

## Databases and ontologies

# Semantic Body Browser: graphical exploration of an organism and spatially resolved expression data visualization

Fritz Lekschas<sup>1,\*</sup>, Harald Stachelscheid<sup>1</sup>, Stefanie Seltsmann<sup>1</sup> and Andreas Kurtz<sup>1,2,\*</sup>

<sup>1</sup>Berlin-Brandenburg Center for Regenerative Therapies, Charité–Universitätsmedizin Berlin, 13353 Berlin, Germany and <sup>2</sup>Seoul National University, College of Veterinary Medicine and Research Institute for Veterinary Science, Seoul 151-742, Republic of Korea

\*To whom correspondence should be addressed.

Associate Editor: Janet Kelso

Received on 16 May 2014; revised on 20 October 2014; accepted on 21 October 2014

## Abstract

**Summary:** Advancing technologies generate large amounts of molecular and phenotypic data on cells, tissues and organisms, leading to an ever-growing detail and complexity while information retrieval and analysis becomes increasingly time-consuming. The Semantic Body Browser is a web application for intuitively exploring the body of an organism from the organ to the subcellular level and visualising expression profiles by means of semantically annotated anatomical illustrations. It is used to comprehend biological and medical data related to the different body structures while relying on the strong pattern recognition capabilities of human users.

**Availability and implementation:** The Semantic Body Browser is a JavaScript web application that is freely available at <http://sbb.cellfinder.org>. The source code is provided on <https://github.com/flekschas/sbb>.

**Contact:** [sbb@cellfinder.org](mailto:sbb@cellfinder.org)

## 1 Introduction

Technological innovations of the last decades have revolutionized current bio-medical research. High-throughput sequence data is widely adopted and repositories like Gene Expression Omnibus (Barrett *et al.*, 2013) or ArrayExpress (Rustici *et al.*, 2013) provide access to large amounts of expression data. Imaging and functional analysis tools improve phenotypic cell and tissue characterization, while text mining facilitates big-data applications with increasing importance in current research. Thus, scientists and clinicians are faced with an ever-growing detail and complexity, resulting in an increasing amount of time and resources that are needed to retrieve and process the right information, which constitutes one of the major obstacles for their efficient use.

Most data resources provide text-based information retrieval only, meaning that the user searches for keywords of a desired target. While this is sufficient when the target is known by name, it often fails or is time-consuming when the exact target name is unknown. Similarly, large tables of numbers are suitable for computational analysis, but require expert knowledge and time to be analysed manually. While graphs or diagrams can greatly summarize numbers, concepts or other data, their representation remains abstract and does not close the gap between complexity of data and intuitiveness of access. In biomedical research reality, it is highly desirable to have quick and intuitive access to a broad range of information and data types, which cannot easily comprehend by conventional means.

The Semantic Body Browser (SBB) exploits the visual pattern recognition capability of human users to provide this access. The SBB uses the anatomy of an organism's body itself as starting point to browse for information associated with organs, tissues, cells or cellular structures. Besides very fundamental computer skills no pre-knowledge is required, to use the SBB. The representation of data is facilitated by means of interactive, annotated anatomical illustrations through a user-friendly web application for fast access to up-to-date computationally derived data. An organism is explored along the four dimensions: resolution/location (gross body to subcellular level), developmental stage (e.g. Carnegie stages), species (e.g. human and mouse) and gender (male and female).

## 2 Description and results

Little to no computational or terminological background knowledge is required to operate the SBB. Retrieving information is facilitated by a mouse click within the region of interest, e.g. an organ, anatomical structure or cell (Fig. 1A, B). Information annotated to the selected entity is then displayed together with further browsing options and links to their related CellFinder (Stachelscheid *et al.*, 2014) entry for further in-depth information. Moreover, a text-based search for biological entities is available.

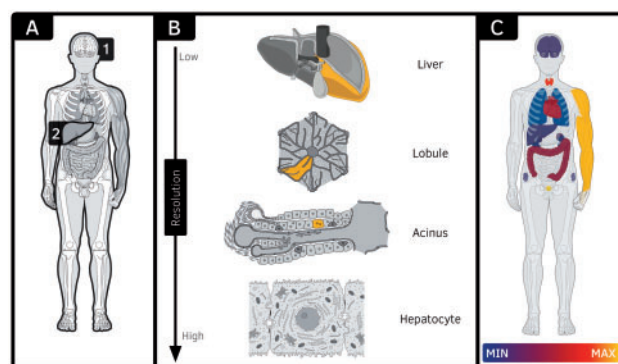
Each biological entity can feature a definition, synonyms and microscopic images as points of reference. To visualize expression profiles associated with an entity, a list of genes can be defined manually using a text-based search for gene symbols. The interactive expression heat maps are displayed within the illustrations to visualize spatial gene expression patterns of different biological entities (Fig. 1C). The generated heat maps and illustrations can be exported as Scalable Vector Graphics (SVG) in modern browsers.

Illustrations were produced by professional biomedical illustrators and validated by experts in anatomy, pathology and cytology. The four dimensions of exploration (resolution, developmental stage, species and gender) are each visualized through different sets of illustrations. All biological entities (e.g. organs, anatomical structures, cells or subcellular components) are annotated using uniform resource identifiers (URI) provided by the Cell: Expression, Localization, Development, Anatomy (CELDA) ontology (Seltmann *et al.*, 2013). Currently 22 illustrations for human and 21 for mouse, featuring three organs (kidney, liver and gall bladder) are implemented, providing 12 levels of resolution (Fig. 1B), 6 developmental stages, 2 species as well as separated views of the male and female human body. The illustrations currently comprise 674 (333 unique) biological entities, validated by experts in liver and kidney biology. In addition to anatomical illustrations, the SBB features high quality microscopic pictures, linked from CellFinder.

## 3 Implementation

The SBB is implemented as a JavaScript web application using the open source framework AngularJS created by Google as the application's backbone. A RESTful API, based on the PHP Slim framework, provides access to the data, which is stored using a MySQL server.

Illustrations are displayed as SVG to provide dynamic interactions as well as semantic annotations. Annotations follow the Resource Description Framework in Attributes (RDFa) 1.1 standard as recommended by the W3C (<http://www.w3.org/TR/rdfa-core/>). Hereby, each illustration is annotated using the `about` attribute. The containing biological entities are semantically integrated using the 'property' and 'resource' attributes. We use the 'has\_part' relation of



**Fig. 1.** (A) The human male body gross view is annotated as a whole (1) by a URI using the RDFa about attribute within the SVG tag. Specific anatomical entities within the illustration, like the liver (2), are precisely selectable, editable and semantically annotated. Thus, the application is aware of the entities contained within an illustration and knows their meaning. (B) Stepwise increased resolution of the yellow areas from the liver gross view down to the subcellular view of the hepatocyte. (C) Heat map of spatially resolved expression profiles is displayed by translating the relative accumulated expression value of a gene selection into a colour ranging from bright yellow (maximal expression value of gene selection) over red to dark blue (minimal expression value of gene selection)

the relations ontology (Smith *et al.*, 2005) as the describing property (Fig. 1A).

The heat map visualization of spatial gene expression profiles currently uses the RNA Seq Atlas (Krupp *et al.*, 2012) and Human BodyMap 2.0, Ensembl release 74 (Flicek *et al.*, 2014) datasets, featuring 11, respectively, 16 healthy human tissues. As shown by Li *et al.* (2010) and Wagner *et al.* (2012) reads per kilo base per million (RPKM) values may be biased in cross-sample comparisons, which is why we provide transcripts per million. Our protocol for assessing count-based expression data is mainly guided by Anders *et al.* (2013) using TopHat (Trapnell *et al.*, 2009), Bowtie2 (Langmead and Salzberg, 2012), Samtools (Li *et al.*, 2009) and HTSeq (Anders *et al.*, 2015).

The SBB has been tested extensively by computer scientists, biological and medical researchers over more than 1 year to ensure quality and compatibility with all modern browsers (Google Chrome  $\geq 19$ , Firefox  $\geq 4$ , Safari  $\geq 4.0.5$ , Opera  $\geq 12.15$ , Internet Explorer  $\geq 9$ ). It is integrated into CellFinder (<http://cellfinder.org/browse>) and also available as a stand-alone web application (<http://sbb.cellfinder.org>). The web application is licensed under GNU GPL 3.0. Unless otherwise stated content is licensed under Creative Commons BY-SA 4.0.

## 4 Discussion

To our knowledge, the SBB is the first web-based tool for browsing and searching large sets of diverse biological information guided by interactive anatomical illustrations that integrate different levels of resolution, developmental stages, gender and species. The two-dimensional (2D) representation of an organism's anatomy using vector graphics facilitates high accuracy, dynamic interactions and element-wise ontological annotations while maintaining simplicity and usability across a wide range of devices. Existing 3D representations like BodyParts3D (Mitsuhashi *et al.*, 2009), Zygote Body (<http://zygotebody.com>) or the Worm Browser (<http://browser.openworm.org/>), conversely, emphasize on graphical representation of the organism's gross anatomy rather than on exploration across many different dimensions.

The intuitive means of finding and displaying spatially resolved gene expression data makes the SBB especially useful for scientists and physicians with less computational background as a quick access point to biological data. The SBB could potentially be adopted in various data repositories or medical applications that benefit from an interactive visual representation along an organism's anatomy. While visualising ontologies in their full spectra, especially in an automated fashion, remains challenging (Carpendale et al., 2014), the SBB is primarily a hand-curated extension to current biomedical-related information retrieval systems using the natural connection between anatomy ontologies and anatomical illustrations. Conceptually, it is an ontologically annotated anatomy visualization.

## 5 Conclusion

To summarize, the SBB is an intuitive, user-friendly web application for graphically browsing an organism's body for cell and tissue associated data. It is meant to enhance the information retrieval and visualization of biological data without requiring computational background knowledge.

## Acknowledgements

The authors thank all CellFinder team members for their excellent comments and support in implementing and validating the SBB.

## Funding

This work was supported by the Deutsche Forschungsgemeinschaft, grant KU 851/3-1 to AK and partially supported by the Research Institute for Veterinary Science, Seoul National University.

*Conflict of Interest:* none declared.

## References

- Anders, S. et al. (2013) Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. *Nat. Protoc.*, **8**, 1765–1786.
- Anders, S. et al. (2015) HTSeq—A Python framework to work with high-throughput sequencing data. *Bioinformatics*, **31**, 166–169.
- Barrett, T. et al. (2013) NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.*, **41**, D991–D995.
- Carpendale, S. et al. (2014) Ontologies in Biological Data Visualization. *IEEE Comput. Graph. Appl.*, **34**, 8–15.
- Flicek, P. et al. (2014) Ensembl 2014. *Nucl. Acids Res.*, **42**, D749–D755.
- Krupp, M. et al. (2012) RNA-Seq Atlas—a reference database for gene expression profiling in normal tissue by next-generation sequencing. *Bioinformatics*, **28**, 1184–1185.
- Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357–359.
- Li, B. et al. (2010) RNA-Seq gene expression estimation with read mapping uncertainty. *Bioinformatics*, **26**, 493–500.
- Li, H. et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
- Mitsuhashi, N. et al. (2009) BodyParts3D: 3D structure database for anatomical concepts. *Nucleic Acids Res.*, **37**, D782–D785.
- Rustici, G. et al. (2013) ArrayExpress update—trends in database growth and links to data analysis tools. *Nucleic Acids Res.*, **41**, D987–D990.
- Selmann, S. et al. (2013) CELDA—an ontology for the comprehensive representation of cells in complex systems. *BMC Bioinformatics*, **14**, 228.
- Smith, B. et al. (2005) Relations in Biomedical Ontologies. *Genome Biol.*, **6**, 46.
- Stachelscheid, H. et al. (2014) CellFinder: a cell data repository. *Nucleic Acid Res.*, **42**, D950–D958.
- Trapnell, C. et al. (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*, **25**, 1105–1111.
- Wagner, G.P. et al. (2012) Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theor. Biosci.*, **131**, 281–285.