

# NON-VERIFIABLE CELL LINES IN CANCER RESEARCH PAPERS DESCRIBING HUMAN GENE RESEARCH



Danielle J. Oste<sup>1,2</sup>; Pranujan Pathmendra<sup>1</sup>; Gracen Johnson<sup>1</sup>; Reese A.K. Richardson<sup>3</sup>; Thomas Stoeger<sup>4,5,6</sup>; Jennifer A. Byrne<sup>1,7</sup>

<sup>1</sup>School of Medical Sciences, Faculty of Medicine and Health, The University of Sydney, New South Wales (NSW), Australia; <sup>2</sup>Sydney School of Veterinary Science, Faculty of Science, The University of Sydney, NSW, Australia; <sup>3</sup>Department of Chemical and Biological Engineering, Northwestern University, Evanston, Illinois, USA; <sup>4</sup>Feinberg School of Medicine in the Division of Pulmonary and Critical Care Medicine, Northwestern University, Chicago, Illinois, USA; <sup>5</sup>The Potosnak Longevity Institute, Northwestern University, Chicago, Illinois, USA; <sup>6</sup>Simpson Querrey Lung Institute for Translational Science, Chicago, Illinois, USA; <sup>7</sup>NSW Health Statewide Biobank, NSW Health Pathology, NSW, Australia. **Correspondence:** jennifer.byrne@health.nsw.gov.au

**Funding:** University of Sydney; National Health and Medical Research Council of Australia Ideas Grant, Grant/Award Number: APP1184263; NIH(USA) Grant/Award Number: AG068544; Moderna Inc.

## WANTED CELL LINE



**BGC803**  
143 ARTICLES  
LAST SEEN IN THE COMPANY  
OF BGC823 AND MGC803  
REWARD:  
BETTER USE OF YOUR RESEARCH \$\$\$

Reproducible laboratory research relies on correctly identified reagents. We have previously described gene research papers with wrongly identified nucleotide sequence(s), including papers studying *miR-145*. Manually verifying reagent identities in **36 recent *miR-145* papers** found **56% and 17% papers described misidentified nucleotide sequences and cell lines**, respectively.

We also found **5 cell line identifiers in *miR-145* papers** with misidentified nucleotide sequences and cell lines, and **18 cell line identifiers published elsewhere**, that **did not represent indexed human cell lines**. These 23 identifiers were described as **non-verifiable (NV)**, as their identities were unclear.

Studying 420 papers that mentioned 8 non-verifiable identifier(s) found **235 papers (56%) that referred to 7 identifiers (BGC-803, BSG-803, BSG-823, GSE-1, HGC-7901, HGC-803, MGC-823) as independent cell lines**. We could not find any publications describing how these cell lines were established. Six cell lines were sourced from cell line repositories with externally accessible online catalogues, but these cell lines **were not indexed as claimed**. Some papers also stated **STR profiles had been generated for 3 cell lines, yet no STR profiles could be identified**.

**Non-verifiable cell lines represent new challenges to research integrity and reproducibility**, further investigations are required to clarify their status and identities.

## Non-verifiable cell lines = 'Miscellings' = Not indexed in Cellosaurus, not available in claimed repositories, no establishment papers, no STR profiles

**Table 1: Original papers describing experiments where NV cell line identifier(s) were described as independent cell line(s)**

NV cell line identifier	Original papers referring to NV identifier as independent cell line <sup>a</sup>	Publication years range	Number of individual journals/publishers	Journal Impact Factor range	Most frequent country of origin proportion (%)	Most frequent institution type proportion (%)	Cell line repository sources	Papers referring to derivation of STR profiles proportion (%)
BGC-803	n=116	2006-2023	n=82/ n=34	0.2-12.7	China 115/116 (99%)	Hospital 73/116 (63%)	Named collaborator/ institute/ laboratory; ATCC; BeNa Culture Collection (Beijing, China); Cell Bank of the Chinese Academy of Sciences; Type Culture Collection of Chinese Academy of Sciences	4/116 (3%)
BSG-803	n=1	2020	n=1/ n=1	4.1	China 1/1 (100%)	Hospital 1/1 (100%)	Not stated	0/1 (0%)
BSG-823	n=14	2015-2022	n=13/ n=11	2.5-6.4	China 14/14 (100%)	Hospital 12/14 (86%)	Named collaborating institute; Cell Bank of the Chinese Academy of Sciences; National Infrastructure of Cell Line Resources of China	0/14 (0%)
GSE-1	n=34	2004-2023	n=30/ n=16	2.0-9.7	China 33/34 (97%)	Hospital 27/34 (79%)	Named collaborating institute; ATCC; BeNa Culture Collection; China Center for Type Culture Collection	2/34 (6%)
HGC-803	n=3	2016-2018	n=3/ n=3	1.7-4.3	China 3/3 (100%)	Hospital 3/3 (100%)	ATCC; Cell Bank of the Chinese Academy of Sciences	0/3 (0%)
HGC-7901	n=7	2015-2023	n=7/ n=5	2.1-8.5	China 7/7 (100%)	Hospital 4/7 (57%)	Cell Bank of Chinese Academy of Sciences	0/7 (0%)
MGC-823	n=34	2005-2023	n=30/ n=16	0.3-11.2	China 33/34 (97%)	Hospital 23/34 (68%)	Named collaborator or collaborating institute; ATCC; Cell Bank of Chinese Academy of Sciences; China Center for Type Culture Collection	5/34 (15%)

<sup>a</sup>Original papers that referred to experiments that employed NV cell line(s) (i) without reference to any similarly-named cell line and/or (ii) with a similarly-named cell line, such that NV cell line(s) were referred to as independent cell line(s)

**Table 2: Literature reviews, commentaries, book chapters, original articles, and preprints where NV cell line identifier(s) were referred to as independent cell line(s) in the absence of experimental results**

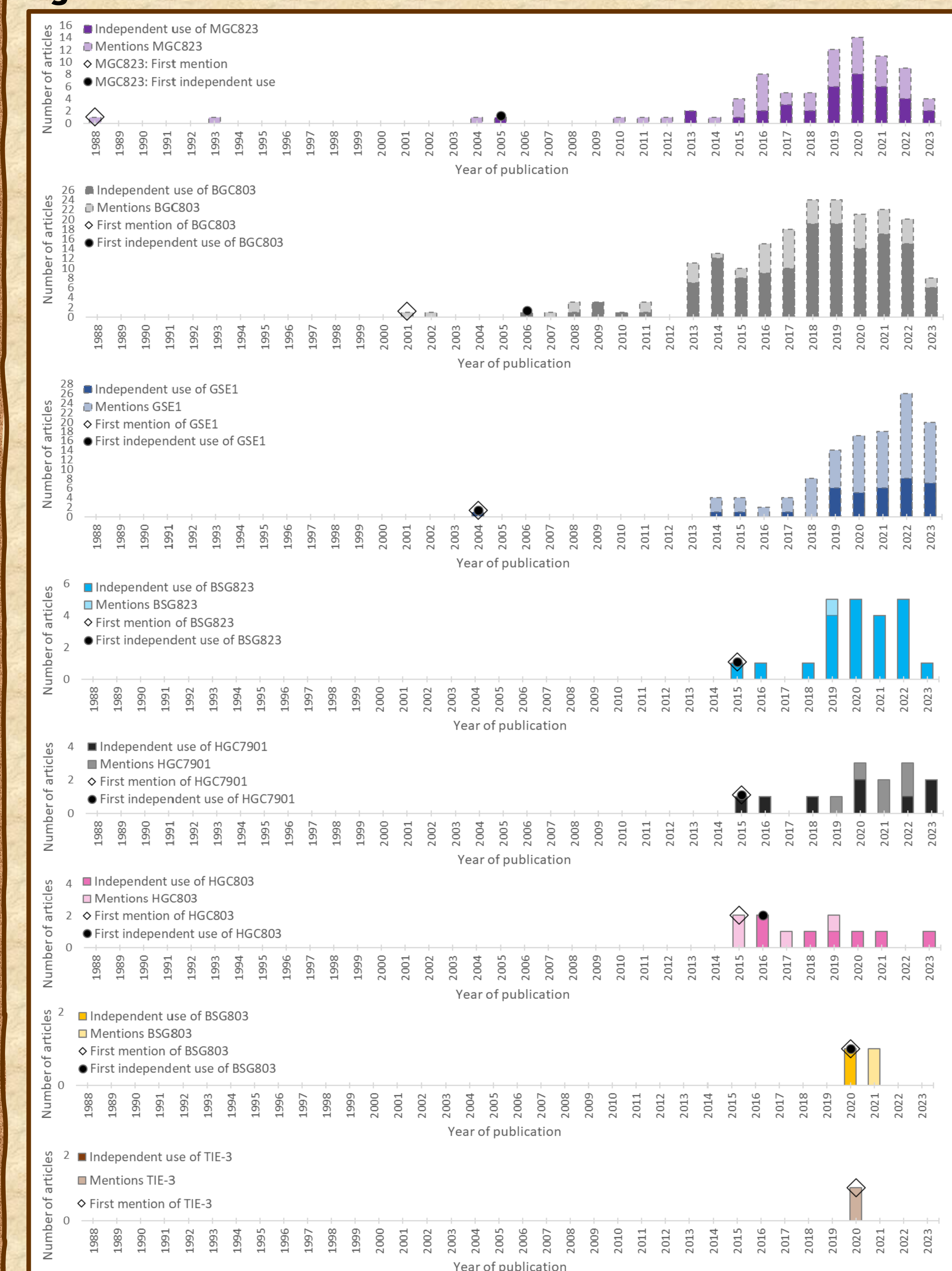
NV cell line identifier	Number of papers (literature reviews) referring to NV identifier as independent cell line	Publication years, range	Number of individual journals/publishers	Journal Impact Factor, range	Most frequent country of origin proportion (%)	Most frequent institution type proportion (%)
BGC-803	n=27 (n=22)	2009-2023	n=22/ n=8	1.8-13.6	China 11/21 (41%)	University 22/27 (81%)
BSG-823	n=8 (n=7)	2016-2023	n=7/ n=5	2.8-13.0	China 4/8 (50%)	University 5/8 (63%)
GSE-1	n=2 (n=2)	2023	n=2/ n=2	No information	N/A <sup>a</sup>	University 2/2 (100%)
HGC-7901	n=1 (n=1)	2020	n=1/ n=1	6.2	China 1/1 (100%)	Hospital 1/1 (100%)
HGC-803	n=4 (n=2)	2019-2023	n=3/ n=3	1.1-6.6	China 2/4 (50%)	N/A <sup>a</sup>
MGC-823	n=2 (n=2)	2019-2021	n=2/ n=2	7.8-16.8	N/A <sup>b</sup>	University 2/2 (100%)

<sup>a</sup>N/A in relation to country/ institution indicates no majority

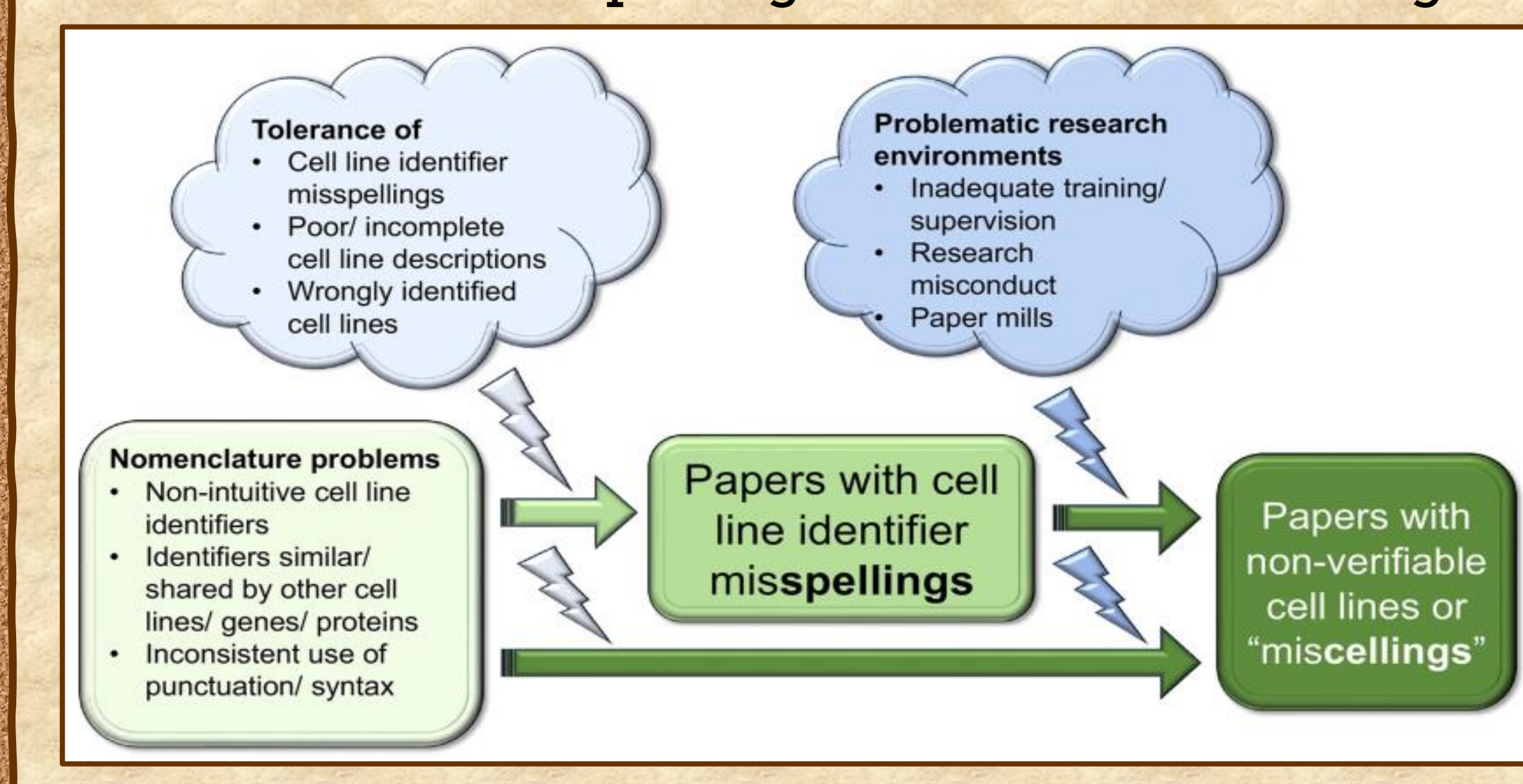
### Actions to reduce descriptions of NV cell lines:

- Establish dynamic register of misspelled cell line identifiers and NV cell lines
- Create screening tools/plug-ins to detect misspelled, NV, and wrongly identified cell lines
- Insist on use of correct cell line identifiers in research publications. Zero tolerance for misspelled cell line identifiers.
- Cell lines to be described in easily searchable publication sections, e.g. title +/- abstract
- Cell lines to be described by at least 2 identifiers, e.g. cell line identifier + RRID
- All source(s) of published cell lines to be fully disclosed
- All cell lines described in Results to also be clearly described in the Methods
- Publications describing new cell line to include description of donor origin, method of establishment, culture conditions, phenotyping, and genotyping data including STR profile
- Verification of all cell line identities prior to peer review
- Immediate publication of expressions of concern or editorial notes for papers that describe or refer to NV cell lines
- Retract original papers describing experiments with NV cell lines
- Published corrections to papers such as literature reviews that refer to NV cell lines

**Figure 1: Timeline of NV identifier use**



**Figure 2: Summary of factors that may predispose cell line identifier to misspellings and NV or 'Miscellings'**



**Acknowledgments:** Yida Ao<sup>1</sup>; Maya D. Arya<sup>1</sup>; Naomi R. Enochs<sup>1</sup>; Muhammed Hussein<sup>1</sup>; Jinghan Kang<sup>1</sup>; Aaron Lee<sup>1</sup>; Jonathan J. Danon<sup>8</sup>; Guillaume Cabanac<sup>9,10</sup>; Cyril Labbé<sup>11</sup>; Amanda Capes Davis<sup>12</sup>

<sup>1</sup>School of Medical Sciences, Faculty of Medicine and Health, The University of Sydney, NSW, Australia; <sup>8</sup>School of Chemistry, Faculty of Science, The University of Sydney, NSW, Australia; <sup>9</sup>IRIT UMR 5505 CNRS, University of Toulouse, Toulouse, France; <sup>10</sup>Institut Universitaire de France (IUF), Paris, France; <sup>11</sup>CNRS, Grenoble INP, Laboratoire d'Informatique de Grenoble, Université Grenoble Alpes, Grenoble, France; <sup>12</sup>CellBank Australia, Children's Medical Research Institute, The University of Sydney, NSW, Australia