





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

Speaker:Yanbo J. WangInformation 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



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



In our study, we focus on MSC investigation





Mesenchymal Stem Cell Investigation Chemical signal **Proliferation** - self renewal Physical **Mesenchymal** Differentiation Mechanical Stem Cell - differentiation fate **Apoptosis** - minor

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.

icdm







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-T} . Let D_T be a (binary-valued) transactional database, and $C = \{c_1, c_2, ..., c_{|C|-1}, c_{|C|}\}$ be a set of predefined class labels, D_{C-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 $(X \cup c_i) = \text{count} (X \cup c_i \text{ in } D_{C-T}) / |D_{C-T}|$.
 - confidence $(X \Rightarrow c_i) = \text{support} (X \cup c_i) / \text{support} (X)$.

icdm









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

weigi:LogOut | P.M. | Search | MSC database | My + | Member's CP + | Statistics | Admin's CP | FAQ -

	data	j'	updated		Se Solation	The biomsking this database. These reserve	ompacible							
USER CP	[all dat	a] nur	mber of all r	ecords: 103,	high-prio	rity: 103, updated within 10 days: 0								
VELCOME TO MSC DATABASE	- The i	nformati	on shown h	ere is just a	small par	t of each record. Click the record number t	o see all d	etails of e	ach record.					
ADD A RECORD	record	high- priority ?	published ?	authors	publish year	paper name	MSC species	BCM	supplement	cultivation type	differentiation type	added by	update time	
MY RECORDS				А.		Tissue-engineered Neogenesis of								
MY MESSAGES[0]	00006	yes	yes	Alhadlaq and J.J.	2003	Human-shaped Mandibular Concyle from	rat	DMEM- LG,	10% FBS, 1% antibiotic-antimycotic, 10 ng/ml TGF-B1,	ЗD	chondrogenic	weiqi	11-12- 2007	1000
DATABASE CP				Mao		Rat Mesenchymai Stem Cells								
NOTICING BOARD	00040	yes	yes	S. Gronthos et al.	1994	The STRO-1+ Fraction of Adult Human Bone Marrow Contains the Osteogenic Precursors	human	aMEM,	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	2002
nis MSC database is for ademic use. When you add ata, please should you make	00041	yes	yes	S. Gronthos et al.	1994	The STRO-1+ Fraction of Adult Human Bone Marrow Contains the Osteogenic Precursors	human	aMEM,	10% FCS, 2 mM L-glutamine,	monolayer	proliferation without differentiation	weiqi	10-12 2007	0.0000
nem as correct and accurate possible. published data are welcome at w time: however, you may	00039	yes	yes	S. Gronthos et al.	1994	The STRO-1+ Fraction of Adult Human Bone Marrow Contains the Osteogenic Precursors	human	aMEM,	10% FCS, 100 nM dexamethasone, 100 µM L-ascorbate-2-phosphate, 2 mM L- glutamine, 1.8 mM KH2PO4, 20 mM HEPPS,	monolayer	osteogenic	weiqi	10-12 2007	1000
ebmaster for double-checking hen you mark your data as	00038	yes	yes	S. Gronthos et al.	1994	The STRO-1+ Fraction of Adult Human Bone Marrow Contains the Osteogenic Precursors	human	aMEM,	20% FCS, 2 mM L-glutamine, 50 μM β- mercaptoethanol,	monolayer	osteogenic	weiqi	10-12 2007	NUM-
anwhile, we highly appreciate you could provide as much	00037	yes	yes	Gronthos et al.	2003	Molecular and cellular characterisation of highly purified stromal stem cells derived from human bone marrow	human	aMEM,	10% FCS, 100 nM dexamethasone, 100 µM L-ascorbate-2-phosphate, 3 mM inorganic phosphate,	monolayer	osteogenic	weigi	10-12 2007	20200
ubilishing information as possible your data are abstracted from ublications. Many thanks for your nd cooperation!	00036	yes	yes	S. Gronthos et al.	1994	The STRO-1+ Fraction of Adult Human Bone Marrow Contains the Osteogenic Precursors	human	aMEM,	20% FCS, 100 μM L-ascorbate-2- phosphate, 2 mM L-glutamine, 50 μM β- mercaptoethanol,	monolayer	osteogenic	weiqi	10-12 2007	No. of Concession, Name
OTE: All the current data are om in vitro experiments. If ou would like to add any in	00035	yes	yes	Gronthos et al.	2003	Molecular and cellular characterisation of highly purified stromal stem cells derived from human bone marrow	human	aMEM,	10% FCS, 10 ng/ml TGF-β3,	monolayer	chondrogenic	weiqi	10-12 2007	10,000
ivo data, please contact the rebmaster and the website rill be upgraded accordingly.	00034	yes	yes	Gronthos et al.	2003	Molecular and cellular characterisation of highly purified stromal stem cells derived from human bone marrow	human	aMEM,	10% FCS, 0.5 mM methylisobutylmethylxanthine, 0.5 μM hydrocortisone, 60 μM indomethacin,	monolayer	adipogenic	weiqi	10-12 2007	NTOTAL OF
QUERY NAME: MSC DATABASE	00033	yes	yes	PAULETTE A. CONGET	2001	Identification of a Discrete Population of Human Bone Marrow-Derived Mesenchymal Cells Exhibiting Properties	human	aMEM,	10% FBS, 50 μg/ml ascorbate-2- phosphate/ascorbic acid-2-phosphate, 100 nM dexamethasone, 10 mM β-	monolayer	osteogenic	weiqi	10-12 2007	ACCOUNT.
VEBMASTER: Weiqi Wang VERSION: 1.0	00032	Ves	Ves	PAULETTE A.	2001	Identification of a Discrete Population of Human Bone Marrow-Derived	human	oMEM.	10% FBS, 1 μM dexamethasone, 100 μα/ml 3-isohutyl-1-methylxanthine , 5	monolaver	adinogenic	weiai	10-12	-
LINKS	00002			et al.		Identification of a Discrete Ropulation of			µg/ml insulin, 60 mM indomethacin,				2007	
	00031	yes	yes	A. CONGET et al.	2001	Mesenchymal Cells Exhibiting Properties of Uncommitted Progenitors	human	aMEM,	10% FBS, 0.6 mg/ml 5-FU,	monolayer	to be determined	weiqi	10-12- 2007	1
	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	aMEM,	20% FBS, none	monolayer	proliferation without differentiation	weiqi	10-12 2007	CARDING .
	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	aMEM,	10% FBS, none	monolayer	proliferation without differentiation	weiqi	10-12 2007	100014
	00028	yes	yes	Timothy R. Brazelton, et al.	2000	From Marrow to Brain: Expression of Neuronal Phenotypes in Adult Mice	mice	none	nonenone	ЗD	to nerve	weiqi	10-12 2007	00000
	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	aMEM,	10% FCS, 1% antibiotic-antimycotic, none	monolayer	proliferation without differentiation	weiqi	10-12 2007	2002
	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	aMEM,	10% FCS, 1% antibiotic-antimycotic, none	monolayer	proliferation without differentiation	weiqi	10-12 2007	Contract of the local data
	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	aMEM,	10% FCS, 1% antibiotic-antimycotic, none	monolayer	proliferation without differentiation	welqi	10-12 2007	ALTERATION OF
	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	aMEM,	10% FCS, 1% antibiotic-antimycotic, none	monolayer	proliferation without differentiation	weiqi	10-12 2007	all second
	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	aMEM,	10% FCS, 1% antibiotic-antimycotic, 50 µg/ml ascorbate-2-phosphate/ascorbic acid-2-phosphate, 10 nM dexamethasone,	monolayer	osteogenic	weiqi	10-12 2007	2002

All times are GMT, the time now is 2-1-2008 23:52

14

TOP

					population doubling time	days		HLAGR: C postre, cf. C reporte C n/a C0146: C postre, cf. C reporte C n/a
OUT	FG Aca	demic Forum		I acobic aot, of (concentration);	sidation percentage of all tics	% please specify, compart with comma ','. e.g. 70% onteogenic, 20% chondrogenic, 10% apoptosis		MHC dass II: C positive, of C nogritive C n/a MHC dass II: C positive, of C negative C n/a
001	LU Aud			Bigleenaphosphate, d : Elfrer, d Elfrer, d	differentiation (leave it if not applicable) *	4		revery: postile, ofC negatile C n/a LPC: C postile, ofC negatile C n/a Dit-E C nontree ofC negatile C n/aC negatile C negatile C negatile C n/a
		weightsplat (PAH) (Seach (MSC database (M)+* (Reminer's CF* (Statutor) Administration	90	🔽 insulin, of 💦 LA-BEA (Indeic add - bowneserun abunin complex), of		collagen 1: C positive, of (percentage of the total coll number, or 'w','m','y'): C negative C n/a		nestia: C postrie, of C regime C n/a 1259: C postrie, of C regime C n/a
UTES Academic Forum + DATA:R	ECORDS +			🔽 PDIF-38 (platekt-denied gowth factor-88), of 📃 🕅 🕅 BMK (pobuthy-methyliamthine), or 1-methyl-3-sobutyliamthine), of		collagen II : C positive, of: C regative C ryla		C0318: C positive, cf. C registre: C rg/s W982 antigen: C positive, cf. C registre: C rg/s
NALEN [1001] [4] 644)	path present sol	12: Best solution for browing the database: 14(11)/1050.				osteocalcin (DC): C postile, of: C negative C n/a		F2D10: C positive, of C registre C n/s 55EA-4: C positive, of C registre C n/s
	áta) (guzea)	AND A DIVISION WITH THE ANTIANAGE		FIGP2, of Findomethach, of Findogen, of		CIEFA1/Rumo2 : C positive, of: C registive C n/a CD45: C positive, of: C registive C n/a		(Prtabalin-III): C postive, at C negative C n/a : NBE: C postive, at C negative C n/a
BER OF BLODME TO MSC DATABASE		Anno dife sensitivo campiono not esti red, and empoieno indirectado e internetino not esti bilas.		540 Enethylicibutylnetholarthine Ehydocartiane		CD105 (SH2): C positive, cf. C negative C n/a CD73 (SH3): C positive, cf. C negative C n/a		GRAP: C postwe, of C negative C n/a _ Probl-4-lacksoglase: C postwe, of C negative C n/a
DD A NECCHD				F Rosigitatione F . of		DD14: C positive, of: C negative C n/a CD34: C positive, of C negative C n/a		Winentin: C positive, of C negative C n/a Desmin: C positive, of C negative C n/a
N NECCHOS	checking option: authentic data de	dantan .		C L25-dhydroxy/tanin 08		CD29: C positive, of: C negative C n/a CD31 (PECAM-1): C positive, of: C negative C n/a		C0137 C poster, st C registre C n/s C0397 C poster, of ; C registre C n/s
ERCH FROMONTARIASE		$^{\rm C}$ yes, I undestand and provide that the information contained in this records auffernit:	supplements and growth factors	T appbale-2-drophite/appbale.org		CD166: C positive, of C registrie C n/a CD13: C positive, of C registrie C n/a		Oblib: C poste, of C nagative C n/a Oblib: C poste, of C nagative C n/a
DATABASE CP	adherik dala?*	C no, 1 do not belene(an not confident about the information contained in this record and) vouid rather delete it from the database (if you are not save whether the databa authentic, you may have to choose this option. The record will not vanish and you will still be able to retrieve it wa	(input concentrations with units)			CD90 (Thy-1): C positive, cf. C negative C n/a CD144: C positive, cf. C negative C n/a		CD41a: C positive, of C registrer C n/a CD10: C positive, of C negative C n/a
NUTCHIC BLARD		costating the webmater.)		F Hydrae, ar:		00621. (L-selectine): C postore, of: C negative C n/a STRO-1: C postore, of: C negative C n/a		CD117 (c-kit): C postive, cf. C registre C n/s Collagen IV: C postive, cf. C registre C n/s
is MSC database is for ademic use. When you add	part 1: publication information (fro	n which paper your data is abstracted)		praine, of:		CD140a (PD6FaR): C postive, of: C negative C n/a CD106: C postive, of: C negative C n/a		Osteopentin (OP): C postive, et. C negative C n/a Osteonectin : C postive, et: C negative C
a, please should you make am as correct and accurate consolite	published data? *	$\theta_{\rm per} \circ \alpha$ - if "to" is chosen, you can jump to part 12, but the webmacter may contact you for clouble-checking		5,8,11,14-eccestetraynoc acid, of :		HLA-DR: C postive, of C negative C n/a CD146: C postive, of C negative C n/a	gene profile / detected mRNA (write 'w' for weak, 'm' for	Bone sialoprotein (BSP): C positio, ch. C negative. C n/a
publified data are welcome at	author(s) *			TTS+2 C gentanicin, of		MHC class E: C positive, of C negative C n/a MHC class II: C positive, of C negative C n/a	medium or s for strong in the bilank)	alkaline phosphultase (ALP): C politive, of C negative C ryla
e trie, rowerer, you nay celve emails from the ebruater for dauble-checking	publish year *			reve gowth factor (NGF), of F brandened neurotropic factor (KCNF), of		PPARy1: C positive, of: C negative C n/a LPL: C positive, of: C negative C n/a		collagen X: C positive, of C negative C n/a #22: C positive, of C negative C n/a
er jou nali jour data as published".	paper name *					Oct-4: C positive, of: C negative C n/a namogr3: C positive, of: C negative C n/a		C/UFBa: C poster, of: C regative C n/a Skat: C poster, of C regative C n/a
nvihle, ve tight approate ou caldigroide as much	volume: issue number & pages					nestin: C positive, of C regative C n/a F209: C positive, of C regative C n/a		defineire C mostries of C mostries C may appropriate types - C positive, at C mostries C may defineire C mostries of C mostries C may
Skring information as possible our data are abstracted from skringers. More therefor the area	ISBN					C0318: C postive, of C negative C n/a W882 antigen: C postive, of C negative C n/a		collegen XI : C postow, of: C negative C n/a matrix gla protein (MSP) : C postow, of: C ne
d caopentioni	1594			other please specify if applicable, leave it blank if not (if more than one, compart them with comme ";")		F2D10: C positive, of: C negative C n/a SSEA-4: C positive, of C negative C n/a		CD9: C postere, of. C regative C n/a CD91: C postere, of C regative C n/a
.E:All the current data are min whto experiments. If as would like to add any in-	001					β-tubulen-tit: C positive, of: C negative C n/a NSE: C positive, of: C negative C n/a		CD48: C postive, of C regative C n/a CD135: C postive, of C regative C n/a
data, please contact the master and the website	nat D ME have to have the			Tak cos in Fish		GitAP: C positive, of: C negative C n/a Probyl-4-hydroxylase: C positive, of: C negative C n/a		NST: C positile, cf. C regizive C n/a Bone-specific alkaline phosphatase (BAP): C positive, cf.
n er upgradett attordingly. Defin	personal construction internation		buffer (solutions / gas)	other-please specify if applicable, leave it blank if not (if more than one, compart them with comma ',')		Venentin: C positive, of: C negative C n/a Desmin: C positive, of C negative C n/a		70: C positive, of C regative C ryla Dollagen III: C positive, of : C regative C ryla
NVE HSCOATIGAE	danar species *	C hunan C sat, C nite, C boine, C sabit, C goat, C other-please-pacify				00138: C postive, of C negative C n/a 0038: C postive, of C negative C n/a		ALDAM: C positive, of C negative C n/a CDS9: C positive, of C negative C n/a
EPRISTER: Weig Wung VERSION: LO	cril line name (if applicable)	Con Carlo Bas	temperature	dagees contigade		CD11b: C positive, of: C negative C n/a CD142: C positive, of: C negative C n/a		CD43: C postere, at C regative C n/a CDMP (cartillage aligometric protein): C postere, of
LINS	danar gender	C tae C tenae M n/s C bore name C R. (co-amétic picenta) C umbical cod block habos. A n/s C other-pices pach	part IV: MSC cultivation methods			wWF: C positive, of C negative C n/a Cytokeratin 19: C positive, of C negative C n/a		Interpretation of large proteophysians: C positive, of: C regative: C ryla 50041; C positive, of: C
	MEC starts			C monolager C 30		CD41a: C postive, of: C negative C n/a CD10: C postive, of: C negative C n/a		nia 1925: C postes of C reades C sia 1920; C postes of C reades C nia
	MEC harvest place in donor's body	C &c cet. C tale and fenue: C ingunal fot pade. R info				CD117 (c-kit): C positive, ef: C negative C n/a Collagen IV: C positive, ef: C negative C n/a		SUX: C posters of C register C n/s PBitP (parathened homone related peptide) : C posters, of :
	- 41	C pter-doos gedy		with the providence of the rest in the	used proteins / cell markers	Disteopentin (OP): C postive, of C negative C n/a Disteomentin : C postive, of C negative C n/a		negates C n/s
	age of dotter	Prote strate if applicable	cultivation type (dimension) *	the vessel used is	blank)	Bone sialoprotein (BSP): C postive, of: C negative C n/a		Partie-secopter: pointe, or regime C na BACIDE C roste of C poste C na
	Have the MECs over been frazen	prime prime prime a particular sector of the sector of the period on the period of the sector of the		C 100 mm auture dah. C 200 mm auture dah. C 75-(cm*cm) faki. C 25-(cm*cm) faki. C 24-wel piate. C 46-wel piate. C 96-wel piate. C 204- wel piate. C 1536-wel piate. C 5-wel piate. C 12-wel piate. C other desse spectly fi sppicable, level it blank if not.	scaring	alkaline phosphatase (ALP): C positive, of: C negative C n/a		86: C postive, pt C regative C n/a 87: C postive, pt C regative C n/a
						collagen X: C postile, of C negatile C n(a 492: C postile, of C negatile C n(a		BB: C positive, at C negative C n/a BP: C positive, at C negative C n/a
	part de rec cator andain rea	aadon jana wadang in da rocuj	cell seeding density	for monoleyer culture: ** the cell seeding density is known to be :		C/EPBax C positive, of C negative C n/a Scal: C positive, of C negative C n/a		90: C positive, of: C negative C n/a 91: C positive, of: C negative C n/a
	in vita or in vita *	C in vito C in vito - please positio the species of the host animal	(vital parameter)	for 30 culture: C the cell seeding density is: cells(in). C 170 these others elastic) C dear life rescales others (their)		PDARy2: C positive, of: C negative C n/a Repopertaine Rease : C positive, of: C negative C n/a		92: C postive, of C regative C n/s 90: C postive, of C regative C n/s
		E DREM, dr (concentration)		 To taxe outerpact, - ges to research use sons) 0.20 moliar date 0.340 moliar date 0.20 moliar date 0.20 moliar date 0.2000 moliar date 0.20000 moliar date 0.2000 moliar date 0.20000 moliar date 0.20000 moliar date 0.20000 moliar date		Adipsin: C positive, of: C regative C n/a aggrecan (ADN): C positive, of: C regative C n/a		94 : C postive, of C negative C n/a 95: C postive, of C negative C n/a
		F (1994), d : F (1994) (1990, d : F (1994), d :	udulada (fi musikum) t	Costed with omthine and Bronetin © a securi accobic add-functionalized poly (nethyl methanylate) (AL49643) © © © © © © © © © © © © © © © © © © ©		collagen XI : C postive, of C negative C n/a GAPOH : C postive, of C negative C n/a		96: C postive, of C negative C n/s 97: C postive, of C negative C n/s
		🕅 MDB-2011. (fram signal), ef:	anteciace (a monosida)	comments clease spectr if applicate, leave it blank if not		CD9 : C postive, of: C regative C n/a CD81: C postive, of C regative C n/a		98: C postive, cf. C negative C n/a 99: C postive, cf. C negative C n/a
		T HE Adptolf TM redun (from HRany Botec) of :				CD48: C positive, of: C negative C n/a CD135: C positive, of: C negative C n/a		300: C positive, ot C regative C rule other-please specify if applicable, leave it blank if not (if more than one, compart them with comma ',')
	obundar."	Ennek mitlet bierriseden Mittel:		f coluges if PEG-based hydrogel if supersion culture (no substrate/scafficial) if sik / Poly(ethyleneoxide) (PEG) if poly-(actic co-glocalic acid)		NST: C positive, of C negative C n/a Bone-specific alkaline phosphatase (BAP): C positive, of C negative C n/a		
	lif only me culture medium is	Encode and a second sec	station of an outpuel *	(PLSA) in poly(coperandme) (PL) in in it in it is a other scatters		70: C positive, of: C negative C n/a collagen III : C positive, of : C negative C n/a	CFU-F induction	colony numbers of CRU-F:
	used leave the concentration black, which by default is 100%)	F HitP22, 00 [2422 (300 Mathema)		comments glease spectry in applicable, leave it bank in not		ALCAM: C postive, of: C negative C n/a CD59: C postive, of: C negative C n/a		induction fold: final cell density: pissee abust the data to fit one of the unit being when applicable:
		E Meiblicht medum (from Steinichel Tec.) E BEIN Eigleis Weimal Essential Medum with Exletis salts)	interval between renewing of			DDH3 : C positive, of: C negative C n/a won Kossa staining : C positive, of: C negative C n/a		unit for monolayer culture: cells/(cm*cm); unit for 3D culture: cells/ml.
		Cot+ABI based redux (how GECI)	solution	i dajs		alizarin S staining: C postive, of: C negative C n/a alcan blue staining: C positive, of: C negative C n/a	expansion of cell number	C or, please provide the final cell number (so that the density will be figured out automaticit):
		other-please specify if applicable, leave it Sank if not (if nove than one, compart them with comma ',')	incubation/culture duration	- Caps		Sudan black IV staining: C positive, cf. C regative C n/a SH4: C positive, cf. C regative C n/a		
			part V: extra conditions			ASMA: C positive, of: C negative C n/a MAB1470 : C positive, of: C negative C n/a		2nd: der anzmu, zei differentation percentage % hotal cell number;
		culture medium volume	passaging enzyme	[mult with concentration]		0056 : C positive, of: C negative C n/a 003 : C positive, of: C negative C n/a		3id: day approx. cell differentiation percentage % total cell number:
		E HS, of Loncentators F PS, of E arbeits armyotic, of	centrifugation rate	(mut with unit: por ram)		CD5 : C positive, cf. C negative C n/a CD7: C positive, cf. C negative C n/a		4th: day approx. cell differentiation percentage 56 total cell number:
		🗸 Utozasii seun substute (hon Utozasii), d 👘 🖓 cigiutanine, d	porosity of substrate/scaffold	(input with unit, or use default unit: sec(100 ml ait)		0015: C positive, of C negative C n/a C016: C positive, of C negative C n/a	dat shel as to one televenters	9th: dayapprox. call differentiation percentage% total call number
		C turnan heah hoam piana (FFP), of	perfusion rate	(rput with unit)		CD19: C positive, of: C negative C n/a CD25: C positive, of C negative C n/a	detailed outcome internation	tith: day approx. cell differentiation percentage% total cel number:
		Elemente de Educe d	osmolarity	(rput with unit)		CD65: C positive, of C negative C n/a tobacine blue sodium borate stateing : C positive, of C negative C		7th: day approx. cel differentiation percentage 5% hotal cel number:
			pH value			n)a		8th: dayapprox. cell differentiation percentage% total cell number
		Elenon, a Elektrikar, a Elektrikar, a	voltage	nie		serraren o scenerg : © postee, of © negative C n/a 93 : C postee, of C negative C n/a		9th: dayapprox. cal differentiation percentaget% total cal number
		Those server, of Coudcust TV server replacement (how invitragen), of	hydrostatic pressure	(rput with unit)		94: 1 postere, of: C negative C n/a 95: C postere, of: C negative C n/a		201x der epror, oel differentation percentage % total cel number:
		🗍 Jenerastratikani, di 👘 🗍 nan-esential anito acidi, di	magnetic field	(input with unit)		99x C postve, ot C negative C n/a 97: C postive, of C negative C n/a	ary additional outcome	
		The suplement from GBCCL of Tarraphenent from GBCCL of		4		98 C postve, of C negative C n/a	information	
		Thuran dourin, d	any additional conditions (e.g.,			100: C positive, of: C negative. C n/a other classe sharfs if averable, laser if Nerk if nerk than one, connact them with connex 10		Ŀ
	supplements and growth factors	11.6 d	cernitugation procedure, etc.)			Control of the second se Second second seco second second sec		۵.
	(input concentrations with units)	Construction of the constr		2		collagen 1: C positive, of (percentage of the total cell number, or 'w','m','y'): C negative C n/a	any conclusion	
		E EA, IN R.3 light JL (I	and Minutes and an a			colligen II : C positive, of: C negative C n/a		-
		ten al latar (57), d erfragalete (57), d	part VI: outcome information			osteorakin (OC) : C positive, of: C negative C n/a		4
		🗍 gadamai gawih lactar (EFF), d		C esteogenic C chandragenic C adpogenic C to muscle C to tendon C to nerve		CBFA1/Runx2 : C postive; of C negative C n/a CD45: C positive; of C negative C n/a	any hypothesis	
		E thankspath (PO), of		C prolifeation without differentiation. (If proliferation, please select what potency(s) do the MECs thil maintain		CD105 (SH2): C positive, of C regative C n/a CD73 (SH3): C positive, of C negative C n/a		
		Recube endstreliël priviti tator (1657), d		E osteogenk E chandragenk E adorgenk E musie E territon E neme;		CD14: C positive, of: C negative C n/a CD34: C positive, of: C negative C n/a	any relevant link	2
			differentiation fate *	other-please specify)		C029: C positive, of C negative C n/a C031 (PECAM-1): C positive, of C negative C n/a		teres teres
		i janzan, w je i informante, ar je i dokonse i, a		C apoptoss		CD166: C postive, of C negative C n/a CD13: C postive, of C negative C n/a		CONTRACTO CONTRACTO
		Diee 1, df 190F-00, df		C multi-differentiation (please provide datale below in "multi-differentiation" colours.)		C090 (Thy-1): C positive, cf. C regative C n/a C044: C positive, cf. C negative C n/a		1
		EDI, d EDI, d EDI, d		C atter-developed		CD628. (8-selectime): C postive, of C negative C n/a STRD-1: C postive, of C negative C n/a		All times are GHT, the time new is 9-3-2009 \$5.54
						CD140a (PDGFaR): C positive, of: C negative C n/a CD106: C positive, of: C negative C n/a	S&& © 2011-2006 Conserve Inc.	Clear Cookies - Contact Us - CUTEG (Confind University Chemical Engin







Data Processing

	Normalised data
A B C D E F 6 H I 1 neesid species NOCsourcePCIII BCN2 BCN4 BCN5 BCN4 BCN5	(1) species = {1} (2) species = {2} (3) MSCsource = {0} (4) MSCsource = {1} (5) MSCsource = {1} (6) MSCsource = {2} (6) MSCsource = {2} (6) MSCsource = {2} (7) BCH1 = {0} (8) BCH2 = {1} (9) BCH2 = {0} (10) BCM3 = {1} (11) BCM3 = {0} (12) BCM4 = {1} (13) BCM4 = {0} (14) BCM5 = {1} (15) BCM5 = {0} (17) BCM6 = {1} (17) BCM6 = {0}
Number of CMAR rules = 78 (1) $(6\rangle \rightarrow (138)$ 100.0%, (28.0, 28.0, 75.0) (2) $(26\rangle \rightarrow (138)$ 100.0%, (28.0, 28.0, 75.0) (3) $(16\rangle \rightarrow (138)$ 100.0%, (28.0, 28.0, 75.0) (4) $(126\rangle \rightarrow (138)$ 100.0%, (28.0, 28.0, 75.0) (5) $(1100 125\rangle \rightarrow (138)$ 100.0%, (21.0, 21.0, 75.0) (6) $(1100 125\rangle \rightarrow (138)$ 100.0%, (12.0, 12.0, 75.0) (7) $(132\rangle \rightarrow (138)$ 100.0%, (12.0, 12.0, 75.0) (8) $(128 \rightarrow (138)$ 100.0%, (12.0, 12.0, 75.0) (9) $(28 46\rangle \rightarrow (138)$ 100.0%, (12.0, 12.0, 75.0) (10) $(28 48) \rightarrow (138)$ 100.0%, (12.0, 12.0, 75.0) (11) $(1132\rangle \rightarrow (138)$ 100.0%, (12.0, 12.0, 75.0) (13) $(128 40) \rightarrow (138)$ 100.0%, (12.0, 12.0, 75.0) (13) $(128 40) \rightarrow (138)$ 100.0%, (12.0, 12.0, 75.0) (14) $(56\rangle \rightarrow (138)$ 100.0%, (11.0, 11.0, 75.0) (15) $(56\rangle \rightarrow (138)$ 100.0%, (11.0, 11.0, 75.0) (16) $(49 49 6) 2> (134)$ 100.0%, (8.0, 8.0, 13.0) (17) $(94 96 125\rangle \rightarrow (134)$ 100.0%, (8.0, 8.0, 13.0) (18) $(49 49 6 98\rangle \rightarrow (134)$ 100.0%, (3.0, 3.0, 13.0) (20) $(89 496\rangle \rightarrow (134)$ 100.0%, (3.0, 3.0, 13.0) (21) $(8 94 96\rangle \rightarrow (134)$ 100.0%, (3.0, 3.0, 13.0) (22) $(18 94 96) \rightarrow (134)$ 100.0%, (3.0, 3.0, 13.0) (23) $(8 30 94 96) \rightarrow (134)$ 100.0%, (3.0, 3.0, 13.0) (24) $(24) - (134)$ 100.0%, (2.0, 2.0, 13.0) (25) $(22) \rightarrow (134)$ 100.0%, (3.0, 3.0, 13.0) (24) $(24) - (134)$ 100.0%, (2.0, 2.0, 4.0) (25) $(22) - (134)$ 100.0%, (2.0, 2.0, 13.0)	 CARM 1 6 26 64 66 70 125 138 4 20 32 34 36 125 138 4 20 32 34 36 125 138 6 26 64 66 90 125 138 8 30 94 96 98 127 134 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

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

GN F:	<pre>:\desktop\desk\cmar\cmar\CMAR LIMITED\cmd.exe</pre>	- 🗆 ×
(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>	<pre>{4 94 96 125> -> <134> 100.0%, <8.0, 8.0, 13.0></pre>	
(20)	(8 94 96) -> (134) 100.02, (3.0, 3.0, 13.0) (8 94 96) -> (134) 100.02, (3.0, 3.0, 13.0)	
(22)	$(1 \ 8 \ 94 \ 96) \rightarrow (134) 100.02, (3.0, 3.0, 13.0)$	
<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>	
<26>	$(112) \rightarrow (136) 100.0\%, (2.0, 2.0, 4.0)$	
(28)	(1 247 -> (1347 100.0%, (2.0, 2.0, 13.0) 224 1251 -> 21343 100.0% 22 0 2 0 12 0)	
(29)	$(1 22) \rightarrow (136) 100.0 \times (2.0, 2.0, 4.0)$	
<30>	<pre><pre><pre><pre><pre><pre><pre><pre></pre></pre></pre></pre></pre></pre></pre></pre>	
<31>	<1 112> -> <136> 100.0%, <2.0, 2.0, 4.0>	
<32>	<4 112> -> <136> 100.0×, <2.0, 2.0, 4.0>	
<33>	$(22 \ 125) \rightarrow (136) \ 100.0\%, \ (2.0, \ 2.0, \ 4.0)$	
(34)	(1 24 125) = 7 (134) 100.02, (2.0, 2.0, 13.0)	
(36)	(121) -> (137) 100.02. (1.0. 1.0. 1.0)	
(37)	<123> -> <137> 100.0%, <1.0, 1.0, 1.0>	
<38>	<94 104> -> {135> 100.0%, <1.0, 1.0, 6.0>	
<39>	<4 114> -> <135> 100.0%, <1.0, 1.0, 6.0>	
<40>	$\langle 8 114 \rangle = \rangle \langle 135 \rangle 100.0\%, \langle 1.0, 1.0, 6.0 \rangle$	
(41)	(104 1147 -7 (1357 100.02, (1.0, 1.0, 5.0) (1) 1265 -5 (1355 100.02 (1.0, 1.0, 6.0)	
(43)	$(12 \ 126) = (135) \ 100.02 \ (1.0, 1.0, 6.0)$	
(44)	<96 126> -> <135> 100.0%, <1.0, 1.0, 6.0>	
<45>	<114 126> -> <135> 100.0%, <1.0, 1.0, 6.0>	
<46>	$(12 \ 127) \rightarrow (135) \ 100.0$, $(1.0, \ 1.0, \ 6.0)$	
<47>	$(114 127) \rightarrow (135) 100.02, (1.0, 1.0, 6.0)$	
(48)	(117 1277 -7 (1357 100.02, (1.0, 1.0, 5.0))	
(50)	$(12, 128) \rightarrow (135) 100.02, (1.0, 1.0, 6.0)$	
<51>	<114 128> -> <135> 100.0%, <1.0, 1.0, 6.0>	
<52>	<117 128} -> <135> 100.0%, <1.0, 1.0, 6.0>	
<53>	<119 128> -> <135> 100.0%, <1.0, 1.0, 6.0>	
<54>	$(12 \ 130) \rightarrow (135) \ 100.0\%, \ (1.0, \ 1.0, \ 6.0)$	
(56)	(74, 130) = 2 (135) 100.02, (1.0, 1.0, 5.0) (296, 130) = 2 (135) 100.02 (1.0, 1.0, 5.0)	
<57>	$(38 \ 130) \rightarrow (135) \ 100.0\%, (1.0, 1.0, 6.0)$	
<58>	<12 131> -> <135> 100.0%, <1.0, 1.0, 6.0>	
<59>	<94 131> -> {135> 100.0%, <1.0, 1.0, 6.0>	
<60>	<96 131> -> <135> 100.0%, <1.0, 1.0, 6.0>	
<61>	$(98 \ 131) \rightarrow (135) \ 100.0\%, (1.0, 1.0, 6.0)$	
(62)	$(2 \ 121) -) \ (137) \ 100.02, \ (1.0, 1.0, 1.0)$	
(64)	$(36) \rightarrow (138) 93.54%, (29.0, 31.0, 75.0)$	
<65>	<1 36> -> <138> 93.54%, <29.0, 31.0, 75.0>	
<66>	(30 114) -> <135> 85.71%, <6.0, 7.0, 6.0>	
<67>	<pre><94 114> -> <135> 85.71×, <6.0, 7.0, 6.0></pre>	
<68>	$\langle 1 \ 30 \ 114\rangle \rightarrow \langle 135\rangle \ 85.71\%, \langle 6.0, 7.0, 6.0\rangle$	
(20)	$(1 \ 94 \ 114) \rightarrow (135) \ 85.712, (6.0, 7.0, 6.0)$	
(71)	(102) -> (134) 83.33, (5.0, 6.0, 13.0)	
(72)	(4 102) -> (134) 83.33%, (5.0, 6.0, 13.0)	
(73)	<10 102> -> <134> 83.33×, <5.0, 6.0, 13.0>	
(74)	<pre>{96 102> -> <134> 83.33×, <5.0, 6.0, 13.0></pre>	
(75)	$\langle 28 \ 100 \rangle \rightarrow \langle 134 \rangle \ 80.0 \times, \langle 4.0, 5.0, 13.0 \rangle$	
(39)		

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 \times 90% of the database).



Rule # 49: {in vitro + monolayer + human donor + DMEM + TGF β 1 + plastic substrate} \Rightarrow {chondrogenesis} [100.0%] which can be interpreted

{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 TGF β 1 (Transforming Growth Factor β 1), on plastic substrate.



Rule # 86: {DMEM + FBS + ascorbate-2-phosphate + Dex} \Rightarrow {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 \times 90% of the database). There were 295 CMAR rules generated, among which many are found to be interesting.

Interesting rules

Rule # 27: {pyruvate + proline} \Rightarrow {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} \Rightarrow {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