

Application of Classification Association Rule Mining for Mammalian Mesenchymal Stem Cell Differentiation

Wei qi Wang¹, Yanbo J. Wang², René Bañares-Alcántara¹, Zhanfeng Cui¹,
and Frans Coenen³

¹ Department of Engineering Science, Parks Road, University of Oxford, OX1 3PJ, UK
{weiqi.wang, rene.banares, zhanfeng.cui}@eng.ox.ac.uk

² Information Management Center, China Minsheng Banking Corp., Ltd.
Room 606, Building No. 8, 1 Zhongguancun Nandajie, Beijing, 100873, China
wangyanbo@cmbc.com.cn

³ Department of Computer Science, University of Liverpool
Ashton Building, Ashton Street, Liverpool, L69 3BX, UK
Coenen@liverpool.ac.uk

Abstract. In this paper, data mining is used to analyze the differentiation of mammalian Mesenchymal Stem Cells (MSCs). A database comprising the key parameters which, we believe, influence the destiny of mammalian MSCs has been constructed. This paper introduces Classification Association Rule Mining (CARM) as a data mining technique in the domain of tissue engineering and initiates a new promising research field. The experimental results show that the proposed approach performs well with respect to the accuracy of (classification) prediction. Moreover, it was found that some rules mined from the constructed MSC database are meaningful and useful.

Keywords: Classification Association Rule Mining, Data Mining, Differentiation, Mesenchymal Stem Cells, Tissue Engineering.

1 Introduction

Mesenchymal Stem Cells (MSCs) have been claimed to be an integral part of tissue engineering due to their proliferation and differentiation potential both in vivo and in vitro [5, 12, 36], and have become one of the most significant research topics in the past several decades. MSCs are able to differentiate along the osteogenic, chondrogenic, adipogenic, myogenic, tendonogenic, neurogenic lineages under appropriate stimuli [28, 31, 33] generating bone, cartilage, fat, muscle, tendon, neuron cells respectively (Fig. 1). The significance of the application of MSCs in clinical therapy has triggered an urgent need for computational prediction of MSC differentiation [14].

A large amount of studies have been carried out with the aim of understanding the mechanisms involved in MSCs' proliferation and differentiation both in vivo and in vitro [4, 17, 21, 25, 26, 27]. However, little has been achieved so far due to the enormous complexity of the intracellular pathways in MSCs, especially during their differentiation process [6].

On the other hand, the experiments and studies which have been executed were not interrelated with each other, i.e. different experiments focused on different combinations of parameters affecting MSC differentiation, including animal species, in vitro vs. in vivo cultures, MSC culture medium, supplements to the culture medium and growth factors, culture type (monolayer vs. 3D culture), cell attaching substrate (for monolayer culture) vs. scaffold (for 3D culture) and, in particular, the differentiation fates of MSCs in terms of the different lineages to which the cells committed [17, 18, 20, 21, 27]. The scattered experimental data results in a large amount of noise in the database and a discrete data structure, which cannot take advantage of traditional mathematical modelling methods.

For this reason, we aim to classify the data according to the different cell fates and predict mammalian MSC differentiation. A series of computational techniques have been compared, and data mining has been found to be a promising method due to its ability of processing discrete and noisy data. In particular, the data mining classification approach known as Classification Association Rule Mining (CARM) [10] is used in this study.

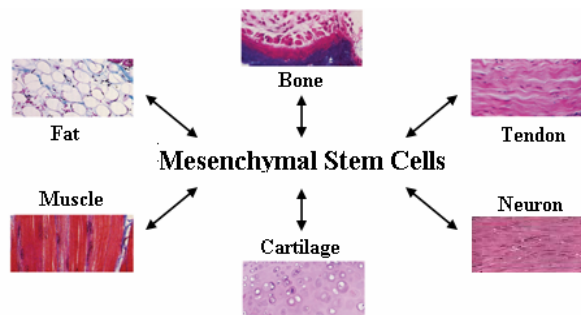


Fig. 1. Differentiation potential of MSCs (modified from [33])

The rest of this paper is organized as follows. In section 2 we describe some related work relevant to this study. Section 3 describes the application of CARM for mammalian MSC differentiation, and the construction of a domain-specific class-transactional database for this study. Experimental results, which demonstrate that the proposed approach performs well with respect to the accuracy of (classification) prediction, are presented in section 4. In section 5, we present our conclusions and discuss open issues for further research.

2 Related Work

2.1 Classification Rule Mining

Classification Rule Mining (CRM) [23, 29] is a technique for identifying Classification Rules (CRs) from a given class database D_c , the objective being to build a classifier to categorize “unseen” data records. Generally D_c is described by a relational database table that includes a class attribute – whose values are a set of predefined

class labels $C = \{c_1, c_2, \dots, c_{|C|-1}, c_{|C|}\}$. The process of CRM consists of two stages: (i) a training phase where CRs are generated from a set of training data instances $D_R \subset D_C$; and (ii) a test phase where “unseen” instances in a test data set $D_E \subset D_C$ are assigned into predefined class groups. A D_C is established as $D_R \cup D_E$, where $D_R \cap D_E = \emptyset$. Both D_R and D_E share the same database attributes except the class attribute. By convention the last attribute in each D_R record usually indicates the predefined class of this record, noted as the class attribute, while the class attribute is missing in D_E .

Mechanisms on which CRM algorithms have been based include: decision trees [29], naive Bayes [24], K -Nearest Neighbor (K -NN) [19], Support Vector Machine (SVM) [7], etc.

- **Decision Tree Induction:** In this approach CRs are mined based on a greedy algorithm. The approach can be separated into two stages. In the first stage the tree is constructed from D_R and this is followed by a tree pruning phase. In the second stage the pruned tree is used in CR generation. C4.5 [29] is the best known decision tree based CRM method and operates by recursively splitting D_R on the attribute that produces the *maximum information gain* to generate the decision tree. This tree is then pruned according to an error estimate. The result is used to classify “unseen” data records.
- **Naive Bayes:** The typical mechanism found in Bayesian CRM approaches such as [13] is naive Bayes [24], which has been widely applied in machine learning. The general idea of naive Bayes is to make use of knowledge of the probabilities involving attribute values and classes in the training dataset to produce a model of a machine learning application that can then be applied to “unseen” data. The term naive is used to refer to the assumption that the conditional probability of a database attribute value given a class is independent of the conditional probability of other attribute values given that class. A naive Bayes classifier [30] is built using D_R , and comprises a set of conditional probabilities for each database attribute and each class $c_i \in C$, so that there are $n \times |C|$ conditional probabilities, where n represents the number of attributes in D_R and $|C|$ is the size function (cardinality) of C . A naive Bayes classifier also comprises a set of prior class probabilities, one for each class. All these probabilities are then used to classify “unseen” data records in D_E according to Bayes’ theorem.
- **K -Nearest Neighbor:** K -NN [19] is a well-known statistical approach used in CRM, and classifies an “unseen” data record $d'_{j'} \in D_E$, by assigning to that record the most frequent class in the set of the K most similar instances to $d'_{j'}$ identified in D_R . To identify the K most similar training-instances for $d'_{j'}$, calculation of the Euclidean distance value between each training data record $d_j \in D_R$ and $d'_{j'}$ is commonly used:

$$\text{distance}(d_j, d'_{j'}) = \sqrt{(\sum_{k=1..n} (d_{j,k} - d'_{j',k})^2)}, \quad (1)$$

where $d_{j,k}$ and $d'_{j',k}$ are the values of the k -th data attribute in D_C for d_j and $d'_{j'}$.

- **Support Vector Machine:** The objective of using SVM [7] is to find a hypothesis \hat{h} which minimizes the *true error* defined as the probability that \hat{h} produces an erroneous result. SVM makes use of linear functions of the form:

$$f(x) = w^T x + b, \quad (2)$$

where w is the weight vector, x is the input vector, and $w^T x$ is the inner product between w and x . The main concept of SVM is to select a *hyperplane* that separates the positive and negative examples while maximizing the smallest margin. Standard SVM techniques produce binary classifiers as opposed to multi-classifiers. Two common approaches to support the application of SVM techniques to the multi-class problem are One Against All (OAA) and One Against One (OAO).

2.2 Association Rule Mining

Association Rule Mining (ARM), first introduced in [1], aims to extract a set of Association Rules (ARs) from a given transactional database D_T . An AR describes an implicative co-occurring relationship between two sets of binary-valued transactional database attributes (items), expressed in the form of an “antecedent \Rightarrow consequent” rule. Cornelis et al. [11] suggest that the concept of mining ARs can be dated back to the work in 1960’s [15].

More generally, we define ARM as follows. Let $I = \{a_1, a_2, \dots, a_{n-1}, a_n\}$ be a set of items, and $\mathcal{F} = \{T_1, T_2, \dots, T_{m-1}, T_m\}$ be a set of transactions (data records), a transactional database D_T is described by \mathcal{F} , where each $T_j \in \mathcal{F}$ comprises a set of items $I' \subseteq I$. In ARM, two threshold values are usually used to determine the significance of an AR:

1. **Support:** A set of items S is called an itemset. The support of S is the proportion of transactions T in \mathcal{F} for which $S \subseteq T$. If the support of S exceeds a user-supplied support threshold σ , S is defined to be a frequent itemset.
2. **Confidence:** Represents how “strongly” an itemset X implies another itemset Y , where $X, Y \subseteq I$ and $X \cap Y = \emptyset$. A confidence threshold α , supplied by the user, is used to distinguish high confidence ARs from low confidence ARs.

An AR $X \Rightarrow Y$ is said to be *valid* when the support for the co-occurrence of X and Y exceeds σ , and the confidence of this AR exceeds α . The computation of support is:

$$\text{support}(X \cup Y) = \text{count}(X \cup Y) / |\mathcal{F}|, \quad (3)$$

where $\text{count}(X \cup Y)$ is the number of transactions containing the set $X \cup Y$ in \mathcal{F} , and $|\mathcal{F}|$ is the size function (cardinality) of the set \mathcal{F} . The computation of confidence is:

$$\text{confidence}(X \Rightarrow Y) = \text{support}(X \cup Y) / \text{support}(X). \quad (4)$$

Informally, “ $X \Rightarrow Y$ ” can be interpreted as: if X is found in a transaction, it is likely that Y also will be found.

2.3 Classification Association Rule Mining

An overlap between ARM and CRM is CARM (Classification Association Rule Mining), which strategically solves the traditional CRM problem by applying ARM techniques. The idea of CARM, first introduced in [2], aims to extract a set of

Classification Association Rules (CARs) from a class-transactional database D_{C-T} . Let D_T be a transactional database, and $C = \{c_1, c_2, \dots, c_{|C|-1}, c_{|C|}\}$ be a set of predefined class labels, D_{C-T} is described by $D_T \times C$. D_{C-T} can also be defined as a special class database D_C , where all database attributes and the class attribute are valued in a binary manner – “*Boolean attributes can be considered a special case of categorical attributes*” [32]. A CAR is a special AR that describes an implicative co-occurring relationship between a set of binary-valued data attributes and a predefined class, expressed in the form of an “ $X \Rightarrow c_i$ ” rule, where X is an itemset found in D_T (as “ $D_{C-T} - C$ ”) and c_i is a predefined class in C .

2.4 Advantages of CARM

CARM offers the following advantages with respect to the CRM techniques mentioned above [3, 35]:

1. The approach is efficient during both the training and categorization phases, especially when handling a large volume of data.
2. The classifier built in this approach can be read, understood and modified by humans.

Furthermore, CARM is relatively insensitive to noise data. CARM builds a classifier by extracting a set of CARs from a given set of training instances. Possible CARs are determined by: (i) a large enough support, and (ii) a large enough confidence. Usually, rules derived from noise in the data will fail to reach these thresholds and will be discarded.

In comparison, CRM approaches other than CARM, i.e. naive Bayes, K -NN, SVM, etc., do not present the classifier in a human readable fashion, so that users do not see why the (classification) predictions have been made. Rules generated by a decision tree based classifier can be read and understood by human; however, Coenen et al. [9] suggest that results presented in the studies of [22] and [23] show that in many cases CARM offers higher classification accuracy than decision tree based classification.

For these reasons it is proposed to use a CARM approach to address the prediction of mammalian MSC differentiation. One of the existing CARM frameworks is the CMAR (Classification based on Multiple Association Rules) algorithm [22]. CMAR generates CARs (from a given set of training instances) through an FP-tree [16] based approach. Experimental results using this algorithm reported in [22] show that it could achieve high classification accuracy for a range of data sets.

3 Database Construction for Mammalian MSC Differentiation

In order to make a (classification) prediction of mammalian MSC differentiation using CARM, a domain-dependent database containing 375 parameters that are believed to influence the MSC differentiation has been built and can be accessed online¹. The parameters in this database include the most significant ones, such as donor species, in vitro vs. in vivo culture, culture medium, supplements and growth factors, culture

¹ <http://www.oxford-tissue-engineering.org/forum/plugin.php?identifier=publish&module=publish>

type (monolayer vs. 3D culture), substrate (for monolayer culture) vs. scaffold (for 3D culture), MSC differentiation fate, as well as other potentially important parameters including age of donor, cell passage number, concentrations of chemicals and buffer, cell seeding density, incubation duration, pH value, expression of cell markers, expansion fold of cell number, etc. All the records are abstracted from previously published papers and each is stated clearly with the corresponding reference.

The current size of this database is 203 records, each containing attributes including experimental conditions and results, i.e. rule antecedents and rule consequent-classes respectively. There are four types of attributes in the database:

- For the qualitative parameters which have only two possible values, such as the presence/absence of insulin 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 integer symbols.
- For the quantitative parameters, such as cell seeding density, the attributes are stored as real numbers.
- For the descriptive parameters, such as additional conditions, the attributes are stored as text/characters.

3.1 Data Filtering

The database contains a large variety of information, among which some parameters are previously known to be more important and effective to the differentiation fates than others. Those parameters and their significance are listed in Table 1.

Other parameters, such as buffer, gas condition, pH value, etc., seem to be just for facilitating the differentiation. We are also aware that procedures such as radiation,

Table 1. The most significant parameters in the MSC database

Parameters	Significance/Description
Donor Species	Same culture conditions on MSCs from different species of mammal may lead to different results.
In Vitro / In Vivo	MSC differentiation varies significantly from in vivo to in vitro environment.
Culture Medium	Most essential environment where MSCs grow, proliferate and differentiate.
Supplements & Growth Factors	Chemicals that maintain MSC differentiation potential or influence their differentiation fate.
Culture Type (2D vs. 3D)	MSC differentiation sometimes differs significantly from monolayer to 3D culture, even under the same culture medium and supplements.
Substrate (for 2D) / Scaffold (for 3D)	Influences cell viability. A chemically modified substrate can even change MSCs' differentiation fate.
Differentiation Fate	The most significant and obvious result after cell culture. Used as the classes in the database and the prediction in this study.

centrifugation and mechanical stimulation sometimes affect MSCs, and that MSCs from bone marrow are sometimes different from those from other sources, e.g. umbilical cord blood. However, in order to identify plausible MSC differentiation rules the essential parameters considered were those in Table 1; other parameters could be excluded if the purpose were to predict the MSC differentiation fate, according to the “unseen” rules.

3.2 Data Normalization and Cleaning

In this study, only the parameters listed in Table 1 were extracted from the MSC database and used for the (classification) prediction. Consequently, the number of attributes in the abstracted database was reduced from 375 to 105. The database was then discretised and normalised using the LUCS-KDD Discretised Normalised (DN) software², so that data are presented in a binary format suitable for use by CARM applications. It should be noted that the database was re-arranged so that occurrences of classes (i.e. osteogenesis, chondrogenesis, etc.) were distributed evenly throughout the database. This then allowed CMAR to be applied to (90% – training set, 10% – test set) divisions of the database with Ten-fold Cross Validation (TCV) accuracy setting. In this study, the discretisation and normalisation process results in a data file with the number of attributes increased to 173.

This discretised and normalised data file contains a number of noisy data, generally caused by the absence of culture conditions such as some growth factors. For example, if the growth factor insulin is absent in a record, this record will have an attribute representing “absence of insulin” after the discretisation and normalisation process. However, the rules that we are looking for in this study are those without the information of “absence”, i.e. those only containing the information of “presence”. Thus, all the attributes representing “absence” were then eliminated from the data file, resulting in the final input data file for CARM. In this final input file, all the records cover five classes in total, i.e. five kinds of MSC fates: “osteogenesis”, “chondrogenesis”, “adipogenesis”, “neurogenesis”, and “proliferation without differentiation”.

4 Results

In this section, we aim to evaluate the usage of CARM for mammalian MSC differentiation. The evaluation was performed using the CMAR algorithm although any other CARM classifier generator could equally well have been used. Experiments were run on a 2.00 GHz Intel(R) Core(TM)2 CPU with 2.00 GB of RAM running under Windows Command Processor. The evaluation undertaken used a confidence threshold value (ϕ) of 50% and a support threshold value (σ) of 1% (as previously used in published CARM evaluations [8, 9, 22, 34]). The overall (classification) prediction accuracy is 77.04%.

There were 163 CMAR rules generated from the input data file, among which many are found to be meaningful and useful. Two rules are described as an example, with the actual confidence values presented in square brackets:

² http://www.csc.liv.ac.uk/~frans/KDD/Software/LUCS-KDD-DN/lucs-kdd_DN.html

1. **Rule # 49:** {in vitro + monolayer + human donor + DMEM + TGFβ1 + plastic substrate} ⇒ {chondrogenesis} [100.0%], which can be interpreted as: in monolayer culture in vitro, human MSCs are most likely to undergo chondrogenesis in the presence of cell culture medium DMEM (Dulbecco’s Modified Eagle’s Medium) and growth factor TGFβ1 (Transforming Growth Factor β1), on plastic substrate (Fig. 2).

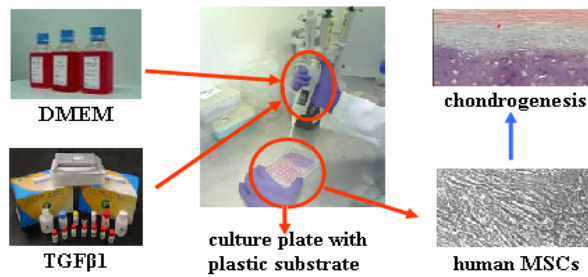


Fig. 2. Description of rule # 49

2. **Rule # 86:** {DMEM + FBS + ascorbate-2-phosphate + Dex} ⇒ {osteogenesis} [93.33%], which can be interpreted as: in DMEM medium supplemented with FBS (Fetal Bovine Serum), MSCs are very likely to be induced to osteogenesis under the stimuli of ascorbate-2-phosphate and Dex (Dexamethasone) together (Fig. 3).

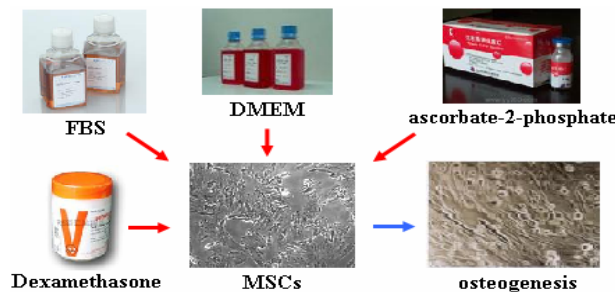


Fig. 3. Description of rule # 86

5 Conclusions

In this study, the (classification) prediction of MSC differentiation has been achieved using CARM, with an accuracy of 77.04%. We have introduced, for the first time, data mining techniques to mesenchymal stem cell, initiating a new promising research field. We are aware, nevertheless, that there is room to increase the (classification) prediction accuracy, as well as to streamline the pruning of generated rules by improving the CARM technique or using alternative mechanisms. Additionally, many of

the identified rules were found to be already known in the laboratory. However, the identification of completely original rules is possible and expected if the size and contents of the MSC database are properly expanded in the future.

Acknowledgments. The authors would like to thank the following colleagues from the Department of Engineering Science at the University of Oxford for their valuable suggestions to this study: Paul Raju, Norazharuddin Shah Abdullah, Prof. James Triffitt, Nuala Trainor, Clarence Yapp, Dr. Cathy Ye, Dr. Xia Xu, Dr. Uday Tirlapur, Yang Liu, Zhiqiang Zhao, Dr. Shengda Zhou and Dr. Renchen Liu.

References

1. Agrawal, R., Imielinski, T., Swami, A.: Mining Association Rules between Sets of Items in Large Database. In: Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, Washington, D.C., USA, pp. 207–216. ACM Press, New York (1993)
2. Ali, K., Manganaris, S., Srikant, R.: Partial Classification using Association Rules. In: Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining, Newport Beach, CA, USA, pp. 115–118. AAAI Press, Menlo Park (1997)
3. Antonie, M.L., Zaiane, O.R.: Text Document Categorization by Term Association. In: Proceedings of the 2002 IEEE International Conference on Data Mining, Maebashi City, Japan, pp. 19–26. IEEE Computer Society Press, Los Alamitos (2002)
4. Battula, V.L., Bareiss, P.M., Treml, S., Conrad, S., Albert, I., Hojak, S., Abele, H., Schewe, B., Just, L., Skutella, T., Buhring, H.J.: Human Placenta and Bone Marrow derived MSC Cultured in Serum-free, b-FGF-containing Medium Express Cell Surface Frizzled-9 and SSEA-4 and Give Rise to Multilineage Differentiation. *Differentiation* 75, 279–291 (2007)
5. Beeres, S.L., Atsma, D.E., van der Laarse, A., Pijnappels, D.A., van Tuyn, J., Fibbe, W.E., de Vries, A.A.F., Ypey, D.L., van der Wall, E.E., Schalijs, M.J.: Human Adult Bone Marrow Mesenchymal Stem Cells Repair Experimental Conduction Block in Rat Cardiomyocyte Cultures. *American College of Cardiology* 46(10), 1943–1952 (2005)
6. Bianco, P., Riminucci, M., Gronthos, S., Robey, P.G.: Bone Marrow Stromal Stem Cells: Nature, Biology, and Potential Applications. *Stem Cells* 19, 180–192 (2001)
7. Boser, B.E., Guyon, I.M., Vapnik, V.N.: A Training Algorithm for Optimal Margin Classifiers. In: Proceedings of the 5th ACM Annual Workshop on Computational Learning Theory, Pittsburgh, PA, USA, pp. 144–152. ACM Press, New York (1992)
8. Coenen, F., Leng, P.: An Evaluation of Approaches to Classification Rule Selection. In: Proceedings of the 4th IEEE International Conference on Data Mining, Brighton, UK, pp. 359–362. IEEE Computer Society Press, Los Alamitos (2004)
9. Coenen, F., Leng, P., Zhang, L.: Threshold Tuning for improved Classification Association Rule Mining. In: Ho, T.-B., Cheung, D., Liu, H. (eds.) PAKDD 2005. LNCS, vol. 3518, pp. 216–225. Springer, Heidelberg (2005)
10. Coenen, F., Leng, P.: The Effect of Threshold Values on Association Rule based Classification Accuracy. *Journal of Data and Knowledge Engineering* 60(2), 345–360 (2007)
11. Comelis, C., Yan, P., Zhang, X., Chen, G.: Mining Positive and Negative Association Rules from Large Databases. In: Proceedings of the 2006 IEEE International Conference on Cybernetics and Intelligent Systems, Bangkok, Thailand, pp. 613–618. IEEE Computer Society Press, Los Alamitos (2006)

12. Derubeis, A.R., Cancedda, R.: Bone Marrow Stromal Cells (BMSCs) in Bone Engineering: Limitations and Recent Advances. *Annals of Biomedical Engineering* 32(1), 160–165 (2004)
13. Domingos, P., Pazzani, M.: On the Optimality of the Simple Bayesian Classifier under Zero-one Loss. *Machine Learning* 29(2/3), 103–130 (1997)
14. Griffith, L.G., Swartz, M.A.: Capturing Complex 3D Tissue Physiology in Vitro. *Nature Reviews Molecular Cell Biology* 7, 211–224 (2006)
15. Hajek, P., Havel, I., Chytil, M.: The GUHA Method of Automatic Hypotheses Determination. *Computing* 1, 293–308 (1966)
16. Han, J., Pei, J., Yin, Y.: Mining Frequent Patterns without Candidate Generation. In: *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data*, Dallas, TX, USA, pp. 1–12. ACM Press, New York (2000)
17. Hanada, K., Dennis, J.E., Caplan, A.I.: Stimulatory Effects of Basic Fibroblast Growth Factor and Bone Morphogenetic Protein-2 on Osteogenic Differentiation of Rat Bone Marrow-derived Mesenchymal Stem Cells. *Journal of Bone and Mineral Research* 12, 1606 (1997)
18. Haynesworth, S.E., Baber, M.A., Caplan, A.I.: Cytokine Expression by Human Marrow-derived Mesenchymal Progenitor Cells in Vitro: Effects of Dexamethasone and IL-1 α . *Journal of Cell Physiology* 166(3), 585–592 (1996)
19. James, M.: *Classification Algorithm*. Wiley Interscience, New York (1985)
20. Kuznetsov, S.A., Friedenstein, A.J., Robey, P.G.: Factors Required for Bone Marrow Stromal Fibroblast Colony Formation in Vitro. *British Journal of Haematology* 97, 561–570 (1997)
21. Lennon, D.P., Haynesworth, S.E., Young, R.G., Dennis, J.E., Caplan, A.I.: A Chemically defined Medium Supports in Vitro Proliferation and Maintains the Osteochondral Potential of Rat Marrow-derived Mesenchymal Stem Cells. *Experimental Cell Research* 219, 211–222 (1995)
22. Li, W., Han, J., Pei, J.: CMAR: Accurate and Efficient Classification based on Multiple Class-association Rules. In: *Proceedings of the 2001 IEEE International Conference on Data Mining*, San Jose, CA, USA, pp. 369–376. IEEE Computer Society Press, Los Alamitos (2001)
23. Liu, B., Hsu, W., Ma, Y.: Integrating Classification and Association Rule Mining. In: *Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining*, New York, USA, pp. 80–86. AAAI Press, Menlo Park (1998)
24. Lowd, D., Domingos, P.: Naive Bayes Models for Probability Estimation. In: *Proceedings of the 22nd International Conference on Machine Learning*, Bonn, Germany, pp. 529–536. ACM Press, New York (2005)
25. Magaki, T., Kurisu, K., Okazaki, T.: Generation of Bone Marrow-derived Neural Cells in Serum-free Monolayer Culture. *Neuroscience Letters* 384, 282–287 (2005)
26. Meuleman, N., Tondreau, T., Delforge, A., Dejefeffe, M., Massy, M., Libertalis, M., Bron, D., Lagneaux, L.: Human Marrow Mesenchymal Stem Cell Culture: Serum-free Medium Allows Better Expansion than Classical α -MEM Medium. *European Journal of Haematology* 76(4), 309–316 (2006)
27. Muller, I., Kordowich, S., Holzwarth, C., Spano, C., Isensee, G., Staiber, A., Viebahn, S., Gieseke, F., Langer, H., Gawaz, M.P., Horwitz, E.M., Conte, P., Handgretinger, R., Dominici, M.: Animal Serum-free Culture Conditions for Isolation and Expansion of Multipotent Mesenchymal Stromal Cells from Human BM. *Cytotherapy* 8, 437–444 (2006)
28. Pittenger, M.F., Mackay, A.M., Beck, S.C., Jaiswal, R.K., Douglas, R., Mosca, J.D., Moorman, M.A., Simonetti, D.W., Craig, S., Marshak, D.R.: Multilineage Potential of Adult Human Mesenchymal Stem Cells. *Science* 284(5411), 143–147 (1999)

29. Quinlan, J.R.: C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, San Mateo (1993)
30. Rish, I.: An Empirical Study of the Naive Bayes Classifier. In: Proceedings of the 2001 IJCAI Workshop on Empirical Methods in Artificial Intelligence, Seattle, WA, USA (2001)
31. Roelen, B.A., Dijke, P.: Controlling Mesenchymal Stem Cell Differentiation by TGFbeta Family Members. *Journal of Orthopaedic Science* 8, 740–748 (2003)
32. Srikant, R., Agrawal, R.: Mining Quantitative Association Rules in Large Relational Tables. In: Proceedings of the 1996 ACM SIGMOD International Conference on Management of Data, Montreal, Quebec, Canada, pp. 1–12. ACM Press, New York (1996)
33. Tuan, R.S., Boland, G., Tuli, R.: Adult Mesenchymal Stem Cell and Cell-based Tissue Engineering. *Arthritis Research & Therapy* 5, 32–45 (2003)
34. Wang, Y.J., Xin, Q., Coenen, F.: A Novel Rule Ordering Approach in Classification Association Rule Mining. In: Perner, P. (ed.) *MLDM 2007*. LNCS, vol. 4571, pp. 339–348. Springer, Heidelberg (2007)
35. Yoon, Y., Lee, G.G.: Practical Application of Associative Classifier for Document Classification. In: Lee, G.G., Yamada, A., Meng, H., Myaeng, S.-H. (eds.) *AIRS 2005*. LNCS, vol. 3689, pp. 467–478. Springer, Heidelberg (2005)
36. Zhang, Y., Li, C., Jiang, X., Zhang, S., Wu, Y., Liu, B., Tang, P., Mao, N.: Human Placenta-derived Mesenchymal Progenitor Cells Support Culture Expansion of Long-term Culture-initiating Cells from Cord Blood CD34+ Cells. *Experimental Hematology* 32, 657–664 (2004)