

## **Cellosaurus newsletter 8 of March 2020**

### **1) What changed from release 29 to release 34**

Since our last newsletter in April 2019, we have deployed five new Cellosaurus releases; the last one is release 34 of March 12, 2020 which contains information on 118,785 cell lines from 671 species. During this 13 months period we added about 7,800 new entries, but also updated the information contained in more than half of the existing entries. We continued our effort to add sequence variation information and we have now more than 12'000 cell lines that contain at least one somatic or genomic mutation description. As we announced in the last newsletter we have introduced (in release 30) a new CC line to store information on "Genome ancestry" as provided by the group of Julie Dutil tool (<http://ecla.moffitt.org/>). In release 34 we added another CC line for "karyotypic information."

In term of STR profile information, there were three noteworthy developments: 1) an increase of about 800 entries with STR profiles thus reaching a total of almost 7'000 cell lines with such information, 2) the introduction of STR profiles for mouse cell lines using the markers defined by the Consortium for Mouse Cell Line Authentication (DOI: [10.1371/journal.pone.0218412](https://doi.org/10.1371/journal.pone.0218412)) and 3) we introduced a new version of the CLASTR tool that can run STR similarity searches on dog and mouse cell lines.

In an important development, we started, with release 34, to map cell lines that are originating from patients suffering from a rare disease or that have been engineered to contain a causal mutation for a rare disease to terms from the Orphanet Rare Disease Ontology (ORDO). We already have 29'574 cell lines mapped to 1'030 ORDO terms and we expect to complete this mapping effort in the next release (release 35). It should be noted that the mapping to ORDO is complementing rather than replacing our mapping of diseases to the NCI thesaurus (NCIt). For the Cellosaurus the advantages of the NCIt ontology compared to any other medical terminologies is that it is inclusive of species other than human, is quite precise in terms of the classifications of cancers and that NCI is highly responsive to our requests in term of the creation of new terms. In the context of this fruitful collaboration, it is interesting to note that the NCIt has a concept of value set, ie a subset of the terminology used by a specific resource and has recently created such a value set for the Cellosaurus: ([https://ncithesaurus.nci.nih.gov/ncitbrowser/ajax?action=create\\_src\\_vs\\_tree&vsd\\_uri=http://evs.nci.nih.gov/valueset/Cellosaurus/C165258](https://ncithesaurus.nci.nih.gov/ncitbrowser/ajax?action=create_src_vs_tree&vsd_uri=http://evs.nci.nih.gov/valueset/Cellosaurus/C165258)). Currently there are 54'666 Cellosaurus entries that are mapped to 1'880 NCIt terms.

The Cellosaurus is now cross-referenced to 89 distinct resources as we have added cross-references to two more cell line collections, namely the China Center for Type Culture (CCTCC) and the Iranian Biological Research Center (IBRC) as well as to four additional resources: the Cancer Dependency Map (DepMap), the Fetal Calf Serum-Free Database (FCS-Free), the Cancer cell Lines GENE fusions portal (LiGeA) and the PharmacoDB integrative pharmacogenomic database.

### **2) Change in the distribution license of the Cellosaurus**

We changed the license of the Cellosaurus from the Creative Commons Attribution-NoDerivatives 4.0 International (CC BY-ND 4.0) license to the Attribution 4.0 International (CC BY 4.0) license which is conformant with the "Open Definition" for content and data from the Open Knowledge Foundation.

### **3) The Cellosaurus on ExPASy**

The traffic toward the Cellosaurus on ExPASy is continuing to increase. Since it was made available in May 2015, it has been visited 1.7 million times by almost 1 million distinct users that have browsed 5.4 million pages.

### **4) New publications**

We published a paper describing the CLASTR tool:

Robin T., Capes-Davis A., Bairoch A.; CLASTR: the Cellosaurus STR similarity search tool -- A precious help for cell line authentication; Int. J. Cancer 146:1299-1306(2020) (DOI: [10.1002/ijc.32639](https://doi.org/10.1002/ijc.32639))

As well as a small mini-review on cell lines that have flown in outer space:

Bairoch A.; Cellosaurus micro-review 1: cellonauts, spacefaring cell lines; OSF Preprints, 29 June 2019 (DOI: [10.31219/osf.io/e5fgj](https://doi.org/10.31219/osf.io/e5fgj))

PS: Subscribe to our Twitter page (<https://twitter.com/Cellosaurus>) for tweets about new developments regarding the Cellosaurus and the universe of cell lines.