Towards a semantic wiki for human and animal cell lines

Paolo DM Romano[1]✉, **Dan M Bolser**[2]

[1]IRCCS AOU San Martino IST, Genova, Italy
[2]European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, United Kingdom

Abstract

Wiki systems have emerged as a social network tool able to stimulate users to contribute to the collaborative building of a common knowledge base. Some of the specific aims of wikis for biology (bio-wikis) include collaborative efforts for the development and sharing of knowledge, the annotation of database contents, and the creation of database contents. Indeed, the increasing volumes of biological data cannot be adequately managed by the existing centralised databases and database curators. On the contrary, information and data must be exchanged by the whole community. Special features may be required to adapt to the specificity of biological data. Moreover, integration with the most relevant biological databases must be implemented. In this context, Semantic technologies promise to be of paramount relevance. In this abstract, we present a preliminary analysis for the implementation of a semantic wiki system for human and animal cell lines derived from the Cell Line Data Base (CLDB) and tightly connected to external relevant resources.

Motivation and Objectives

It is well known that human cell lines constitute one of the most useful biological resources for biomedical research since they represent an optimal model for many assays. For their best exploitation, it is essential that cell lines are properly stored, characterised, maintained, distributed and used. Biological Resources Centres offer an adequate infrastructure for these aims. More and more, their resources and services are promoted through the Internet, by means of web sites, databanks and catalogues. However, knowledge concerning cell lines is not restricted to Biological Resources Centers, and a great wealth of knowledge, information, and data, is owned by researchers throughout the world: for example, those developing cell lines, investigating the properties of cell lines, or using cell lines in their own experiments. Not to mention the knowledge and facts relevant to cell lines that are to be found in the scientific literature. This additional information may be highly relevant and interesting for researchers, stimulating the search and investigation of known characteristics of resources before their selection, the comparison of different resource features, the analysis of previous behaviours and responses from cell lines, and, finally, the selection of the most adequate and effective research tool. To this end, a knowledge base, able to store and organise this collective information could be a valid contribution.

Wiki systems have recently emerged as a network tool able to stimulate users to contribute to the collaborative building of a common knowledge base. Notorious examples exist, e.g., [Wikipedia](http://wwvw.wikipedia.org/)[1], that demonstrates this concrete opportunity. In the Life Sciences, it has already been demonstrated that wiki systems offer a variety of advantages for the management of biological data and information. These include, just to mention a notable few, [Gene Wiki](http://en.wikipedia.org/wiki/Gene_Wiki)[2] (Huss *et al.*, 2008; Huss *et al.*, 2010; Good *et al.*, 2012), a [specialised section of Wikipedia](http://en.wikipedia.org/wiki/Portal:Gene_Wiki)[3] aimed at re-organising, extending, and completing its articles related to human genes, [WikiGenes](http://www.wikigenes.org/)[4] (Hoffmann, 2008), a wiki system whose main goal is to encourage the collaborative creation of scientific papers by associating every single text to its author, and [WikiPathways](http://www.wikipathways.org/)[5] (Pico *et al.*, 2008; Kelder *et al.*, 2012), a wiki system aimed at complementing the existing databases of metabolic pathways (KEGG, Reactome, Pathway Commons). A wiki based database of biological databases was also implemented (Bolser *et al.*, 2011).

Some of the specific aims of wikis for biology (bio-wikis) include collaborative efforts for the development and sharing of knowledge, the annotation of database contents, and the creation of database contents. These aims stem from the realisation that the increasing volumes of biological data cannot be adequately managed by the existing centralised databases and database curators. Information and data must be exchanged by the whole community. In the

---

1   http://wwvw.wikipedia.org/
2   http://en.wikipedia.org/wiki/Gene__Wiki
3   http://en.wikipedia.org/wiki/Portal:Gene__Wiki
4   http://www.wikigenes.org/
5   http://www.wikipathways.org/

future, many collaborative wiki systems can be envisaged.

However several important issues still need to be addressed. For example, i) how reliable are user contributions? ii) what format should annotations take? iii) how can user provided information be feed back into 'authoritative' databases? Special features may be required to cater for the specificity of biological data: textual information is only a small part of biological data, we must cater for the numerous and heterogeneous biological data types, for example, images, plots, and diagrams.

In this abstract, we present some considerations and a preliminary analysis of a possible implementation of a wiki system for human and animal cell lines tightly connected to the Cell Line Data Base[6] (CLDB).

Methods

CLDB (Romano, 2009) and its hypertextual version HyperCLDB[7] are a long established service, offering information on human and animal cell lines since 1990. Included in CLDB, are data on

availability of cell lines in some of the most well known European collections and in many Italian research laboratories, together with their main biological characterisation.

We are developing a wiki system (CellLinesWiki) as a collaborative knowledge base for human and animal cell lines for the community of Biological Resources Centers, biobanks, collections and researchers active in this area. CellLinesWiki consists of three layers. Database information is automatically uploaded from CLDB and constitutes the first layer of authoritative information. The second layer is composed of a curated set of contributions about the cell lines described in the database. It is maintained by a limited number of nominated experts in the field, and the creators of cell lines. The third layer is built from end users, authenticated, but not necessarily trusted at the same level, authorised to provide less specific information on cell lines.

MediaWiki[8], an open source package for wiki development written in PHP, was chosen as the starting point for the development of CellLinesWiki due to its wide user and implementation base.

Figure 1. Display of data entry 10429. Protein attributes, features and sequence with transmembrane section (382-40v2) are shown.

---

6   http://bioinformatics.istge.it/cldb/cldb.php
7   http://bioinformatics.istge.it/hypercldb/

8   http://www.mediawiki.org/wiki/MediaWiki

As we required to store structured data, its extension Semantic MediaWiki[9] (SMW) was also used. Together these tools provide for the collaborative authorship of structure data within the wiki system. Using various wiki extensions we designed a system whereby the data in CLDB can be browsed, queried and updated by known users at the three different levels described.

Results and Discussion

The CellLinesWiki currently contains already information for the 6,632 cell lines, housed within 82 laboratories.

Each cell line may carry information about its associated literature, quality control information, purchasing information and treatments, as well as the associated biological meta-data about the cell lines origin, and transformation details.

Each section of the wiki can be edited by the administrators, while recognised researchers can add further details to their lines of interest. Finally, general users can comment on the information found. Information coming from these three sources is clearly delineated.

In Figure 1, the form that allows granted users to update information on a given cell line is shown. Four distinct tabs include information on data subsets. A contextual help allow users to select one item value from a list as they key some preliminary characters (in the figure, having typed "adeno" four possible values are presented).

As the wiki is in the early stages of development, so far it has only been used internally. The preliminary version of the CellLinesWiki will soon be made publicly available on-line. By promoting the wiki, we hope to engage the community in contribution.

References

Bolser DM, Chibon P-Y, *et al.* (2012) MetaBase - The wiki-database of biological databases. *Nucl. Acids Res.* **40**(Database issue), D1250-D1254. doi:10.1093/nar/gkr1099

Good BM, Clarke EL, *et al.* (2012) The Gene Wiki in 2011: community intelligence applied to human gene annotation. Nucleic Acids Res. 40(Database issue), D1255-1261. doi:10.1093/nar/gkr925

Hoffmann R. (2008) A wiki for the life sciences where authorship matters. *Nat Genet* **40**, 1047-1051. doi:10.1038/ng.f.217

Huss JW, Lindenbaum P, *et al.* (2010) The Gene Wiki: community intelligence applied to human gene annotation. *Nucleic Acids Res.* **38**(Database issue), D633-639. doi:10.1093/nar/gkp760

Huss JW III, Orozco C, *et al.* (2008) A Gene Wiki for Community Annotation of Gene Function. *PLoS Biology* **6**(7), e175. doi:10.1371/journal.pbio.0060175

Kelder T, van Iersel MP, *et al.* (2012) WikiPathways: building research communities on biological pathways. *Nucleic Acids Res.* **40**(Database issue), D1301-1307. doi:10.1093/nar/gkr1074

Pico AR, Kelder T, *et al.* (2008) WikiPathways: Pathway Editing for the People. *PLoS Biol* **6**(7), e184. doi:10.1371/journal.pbio.0060184

Romano P, Manniello A, *et al.* (2009) Cell Line Data Base: structure and recent improvements towards molecular authentication of human cell lines. *Nucleic Acids Res.* **37**(Database issue), D925-D932. doi:10.1093/nar/gkn730

---

9   http://semantic-mediawiki.org/wiki/Semantic__MediaWiki