Supplemental Information

A Standard Nomenclature for Referencing and Authentication of Pluripotent Stem Cells

Andreas Kurtz, Stefanie Seltmann, Amos Bairoch, Marie-Sophie Bittner, Kevin Bruce, Amanda Capes-Davis, Laura Clarke, Jeremy M. Crook, Laurence Daheron, Johannes Dewender, Adam Faulconbridge, Wataru Fujibuchi, Alexander Gutteridge, Derek J. Hei, Yong-Ou Kim, Jung-Hyun Kim, Anja Kolb- Kokocinski, Fritz Lekschas, Geoffrey P. Lomax, Jeanne F. Loring, Tenneille Ludwig, Nancy Mah, Tohru Matsui, Robert Müller, Helen Parkinson, Michael Sheldon, Kelly Smith, Harald Stachelscheid, Glyn Stacey, Ian Streeter, Anna Veiga, and Ren-He Xu

Supplementary information:

SI 1 (related to table 2): Nomenclature Example

	XXXXXXi001-A-1
Example	**SCI**i*001*-A-1
Explanation	**Stem Cell Institute** iPSC *first donor* – donor's first cell line – subline 1
Cell lines from different donors	SCIi001-A SCIi002-A
Cell lines sharing a donor	SCIi001-A SCIi001-B
Parental cell line and sublines (subclones)	SCIi001-A SCIi001-A-1 SCIi001-A-2

The generation of standard names for the fictional Stem Cell Institute.

Generator: The first element is chosen by the user/generator and consists of 2 to 6 capital letters (321,272,380 permutations, of which a few such as XXXXXX are unlikely to be used). Users and institutions are verified by the Registry using an automated email verification and a manual process involving a network on national representatives. The system blocks the use of acronyms already in use.

Cell type: Coded by one lowercase letter, e.g. i for iPSC.

Donor code: Each of the generator acronyms allows at least 46,655 donor codes (excluding 000) for each cell type. If necessary, this number can be increased by allowing a longer string of digits in this name element.

Clones: Using one or two letters, 702 clones or lines derived from a donor are currently possible.

Sublines (subclones): Naming a subclone / subline requires a reference parental cell line. To indicate a subline, a hyphen followed by a 1-2 digit alphanumeric code is added to the end of a name, currently allowing 1,330 possibilities. Sublines are lines with stable genetic differences from the reference line, for example transgenic lines, isogenic clones or lines that arise spontaneously due to karyotype changes. Anecdotal changes in the cell line properties, which are not thoroughly documented by experimental data, can be recorded in a free-text comment associated to the reference line. If there are substantial differences from the reference line, the user is encouraged to register a new subline/subclone of the "reference" line. For example, "PSC line has low efficiency for differentiation to mesodermal lineages compared to endoderm lineages" is an anecdotal remark if no genetic evidence is provided, while "duplication of chromosome 12 shown by karyotyping" could be encouraged as a new subline submission.

Registration and naming of a subline requires previous registration of a parental reference line. However, some times no reference line may have been registered. In this case the reference line must be registered first to generate a name, or, in exceptional cases when the identity of the reference line can not be established, a placeholder name will be generated manually and a explanatory comment linked. The associated data could be registered retrospectively.

To sum up, in theory more than 87 billion cell lines per generator and over 28×10^{18} names could be generated altogether. More information is available at (https://hpscreg.eu/about/naming-tool).

SI 2: Semantic URLs

The semantic URLs in hPSCreg follow the schema: <ORGANIZATION> /<CATEGORY> /<IDENTIFIER>. Hereby, the organization and category will be https://hpscreg.eu and cell-line respectively, while the identifier is the standardized name. For example, for finding qualitative meta information about WAe009-A (hPSCreg-derived standard name; synonymous with WA09 or H9), the semantic URL provided by hPSCreg is https://hpscreg.eu/cell-line/WAe009-A.