# Construction and Application of a Public-Domain Mesenchymal Stem Cell Database

Weiqi Wang, René Bañares-Alcántara, Zhanfeng Cui*, Yanbo Justin Wang, and Frans Coenen

***Abstract*—Mesenchymal Stem Cells (MSCs) are important to stem cell therapy and tissue engineering due to their differentiation potentials both in vivo and in vitro. The variety of MSC culture scenarios has triggered an urgent need for the integration of MSC experimental data. In this study, a public-domain database comprising the key parameters which influence the behaviors and fate of mammalian MSCs has been constructed. The data from this database have been analyzed using data mining techniques, the results of which show the significance of this database and its potential application in the future. As the first public-domain MSC database, it contributes to MSC data integration and has potential contribution to future research on MSC.**

## I. Introduction

Mesenchymal Stem Cells (MSCs) are important to stem cell therapy and tissue engineering due to their pluripotent differentiation potentials [1]-[9], and have become one of the most studied types of stem cells nowadays. The pluripotency of MSCs includes osteogenesis, chondrogenesis, adipogenesis, myogenesis, tendonogenesis, and neurogenesis, besides trans-differentiation [1]-[8]. The plasticity and immunologic properties of MSC have further increased their clinical applications [9]-[14]. In order to pursue a better understanding of MSCs, a huge number of studies have been carried out both in vivo and in vitro [15]-[24]. However, due to the diversity of culture conditions for MSC differentiation in different experiments, the cell behaviors, especially the differentiation fates of MSCs were also different. Many alternative culture conditions are possible, for example in terms of donor species, culture medium, supplement and growth factor, culture dimension (monolayer vs. 3D culture), and substrate (for monolayer culture) vs. scaffold (for 3D culture) [4]-[24]. They result in a large yet scattered spectrum of MSC differentiation scenarios, which compromises the advancement in achieving a better overview for MSC differentiation.

For this reason, an urgent need for the collection and integration of MSC experimental data has been triggered. Although many bioinformatics databases have been made accessible online, no public-domain database specifically on MSC differentiation was available until recently. In this study, a database containing over 500 parameters organized in more than 30 parameter groups that are believed to influence the MSC differentiation has been built and put in a public domain so that it can be freely accessed online by MSC researchers all over the world. All the data in this online database have been published in the literature and can be viewed by worldwide researchers. Moreover, new data can be entered by any registered user, to increase the size of the database and improve data integration for MSC research.

With the help of this database, scattered data on MSC differentiation which can not be utilized by traditional modeling techniques are made usable to pattern recognition -based computational approaches, such as data mining techniques, to make prediction on MSC behaviors. With the expansion of this database, a better overview for our current understanding on MSC can be achieved.

This paper is organized as follows: Section II describes the A) parameters in the database, B) construction of the database in MYSQL, C) access to the online database, and D) data analysis using data mining techniques as an application of the database. Section III is the conclusions.

## II. Construction of the Public-Domain MSC Database

### A. Selection and Ranking of Parameters in the Database

According to the known experimental scenarios in MSC differentiation, the parameters representing culture conditions together with experimental outcomes have been represented in the database. These parameters include both qualitative parameters such as various types of donor species, culture media, supplements, substrates, cell markers, gene profiles, and quantitative parameters such as growth factor dosage, seeding density, perfusion rate, culture duration, population doubling time, etc. All the parameters are categorized into 36 groups (a parameter group may contain only one parameter), as listed and briefly described in Table I. All the parameter groups have been ranked into four levels according to their significance to experimental outcome, based on a priori knowledge. The parameter groups which are believed to be most significant, such as donor species, in vitro vs. in vivo culture, culture medium, supplement and growth factor, culture dimension (monolayer vs. 3D culture), substrate (for monolayer culture) vs. scaffold (for 3D culture), are marked as Rank 1. Those which are believed to be

Manuscript received July 30, 2009.

W. Wang (e-mail: weiqi.wang@eng.ox.ac.uk), R. Bañares-Alcántara (e-mail: rene.banares@eng.ox.ac.uk) and Z. Cui are with the Department of Engineering Science, Parks Road, University of Oxford, OX1 3PJ, UK.

*Z. Cui is the corresponding author (phone: +44 1865 273118/273017; fax: +44 1865 273905; e-mail: zhanfeng.cui@eng.ox.ac.uk).

Y. J. Wang is with the Information Management Center, China Minsheng Banking Corporation Limited, Room 606, Building No. 8, 1 Zhongguancun Nandajie, Beijing, China, 100873.

F. Coenen is with the Department of Computer Science, University of Liverpool, Ashton Building, Ashton Street, Liverpool, L69 3BX, UK.

potentially important, such as age of donor, cell passage number, cell seeding density, incubation duration, are marked as Rank 2. The least important parameter groups, which usually act as supplementary comments, such as donor gender, MSC harvest place, are marked as Rank 3. Lastly, the parameter groups representing cell behaviors as experimental outcome are marked as Rank 4, including MSC differentiation fate, expression of cell markers, gene profiles, expansion fold of cell number, etc.

TABLE I
PARAMETER GROUPS IN THE DATABASE

| PG [a] | Descriptions | R |
|---|---|---|
| Donor species | The current database covers human, rat, mouse, bovine, ovine and rabbit. | 1 |
| Vitro/vivo | MSC differentiation varies from in vitro to in vivo. | 1 |
| Culture medium | The most popular culture media include DMEM, DMEM-LG, DMEM-HG, α-MEM, and RPMI. | 1 |
| Supplement and growth factor | Common supplements include FBS, FCS and chemicals capable of stimulating cellular growth, proliferation and differentiation, such as dexamethasone, insulin, ascorbic acid, etc. | 1 |
| Culture dimension (2D / 3D) | MSC differentiation sometimes differs significantly from 2D to 3D culture, even in the presence of the same culture medium and supplements [25]. | 1 |
| Substrate /scaffold | In 2D culture, common substrates are plastic coated with gelatine or fibronectin. It has been claimed that chemically modified glass surfaces can influence MSC differentiation [24]. In 3D culture, common scaffolds usually include various kinds of collagen or hydrogel. | 1 |
| Cell line | Some experiments use highly purified MSC cell lines. | 2 |
| MSC source | MSC can be obtained from bone marrow, adipose tissue, peripheral blood, umbilical cord blood, fetal hepatic, and placenta [9], [10], [26]-[28] with different differentiation capacities. | 2 |
| Seeding density | Seeding density can potentially influence the velocity of cell growth and extent of MSC differentiation. | 2 |
| Age of donor | Normally, the proliferating capacity and pluripotency of MSCs decreases when the age of donor increases. | 2 |
| Passage number | Normally, the proliferating capacity and pluripotency of MSCs decreases when the passage number increases. | 2 |
| Culture duration | Commonly used time points for observation are 7, 10, 14, 21 and 28 days after a new passage. | 2 |
| Cryopreservation | Cryopreserved MSCs may have some different properties from non-cryopreserved MSCs. | 2 |
| Harvest place | MSCs can be obtained from iliac crest, tibias, femurs, and so on. | 3 |
| Gender | Gender of donor (male or female). | 3 |
| Buffering agent | Chemicals which keep a pH value in the medium. Popular buffers include $CO_2$, phosphate, and HEPES. | 3 |
| Temperature | Most experiments are undertaken at 37℃, while some others are undertaken at room temperature. | 3 |
| Culture vessel | Common culture vessels include Petri dishes, culture flasks, culture plates, micro bioreactors, etc. | 3 |
| Medium renewal interval | Culture medium is normally renewed every 2-3 days to ensure the presence of nutrients that keeps cells alive, while eliminating the metabolic waste secreted by cells. | 3 |
| Passaging enzyme | Passaging enzymes are not expected to affect MSC differentiation. A common enzyme is trypsin/EDTA. | 3 |
| Centrifugation rate | MSCs have to be centrifuged to be isolated from culture medium when passaging. Some experiments also require centrifugation to facilitate cartilage formation. | 3 |
| Substrate /scaffold Porosity | A substrate or scaffold with larger porosity can provide a better environment for nutrients to be evenly distributed within the culture medium. | 3 |
| Perfusion rate | MSCs are sometimes cultured in a perfusion system where culture medium is constantly renewed [29]. | 3 |
| OM [b] | Osmolarity of the culture medium. | 3 |
| pH | pH may have potential effects on MSC growth and differentiation (optimal pH for metabolism is 7.4). | 3 |
| HP [c] | Hydrostatic pressure of culture medium. | 3 |
| Additional info. | Supplementary details to culture conditions, such as doze/concentration of supplements and growth factors. | 3 |
| Differentiation fate | Common fates of MSC include osteogenesis, chondrogenesis, adipogenesis, myogenesis, and tendogenesis. | 4 |
| Future potency | Potency of MSCs after the respect experiment. | 4 |
| Doubling time | Time after which the MSC population doubles. | 4 |
| DP [d] | Percentage of differentiated MSCs with respect to the total number of cells. | 4 |
| Cell marker expression | Cell marker is the basis according to which the differentiation can be confirmed. Cell markers expressed on MSC vary during the MSC differentiation process, indicating the extent of MSC maturation. | 4 |
| Gene profile | Measurement of the activity of genes provides a global picture of cellular function, indicating the current state of MSCs and showing how MSC react to the chemicals in the current environment. | 4 |
| number of CFU-F Colonies | Number of CFU-F (Colony-Forming Unit Fibroblast) colony is an indication of how well MSCs grow in the current medium. The more colonies are formed, the better MSCs grow. | 4 |
| Expansion fold | Expansion folds of MSC with respect to the initial cell number in the current passage. It is a good indicator of MSC growth velocity. | 4 |
| Conclusions, etc. | Conclusions/hypotheses derived from experimental results. | 4 |

[a] PG = parameter group. [b] OM = osmolarity.
[c] HP = hydrostatic pressure. [d] DP = differentiation percentage.

The parameters in this database are constantly being updated with the expansion of the database. The current size of the database is 501 records with over 500 parameters maximum for each record. All the records in the database are abstracted from previously published papers and each is associated with its corresponding reference.

### *B. Construction of the Database in MySQL*

The database containing the parameters listed above was constructed with MySQL 5.0, as MySQL is a commonly used database management system which can be conveniently accessed and managed via PHP scripts on a common Linux server. The parameters in the database are stored as follows:

- For the qualitative parameters with only two possible values, such as the presence/absence of a growth factor in the culture medium, the corresponding attributes in the database are binary, i.e. '1' refers to 'presence' and '0' refers to 'absence'.
- For the qualitative parameters which have a set of possible categorical values, such as animal species, the attributes are stored as integers, e.g., '1' refers to 'human', '2' refers to 'rat', etc.
- For the quantitative parameters, such as cell seeding density, the attributes are stored as real numbers.
- For descriptive parameters, such as additional conditions, the attributes are stored as strings of characters.

The digitalization for most of the parameters in this database is to facilitate the data exportation and utilization by other software for data analysis (see section D).

### C. *Access to the Online Database*

The database was integrated into the website of OUTEG (Oxford University Tissue Engineering Group), accessible at http://www.oxford-tissue-engineering.org/forum/plugin.php?identifier=publish&module=publish. The OUTEG website is a public-domain academic forum powered by DISCUZ! 5.0.0 (© 2001-2006 Comsenz Inc.); the MSC database was integrated into the website as a plug-in. A typical user interface of the database is shown in Figure 1, where only brief information of the records is shown. The details of each record can be accessed by clicking the record number.

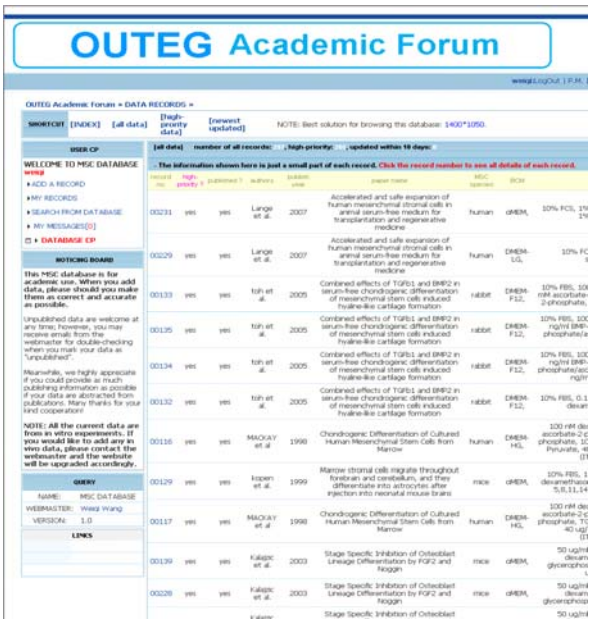


Fig. 1. User interface for the public-domain MSC database (partially). Details of each record can be accessed by clicking the record number.

The online MSC database is purely for academic purposes. It allows all users to view the data online, while giving only registered users the authority to add new data. With the intention to protect the database as well as to guarantee the authenticity of the data, the user registration is supervised by the webmaster. Only users having a valid email address with a domain of a university or a research institute can register. Registered users can get access to the data submission page, shown in Figure 2. In order to facilitate the input, all the parameters are divided into the following six sections (colored in purple in Figure 2): 1) publication information, 2) MSC harvest information, 3) culture media and supplements, 4) cultivation methods, 5) extra culture conditions, 6) outcome information. The parameter groups with Rank 1 (as listed in Table 1) are colored in red, the information of which have to be provided in order to accomplish a successful submission, as those parameters are believed to be essential for each record. All the records are tagged with their respective submitters, who can easily retrieve and edit their submissions. In order to avoid repeated data input, a search function has been provided online; in the meantime the database is periodically maintained and updated.

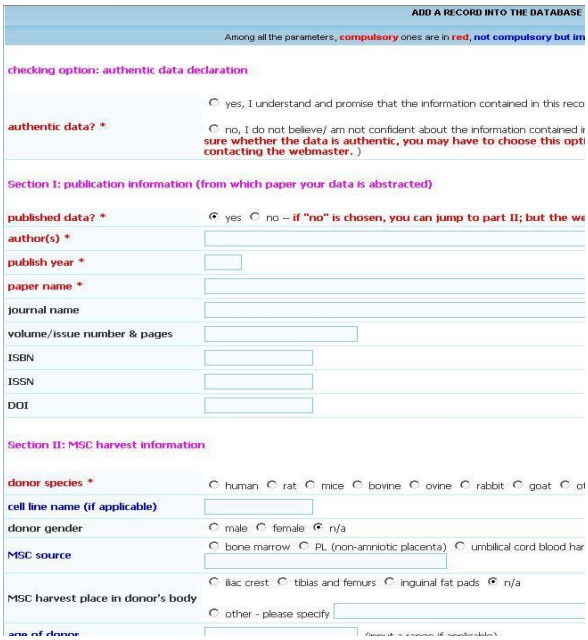


Fig. 2. Data submission page (partially). The parameters are divided into six sections (section names in purple). The parameters with Rank 1 (colored in red) are compulsory input, while those with Rank 2 (colored in blue) are highly recommended to be input. With the aim to guarantee the authenticity of the data, an "authentic data declaration" is compulsory in each submission.

### D. *Case Study: Application of the Database*

As mentioned in section B, the data in the online database can be easily exported and utilized by other software. In this study, two data mining frameworks with the names of CMAR (Classification based on Multiple Association Rules) [30] and ARM (Association Rule Mining) [31] have been chosen as examples to show the application of the database.

As the first application, CMAR is used to analyze MSC differentiation in vitro. In this application, six parameter groups were believed to contain most essential information on MSC differentiation in vitro, and were hence abstracted to be a sub-database: donor species, culture medium, supplement and growth factor, culture dimension (2D/3D), substrate/scaffold, and differentiation fate. After the sub-database is analyzed using CMAR (see Figure 3), some meaningful rules that abstract known protocols for culturing MSC were obtained. For example:

**Rule # 154:** {human + ascorbic acid + insulin + TGF-β} => {chondro} [96.42%] reports the frequent use of ascorbic acid, insulin, and TGF-β when inducing human MSC into chondrogenesis (the percentage in the bracket is the confidence of the rule). Rules like this could be beneficial for educational purposes as they summarize our knowledge in lab.

On the other hand, some interesting rules suggesting hypotheses that have not been clearly raised and concluded in previous literature were obtained. For example:

**Rule # 188:** {transferrin + selenous acid + dexamethasone} => {chondro} [91.17%] suggests that dexamethasone,

transferrin, and selenous acid altogether may contribute to chondrogenesis. Rules like this, not yet validated, can be used to suggest new research directions that could reveal the underlying intracellular pathways in MSCs, and hence are vital to MSC study and tissue engineering.

The reader is reminded that all the rules obtained through data mining are extracted from patterns present in the experimental data.

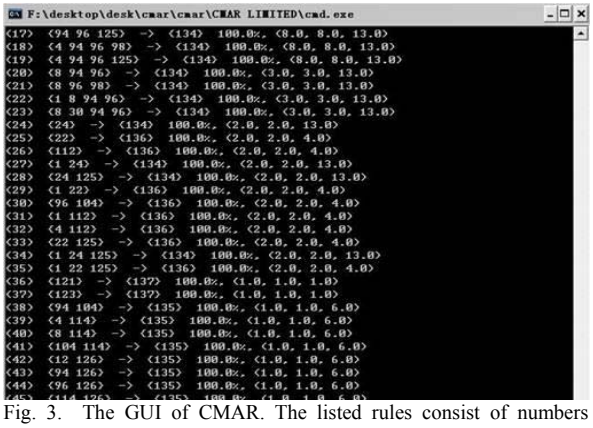


Fig. 3. The GUI of CMAR. The listed rules consist of numbers corresponding to the parameters in the MSC database, which can be decoded to a human readable format, such as the examples above.

For data analysis using ARM as the second application, data on cell marker expression were abstracted from the online database and added into the sub-database. Rules relating chemical reagents and cell marker expression, as well as showing co-expression patterns of cell markers were obtained. For example:

**Rule # 126**: {heparin sulfate} => {CD105 (SH2) + CD90 (Thy-1) - CD45} [100.0%] suggests that heparin sulfate plays a role in the expression of CD105 and CD90, but not CD45;

**Rule # 130:** {CD105 (SH2)} => {CD90 (Thy-1)} [95.0%] suggests that CD105 is often co-expressed with CD90 on the MSC surface, which means that one may not have to test the presence of CD90 if the presence of CD105 has been detected. Rules like these have the potential to save considerable time and money in the laboratory.

Furthermore, such rules as above may be significant in the understanding and prediction of cell marker expression on MSC, and have enormous potential to initiate new research directions in the area of MSC. With the database expanded in future, an increasing contribution to MSC research and tissue engineering is expected.

## III. Conclusions

In this study, a public-domain database containing the key parameters that influence MSC differentiation and proliferation has been built. All the data in this online database is obtained from the published literature and can be freely accessed by stem cell researchers worldwide. Registered users can also conveniently input their data to expand this database. Rules obtained from the data analysis using data mining techniques have shown one application of this database, demonstrating the significance of this database to further research in MSC and tissue engineering.


## Acknowledgements

The authors would like to thank Prof. Jian Lu from the School of Physics & Astronomy at the University of Manchester, and colleagues from the Department of Engineering Science at the University of Oxford.


## References


[1] S. L. Beeres et al., J Am Coll Cardiol, vol. 46, pp. 1943-52, Nov 15 2005.

[2] A. R. Derubeis and R. Cancedda, Ann Biomed Eng, vol. 32, pp. 160-5, Jan 2004.

[3] R. S. Tuan, G. Boland, and R. Tuli, Arthritis Res Ther, vol. 5, pp. 32-45, 2003.

[4] M. F. Pittenger et al., Science, vol. 284, pp. 143-7, Apr 2 1999.

[5] B. A. Roelen and P. Dijke, J Orthop Sci, vol. 8, pp. 740-8, 2003.

[6] A. M. Mackay et al., Tissue Engineering, vol. 4, pp. 415-428, Win 1998.

[7] N. Jaiswal, S. E. Haynesworth, A. I. Caplan, and S. P. Bruder, Journal of Cellular Biochemistry, vol. 64, pp. 295-312, Feb 1997.

[8] J. M. Gimble, S. Zvonic, Z. E. Floyd, M. Kassem, and M. E. Nuttall, Journal of Cellular Biochemistry, vol. 98, pp. 251-266, May 15 2006.

[9] I. Muller et al., Cytotherapy, vol. 8, pp. 437-44, 2006.

[10] M. Krampera et al., Blood, vol. 101, pp. 3722-3729, May 1 2003.

[11] L. G. Griffith and M. A. Swartz, Nat Rev Mol Cell Biol, vol. 7, pp. 211-24, Mar 2006.

[12] A. I. Caplan, Journal of Orthopaedic Research, vol. 9, pp. 641-650, Sep 1991.

[13] P. Bianco, M. Riminucci, S. Gronthos, and P. G. Robey, Stem Cells, vol. 19, pp. 180-92, 2001.

[14] S. M. O'Connell et al., Wound Repair and Regeneration, vol. 16, pp. 749-756, Nov-Dec 2008.

[15] V. L. Battula et al., Differentiation, vol. 75, pp. 279-91, Apr 2007.

[16] K. Hanada, J. E. Dennis, and A. I. Caplan, J Bone Miner Res, vol. 12, pp. 1606-14, Oct 1997.

[17] D. P. Lennon et al., Exp Cell Res, vol. 219, pp. 211-22, Jul 1995.

[18] T. Magaki, K. Kurisu, and T. Okazaki, Neurosci Lett, vol. 384, pp. 282-7, Aug 26 2005.

[19] N. Meuleman et al., Eur J Haematol, vol. 76, pp. 309-16, Apr 2006.

[20] F. Celotti et al., Wound Repair and Regeneration, vol. 14, pp. 195-202, Mar-Apr 2006.

[21] B. Johnstone et al., Experimental Cell Research, vol. 238, pp. 265-272, Jan 10 1998.

[22] S. E. Haynesworth, M. A. Baber, and A. I. Caplan, J Cell Physiol, vol. 166, pp. 585-92, Mar 1996.

[23] S. A. Kuznetsov, A. J. Friedenstein, and P. G. Robey, Br J Haematol, vol. 97, pp. 561-70, Jun 1997.

[24] J. M. Curran, R. Chen, and J. A. Hunt, Biomaterials, vol. 27, pp. 4783-4793, Sep 2006.

[25] J. E. Barralet et al., Journal of Materials Science-Materials in Medicine, vol. 16, pp. 515-519, Jun 2005.

[26] A. Erices, P. Conget, and J. J. Minguell, British Journal of Haematology, vol. 109, pp. 235-242, Apr 2000.

[27] Y. Zhang et al., Exp Hematol, vol. 32, pp. 657-64, Jul 2004.

[28] C. H. Liu, M. L. Wu, and S. M. Hwang, Biochemical Engineering Journal, vol. 33, pp. 1-9, Jan 2007.

[29] M. H. Wu et al., Biomedical Microdevices, vol. 10, pp. 309-319, Apr 2008.

[30] F. Coenen, LUCS KDD implementation of CMAR (Classification based on Multiple Association Rules). http://www.csc.liv.ac.uk/~frans/KDD/Software/CMAR/cmar.html, Department of Computer Science, The University of Liverpool, UK, 2004.

[31] F. Coenen, The LUCS-KDD Apriori-T Association Rule Mining Algorithm, http://www.cxc.liv.ac.uk/~frans/KDD/Software/Apriori_T/aprioriT.html, Department of Computer Science, The University of Liverpool, UK., 2004.