



Naming a Cell Line, Essential Requirements

Cell lines with different names and the same identity are just as misleading as cell lines with the same name and different identities. Misidentified cell lines and name conflicts, both lead to non-reproducible data in life sciences.

If your work entails generating new cell lines, it will be necessary to give them a designation. Historically researchers have had the tendency to use very short names. This practice has led to many name "conflicts"; for example, there are 5 cell lines that have been named "CF" and 4 that have been named "TK". For a full list of recorded name conflicts see:

https://ftp.expasy.org/databases/cellosaurus/cellosaurus_name_conflicts.txt

But short names are not the only problem to avoid, thus we have developed five rules regarding cell line names:

- 1) Cell line names should be at least six characters long but not a recognizable word (like "glioma"). Characters that should be used are upper and lower case letters, Arabic numerals (0 to 9), dashes (-) and parenthesis "...". Why other characters should be avoided is explained in the appendix of this document.
- 2) For human cell lines do not use the donor's name, initials or other personal identifiers (e.g., date of birth) to avoid compromising patient confidentiality.
- 3) Check that the proposed name was not already used to describe another cell line. Do so by searching in the Cellosaurus cell line database (<https://www.cellosaurus.org>)
- 4) Check that the proposed name is not identical to the name of a gene, protein, antibody or another other biological entity. Do so by searching in Google Scholar (<https://scholar.google.com>)
- 5) Except if absolutely necessary because of a lack of space (as it often the case in figures), abbreviating the chosen cell line name should be avoided and the full name should always be used in all sections of publications (including supplementary material), presentations and also when submitting data sets in experimental repositories.

Once chosen, the cell line name should be used consistently and not changed unnecessarily.

Standardized nomenclatures

Currently the only standardized nomenclature which is widely used is the one developed for human embryonic stem cell (ESC) and induced pluripotent stem cell lines (iPSC) [1]. This nomenclature is implemented and enforced in the Human Pluripotent Stem Cell Registry (hPSCReg, <https://hpscereg.eu>) and we strongly encourage you to register any new human ESC or iPSC in this resource.

[1] Kurtz A. et al; A standard nomenclature for referencing and authentication of pluripotent stem cells. *Stem Cell Reports* 10:1-6(2018) (PubMed=29320760; DOI=10.1016/j.stemcr.2017.12.002).

Recommended style for cell line names

The recommended style for cell lines other than those registrable in hPSCReg is as follows:

- 1) Use an origin identifier relevant to the institute or laboratory in which the cell line was established. If possible, use the abbreviation already in common use by your institution. Example: MSKCC for



Memorial Sloan Kettering Cancer Center, EPFL for École Polytechnique Fédérale de Lausanne, etc.;

- 2) Use an abbreviation for the tissue, disease or species. Examples: Lu for lung, Mel for melanoma, Macmu for Macaca mulatta;
- 3) Use a number to identify the specific cell line, e.g., the number taken from the sample log completed when the tissue is received;
- 4) Another number, though preferably an alphanumeric binomial, would then define a subline or clone;
- 5) Separate the different component of the name with dashes "-"

Example: a cell line from Medical Oncology, Glasgow, isolated from lung, tissue biopsy #113 would become MOG-Lu-113, and a clone derived from it could be MOG-Lu-113-C4.

Recommended name style for cell lines where one or more genes have been knocked out

Cell_line_Name KO:Gene_name

Example: HeLa KO:TP53

- Do not forget to use the correct casing of the gene name in the species where your cell line was established. Example: BRCA1 in human, Brca1 in mouse or rat.
- When more than one gene has been knocked out, list the genes in alphabetical order. Example: HEK293 KO:FOS KO:MYC
- For cases where only one of the alleles of a gene has been silenced, we propose to use "KD" (Knock Down) instead of "KO". Example: FM3A KD:Kras

Recommended name style for cell lines selected for resistance to a toxic chemical compound

Cell_line_Name Res:Compound_name

Example: MCF-7 Res:Taxol

We encourage using the full name of the compound and not an abbreviation except when that abbreviation is widely used by the community (examples are 5-FU for 5-fluorouracil, CDDP for cisplatin and MTX for methotrexate).

Recommended name style for cell lines transfected/transduced with a construct

Cell_line_Name Added:Construct_description

Example: Caco-2 Added:eGFP-PYGM



Appendix: Characters to avoid using in cell line names

- Never use Greek letters as they will not be transliterated correctly and will often be either replaced at best by a Latin letter or by an undecipherable symbol.
- Never use the underscore as it will often be confused with the dash.
- Never use sub- or superscripts as they will either be ignored, and the two "words" will be glued together or be put between brackets.
- Avoid using punctuation signs (period, comma, semicolon, query and exclamation marks). Many computer indexing tools will recognize these characters as places where they can "break up" chunks of text and thus a name that contains these characters may be not recognized by search engines.
- Avoid using a mixture of uppercase and lowercase characters. This is not a major issue but in general most people and even cell collections will ignore the distinction and will write it according to their whims. Thus "HeLa" is often transliterated as "Hela" or "HELA".
- Avoid using the capital letter "O" as it is often confused for a zero. Similarly, be aware that lowercase "l" and "i" are often confused with each other and with "1" thus leading to errors. Example "32Dcl3" has sometimes been incorrectly termed "32Dc13".