

Widespread Use of Misidentified Cell Line KB (HeLa): Incorrect Attribution and Its Impact Revealed through Mining the Scientific Literature



Liwen Vaughan¹, Wolfgang Glänzel^{2,3}, Christopher Korch⁴, and Amanda Capes-Davis⁵

Abstract

Continuous cell lines are widely used, but can result in invalid, irreproducible research data. Cell line misidentification is a common problem that can be detected by authentication testing; however, misidentified cell lines continue to be used in publications. Here we explore the impact of one misidentified cell line, KB (HeLa), on the scientific literature. We identified 574 articles between 2000 and 2014 that provided an incorrect attribution for KB, in accordance with its false identity as oral epidermoid carcinoma, but only 57 articles that provided a correct attribution for KB, as HeLa or cervical adenocarcinoma. Statistical analysis of 57 correct and 171 incorrect articles

showed that the number of citations to these articles increased over time. Content analysis of 200 citing articles showed there was a tendency to describe the cell line in accordance with the description in the cited paper. Analysis of journal impact factor showed no significant difference between correct and incorrect groups. Articles using KB or citing that usage were most frequently published in the subject areas of pharmacology, pharmacy, oncology, and medicinal chemistry. These findings are important for science policy and support the need for journals to require authentication testing as a condition of publication. *Cancer Res*; 77(11); 2784–8. ©2017 AACR.

Cell Line Misidentification Is an Important Cause of Invalid Data

Cancer cell lines, such as the cervical carcinoma cell line HeLa (1), are used as models by many laboratories to explore cancer biology and test therapeutic agents (2, 3). Cross-contamination results in misidentified or false cell lines, which no longer correspond to the original donor, but instead come from a different donor or species. This common cell culture problem affects the validity of cell lines as cancer models (4), even if a cell line is widely used, it may not be a valid model for the tumor type from which it was reportedly established. Failure to address this concern is an important cause of erroneous and misleading research data (4).

Some journals and funding bodies recognize the problem and have moved to require cell line authentication testing (5, 6). However, many others do not require authentication, allowing

continued publication of work using misidentified cell lines (4, 7). Short tandem repeat (STR) genotyping is now widely available to authenticate human cell lines, but is not yet widely performed by research laboratories. Consequently, as many as 18%–48% of cell lines are believed to be misidentified (8, 9) and this remains an important cause of invalid data in preclinical research.

Assessing the impact of cell line misidentification in the scientific literature

Review of the scientific literature has uncovered more than 400 misidentified cell lines (ref. 10; iclac.org/databases/cross-contaminations/). It can be difficult to track their usage and understand their impact over time. Cell line names are frequently short (e.g., "FL," "KB"), include common terms or descriptions (e.g., "Chang liver"), and can change over time (Supplementary Table S1; refs. 7, 11). To understand the impact of misidentification, we studied a single misidentified cell line where we could curate the reference dataset to improve its validity.

The KB cell line was established in 1955 by Harry Eagle (12), reportedly from an epidermoid carcinoma (now known as squamous cell carcinoma, SCC), from the larynx of a male donor. Eagle was also working with HeLa at that time (13). In 1966, Gartler demonstrated that KB was misidentified and corresponded to HeLa (14). Gartler's finding has been confirmed multiple times using multiple methods (15, 16). Original stocks of KB, deposited at the ATCC by Harry Eagle (www.atcc.org/Products/All/CCL-17.aspx#history), were shown to be HeLa (17). After 50 years of testing, no authentic material has been found.

Many publications describe the cell line incorrectly, even as recently as in *Cancer Research* 75th Anniversary Commentaries, where the cell line was described as coming from nasopharynx (18). Despite reviews and letters that urge a halt to its use as an oral

¹Faculty of Information and Media Studies, University of Western Ontario, London, Ontario, Canada. ²Centre for R&D Monitoring (ECCOM) and Department of MSI, KU Leuven, Leuven, Belgium. ³Department of Science Policy & Scientometrics, LHAS, Budapest, Hungary. ⁴Division of Medical Oncology, University of Colorado Anschutz Medical Campus, Aurora, Colorado. ⁵CellBank Australia, Children's Medical Research Institute, The University of Sydney, Westmead, New South Wales, Australia.

Note: Supplementary data for this article are available at Cancer Research Online (<http://cancerres.aacrjournals.org/>).

Corresponding Author: Amanda Capes-Davis, Children's Medical Research Institute, 214 Hawkesbury Road, Westmead, NSW 2145, Australia. Phone: 612-8865-2800; Fax: 612-8865-2801; E-mail: acapes-davis@cmri.org.au

doi: 10.1158/0008-5472.CAN-16-2258

©2017 American Association for Cancer Research.

cancer model, the KB cell line continues to be widely used (3, 4, 19–21). Some other publications describe the cell line correctly as cervical carcinoma or as a HeLa derivative; these may help to raise awareness of KB as a misidentified cell line.

Citation analysis can trace and assess the dissemination of information in the framework of documented scholarly communication, where it is a long-used and proven method (22–24). Erroneous ideas, incorrect results, and even fraud can be cited (25, 26) and thus have an impact on their respective research fields. Using KB as a model, we explored the impact of misidentified cell lines on the scientific literature by generating curated reference datasets to determine how many articles refer to KB correctly and incorrectly and how they are cited.

The KB cell line was used in more than 600 journal articles published in 2000–2014

We conducted two searches for the term "KB" using the PubMed database (www.ncbi.nlm.nih.gov/pubmed/) in the period 2000 to 2014. The first search examined "Correct" usage of the KB cell line, which we defined as an accurate description using the terms "cervical," "cervix," or "adenocarcinoma." The second search examined "Incorrect" usage of the KB cell line, which we defined as an inaccurate or misleading description using the terms "head and neck," "oral," "SCC," "squamous," or "epidermoid." "Correct" and "Incorrect" were defined according to how the cell line is described, because readers are guided by those descriptions when choosing it for their own work.

Search results were curated manually to confirm use of the KB cell line and exclude other uses of the term "kb." Curation was performed by examining titles and abstracts. It should be noted that titles and abstracts were typically examined without reference to the text body. Many articles refer to cell lines only in their Methods sections, so this is an important limitation to this study arising from insufficient time and journal access. References with unclear (indeterminant) descriptions were excluded from both datasets, creating a clear distinction between Correct and Incorrect groups.

This approach generated a total of 631 journal articles that used KB cells in the period 2000–2014, separated into "Correct" and "Incorrect" reference datasets for further analysis. More information on reference collection is supplied in the Supplementary Methods. Datasets used in this study are provided in full in Supplementary Tables S2 and S3.

KB cells were incorrectly described as oral or squamous cell (epithelioid) carcinoma in 574 articles and correctly described in only 57 articles

Figure 1A shows the distribution of journal articles that used the KB cell line in the period 2000–2014 and the ratio of Correct to Incorrect papers. We identified 574 articles in which the identity of the cell line KB was described incorrectly ("KB Incorrect" or simply "Incorrect" articles) and 57 articles in which the identity was described correctly ("KB Correct" or simply "Correct" articles). The overall ratio of Correct:Incorrect was 1:10. The ratio increased in the last five years (2010–2014), indicating that proportionally more correct descriptions were published in recent years. However, the pool of "Incorrect" articles continued to greatly exceed the total number of "Correct" articles.

Considering the overwhelming number of articles that incorrectly described KB as oral or squamous cell (epithelioid) carcinoma, we examined whether any were associated with a correc-

tion statement or retraction notice. We found seven corrections, all in our Incorrect dataset; none of these corrections addressed cell line information or usage (data available on request). We were unable to find any retractions within our datasets. We are aware of only one article that has been retracted for use of KB, published after our study period in 2015 (27, 28), and yet the retraction notice describes KB incorrectly as being of oral origin. This finding fits with other work showing that fewer than 20 journal articles have been retracted because of misidentified cell lines (unpublished data; ref. 29). A common cause of retraction is due to not authenticating selected drug-resistant subclones (30).

Considering that STR profiling clearly shows if a cell line corresponds to HeLa, we examined whether any articles referred to authentication testing. For this question, we used the HighWire database (highwire.stanford.edu/cgi/search; ceased operations in January 2017), which includes the body of the manuscript in its searches if available. We found only three articles that referred to authentication testing in our datasets, two in the Incorrect group and one in the Correct group (data available on request).

Articles using KB cells are increasingly cited, resulting in a greater impact when incorrect descriptions are used at greater frequency

To understand the impact of publications using KB cells, we retrieved all journal articles that cited Correct or Incorrect original articles, using the 2001 to 2015 annual volumes of the Web of Science Core collection database. Each Correct original article was matched with three Incorrect original articles, selected at random from the same publication year, giving a total of 228 articles for citation analysis (57 Correct, 171 Incorrect; Supplementary Table S4). More information on citation analysis can be found in the Supplementary Methods.

Figure 1B shows the distribution of articles citing Correct and Incorrect datasets, and the normalized ratio of citations to Correct: Incorrect groups. A total of 1,418 articles cited the Correct dataset, and 3,074 articles cited the Incorrect dataset (mean citation counts = 24.9 for Correct and 18.0 for Incorrect groups; the difference is statistically significant, $P < 0.05$). Very few articles were not cited at least once (2/57 Correct and 7/171 Incorrect; Supplementary Table S5). Citations grew considerably over time (Fig. 1B). The high proportion of Incorrect articles (Fig. 1A) would indicate that publications describing KB incorrectly are cited by a greater number of researchers over time.

Publications are cited for many reasons, not all of which relate to cell line usage. We examined the content of articles that cited Correct and Incorrect datasets (100 articles per group) in 2014–2015, to determine if KB cells were used in citing articles (data available on request). Results are summarized in Supplementary Table S6. We found that 31 of 100 articles citing a "Correct" article also used KB cells, with 20 of these describing the cells correctly. In contrast, 9 of 100 articles citing an "Incorrect" article used KB cells, with five of these describing the cells incorrectly. We found that the citing articles tended to use KB in accordance with the description in the original article, i.e. Incorrect tend to cite Incorrect and Correct tend to cite Correct (χ^2 test, $df = 3$, $P < 0.001$). Some "indeterminant" usage was also noted, with references describing KB cells in a way that avoided differences in terminology, for example "cells" or "cancer." It should be noted that this is a small sample size, due to the labor intensive nature of the analysis (reading each paper), so we must be cautious when drawing conclusions from these data.

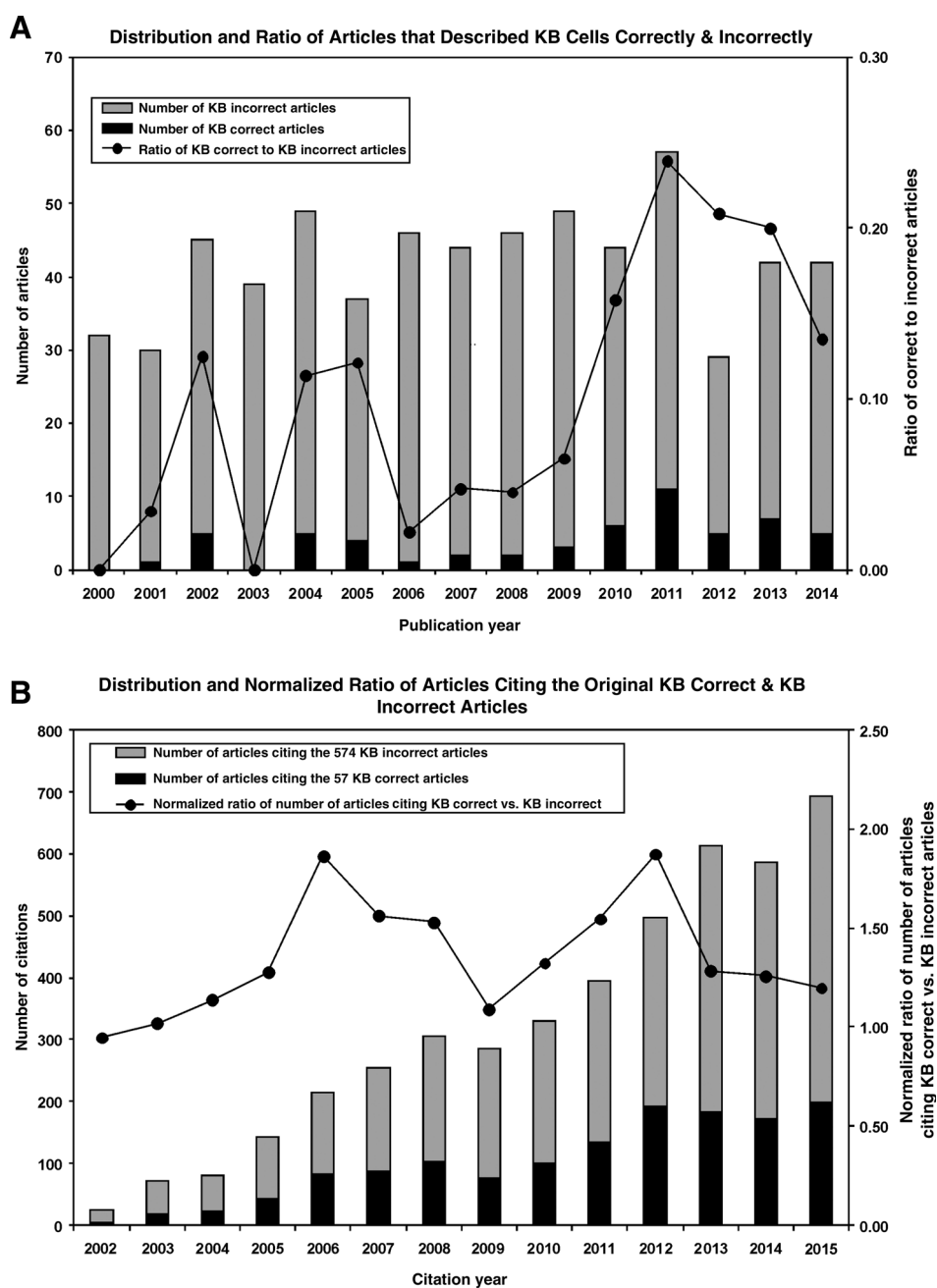


Figure 1. **A**, Distribution and ratio of articles that described KB cells correctly and incorrectly in the period 2000 to 2014. **B**, Distribution and normalized ratio of articles citing the original KB Correct and KB Incorrect articles between 2002 and 2015. (Data sourced from Thomson Reuters Web of Science Core Collection.)

We collected data on Journal Impact Factor (JIF) for all Correct and Incorrect original articles, and for all citing articles. Although JIF is not a primary measure of journal quality and should not be used to evaluate individual articles or their authors, it has proven to be an accepted and objective measure of journal prestige and performance, notably at aggregate levels (31). We found no significant differences ($P > 0.05$) between original articles in Correct and Incorrect groups or between the articles that cited them (Supplementary Table S7). This implies that, as measured by JIF, there were no significant differences in the quality of journals in which the Correct and Incorrect original articles, or their associated citing articles, were published.

Finally, we collected data on journal subject areas for all Correct and Incorrect original articles, and for all citing articles. The top three research fields for all groups were Pharmacology and Pharmacy, Medicinal Chemistry, and Oncology (Supplementary Table S8). Looking in more detail, 12% (71/574) of articles describing KB incorrectly were published in 27 journals with oral, dental, or laryngeal orientation, whereas none of the articles describing KB correctly appeared in these journals (data available on request). More than 78% of articles from both Correct and Incorrect groups were used to study the effects of drugs, as judged by their Medical Subject Headings (MeSH) classification in PubMed (data available on request). We concluded that these fields are most affected by

usage of KB cells; a great deal of research on oral cancer, SCC, oral cells, or the effects of drugs on cell responses are based on incorrect descriptions of KB. There is the potential for incorrect results in these fields, negatively affecting treatment choices for patients.

Authentication testing and accurate cell line descriptions are essential for cell lines to be correctly used as research models

Cell lines are chosen as research models for many different reasons. Cell lines may be used to model (i) a specific organism (human in the case of KB); (ii) a specific type of tissue from that organism; (iii) a specific biological process; (iv) a specific disease; or (v) the effects of drugs on specific biological processes. The identity or authenticity of the cell line is important for all these reasons, as part of ensuring that it is an appropriate model and results are reproducible.

Although the number of journals that require authentication is increasing, very few journal articles currently provide information on authentication testing. Researchers must rely on cell line information supplied by the authors to judge whether a cell line is appropriate for their own work. The tissue type or disease state may not appear to be relevant for all studies, but readers will use that information to make their own cell line choices. Cell line information must be accurate or the research community will be misled when drawing conclusions from that work. This is particularly of concern when therapeutic substances are tested in cell line models.

Analysis of KB cell line information shows a complex picture with diverse descriptions. KB cells are described as coming from many different tissues including nasopharynx, gastrointestinal tract, and kidney (Supplementary Table S1). Authors frequently fail to describe KB cells as HeLa variants, even when including correct references to sources that reported this information, or provide source information for the originating laboratory without providing literature references to clarify the nature of the cells. Incorrect attribution as "epidermoid carcinoma" rather than "adenocarcinoma" is particularly common. ATCC referred to KB cells as epidermoid carcinoma until relatively recently, making this a likely contributing factor.

Recent efforts have been made to improve the quality of cell line information and these may contribute to the increased number of papers that refer to KB correctly (Fig. 1A). Efforts include agreement on a consensus method for authentication of human cell lines (4); recommendations for standardized terminology (11); efforts by journals to encourage or require authentication (6); the establishment of the International Cell Line Authentication Committee (iclac.org); and the development of Cellosaurus, a resource for cell line information (web.expasy.org/cellosaurus/description.html). However, any improvements in citation practices will lag behind changes in publication practice and leave much incorrect information to plague the research literature.

New initiatives are needed, such as the NIH-proposed principles and guidelines for reporting preclinical research (www.nih.gov/about/reporting-preclinical-research.htm), which

would implement authentication of cell lines as a requirement for grants and journals. More than 80 journals and societies agreed to adhere to these principles. Recently, the NIH issued three notices that mandate cell line authentication as a requirement for submission of different types of grants starting in January 2016 (NOT-OD-15-103, NOT-OD-16-011, and NOT-OD-16-012). Implementation of these requirements into policy should result in detection of all KB usage in articles published from that point in time.

Although accurate reporting of cell line information is important, testing and further characterization are also required. Even where authentication testing has been reported, cell lines may fail to mimic the tissues or diseases from which they originated because of other variables. For example, lack of reproducibility in toxicity studies can be traced back to variations in cell line sources and reagents (32, 33). A cell line's validity as a research model should never be taken for granted. Evaluation and optimization of cell line choices, culture conditions, and reporting will improve our preclinical research models.

A limitation of this study should be noted: we have only studied recent literature using a single example from more than 450 known misidentified cell lines (iclac.org/databases/cross-contaminations/). The findings of this study should not be generalized to the literature of other cell lines without further validation. Hopefully, the implementation of policies requiring cell line authentication for journals and grants will improve the quality of the literature using cell lines. However, should the pool of incorrect literature for other cell lines be found to be greater than that for correct literature, the use of such false cell lines will persist and pervade the scientific literature for the foreseeable future.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

Conception and design: L. Vaughan, W. Glänzel, C. Korch, A. Capes-Davis
Development of methodology: L. Vaughan, W. Glänzel, C. Korch, A. Capes-Davis
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): C. Korch, A. Capes-Davis
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): L. Vaughan, W. Glänzel, C. Korch, A. Capes-Davis
Writing, review, and/or revision of the manuscript: L. Vaughan, W. Glänzel, C. Korch, A. Capes-Davis
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): L. Vaughan, C. Korch, A. Capes-Davis

Acknowledgments

We would like to thank Bart Thijs (ECOOM, KU Leuven) for his kind assistance in processing the bibliometric data.

Received August 22, 2016; revised September 26, 2016; accepted March 24, 2017; published OnlineFirst April 28, 2017.

References

- Gey GO, Coffman WD, Kubicek MT. Tissue culture studies of the proliferative capacity of cervical carcinoma and normal epithelium. *Cancer Res* 1952;12:264–65.
- Gillet JP, Varma S, Gottesman MM. The clinical relevance of cancer cell lines. *J Natl Cancer Inst* 2013;105:452–8.
- Masters JR. Human cancer cell lines: fact and fantasy. *Nat Rev Mol Cell Biol* 2000;1:233–6.
- American Type Culture Collection Standards Development Organization Workgroup ASN-0002. Cell line misidentification: the beginning of the end. *Nat Rev Cancer* 2010;10:441–8.
- Lorsch JR, Collins FS, Lippincott-Schwartz J. Cell Biology. Fixing problems with cell lines. *Science* 2014;346:1452–3.
- Potash J, Anderson KC. What's your line? *Clin Cancer Res* 2009;15:4251.

7. Neimark J. Line of attack. *Science* 2015;347:938–40.
8. MacLeod RA, Dirks WG, Matsuo Y, Kaufmann M, Milch H, Drexler HG. Widespread intraspecies cross-contamination of human tumor cell lines arising at source. *Int J Cancer* 1999;83:555–63.
9. Schweppe RE, Klopfer JP, Korch C, Pugazhenthii U, Benezra M, Knauf JA, et al. Deoxyribonucleic acid profiling analysis of 40 human thyroid cancer cell lines reveals cross-contamination resulting in cell line redundancy and misidentification. *J Clin Endocrinol Metab* 2008;93:4331–41.
10. Capes-Davis A, Theodosopoulos G, Atkin 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.
11. Yu M, Selvaraj SK, Liang-Chu MM, Aghajani S, Busse M, Yuan J, et al. A resource for cell line authentication, annotation and quality control. *Nature* 2015;520:307–11.
12. Eagle H. Propagation in a fluid medium of a human epidermoid carcinoma, strain KB. *Proc Soc Exp Biol Med* 1955;89:362–4.
13. Eagle H. The specific amino acid requirements of a human carcinoma cell (Strain HeLa) in tissue culture. *J Exp Med* 1955;102:37–48.
14. Gartler SM. Genetic markers as tracers in cell culture. *Natl Cancer Inst Monogr* 1967;26:167–95.
15. Nelson-Rees WA, Daniels DW, Flandermeyer RR. Cross-contamination of cells in culture. *Science* 1981;212:446–52.
16. 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.
17. Lavappa KS. Survey of ATCC stocks of human cell lines for HeLa contamination. *In Vitro* 1978;14:469–75.
18. Sporn MB. Commentary on eagle and foley: "cytotoxicity in human cell cultures". *Cancer Res* 2016;76:989–90.
19. Jiang L, Zeng X, Wang Z, Chen Q. Cell line cross-contamination: KB is not an oral squamous cell carcinoma cell line. *Eur J Oral Sci* 2009;117:90–1.
20. O'Neill ID. Continued misrepresentation of KB cells as being of oral cancer phenotype requires action. *Oral Oncol* 2009;45:e117–8.
21. Stokes A. Comment on "Oral Oncology editorial–Continued misrepresentation of KB cells as being of oral cancer phenotype requires action" [O'Neill ID. *Oral Oncol* 2009;45:e117–118]. *Oral Oncol* 2010;46:e38.
22. Westney LCH. Historical rankings of science and technology: a citationist perspective; 1998. Available from: <https://quod.lib.umich.edu/cgi/p/pod/dod-idx/historical-rankings-of-science-and-technology-a-citationist.pdf?c=jahc;idno=3310410.0001.105>.
23. Smith LC. Citation analysis. *Libr Trends* 1981;30:83–106.
24. Glänzel W, Schoepfin U. A bibliometric study of reference literature in the sciences and social sciences. *Inf Process Manag* 1999;35:31–44.
25. Campanario JM. Fraud: retracted articles are still being cited. *Nature* 2000;408:288.
26. van der Vet PE, Nijveen H. Propagation of errors in citation networks: a study involving the entire citation network of a widely cited paper published in, and later retracted from, the journal *Nature*. *Res Integr Peer Rev* 2016;1:1–10.
27. He Y, Chen F, Cai Y, Chen S. Knockdown of tumor protein D52-like 2 induces cell growth inhibition and apoptosis in oral squamous cell carcinoma. *Cell Biol Int* 2015;39:264–71.
28. He Y, Chen F, Cai Y, Chen S. Retracted: Knockdown of tumor protein D52-like 2 induces cell growth inhibition and apoptosis in oral squamous cell carcinoma. *Cell Biol Int* 2016;40:361.
29. Casadevall A, Steen RG, Fang FC. Sources of error in the retracted scientific literature. *FASEB J* 2014;28:3847–55.
30. Clement V, Marino D, Cudalbu C, Hamou MF, Mlynarik V, de Tribolet N, et al. Retraction: Marker-independent identification of glioma-initiating cells. *Nat Methods* 2013;10:1035.
31. Garfield E. The history and meaning of the journal impact factor. *JAMA* 2006;295:90–3.
32. Kassotis CD, Masse L, Kim S, Schlezinger JJ, Webster TF, Stapleton HM. Characterization of adipogenic chemicals in three different cell culture systems: implications for reproducibility based on cell source and handling. *Sci Rep* 2017;7:42104.
33. Kleensang A, Vantangoli MM, Odwin-DaCosta S, Andersen ME, Boekelheide K, Bouhifd M, et al. Genetic variability in a frozen batch of MCF-7 cells invisible in routine authentication affecting cell function. *Sci Rep* 2016;6:28994.