Cell line authentication to improve preclinical cancer research: Methods in cell line authentication, quality control, and annotation

Stephen P. Ethier and Richard M. Neve

Background

Cell lines are essential tools used in all aspects of basic and applied biomedical research. Unfortunately, a large number of cell lines are contaminated or mislabeled resulting in far-reaching consequences. In addition to damaged reputations and misleading data, a significant sum of research funds are wasted and, more importantly, delivery of essential therapeutics to patients is delayed (1–3). Despite the knowledge that use of contaminated cell lines is widespread, only a minority of scientists perform adequate quality controls on cell lines (4), and there has been an astounding lack of impetus to provide meaningful solutions to this problem. As awareness grows, practices improve, but it will require a concerted, coordinated effort from research institutions, funding agencies, and journals to insist upon cell line authentication as standard scientific practice (2). Such a fundamental change needs to come through a balance of regulation and education. Remarkably, there are simple, inexpensive solutions that can be easily implemented in any research laboratory for quality control of cell line stocks (5). In this session, we will highlight the major sources of cell line contamination and misidentification, and outline potential solutions. Discussion topics cover problems associated with cell line annotation, contamination by adventitious organisms, intra- and inter-species cell line contamination, and the pros and cons of two technologies to DNA profile biosamples.

Discussion

Cell line nomenclature

Historically, there have been no naming conventions or rules guiding associated descriptive data for biosamples. Thus, cell lines of completely different origin may have similar or identical names, and confusing or misleading descriptions of tissue of origin or tumor pathology (6). The use of incorrect spelling and capitalization is apparent in PubMed as well as in the catalogs of cell lines at cell line repositories. Recently guidelines have been drafted to aid naming of cell lines and their derivatives (7). HyperCLDB and CLIMA (6) are two resources that catalog cell lines linked to associated short-tandem repeats (STR) profiles. The Cell Line Knowledge Base (CLKB) is a resource which draws from ATCC and HyperCLDB to structure the information
into a defined ontology. These resources provide an excellent centralized knowledge base for cell line information however, redundancies are still present. To help to address this, a database of highly curated information including cell line annotation using a controlled vocabulary was recently developed for more than 3,000 cell lines (8).

Cell line authentication

STR profiling constitutes the American National Standards Institute (ANSI) standard test for authenticating cell lines as recommended by the ATCC Standards Development Organization Workgroup ASN-0002 (9, 10). Single nucleotide polymorphisms (SNP) analysis is a less-widely adopted but equally accurate method to DNA profile human biosamples (8, 11). There are a number of advantages and disadvantages to using each method of which users should be aware (12). Cell line heterogeneity, occurring through acute or chronic selection of cells in culture due to a variety of factors (12, 13), can lead to loss of heterozygosity (LOH), microsatellite instability (MSI), and aneuploidy in addition to cross-contamination that collectively can make identity calls challenging (14). Technical artifacts due to each procedure can affect results whereas incorrect typing of male cell lines as female is common (up to 45 percent of samples) due to deletion of the Y copy of amelogenin or complete loss of the Y chromosome (8, 11, 15). STR databases exist that allow comparison of cell line STRs to databases of STR profiles (16). SNP profiles for over a thousand cell lines are available from Genentech (8, 11) and the Sanger institute (17), albeit using different platforms. A reference library removing redundancy and resolving discrepancies is available for STRs and SNPs (8). Several methods to calculate identity scores for STR and SNP to match as sample to a reference exist (10, 18). These can be implemented in a number of ways, which can cause some confusion over what constitutes a “match.” Based on this published work, we provided an optimized method for comparisons (8).

Non-human cell lines

Forensic STR tests are available for humans, horses, cattle, and canines, but none that are relevant to cell culture. Although mouse STR primers are also available, a standardized STR test for murine has yet to be developed. A challenge remains that many murine cell lines are derived from a handful of inbred strains; however, a recent SNP-based analysis indicates a test could be developed (19, 20). Therefore, the chance of detecting mouse intra-species cross-contamination is low and development of a reliable test is needed. Synonymous cell lines are those that share the same DNA profile. These can be true relationships—such as lines derived directly from the same patient or subsequently from a cell line in culture (derivatives). They also occur as a result of human error by cross-contamination or misidentification. The International Cell Line Authentication Committee (ICLAC) and cell line repositories track synonymous cell lines, although reporting is scattered and sometimes inconsistent (21). Recently, a reference list of synonymous cell lines encompassing legitimate synonymous lines as well as contaminated and misidentified lines was published covering more than 1200 cell lines (8).

Cross-contamination of cell lines by adventitious organisms (fungi, mold, bacteria, viruses) or other cell lines is a common event. Careful observation and frequent testing for mycoplasma is
good laboratory practice (22). Human (intra-species) cell line contamination is frequently observed (21), yet small numbers of contaminating cells are difficult to detect (23) and can vary due to a number of factors (11). Cells placed under selective pressure (drug resistance, antibiotic selection) can frequently select for a previously undetectable, underlying contaminant population; thus, frequent DNA profiling is advised. Non-human (inter-species) cell line contamination is less well publicized but thought to affect ~6 percent of cultures (24). Moreover, current STR and SNP profiling does not detect this form of contamination. Although a number of methods are available to detect non-human contaminants of cell lines, PCR-based tests are favored for sensitivity and utility (25).

**Future Directions**

While we have a powerful toolbox of molecular tools to perform quality control on cell lines and biosamples, there are several important areas that require development so researchers can be completely certain they are using the right model system.

1. At present, there is no STR or SNP equivalent for DNA profiling non-human cell lines. This is particularly problematic when dealing with cell lines derived from mice. If one assumes that the prevalence of contamination in human cell lines occurs at a similar rate in non-human cell lines, this is a serious problem with no fast, affordable test to prove otherwise.

2. No commercial test exists for inter-species cross contamination. PCR-based tests have been published, but take some time to implement and optimize in a laboratory. Developing a higher-throughput, multiplexed assay using, for example, microfluidic technology would provide a valuable tool.

3. Cell line origin. Many cell lines have been in culture for many years or even decades. Even when a DNA profile confirms the expected ancestry, there may still be some doubt as to the actual tissue of origin of a cell line. For example, cell lines derived from distant sites of metastasis (e.g., the pleural cavity) might not derive from what was thought to be the primary tumor and human error at the time of deriving a cell line are likely causes of such an error. The potential of lineage-specific genes or long non-coding RNAs could hold promise in resolving this problem.

4. Constant vigilance. In our experience, a single assessment of cell line identity is insufficient to ensure that mistakes are not made. By following good cell culture practice, mistakes can be minimized. Nevertheless, we have found that mistakes can occur in any laboratory and certain practices such as antibiotic selection and generating drug resistant lines are associated with a higher-than-average incidence of cross-contamination. Therefore, access to an affordable high throughput DNA profiling technology is necessary to maintain the level of vigilance required to monitor cell contamination throughout a research project.

5. Emerging technologies. With the precipitous drop in the cost of sequencing and development of streamlined pipelines for analysis of sequencing data, it is very likely that a fast, sensitive, and affordable method can be developed combining biosample identity and cross-contamination in a single assay.
6. Awareness of the unknown. There are still deficits in our ability to fully assess the quality of our cell lines. For example, most PCR-based mycoplasma assays only detect the most common forms of mycoplasma—far from a comprehensive assessment of all mollicute species. Also, many laboratories do not perform a comprehensive test for presence of viruses that can both alter cells and present a potential health hazard to the researcher. Finally, the sensitivity of current DNA profiling tests to detect intra-human cross-contamination is at best 2–3 percent, meaning an underlying, as yet undetectable contaminant can easily emerge given the right conditions.

References

1. Lacroix M. Persistent use of “false” cell lines. Int J Cancer 2008;122:1–4.
2. Masters JR. False cell lines: The problem and a solution. Cytotechnology 2002;39:69–74.
3. MacLeod RA, Dirks WG, Maruao Y, Kaufmann M, Milich H, Drexler HG. Widespread intraspecies cross-contamination of human tumor cell lines arising at source. Int J Cancer 1999;83:555–63.
4. Buhring GC, Eby EA, Eby MJ. Cell line cross-contamination: How aware are Mammalian cell culturists of the problem and how to monitor it? In Vitro Cell Dev Biol Anim 2004;40:211–5.
5. Geraghty RJ, Capes-Davis A, Davis JM, Downward J, Freshney RJ, Knezevic I, et al. Guidelines for the use of cell lines in biomedical research. Br J Cancer 2014;111:1021–46.
6. Romano P, Manniello A, Aresu O, Armento M, Cesaro R, Parodi B. Cell Line Data Base: Structure and recent improvements towards molecular authentication of human cell lines. Nucleic Acids Res 2009;37:D925–32.
7. ICLAC. Naming a Cell Line. 2014; Available from: http://iclac.org/resources/cell-line-names/
8. Yu M, Selvaraj SK, Liang-Chu MMY, Aghajani S, Busse M, Yuan J, et al. A resource for cell line authentication, annotation, quality control and data integration. Nature 2015;In Press.
9. Kerrigan L, Nims RW. Authentication of human cell-based products: The role of a new consensus standard. Regen Med 2011;6:255–60.
10. Masters JR, Thomson JA, Daly-Burns B, Reid YA, Dirks WG, Packer P, et al. Short tandem repeat profiling provides an international reference standard for human cell lines. Proc Natl Acad Sci U S A 2001;98:8012–7.
11. Liang-Chu MMY, Yu M, Havery PM, Koeman J, Ziegler J, Lee M, et al. Human biosample authentication using the high-throughput, cost-effective SNP trace system. PLoS ONE 2015;In Press.
12. Parson W, Kirchelner B, Muhsman R, Renner K, Kofler A, Schmidt S, et al. Cancer cell line identification by short tandem repeat profiling: Power and limitations. Faseb J 2005;19:434–6.
13. Capes-Davis A, Reid YA, Kline MC, Stotts DR, Strauss E, Dirks WG, et al. Match criteria for human cell line authentication: Where do we draw the line? Int J Cancer 2013;132:2510–9.
14. Much M, Buza N, Hui P. Tissue identity testing of cancer by short tandem repeat polymorphism: Pitfalls of interpretation in the presence of microsatellite instability. Hum Pathol 2014;45:549–55.
15. Santos FR, Pandya A, Tyler-Smith C. Reliability of DNA-based sex tests. Nat Genet 1998;18:103.
16. Dirks WG, Drexler HG. Online verification of human cell line identity by STR DNA typing. Methods Mol Biol 2011;731:45–55.
17. Institute WTS. The Cell Lines Project. 2015; Available from: http://cancer.sanger.ac.uk/cancergenome/projects/cell_lines/about
18. Tanabe H, Takada Y, Minegishi D, Kurematsu M, Masui T, Mizusawa H. Cell line individualization by STR multiplex system in the cell bank found cross-contamination between ECV304 and EJ-1/T24. Tiss Cult Res Commun 1999;18:329–38.
19. Almeida JL, Hill CR, Cole KD. Mouse cell line authentication. Cytotechnology 2014;66:133–47.
20. Didion JP, Buus RJ, Nakhshatar Z, Threadgill DW, Morse HG, de Villena FP. SNP array profiling of mouse cell lines identifies their strains of origin and reveals cross-contamination and widespread aneuploidy. BMC Genomics 2014;15:847.
21. Capes-Davis A, Theodosopoulos G, Arkin I, Drexler HG, Kohara A, MacLeod RA, et al. Check your
cultures! A list of cross-contaminated or misidentified cell lines. Int J Cancer 2010;127:1–8.

22. Volokhov DV, Graham LJ, Bronson KA, Chizhikov VE. Mycoplasma testing of cell substrates and biologics: Review of alternative non-microbiological techniques. Mol Cell Probes 2011;25:69–77.

23. Castro F, Dirks WG, Fahnrich S, Hotz-Wagenblatt A, Pawlita M, Schmitt M. High-throughput SNP-based authentication of human cell lines. Int J Cancer 2013;132:308–14.

24. Dirks WG, Drexler HG. STR DNA typing of human cell lines: Detection of intra- and interspecies cross-contamination. Methods Mol Biol 2013;946:27–38.

25. Cooper JK, Sykes G, King S, Cottrill K, Ivanova NV, Hanner R, et al. Species identification in cell culture: A two-pronged molecular approach. In Vitro Cell Dev Biol Anim 2007;43:344–51.
