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

Speaker: Yanbo J. Wang

Information Management Center, China Minsheng Banking Corp., Ltd.

Co-work with: Weiqi Wang, René Bañares-Alcántara, Zhanfeng Cui

Department of Engineering Science, University of Oxford

Frans Coenen

Department of Computer Science, University of Liverpool

icdm

Industrial Conference on Data Mining

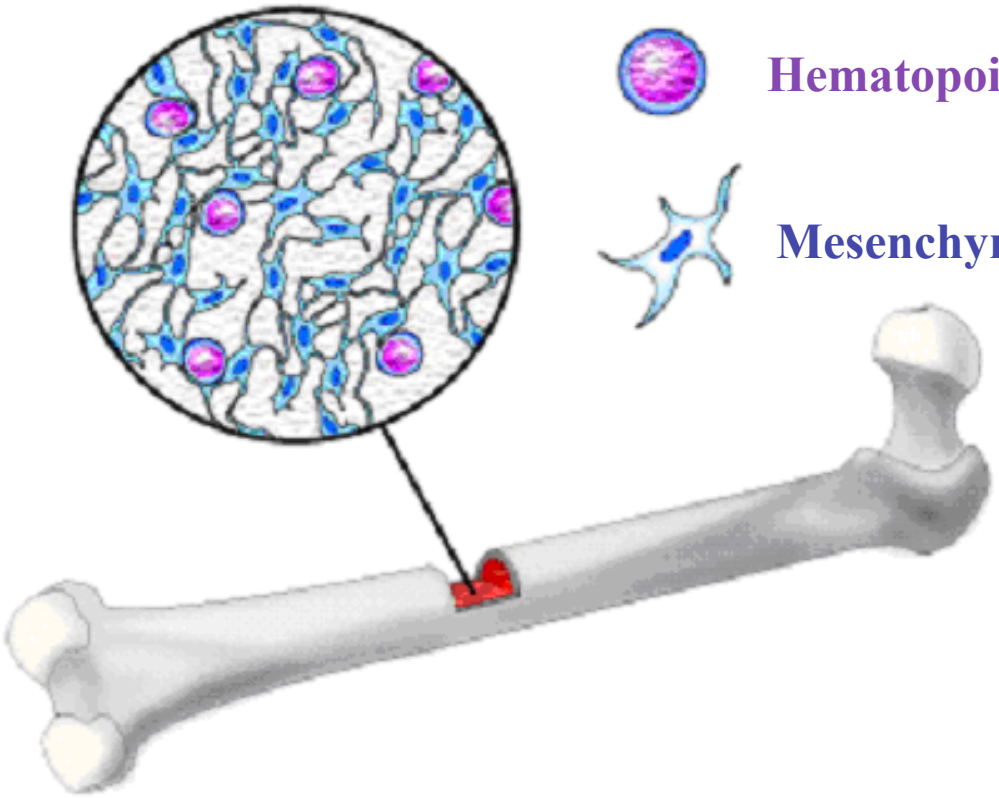
ICDM'2009 ▪ Leipzig ▪ Germany ▪ July 2009

OUTLINE

- **BACKGROUND**
  - Bone Marrow Stem Cells
  - Mesenchymal Stem Cell Investigation
  - Mesenchymal Stem Cell Differentiation
- **MOTIVATION**
- **CLASSIFICATION ASSOCIATION RULE MINING**
- **PROCEDURES**
  - Online MSC Database
  - Data Processing
- **RESULTS**
  - Classification Accuracy
  - Interesting Rules
- **CONCLUSION & FUTURE WORK**

BACKGROUND

Bone Marrow Stem Cells

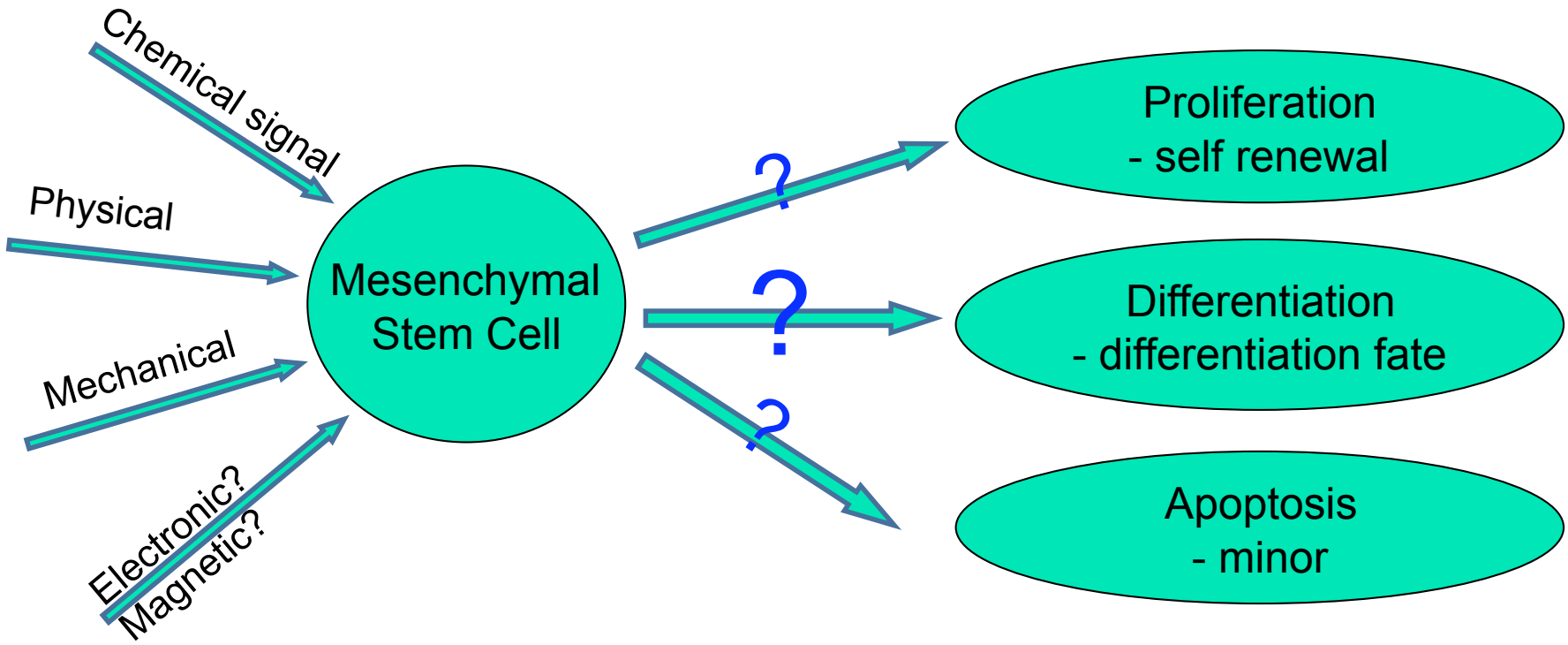


Hematopoietic Stem Cell (HSC)

Mesenchymal Stem Cell (MSC)

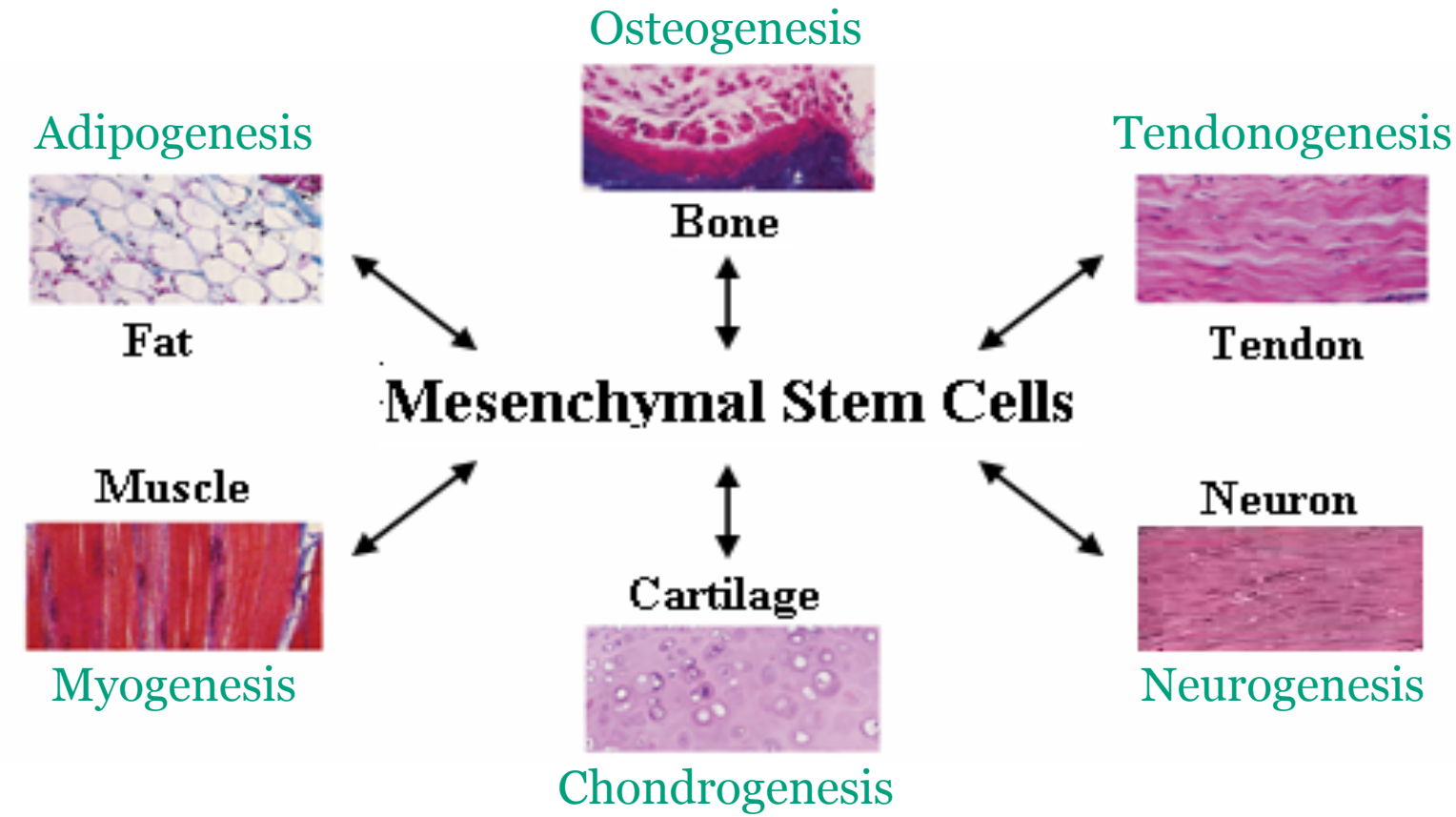
In our study, we focus on MSC investigation

Mesenchymal Stem Cell Investigation



Our study is mainly concentrated on MSC differentiation

Mesenchymal Stem Cell Differentiation



Differentiation Potential (Fate) of MSCs

MOTIVATION

Motivation

- We are interested in finding the way how MSC to be differentiated.
- The scattered data on this (MSC) study is available online – we can extract and collect such data from online research/academic repositories, e.g. MEDLINE.
- In general, such differentiation problem can be simply modelled as a Classification problem in Data Mining.
- Our study is concerned with the *single-label* Classification task – assigning exactly one predefined class (differentiation fate) to each "unseen" (MSC) data record.
- There are many Classification approaches/mechanisms available, i.e. Artificial Neural Network, Support Vector Machine, Naive Bayes, K-Nearest Neighbour, Classification Association Rule Mining, etc.

CLASSIFICATION ASSOCIATION RULE MINING

Classification Association Rule Mining

- In our study, we select to use the Classification Association Rule Mining (CARM) approach.
- CARM offers the following advantages:
  - The approach is efficient during both the training and categorisation phases, especially when handling a large volume of data.
  - The classifier built in this approach can be read, understood and modified by humans, whereas other classifiers cannot.
  - CARM is relatively insensitive to noise data.
  - In previous studies, CARM was reported to offer good classification accuracy.
- CARM strategically solves the traditional Classification problem by applying Association Rule Mining (ARM) techniques.

continue...

- CARM aims to extract a set of Classification Association Rules (CARs) from a class-transactional database $D_{C\text{-}T}$. Let D_T be a (binary-valued) transactional database, and $C = \{c_1, c_2, \ldots, c_{|C|-1}, c_{|C|}\}$ be a set of predefined class labels, $D_{C\text{-}T}$ is described by $D_T \times C$.
- A CAR 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 and c_i is a predefined class in C.
- A CAR is said to be *valid* when the support of X and c_i exceeds a user supplied *minsupp* (support threshold), and the confidence of this CAR exceeds a user supplied *minconf* (confidence threshold).
  - support $(X \cup c_i)$ = count $(X \cup c_i$ in $D_{C\text{-}T}) / |D_{C\text{-}T}|$.
  - confidence $(X \Rightarrow c_i)$ = support $(X \cup c_i)$ / support (X) .

PROCEDURES

Online MSC Database

- A domain-dependent database containing total 375 parameters that are believed to influence the MSC differentiation has been built and can be accessed online at "http://www.oxford-tissue-engineering.org/forum/plugin.php?identifier=publish&module=publish".
- Each record in the MSC database that describes a (real-life) experiment of MSC differentiation, was read, extracted and collected from such research/academic papers (note that paper reference is also recorded in the database).
- The size of this database was 203 records (as reported in the conference paper), and now it has been increased up to 501 records.
- In this database, the key (most significant) parameters include: in vivo/vitro, culture medium, supplement, monolayer/3D culture, substrate/scaffold, cell seeding density, result, etc.

OUCEG Academic Forum

USER CP

WELCOME TO MSC DATABASE weiqi

- ADD A RECORD
- MY RECORDS
- MY MESSAGES[0]
- DATABASE CP

NOTICING BOARD

This MSC database is for academic use. When you add data, please should you make them as correct and accurate as possible.

Unpublished data are welcome at any time; however, you may receive emails from the webmaster for double-checking when you mark your data as "unpublished".

Meanwhile, we highly appreciate if you could provide as much publishing information as possible if your data are abstracted from publications. Many thanks for your kind cooperation!

NOTE: All the current data are from in vitro experiments. If you would like to add any in vivo data, please contact the webmaster and the website will be upgraded accordingly.

QUERY

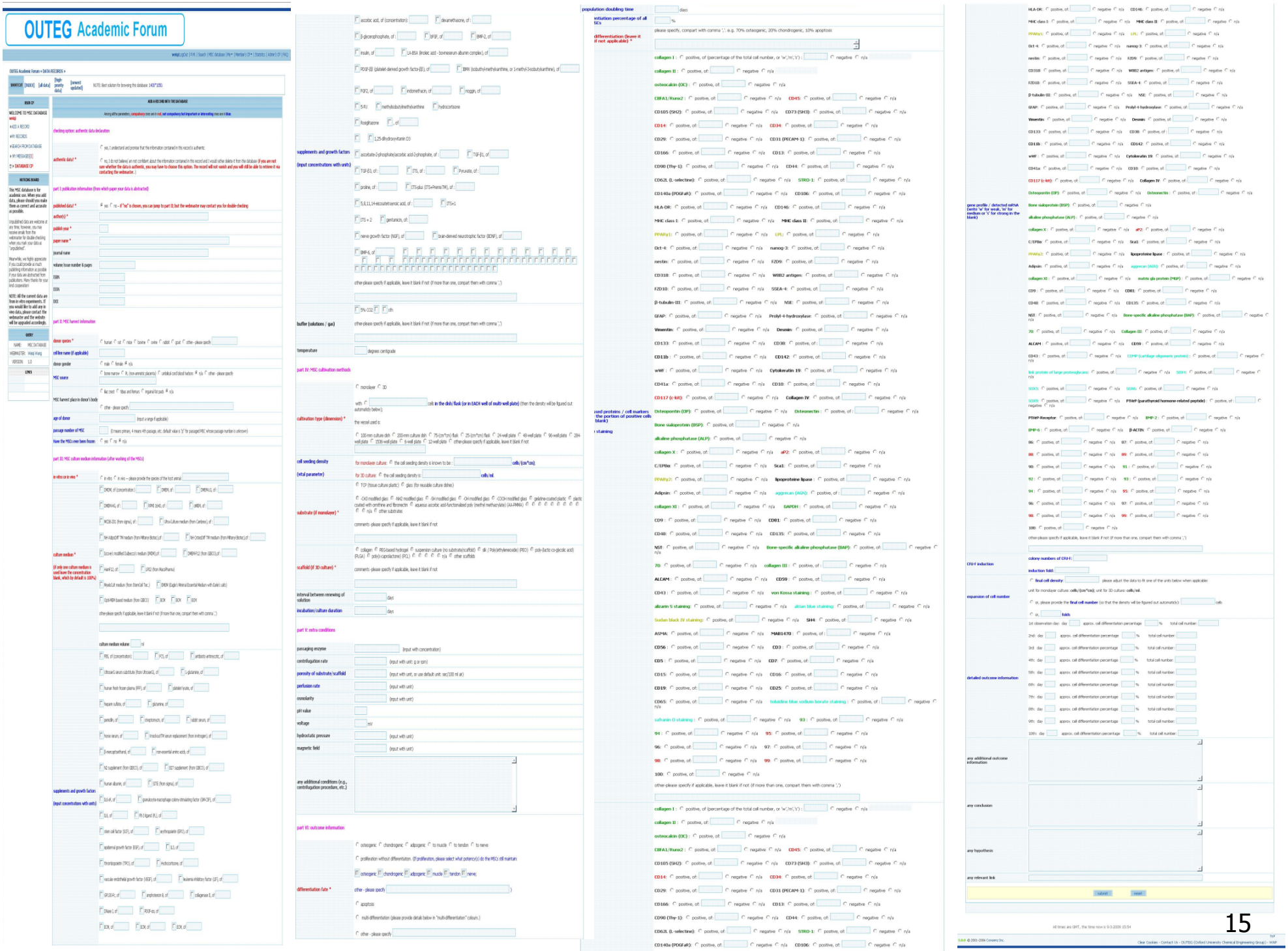
NAME:	MSC DATABASE
WEBMASTER:	Weiqi Wang
VERSION:	1.0

LINKS

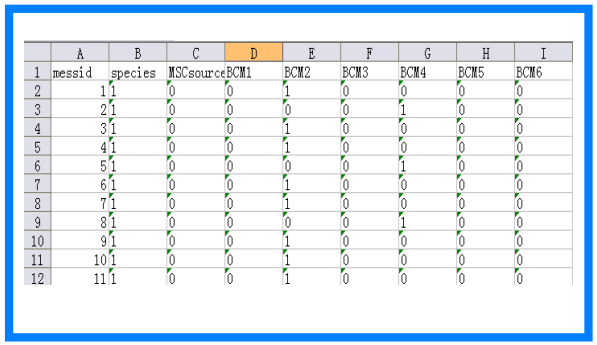
[all data] number of all records: 103, high-priority: 103, updated within 10 days: 0

- The information shown here is just a small part of each record. Click the record number to see all details of each record.

record no.	high-priority ?	published ?	authors	publish year	paper name	MSC species	BCM	supplement	cultivation type	differentiation type	added by	update time	
00006	yes	yes	A. Alhadlaq and J.J. Mao	2003	Tissue-engineered Neogenesis of Human-shaped Mandibular Condyle from Rat Mesenchymal Stem Cells	rat	DMEM-LG,	10% FBS, 1% antibiotic-antimycotic, 10 ng/ml TGF-β1,	3D	chondrogenic	weiqi	11-12-2007	edit
00040	yes	yes	S. Gronthos et al.	1994	The STRO-1+ Fraction of Adult Human Bone Marrow Contains the Osteogenic Precursors	human	αMEM,	10% FCS, 100 nM dexamethasone, 100 μM L-ascorbate-2-phosphate, 2 mM L-glutamine, 1.8 mM KH2PO4,	monolayer	osteogenic	weiqi	10-12-2007	edit
00041	yes	yes	S. Gronthos et al.	1994	The STRO-1+ Fraction of Adult Human Bone Marrow Contains the Osteogenic Precursors	human	αMEM,	10% FCS, 2 mM L-glutamine,	monolayer	proliferation without differentiation	weiqi	10-12-2007	edit
00039	yes	yes	S. Gronthos et al.	1994	The STRO-1+ Fraction of Adult Human Bone Marrow Contains the Osteogenic Precursors	human	αMEM,	10% FCS, 100 nM dexamethasone, 100 μM L-ascorbate-2-phosphate, 2 mM L-glutamine, 1.8 mM KH2PO4, 20 mM HEPES,	monolayer	osteogenic	weiqi	10-12-2007	edit
00038	yes	yes	S. Gronthos et al.	1994	The STRO-1+ Fraction of Adult Human Bone Marrow Contains the Osteogenic Precursors	human	αMEM,	20% FCS, 2 mM L-glutamine, 50 μM β-mercaptoethanol,	monolayer	osteogenic	weiqi	10-12-2007	edit
00037	yes	yes	Gronthos et al.	2003	Molecular and cellular characterisation of highly purified stromal stem cells derived from human bone marrow	human	αMEM,	10% FCS, 100 nM dexamethasone, 100 μM L-ascorbate-2-phosphate, 3 mM inorganic phosphate,	monolayer	osteogenic	weiqi	10-12-2007	edit
00036	yes	yes	S. Gronthos et al.	1994	The STRO-1+ Fraction of Adult Human Bone Marrow Contains the Osteogenic Precursors	human	αMEM,	20% FCS, 100 μM L-ascorbate-2-phosphate, 2 mM L-glutamine, 50 μM β-mercaptoethanol,	monolayer	osteogenic	weiqi	10-12-2007	edit
00035	yes	yes	Gronthos et al.	2003	Molecular and cellular characterisation of highly purified stromal stem cells derived from human bone marrow	human	αMEM,	10% FCS, 10 ng/ml TGF-β3,	monolayer	chondrogenic	weiqi	10-12-2007	edit
00034	yes	yes	Gronthos et al.	2003	Molecular and cellular characterisation of highly purified stromal stem cells derived from human bone marrow	human	αMEM,	10% FCS, 0.5 mM methylisobutylmethylxanthine, 0.5 μM hydrocortisone, 60 μM indomethacin,	monolayer	adipogenic	weiqi	10-12-2007	edit
00033	yes	yes	PAULETTE A. CONGET et al.	2001	Identification of a Discrete Population of Human Bone Marrow-Derived Mesenchymal Cells Exhibiting Properties of Uncommitted Progenitors	human	αMEM,	10% FBS, 50 μg/ml ascorbate-2-phosphate/ascorbic acid-2-phosphate, 100 nM dexamethasone, 10 mM β-glycerophosphate,	monolayer	osteogenic	weiqi	10-12-2007	edit
00032	yes	yes	PAULETTE A. CONGET et al.	2001	Identification of a Discrete Population of Human Bone Marrow-Derived Mesenchymal Cells Exhibiting Properties of Uncommitted Progenitors	human	αMEM,	10% FBS, 1 μM dexamethasone, 100 μg/ml 3-isobutyl-1-methylxanthine , 5 μg/ml insulin, 60 mM indomethacin,	monolayer	adipogenic	weiqi	10-12-2007	edit
00031	yes	yes	PAULETTE A. CONGET et al.	2001	Identification of a Discrete Population of Human Bone Marrow-Derived Mesenchymal Cells Exhibiting Properties of Uncommitted Progenitors	human	αMEM,	10% FBS, 0.6 mg/ml 5-FU,	monolayer	to be determined	weiqi	10-12-2007	edit
00030	yes	yes	PAULETTE A. CONGET et al.	2001	Identification of a Discrete Population of Human Bone Marrow-Derived Mesenchymal Cells Exhibiting Properties of Uncommitted Progenitors	human	αMEM,	20% FBS, none	monolayer	proliferation without differentiation	weiqi	10-12-2007	edit
00029	yes	yes	PAULETTE A. CONGET et al.	2001	Identification of a Discrete Population of Human Bone Marrow-Derived Mesenchymal Cells Exhibiting Properties of Uncommitted Progenitors	human	αMEM,	10% FBS, none	monolayer	proliferation without differentiation	weiqi	10-12-2007	edit
00028	yes	yes	Timothy R. Brazelton, et al.	2000	From Marrow to Brain: Expression of Neuronal Phenotypes in Adult Mice	mice	none	nonenone	3D	to nerve	weiqi	10-12-2007	edit
00027	yes	yes	oreffo et al.	1999	Human Bone Marrow Osteoprogenitors Express Estrogen Receptor-Alpha and Bone Morphogenetic Proteins 2 and 4 mRNA During Osteoblastic Differentiation	human	αMEM,	10% FCS, 1% antibiotic-antimycotic, none	monolayer	proliferation without differentiation	weiqi	10-12-2007	edit
00026	yes	yes	oreffo et al.	1999	Human Bone Marrow Osteoprogenitors Express Estrogen Receptor-Alpha and Bone Morphogenetic Proteins 2 and 4 mRNA During Osteoblastic Differentiation	human	αMEM,	10% FCS, 1% antibiotic-antimycotic, none	monolayer	proliferation without differentiation	weiqi	10-12-2007	edit
00025	yes	yes	oreffo et al.	1999	Human Bone Marrow Osteoprogenitors Express Estrogen Receptor-Alpha and Bone Morphogenetic Proteins 2 and 4 mRNA During Osteoblastic Differentiation	human	αMEM,	10% FCS, 1% antibiotic-antimycotic, none	monolayer	proliferation without differentiation	weiqi	10-12-2007	edit
00024	yes	yes	oreffo et al.	1999	Human Bone Marrow Osteoprogenitors Express Estrogen Receptor-Alpha and Bone Morphogenetic Proteins 2 and 4 mRNA During Osteoblastic Differentiation	human	αMEM,	10% FCS, 1% antibiotic-antimycotic, none	monolayer	proliferation without differentiation	weiqi	10-12-2007	edit
00023	yes	yes	oreffo et al.	1999	Human Bone Marrow Osteoprogenitors Express Estrogen Receptor-Alpha and Bone Morphogenetic Proteins 2 and 4 mRNA During Osteoblastic Differentiation	human	αMEM,	10% FCS, 1% antibiotic-antimycotic, 50 μg/ml ascorbate-2-phosphate/ascorbic acid-2-phosphate, 10 nM dexamethasone,	monolayer	osteogenic	weiqi	10-12-2007	edit


OUTEG Academic Forum
OUTEG Academic Forum » DATA RECORDS »
SEARCH
[high-priority data]
[lowest updated]
[all data]
USER CP
WELCOME TO MSC DATABASE
MY RESEARCH
DATABASE CP
NOTICING BOARD
QUERY
LINKS
ADD A RECORD INTO THE DATABASE
checking option: authentic data declaration
part I: publication information (from which paper your data is abstracted)
part II: MSC harvest information
part III: MSC culture medium information (after washing of the MSCs)
part IV: MSC cultivation methods
part V: extra conditions
part VI: outcome information
supplements and growth factors
(input concentrations with units)
buffer (solutions / gas)
temperature
cultivation type (dimension) *
cell seeding density
(vital parameter)
substrate (if monolayer) *
scaffold (if 3D culture) *
culture medium *
serum or in vivo *
passaging enzyme
centrifugation rate
porosity of substrate/scaffold
perfusion rate
osmolarity
pH value
voltage
hydrostatic pressure
magnetic field
any additional conditions (e.g., centrifugation procedure, etc.)
differentiation fate *
population doubling time
differentiation (leave it if not applicable) *
gene profile / detected mRNA (write 'w' for weak, 'm' for medium or 's' for strong in the blank)
CFU-F induction
expansion of cell number
detailed outcome information
any additional outcome information
any conclusion
any hypothesis
any relevant link
submit
reset

Data Processing



	A	B	C	D	E	F	G	H	I
1	messid	species	MSCsource	BCM1	BCM2	BCM3	BCM4	BCM5	BCM6
2	1	1	0	0	1	0	0	0	0
3	2	1	0	0	0	0	1	0	0
4	3	1	0	0	1	0	0	0	0
5	4	1	0	0	1	0	0	0	0
6	5	1	0	0	0	0	1	0	0
7	6	1	0	0	1	0	0	0	0
8	7	1	0	0	1	0	0	0	0
9	8	1	0	0	0	0	1	0	0
10	9	1	0	0	1	0	0	0	0
11	10	1	0	0	1	0	0	0	0
12	11	1	0	0	1	0	0	0	0

Original data in MSC database

```
(1) species = {1}
(2) species = {2}
(3) MSCsource = {0}
(4) MSCsource = {1}
(5) MSCsource = {2}
(6) MSCsource = {3}
(7) BCM1 = {0}
(8) BCM2 = {1}
(9) BCM2 = {0}
(10) BCM3 = {1}
(11) BCM3 = {0}
(12) BCM4 = {1}
(13) BCM4 = {0}
(14) BCM5 = {1}
(15) BCM5 = {0}
(16) BCM6 = {1}
(17) BCM6 = {0}
```

Schema

Normalised data

```
1 3 7 8 11 13 15 17 19 21 23 25 27 29 30 33 35 37 39
1 3 7 8 11 13 15 17 19 21 23 25 27 29 30 33 35 37 39
1 3 7 8 11 13 15 17 19 21 23 25 27 29 30 33 35 37 39
1 3 7 8 11 13 15 17 19 21 23 25 27 29 30 33 35 37 39
1 3 7 8 11 13 15 17 19 21 23 25 27 29 30 33 35 37 39
1 3 7 8 11 13 15 17 19 21 23 25 27 29 30 33 35 37 39
1 3 7 8 11 13 15 17 19 21 23 25 27 29 30 33 35 37 39
1 3 7 9 11 12 15 17 19 21 23 25 27 29 30 33 35 37 39
2 4 7 9 10 13 15 17 19 21 23 25 27 28 31 32 35 37 39
2 4 7 9 10 13 15 17 19 21 23 25 27 28 31 32 35 37 39
```

De-noise

```
1 6 26 64 66 70 125 138
1 4 20 32 34 36 125 138
1 4 20 32 34 36 125 138
1 6 26 64 66 90 125 138
1 8 30 94 96 98 127 134
1 5 8 36 54 56 58 100 125 138
2 4 10 28 125 138
2 4 10 18 32 104 106 108 125 138
2 4 46 48 60 62 100 133 138
```

Input data

CARM

```
Number of CMAR rules    = 78
(1)  {6}  ->  {138}  100.0%, (28.0, 28.0, 75.0)
(2)  {26}  ->  {138}  100.0%, (28.0, 28.0, 75.0)
(3)  {1 6}  ->  {138}  100.0%, (28.0, 28.0, 75.0)
(4)  {1 26}  ->  {138}  100.0%, (28.0, 28.0, 75.0)
(5)  {1 100}  ->  {138}  100.0%, (21.0, 21.0, 75.0)
(6)  {1 100 125}  ->  {138}  100.0%, (13.0, 13.0, 75.0)
(7)  {132}  ->  {138}  100.0%, (12.0, 12.0, 75.0)
(8)  {1 28}  ->  {138}  100.0%, (12.0, 12.0, 75.0)
(9)  {28 46}  ->  {138}  100.0%, (12.0, 12.0, 75.0)
(10)  {28 48}  ->  {138}  100.0%, (12.0, 12.0, 75.0)
(11)  {1 132}  ->  {138}  100.0%, (12.0, 12.0, 75.0)
(12)  {36 132}  ->  {138}  100.0%, (12.0, 12.0, 75.0)
(13)  {1 28 46}  ->  {138}  100.0%, (12.0, 12.0, 75.0)
(14)  {54}  ->  {138}  100.0%, (11.0, 11.0, 75.0)
(15)  {56}  ->  {138}  100.0%, (11.0, 11.0, 75.0)
(16)  {4 94 96}  ->  {134}  100.0%, (8.0, 8.0, 13.0)
(17)  {94 96 125}  ->  {134}  100.0%, (8.0, 8.0, 13.0)
(18)  {4 94 96 98}  ->  {134}  100.0%, (8.0, 8.0, 13.0)
(19)  {4 94 96 125}  ->  {134}  100.0%, (8.0, 8.0, 13.0)
(20)  {8 94 96}  ->  {134}  100.0%, (3.0, 3.0, 13.0)
(21)  {8 96 98}  ->  {134}  100.0%, (3.0, 3.0, 13.0)
(22)  {1 8 94 96}  ->  {134}  100.0%, (3.0, 3.0, 13.0)
(23)  {8 30 94 96}  ->  {134}  100.0%, (3.0, 3.0, 13.0)
(24)  {24}  ->  {134}  100.0%, (2.0, 2.0, 13.0)
(25)  {22}  ->  {136}  100.0%, (2.0, 2.0, 4.0)
```

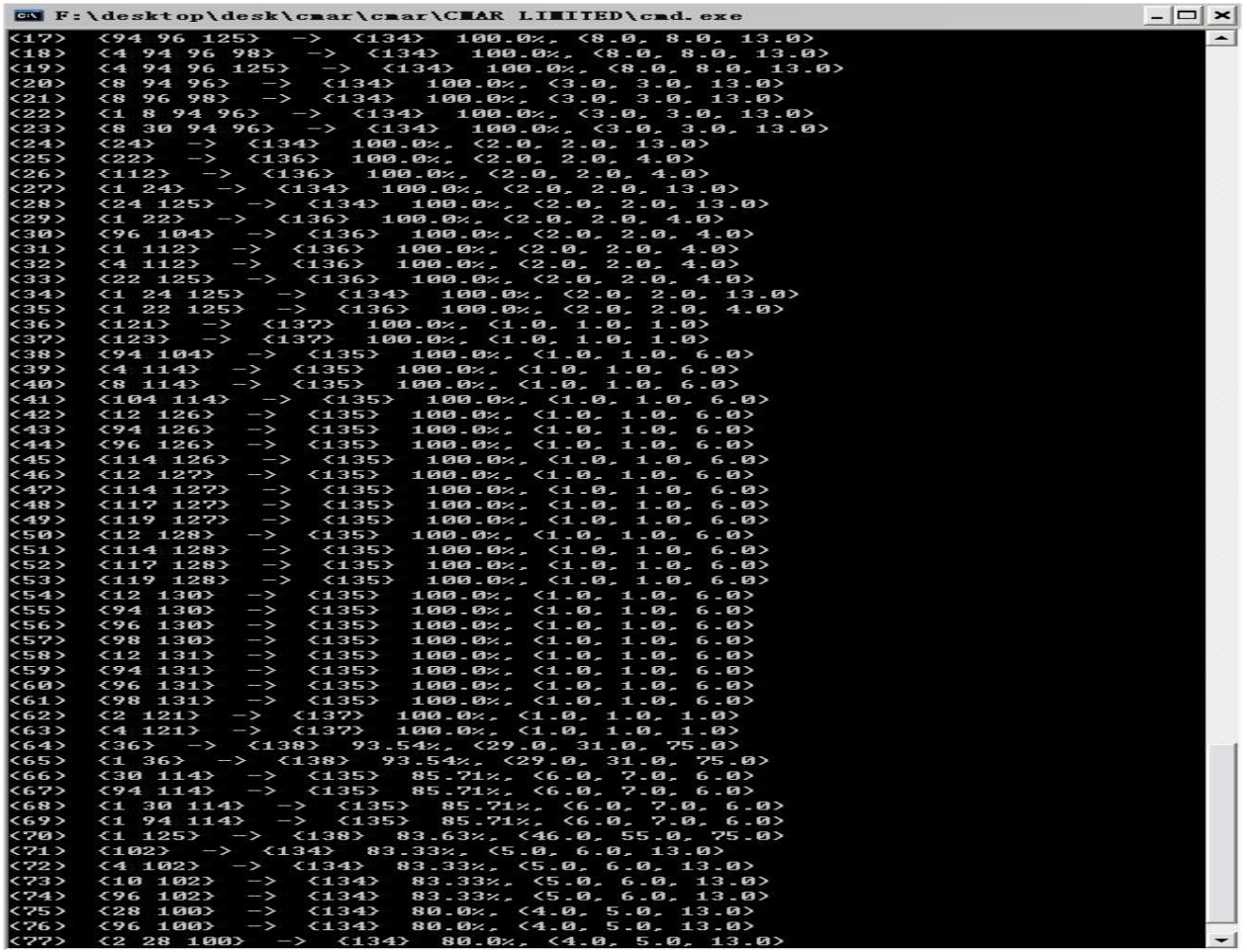
Generated rules

RESULTS

Classification Accuracy

- Experiments were run on a 2.00 GHz Intel(R) Core(TM)2 CUP with 2.00 GB of RAM running under Windows Command Processor.
- The evaluation was performed using the CMAR (Classification based on Multiple Association Rules) algorithm although any other CARM classifier generator (i.e. CBA, CPAR, TFPC, etc.) could equally well have been used. The CMAR software can be download from "http://www.csc.liv.ac.uk/~frans/KDD/Software/CMAR/cmar.html".
- The evaluation undertaken used a support threshold value (*minsupp*) of 1% and a confidence threshold value (*minconf*) of 50%.
- The evaluation was performed with the Ten-fold Cross Validation (TCV) accuracy setting.
- The classification (prediction) accuracy was 77.04% with 203 data records (as reported in the conference paper).
- The up-to-dated classification accuracy is 90.4% with 501 data records.

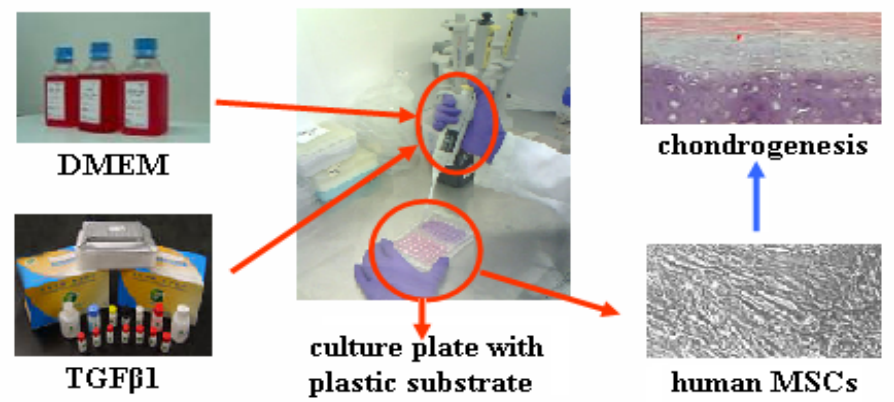
Interesting Rules



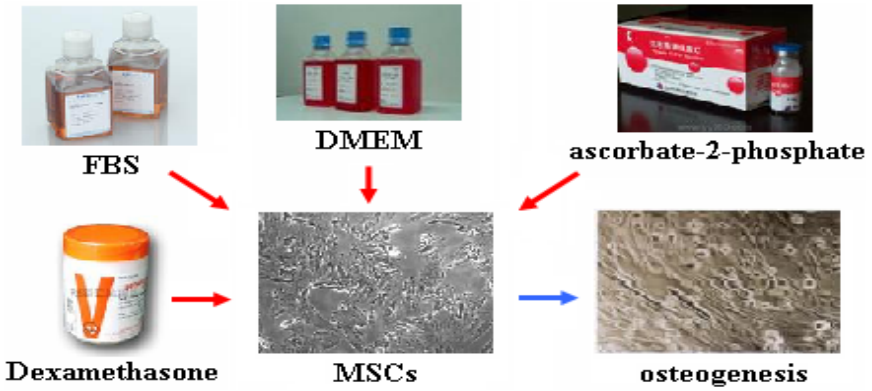
Rules that are generated by the CMAR software

continue...

With regard to a particular fold in the TCV process, there were 163 CMAR rules generated from the input data, which is around 182 data records (203 records × 90% of the database).



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.



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.

continue…

Based on the total 501 data records, the new training dataset consists of around 450 data records (501 records × 90% of the database). There were 295 CMAR rules generated, among which many are found to be interesting.

Interesting rules

Rule # 27: {pyruvate + proline} ⇒ {chondro} [100.0%]

- Pyruvate: important in metabolic pathways, may potentially help promote MSC chondrogenesis.
- Proline: a catalyst in biochemical reactions, may facilitate chondrogenesis.
- Not 100% sure, but we just did not realise them yet!

Rule # 188: {transferrin + selenous acid + dexamethasone} ⇒ {chondro} [91.17%]

- Transferrin: participate in the immune system, prevent bacteria from survival, not contrary for transferrin to support chondrogenesis.
- Selenous acid: highly toxic, usually fatal, needs further investigation.

DISCOVER NEW RULES
RAISE OPEN QUESTION TO INVESTIGATE

CONCLUSION & FUTURE WORK

Conclusion and Future Work

- CARM is a promising method to discover rules involved in MSC differentiation.
- The classification accuracy of this approach is good.
- Some rules have been found to be interesting, and need further investigation.

- In the future, we would like to continuously increase/expand the size of the MSC database.
- For this (MSC differentiation) study, we also want to compare the performance among various CARM algorithms, i.e. CMAR, CBA, CPAR, TFPC, etc.

THANK YOU!

We are the most grateful if you can share your MSC data online!

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