

## Cellosaurus newsletter of July 2016

Last week release 18 of the Cellosaurus was made available on ExpASY (<http://web.expasy.org/cellosaurus/>). It contains information on more than 70'000 cell lines from 557 species. Since the first issue of this newsletter in March this year, we have provided two releases of the Cellosaurus and introduced many new developments in terms of the scope, depth and use of the resource as summarized in this newsletter.

As announced in March we added STR (short tandem repeat) DNA profiles of human cell lines in the Cellosaurus. After an intensive search for sources of such data we managed to get hold of almost 8'000 distinct STR profiles relevant to 4'630 cell lines. To the best of our knowledge this is the biggest compendium of STR profiles. We have also asked a number of investigators and organizations to provide us with additional STR profiles that have not yet been made available to the community.

Another recent key development concerns the participation of the Cellosaurus in the Resource Identification Initiative (RII) (<https://www.force11.org/group/resource-identification-initiative>). The RII aims to promote research resource identification, discovery, and reuse. To do so it introduced the concept of Research Resource Identifiers (RRIDs) - persistent and unique identifiers for referencing biological resources such as antibodies, cell lines, or tools in the biomedical literature and elsewhere. To help researchers to search for relevant RRIDs, the initiative has put in place a Resource Identification Portal (<https://scicrunch.org/resources>). The Cellosaurus is integrated in the portal and the first papers referring to cell line RRIDs have started to appear.

In term of external links, we have added cross-references to 4 cell lines collections: Cell Line Services (CLS), the European Bank for induced pluripotent Stem Cells (EBiSC), the Kunming Cell Bank of Type Culture Collection (KCB) and the Tick Cell Biobank (TCB). We also added cross-references to the Beta Cell Genomics Ontology (BCGO) and have started introducing cross-references to two additional experimental repositories: the Gene Expression Omnibus (GEO) and the PRoteomics IDentifications archive (PRIDE). Currently the Cellosaurus cross-references 56 different resources. Reciprocally, the Cellosaurus is now cross-referenced from a number of resources, namely, ChEMBL, ENCODE, Ximbio and the Human Protein Atlas cell line page (<http://www.proteinatlas.org/learn/cellines>). In addition to these resources "LinkOut" from NCBI PubMed and BioSample records link to the Cellosaurus when cited.

On the ExpASY server we introduced two options that respectively allow to browse cell lines by groups (examples: Insect cell lines or Haploid karyotype cell lines) or by panels (example: NCI60 cancer cell line panel).

Our "to-do" list for the coming months is quite busy and includes the development of an XML version of the Cellosaurus to facilitate the parsing and integration of the data in external resources as well as the progressive introduction of information on the tissue/organ origin of the cell lines.

As stated in our previous newsletter we wish to thank individual researchers and representatives of cell line collections, organizations and companies for their support and for answering our numerous email queries. It is thanks to the involvement of the research community that we can go forward in providing a resource as comprehensive as possible and with the highest quality standard that are the hallmark of the Swiss Institute of Bioinformatics resources (UniProtKB/Swiss-Prot, neXtProt, etc.).