Stem Cell Research (2009) 3, 88–95

available at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/scr

REGULAR ARTICLE

The Characterization Tool: A knowledge-based stem cell, differentiated cell, and tissue database with a web-based analysis front-end

Inken Wohlers [a,b,1], Harald Stachelscheid [a,b,*,1], Joeri Borstlap [b], Katrin Zeilinger [a,b], Jörg C. Gerlach [a,b,c]

[a] *Department of Experimental Surgery, Charité-Universitätsmedizin Berlin, Campus Virchow-Klinikum, Germany*
[b] *Berlin-Brandenburg Center for Regenerative Therapies, Charité-Universitätsmedizin Berlin, Campus Virchow-Klinikum, Germany*
[c] *McGowan Institute for Regenerative Medicine, University of Pittsburgh, Pittsburgh, PA, USA*

Received 20 October 2008; received in revised form 13 May 2009; accepted 14 May 2009

Abstract In the rapidly growing field of stem cell research, there is a need for universal databases and web-based applications that provide a common knowledge base on the characteristics of stem cells, differentiated cells, and tissues by collecting, processing, and making available diverse types of characterization data. The Characterization Tool is such a novel knowledge database that allows the storage of various characteristics of cells, cell lines, and tissues across different species as well as the analysis of associated marker profiles. Its broad ontology-based framework facilitates the integration of characterization data on the morphological, molecular, and functional level acquired *in vivo* and *in vitro* including published marker expressions, cross-references to other databases, text descriptions, information on characterization experiments, and image storage. Data input and modification are recorded on the basis of a secure user management. By means of several easy-to-use data mining tools, marker profiles can be searched and analyzed. The Characterization Tool will aid in the establishment of standards for cell characterization, needed, for example, for stem cell isolation, propagation, and differentiation. The Characterization Tool is available at http://characterizationtool.cellnet.org. It currently holds more than 7000 marker expressions for different human embryonic stem cell lines, adult stem cells, and differentiated cells.
© 2009 Elsevier B.V. All rights reserved.

Introduction

Cells and tissues can be characterized on gene, protein, functional, and morphological levels, where the particular characteristics of interest depend on manifold objectives and on differing measurement techniques. Most notably, the relatively young field of stem cell research makes extensive use of cell and tissue characterization techniques with the goal of identifying tissue stem cells and defining marker patterns for different developmental cell states and degrees of cell differentiation. Activities in stem cell research, especially on human embryonic stem cells, focus on the development of differentiation protocols to yield fully differentiated cells that can be utilized for regenerative therapies such as cell transplantation, for example, in hepatic (Asahina et al., 2006; Fiegel et al., 2006), pancreatic (Burns et al., 2006), and neuronal (Bambakidis et al., 2008; Zhang et al., 2008)

* Corresponding author. AG Experimentelle Chirurgie/BCRT, Biomedical Research Center, Charité-Universitätsmedizin Berlin, Campus Virchow-Klinikum, Augustenburger Platz 1, 13353 Berlin, Germany. Fax: +49 30 450 559 909.
E-mail address: Harald.Stachelscheid@charite.de (H. Stachelscheid).
[1] Both authors contributed equally to this work.

1873-5061/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.scr.2009.05.001

diseases. Other lines of investigation focus on extracorporeal applications (Gerlach, 2006), toxicity testing, and pharmaceutical drug screening. For directed differentiation it is necessary to distinguish different cell types and to characterize specific cells, which have been subjected to a differentiation protocol. In the field of adult stem cell research the identification of specific markers, or descriptive marker sets that can be used to establish isolation protocols from primary cell suspensions, is a further challenge. However, the characterization approaches applied for these tasks are fragmented and corresponding data are widely spread, which leads to the need for a common ground (Blow, 2008): A common reference is desirable to aid in the establishment of standards for cell characterization.

Currently there are many specialized databases and web applications that provide data, which are frequently limited to one measurement technique or focused on individual species and/or cell and tissue types (Haudry et al., 2008; Ringwald et al., 1997; Edgar et al., 2002). To date, however, there is no database or interface that incorporates characterization data from different species, measurement techniques, and sources from the tissue down to the level of single cells *in vivo* as well as *in vitro* into a single comprehensive framework. To address this task, the Characterization Tool (CT) was developed.

Basic concept and database design

The Characterization Tool has four objectives: The major objective is to facilitate storage of biological data from the tissue down to the single-cell level of different species, obtained through various measurement techniques. Secondly, it provides a framework that is capable of integrating or linking to data from various publicly available data sources, especially data from expression databases, which are usually set up for only one species, but are available for a number of model organisms. Its framework also serves as an access point to external web applications and databases. Thirdly, it offers several easy-to-use modules that enable searching and analyzing the data, e.g., by means of comparison of marker profiles using tables, heat maps, and hierarchical clustering. Finally its generic concept permits the easy extension of the database for the storage of additional information that sets the data in a broader context, e.g., by incorporation of experimental raw data.

In order to organize and automatically search data, biological databases use controlled vocabularies, so-called ontologies, which link, for example, expression data to a set of common anatomical entities. For data in different developmental stages, ontology terms might represent tissue types divided into Theiler (mouse) (Theiler, 1989) or Carnegie (human) (Gasser, 1975) stages. In contrast to fixed terms, the CT allows the user to either define a custom term, which is not subject to the rather strict requirements of ontologies and can be manually linked to several terms of multiple ontologies, or choose exactly one existing term of a suitable and most likely species-specific ontology. By this means the CT can hold data from different species and developmental stages for tissues, as well as cells *in vivo* and *in vitro,* and as such has a broader setup than other databases. Currently the Foundational Model of Anatomy (FMA) ontology (Rosse and Mejino, 2003), a broad ontology of anatomical entities that also includes cells, is integrated into the CT to organize cells and tissues. Markers on the RNA and protein level are automatically linked with the corresponding gene ontology (GO) terms that provide descriptions in terms of associated biological processes, cellular components, and molecular functions in a species-independent manner (Ashburner et al., 2000). An overview on the structure of the data in the database and the cross-linking and incorporation of external databases is shown in Fig. 1.

Markers

To provide a definite characterization of all cell types of different organisms, their morphological, molecular, and

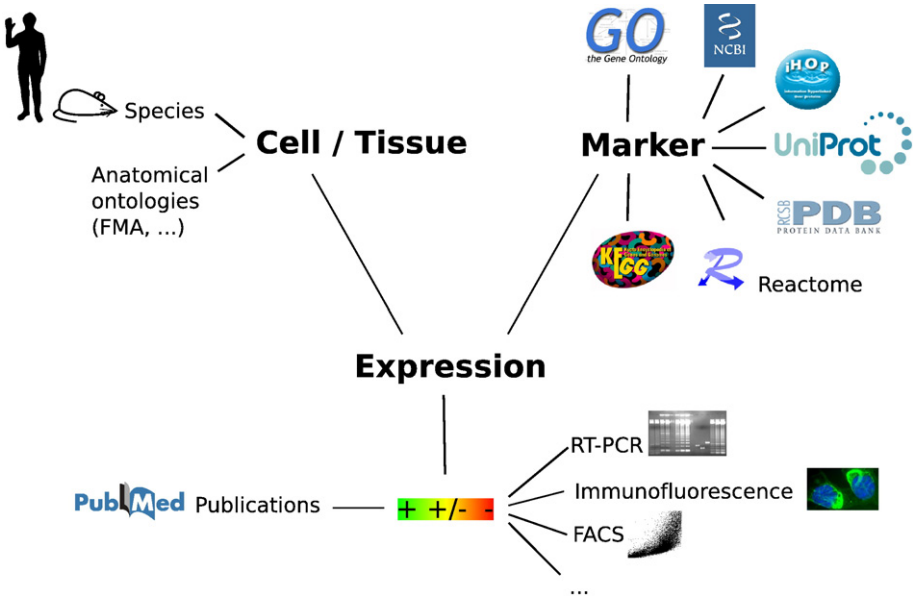


Figure 1 Data structure of the Characterization Tool. Relationships and cross-linking of cells and tissue, markers and expression.

functional features must be defined. In the context of cell or tissue characterization the above-noted features are termed markers. One specific cell is characterized by the sum of its markers. A marker is defined by the probe and the measurement method used for detection and can be, for example, an antibody detecting a specific epitope, a primer pair that is specific for an expressed mRNA/cDNA, or a functional assay. Since different probes for the same gene can detect different gene products, e.g., gene products derived from one gene by mechanisms such as alternative splicing or differential translation, more than one marker can be mapped to a corresponding gene. The concept of using "markers," and not only genes and gene products for cell and tissue characterization, is therefore much more flexible and comprehensive.

In the CT each marker holds a description with revision control, if available, a gene symbol, synonyms, and a list of cells and tissues, to which the marker has been associated. Cross-references to other databases and web-based applications are generated automatically (Fig. 1). Currently the Characterization Tool's markers are linked to the gene and protein databases Genbank (Benson et al., 2008) and UniProt (Wu et al., 2006), the pathway databases Reactome (Joshi-Tope et al., 2005) and KEGG (Wixon and Kell, 2000), the literature search tool iHOP (Hoffmann and Valencia, 2004), the structure database PDB (Berman et al., 2000), and to IDClight that performs conversion and annotation of gene and protein IDs (Alibes et al., 2007).

Markers are listed either according to user-defined categories, to a subset of GO terms, or alphabetically according to name or gene symbol. The list of markers can be either browsed or searched.

Cells and tissues

The Characterization Tool provides a framework for marker profiling of tissues as well as cells from various species using diverse types of markers. This way it is possible to integrate data from more specific databases at a later date, especially with respect to developmental biology and stem cell differentiation.

Cells or tissues in the CT are displayed in species-specific tables (Fig. 2c), where each cell's or tissue's marker profile depends on the corresponding species, permitting cross-species comparisons of specific marker profiles. Each cell or tissue record holds a description with revision control, a list of markers, synonyms, and annotated images that can serve as a reference for morphological characterization (Fig. 2f). The CT is capable of using several ontologies of anatomical entities or developmental stages in parallel. The corresponding ontology terms associated with a cell or tissue can be visualized as a network graph (Fig. 2e).

Expression

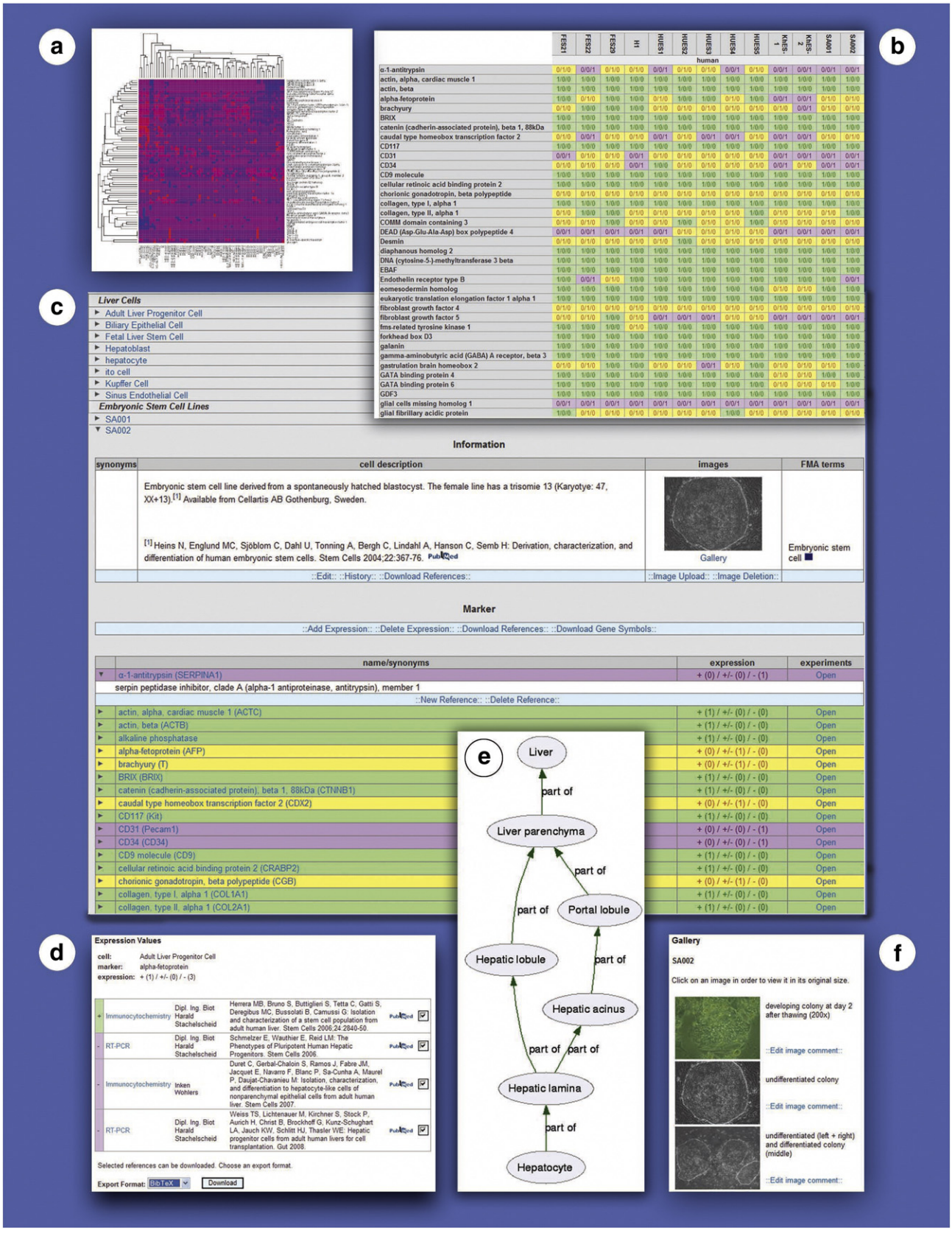
The expression of a biomolecular marker is verified or disproved in an experiment by applying a biomolecular technique such as immunostaining, RT-PCR, or FACS. In the field of stem cell research, for example, such experiments are conducted on a large scale in order to characterize stem cell lines in their undifferentiated state as well as under various differentiation conditions. Additionally, the expression of markers can also be extracted from experiments where the primary intention has not been the characterization of cells or tissues. Usually in the context of a publication the result of an experiment is interpreted in a qualitative way: it is reported that either the marker has been verified or disproved or the result was ambiguous.

In the Characterization Tool, the expression of an individual marker consists of such a qualitative classification of the marker expression, whereas '+' means expressed, '−' not expressed, and '+/−' maybe expressed. The experimental data that underlie the marker expressions originate from different measurement methods, which may have very different output formats and precision. Therefore, the raw data of different measurement techniques usually cannot be compared. Consequently, the classification of marker expression ('+', '−' or '+/−') is based on the interpretation of experimental results within the publication—the CT thus uses the author's assessment of the data. If different authors apply different criteria, in the extreme case this approach might lead to different thresholds during transformation from quantitative experimental data to qualitative marker expressions. For this reason only the most basic classification of marker expression, '+', '−' and '+/−' has been used, thereby minimizing the variation in the data due to subjective interpretation while maintaining reasonable validity. Besides the qualitative marker expression, the Characterization Tool stores the biomolecular technique that has been applied to detect the marker and a reference to the publication in which the corresponding experiment has been reported (Fig. 1 and Fig. 2d). Therefore, the criteria that originally have been applied by the authors to derive the qualitative marker expression ('+', '−' or '+/−') can easily be accessed. Furthermore, the Characterization Tool records the submitting user of each marker expression and also provides storage space for additional information about the corresponding experiment, e.g., for images or protocols.

If several experiments have been conducted for the same marker for a specific cell type, a cumulative qualitative marker expression is computed. Therefore, the number of experiments in which the marker has been interpreted as being expressed ('+') is counted, as well as the number of experiments in which the marker has been interpreted as being not expressed ('−'). If the majority of experiments

Figure 2 The web-based analysis front-end. (a) Heat map representing the result of a cluster analysis of different cell types based on their expression profiles. (b) Matrix that compares different cells based on their marker expression. (c) Main view of cells and tissues that includes a description, a representative picture, connected ontology terms, and a color-coded list of markers. (d) Detailed view of a cell type's markers, which also lists references and the described methods. (e) Graph of relationships of ontology terms. (f) Gallery view of annotated images of a certain cell type.

reports an expression, the overall marker expression is set to '+'; likewise, if a majority of experiments report that the marker is not expressed, the overall marker expression is set to '–'; in all other cases the overall marker expression is set to '+/–'. Intuitive color coding is applied for the survey of marker expression profiles.



Data exploration and analysis

The basic concept for recording cells, tissues, and markers provides a framework that permits computational analysis by means of online analysis modules for searching, comparison, and evaluation.

Several such Characterization Tool modules focus on the comparison of marker expression profiles of different cells and tissues. Here, basic visualizations are displayed in the shape of expression matrices for a user-defined set of markers from different cells and tissues from one or several species (Fig. 2b). In these matrices information on the origin of the marker expression in the shape of associated publications and experiments are linked to the corresponding expression, enabling easy access.

Furthermore, marker profiles described in various publications for one specific cell or tissue type can be compared. This is, for example, useful for comparing cells derived by stem cell differentiation toward a specific cell type using different protocols with cells of the corresponding primary cell type. Finally, marker profiles can be compared according to the GO terms (molecular function, cellular component, or biological process) of markers or the FMA terms (superordinate anatomical entities) of cells or tissues. This way, for example, it is possible to specify all cell surface markers by selecting markers associated with the gene ontology term "plasma membrane," and to retrieve cells and tissues where these markers are expressed.

Expression matrices generated with analysis modules can be further analyzed by hierarchical clustering of markers, cells, or tissues, e.g., to detect markers with similar expression patterns or to examine the relation between different cells and tissues. The clustering algorithm that is used within the Characterization Tool was developed by Eisen at al. (Eisen et al., 1998) and is commonly used for hierarchical clustering of gene expression data. Its implementation is provided by the R package MADE4 (Culhane et al., 2005). Overall qualitative marker expressions are used as input for the clustering algorithm. If the majority of experiments report a marker expression and thus the overall expression is '+', the value for clustering is 1. Likewise, if in the majority of cases the marker could not be detected and therefore the overall expression is '−', the value for clustering is −1. In all other cases—also in the case of missing data—the value for clustering is 0. As a result of cluster analysis, a heat map with dendrograms is generated (Fig. 2a). This is particularly useful for the identification of common marker sets that characterize distinct cell populations such as adult stem cells of a certain tissue, or for the investigation of the relatedness of different stem cell lines.

Another Characterization Tool module is used for querying specific markers that show exclusive expression for a specific cell type, thereby, for example, aiding in the process of identifying markers for the development of cell isolation strategies utilizing cell-sorting techniques. This function has the option to compare individual cell types with a defined number of other cell types or with all cell types available in the database.

Technical implementation

The Characterization Tool is web-based and therefore completely platform independent. It uses PHP, JavaScript/AJAX, and a relational MySQL database and runs on a Linux server. The integration of specific functionalities is achieved by the utilization of various open-source modules (Culhane et al., 2005; Gentleman et al., 2004; Emden and Stephen, 2000; http://www.imagemagick.org, http://tinymce.moxiecode.com, http://biborb.glymn.net). The FMA ontology has been integrated with a PHP script, which converts arbitrary flat files from Open Biomedical Ontology (OBO) format (http://obofoundry.org) into MySQL tables. A local version of the Gene Ontology has been set up using a Gene Ontology database dump. Browsing FMA and GO is achieved by means of AJAX; the FMA can additionally be searched for string literals. Heat maps and hierarchical clusterings are computed, and graphics are created using the statistics software R (R Development Core Team, 2005). For handling bibliographies, several open source modules are used and extended to allow automatic retrieval of references from PubMed (http://www.pubmed.gov) as well as to export to all common bibliographic formats. All of the Characterization Tool's database description entries have revision control, and the image upload supports all common formats. For each cell type a list of gene symbols of expressed markers can be downloaded as a flat file for use in external programs. The Tool is available at http://characterizationtool.cellnet.org.

Data input and validation

Browsing, searching, and using the analysis functions is feely available through the Internet, but adding and modifying data is controlled by a user management. After login the user has free access to all data modification functions. This is unlike open access databases such as Wikipedia (http://www.wikipedia.org) were data can be changed without login. Accordingly, the general mechanism for data validation is a closed wiki model which allows only a group of trusted users to insert, delete, alter or reform entries or to revert changes made during a previous edit. To get an account a registrant contacts the administrators and supplies his contact details. After validating that the applicant is bona fide user—i.e., from a recognized research organization—the administrators open a new user account, which is used to access the database. The user chooses a login name and a password.

Marker expressions are validated by publications entered by the user. Along with the publication, the submitting user is recorded, allowing one to track data input and modification back to the respective user.

Textual descriptions provide the functionality to include references that are automatically linked to their PubMed entry. A version control records all changes that have been made in any text field and allows a review and, if necessary, to revert any changes made.

Discussion

The field of stem cell research has a high demand for a comprehensive knowledge database on the characteristics of

stem cells, differentiated cells, and tissues from different species, on the morphological, molecular, and functional level *in vivo* and *in vitro*. Such a knowledge base could be a common point of reference and would help in the establishment of standards for cell characterization, needed, for example, for stem cell isolation, propagation, and differentiation.

To fully characterize specific cells or tissues, analysis must be performed on the gene, protein, functional, and morphological level. A database for the storage of results from such analysis must deal with many different types of data. None of the currently available databases and web applications sufficiently fulfills this requirement because they are limited to store data of one measurement technique, or focus on individual species and/or cell and tissue types, for example, StemBase that provides data of stem cells from different species but is limited to RNA expression data (Porter et al., 2007). 4Dxpress stores gene expression data from *in situ* hybridization, antibody, and transgenic experiments down to the tissue level from drosophila, zebrafish, medaka, and mouse but not from human and not down to the single-cell level (Haudry et al., 2008). Another example is the Gene Expression Database (GXD), which only provides gene expression data from mouse (Ringwald et al., 1997). Therefore, the Characterization Tool was developed, which provides a user-friendly, flexible, and upgradeable web-based and thus platform-independent framework that enables storage and provision for a broad range of information for cell and tissue characterization and integration of various databases and applications available online. Additionally, several functions for display, search, and evaluation of data were implemented in the CT.

The primary data source of the CT is data that is extracted from publications, making a huge amount of valuable data available for searching and analysis which, until now, was only available inside the text of publications or downloadable as supplemental data from the publisher's or author's web sites as flat files.

The currently implemented analysis modules focus on data comparison. Comparative data analysis leads to the identification of specific makers or marker sets for specific cell types and are of special interest in stem cell research. One application where specific markers are needed is the development of cell isolation and purification strategies like cell isolation from primary tissues, or purification of mixed cell populations derived by stem cell differentiation. This is particularly useful in the separation of remaining pluripotent cells in cell suspensions derived by embryonic stem cell differentiation.

Another application of cell specific marker sets is their use as reference markers to compare a primary cell type with its analog derived by differentiation of stem cells *in vitro*. In such comparisons, functional markers that can be stored in the tool are of greatest value, especially when dealing with cells characterized by highly specialized functions like hepatocytes. For example, the CT currently contains characterization data from publications describing hepatic differentiation approaches of mouse and human embryonic stem cells.

In the field of adult stem cell research, the CT can aid in the identification of specific markers, which then can be used as a standard in further studies on specific stem cell populations. Here, an example is a dataset on adult liver stem cells in the CT. There are numerous publications on identification and isolation approaches of adult liver stem cells (for a review, see Walkup and Gerber, 2006; Dan and Yeoh, 2008) but so far no specific marker for these cells has been identified. Therefore, the markers used for cell characterization highly vary among the individual studies. In addition, many studies use quite unspecific markers that are also expressed on other hepatic or extrahepatic cell types. Using the CT it is now possible to easily compare the described cell populations from different studies and identify differences and commonalities.

Currently the CT holds more than 7000 marker expressions for different human embryonic stem cell lines, adult stem cells, and differentiated cells and tissues. For example it holds a dataset from the International Stem cell Characterization Initiative (ISCI-1), in which 17 laboratories from 11 countries characterized 59 human embryonic stem cell lines by means of standardized techniques and a set of 100 markers (Adewumi et al., 2007); this dataset can now be used to validate newly derived stem cell lines because their characteristics can be easily compared with those of established lines.

Future directions

The further development of the CT will be the integration of or linking to experimental raw data, e.g., RT-PCR, FACS, ELISA, microarray, SAGE, MPSS, and RNA-Seq data as well as EST libraries. This will add a quantitative level of marker profiling to the existing qualitative level and give rise to various new modules for analysis and to more sophisticated querying of data. Recently, several sets of experimental data that focus on stem cells have been published (e.g. http://www.stemcellmatrix.org; http://www.stemdb.org); in a first step these may be included into the Characterization Tool via linkage. On integration of experimental raw data the user will be able to specify precisely how the data are interpreted and converted to the qualitative marker expressions '+', '−' and '+/−', for example, by specification of thresholds. Dedicated sections for each type of supported measurement technique in the CT will store experiments in a standardized way, using, if available, the established method's minimum information (MI) reporting guidelines, e.g., MISFISHIE (Deutsch et al., 2008) and MIFlowCyt (http://flowcyt.sourceforge.net/miflowcyt).

The CT will therefore provide an easy and platform-independent way to store, exchange, and disseminate experimental data within a defined community. This may be achieved by modular integration into individual intranet environments of research consortia to offer a virtual and standardized data space to topographically spread research groups, promoting standardization and leading to better collaboration. After setup of a public master database, sets of experimental data and research findings from the contents of such intranet databases may be shared on the public CT platform.

Furthermore future work lies on cross-linking Characterization Tool data with marker expression data from the European Human Embryonic Stem Cell Registry (www.hescreg.eu) (Borstlap et al., 2008). hESCreg offers the research community, legislators, regulators, and the general

public at large an in-depth overview on the current status of hESC research in Europe. The freely accessible database contains information about approximately 600 hESC lines that have been derived in Europe and beyond (as of March 2009).

Acknowledgments

This work was supported by the German Federal Ministry of Education and Research (BMBF, 01GN0526, 01GN0526, 0313911) and the European Commission (STREP-CT-2005-018940). We thank Alexander Damaschun for helpful discussions and critical review of the manuscript and Martin Pietsch for technical support.

References

Adewumi, O., Aflatoonian, B., Ahrlund-Richter, L., Amit, M., Andrews, P.W., Beighton, G., Bello, P.A., Benvenisty, N., Berry, L.S., Bevan, S., Blum, B., Brooking, J., Chen, K.G., Choo, A.B.H., Churchill, G.A., Corbel, M., Damjanov, I., Draper, J.S., Dvorak, P., Emanuelsson, K., Fleck, R.A., Ford, A., Gertow, K., Gertsenstein, M., Gokhale, P.J., Hamilton, R.S., Hampl, A., Healy, L.E., Hovatta, O., Hyllner, J., Imreh, M.P., Itskovitz-Eldor, J., Jackson, J., Johnson, J.L., Jones, M., Kee, K., King, B.L., Knowles, B.B., Lako, M., Lebrin, F., Mallon, B.S., Manning, D., Mayshar, Y., McKay, R.D.G., Michalska, A.E., Mikkola, M., Mileikovsky, M., Minger, S.L., Moore, H.D., Mummery, C.L., Nagy, A., Nakatsuji, N., O'Brien, C.M., Oh, S.K.W., Olsson, C., Otonkoski, T., Park, K., Passier, R., Patel, H., Patel, M., Pedersen, R., Pera, M.F., Piekarczyk, M.S., Pera, R.A.R., Reubinoff, B.E., Robins, A.J., Rossant, J., Rugg-Gunn, P., Schulz, T.C., Semb, H., Sherrer, E.S., Siemen, H., Stacey, G.N., Stojkovic, M., Suemori, H., Szatkiewicz, J., Turetsky, T., Tuuri, T., van den Brink, S., Vintersten, K., Vuoristo, S., Ward, D., Weaver, T.A., Young, L.A., Zhang, W., 2007. Characterization of human embryonic stem cell lines by the International Stem Cell Initiative. Nat. Biotechnol. 25, 803–816.

Alibes, A., Yankilevich, P., Canada, A., Uriarte, R.D., 2007. IDconverter and IDClight: Conversion and annotation of gene and protein IDs. BMC Bioinform. 8.

Asahina, K., Teramoto, K., Teraoka, H., 2006. Embryonic stem cells: hepatic differentiation and regenerative medicine for the treatment of liver disease. Curr. Stem Cell Res. Ther. 1, 139–156.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G., 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. 25, 25–29.

Bambakidis, N.C., Butler, J., Horn, E.M., Wang, X., Preul, M.C., Theodore, N., Spetzler, R.F., Sonntag, V.K., 2008. Stem cell biology and its therapeutic applications in the setting of spinal cord injury. Neurosurg. Focus 24.

Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Wheeler, D.L., 2008. GenBank. Nucleic Acids Res. 36, D25–D30.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E., 2000. The Protein Data Bank. Nucleic Acids Res. 28, 235–242.

Blow, N., (2008), Stem cells: in search of common ground. Nature 451 855–858.

Borstlap, J., Kurtz, A., Stacey, G., Elstner, A., Damaschun, A., Aran, B., Gerlach, J., Izpisua, J., Veiga, A., 2008. Development of a European human embryonic stem cell registry. Regen. Med. 3, 945–951.

Burns, C.J., Persaud, S.J., Jones, P.M., 2006. Diabetes mellitus: a potential target for stem cell therapy. Curr. Stem Cell Res. Ther. 1, 255–266.

Culhane, A.C., Thioulouse, J., Perrière, G., Higgins, D.G., 2005. MADE4: an R package for multivariate analysis of gene expression data. Bioinformatics 21, 2789–2790.

Dan, Y.Y., Yeoh, G.C., 2008. Liver stem cells: a scientific and clinical perspective. J. Gastroenterol. Hepatol. 23, 687–698.

Deutsch, E.W., Ball, C.A., Berman, J.J., Bova, G.S., Brazma, A., Bumgarner, R.E., Campbell, D., Causton, H.C., Christiansen, J.H., Daian, F., Dauga, D., Davidson, D.R., Gimenez, G., Goo, Y.A., Grimmond, S., Henrich, T., Herrmann, B.G., Johnson, M.H., Korb, M., Mills, J.C., Oudes, A.J., Parkinson, H.E., Pascal, L.E., Pollet, N., Quackenbush, J., Ramialison, M., Ringwald, M., Salgado, D., Sansone, S.A., Sherlock, G., Stoeckert, C.J., Swedlow, J., Taylor, R.C., Walashek, L., Warford, A., Wilkinson, D.G., Zhou, Y., Zon, L.I., Liu, A.Y., True, L.D., 2008. Minimum information specification for in situ hybridization and immunohistochemistry experiments (MISFISHIE). Nat. Biotechnol. 26, 305–312.

Edgar, R., Domrachev, M., Lash, A.E., 2002. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. 30, 207–210.

Eisen, M.B., Spellman, P.T., Brown, P.O., Botstein, D., 1998. Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. USA 95, 14863–14868.

Emden, R.G., Stephen, C.N., 2000. An open graph visualization system and its applications to software engineering. Softw. Pract. Exp. 30, 1203–1233.

Fiegel, H.C., Lange, C., Kneser, U., Lambrecht, W., Zander, A.R., Rogiers, X., Kluth, D., 2006. Fetal and adult liver stem cells for liver regeneration and tissue engineering. J. Cell. Mol. Med. 10, 577–587.

Gasser, R.F., 1975. Atlas of Human Embryos. Harper and Row, Hagerstown.

Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., andLaurent Gautier, B.E., Ge, Y., Gentry, J., Hornik, K., Hothorn, T., Huber, W., Iacus, S., Leisch, R.I.F., Li, C., Maechler, M., Sawitzki, A.J.R.G., Smith, C., Smyth, G., Yang, L.T. J.Y.H., Zhang, J., 2004. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 5, R80.

Gerlach, J.C., 2006. Bioreactors for extracorporeal liver support. Cell Transplant. 15 (Suppl. 1).

Haudry, Y., Berube, H., Letunic, I., Weeber, P.D., Gagneur, J., Girardot, C., Kapushesky, M., Arendt, D., Bork, P., Brazma, A., Furlong, E.E., Wittbrodt, J., Henrich, T., 2008. 4DXpress: a database for cross-species expression pattern comparisons. Nucleic Acids Res. 36, 847–853.

Hoffmann, R., Valencia, A., 2004. A gene network for navigating the literature. Nat. Genet. 36, 664.

Joshi-Tope, G., Gillespie, M., Vastrik, I., D'Eustachio, P., Schmidt, E., de Bono, B., Jassal, B., Gopinath, G.R., Wu, G.R., Matthews, L., Lewis, S., Birney, E., Stein, L., 2005. Reactome: a knowledgebase of biological pathways. Nucleic Acids Res. 33, D428–D432.

Porter, C.J., Palidwor, G.A., Sandie, R., Krzyzanowski, P.M., Muro, E.M., Perez-Iratxeta, C., Andrade-Navarro, M.A., 2007. StemBase: a resource for the analysis of stem cell gene expression data. Methods Mol. Biol. 407, 137–148.

R Development Core Team, 2005. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Ringwald, M., Davis, G.L., Smith, A.G., Trepanier, L.E., Begley, D.A., Richardson, J.E., Eppig, J.T., 1997. The mouse gene expression database GXD. Semin. Cell Dev. Biol. 8, 489–497.

Rosse, C., Mejino, J.L.V., 2003. A reference ontology for biomedical informatics: the foundational model of anatomy. J. Biomed. Inform. 36, 478–500.

Theiler, K., 1989. The House Mouse. Springer-Verlag, NY.
Walkup, M.H., Gerber, D.A., 2006. Hepatic stem cells: in search of. Stem Cells 24, 1833–1840.
Wixon, J., Kell, D., 2000. The Kyoto encyclopedia of genes and genomes—KEGG. Yeast 17, 48–55.
Wu, C.H., Apweiler, R., Bairoch, A., Natale, D.A., Barker, W.C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., Martin, M.J., Mazumder, R., O'Donovan, C., Redaschi, N., Suzek, B., 2006. The Universal Protein Resource (UniProt): an expanding universe of protein information. Nucleic Acids Res. 34, D187–D191.
Zhang, S.C., Li, X.J., Johnson, M.A., Pankratz, M.T., 2008. Human embryonic stem cells for brain repair? Philos. Trans. R. Soc. Lond., B Biol. Sci. 363, 87–99.