines as described herein and deciding to publish human STR genotyping data, patient confidentiality may need to be considered. The human STR genotyping assays were developed for identification of forensic samples, with newer kits detecting the alleles in 25 to 30 STR loci. These alleles are in noncoding chromosomal regions and are of minimal prognostic/diagnostic value. The original STR assays used for cell line identification between 1999 and 2013 [6, 59, 72, 81] were based on kits using nine or fewer STR loci. Between 2013 and 2021 it became evident that data for additional loci was needed to distinguish between cell lines, especially ones that had many monoallelic STR loci. To address this issue, the 2022 ANSI-ATCC standard for authentication of human cell lines recommended that data from a minimum of thirteen STR loci be used for the identification of cell lines [48].

This expansion of available genetic data increases the potential for identifying a specific patient and familial relatives, which means their patient confidentiality might be breached. To address this concern (as discussed in detail in the 2022 ANSI-ATCC cell line authentication standard [48]) the M.D. Anderson Cancer Center and the Japanese Protection of Personal Information Act require that the alleles at only eight STR loci be published as opposed to alleles at thirteen STR loci recommended by the 2022 ANSI-ATCC standard. The MD Anderson rule is self-mandated, whereas in Japan it is stipulated by national law. However, as of November 2022, there are already 8,327 human STR genotypes publicly available in Cellosaurus, which were obtained from published and unpublished sources and can include data for up to 31 loci, with most profiles containing data for between 8 and 17 loci.

Under the 2018 General Data Protection Regulation (GDPR) genetic data privacy is regulated in the European Union, but it is not clear to what extent each country can set their own rules and how much genetic information may be revealed in biospecimens. In Australia, there are rules governing the ethical use of human biospecimens, including cell lines. In the USA, patient confidentiality and medical insurability are regulated by the Patient Protection and Affordable Care Act (AFA), Health Insurance Portability and Accountability Act (HIPAA), the Genetic Information Nondiscrimination Act (GINA), and the American Disabilities Act (ADA) [23, 48]. However, to our knowledge STR genotyping is not specifically addressed in Australia, Europe, or the US.

We are not aware of any journal that limits the number of loci for which STR data can be reported. The International Journal of Cancer (IJC) requests data for a minimum of eight core STR loci as outlined by Souren et al. [78] and described in the IJC instructions to authors. Therefore, the decision of how much human STR genotyping data to publish rests with individual laboratories and research organizations. It is worth considering the recommendation of Souren et al. [78] that only the tabulated STR data be published. The actual electropherograms should not be published, but only be provided to the editor and reviewers for their use. This is because the IJC received several manuscripts in which the authors had copy-and-pasted electropherograms from articles published by other authors.

Section 6 - Recommended Online Resources

Section 6 .1 International Cell Line Authentication Committee (ICLAC) Website

The ICLAC website has several recommendations for cell line authentication listed under Resource Documents, including:

- Cell Line Policy for Research Institutions
- Match Criteria Worksheet for Human Cell Line Authentication.
- Cell Line Checklist for Manuscripts and Grant Applications
- Guide to Human Cell Line Authentication
- Guide to Mouse Cell Line Authentication
- Advice to Scientists; Incorporating Authentication into Everyday Culture Practice
- Naming a Cell Line
- Definitions – Authentication, Cross-contamination, Misidentification
- Register of cross-contaminated or misidentified cell lines
- Links to STR profile databases for Cell Lines at the ATCC, Cellosaurus, the German DSMZ tissue culture facility, and the CLIMA database of STR profile.

Section 6.2 - Cellosaurus Website Cell Line Knowledge Resource Database with the CLASTR Search Tool for Finding Cell Lines with Matching STR Genotypes

Cellosaurus is a knowledge resource for cell lines. It attempts to describe all cell lines used in biomedical research. Its scope includes:

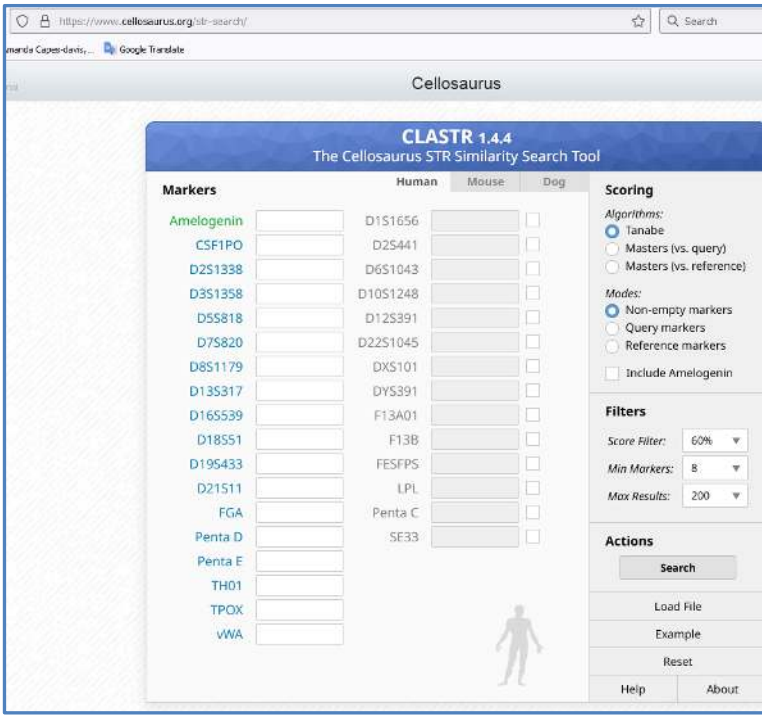
- Immortalized cell lines
- Naturally immortal cell lines (example: stem cell lines)
- Finite life cell lines when those are distributed and used widely
- Vertebrate cell line with an emphasis on human, mouse and rat cell lines
- Invertebrate (insects and ticks) cell lines

Its scope does not include:

- Primary cell lines (with the exception of the finite life cell lines described above)
- Plant cell lines

In addition, it offers the CLASTR - STR similarity search tool (see adjacent figure) which allows the comparison of STR genotypes of human cell lines, mouse cell lines, and dog cell lines. The different genotypes can be compared using three different algorithms:

- Tanabe algorithm,
- Masters algorithm for comparing reference profiles to query profiles, and
- Alternative Masters for comparing query profiles to reference profiles.



CLASTR STR search tool, available at Cellosaurus (https://www.cellosaurus.org/str-search/)

As of September, 2022, Release 43 contains information about 141,885 cell lines (106,170 of human origin, 21,928 from mouse, and 2,637 from rat). The CLASTR tool can compare one or more query genotype against genotypes for 8,327 human cell lines, 83 mouse cell lines, and 36 dog cell lines.

Section 6.3 - Published Guidelines for Culturing and Handling of Cell Lines

- The Guidelines for the use of cell lines in biomedical research by Geraghty et al. [37].
- The 7th edition of *Culture of Animal Cells: A Manual of Basic Technique and Specialized Applications* by R. Ian Freshney [34].
- The 8th edition *of Freshney's Culture of Animal Cells: A Manual of Basic Technique and Specialized Applications* by Amanda Capes-Davis and R. Ian Freshney [19]. This replaces the 7th edition cited above.
- The ANSI/ATCC ASN-0002-2022 Authentication of Human Cell Lines: Standardization of STR Profiling of 2021 by Korch et al. [48] replaces the standard with a similar name that was published in 2011 [6]. This new Standard is available online at: https://webstore.ansi.org/standards/atcc/ansiatccasn00022022

- The Assay Guidance Manual (AGM) for Authentication of Human and Mouse Cell Lines by Short Tandem Repeat (STR) DNA Genotype Analysis by Almeida and Korch [4], which is an update of the AGM by Reid et al. [72].

Section 6.4 - National Institutes of Health (NIH) Notices Pertaining to Cell Lines

Notification Regarding Authentication of Cultured Cell Lines, November 28, 2007. NOT-OD-08-017. In addition, NIH Updated policy (June 2015) *Notification Enhancing Reproducibility through Rigor and Transparency.* and NIH Notices NOT-OD-15-103, NOT-OD-16-011, and NOT-OD-16-012 and the *Principles and Guidelines for Reporting Preclinical Research* .

Section 6.5 - Other Useful Links including Cell Line Repositories

- **American Type Culture Collection (ATCC)** - https://www.atcc.org/
- **Biocompare** - https://www.biocompare.com/Cell-Line-Authentication/ and https://www.biocompare.com/Reproducibility/Cell-Line-Authentication/
- **Biosample** - https://www.ncbi.nlm.nih.gov/biosample
- **Catalogue of Somatic Mutations in Cancer (COSMIC)**. Version 97 (29 November 2022) contains curated DNA sequence information on approximately 1,000 cell lines: https://cancer.sanger.ac.uk/cell_lines
- **Leibniz Institute, DSMZ-German Collection of Microorganisms and Cell Cultures GmbH (DSMZ)** – This organization has a searchable collection of cell line STR genotyping data from DSMZ, ATCC, JCRB, and RIKEN. These data have now been integrated into Cellosaurus. https://www.dsmz.de and https://www.dsmz.de/services/human-and-animal-cell-lines/online-str-analysis.
- **The** M.D. Anderson's Cell Line Authentication Policy (ACA#1044) requires all researchers to validate their cell lines at least once per year. It is described at the following site: https://www.mdanderson.org/content/dam/mdanderson/documents/core-facilities/Characterized%20Cell%20Line%20Core%20Facility/CCLC_%20Policy_ACA1044.pdf
- **JCRB Cell Bank (Japanese Collection of Research Bioresources Cell Bank) of the** National Institutes of Biomedical Innovation, Health and Nutrition - https://cellbank.nibiohn.go.jp/english/ is one of the two major repository of cell lines in Japan.
- **RIKEN (Rikagaku Kenkyūsho) BioResource Research Center** – The cell line repository for a nation-wide group of Japanese institutions that perform research in multiple scientific disciplines, including biomedical research (https://cell.brc.riken.jp/en/). The STR genotypes of its many cell lines are accessible in the databases of STR profiles maintained by the DSMZ and Cellosaurus.

References

1. **Aldecoa-Otalora E, Langdon W, Cunningham P, Arno MJ.** 2009. Unexpected presence of mycoplasma probes on human microarrays. Biotechniques **47:**1013-1015. PMID: 20047202
2. **Almeida JL, Dakic A, Kindig K, Kone M, Letham DLD, Langdon S, Peat R, Holding-Pillai J, Hall EM, Ladd M, Shaffer MD, Berg H, Li J, Wigger G, Lund S, Steffen CR, Fransway BB, Geraghty B, Natoli M, Bauer B, Gollin SM, Lewis DW, Reid Y.** 2019. Interlaboratory study to validate a STR profiling method for intraspecies identification of mouse cell lines. PLoS One **14:**e0218412. PMID: 31220119
3. **Almeida JL, Hill CR, Cole KD.** 2014. Mouse cell line authentication. Cytotechnology **66:**133-147. PMID: 23430347
4. **Almeida JL, Korch CT.** 2022. Authentication of Human and Mouse Cell Lines by Short Tandem Repeat (STR) DNA Genotype Analysis. *In* Sittampalam GS GA, Brimacombe K, et al. (ed), Assay Guidance Manual [Internet]. National Center for Advancing Translational Sciences; 2004-. Bethesda, MD.
5. **American Type Culture Collection Standards Development Organization Workgroup ASN-0002.** 2010. Cell line misidentification: the beginning of the end. Nat Rev Cancer **10:**441-448. PMID: 20448633
6. **ATCC SDO.** 2011. ASN-0002. Authentication of Human Cell Lines: Standardization of STR Profiling. ANSI eStandards Store, *on* ATCC-Standards Development Organization (SDO). https://webstore.ansi.org/standards/atcc/ansiatccasn00022011. Accessed September, 2021.
7. **ATCC SDO.** 2015. ASN-0003 Species-level identification of animal cells through mitochondrial cytochrome c oxidase subunit 1 (CO1) DNA barcodes, *on* ATCC-Standards Development Organization (SDO). https://webstore.ansi.org/Standards/ATCC/ANSIATCCASN00032015. Accessed
8. **Babic Z, Capes-Davis A, Martone ME, Bairoch A, Ozyurt IB, Gillespie TH, Bandrowski AE.** 2019. Incidences of problematic cell lines are lower in papers that use RRIDs to identify cell lines. Elife **8**. PMID: 30693867
9. **Bacher JW, Sievers CK, Albrecht DM, Grimes IC, Weiss JM, Matkowskyj KA, Agni RM, Vyazunova I, Clipson L, Storts DR, Thliveris AT, Halberg RB.** 2015. Improved detection of microsatellite instability in early colorectal lesions. PLoS One **10:**e0132727. PMID: 26252492
10. **Bairoch A.** 2018. The Cellosaurus, a cell-line knowledge resource. J Biomol Tech **29:**25-38. PMID: 29805321
11. **Bairoch A.** 2018. The Cellosaurus: a cell line knowledge resource, *on* The CALIPHO group at the Swiss Institute of Bioinformatics, Geneva, Switzerland. https://web.expasy.org/cellosaurus/. Accessed May 2019.
12. **Barile MF.** 1968. Mycoplasma and cell cultures. Natl Cancer Inst Monogr **29:**201-204. PMID: 4976863
13. **Baust JM, Buehring GC, Campbell L, Elmore E, Harbell JW, Nims RW, Price P, Reid YA, Simione F.** 2017. Best practices in cell culture: an overview. In Vitro Cell Dev Biol Anim **53:**669-672. PMID: 28808859
14. **Begley CG.** 2013. Six red flags for suspect work. Nature **497:**433-434. PMID: 23698428
15. **Ben-David U, Siranosian B, Ha G, Tang H, Oren Y, Hinohara K, Strathdee CA, Dempster J, Lyons NJ, Burns R, Nag A, Kugener G, Cimini B, Tsvetkov P,**

Maruvka YE, O'Rourke R, Garrity A, Tubelli AA, Bandopadhayay P, Tsherniak A, Vazquez F, Wong B, Birger C, Ghandi M, Thorner AR, Bittker JA, Meyerson M, Getz G, Beroukhim R, Golub TR. 2018. Genetic and transcriptional evolution alters cancer cell line drug response. Nature **560:**325-330. PMID: 30089904

16. **Biocompare.** 2018. Reproducibility issues in life science research: The importance of cell line authentication, First ed, vol 2022, p Video interview of 13 scientists working on the issue of cell line authentication., San Francisco, CA 94080, USA. PMID:
17. **Callari M, Batra AS, Batra RN, Sammut SJ, Greenwood W, Clifford H, Hercus C, Chin SF, Bruna A, Rueda OM, Caldas C.** 2018. Computational approach to discriminate human and mouse sequences in patient-derived tumour xenografts. BMC Genomics **19:**19. PMID: 29304755
18. **Callaway E.** 2014. Contamination hits cell work. Nature **511:**518. PMID: 25079533
19. **Capes-Davis A, Freshney RI.** 2021. Freshney's Culture of Animal Cells: A Manual of Basic Technique and Specialized Applications, 8th Edition, Eighth ed. Wiley-Blackwell, Hoboken, NJ, USA.
20. **Capes-Davis A, Reid YA, Kline MC, Storts DR, Strauss E, Dirks WG, Drexler HG, MacLeod RA, Sykes G, Kohara A, Nakamura Y, Elmore E, Nims RW, Alston-Roberts C, Barallon R, Los GV, Nardone RM, Price PJ, Steuer A, Thomson J, Masters JR, Kerrigan L.** 2013. Match criteria for human cell line authentication: where do we draw the line? Int J Cancer **132:**2510-2519. PMID: 23136038
21. **Capes-Davis A, Theodosopoulos G, Atkin I, Drexler HG, Kohara A, MacLeod RA, Masters JR, Nakamura Y, Reid YA, Reddel RR, Freshney RI.** 2010. Check your cultures! A list of cross-contaminated or misidentified cell lines. Int J Cancer **127:**1-8. PMID: 20143388
22. **Center for Disease Control and Prevention CDC.** 2009. CDC Biosafety in Microbiological and Biomedical Laboratories (BMBL), 5th edition *on* USA Department of Human Health and Services. https://www.cdc.gov/labs/BMBL.html). Accessed May 2019.
23. **Clayton EW, Evans BJ, Hazel JW, Rothstein MA.** 2019. The law of genetic privacy: applications, implications, and limitations. J Law Biosci **6:**1-36. PMID: 31666963
24. **Cooper JK, Sykes G, King S, Cottrill K, Ivanova NV, Hanner R, Ikonomi P.** 2007. Species identification in cell culture: a two-pronged molecular approach. In Vitro Cell Dev Biol Anim **43:**344-351. PMID: 17934781
25. **Coriell LL.** 1962. Detection and elimination of contaminating organisms. Natl Cancer Inst Monogr **7:**33-53. PMID: 13881339
26. **Didion JP, Buus RJ, Naghashfar Z, Threadgill DW, Morse HC, 3rd, de Villena FP.** 2014. SNP array profiling of mouse cell lines identifies their strains of origin and reveals cross-contamination and widespread aneuploidy. BMC Genomics **15:**847. PMID: 25277546
27. **Drexler HG, Dirks WG, MacLeod RA, Uphoff CC.** 2017. False and mycoplasma-contaminated leukemia-lymphoma cell lines: time for a reappraisal. Int J Cancer **140:**1209-1214. PMID: 27870004
28. **Drexler HG, Uphoff CC.** 2002. Mycoplasma contamination of cell cultures: Incidence, sources, effects, detection, elimination, prevention. Cytotechnology **39:**75-90. PMID: 19003295

29. **Drexler HG, Uphoff CC, Dirks WG, MacLeod RA.** 2002. Mix-ups and mycoplasma: the enemies within. Leuk Res **26:**329-333. PMID: 11839374
30. **DSMZ Leibniz Institute German Collection of Microorganisms and Cell Cultures GmbH.** 2019. Online STR Analysis - Short Tandem Repeat (STR) Profile Search, *on* DSMZ. https://www.dsmz.de/services/human-and-animal-cell-lines/online-str-analysis. Accessed 8 July 2019.
31. **Freedman LP, Cockburn IM, Simcoe TS.** 2015. The Economics of Reproducibility in Preclinical Research. PLoS Biol **13:**e1002165. PMID: 26057340
32. **Freedman LP, Gibson MC, Ethier SP, Soule HR, Neve RM, Reid YA.** 2015. Reproducibility: changing the policies and culture of cell line authentication. Nat Methods **12:**493-497. PMID: 26020501
33. **Freedman LP, Gibson MC, Wisman R, Ethier SP, Soule HR, Reid YA, Neve RM.** 2015. The culture of cell culture practices and authentication--Results from a 2015 Survey. Biotechniques **59:**189-190, 192. PMID: 26458546
34. **Freshney RI.** 2016. Culture of animal cells: a manual of basic technique and specialized applications, 7th ed. Wiley-Blackwell, Hoboken, New Jersey, U. S. A.
35. **Fusenig NE, Capes-Davis A, Bianchini F, Sundell S, Lichter P.** 2017. The need for a worldwide consensus for cell line authentication: Experience implementing a mandatory requirement at the International Journal of Cancer. PLoS Biol **15:**e2001438. PMID: 28414712
36. **Gartler SM.** 1967. Genetic markers as tracers in cell culture. Natl Cancer Inst Monogr **26:**167-195. PMID: 4864103
37. **Geraghty RJ, Capes-Davis A, Davis JM, Downward J, Freshney RI, Knezevic I, Lovell-Badge R, Masters JR, Meredith J, Stacey GN, Thraves P, Vias M.** 2014. Guidelines for the use of cell lines in biomedical research. Br J Cancer **111:**1021-1046. PMID: 25117809
38. **Gorphe P.** 2019. A comprehensive review of Hep-2 cell line in translational research for laryngeal cancer. Am J Cancer Res **9:**644-649. PMID: 31105993
39. **Horbach SPKM, Halffman W.** 2017. The ghosts of HeLa: How cell line misidentification contaminates the scientific literature. PLoS One **12:**e0186281. PMID: 29023500
40. **Huang Y, Liu Y, Zheng C, Shen C.** 2017. Investigation of Cross-Contamination and Misidentification of 278 Widely Used Tumor Cell Lines. PLoS One **12:**e0170384. PMID: 28107433
41. **International Cell Line Authentication Committee (ICLAC).** 14 October 2018 2018. Database of cross-contaminated or misidentified cell lines. http://iclac.org/databases/cross-contaminations/. Accessed 09 December 2018.
42. **Jacobsen BM, Harrell JC, Jedlicka P, Borges VF, Varella-Garcia M, Horwitz KB.** 2006. Spontaneous fusion with, and transformation of mouse stroma by, malignant human breast cancer epithelium. Cancer Res **66:**8274-8279. PMID: 16912208
43. **Kleensang A, Vantangoli MM, Odwin-DaCosta S, Andersen ME, Boekelheide K, Bouhifd M, Fornace AJ, Jr., Li HH, Livi CB, Madnick S, Maertens A, Rosenberg M, Yager JD, Zhaog L, Hartung T.** 2016. Genetic variability in a frozen batch of MCF-7 cells invisible in routine authentication affecting cell function. Sci Rep **6:**28994. PMID: 27456714

44. **Korch C, Hall EM, Dirks WG, Ewing M, Faries M, Varella-Garcia M, Robinson S, Storts D, Turner JA, Wang Y, Burnett EC, Healy L, Kniss D, Neve RM, Nims RW, Reid YA, Robinson WA, Capes-Davis A.** 2018. Authentication of M14 melanoma cell line proves misidentification of MDA-MB-435 breast cancer cell line. Int J Cancer **142:**561-572. PMID: 28940260
45. **Korch C, Spillman MA, Jackson TA, Jacobsen BM, Murphy SK, Lessey BA, Jordan VC, Bradford AP.** 2012. DNA profiling analysis of endometrial and ovarian cell lines reveals misidentification, redundancy and contamination. Gynecol Oncol **127:**241-248. PMID: 22710073
46. **Korch C, Varella-Garcia M.** 2018. Tackling the Human Cell Line and Tissue Misidentification Problem Is Needed for Reproducible Biomedical Research. Adv Mol Pathol **1:**209-228.e236. PMID: https://doi.org/10.1016/j.yamp.2018.07.003
47. **Korch CT, Capes-Davis A.** 2021. The Extensive and Expensive Impacts of HEp-2 [HeLa], Intestine 407 [HeLa], and Other False Cell Lines in Journal Publications. SLAS Discov **26:**1268-1279. PMID: 34697958
48. **Korch CT, Hall EM, Dirks WG, Sykes GR, Capes-Davis A, Butler JM, Neve RM, Nims RW, Storts DR, Tian F, Nardone RM.** 2021. Human Cell Line Authentication. Standardization of Short Tandem Repeat (STR) Profiling. ASN-0002 Revised 2022, November 2022 ed. American National Standards Institute (ANSI) - American Type Culture Collection (ATCC) Standards Development Organization, Manassas, Virginia, United States.
49. **Langdon WB.** 2014. Mycoplasma contamination in the 1000 Genomes Project. BioData Min **7:**3. PMID: 24872843
50. **Liang-Chu MM, Yu M, Haverty PM, Koeman J, Ziegle J, Lee M, Bourgon R, Neve RM.** 2015. Human biosample authentication using the high-throughput, cost-effective SNPtrace(TM) system. PLoS One **10:**e0116218. PMID: 25714623
51. **Lichter P, Allgayer H, Bartsch H, Fusenig N, Hemminki K, von Knebel Doeberitz M, Kyewski B, Miller AB, zur Hausen H.** 2010. Obligation for cell line authentication: appeal for concerted action. Int J Cancer **126:**1. PMID: 19882693
52. **Linger RM, DeRyckere D, Brandao L, Sawczyn KK, Jacobsen KM, Liang X, Keating AK, Graham DK.** 2009. Mer receptor tyrosine kinase is a novel therapeutic target in pediatric B-cell acute lymphoblastic leukemia. Blood **114:**2678-2687. PMID: 19643988
53. **Linger RM, Graham DK.** 2014. A Common Case of Cell Line Misidentification. https://thewinnower.com/discussions/9-a-common-case-of-cell-line-misidentification. Accessed
54. **Linger RM, Lee-Sherick AB, DeRyckere D, Cohen RA, Jacobsen KM, McGranahan A, Brandao LN, Winges A, Sawczyn KK, Liang X, Keating AK, Tan AC, Earp HS, Graham DK.** 2013. Mer receptor tyrosine kinase is a therapeutic target in pre-B-cell acute lymphoblastic leukemia. Blood **122:**1599-1609. PMID: 23861246
55. **Linger RMA, DeRyckere D, Brandao L, Sawczyn KK, Jacobsen KM, Liang X, Keating AK, Graham DK.** 2012. Retraction. Linger RMA, DeRyckere D, Brandao L, Sawczyn KK, Jacobsen KM, Liang X, Keating AK, Graham DK. Mer receptor tyrosine kinase is a novel therapeutic target in pediatric B-cell acute lymphoblastic leukemia. Blood. 2009;114(13):2678-2687. Blood **120:**1533. PMID: 22753864

56. **Liu Y, Mi Y, Mueller T, Kreibich S, Williams EG, Van Drogen A, Borel C, Frank M, Germain PL, Bludau I, Mehnert M, Seifert M, Emmenlauer M, Sorg I, Bezrukov F, Bena FS, Zhou H, Dehio C, Testa G, Saez-Rodriguez J, Antonarakis SE, Hardt WD, Aebersold R.** 2019. Multi-omic measurements of heterogeneity in HeLa cells across laboratories. Nat Biotechnol **37:**314-322. PMID: 30778230
57. **Masters JR.** 2002. False cell lines: The problem and a solution. Cytotechnology **39:**69-74. PMID: 19003294
58. **Masters JR.** 2002. HeLa cells 50 years on: the good, the bad and the ugly. Nat Rev Cancer **2:**315-319. PMID: 12001993
59. **Masters JR, Thomson JA, Daly-Burns B, Reid YA, Dirks WG, Packer P, Toji LH, Ohno T, Tanabe H, Arlett CF, Kelland LR, Harrison M, Virmani A, Ward TH, Ayres KL, Debenham PG.** 2001. Short tandem repeat profiling provides an international reference standard for human cell lines. Proc Natl Acad Sci U S A **98:**8012-8017. PMID: 11416159
60. **McLaren RS, Reid Y, Storts DR.** 2013. Human cell line authentication: the critical first step in any project using human cell lines. Methods Mol Biol **963:**341-353. PMID: 23296621
61. **Morgan AP, Fu CP, Kao CY, Welsh CE, Didion JP, Yadgary L, Hyacinth L, Ferris MT, Bell TA, Miller DR, Giusti-Rodriguez P, Nonneman RJ, Cook KD, Whitmire JK, Gralinski LE, Keller M, Attie AD, Churchill GA, Petkov P, Sullivan PF, Brennan JR, McMillan L, Pardo-Manuel de Villena F.** 2015. The Mouse Universal Genotyping Array: From Substrains to Subspecies. G3 (Bethesda) **6:**263-279. PMID: 26684931
62. **Nardone RM.** 2007. Eradication of cross-contaminated cell lines: a call for action. Cell Biol Toxicol **23:**367-372. PMID: 17522957
63. **Neimark J.** 2014. Trial and Error - The Dirty Little Secret of Cancer Research. Discover Magazine**:**44-51. PMID:
64. **Neimark J.** 2015. Line of attack. Science **347:**938-940. PMID: 25722392
65. **Nerlich AG, Bachmeier BE.** 2013. Density-dependent lineage instability of MDA-MB-435 breast cancer cells. Oncol Lett **5:**1370-1374. PMID: 23599796
66. **O'Donoghue LE, Rivest JP, Duval DL.** 2011. Polymerase chain reaction-based species verification and microsatellite analysis for canine cell line validation. J Vet Diagn Invest **23:**780-785. PMID: 21908323
67. **Oliver T.** June 11, 2018 2018. Combating Misidentified and Contaminated Cell Lines, *on* Biocompare. https://www.biocompare.com/Editorial-Articles/348977-Authenticate-Combating-Misidentified-and-Contaminated-Cell-Lines/. Accessed September 23, 2022.
68. **Ono K, Satoh M, Yoshida T, Ozawa Y, Kohara A, Takeuchi M, Mizusawa H, Sawada H.** 2007. Species identification of animal cells by nested PCR targeted to mitochondrial DNA. In Vitro Cell Dev Biol Anim **43:**168-175. PMID: 17516125
69. **Parson W, Kirchebner R, Muhlmann R, Renner K, Kofler A, Schmidt S, Kofler R.** 2005. Cancer cell line identification by short tandem repeat profiling: power and limitations. FASEB J **19:**434-436. PMID: 15637111
70. **Petkov PM, Cassell MA, Sargent EE, Donnelly CJ, Robinson P, Crew V, Asquith S, Haar RV, Wiles MV.** 2004. Development of a SNP genotyping panel for genetic monitoring of the laboratory mouse. Genomics **83:**902-911. PMID: 15081119

71. **Price PJ.** 2017. Best practices for media selection for mammalian cells. In Vitro Cell Dev Biol Anim **53:**673-681. PMID: 28726187
72. **Reid Y, Storts D, Riss T, Minor L.** 2013. Authentication of Human Cell Lines by STR DNA Profiling Analysis. https://www.ncbi.nlm.nih.gov/books/NBK144066/. Accessed
73. **Reid YA.** 2017. Best practices for naming, receiving, and managing cells in culture. In Vitro Cell Dev Biol Anim **53:**761-774. PMID: 28986713
74. **Robin T, Capes-Davis A, Bairoch A.** 2019. CLASTR: the Cellosaurus STR Similarity Search Tool A Precious Help for Cell Line Authentication. Int J Cancer 10.1002/ijc.32639. PMID: 31444973
75. **Schweppe RE, Korch C.** 2018. Challenges and advances in the development of cell lines and xenografts. Adv Mol Pathol **1:**239-251. PMID:
76. **Sikora MJ, Johnson MD, Lee AV, Oesterreich S.** 2016. Endocrine response phenotypes are altered by charcoal-stripped serum variability. Endocrinology **157:**3760-3766. PMID: 27459541
77. **Smits BM, Guryev V, Zeegers D, Wedekind D, Hedrich HJ, Cuppen E.** 2005. Efficient single nucleotide polymorphism discovery in laboratory rat strains using wild rat-derived SNP candidates. BMC Genomics **6:**170. PMID: 16316463
78. **Souren NY, Fusenig NE, Heck S, Dirks WG, Capes-Davis A, Bianchini F, Plass C.** 2022. Cell line authentication: a necessity for reproducible biomedical research. Embo j **41:**e111307. PMID: 35758134
79. **Stern AM, Casadevall A, Steen RG, Fang FC.** 2014. Financial costs and personal consequences of research misconduct resulting in retracted publications. Elife **3:**e02956. PMID: 25124673
80. **Steube KG, Koelz AL, Drexler HG.** 2008. Identification and verification of rodent cell lines by polymerase chain reaction. Cytotechnology **56:**49-56. PMID: 19002841
81. **Tanabe H, Takada Y, Minegishi D, Kurematsu M, Masui T, Mizusawa H.** 1999. Cell line individualization by STR multiplex system in the cell bank found cross-contamination between ECV304 and EJ-1/T24. https://www.jstage.jst.go.jp/article/jtca1981/18/4/18_329/_article/-char/en. Accessed September, 2021.
82. **Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, Boutselakis H, Cole CG, Creatore C, Dawson E, Fish P, Harsha B, Hathaway C, Jupe SC, Kok CY, Noble K, Ponting L, Ramshaw CC, Rye CE, Speedy HE, Stefancsik R, Thompson SL, Wang S, Ward S, Campbell PJ, Forbes SA.** 2019. COSMIC: the Catalogue Of Somatic Mutations In Cancer. Nucleic Acids Res **47:**D941-d947. PMID: 30371878
83. **University of Texas MACC.** 2013. Cell Line Authentication Policy UTMDACC Institutional Policy # ACA1040. https://www.mdanderson.org/documents/core-facilities/Characterized%20Cell%20Line%20Core%20Facility/CCLC_%20Policy_ACA1044.pdf. Accessed
84. **Vaughan L, Glanzel W, Korch C, Capes-Davis A.** 2017. Widespread use of misidentified cell line KB (HeLa): Incorrect attribution and its impact revealed through mining the scientific literature. Cancer Res **77:**2784-2788. PMID: 28455420
85. **Vaught J, Korch CT.** 2021. The Continuing Saga of Cell Line Misidentification. Biopreserv Biobank **19:**357-358. PMID: 34637628
86. **World Health Organization WHO.** 2004. Laboratory Biosafety Manual, 3rd edition, *on* World Health Organization (WHO).

https://www.who.int/csr/resources/publications/biosafety/WHO_CDS_CSR_LYO_2004_11/en/. Accessed May 2019.

87. **Ye F, Chen C, Qin J, Liu J, Zheng C.** 2015. Genetic profiling reveals an alarming rate of cross-contamination among human cell lines used in China. FASEB J **29:**4268-4272. PMID: 26116706

88. **Yu M, Selvaraj SK, Liang-Chu MM, Aghajani S, Busse M, Yuan J, Lee G, Peale F, Klijn C, Bourgon R, Kaminker JS, Neve RM.** 2015. A resource for cell line authentication, annotation and quality control. Nature **520:**307-311. PMID: 25877200

Cell Line Authentication Flow Chart