Reproducibility: changing the policies and culture of cell line authentication

Leonard P Freedman[1], Mark C Gibson[1], Stephen P Ethier[2], Howard R Soule[3], Richard M Neve[4] & Yvonne A Reid[5]

Quality control of cell lines used in biomedical research is essential to ensure reproducibility. Although cell line authentication has been widely recommended for many years, misidentification, including cross-contamination, remains a serious problem. We outline a multi-stakeholder, incremental approach and policy-related recommendations to facilitate change in the culture of cell line authentication.

Advances in life science research build upon the reproducibility of previously published data and findings, yet irreproducibility in basic and preclinical biological research is a pervasive, expensive and increasingly well-recognized problem[1,2]. Also called replication, validation, verification or reanalysis[3], in simplest terms, reproducibility means that an experiment should be able to be confirmed in an independent laboratory with results that broadly support the conclusions of the original scientist. Excluding deliberate scientific misconduct[4], irreproducibility typically results from errors or flaws in one or more of the following areas of the research process: reference materials, study design, laboratory protocols, and data analysis and reporting[5,6]. Irreproducible preclinical research contributes to both delays and increased costs in drug discovery.

One common contributor to irreproducibility in the life sciences is the widespread use of misidentified (including by intraspecies and interspecies cross-contamination and simple mislabeling) or microbially contaminated cell lines isolated from various human tissues[7,8]. Cell lines have been used for decades to study basic biological mechanisms and serve as preclinical models for drug target discovery and to generate diagnostic hypotheses in multiple areas of biomedical research[9]. In these uses, accurate documentation of species, sex and tissue of origin is integral to interpretation and validity of research results. It is also important to monitor genotypic or phenotypic changes (i.e., drift) that might occur over time[10].

Correct identification of the origin of a cell line is simple. Cell line authentication can be achieved by determining the genetic signature (by profiling or fingerprinting) and comparing it with established databases to confirm identity[11]. From 2011 to 2012, an international group of scientists from multiple stakeholder groups collaborated to develop an accredited standard that describes optimal cell line authentication practices based on STR (short tandem repeat) profiling[12]. The International Cell Line Authentication Committee was formed following publication of the STR profiling standard to make cell line misidentification more visible and to promote awareness and use of authentication testing. However, there is little evidence that authentication is widely used in the life sciences. Many researchers are simply unaware of the need to establish and carefully maintain cell cultures and techniques, or they are aware but do nothing to validate their cell lines. Several journals, including *Nature*[13], now require or strongly recommend cell line authentication for studies they publish. Yet, despite these efforts, new reports of misidentified or contaminated cell lines still appear in the scientific literature[14,15].

The Global Biological Standards Institute (GBSI) formed the Cancer Cell Authentication and Standards Task Force in 2014 to (i) identify and overcome existing barriers to the use of currently available cell authentication tools and (ii) support the development, evaluation and application of policies, novel technologies, and standards for expanded cell authentication. Cell line misidentification and microbial contamination are not new problems, and although several consistent solutions involving authentication have been offered (and in some limited venues, adopted) over the years[16], such issues have proven surprisingly difficult to eradicate. This paper reflects the deliberations and reports the conclusions and recommendations of the Task Force to affect meaningful change in both the policies and culture of cancer cell line use and authentication across the entire biomedical research community. Although microbial contamination of culture systems by bacteria (particularly mycoplasma[17]), fungi, viruses and other microorganisms continues to be highly problematic and expensive[18], a detailed discussion of those issues is beyond the scope of this paper.

The problems

Prevalence of misidentified cell lines. Reports from major cell repositories and research laboratories indicate that a wide variety of cell lines submitted or tested are misidentified (**Table 1**). A widely cited review examining the prevalence of this problem from 1968 to 2007 estimated that between 18% and 36% of cell lines might be

[1]Global Biological Standards Institute, Washington, DC, USA. [2]Hollings Cancer Center, Medical University of South Carolina, Charleston, South Carolina, USA. [3]Prostate Cancer Foundation, Milken Institute, Santa Monica, California, USA. [4]Department of Discovery Oncology, Genentech Inc., South San Francisco, California, USA. [5]ATCC, Manassas, Virginia, USA.
e-mail: lfreedman@gbsi.org

Table 1 | Select reports of misidentified or cross-contaminated cell lines by major cell repositories

Cell type	Total number of lines	Number of false cell lines	Percentage of false cell lines	Ref.
Lymphoma, leukemia	550	82	15	39
Ovarian cancer	51	15	30	40
Adenoid cystic carcinoma	6	6	100	41
Thyroid cancer	40	17	43	42
Head, neck cancer	122	37	30	43
Esophageal adenocarcinoma	14	3	21	44
Total	783	160	20 (average)	

misidentified or cross-contaminated, with only a small improvement in rates over time[19]. A recent study reported that only 43% of cell lines could be uniquely identified (i.e., an unambiguous name or identifier was provided as well as a source for the line such as a vendor or repository) by their description in an evaluation of over 200 biomedical papers[20]. Furthermore, these surveys do not reflect (i) cell lines that were incorrectly identified when they were first cultured, (ii) cell lines that have been displaced (cross-contaminated) by unknown cell types, or (iii) cell lines that were mislabeled or contaminated in individual laboratories. For these reasons, the true incidence of misidentified cell lines is probably even larger.

Several organizations[21,22] promote and recommend best practices for handling biospecimens and biological resources, including cell lines and cell banking[23,24], but none is recommended by the majority of key life science stakeholders. The expanded adoption of best practices should, over time, lead to additional consensus-based standards for obtaining and handling authenticated and contaminant-free cell lines from any biological resource center (BRC), reputable cell bank or, eventually, smaller repository. If this fails to take place, it is likely that biological materials in the public domain will remain or eventually become compromised.

Use and cost of misidentified cell lines. Despite the availability of affordable commercial kits to profile cells and measure contamination, it is unknown how many laboratories authenticate or conduct quality control (QC) on their cell lines and how often. A decade-old survey[25] reported that just one-third of laboratories tested their cell lines for identity. A recent *Nature Cell Biology* editorial reported that an audit of papers with data generated using cell lines published between August and December 2013 revealed that only 19% of published papers conducted (or at least reported conducting) cell line authentication[26].

HeLa is by far the most common contributor to cross-contaminated cell lines[27]. Another poster child for misidentified cell lines is MDA-MB-435. More than 1,000 articles have been published on this cell line as a triple-negative, metastatic breast cancer cell line that grows well *in vivo*. It was later reported to have an identical STR profile to that of the melanoma cell line, M14. Since then, a controversy has raged as to their provenance and the validity of results using either cell line[28]. Similarly, over 300 studies used the Adriamycin-resistant breast adenocarcinoma cell lines MCF-7/ADR before they were found to be derived from human ovarian carcinoma cells (now redesignated NCI/ADR-RES)[29]. For perspective, given the cost of an average US National Institutes of Health–funded breast cancer research grant in 2013 ($370,000), as much as $100 million of research funding may have been spent using this misidentified cell line alone.

The need for new cell authentication techniques. If we exclude ignorance and indifference, cost and simplicity of assays appear to be the biggest roadblocks to universal use of cell authentication. The introduction of cheap, commercially available assays that can be easily implemented and interpreted in any laboratory will greatly increase the likelihood of universal adoption. Because STR profiling alone may never achieve those goals, alternative methods are needed.

Single-nucleotide polymorphisms (SNPs) are genetic variations between members of the same species. SNPs within a specific locus are conserved during evolution and can be used as a genetic test of identity. SNP typing assays have been published for forensic and cell line authentication applications[30,31]. 52-plex SNP assays appear to have the same rate of discrimination as 16-plex STR, whereas smaller panels of 24 SNPs are equivalent to 8-plex STR panels[32]. Commercial kits are available (for example, iPLEX Pro Sample ID Panel from Agena Bioscience and SNP Trace Panel from Fluidigm), but adoption as a general method for cell line authentication is not yet widespread, most likely because of the lack of a centralized, online reference database. As with STR profiling, there are pros and cons to SNP-based assays (**Table 2**); however, notwithstanding initial equipment costs, significant savings of time, cost per sample and data analysis can be achieved and may make SNPs attractive to academic and other organizations.

Table 2 | Short tandem repeat (STR) versus single-nucleotide polymorphism (SNP) profiling attributes

Attributes	STR	SNP
Level of discrimination (PowerPlex 18D)[a,b]	3.47×10^{-22}	$\sim 1.0 \times 10^{-18}$
Number of alleles per locus	Multiallelic	Biallelic
Inclusion of size markers during electrophoresis	Yes	Not applicable
Inclusion of allelic ladders during electrophoresis	Yes	Not applicable
Commercially available kits	Yes	Yes
Commercially available testing services	Yes	Yes
Public, online database(s)	Yes	Pending
Cost per sample (in lab; vendor)	$15–30; ~$150–300	$6; not available

[a]STR: http://www.promega.com/products/pm/genetic-identity/population-statistics/power-of-discrimination/. [b]SNP: http://agenabio.com/iplex-pro-sample-id-panel.

Authentication of nonhuman cell lines. Although STR and SNP can, in theory, be used for all animals, currently used markers in cell banks are all human specific[33]. Little is known about the level of misidentification among animal cell lines (interspecies contamination). STR markers have been identified for some organisms (for example, mouse[34]), but currently no commercial kit is available, whereas other methods (for example, isoenzyme analysis[35]) are

laborious and likely exceed the scope and means of a typical biomedical research laboratory. Much research is needed to apply profiling technologies or to develop alternative technologies for the identification and authentication of nonhuman cell lines.

Authentication versus characterization. Cell profiling is only a minor part of understanding the complex molecular and phenotypic properties of a cell line. Cells in culture are not a uniform, clonal population but rather comprise a heterogeneous population with various states of genomic and phenotypic instability, oxidative stress and impaired checkpoint mechanisms. As such, a cell responds to its environment and adjusts accordingly with altered mutations and growth properties[36]. Many cellular behaviors can be highly sensitive to changes in experimental and growth conditions. In this regard, most studies using cell lines do not sufficiently control for or report all variables affecting experimental outcomes. Genetic tests such as STR or SNP profiling identify only cell line origin. Fully characterizing a cell would require a detailed genomic, proteomic and phenotypic analysis, which is at present implausible and costly. Nonetheless, cell line authentication and QC (for example, mycoplasma detection) is the first step to ensure continued integrity of cell cultures and contribute to enhanced reproducibility of results.

Sharing cell lines. Sharing cell lines is endemic in research and should be strongly discouraged if not actively banned by the entire biomedical research community. Is saving a few hundred dollars and a few days of time worth finding out 6 months later that your cells are not what you thought they were? At a minimum, if a researcher receives a cell line from a colleague, not directly from a reputable cell bank or BRC, then it is imperative to conduct minimum QC to ensure provenance and absence of contamination.

Lack of detailed methodology in publications. Reporting detailed methods and QC of reagents facilitates independent reproduction of experiments, yet there is no consensus on the level of documentation required in the literature. Historically, space limitations in journals or simple oversight missed by peer review has resulted in omission of critical methodological details. For example, the unfortunate persistence of papers that state "as performed in ref. [x]" often leads to even more references or ends in an unrelated article or one that cannot be found or accessed. Fortunately, technology has enabled life science journals—including *Nature Methods*—to expand the space available for researchers to describe their methods and analyses, usually in the form of online materials or databases such as *Nature*'s Protocol Exchange.

Ultimately, research institutes, funding agencies and journals must develop, implement and enforce the use of simple reporting guidelines for demonstrating that cell lines are authenticated to help ensure the credibility, reproducibility and translatability of the data and results. Merely reporting that cell lines were authenticated (i.e., a box-checking exercise) is not sufficient. At a minimum, authors should provide the method of authentication, use of reference materials, and passaging or population doubling time (PDL) information. The existing honor system that assumes researchers have authenticated their cell lines is not working. Insisting on a certificate of authenticity or the equivalent, perhaps issued by an independent testing organization such as required by the Prostate Cancer Foundation's Cell Line Authentication Initiative, will help change the culture of cancer cell authentication—provided such policies are not prohibitively expensive and onerous to establish and implement—particularly for organizations that maintain reputable cell banking QC. Oversight of whether guidelines requiring authentication are being followed is more likely to be effective when conducted by the host institute or organization where the laboratory resides.

We note that the Nature journals recently strengthened their cell authentication policies by asking authors to (i) verify that they are not working on cells known to have been misidentified or cross-contaminated and (ii) provide additional details about the source and testing of their cell lines.

Table 3 | Technical considerations affecting cell-based assays and cell culture

Factor	Consideration	Solutions	Refs.
Medium/serum	Lot-to-lot variation; antibiotics and growth supplements; expired reagents	Testing multiple lots, multiple cell lines; question vendors; monitor and document reagent shelf life	45
Passage number	No consensus method for determining passage number; passage number not tracked; phenotypic drift	Track population doubling times; profile growth curve characteristics periodically	46,47
Cryopreservation	Cell damage or altered physiology; toxicity of cryopreservant	Strict adherence to standard operating procedures; thaw QC; monitor potential toxicities	23,48

Differences in experimental design and technique. In a cell-based assay, it is best to culture the cells under the specified (optimal) conditions to help ensure the best and most reproducible outcome. This includes the selection and handling of raw materials, including media and sera; preparation and QC of reagents; use of cells at low passage number or PDL (avoiding long-term serial passaging of cells in culture); and use of cryopreservation (**Table 3**).

Reporting and training. Detailed laboratory notes recording events such as cell growth, response to perturbations and morphological appearance are essential to quickly identify problems with cells in culture and potential population drift. This maintains a 'laboratory memory' in any institution and helps counteract loss of knowledge as personnel change. Documenting the history of a batch of cells, how they were maintained, the passage number, who handled them, and which experiments were run and when can expedite troubleshooting if experiments are not readily reproducible. Anyone handling cells should have a foundation of rigorous training in cell biology and cell culture, which ideally should begin with undergraduate classes in biological sciences and continue through and after graduate school. All too often, poor skill sets and incomplete or erroneous knowledge within a laboratory are self-perpetuating.

Cell line naming and associated metadata. Nomenclature in the life sciences (for example, naming of genes) is a longstanding problem that also applies for naming cell lines. Typically, names are assigned by the originator, and some names appear nonsensical or simply confusing. For example, COLO 699N is a lung line

and not colorectal[37]; however, these lines were named because they were derived by researchers in Colorado, USA. Other cell line names are not unique and are thus difficult to search (for example, A9). In addition, completely different terminology can be used for lines derived from a given parental cell line or from the same patient (for example, LUDU-23 and WILCL). More problematic are the lack of adherence to established syntax in published manuscripts (for example, MDA-MB-231 can be found as MDAMB231, MDA-MB231 and MDAMB-231) and blatant misspellings that can also be found in PubMed and other databases. For these reasons, defining a simple, consistent and consensus-based set of rules, as was established for gene names, would greatly improve the terminology of cell lines; stricter adherence to standardized terms by journals and researchers is also needed. In addition, associated metadata naming issues such as tissue type, pathology terms (for example, disease) and patient descriptors (for example, ethnicity) can also vary dramatically in the scientific literature in general and between cell banks, specifically.

Solutions

The task of fixing the problem of misidentified and contaminated cell lines is daunting and may seem overwhelming. The existing honor system that assumes scientists have authenticated cell lines is not working. Despite a written, consensus standard for authenticating human cell lines on the basis of STR profiling and the availability of affordable fee-for-service and commercial kits for STR, there is still not universal adoption of authentication.

It has been 10 years since Buehring *et al.*'s survey reported that only one-third of laboratories typically tested their cell lines for identity[25]. GBSI launched a survey in mid-April 2015 to improve our understanding of why cell authentication in general, and the STR standard specifically, is not broadly used in the life science research community. Understanding the barriers to cell authentication will further inform future policies to effectively address these issues.

Resolving this pervasive and stubborn problem demands a systematic approach with commitment by all stakeholders, starting with basic issues related to cell authentication and then transitioning to more advanced authentication technologies and issues associated with cell characterization. The life science research community must also commit sufficient time, resources and expertise in training and educating young scientists about both good laboratory practices and cell culturing techniques. Ultimately, the results of these changes will trickle down and possibly accelerate by explicitly recognizing and incentivizing the publication of reproducible data and results[38].

To this end, we provide the following recommendations to (i) raise awareness and help change the current culture of cancer cell line authentication; (ii) address issues associated with STR, SNP and related technologies, standards and best practices; and (iii) improve the credibility, reproducibility and translatability of life science, particularly basic and preclinical, research.

1. Change the culture: attitudes and practices of the research community should shift to embrace the importance of obtaining—and paying for—cell lines from BRCs or reputable cell banks.

2. Use standards and best practices: authentic, contaminant-free cell lines should be documented using existing authentication standards and tools such as STR profiling or through development and use of novel technologies such as SNP profiling. Because the majority of contamination and genotypic and phenotypic drift of cell lines occurs after they have been obtained, at a minimum, authentication testing should be done at the beginning of the research project and continue after three subcultures.

3. Establish dedicated funding: costs to address cell misidentification and contamination should be explicitly covered in research grants. Model policies such as those implemented by the Prostate Cancer Foundation that require annual authentication as a condition of funding should be adopted by other funders, including private foundations and particularly the US National Institutes of Health.

4. Authenticate to publish: documentation of cell authentication (for example, certificate of analysis or the equivalent, perhaps issued by an independent, certified testing organization) that occurred during the course of the research process should be required as part of journal submission. Requiring authentication only at the end of the publication process is arguably too late.

5. Commit to train: training and education are critical to raise awareness of the importance of authentication and characterization. Basic cell biology training and mentoring for graduate and postdoctoral students should be expanded to include the importance of reagent validation broadly and cell authentication specifically.

6. Invest in new technologies: expanded, targeted funding is needed for the development and detailed assessment of novel technologies for purity, cell authentication and characterization of cell lines.

ACKOWLEDGEMENTS

The authors thank the Global Biological Standards Institute staff for support during the preparation of this paper.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Begley, C.G. *Nature* **497**, 433–434 (2013).
2. Collins, F.S. & Tabak, L.A. *Nature* **505**, 612–613 (2014).
3. Gómez, O.S, Juristo, N. & Vegas, S. in *Proc. Int. Workshop Replication Empir. Softw. Eng. Res.* (RESER, 2010).
4. Stern, A.M., Casadevall, A., Steen, R.G. & Fang, F.C. *eLife* **3**, e02956 (2014).
5. Freedman, L.P. & Inglese, J. *Cancer Res.* **74**, 4024–4029 (2014).
6. Health Advances & Feinstein Kean Healthcare. The case for standards in life science research: seizing opportunities at a time of critical need. http://www.gbsi.org/publications/the-case-for-standards (Global Biological Standards Institute, Washington, DC, 2013).
7. Lorsch, J.R., Collins, F.S. & Lippincott-Schwartz, J. *Science* **346**, 1452–1453 (2014).
8. Perkel, J.M. *Biotechniques* **51**, 85–90 (2011).
9. Lichter, P. *et al. Int. J. Cancer* **126**, 1 (2010).
10. Masters, J.R. & Stacey, G.N. *Nat. Protoc.* **2**, 2276–2284 (2007).
11. Reid, Y.A. *Methods Mol. Biol.* **731**, 35–43 (2011).
12. ATCC/SDO. ASN-0002: authentication of human cell lines: standardization of STR profiling. (ATCC-Standards Development Organization (SDO), Manassas, Virginia, USA, 2012).
13. Anonymous. *Nature* **496**, 398 (2013).
14. Schweppe, R.E. *J. Clin. Endocrinol. Metab.* **98**, 956–957 (2013).
15. Kuo, S.H. *et al. Genes Chromosom. Cancer* **53**, 211–213 (2014).
16. Nims, R.W., Sykes, G., Cottrill, K., Ikonomi, P. & Elmore, E. *In Vitro Cell. Dev. Biol. Anim.* **46**, 811–819 (2010).
17. Rottem, S., Kosower, N.S. & Kornspan, J.D. in *Biomedical Tissue Culture* (eds. Ceccherini-Nelli, L. & Matteoli, B.) (Intech, 2012).
18. Olarerin-George, A.O. & Hogenesch, J.B. *Nucleic Acids Res.* **11**, 2535–2542 (2015).
19. Hughes, P., Marshall, D., Reid, Y., Parkes, H. & Gelber, C. *Biotechniques* **43**, 575–586 (2007).
20. Vasilevsky, N.A. *et al. PeerJ* **1**, e148 (2013).
21. OECD. OECD best practice guidelines for biological resource centres. http://www.oecd.org/sti/biotech/38777417.pdf (Organisation for Economic Co-operation and Development, Paris, 2007).
22. Office of Biorepositories and Biospecimen Research, NCI. NCI best practices for biospecimen resources. (National Cancer Institute, Rockville, Maryland, USA, 2011).
23. Geraghty, R.J. *et al. Br. J. Cancer* **111**, 1021–1046 (2014).

24. Wrigley, J.D. *et al. Drug Discov. Today* **19**, 1518–1529 (2014).
25. Buehring, G.C., Eby, E.A. & Eby, M.J. *In Vitro Cell. Dev. Biol. Anim.* **40**, 211–215 (2004).
26. Anonymous. *Nat. Cell Biol.* **16**, 385 (2014).
27. Kniss, D.A. & Summerfield, T.L. *Reprod. Sci.* **21**, 1015–1019 (2014).
28. Chambers, A.F. *Cancer Res.* **69**, 5292–5293 (2009).
29. Liscovitch, M. & Ravid, D. *Cancer Lett.* **245**, 350–352 (2007).
30. Demichelis, F. *et al. Nucleic Acids Res.* **36**, 2446–2456 (2008).
31. Sanchez, J.J. *et al. Electrophoresis* **27**, 1713–1724 (2006).
32. Castro, F. *et al. Int. J. Cancer* **132**, 308–314 (2013).
33. Reid, Y., Storts, D., Riss, T. & Minor, L. in *Assay Guidance Manual [Internet]* (eds. Sittampalam, G.S. *et al.*) (Eli Lilly & Company and the National Center for Advancing Translational Sciences, Bethesda, Maryland, USA, 2013).
34. Almeida, J.L., Hill, C.R. & Cole, K.D. *Cytotechnology* **66**, 133–147 (2014).
35. Nims, R.W., Shoemaker, A.P., Bauernschub, M.A., Rec, L.J. & Harbell, J.W. *In Vitro Cell. Dev. Biol. Anim.* **34**, 35–39 (1998).
36. Matsumura, T. *Adv. Exp. Med. Biol.* **129**, 31–38 (1980).
37. Semple, T.U., Quinn, L.A., Woods, L.K. & Moore, G.E. *Cancer Res.* **38**, 1345–1355 (1978).
38. Hartshorne, J.K. & Schachner, A. *Front. Comput. Neurosci.* **6**, 8 (2012).
39. Drexler, H.G., Dirks, W.G., Matsuo, Y. & MacLeod, R.A. *Leukemia* **17**, 416–426 (2003).
40. Korch, C. *et al. Gynecol. Oncol.* **127**, 241–248 (2012).
41. Phuchareon, J., Ohta, Y., Woo, J.M., Eisele, D.W. & Tetsu, O. *PLoS ONE* **4**, e6040 (2009).
42. Schweppe, R.E. *et al. J. Clin. Endocrinol. Metab.* **93**, 4331–4341 (2008).
43. Zhao, M. *et al. Clin. Cancer Res.* **17**, 7248–7264 (2011).
44. Boonstra, J.J. *et al. J. Natl. Cancer Inst.* **102**, 271–274 (2010).
45. Mathieson, W., Kirkland, S., Leonard, R. & Thomas, G.A. *J. Cell. Biochem.* **112**, 2170–2178 (2011).
46. Briske-Anderson, M.J., Finley, J.W. & Newman, S.M. *Proc. Soc. Exp. Biol. Med.* **214**, 248–257 (1997).
47. Yamashita, S. *et al. Pharm. Res.* **14**, 486–491 (1997).
48. Stacey, G.N. & Masters, J.R. *Nat. Protoc.* **3**, 1981–1989 (2008).