

Genexpressionsanalyse boviner mesenchymaler Stammzellen und deren *in vitro* differenzierten Folgelinien

Dissertation

zur Erlangung des Grades

Doktor der Naturwissenschaften

(Dr. rer. nat.)

am Fachbereich Biologie der
Johannes Gutenberg-Universität in Mainz



Martin Schulze

Geboren am 05. April 1981 in Essen

Mainz, August 2015

Dekan: Prof. Dr. XXX

1. Berichterstatter: Prof. Dr. XXX

2. Berichterstatter: Prof. Dr. XXX

Tag der mündlichen Prüfung:

INHALTSVERZEICHNIS

1	Einleitung	1
1.1	Stammzellen und ihre Nutzung in Medizin und Forschung	1
1.2	Mesenchymale Stammzellen als Vertreter der adulten Stammzellen	6
1.3	Genexpressionsanalyse im Hochdurchsatz mittels Microarrays und Sequenziertechniken der neuen Generation	11
1.4	Bioinformatische Auswertung von NGS Daten zur Genexpressionsanalyse 21	
1.5	Zielsetzung	23
2	Material und Methoden	25
2.1	Isolation Mesenchymaler Stammzellen aus dem Knochenmark des Rindes 25	
2.2	Charakterisierung der Zellkulturen	26
2.2.1	Immunohistochemische Färbungen von undifferenzierten MSC	27
2.2.2	Charakterisierung von MSC durch Nachweis zelllinientypischer Transkripte mittels RT-PCR und Sanger Sequenzierung	28
2.3	Differenzierung Mesenchymaler Stammzellen	31
2.3.1	<i>In vitro</i> Differenzierung von undifferenzierten MSC zu Adipozyten .	31
2.3.2	<i>In vitro</i> Differenzierung von undifferenzierten MSC zu Osteoblasten	32
2.3.3	<i>In vitro</i> Differenzierung von undifferenzierten MSC zu Chondrozyten 33	
2.4	Charakterisierung von aus MSC differenzierten Folgelinien	34
2.4.1	Nachweis chondrogener Differenzierung durch Alcian Blue Färbung 34	
2.4.2	Nachweis Adipogener Differenzierung durch Oil Red O Färbung	35
2.4.3	Nachweis Osteogener Differenzierung durch Alzarin Red S Färbung	35
2.5	Herstellung, Hybridisierung und Auswertung von Plasmid – DNA Microarrays zur Genexpressionsanalyse	36

2.5.1	Gewinnung der Plasmid – DNA Sonden durch Retransformation chemisch kompetenter Zellen.....	37
2.5.2	Isolation von Plasmid-DNA im 96er Format	38
2.5.3	Herstellen der Plasmid-DNA Microarrays durch Spotten.....	38
2.5.4	Lineare Amplifikation von mRNA in zwei aufeinanderfolgenden Schritten.....	39
2.5.5	Hybridisierung, Scan und Auslesen der Plasmid-DNA Microarrays .	41
2.5.6	Auswertung der Plasmid-DNA Microarrays.....	42
2.6	Genexpressionsanalyse mittels Illumina NGS Sequenzierungen (RNA – Seq)	43
2.6.1	Kartierung von Sequenzen gegen annotierte Bereiche des Rindergenoms mit Hilfe der <i>CLC Genomics Workbench</i>	46
2.6.2	BLASTn Abgleich der Illumina Sequenzen mit der NCBI Unigene Datenbank des Rindes	47
2.6.3	Assemblierung von Illumina Sequenzen und Annotation der Contigs mittels BLASTn Abgleich mit der Refseq RNA Datenbank.....	49
2.6.4	Bestimmung differenzieller Genexpression	50
2.7	Einordnung der Genexpressionsprofile in einen übergeordneten biologischen Kontext.....	51
2.7.1	Globaler Vergleich der Genexpressionsprofile.....	52
2.7.2	Analyse der Genexpression – FAC der 100 stärkstexprimierten Gene mittels DAVID	53
2.7.3	Analyse differenziell exprimierter Gene – Abgleich mit biologischen Pathways und Zuordnen der Genontologie.....	54
2.7.4	Assemblierung von unkartierten Sequenzen und Charakterisierung noch nicht annotierter Transkripte	55
2.8	Übersicht über die bioinformatischen Hilfsmittel.....	57
2.9	Puffer und Lösungen.....	59
3	Ergebnisse.....	60
3.1	Isolierung und <i>in vitro</i> Kultur Mesenchymaler Stammzellen.....	60
3.2	Charakterisierung Mesenchymaler Stammzellen	61
3.2.1	Nachweis MSC typischer Antigene mittels Immunohistochemie und Test auf Mycoplasmen Kontamination	61
3.2.2	Nachweis MSC typischer Transkripte durch RT-PCR.....	62
3.3	<i>In vitro</i> Differenzierung der MSC	65
3.3.1	Adipogene Differenzierung der MSC	65

3.3.2	Osteogene Differenzierung der MSC	66
3.3.3	Chondrogene Differenzierung der MSC.....	68
3.4	RNA - Präparation und RNA - Amplifizierung	69
3.4.1	Präparation von Gesamt-RNA aus Zellkulturen.....	70
3.4.2	Lineare Amplifikation von RNA	71
3.5	Bestimmung differenzieller Genexpression mittels Plasmid - DNA Microarrays.....	73
3.5.1	Herstellung der Plasmid-DNA Microarrays	74
3.5.2	Bestimmung differenzieller Genexpression mittels Microarrays – Vergleich von undifferenzierten MSC und Osteoblasten.....	76
3.5.3	Bestimmung differenzieller Genexpression mittels Microarrays – Vergleich von undifferenzierten MSC und Chondrozyten.....	77
3.5.4	Bestimmung differenzieller Genexpression mittels Microarrays – Vergleich von undifferenzierten MSC und Adipozyten	79
3.6	Genexpressionsanalyse mittels Illumina Hochdurchsatz Sequenzierung – Prozessieren der Rohdaten.....	79
3.6.1	Prozessieren der Rohdaten und Annotation der Sequenzen – Kartierung gegen das Rindengenom.....	81
3.6.2	Prozessieren der Rohdaten und Annotation der Sequenzen – BLASTn Abgleich mit der UG-Datenbank des Rindes	82
3.6.3	Prozessieren der Rohdaten, Assemblierung und Annotation der Konsensussequenzen – BLASTn Abgleich mit der refseq-RNA Datenbank.....	83
3.6.4	Globaler Vergleich der Expressionsprofile aller Zelllinien.....	84
3.7	Read-Count basierte Analyse der Genexpression mittels RNA-Seq	86
3.7.1	Expression von zelllinientypisch exprimierten Markergenen.....	87
3.7.2	Funktionelle Annotation stark exprimierter Gene – Junge MSC	90
3.7.3	Funktionelle Annotation stark exprimierter Gene – Adipogenese	92
3.7.4	Funktionelle Annotation stark exprimierter Gene – Osteogenese	94
3.7.5	Funktionelle Annotation stark exprimierter Gene – alte MSC	97
3.7.6	Funktionelle Annotation stark exprimierter Gene – Chondrogenese .	99
3.8	Analyse differenziell exprimierter Gene mittels RNA-Seq.....	100
3.8.1	Differenzielle erhöhte Genexpression in jungen MSC.....	103
3.8.2	Differenziell erhöhte Genexpression in Adipozyten	105
3.8.3	Differenziell erhöhte Genexpression in Osteoblasten und alten MSC 107	107
3.8.4	Differenziell erhöhte Genexpression in Chondrozyten.....	110

3.9	Charakterisierung unbekannter Transkripte	110
4	Diskussion	116
4.1	Primärkultur undifferenzierter MSC	117
4.2	<i>In vitro</i> Differenzierung von MSC und Genexpressions-analyse differenzierungsspezifischer Markergene.....	120
4.2.1	Adipogene Differenzierung von MSC	121
4.2.2	Osteogene Differenzierung von MSC	122
4.2.3	Chondrogene Differenzierung von MSC.....	122
4.2.4	Langzeit-Zellkultur undifferenzierter MSC im Multilayer.....	125
4.2.5	RNA Präparation und RNA Amplifikation	126
4.3	Bestimmung differenzieller Genexpression mittels Plasmid - DNA Microarrays.....	130
4.3.1	Differenziell exprimierte Gene im Vergleich undifferenzierter MSC mit Osteoblasten.....	132
4.3.2	Differenziell exprimierte Gene im Vergleich undifferenzierter MSC mit Chondrozyten	134
4.3.3	Differenziell exprimierte Gene im Vergleich undifferenzierter MSC mit Adipozyten.....	136
4.4	Genexpressionsanalyse mittels Illumina Hochdurchsatz Sequenzierung – Prozessierung der Rohdaten	137
4.5	Umfassender Vergleich der Genexpressionsprofile aller Zelllinien.....	142
4.6	Funktionelle Annotation stark exprimierter Gene - Junge MSC.....	144
4.7	Funktionelle Annotation stark exprimierter Gene - Adipozyten.....	145
4.8	Funktionelle Annotation stark exprimierter Gene – Osteoblasten und alte MSC	146
4.9	Funktionelle Annotation stark exprimierter Gene – Chondrozyten.....	147
4.10	Analyse differenziell exprimierter Gene mittels RNA-Seq.....	148
4.10.1	Differenziell erhöhte Genexpression in jungen MSC.....	149
4.10.2	Differenziell erhöhte Genexpression in Adipozyten	154
4.10.3	Differenziell erhöhte Genexpression in Osteoblasten und alten MSC 156	156
4.10.4	Differenziell erhöhte Genexpression in Chondrozyten.....	157
4.11	Charakterisierung unbekannter Transkripte	158
5	Zusammenfassung und Ausblick	163
6	Abkürzungsverzeichnis.....	166
7	Abbildungsverzeichnis.....	167

8	Tabellenverzeichnis.....	169
9	Digitaler Anhang.....	171
10	Anhang – Differenziell exprimierte Gene.....	172
11	Literaturverzeichnis.....	174
12	Curriculum Vitae.....	189
13	Danksagung.....	190
14	Eidesstattliche Erklärung.....	191

1 EINLEITUNG

1.1 STAMMZELLEN UND IHRE NUTZUNG IN MEDIZIN UND FORSCHUNG

Jede Zelle des Körpers lässt sich, wenn man nur genügend Zellteilungen und Entwicklungsschritte zurückgeht, letztendlich auf eine einzige Zelle zurückführen. Die befruchtete Eizelle ist diese erste Stammzelle, aus ihr entwickelt sich der komplette Organismus. Nach der sog. symmetrischen Zellteilung sind die gebildeten Tochterzellen identisch zur Ursprungszelle, die Stammzelle erneuert sich als exakte Kopie ihrer selbst (siehe Abbildung 1 - 1). Die asymmetrische Zellteilung einer Stammzelle bringt hingegen eine zur Ursprungszelle identische Stammzelle,

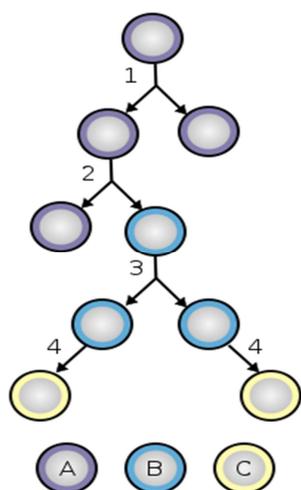


Abbildung 1 - 1 Modell der Stammzellteilung

Quelle: Wikipedia

- 1, 3 – Symmetrisch Teilung
- 2 – Asymmetrisch Teilung
- 4 – Differenzierung
- A – Totipotente SZ
- B – Progenitorzelle
- C – Ausdifferenzierte Zelle

sowie eine, in ihrem Differenzierungspotential bereits weiter eingeschränkte, Progenitorzelle hervor. Per Definition können sich Stammzellen durch Zellteilung beliebig oft teilen, oder aber zu weiter spezialisierten Zellen differenzieren [1]. Eine totipotente Zygote kann sowohl embryonales, als auch extraembryonales Gewebe hervorbringen. Aus den gemeinhin als Embryonale Stammzellen (ES – Zellen) bezeichneten, pluripotenten Zellen der inneren Zellmasse des Blastozysten hingegen entwickelt sich zwar ein vollständiger Organismus, zu extraembryonalen Strukturen können diese Stammzellen jedoch nicht mehr differenziert werden. Im Allgemeinen lässt sich daher die Aussage treffen, daß mit der zunehmenden Spezialisierung der Tochterzellen einer Stammzelle deren Differenzierungspotential immer weiter abnimmt. Die Benutzung des Wortes Stammzelle geht auf den deutschen Biologen Ernst Haeckel zurück [2], der bereits 1868 in seiner an Charles Darwins Ideen

angelehnten Hypothese eine Urstammzelle als Ursprungsorganismus postulierte, von dem aus sich phylogenetisch alle vielzelligen Organismen ableiten. Haeckel war es auch, der einige Jahre später den Begriff der Stammzelle als erster in der heute gebräuchlichen, ontogenetischen Bedeutung für eine befruchtete Eizelle als Ursprung eines Organismus prägte [3]. Der wissenschaftliche Pionier und spätere Nobelpreisträger Sir Robert Geoffrey Edwards und Kollegen schrieben in den 1960er Jahren bemerkenswerte Veröffentlichungen über Ihre praktischen Erfahrungen der *in vitro* Reifung von Eizellen des Menschen und anderer Säuger [4, 5]. Die künstliche Befruchtung beim Menschen war erklärtes Fernziel. Zwar war bereits 1959 zumindest für Kaninchen eine *in vitro* Befruchtung nachgewiesen worden [6]. Dennoch stellten Chromosomenaberrationen und sonstige abnormale Entwicklungen in frühen Teilungsstadien, sowie eine schwer zu kontrollierende hormonelle Regulation zunächst echte Herausforderungen dar. Die erste gelungene künstliche Befruchtung beim Menschen konnten Edwards und sein Kollege Steptoe erst mit der Geburt von Louise Joy Brown im Jahr 1978 der Welt vorstellen [7]. Heute ist die Reproduktionsmedizin in entwickelten Ländern ein fest etablierter Teil der medizinischen Grundversorgung und kann auch als Quelle zur Gewinnung von neuen ES – Zelllinien heran gezogen werden. Für die medizinische Forschung sind prinzipiell pluripotente Zellen von großem Interesse. Da dieser Zelltyp hinsichtlich seiner Differenzierbarkeit keinerlei Einschränkungen unterliegt, sind die medizinisch denkbaren Anwendungsmöglichkeiten mittels der sog. Stammzelltherapie sehr vielfältig. Embryonale Stammzellen aus Kaninchen in diversen Teilungsstadien wurden schon 1966 in Kultur genommen, auch ließen sich im Rahmen dieser Experimente spontane Differenzierungsprozesse beobachten [8]. Dennoch waren die Ergebnisse noch in erheblichem Maße vom Zufall abhängig. Gezieltes wissenschaftliches Arbeiten an ES - Zellen wurde erst 1981 durch die Isolation der ersten ES Zelllinie aus der Maus durch M. J. Evans und Kollegen möglich [9]. Diese ES-Zellen waren für 30 Passagen in Kultur stabil. Die Pluripotenz der Zellen bewies Evans durch Injektion der über einen längeren Zeitraum in Kultur gehaltenen Zellen in Mäuse. Die Stammzellen begannen dort spezielle Tumore, sog. Teratome zu bilden. Diese Geschwüre enthalten Gewebe aller Keimblätter und es können sich beispielsweise komplette Zähne und Haare darin ausbilden. Damit war auch ein direkter Zusammenhang zwischen Tumoren und unkontrolliert wachsenden Stammzellen gegeben, der mit der Entdeckung von entarteten, hämatopoetischen Stammzellen als Ursache für die akute myeloische Leukämie (AML) auch unter natürlichen Bedingungen nachgewiesen werden konnte [10].

Neben dem Risiko einer Tumorbildung das eine Stammzelltherapie potentiell mit sich bringt [11], spielen, besonders im Hinblick auf die ES – Zellen, auch erhebliche ethische Bedenken eine Rolle. Für die Etablierung einer nativen humanen ES – Zelllinie muss sich entwickelndes menschliches Leben vernichtet werden. Die hierfür genutzten, befruchteten Eizellen stammen wie oben erwähnt meist aus der Reproduktionsmedizin. Bei der *in vitro* Fertilisation werden nicht alle zur Verfügung stehenden befruchteten Eizellen auch in die Gebärmutter eingesetzt. Diese „Überschüsse“ können als Quelle neuer ES – Zelllinien dienen, so daß sich der Gesetzgeber in Deutschland aus ethischen Gründen für einen Stopp der Etablierung solcher ES – Zelllinien entschieden hat. Generell erscheinen ethische Bedenken in Zusammenhang mit der Forschung an ES – Zellen in vielerlei Hinsicht angebracht, so ist schon in der oben erwähnten Publikation von Edwards von 1965 der Vorschlag nachzulesen, [4] im Falle eines Kinderwunsches bei familiärer X-chromosomaler Krankheitsgeschichte vor Implantation männliche Blastozysten zu verwerfen. Mit dem heutigen Stand der Technik lassen sich Stammzellen in Kultur auch leicht genetisch manipulieren, die resultierenden medizinischen Möglichkeiten wurden 2002 eindrucksvoll durch die Heilung von Mäusen eines immundefizienten Stammes mittels kombinierter Gen/Stammzelltherapie bewiesen [12]. Ein konsequenter Verzicht auf die ES - Stammzelltherapie birgt folglich ebenfalls ein ethisches Dilemma. Sich aus der Stammzellforschung ergebenden Konflikte sind heute ein viel diskutiertes politisches und gesellschaftliches Thema, das in unterschiedlichen Ländern sehr unterschiedlich strikte gesetzliche Regelungen hervor gebracht hat. Eine sehr elegante Lösung für ethische Probleme, ohne gleichzeitig auf pluripotente Stammzellen verzichten zu müssen, ergab sich im Jahr 2006 aus den Forschungsergebnissen der japanischen Arbeitsgruppe um Takahashi. Durch deren Experimente konnte erstmals an Mäusen bewiesen werden, daß adulte, ausdifferenzierte Fibroblasten durch von außen induzierte Aktivierung von bestimmten ES-Zell assoziierten Genen (*oct3/4*, *sox2*, *c-myc*, und *klf4*) zu sog. induzierten, pluripotenten Stammzellen (iPS) umprogrammiert werden können [13]. Wenig später gelang dies auch mit humanen somatischen Zellen, wobei die Auswahl an Genen, die zur Reprogrammierung benötigt wurden, noch eingeschränkt werden konnte [14-16]. Da dieser Erfolg auf der gentechnischen Manipulation der Expression oben genannter Gene mittels retroviraler Transfektion beruht, sind damit für die direkte Übertragung der Methoden auf menschliche Zellen für Therapiezwecke schwer abwägbare Risiken verbunden, die über das generelle cancerogene Potential transplantiertter Stammzellen noch hinaus gehen.

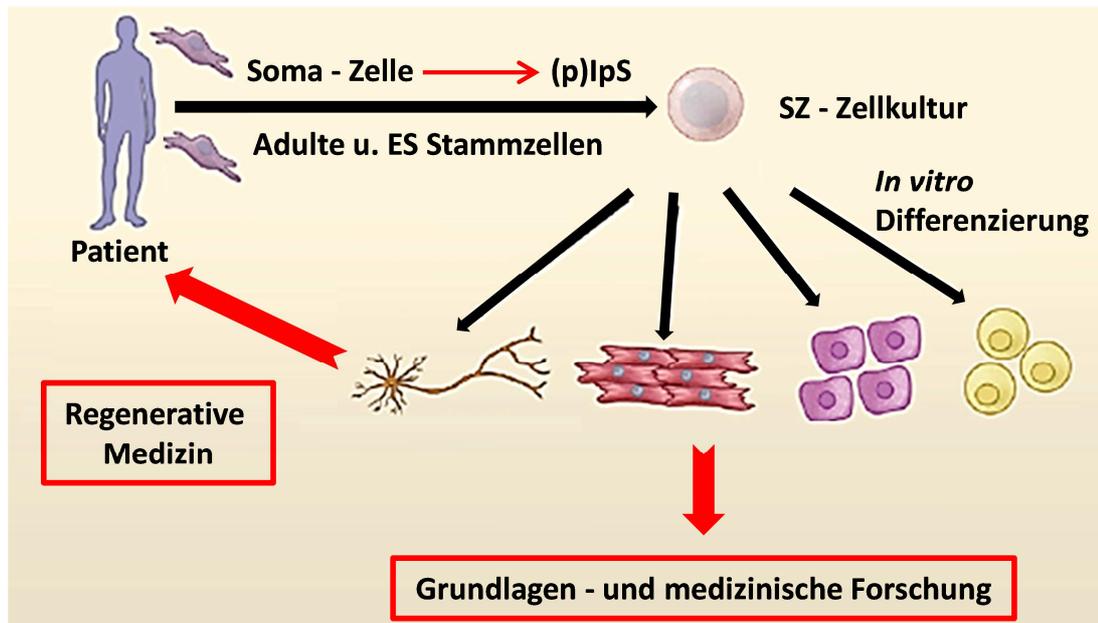


Abbildung 1 - 2 Stammzellen in Medizin und Forschung

Die Abbildung zeigt schematisch die Anwendungsmöglichkeiten für Stammzellen in Medizin und Forschung. Unabhängig von der Art der verwendeten Stammzelle oder dem angestrebten Verwendungszweck wird stets eine Stammzellen – Zellkultur, oft mit angeschlossener *in vitro* Differenzierung, als Zwischenschritt benötigt (Abb. verändert nach Yamanaka et al [17]).

Ein größeres Maß an Kontrolle über die genetischen Modifikation im Zuge der Reprogrammierung erlaubte das Vorgehen von Woltjen und Kollegen, die durch Einsatz des sog. piggyBAC Transposon/Transposase Systems Stammzellgene in der Zielzelle exprimieren, und diese nach Erschaffung einer stabilen iPS Linie auch wieder aus dem Genom entfernen konnten [18]. Dass die Bildung von gentechnisch veränderten Zellen bei der Reprogrammierung adulter Zellen zu Stammzellen auch zu umgehen ist, konnte in Versuchen von Zhou und Kollegen gezeigt werden. Statt Transfektion und Expression der Stammzell – Gene innerhalb der zu reprogrammierenden Zielzelle, schleusten diese direkt die korrespondierenden Proteine ein [19]. Die mit einer Stammzelltherapie einhergehenden Risiken können so nicht gänzlich verhindert, aber doch minimiert werden. Zudem ist die Herstellung von (p)IPs Zelllinien ein komplexes Verfahren, welches mit hohen Kosten und niedriger Ausbeute einhergeht. Es bleibt die Beobachtung, daß pluripotente Stammzellen, egal ob nativer oder induzierter Art, unter Kulturbedingungen in ihrer Genomstruktur weit weniger stabil zu sein scheinen als nicht pluripotente Zellen [20]. Betroffen sind hier vor allem die Anzahl von Genkopien von Onkogenen und Tumorsuppressorgenen. Über die Transplantation

von Knochenmark hinausgehende Stammzelltherapien wurden bis zum heutigen Tage sowohl an Tieren als, auch am Menschen durchgeführt, wobei sowohl adulte, als auch ES – sowie iPS Zellen (noch nicht am Menschen) zum Einsatz [21] kamen (siehe Abbildung 1 - 2).

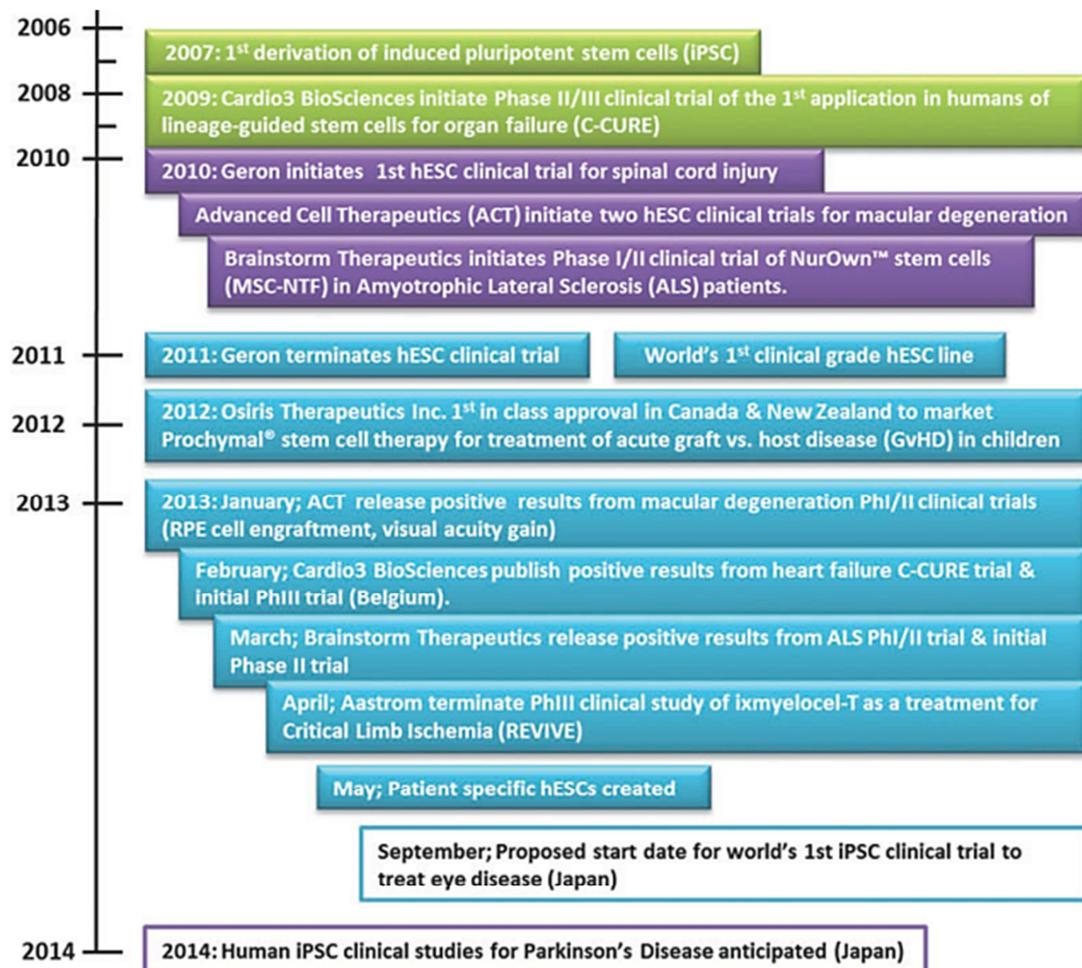


Abbildung 1 - 3 Meilensteine der Stammzelltherapie

Die Abbildung zeigt eine Auswahl kommerziell orientierter Ansätze der Stammzelltherapie auf. Die sichere Anwendung standardisierter Stammzelltherapieansätze birgt ein erhebliches wirtschaftliches und medizinisches Potential. (Quelle: Ratcliffe et al [22])

In den USA sind Versuche mit nativen humanen ES-Zellen am Menschen für distinkte Krankheitsbilder seit 2009 unter strengen Auflagen erlaubt. Dennoch ist der Erfahrungsumfang solcher Therapien am Menschen aus den bereits aufgeführten Gründen noch immer begrenzt. Humanstudien sind bislang vor allem für neuronale Erkrankungen wie Parkinson, Alzheimer, Amyotrophe Lateralsklerose, Verletzungen des Rückenmarks und Makuladegeneration

beschrieben worden, aber auch Herzerkrankungen und die regenerative Medizin im Allgemeinen stehen im Focus der Forschung. Die ersten Ergebnisse dieser Studien sind vielversprechend. Sie lassen für Stammzelltherapien eine wichtige Rolle in der zukünftigen medizinischen Versorgung erahnen und geben erste Hinweise, daß die Risiken dieser neuartigen Therapie zumindest prinzipiell beherrschbar sind [21]. Dabei wird auch von kommerzieller Seite her großes Interesse an der raschen Fortentwicklung der Technik gehegt. In Großbritannien hat der vom regierungsnahen „*Technology Strategy Board*“ gegründete Zusammenschluss von mit Stammzelltherapie assoziierten Unternehmen für das Jahr 2011 einen geschätzten weltweiten Umsatz von einer Milliarde Dollar für Stammzelltherapie Produkte bekannt gegeben. Dieser soll bis zum Jahr 2025 auf 20 Milliarden Dollar ansteigen [22]. Abbildung 1 - 3 zeigt, daß die Stammzelltherapie noch 2009, zeitgleich mit der ersten Erlaubnis Ihrer Anwendung am Menschen in den USA, von kommerzieller Seite betrieben in die klinische Testphase einging. Trotz erster Erfolge ist von den in diesem Absatz beleuchteten Problemen, die die Stammzelltherapie mit sich bringt, noch keines zufriedenstellend gelöst. Weitere Grundlagenforschung, auch im Bereich der Alternativen zu pluripotenten Stammzellen, ist daher unabdingbar um in Zukunft einen sicheren und umfassenden Einsatz von Stammzellen in der Medizin gewährleisten zu können.

1.2 MESENCHYMALE STAMMZELLEN ALS VERTRETER DER ADULTEN STAMMZELLEN

Erste, die Stammzellhypothese überprüfende Beobachtungen und darauf aufbauende initiale Experimente zur Einschätzung des Potentials von Stammzellen für wissenschaftliche und medizinische Zwecke fanden bereits im frühen 20ten Jahrhundert statt. Dabei wurde die Vermutung aufgestellt, daß nicht ausschließlich embryonale Stammzellen zu weiter spezialisierten Zellen differenzieren können. Alexander A. Maximov brachte 1909 aufgrund lichtmikroskopischer Beobachtungen die (inkorrekte) These auf, daß der Lymphozyt eine blutbildende Stammzelle darstellt, die sowohl im Embryo, als auch in der postfetalen Entwicklung eines Organismus als Stammzelle aktiv bleibt [23]. Die Blutbildung blieb lange Zeit im Fokus der Forschung, hier war auch am Tiermodell leicht Erfahrung zu sammeln. Der Beweis der klonalen Natur der blutbildenden Stammzellen konnte von Becker und Mc Culloch an Mäusen gezeigt werden [24]. Diese Forschungen bewiesen, daß

neben embryonalen Stammzellen auch im adulten Organismus Stammzellen mit zumindest eingeschränktem Differenzierungspotential existieren mussten. Neben den erwähnten Hämatopoetischen Stammzellen (HSC) findet sich im Organismus eine große Vielzahl weiterer adulter Stammzellen. So wurden bislang adulte Neuronale, Hepatische, Olfaktorische, Dentale, Mesenchymale, Intestinale und weitere Stammzellen beschrieben. Sie alle scheinen spezifische Funktionen, die hauptsächlich mit der Regeneration oder Neubildung der entsprechenden Gewebe assoziiert sind, auszuführen und sind Gegenstand aktueller Forschung. Unter den adulten Stammzellen nehmen Mesenchymale Stammzellen (MSC) eine besondere Stellung ein. Es handelt sich bei MSC um sehr proliferative, in Kultur adhärent wachsende Zellen mit fibroblastischer Morphologie, die zum ersten Mal durch Friedenstein und Kollegen in den 1970er Jahren beschrieben wurden [25]. In den Focus der Stammzellforschung nach heutigen Maßstäben rückten die MSC durch die Arbeiten von Pittenger und Kollegen, die 1999 erfolgreich MSC aus Knochenmark isolierten und *in vitro* in Knochen, Knorpel und Fettzellen differenzierten [26]. MSC werden vornehmlich aus Knochenmark gewonnen, wo sie sich dieselbe Wachstums Umgebung mit den blutbildenden Stammzellen teilen. Zwar wurden MSC auch aus anderen Quellen wie Blut, Plazenta und Muskeln bereits erfolgreich isoliert [27], dennoch ist Ihre Gewinnung aus Knochenmark die wohl am besten etablierte und einfachste Methode. Derzeit geht man im Knochenmark von einer Art „Stammzellnische“ als determinierende lokale Umgebung aus, in der die beiden unterschiedlichen Stammzelltypen differenzieren und dabei vermutlich auch funktionell interagieren [28]. So werden MSC häufig als HSC bei der Blutbildung unterstützende [29] oder sogar lenkende Zellen [30] beschrieben. *In vitro* ließ sich durch Co-Kultivierung beider Zelltypen diese Stammzellnische ansatzweise nachbilden. Es war deutlich ein positiver Effekt der MSC auf die Proliferation der HSC nachzuweisen [31], auch wenn dies sicher nicht bedeutet, daß die Funktion der MSC auf Ihren Einfluß auf HSC beschränkt ist. Kolf und Kollegen [32] gehen davon aus, daß sowohl Integrinrezeptor gesteuerte Einflüsse aus der Extrazellulären Matrix, als auch die Cadherin vermittelte Zell-Zell Interaktion die Stammzellnische mit definieren. Auch wird ein Einfluss der abnehmenden Sauerstoffkonzentration entlang eines Gradienten in Abhängigkeit zum Abstand des nächsten Blutgefäßes als Teil der Nische in Betracht gezogen (siehe nachfolgende Abbildung 1 - 4).

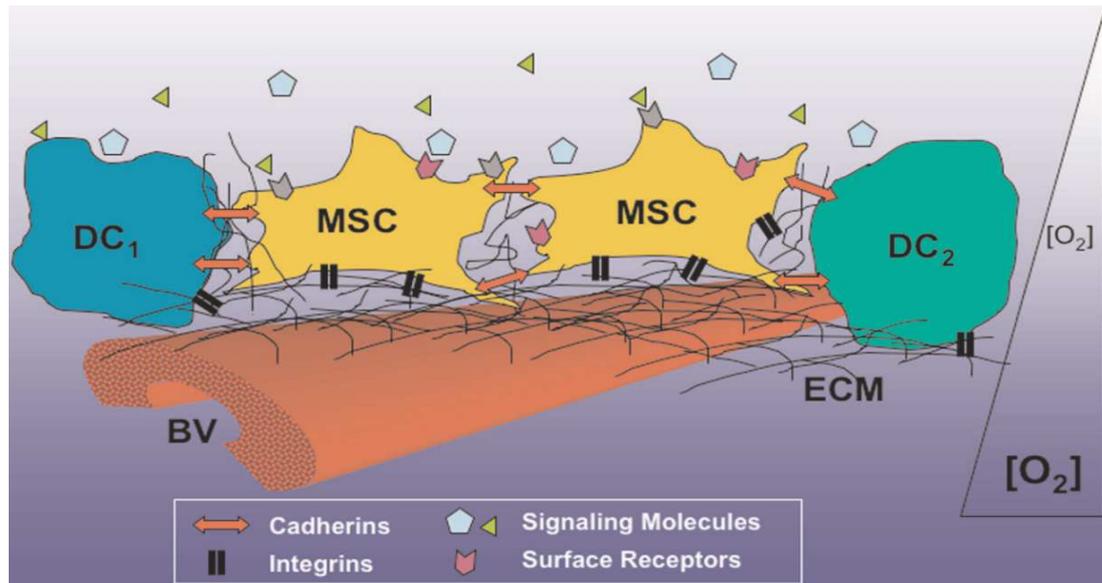


Abbildung 1 - 4 Modell der Stammzellnische

Die Abbildung zeigt ein Modell der sog. Stammzellnische. Als Stammzellnische wird die unmittelbare lokale Umgebung der MSC bezeichnet, in der diese unter physiologischen Bedingungen vorkommen. Innerhalb der Stammzellnische steuern multiple externe Faktoren die Entwicklung der MSC und legen fest, ob sich die Zellen selbst erneuern oder differenzieren. Nicht in der Legende aufgeführte Symbole: $DC_{1/2}$ = Aus MSC differenzierte Zellen Typ $_{1/2}$ | BV = Blutgefäß | ECM = Extrazelluläre Matrix | O_2 = mit der Entfernung zum BV sinkende Sauerstoffsättigung. Abbildung aus Kolf et al. 2007 [32].

Die erstaunliche Plastizität der MSC bedingt ihr hohes medizinisch nutzbares Potential. MSC zeigen in Studien eine sehr starke Heterogenität bezüglich ihres Differenzierungspotentials [33]. Sie gelten als tri- bis multipotent und scheinen sich sogar in Zellen außerhalb der mesenchymalen Keimblattlinie transdifferenzieren zu können [27]. Um einheitliche Standards in der Benennung und klare Richtlinien bezüglich der Differenzierbarkeit von humanen MSC festzulegen, veröffentlichte die „International Society for Cellular Therapy“ 2006 entsprechende minimale Kriterien für humane MSC. Demnach müssen MSC adherent auf Plastikoberflächen der Kulturgefäße anwachsen, sich mindestens zu Adipozyten, Chondroblasten und Osteoblasten differenzieren lassen, sowie ein spezifisches Oberflächenantigenprofil aufweisen. Besonders das Fehlen typischer hämatopoetischer Marker wie CD45 ist hierbei ein wichtiges Merkmal, das die Unterscheidung zu blutbildenden Zellen gewährleistet. Des Weiteren dürfen humane MSC keine Expression der Oberflächenantigene CD34, CD14 und CD19 aufweisen, während die Marker CD105, CD73 und CD90 nachzuweisen sein müssen. Diese Definition ist zunächst einmal gültig, auch wenn nicht unwahrscheinlich ist, daß zukünftige Erkenntnisse eine Modifizierung der genannten Kriterien nach sich ziehen [27]. Grund hierfür

dürften die über die genannten minimalen Kriterien von MSC hinaus gehenden Eigenschaften der Stammzellen sein. Zwar konnte sich eine echte Pluripotenz, die sich durch die Entstehung eines chimären Embryos nach Injektion von MSC in eine Blastozyste erreichen ließe, bislang nicht nachweisen lassen [27]. Die erfolgreiche Differenzierung von MSC zu anderen als zu Fett, Knochen und Knorpelzellen wurde aber bereits mehrfach beschrieben [32]. Demnach lassen sich MSC in unterschiedlichen Studien durch Co-Kultivierung mit Kardiomyozyten sowie durch externe Stimulation mit aktiviertem *notch1* oder dessen Ligand *jagged1* zur Myogenese anregen. In anderen Versuchen konnte weiteres Potential zur Transdifferenzierung von MSC zum Beispiel in neuronale Zellen [34] nachgewiesen werden. Eine Hypothese ist hierbei, daß diese Fähigkeit zur Transdifferenzierung vor allem durch direkten Kontakt von MSC eingebettet in Zellen anderen Typs hervorgebracht werden kann [35]. Dies stützt die Theorie, daß die Plastizität (siehe Abbildung 1 - 5) der MSC durch den externen Einfluss der neuen lokalen Nische vorgegeben wird.

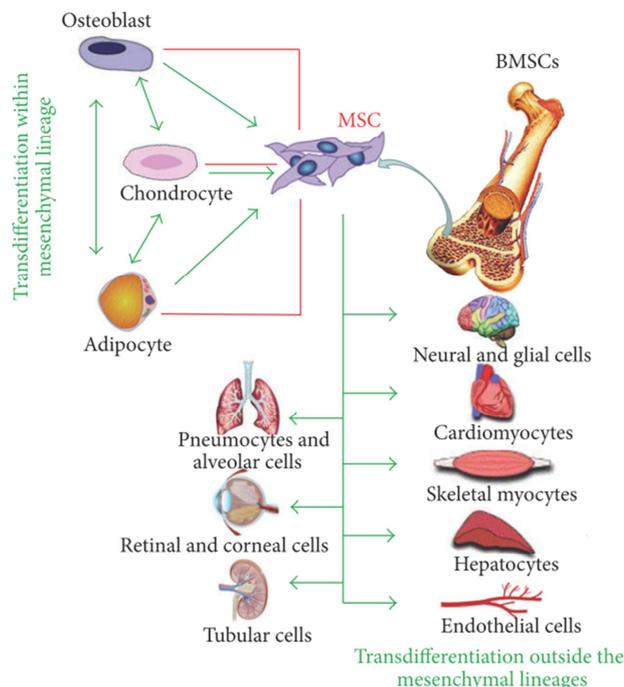


Abbildung 1 - 5 Plastizität von MSC

Die Abbildung zeigt im oberen, linken Teil die hinreichend belegte Differenzierung von MSC zu Osteoblasten, Chondrozyten und Adipozyten. Auch scheint eine Transdifferenzierung zwischen den differenzierten Linien, vermutlich über einen Dedifferenzierungsschritt, möglich. Experimentell konnte die Bildung von Leber-, Nieren-, Pulmonalen-, Retinalen-, Muskel-, Nerven- und Endothelzellen aus MSC beobachtet werden. Abbildung verändert nach Catacchio et al. 2013 [35].

Es ist nicht klar, ob solche Transdifferenzierungsschritte eine komplette Dedifferenzierung von bereits differenzierten MSC voraussetzen. Als gesichert kann jedoch angenommen werden, daß Dedifferenzierungsprozesse bei eigentlich ausdifferenzierten MSC möglich sind [36]. Ob MSC unter physiologischen Bedingungen *in vivo* alle in der Zellkultur oder im Tierversuch nachweisbaren

Eigenschaften auch wirklich abrufen, ist zudem fraglich. Für die medizinische Forschung sind MSC aufgrund ihrer beschriebenen Eigenschaften bereits heute Gegenstand zahlreicher klinischer Studien mit sehr unterschiedlichen Schwerpunkten (siehe Abbildung 1 - 6).

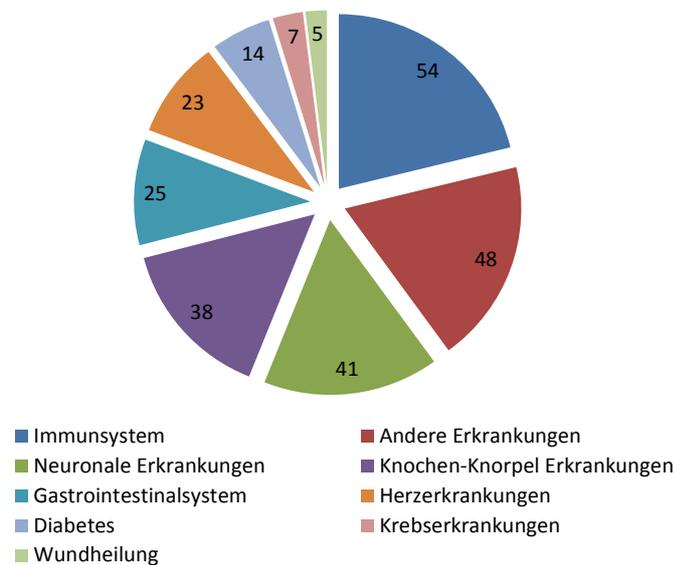


Abbildung 1 - 6 Klinische Studien mit MSC gemäß des "US National Institute of Health"

Gezeigt wird die beachtliche Diversität der klinischen Fragestellungen, die bereits heute unter Verwendung von MSC adressiert werden. Man beachte, daß der zweitgrößte Posten derart viele unterschiedliche Studien beinhaltet, daß er sinnvollerweise unter „Andere Erkrankungen“ zusammengefasst werden musste. Quelle: www.clinicaltrials.gov

Derzeit sind im amerikanischen „National Institute of Health“ (NIH) 255 klinische Studien bei denen MSC verwendet werden gelistet, dabei werden 51 % der gelisteten Studien außerhalb der USA durchgeführt. Das Feld scheint zudem einen sehr schnellen Zuwachs zu verzeichnen, im Mai 2011 lag die Zahl noch bei 123 Studien [37]. Dieser Trend ist ungebrochen, aktuelle Zahlen dürften folglich über die aufgeführten 255 Studien hinausgehen. Die Art der unterschiedlichen Versuche umfasst dabei das Spektrum der klinischen Phasen I bis III, somit werden nennenswerte Patientenzahlen erreicht. Eine deutliche Gefahr geht durch das Verpflanzen von MSC in Menschen zu therapeutischen Zwecken folglich nicht aus, besonders die Differenzierung in Kultur und die Applikation der differenzierten Zellen umgehen eine mögliche, ungewollte Differenzierung, die dem Therapieziel entgegen wirken könnte. Langzeitstudien hierzu sind in Anbetracht der Aktualität der Studien jedoch nicht verfügbar. Immunologisch ist es zudem von Vorteil, patienteneigene Zelllinien zu verwenden, da hier keinerlei Abstoßungsreaktionen

zu erwarten sind. Die Möglichkeiten, das Differenzierungspotential von MSC klinisch nutzbar zu machen, dürften auch in naher Zukunft noch zunehmen, und das, obwohl über die zugrundeliegenden biologischen Mechanismen der Differenzierbarkeit derzeit kein umfassendes Wissen vorliegt. Allgemein anerkannt ist die These, daß ein Differenzierungsprozess eine Veränderung in der Genexpression der MSC grundlegend voraussetzt. Die Expression von Genen, die mit der Selbsterneuerung von Stammzellen assoziiert werden, muss eingestellt, und die für die jeweilige Richtung der Differenzierung benötigte Gene müssen spezifisch gesteuert vermehrt transkribiert werden. Epigenetische Modifikationen in Form von Methylierung der Promotorregionen einzelner Gene, sowie eine Veränderung der Heterochromatinstruktur durch Histonmodifikationen scheinen hierbei eine Schlüsselrolle zu spielen [38]. Die Expression einzelner mit den Stammzeleigenschaften und mit den drei Differenzierungen (Adipogenese, Chondrogenese und Osteogenese) in Zusammenhang stehender Gene wurden bereits beschrieben (Details siehe Tabelle 3 - 1 und Tabelle 3 - 11). Eine transkriptomweite Analyse der Genexpression undifferenzierter und differenzierter MSC liefert darüber weit hinaus gehende Informationen und erlaubt einen Einblick in die zugrundeliegenden biologischen Vorgänge. Die bisherige Forschung hat klar gezeigt, daß gerade für die oft potentiell folgenschweren Experimente mit Stammzellen Modellorganismen gute Dienste leisten. Es liegen bislang keine Erkenntnisse vor, die gegen eine Übertragbarkeit von Ergebnissen der Stammzellforschung zwischen Säugern sprechen.

1.3 GENEXPRESSIONSANALYSE IM HOCHDURCHSATZ MITTELS MICROARRAYS UND SEQUENZIERTECHNIKEN DER NEUEN GENERATION

Für die Analyse der Genexpression steht heute eine Vielzahl sehr unterschiedlicher Methoden zur Verfügung. Besonders die schon länger etablierten Verfahren beruhen dabei auf dem Nachweis oder sogar der Quantifizierung einzelner weniger Transkripte. So lassen sich mittels eines „Northern Blot“ [39] oder durch „reverse Transkriptase PCR“ [40] sehr schnell Erkenntnisse über die Expressionsstärke und sogar die Spleißvarianten eines Gens gewinnen. Um umfassende, quantitative Genexpressionsanalysen durchführen zu können, muss die Identität und Häufigkeit möglichst vieler Transkripte in den Zielzellen bestimmt werden. Betrachtet man auf aktuelle Sequenzierabläufe der nächsten Generation (NGS = *Next Generation*

Sequencing) zugeschnittene Anreicherungssysteme, die das gezielte Sequenzieren aller Exons der bekannten Gene des Menschen aus genomischer DNA heraus ermöglichen, so geben die Hersteller den Zielbereich mit einer DNA Menge von 50 Megabasen bei 21.522 Genen mit 357.999 Exons an (Quelle: Agilent Sure Select All Exon V5 System). Somit umfasst der proteinkodierende Bereich des Genoms bei einem Säuger wie dem Menschen in etwa 1,5% der Erbinformation. Zwar werden in keiner Zelle alle Gene transkribiert, dennoch benötigt man für eine umfassende Genexpressionsanalyse einen sehr hohen Durchsatz um dem Anspruch „transkriptomweit“ gerecht werden zu können. Nukleinsäuren ließen sich bis in die Mitte der 1970er Jahre nur indirekt, beispielsweise durch Restriktionskartierung und Gelelektrophorese oder mittels photometrischer Messungen charakterisieren. Mit diesen Methoden kann über die Sequenz, also die Basenabfolge des genetischen Kodes, nur sehr eingeschränkt Informationen gewonnen werden. Die ersten DNA Sequenzierverfahren wurden parallel zueinander in der zweiten Hälfte der 70er Jahre entwickelt. Dabei kamen zwei grundlegend verschiedene Techniken zum Einsatz. Allan Maxam und Walter Gilbert entwickelten eine chemische Methode, radioaktiv markierte DNA basenspezifisch zu spalten und die Fragmente anschließend durch Gelelektrophorese auf zu trennen [41]. Die nötigen chemischen Verfahren sind aufwändig, die Methode konnte sich gegen die noch heute gebräuchliche von Frederick Sanger entwickelte Technik der enzymatischen Kettenabbruch-Sequenzierung nicht durchsetzen [42]. Mittels Sanger-Sequenzierung ist heute eine Leseweite von ~1200 bp und mehr möglich, wobei der Preis einer Sequenz bei kommerziellen Anbietern heute mit weniger als 5 Euro zu Buche schlägt. Dementsprechend waren tiefe Genexpressionsanalysen, die auf der Sanger Methode beruhten schon aus monetären Gründen aufgrund des im Vergleich zur Transkriptomkomplexität verschwindend geringen Datenaufkommens pro Reaktion nicht realisierbar. Der bis zur Mitte der 1990er Jahre beste Ansatz der Expressionsanalyse beruhte daher auf der Sanger Sequenzierung einer Stichprobe transkriptomischer Information mittels sog. EST-Sequenzierungen (EST = „*Expressed Sequence Tag*“) [43]. Dabei wird der proteinkodierende „*messenger*“ RNA (mRNA) - Anteil der Gesamt - RNA aus den Zielzellen mittels einer reversen Transkriptase vom poly-A Ende der mRNA ausgehend in komplementäre DNA (cDNA) umgeschrieben. Diese wird mit geeigneten Plasmiden kloniert bzw. in Bakterien vermehrt. Von einem im Plasmid liegendem Sequenzmotiv ausgehend können die EST im Sanger-Verfahren sequenziert werden. Zur Identifikation werden die Sequenzen mit den in

öffentlichen Datenbanken verfügbaren Referenzen abgeglichen. Je mehr EST sequenziert werden, desto detaillierter wird die Genexpression abgebildet. Schwach exprimierte Gene werden in Abhängigkeit zur Größe der Stichprobe eventuell nicht erfasst. Die genomische Sequenz der Zielzellen muss bei diesem Verfahren nicht bekannt sein, was EST Sequenzierprojekte auch für nicht Modellorganismen einsetzbar macht. Zudem hat die Methode den Vorteil, daß sowohl Spleißvarianten, als auch unbekannte Transkripte detektiert werden können. Dem klassischen EST Projekt ähnlich ist das sog. SAGE („*S*erial *A*nalysis of *G*ene *E*xpression“) Verfahren [44]. Hierbei werden statt cDNA Molekülen voller Länge nur kurze Fragmente (10-26 bp) kloniert und sequenziert, was im Vergleich zum klassischen EST Projekt sowohl Kosten als auch Zeit spart [45]. Ein weiterer Schritt zu Genexpressionsanalysen mit hoher Auflösung bei vertretbaren Kosten wurde durch das 1995 von Schena und Kollegen entwickelte Microarray Verfahren möglich [46]. Dabei werden auf einem speziell beschichteten Glasobjektträger DNA-Sonden in einem festgelegten Arraymuster immobilisiert. Die Herstellung der Microarrays kann durch sog. Spotter mit speziellen Nadeln erfolgen, die von einem Roboter in die Sondenlösung getaucht werden und die Sonden anschließend punktgenau auf dem Objektträger platzieren (sog. „*S*potting“). Auch ein Aufdrucken der Sonden ähnlich eines Tintenstrahldrucks ist möglich. Wenn kurze Oligonukleotide als Sonde dienen sollen, können diese auch durch ein photolithografisches Verfahren direkt auf dem Objektträger synthetisiert werden. Im Falle der Nutzung eines Microarrays zur Genexpressionsanalyse ist die DNA-Sequenz der Sonden transkriptkodierend und muss bekannt sein. An diese Sonden werden in einem nachfolgenden Schritt fluoreszenzmarkierte cDNA Moleküle spezifisch hybridisiert (siehe Abbildung 1 - 7). Jedes cDNA Molekül kann unter den richtigen Hybridisierungsbedingungen nur an die jeweils komplementäre Sonde binden. Nach der Hybridisierung kann mittels eines Laserscanners die Intensität des Signals an allen Sondenpositionen ausgelesen werden. Je heller das Signal an einer Sonde ist, desto mehr cDNA wurde dort hybridisiert und desto stärker wurde das entsprechende Transkript in den Zielzellen exprimiert. Die Menge an Sonden-DNA, die auf dem Objektträger zur Hybridisierung zur Verfügung steht, darf nicht limitierend sein. Daher ist eine vergleichende Analyse der Genexpression zweier unterschiedlicher Proben auf demselben Microarray möglich. Die cDNA aus den zu vergleichenden Zielzellen wird mit zwei unterschiedlichen Fluoreszenzfarbstoffen markiert und die Hybridisierung an die Sonden des Arrays erfolgt simultan und kompetitiv. Microarrays wurden für unterschiedliche Fragestellungen über die

Jahre stetig weiter entwickelt, heutzutage ist die Analyse aller bekannten Transkripte zumindest für die gängigen Modellorganismen mittels Microarrays möglich. Hierbei müssen die unterschiedlichen Spleißvarianten eines Gens mit einzelnen Sonden abgedeckt werden, die Entdeckung neuer Gene ist damit nicht möglich.

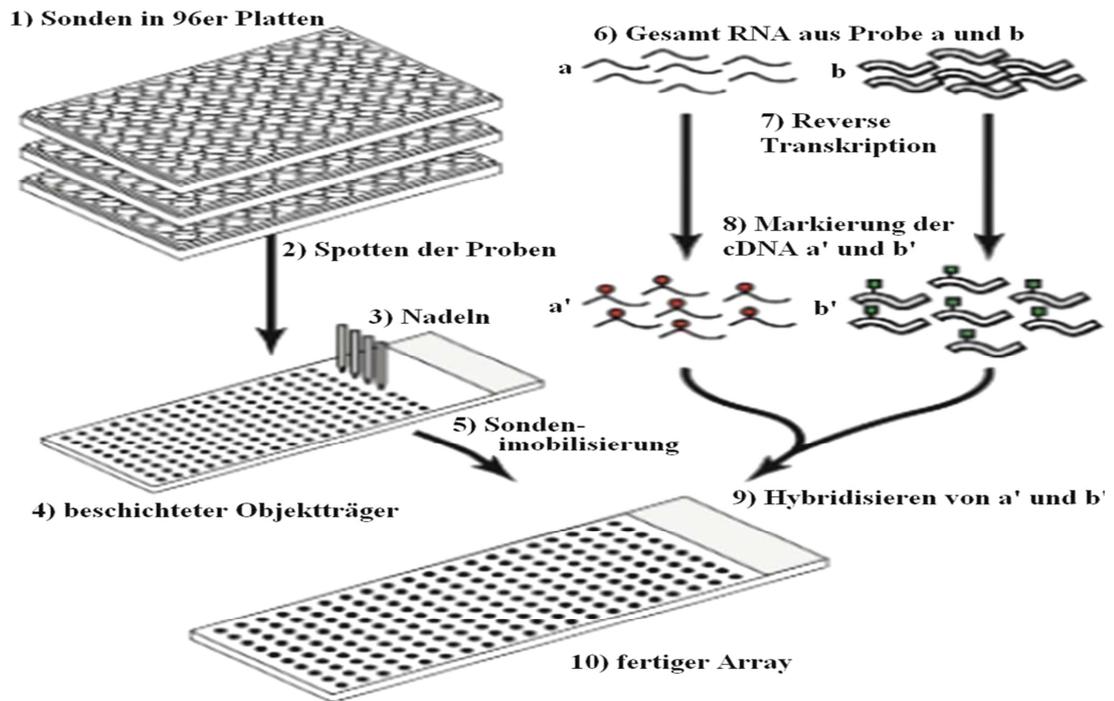


Abbildung 1 - 7 Konstruktion und Hybridisierung eines DNA – Microarrays

Konstruktion eines Microarrays und die nötigen Schritte der Probenbehandlung bis zur Hybridisierung. Aus Microtiter (1) werden die Sonden vom Spottingroboter (2) mittels spezieller Nadeln (3) auf den beschichteten Objektträger (4) gespottet und durch UV-Strahlung immobilisiert (5). Die Konstruktion des Microarrays ist damit abgeschlossen (10). Die mRNA der zu vergleichenden Proben a und b (6) wird durch reverse Transkription mit Oligo-dT-Primern (7) zu cDNA a' und b' umgeschrieben und dabei fluoreszenzmarkiert (8). Die unterschiedlich markierten Proben werden auf den fertigen Array (10) hybridisiert (9). Abbildung aus eigener Diplomarbeit übernommen, dort verändert nach Duggan et al. [47].

Zudem sind der Sensitivität des Verfahrens Grenzen gesetzt, was sich besonders bei sehr schwach (kein Signal) und sehr stark exprimierten Genen (Sättigungsgrenze) bemerkbar macht. Auch ist eine gewisse Redundanz in den Versuchen aus statistischen Gründen angebracht, die Rohdaten müssen vor Verwendung noch normalisiert werden. Durch die Entwicklung der Sequenziertechnologien der nächsten Generation wurden die Möglichkeiten der Genexpressionsanalyse noch einmal deutlich erweitert. Dies hängt zum einen mit dem enormen Zuwachs der Größe der öffentlichen Datenbanken zusammen, in welchen DNA Sequenzen für

vergleichende Analysen zugänglich gemacht werden. Für eine Vielzahl von Organismen und natürlich den Menschen sind heutzutage die Sequenzen kompletter Genome mit einer sehr guten Annotation der darin befindlichen Gene verfügbar. Die „GenBank“ Datenbank des NCBI umfasst derzeit in ihrer 200sten Version vom Februar 2014 ca. $1,5E^{11}$ Basenpaare an Erbinformation. Genome werden in einer eigenen Sparte organisiert, hier stehen die Nukleotidsequenzen kompletter eukaryotischer Genome von 225 Spezies zur Verfügung. Im Durchschnitt verdoppelt sich der Datenbestand in „GenBank“ seit deren Bestehen im Jahr 1982 etwa alle 18 Monate. Wie die nachstehende Grafik verdeutlicht (Abbildung 1 - 8), konnte ein solcher Zuwachs nicht allein mit der Aufstockung der Maschinenparks im klassischen Sanger Verfahren bewerkstelligt werden.

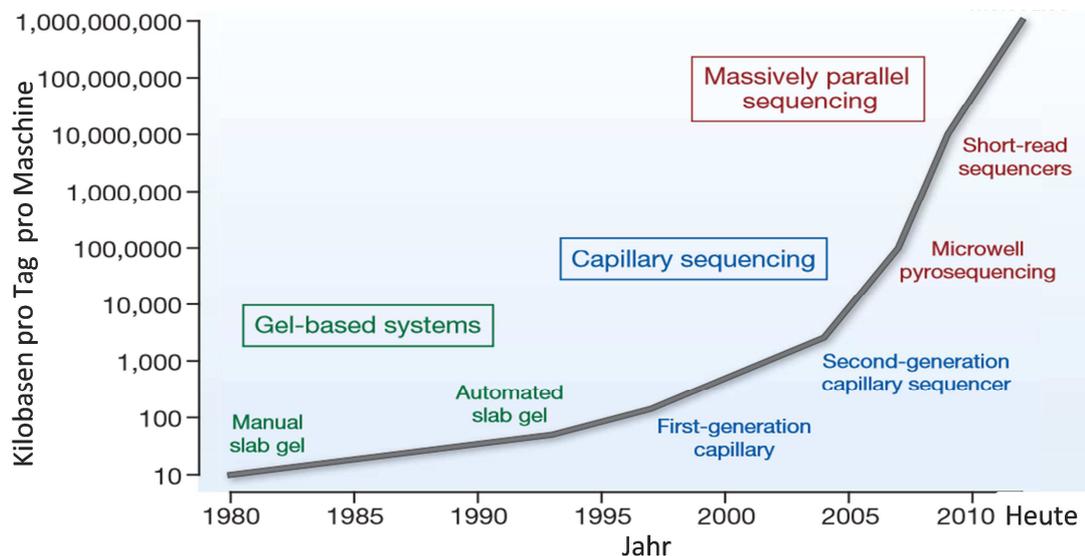


Abbildung 1 - 8 Entwicklung der DNA – Sequenzierung – Der Trend ist ungebrochen

Die Abbildung zeigt, wie sich die Effizienz der unterschiedlichen Sequenzieretechniken über die Zeit massiv steigern ließ. In Grün sind auf Acrylamid-Gelen basierende Systeme, wie sie nach der Entwicklung der Methode durch Sanger Verwendung fanden, aufgeführt. Mit zunehmender Automatisierung und durch Einsatz von Kapillargelelektrophorese ließ sich eine deutliche Steigerung der Tagesleistung eines Gerätes im Sanger Verfahren erreichen (blaue Farbe). Dank der stetigen Weiterentwicklung der NGS Techniken (rote Farbe) ist es heute möglich ein humanes Genom für 1000 \$ in wenigen Tagen zu sequenzieren (Illumina HiSeq X Ten™ System). Abb. verändert nach Stratton et. al [48].

Um das Jahr 2006 erreichten mehrere Sequenzieretechniken der nächsten Generation die Marktreife. Ihnen allen ist gemein, daß die Kosten pro Megabase sequenzierter DNA, sowie die Menge an sequenzierten Nukleotiden pro Lauf den im klassischen Sanger Verfahren erreichbaren Werten deutlich überlegen sind. Ebenso muss die zu sequenzierende DNA/cDNA für alle unterschiedlichen Verfahren in eine geeignete

Form gebracht werden, man erzeugt sog. „*Libraries*“. Bei deren Erstellung werden meist eine Amplifikation des zu sequenzierenden Materials, sowie eine Größenselektion der Fragmente passend zur Leseweite der verwendeten Technik durchgeführt. Ferner werden für den Vorgang der Sequenzierung an sich benötigte, kurze DNA Adaptoren an die DNA angefügt [49]. Auch zur Identifizierung einzelner Proben aus einem Probengemisch das gleichzeitig sequenziert werden soll, können Adapter als DNA – Barcode angefügt werden. Die von der Firma Roche vermarktete 454/FLX Technologie basiert auf einer Emulsions – PCR an magnetischen Kügelchen (sog. emPCR an „magnetic beads“) [50], die Sequenzierung erfolgt in einer miniaturisierten Mikrotiter Platte im Pyrosequenzierverfahren [51]. Dabei wird pro „*Bead*“ statistisch nur ein DNA Molekül gebunden und klonal amplifiziert, das Verteilen einzelner Beads in je eine Vertiefungen der Platte ermöglicht eine räumliche Trennung und Detektion der einzelnen Sequenzierreaktionen. Es handelt sich um eine Sequenzierung durch Neusynthese des zu sequenzierenden Strangs (SBS „*Sequencing by synthesis*“), aufgeteilt in vier Zyklen. Der DNA-Polymerase steht immer nur ein Nukleotid eines Typs zur Verfügung. Bei der Pyrosequenzierung wird Chemilumineszenz detektiert, welche beim Abspalten des Pyrophosphates nach Strangelongation durch die Polymerase frei wird. Die Menge an freigesetztem Pyrophosphat wird enzymatisch durch eine Sulfurylase/Luciferase Reaktion bei der Umwandlung von APS zu ATP und in Folge dessen von Luciferin zu Oxyluciferin quantitativ in Form von Lichtblitzen detektierbar. Der große Vorteil der Methode ist die im Verhältnis zu den übrigen NGS Technologien mit bis zu 1000 Basen recht lange Leseweite, die Datenmenge eines Laufes liegt bei ca. 700 MB generiert in 24 h. DNA Fragmente können in beide Richtungen (sog. „*paired end sequencing*“) sequenziert werden. Die Roche 454/FLX Methode weist deutliche Schwächen bei der Sequenzierung von Homopolymeren auf, da die Quantifizierung der Lichtblitze ab einem Einbau von mehr als 5 aufeinanderfolgenden identischen Basen nicht mehr sicher möglich ist. Auch die Probenvorbereitung mittels Emulsions-PCR erscheint technisch anspruchsvoll, zudem ist das Verfahren im Vergleich zu Konkurrenzprodukten mit ungefähr 10 \$ pro Megabase am kostenintensivsten [52]. Roche wird diese erste kommerziell erfolgreiche NGS Technologie nur noch bis ins Jahr 2016 hinein unterstützen. Nur wenig später als die 454 Technologie wurde die heute am weitesten verbreitete Sequenziertechnik der nächsten Generation auf dem Markt verfügbar. Auch hierbei handelt es sich technisch um einen SBS Ansatz, der ursprünglich durch die Firma Solexa entwickelt wurde, seine heutige Marktposition

jedoch erst ab 2007 nach der Übernahme von Solexa durch die Firma Illumina erlangte. Zu sequenzierende DNA Fragmente werden auf einem Glasobjektträger, der sog. „Flow Cell“, in einem zufälligen Verteilungsmuster immobilisiert. Um für die spätere Detektion der Fluoreszenz bei der Sequenzierung genügend Material zur Verfügung zu haben, werden diese Moleküle in einem als „Bridge Amplification“ bezeichneten Prozess am Ort ihrer Immobilisierung in einem PCR Verfahren klonal amplifiziert [53]. Die DNA wird hierfür zunächst modifiziert, ein typisches Fragment, das auf die „Flow Cell“ in ein Illumina Gerätes geladen wird, ist schematisch in Abbildung 1 - 9 dargestellt. Die doppelsträngigen Fragmente können in beiden Richtungen sequenziert werden, wobei sich in Abhängigkeit von der Fragmentgröße und Leseweite nicht zwingend eine Überlappung ergibt. Die derzeit maximale Leseweite auf Illumina Geräten beträgt 2 X 300 bp, für Hochdurchsatz Sequenzierungen sind 50 – 125 bp üblich.



Abbildung 1 - 9 Ein DNA Fragment aus einer Illumina „Library“

Zu sequenzierende DNA/cDNA Fragmente (grün) werden im Rahmen der „Library“ Erstellung auf eine geeignete Größe selektiert (~ 300 – 700 bp). Für die „Bridge Amplification“ und die eigentliche Sequenzierung, sowie zur Identifikation aller Sequenzen einer „Library“ in einem Probengemisch (optional), werden Adaptersequenzen durch Ligation oder durch ein auf Transposons beruhendes Verfahren, jeweils in Kombination mit einer nachgeschalteten PCR, angefügt [54]. Die Adapter P5 (rosa) und P7 (hellblau) dienen der Hybridisierung an Oligomere auf der Flow Cell. Der Index (lila) ist eine Art Barcode aus meist 8 Basen und dient der Probenidentifizierung. Die Sequenzierprimer für den Index und die reverse Sequenzierung sind gegensätzlich orientiert und binden im selben Motiv (dunkelblau). Die erste Sequenz startet an einem Primer der am als Read1SeqP bezeichneten Bereich bindet (orange). Die Pfeile geben die Orientierung der Primer an (5'-3').

Die Amplifikation und Sequenzierung der oben abgebildeten Fragmente auf der „Flow Cell“ erfolgt in mehreren Schritten. Zunächst werden die Fragmente an fest auf der Oberfläche des Glasobjektträgers verankerte Oligonukleotide hybridisiert. Es gibt zwei Typen von Oligonukleotiden auf der Oberfläche der „Flow Cell“ die entweder zum P5 oder zum P7 Adapter komplementär sind. Am freien 3'OH Ende dieser Oligonukleotide kann nun eine Elongationsreaktion mittels DNA – Polymerase erfolgen, der neu synthetisierte Strang ist komplementär zum ursprünglich hybridisierten Fragment, welches abgeschmolzen wird. Die freien Enden der nun fest verankerten Fragmente können durch Bildung einer Brücke an

die jeweils passenden Oligonukleotide auf der „FlowCell“ hybridisieren. Eine erneute Elongationsphase erlaubt die Generierung von komplementären Kopien. Es bilden sich nach mehreren Wiederholungen sog. Cluster aus, das sind klonal amplifizierte Fragmente desselben Moleküls, die räumlich getrennt vom Cluster anderer Fragmente auf der Oberfläche verteilt auftreten. Zur Sequenzierung werden zunächst enzymatisch alle DNA Stränge wieder von der Oberfläche entfernt, die in Ihrer Orientierung nicht das Adaptermotiv für den ersten Sequenzierprimer auf der der „Flow Cell“ abgewandten Seite haben (siehe Abbildung 1 - 10). So wird eine unidirektionale Sequenzierung aller verbleibenden Fragmente gewährleistet.

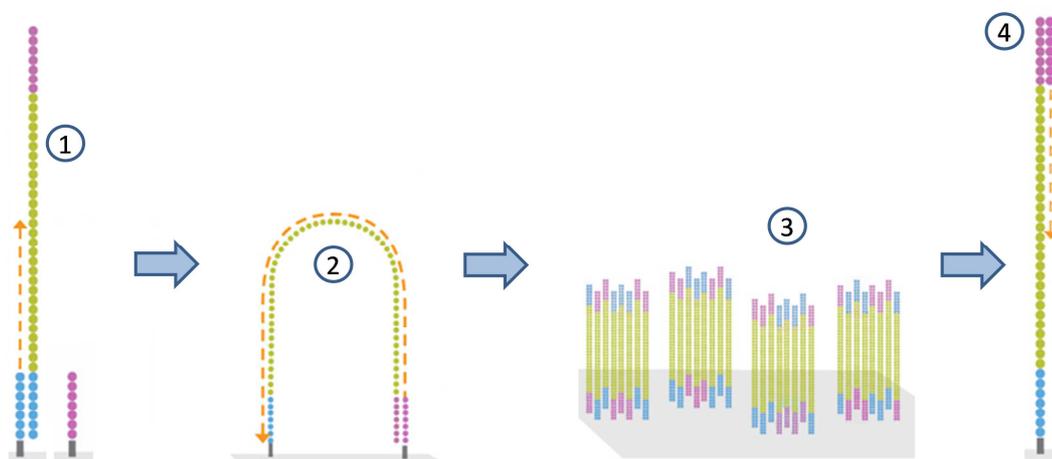


Abbildung 1 - 10 Illumina „Bridge Amplification“ und Sequenzierung

Die Abbildung zeigt schematisch den Ablauf der „Bridge Amplification“ bis zum Start der Sequenzierung. Fragmente hybridisieren an zu den Adaptern P5/P7 komplementären Oligonukleotiden auf der „Flow Cell“. Durch Elongation werden Kopien der Fragmente kovalent auf der Oberfläche gebunden (1). Es bildet sich eine Brücke zu den zum anderen Ende der Fragmente komplementären Oligonukleotiden auf der „Flow Cell“ aus (2), es folgt eine mehrzyklige klonale Amplifikation in deren Folge räumlich getrennte Cluster vieler Fragmente entstehen (3). Durch enzymatische Spaltung werden nur Fragmente der gewünschten Orientierung auf der „Flow Cell“ belassen, nach Denaturierung und Hybridisierung des Sequenzierprimers beginnt die eigentliche Sequenzierung (4).

Nachdem die Fragmente denaturiert sind, werden zunächst sämtliche freie 3'OH Gruppen inaktiviert, anschließend wird der erste Sequenzierprimer hybridisiert. Die Sequenzierung erfolgt in aufeinander folgenden Zyklen standardgemäß mittels 4 Fluoreszenzfarbstoffen (2 Farbstoffen im Illumina Gerät „Next Seq 500“ mit leicht abgewandelter Technik) ähnlich dem Sanger Verfahren. Es werden reversible Terminatoren verwendet, dies gewährleistet eine Sequenzierung Base um Base.

Homopolymere stellen entsprechend kein Problem dar. Die Detektion erfolgt über eine hochauflösende Kamera, die Position jedes Clusters auf der „Flow Cell“ wird dabei exakt bestimmt. Durch Übereinanderlegen der Bilder aus jedem Zyklus kann eine Sequenz zusammengesetzt werden. Für eine Sequenzierung der Fragmente von beiden Seiten aus (sog. „paired end“ Modus) wird die „Bridge Amplification“ nach einer festgelegten Anzahl Zyklen (= sequenzierte Basen) wiederholt, durch enzymatischen Verdau werden diesmal die in der ersten Runde nicht sequenzierten Fragmente auf der „Flow Cell“ belassen. Aus technischen Gründen sowie in Abhängigkeit zum verwendeten Illumina Gerät und der gewählten Fragmentgröße sind derzeit Leseweiten von 25 – 300 Basen in einer Richtung möglich. Hochdurchsatzgeräte wie der HiSeq 2500 sequenzieren binnen 6 Tagen eine Terabase DNA auf zwei Flow Cells parallel im „paired end“ Modus. Jedes Fragment wird dann von beiden Seiten 125 Basen weit sequenziert. Die Illumina Systeme sind die derzeit kosteneffektivsten auf dem Markt, der Preis pro Megabase sequenzierter DNA liegt im geschilderten Experiment auf dem HiSeq 2500 in etwa bei 0,03 \$ [55]. Die Firma Illumina hat unlängst angekündigt mit dem Gerätepark vertrieben unter dem Namen HiSeq X Ten™ erste das menschliche Genom im Endkundenbereich für 1000 \$ sequenzieren zu können. Die Investitionskosten liegen allerdings bei 10 Mio \$ und das Gerät ist derzeit nur für ausgesuchte Kunden verfügbar. Zudem wird es nur für die Sequenzierung menschlicher DNA vom Hersteller unterstützt. Neben den im Detail näher erläuterten NGS Methoden sind noch mindestens 3 weitere technische Verfahren kommerziell erfolgreich. Das als SOLID™ („Sequencing by Oligo Ligation and Detection“) bezeichnete System der Firma ABI/Life Technologies basiert, ähnlich wie das Roche 454 Verfahren, auf einer emPCR an magnetischen Kügelchen. Die Sequenzierung erfolgt durch die Ligation fluoreszenzmarkierter, degenerierter Oligonukleotid – Octamere an einen Sequenzierprimer [56]. Es handelt sich um ein 4-Farben Fluoreszenzsystem, bei dem die zwei ersten Basenfolgen eines Octamers durch einen Farbstoff kodiert werden. Ligation und Detektion erfolgen in mehreren Runden, wobei der Sequenzierprimer am 5' Ende um eine Base auf dem Template verschoben wird (n-1 System), sobald die gewünschte Leseweite ausgehend vom ersten Primer erreicht wurde. Aus den sich ergebenden Kombinationen nach 5 Primer-Verschiebungen kann die Zielsequenz mit äußerst geringer Fehlerrate bestimmt werden, da jede Position redundant sequenziert wurde. Das System „SOLID 5500 XL“ sequenziert binnen 8 Tagen 155 GB DNA, die Leseweite liegt bei 110 bp und die Technik ist mit 0,07 \$ pro Megabase durchaus konkurrenzfähig [55]. Die Firma Ion Torrent entwickelte ein weiteres

Verfahren, bei dem ebenfalls zunächst eine emPCR zur Probenvorbereitung eingesetzt wird. Die Sequenzieretechnik kommt ohne Fluoreszenzfarbstoffe aus, welche einen wesentlichen Kostenfaktor bei den übrigen Techniken darstellen. Das Auslesen der Sequenzinformation erfolgt mittels Halbleitertechnik und einem ionensensitiven Feldeffekttransistors. Dies ermöglicht ähnlich einem pH Meter die Detektion des Protons welches beim enzymatischen Einbau eines Nukleotids im Rahmen der Elongation eines DNA Stranges frei wird [57]. Verkauft werden Chips mit unterschiedlicher Sequenzierkapazität, die technisch gesehen quasi den Sequenzierer darstellen. Die Sequenzierung ist in wenigen Stunden abgeschlossen, die Leseweite liegt bei bis zu 400 bp. Den größten Durchsatz liefert derzeit das Modell „Proton I“ mit ca. 12,5 GB Sequenzdaten pro Lauf mit 175 bp Sequenzlänge für 0,08 \$ pro Megabase DNA [55]. Eine „paired end“ Sequenzierung ist mit Ion Torrent Technologie nicht möglich, zudem teilt sich die Technik mit dem Roche 454 Ansatz das Problem der inkorrekten Detektion von Homopolymeren im Zielfragment. Die maximale Leseweite ist bei fast allen NGS Techniken auf relativ kurze Sequenzlängen beschränkt. Einzig die Firma „Pacific Biosciences“ bietet derzeit eine Technik mit einer Leseweite von bis zu >30.000 bp an, und liegt damit sogar weit über den im klassischen Sanger Verfahren erreichbaren Werten. Das System arbeitet mit einer in einem Mikrokompartiment immobilisierten DNA-Polymerase. Die Sequenzierung erfolgt nach Zugabe von fluoreszenzmarkierten Nukleotiden und einem langen DNA Fragment („single molecule sequencing“) während der Replikation eines einzelnen DNA Stranges [58] in nur 2 Stunden Laufzeit. Die Anzahl an Sequenzen pro Lauf fällt entsprechend mit „nur“ 30.000 Stück vergleichsweise gering aus und der Preis pro Megabase DNA ist mit 1,11 \$ vergleichsweise hoch [55]. Die Vorteile für *de novo* Sequenzierprojekte gegenüber den anderen NGS Techniken sind aufgrund der Sequenzlänge natürlich erheblich, auch Haplotypisierungen, wie das Bestimmen von zusammengesetzt heterozygoten Genotypen, sind ohne weiteres möglich. Für Genexpressionsanalysen spielt die Methode hingegen kaum eine Rolle. Die Fragestellungen zu deren Beantwortung NGS Techniken eingesetzt werden können sind vielfältig. Neben der in dieser Arbeit durchgeführten Expressionsanalyse können natürlich neue Genome assembliert [59] oder bekannte Genome resequenziert werden [60]. Auch medizinische Fragestellungen, wie beispielsweise die Detektion von SNPs oder CNVs werden patientenindividuell durchgeführt [61]. Aufgrund der sehr tiefen Sequenzierung sind beispielsweise auch seltene somatische Varianten gut detektierbar, was besonders der Krebsforschung Möglichkeiten eröffnet [62]. Fast

allen NGS Techniken gemein ist, daß die für eine transkriptomweite Analyse der Genexpression benötigte Sequenziertiefe verhältnismäßig kostengünstig und in kurzen Zeitspannen erreicht wird. Dies führt wiederum zu technischen Herausforderungen diese immensen Mengen an Rohdaten in interpretierbare Werte umzurechnen.

1.4 BIOINFORMATISCHE AUSWERTUNG VON NGS DATEN ZUR GENEXPRESSIONSANALYSE

Die gängigen, für die Genexpressionsanalyse geeigneten NGS Verfahren liefern als Rohdaten Sequenzinformation im Gigabyte Maßstab, es fallen oft weit mehr als 100 Millionen Einzelsequenzen von 50 bis 150 Basenpaaren Länge an. Die Frage, welche Gene im Zielgewebe exprimiert werden und wie häufig ein Transkript gebildet wird, steht bei Genexpressionsanalysen meist im Vordergrund. Jede Sequenz muss hierfür einem Transkript sicher zugeordnet werden, anschließend kann die Anzahl der Sequenzen pro Transkript ähnlich wie beim klassischen EST Projekte „gezählt“ und die Daten gegebenenfalls normalisiert werden. Die aus der Prä-NGS Zeit bekannten Algorithmen leisten nach wie vor gute Dienste. So kann die Sequenzidentität beispielsweise durch den BLASTn/x Algorithmus [63] mit entsprechenden Datenbanken abgeglichen werden, limitierend ist hier vor allem die sehr hohe Rechenkapazität die benötigt wird, wenn die Analysen mit Großdatensätzen in annehmbarer Zeit durchgeführt werden sollen. Aus diesem Grund werden Analysen von NGS-Daten heute üblicherweise in einer hochparallelierten Umgebung durchgeführt, wobei die Verwendung von Computerclustern oder (Hoch)Parallelrechnern möglich ist und die Rechenzeit nachhaltig beschleunigt. Die altbekannten Algorithmen sind dennoch am ehesten für die Sequenzen, wie sie aus klassischen Sanger Projekten hervorgehen, geeignet. Es wurden daher neue Strategien zur Verarbeitung großer Mengen kurzer Sequenzen entwickelt. Diese beruhen auf der Verarbeitung kurzer Subfragmente (sog- K-Mere) jeder Sequenz auf der Basis von Hash Tabellen. Die mathematische Grundlage des am weitesten verbreiteten Algorithmus hierfür basiert auf dem bereits 1946 veröffentlichten De Bruijn Graphen [64]. Hiermit lassen sich kurze Sequenzen besonders effektiv zu einem Contig assemblieren [65], es sind viele Programme sowohl als freie Software [66] als auch kommerzieller Art erhältlich. Das Assemblieren ist die einzige Möglichkeit, ohne geeignete Referenz aus kleinen

Sequenzen, große Abschnitte oder sogar komplette Genome zusammzusetzen. Für cDNA Sequenzierungen können so im Idealfall vollständige Transkripte zusammengesetzt werden. Dies ermöglicht bislang unbeschriebene Transkripte oder Transkriptvarianten bereits bekannter Gene zu finden. Wesentlich schneller lassen sich NGS Sequenzen durch Abgleich mit bereits annotierten Referenzgenomen verarbeiten. Hier kommt ein als „Kartieren“ (engl. „*mapping*“) bezeichnetes Verfahren der Sequenzen zum Einsatz. Der große Vorteil liegt in den vergleichsweise geringen benötigten Rechenressourcen. Ein höherwertiger, handelsüblicher PC ist bereits in der Lage mittels Kartierung große NGS Datensätze binnen weniger Stunden bis Tage zu verarbeiten. Daher liegt hier auch der Schwerpunkt kommerzieller Programmpakete, die ohne bioinformatische Expertenkenntnisse die Analyse von NGS Daten erlauben. Für unterschiedliche Sequenzieretechniken und Datensätze können in den meisten Programmpaketen zur NGS Analyse entsprechend angepasste Einstellungen gewählt werden. Es muss jedoch bedacht werden, daß auch technische Begebenheiten einen großen Einfluß auf die Qualität der Ergebnisse haben können. So bewirkt beispielsweise die Art der cDNA Synthese mittels eines Oligo-T Nukleotides ausgehend vom Poly-A Schwanz der mRNA eine Bevorzugung von 3' nahe gelegenen Fragmenten, verursacht durch die Tendenz der Reversen Transkriptase, die cDNA Synthese schon vor Erreichen des 5' Endes eines Trankriptes zu beenden [67]. Eine Normalisierungsmethode, wie die Bildung eines RPKM Wertes [68] (engl. für Anzahl der Fragmente pro Kilobase Transkriptlänge pro eine Millionen Sequenzen) welcher die Länge der Transkripte beim Berechnen der Expressionsstärke berücksichtigt, kann so zu einer Verfälschung der Ergebnisse führen. Zudem enthalten cDNA Sequenzierungen oft einen gewissen Anteil ribosomaler RNA Sequenzen und auch transkribierte, repetitive Elemente. Mit einer geeigneten Strategie bei der Auswertung können solche technisch- bzw. biologischen Schwierigkeiten jedoch gut abgefangen werden.

1.5 ZIELSETZUNG

Mesenchymale Stammzellen (MSC) haben in frühen klinischen Tests unlängst ihren Nutzen in der Stammzelltherapie bei diversen schweren Erkrankungen des Menschen bewiesen [37]. Um die Mechanismen der Differenzierung und damit die potentiellen Anwendungsmöglichkeiten dieser Zellen für medizinische Zwecke besser verstehen zu können, ist umfassendes Wissen über die mit den Differenzierungsprozessen einhergehenden Veränderungen in der Genexpression dieser Stammzellen nötig. Knochenmark stellt eine ideale Quelle für native MSC dar, ist jedoch aus humanen Quellen nur in unzureichendem Maße verfügbar. Bislang publizierte Studien im Bereich der Stammzellforschung legen jedoch eine Übertragbarkeit der Ergebnisse zwischen Säugern nahe. Aus diesem Grund wurde entschieden, die vorliegende Arbeit an MSC des Rindes durchzuführen, das für die zu klärende Fragestellung in vielerlei Hinsicht einen geeigneten Modellorganismus darstellt. Genetisch weist es eine hohe Ähnlichkeit zum Menschen auf [69] und ist in den öffentlichen Datenbanken sowohl durch ein gut annotiertes Genom, als auch eine große Anzahl transkriptomischer und proteomischer Datensätze belegt. Zudem ist bovines Knochenmark durch die Schlachtung der Tiere zu Nahrungszwecken jederzeit gut verfügbar.

Das Ziel der vorliegenden Arbeit ist zunächst die Isolation und das Anlegen von undifferenzierten Primärzellkulturen Mesenchymaler Stammzellen aus dem Knochenmark des Rindes. Diese Stammzellen sollen *in vitro* zu Adipozyten, Chondrozyten und Osteoblasten differenziert werden. Der Erfolg der Differenzierungsexperimente wird anhand histologischer Färbungen nachgewiesen. Zudem soll die Expression zelllinienspezifischer Markergene, sowohl in den undifferenzierten MSC, als auch in den differenzierten Folgelinien, den Differenzierungsstatus der Zellen einzuschätzen helfen. Um ein das gesamte Transkriptom umfassendes Genexpressionsprofil erstellen zu können sollen sowohl eigenproduzierte Plasmid-DNA Microarrays, als auch Sequenzierverfahren der nächsten Generation (Illumina NGS) zum Einsatz kommen.

Der Schwerpunkt der Arbeit liegt auf dem Vergleich der Genexpressionsprofile differenzierter mit undifferenzierten MSC. Dabei werden möglichst kurz unter

Kulturbedingungen gehaltene, sowie eine Langzeitkultur undifferenzierter MSC getrennt voneinander untersucht. Typische Transkripte sollen gruppiert und mit bekannten biologischen Pathways und Genontologien abgeglichen werden. Zuletzt sollen noch nicht charakterisierte Transkripte identifiziert und auf Ihre Funktion hin bioinformatisch untersucht werden.

2 MATERIAL UND METHODEN

2.1 ISOLATION MESENCHYMALER STAMMZELLEN AUS DEM KNOCHENMARK DES RINDES

Bovines Knochenmark kann in Mengen zwischen 5 bis 30 Gramm leicht aus den Röhrenknochen der Extremitäten der Tiere isoliert werden. Die Isolation der Zellen erfolgte nach einem veränderten Protokoll von Bosnakowski et al. beschriebenen Methode [70]. In dieser Arbeit wurden ausschließlich Knochenmarksproben aus dem Schienbein (lat. *Tibia*) entnommen, diese stellten freundlicherweise die Schlachthöfe „Schlachthof Alzey“ (Alzey) und die „Großschlachtereier Frühwein“ (Brensbach) zur Verfügung. Die Scheinbeine wurden im oder knapp über dem Kniegelenk bis einschließlich des Hufes vom Körper abgetrennt und direkt nach der Tötung der Tiere gekühlt gelagert. Der Transport erfolgte stets auf Eis. Das Alter und Geschlecht der Tiere wurde von Seiten des Schlachthofes nicht protokolliert. Die geschlossenen Röhrenknochen wurden im Labor vom Huf, Fell und Gewebe frei präpariert und zunächst in Wasser gewaschen. Zur Desinfektion wurden die gereinigten Knochen für 3 min in ein Bad aus 70 % Ethanol gelegt und mehrfach hin und her geschwenkt. Die Knochen wurden sofort unter eine Sterilbank (Faster BHA 72, Ferrara) verbracht und mit einer handelsüblichen, zuvor hitzesterilisierten Metallsäge in der Mitte geteilt. Mit einem sterilen Spatel konnte das Knochenmark aus je einer Schienbeinhälfte in ein 50 ml Zentrifugenröhrchen (Sarstedt, Nürnberg) überführt werden. Das Gewicht des isolierten Knochenmarks wurde gemessen und die Probe bis zum Anlegen der Primärkultur auf Eis gekühlt. Um die Zellen von übrigen Bestandteilen des Knochenmarks abzutrennen, wurde jede Probe mit demselben Volumen 37°C warmen Anzuchtmedium (siehe Tabelle 2 - 1) aufgefüllt und anschließend für 10 min kräftig geschüttelt, bis sich eine homogene Suspension bildete. Die dabei in Lösung gehenden Zellen können durch 5 minütige Zentrifugation im Schwingrotor bei 300 g auf dem Boden des Zentrifugenröhrchens pelletiert werden (Eppendorf Typ 5702, Hamburg).

Tabelle 2 - 1 Primärkultur Anzuchtmedium MSC

Das Anzuchtmedium enthält für die ersten 2 Tage die doppelte übliche Menge Antibiotikum, ein Antimykotikum (alles PAA, Marburg) sowie einen Gerinnungshemmer (Sigma-Aldrich, München). Ab dem ersten Medienwechsel am Tag 2 wird dem Quantum 333 nur noch 1 X Pen/Strep zugesetzt.

Bestandteil	Konzentration	Funktion
Quantum 333	Basisbestandteil	Nährmedium
Penicillin/Streptomycin	2 X	Antibiotikum
Amphotericin B	1 X	Antimykotikum
Heparin	300 U/ μ l	Blutgerinnungshemmer

Der stark fetthaltige Überstand wurde mit einer Vakuumpumpe unter sterilen Bedingungen abgesaugt, das Pellet in 10 bis 15 ml Anzuchtmedium aufgenommen und für 2 Tage bei 37°C und 5 % CO₂ in 25 bzw. 75 cm² Zellkulturflaschen (Greiner Bio-One, Frickenhausen) mit gasdurchlässigem Deckel im Brutschrank (Jouan IG150, Unterhachingen) als Primärkultur (PK) in wasserdampfgesättigter Atmosphäre inkubiert. Nach 2 Tagen Inkubation wurde der erste Medienwechsel durchgeführt. Ab diesem Zeitpunkt wurden die Zellen in Quantum 333 versetzt mit 1 X Pen/Strep als Standardmedium gezogen. Da im Überstand erfahrungsgemäß nach 2 Tagen noch immer nennenswerte Mengen noch nicht adhärent angewachsener MSC in Suspension sind, wurde der Überstand mit 20 ml frischem Medium versetzt als Suspensions-Primärkultur (SPK) erneut in Kultur genommen, um die Ausbeute für Differenzierungen nutzbarer Zellen zu erhöhen. Die Zellkulturen wurden regelmäßig im Mikroskop (Carl Zeiss Axiovert 40 CFL, Jena) kontrolliert und die Fortschritte mit einer Digitalkamera (Canon Power Shot G9, Krefeld) dokumentiert. Medienwechsel erfolgten alle 2 bis 3 Tage, ab dem zweiten Medienwechseln wurden die Zellen mit 8 ml 37 °C warmen, Ca²⁺ und MG²⁺ freiem 1 X PBS Puffer (PAA, Marburg) gewaschen, um nicht adhärente Zellen und Debris aus der Kultur zu entfernen. Nach Anlegen der SPK Kulturen wurde der Überstand beim Medienwechsel stets verworfen. Eine Übersicht aller Zellkulturen befindet sich im digitalen Anhang D1.

2.2 CHARAKTERISIERUNG DER ZELLKULTUREN

Vor Beginn der Differenzierungsexperimente wurden die undifferenzierten Zellen durch spezifische Antikörperfärbung auf die Expression des für MSC typischen CD-Oberflächenantigens CD44 überprüft. Da MSC im Knochenmark vergesellschaftet mit hämatopoetischen Stammzellen auftreten, wurde die Reinheit der MSC

Primärkultur durch das Fehlen der Expression des für hämatopoetische Stammzellen typischen CD-Oberflächenantigens CD45 nachgewiesen [71-73]. Zur weiteren Charakterisierung wurde Total - RNA aus undifferenzierten Zellen isoliert, in cDNA transkribiert und die Expression von für MSC typischen Gene per RT-PCR nachgewiesen. Zur Isolation der Gesamt-RNA wurde das „High Pure RNA Isolation Kit“ (Roche Applied Science, Penzberg) oder das „NucleoSpin® Extract II Kit“ (Macherey-Nagel, Düren) gemäß der Herstellerangaben verwendet. Um den Zellen möglichst wenig Gelegenheit zu geben ihre Genexpressionsmuster bis zur Zelllyse zu verändern, wurden die Kulturflaschen direkt vor der RNA Isolation auf Eis gestellt. Die Zellen wurden mittels eines Schabers vom Flaschenboden abgelöst und direkt weiter verarbeitet. Die Qualitätsbewertung und Quantifizierung der Gesamt-RNA erfolgte mittels des Agilent 2100 Bioanalyzer Chip Systems „RNA 6000 Nano“ (Agilent Technologies, Santa Clara, USA) gemäß Herstellerangaben.

2.2.1 IMMUNOHISTOCHEMISCHE FÄRBUNGEN VON UNDIFFERENZIIERTEN MSC

Um einen Nachweis über die Expression und Translation MSC typischer CD Oberflächenantigene führen zu können, wurden die Zellen in 24er Lochplatten ausgesät (Greiner Bio-One, Frickenhausen) und bis zum Erreichen von ca. 70% Konfluenz unter Standardbedingungen angezogen (vergleiche Punkt 2.1). Das Kulturmedium wurde abgehoben und die Zellen anschließend zweimal für je 5 Minuten mit 1 X PBS gewaschen. Nach Entfernen des PBS wurden die Zellen mit eiskaltem Methanol bedeckt und für 8 min bei -20°C fixiert. Auf die Fixierung folgten erneut zwei je 5 minütige Waschschrte mit 1 X PBS. Unspezifische Bindestellen wurden durch eine 20 minütige Inkubation in 10% „Fetal Bovine Serum Gold“ (FBS) (PAA, Marburg) in 1 X PBS bei Raumtemperatur blockiert. Anschließend wurden die Zellen mit je 100 µl des primären Antikörpers (AK) bei 4°C über Nacht inkubiert (siehe Tabelle 2 - 2).

Tabelle 2 - 2 Übersicht der verwendeten Antikörper

Die Verdünnung der AK erfolgte in 1 X PBS / 10% FCS. Sämtliche AK stammen von der Fa. Acris, Hiddenhausen

Antikörper	Verdünnung	Expression MSC	Spezifität	Kopplung
CD44 (primär)	1:20	+	Schaf, Ziege, Rind	-
CD45 (primär)	1:20	-	Rind	-
Rabbit anti-mouse IgG (sek.)	3:100	-	Maus	FITC

Die 2,5 stündige Inkubation mit je 100 µl sekundärem Antikörper am Folgetag konnte direkt nach zwei weiteren fünfminütigen Waschschritten mit 1 X PBS durchgeführt werden. Um eine Abschätzung des Verhältnisses von Zellen zu angefärbten Zellen und damit der Reinheit der MSC Primärkultur erhalten zu können, wurde direkt im Anschluss an die Inkubation mit dem sek. AK eine DAPI Färbung der Zellkerne durchgeführt. Dies dient zusätzlich einer Kontrolle der Kulturen auf Mycoplasmen Kontaminationen [74]. Zur DAPI Färbung wurden die Zellen erneut für 2 X 5 min mit 1 X PBS gewaschen um ungebundenen sek. AK zu entfernen. Die Zellen wurden mit 100 µl in Wasser 1/1000 verdünnter DAPI Stammlösung (Roche Applied Science, Penzberg) für 10 min gefärbt. Vor dem Mikroskopieren wurde die DAPI Lösung durch zwei letzte, 5 minütige Waschschriffe mit 1 X PBS wieder entfernt.

2.2.2 CHARAKTERISIERUNG VON MSC DURCH NACHWEIS ZELLINIENTYPISCHER TRANSKRIPTE MITTELS RT-PCR UND SANGER SEQUENZIERUNG

Zum Nachweis der Expression MSC typischer Transkripte (sowie der Expression von differenzierungstypischen Markergenen) wurde der mRNA Anteil aus 1 µg Total - RNA in cDNA umgeschrieben. Hierfür wurde das Kit Super Script III (Invitrogen, Karlsruhe) genutzt. Das Priming erfolgte über Oligo(dT)₂₀-Primer.

Tabelle 2 - 3 cDNA Erststrangsynthese

Komponenten	Volumen
Gesamt-RNA	1 µg
dNTPs (10 mM)	1 µl
Oligo(dT) ₂₀ -Primer (50 µM)	1 µl
RNAse-freies H ₂ O	Auffüllen auf 13 µl TV
5 Minuten Denaturierung bei 65° C, 1 Minute auf Eis kühlen	
5 x First-Strand Puffer	4 µl
DTT (0,1 mM)	1 µl
RNase Out (40 U/µl) (RNAse Inhibitor)	1 µl
Superscript III RT (200 U/µl)	1 µl
60 min Elongation bei 50°C, 30 min Elongation bei 55°C, 20 min Inaktivierung bei 70°C	

Zur Durchführung genspezifischer PCRs wurde, je nach erwarteter Expressionsstärke des Zielgens, 1 µl unverdünntes bis 1:100 verdünntes Produkt

aus der cDNA Erstrangsynthese direkt in die Reaktion eingesetzt. Eine Liste aller verwendeten Primer ist nachfolgend in Tabelle 2 - 4 aufgeführt.

Tabelle 2 - 4 Primer für die RT - PCR

Gen	Primersequenz 5' - 3'
Bt_gapdh_for	AACACCCTCAAGATTGTCAG
Bt_gapdh_rev	CATGTGGAAGTCAGGAGATT
Bt_scf/kitlg_for	GATAAGCGAGATGGTGGAAAC
Bt_scf/kitlg_rev	ATTGAGACCACAACCTCCCTG
Bt_sox2_for	AATCCCATCCACACTCACG
Bt_sox2_rev	TACATGGATTCTCGGCAGAC
Bt_fgf2_for	GCACTGAAACGAACTGGGC
Bt_fgf2_rev	CAGGAGAGACCAGGGATG
Bt_lif_for	CAACATCACGAGAGACCAG
Bt_lif_rev	CCACAGCTCCCCTTCTAG
Bt_myc_for	AAACCTCCTCACAGCCCG
Bt_myc_rev	TCTGCTTGGACCGACAGG
Bt_klf4_for	ACTCGCTGTCCCATCAGG
Bt_klf4_rev	GTGCCTCTTCATGTGTAAGG
Bt_yy1_for	GAGCAGAAGCAGGTGCAG
Bt_yy1_rev	CTCCTGTTGGGACCACAC
Bt_yy2_for	ACGACAGCTACTTTGAGCAG
Bt_yy2_rev	GCACCAGTTGATGACGTTTC
Bt_oct4-Nested_for	AGTGGGTGGAGGAAGCTG
Bt_oct4-Nested_rev	TCGTTTGCCCTTCTGGCG
Bt_nanog_for	AAGACCTGGTTCCAGAACC
Bt_nanog_rev	TGTTCCAGGAGTGTTG
Bt_CD105_for	TGGGTCCAGGTTACAATGAC
Bt_CD105_rev	TCCTCGGAGAGCAGTAGC
Bt_CD45_for	CTACTAAGCCACCATGTGAG
Bt_CD45_rev	TCAGGACGTTTACAGCTCAC
Bt_CD44_for	CCCACTGAGAAGCGTCTC
Bt_CD44_rev	GGGTATTTCCAGAAGCATC

Die Primer wurden von Hand ausgesucht, Tabelle 2 - 12 listet die bioinformatischen Werkzeuge zum Berechnen der Annealingtemperatur und zum Aufklären von etwaigen Primer homo- und heterodimeren Strukturen auf. Alle Primer wurden mit einer Annealingtemperatur von ca. 58°C ausgewählt und mittels BLASTn Analysen auf Ihre Zielgen-Spezifität hin untersucht. Die PCR wurde stets gemäß den in Tabelle 2 - 5 aufgelisteten Ansätzen durchgeführt.

Tabelle 2 - 5 Pipettierchema PCR

Je nach Amplifizierbarkeit des Zielgens konnten unterschiedliche Polymerasen getestet werden.

Chemikalie	GoTaq	Sigma Taq	Klear Taq
HPLC-H ₂ O	31,5 µl	40,7 µl	40,8 µl
PCR-Puffer (5 X bzw. 10 X)	10 µl	5 µl	5 µl
dNTPs (10 mM)	1 µl	1 µl	1 µl
Primer F (10 pmol/µl)	1 µl	1 µl	1 µl
Primer R (10 pmol/µl)	1 µl	1 µl	1 µl
MgCl ₂ (25 mM)	4 µl	In Puffer enth.	In Puffer enth.
Polymerase	0,5 µl	0,3 µl	0,2 µl
cDNA (teilw. verdünnt)	1 µl	1 µl	1 µl
Gesamtvolumen	50 µl	50 µl	50 µl

Die Amplifikation fand je nach Verfügbarkeit in unterschiedlichen Thermocyclern statt (Primus 96 advanced Gradient Fa. PeqLab, Erlangen, Deutschland; Professional Thermocycler Fa. Biometra, Göttingen, Deutschland; PTC DNA Engine Fa. Bio-Rad, Hercules, USA). Einzelne Parameter wurden bei Bedarf verändert, das Standardprogramm für alle PCR-Reaktionen gibt Tabelle 2 - 6 wieder.

Tabelle 2 - 6 Standard PCR-Programm

Schritt	Temperatur	Dauer	Zyklen
Initiale Denaturierung	96°C	5 - 15 min	-
Denaturierung	96°C	15 sec	
Annealing	55°C	30 sec	35-40 x
Elongation	72°C	90 sec	
Finale Elongation	72°C	4 min	-
Aufbewahrung	16 °C	∞	-

Nach Abschluss der PCR wurde 1/10 Volumen der Reaktion auf ein 1 %iges Agarosegel aufgetragen (GENTerphorese, GENTerprise Mainz | 100 bp Marker Fermentas GeneRuler™ DNA Ladder Mix). Bei passender Größe wurde das PCR Produkt mittels Silica-Säulen gemäß Herstellerangaben aufgereinigt (Qiagen, Maccherey Nagel) und zur Sanger-Sequenzierung bei der Fa. Starseq (Mainz) eingereicht. Die Sequenzen wurden sowohl im Textformat als auch als Chromatogramme zur Verfügung gestellt. Zur Betrachtung der Chromatogramme wurde FinchTV der Fa. Geospiza (Seattle, USA) verwendet. Die Sequenzidentität wurde durch manuellen BLASTn/x Abgleich gegen die NCBI Datenbanken ermittelt.

2.3 DIFFERENZIERUNG MESENCHYMALER STAMMZELLEN

Die Primärkulturen (PK/SPK) bewachsen nach 5 - 9 Tagen konfluent den Boden der Zellkulturflaschen. Vor Beginn der Differenzierungsexperimente wurden die Zellen passagiert, um die Anzahl der zur Differenzierung und für Färbeversuche zur Verfügung stehenden Ansätze zu erhöhen. Zur Passage wurde das Medium aus den 75 cm² Zellkulturflaschen abgenommen und verworfen. Anschließend wurden die Zellen mit 10 ml PBS gewaschen, um verbleibende Medienreste sicher zu entfernen. Das PBS wurde abgenommen und der Boden der Zellkulturflaschen mit 9 ml Trypsin/EDTA-Lösung vollständig bedeckt. Es folgte eine 3 minütige Inkubation bei 37°C. Um das Trypsin zu inaktivieren, wurden die nun vom Boden abgelösten Zellen in 5 ml vorgewärmtem Medium aufgenommen und in 15 ml Zentrifugenröhrchen (Sarstedt, Nürnberg) überführt. Die Zellen wurden für 5 min bei 300g zentrifugiert, der Überstand abgehoben und verworfen und das Pellet in 12 ml Medium erneut aufgenommen. Zur Passage wurden die in 12 ml gelösten Zellen zu gleichen Teilen auf vier 25 cm² Zellkulturflaschen aufgeteilt und mit 3 ml Medium auf 6 ml Volumen aufgefüllt. Sobald diese Schalen wiederum beinahe konfluent bewachsen waren, konnten die Differenzierungsexperimente beginnen. Zur Bevorratung wurden einzelne Ansätze als Cryostocks in CryomaxS Medium (PAA, Marburg) in flüssigem Stickstoff konserviert. Die im späteren Verlauf der Experimente als „junge MSC“ (bzw. msc_neu) bezeichnete Kultur ist unpassagiert im konfluentem Zustand direkt zur RNA Extraktion verwendet worden. Die als „alte MSC“ (bzw. msc_alt) bezeichnete Kultur wurde nach der ersten Passage am neunten Tag über eine Dauer von 31 Tagen bei stetem Medienwechseln ebenfalls unpassagiert in einem mehrschichtigen, gewebeähnlichen Zellverband kultiviert.

2.3.1 *IN VITRO* DIFFERENZIERUNG VON UNDIFFERENZIIERTEN MSC ZU ADIPOZYTEN

Die Differenzierung zu Adipozyten erfolgte in 25 cm² Zellkulturflaschen. Die Zellen wurden bis zum Erreichen der Konfluenz in Quantum 333 Medium kultiviert, die Induktion der Differenzierung erfolgte durch Austausch des Standardmediums durch Adipozyten-Differenzierungsmedium [70]. Zur Differenzierung wurden zwei unterschiedliche Medien verwendet (siehe Tabelle 2 - 7). Auf drei aufeinanderfolgende Medienwechsel mit Adipogenese-Induktionsmedium erfolgte ein Medienwechseln mit Adipogenese-Maintenance-Medium. Medienwechsel

erfolgten alle 2 Tage über einen Zeitraum von 28 Tagen. Die Differenzierung erfolgte unter 5 % CO₂ in wasserdampfgesättigter Atmosphäre bei 37°C. Der Verlauf der Differenzierung wurde lichtmikroskopisch beobachtet und photographisch dokumentiert.

Tabelle 2 - 7 Differenzierungsmedium Adipogenese

Alle Differenzierungszusätze stammen von Sigma Aldrich (München). IBMX = Isobutyl-Methylxanthin

Adipogenese-Induktions-Medium		Adipogenese-Maintenance-Medium	
Bestandteil	Konzentration im Medium	Bestandteil	Konzentration im Medium
Quantum 333	Basisbestandteil	Quantum333	Basisbestandteil
Pen/Strep	1 X	Pen/Strep	1 X
Insulin	10 µg/ml	Insulin	10 µg/ml
Dexamethason	1 µM		
IBMX	0,5 mM		
Indomethacin	0,5 mM		

2.3.2 *IN VITRO* DIFFERENZIERUNG VON UNDIFFERENZIIERTEN MSC ZU OSTEOLASTEN

Die Differenzierung zu Osteoblasten erfolgte in 25 cm² Zellkulturflaschen. Die Zellen wurden bis zum Erreichen der Konfluenz in Quantum 333 Medium kultiviert. Die Induktion der osteogenen Differenzierung [70] erfolgte durch Inkubation undifferenzierter MSC in entsprechendem Differenzierungsmedium (siehe Tabelle 2 - 8). Medienwechsel erfolgten alle 2 Tage über einen Zeitraum von 25 Tagen.

Tabelle 2 - 8 Differenzierungsmedium Osteogenese

Alle Differenzierungszusätze stammen von Sigma Aldrich (München).

Bestandteil	Konzentration im Medium
Quantum 333	Basisbestandteil
Pen/Strep	1 X
Dexamethason	100 nM
Natrium-β-Glycerophosphat	10 mM
L-Ascorbinsäure-2-Phosphat	50 µM
1,25-Dihydroxyvitamin D3	10 nM

Die Differenzierung erfolgte unter 5 % CO₂ in wasserdampfgesättigter Atmosphäre bei 37°C. Der Verlauf der Differenzierung wurde lichtmikroskopisch beobachtet und photographisch dokumentiert.

2.3.3 *IN VITRO* DIFFERENZIERUNG VON UNDIFFERENZIIERTEN MSC ZU CHONDROZYTEN

Die Differenzierung zu Chondrozyten erfolgte in speziellen, das adhärenente Wachstum von Zellen fördernden Petrischalen des Typs „Advanced TC™“ durchgeführt (Greiner Bio-One, Frickenhausen). Es wurden Zellen der Passage 1 verwendet. Die Zellen wurden nach der Aussaat bis zum Erreichen der Konfluenz in Quantum 333 Medium kultiviert. Medienwechsel erfolgten 24 h nach der Aussaat, danach alle 2 Tage über einen Differenzierungszeitraum von 5 Tagen. Die Induktion der chondrogenen Differenzierung [70] erfolgte durch Inkubation der undifferenzierten MSC in entsprechendem Differenzierungsmedium (siehe Tabelle 2 - 8). Die Differenzierung erfolgte unter 5 % CO₂ in wasserdampfgesättigter Atmosphäre bei 37°C. Der Verlauf der Differenzierung wurde lichtmikroskopisch beobachtet und photographisch dokumentiert.

Tabelle 2 - 9 Differenzierungsmedium Chondrogenese

Alle Differenzierungszusätze außer der „Human Transforming Growth Factor β “ (Fa. Enzo Lifes Sciences, Farmingdale, NY, USA) stammen von Sigma Aldrich (München). ITS = Insulin, apo-Transferin, Sodium-Selenit.

Bestandteil	Konzentration im Medium
Quantum 333	Basisbestandteil
Pen/Strep	1 X
TGF β	5 ng/ml
L-Ascorbinsäure-2-Phosphat	50 μ g/ml
Natriumpyruvat	1 mM
BSA	1,25 mg/ml
Linolsäure	5,35 μ g/ml
ITS	Je 6,25 μ g/ml

2.4 CHARAKTERISIERUNG VON AUS MSC DIFFERENZIIERTEN FOLGELINIEN

Der Erfolg der Differenzierungen wurde zum einen durch das unter Punkt 2.2.2 beschriebene Vorgehen zum Nachweis der Expression von Markergenen mittels RT-PCR überprüft. Eine Liste der Primer befindet sich in Tabelle 2 - 4, die Zielgene sind unter Punkt 3.2.2 aufgeführt. Zusätzlich wurden differenzierte Zellen durch histologische Färbungen lichtmikroskopisch auf die Einlagerung von Lipiden (Adipogenese), das Ausbilden einer sauren, sulfatreiche Proteoglycane enthaltenden Matrix (Chondrogenese) und das Auftreten von Calciumablagerungen (Osteogenese) hin untersucht. Zur Kontrolle wurden stets gleich alte, undifferenzierte MSC, welche unter identischen Wachstumsbedingungen in Standardmedium kultiviert wurden, mit angefärbt. Es wurden unterschiedlich modifizierte Färbeprotokolle getestet. Nachfolgend sind die für die im Ergebnisteil gezeigten Abbildungen verwendeten Färbeprotokolle aufgeführt.

2.4.1 NACHWEIS CHONDROGENER DIFFERENZIERUNG DURCH ALCIAN BLUE FÄRBUNG

Die Färbung chondrogen differenzierter Zellen erfolgte gemäß einem nach Song et al. [75] veränderten Färbeprotokoll. Gefärbt wurden 9 Tage lang zu Chondrozyten differenzierte Zellen sowie gleichalte Kontrollen. Der Farbstoff Alcian Blue (Sigma Aldrich, München) färbt im Falle einer gelungenen chondrogenen Differenzierung saure Proteoglykane wie sie in der extrazellulären Matrix typischerweise zu finden sind blau an. Hierfür wurde das Kulturmedium abgenommen und die Kulturen dreimal je 5 Minuten mit 1 X PBS gewaschen. Darauf folgte eine Fixierung in 100% eiskaltem Methanol für 10 min bei -20 °C. Die Färbung fand durch 45 minütige Inkubation in 1 % Alcian Blue gelöst in 0,1 M HCL bei RT statt. Vor der photographischen Dokumentation im Lichtmikroskop wurden die Präparate viermal je 5 min mit 1 X PBS gewaschen. Die in dieser Arbeit gezeigten Aufnahmen stammen aus der Diplomarbeit von Jan Rettich (2011), IMSB Mainz.

2.4.2 NACHWEIS ADIPOGENER DIFFERENZIERUNG DURCH OIL RED O FÄRBUNG

Die Färbung adipogen differenzierter Zellen erfolgte gemäß einem nach Ramirez-Zacarias [76] veränderten Färbeprotokoll. Gefärbt wurden über einen Zeitraum von 21 Tagen adipogen differenzierte MSC sowie gleich alte Kontrollen. Zunächst wurde das Kulturmedium abgehoben und die Zellen anschließend in 4 %iger Paraformaldehyd (1 X PBS Lösung pH 7,2) für 40 min bei RT fixiert. Das Fixativ wurde durch dreimaliges Waschen mit 1 X PBS und zweimaligem Waschen mit VE-H₂O (5 min je Waschung) entfernt. Vor Zugabe der Färbelösung wurden sämtliche Flüssigkeitsreste entfernt und die trockene Kulturflasche kurz mit 60 % Isopropanol gespült. Die rote Färbung der für Fettzellen typischen Triglyceride und Lipide erfolgte durch 50 minütige Inkubation in 0,36 % Oil Red O (Sigma Aldrich, München) Lösung in 60 % Isopropanol. Die Dokumentation erfolgte ebenfalls lichtmikroskopisch. Die in dieser Arbeit gezeigten Aufnahmen stammen aus der Diplomarbeit von Anne Dubberke (2009), IMSB Mainz.

2.4.3 NACHWEIS OSTEOGENER DIFFERENZIERUNG DURCH ALZARIN RED S FÄRBUNG

Die Färbung osteogener Zellen und gleich alter Kontrollen erfolgte nach 21 Tagen Differenzierung. Alizarin Red S (Sigma Aldrich, München) färbt hierbei die für Osteoblasten typischen Calciumablagerungen der mineralisierten Matrix rot an. Das Kulturmedium wurde entfernt und die Zellen anschließend für 15 min in mit 1 X PBS frisch angesetzter 2,5 % Glutaraldehydlösung (Sigma Aldrich, München) bei RT fixiert. Anschließend wurden die Zellen zweimal für je 5 min mit 1 X PBS pH 4,2 gewaschen. Die Färbung erfolgte für 20 min in 2 % Alizarin Red S (gelöst in 1 X PBS pH 4,2) bei 37°C. Eine spezifische rote Färbung der Osteoblasten konnte nach mehrmaligen Waschungen in 1 X PBS pH 7,2 im Lichtmikroskop dokumentiert werden. Die in dieser Arbeit gezeigten Aufnahmen stammen aus der Diplomarbeit von Jan Rettich (2011), IMSB Mainz.

2.5 HERSTELLUNG, HYBRIDISIERUNG UND AUSWERTUNG VON PLASMID – DNA MICROARRAYS ZUR GENEXPRESSIONSANALYSE

Die Analyse der Genexpression erfolgte in dieser Arbeit neben der Illumina „Next Generation Sequencing“ Technologie auch durch Hybridisierung eigens angefertigte DNA Microarrays mit RNA aus den unterschiedlichen Zelllinien. Das Vorgehen entspricht dabei in leicht abgewandelter Form der durch Schena et al. [46] entwickelten Methode. Besonders hervorzuheben ist die Eigenart, bei der Herstellung der Microarrays auf den Objektträgern Plasmid - DNA zu spotten. Jedes Plasmid enthält als Integrat ein EST Molekül, welches aus bovinem Gewebe der fötalen Wachstumsfuge isoliert wurde. Das EST Projekt ist Teil der Promotionsarbeit von Steffen Rapp, der seine EST Bank dankeswerterweise zur Verfügung gestellt hat. Da durch Sanger Sequenzierung die Identität jedes der von Herrn Dr. Rapp gerichtet in pSPORT1 Vektor klonierten EST Moleküle bekannt war, konnte eine für diese Arbeit zielführende Auswahl an Plasmiden (\approx Transkripten) ausgesucht und durch Retransformation in chemisch kompetenten Zellen vermehrt werden. Die ursprüngliche EST Bank wurde so erhalten und gleichzeitig ausreichend Material für den Microarray gewonnen. Das als Grundlage für die Microarrays zur Verfügung gestellte EST Projekt umfasst annähernd 4000 sequenzierte Plasmide verteilt auf insgesamt fünfzig 96er Lochplatten. Von Interesse für Genexpressionsanalysen sind nur solche Plasmide, die als Integrat über die Differenzierung oder den Stammzellstatus potentiell aussagekräftige Transkripte aufweisen könnten. Zunächst folgte also eine Ausselektion all jener Plasmide, die kein Integrat aufweisen, deren Integrat für ribosomale RNA kodiert, deren Sequenzierung nicht erfolgreich verlief oder deren Integrat nur sehr wenige Basen aufweist. Die übrig gebliebenen Plasmide enthalten entweder in fötaler Wachstumsfuge exprimierte und eindeutig identifizierte Transkripte, oder aber Transkripte bislang unbekannter Funktion. Insgesamt wurden 2242 Plasmide zur Retransformation als Sonden für den Plasmid – DNA Microarray ausgewählt. Da ein EST Projekt die Expressionsstärke eines Gens anhand der Anzahl der cDNA Fragmente für dasselbe Transkript bestimmt, sind gerade in boviner fötaler Wachstumsfuge stark exprimierte Gene mehrfach redundant als Sonde auf dem Microarray vorhanden. Dabei sind Sonden für dasselbe Gen nicht notwendigerweise sequenzidentisch. Die fertig gespotteten Plasmid – DNA Microarrays wurden stets in der Kombination von cDNA aus undifferenzierten MSCs vs. cDNA differenzierter Zelllinien hybridisiert. Das EST Integrat in den

Plasmiden gewährleistet dabei eine genspezifische, kompetitive Hybridisierung der verglichenen RNAs. Da die RNA Menge, welche laut Herstellerangaben (Invitrogen, Carlsbad USA) in die cDNA Markierungsreaktion einzusetzen ist, mit 5 – 20 µg Gesamt-RNA oder wahlweise 0,4 – 2 µg mRNA verhältnismäßig groß ist, musste der mRNA Anteil der aus den Zellen isolierten Gesamt-RNA für diese Experimente in zwei aufeinander folgenden Reaktionen amplifiziert werden. Die so gewonnene, doppelt amplifizierte mRNA wurde auch im NGS Sequenzierverfahren eingesetzt.

2.5.1 GEWINNUNG DER PLASMID – DNA SONDEN DURCH RETRANSFORMATION CHEMISCH KOMPETENTER ZELLEN

Um die EST haltigen Plasmide retransformieren zu können, wurden zunächst chemisch kompetente Bakterienzellen hergestellt. Hierfür wurde ein einzelner Klon des Bakteriums *E. coli* Stamm RRI in einer 20 ml L-Medium Übernachtskultur bei 37°C ohne Antibiotikum angezchtet. Am folgenden Morgen wurden 10 ml frisches L-Medium mit 1 ml der über Nacht gewachsenen Bakterienkultur angeimpft und für 100 min bei 37°C geschüttelt. Diese Kultur wurde anschließend bei 4°C für 5 min bei 3500 UpM abzentrifugiert. Das Bakterienpellet wurde in 10 ml einer 10 mM eiskalten CaCl₂ Lösung resuspendiert und für 20 - 45 min auf Eis ruhen gelassen. Anschließend wurden die Zellen erneut für 5 min bei 3500 UpM zentrifugiert und das Pellet in 2 ml einer eiskalten 10 mM CaCl₂ resuspendiert. Die chemisch kompetenten Zellen können durch Einfrieren bei -80°C in wässriger 10 % Glycerinlösung für wenige Monate konserviert werden. Zur Transformation wurden zunächst 2 µl Plasmidlösung aus den Platten des EST Projekts in eine 2 ml 96er Lochplatte (Abgene, Epsom Großbritannien) vorgelegt und die Positionen elektronisch dokumentiert. Die 96er Lochplatte wurde auf Eis gekühlt und je Vertiefung 20 µl chemisch kompetente Zellen hinzugegeben. Nach einer Inkubationszeit von 45 min auf Eis erfolgte ein 3 min Hitzeschock bei 42°C. Anschließend wurden je Vertiefung 1,15 ml vorgewärmtes 2 x L – Medium versetzt mit Ampicillin (0,1 mg / ml) hinzugegeben. Die Inkubation der transformierten Zellen erfolgte für 18-24 Stunden auf einem Schüttler bei 37°C. Die 96er Lochplatten wurden für die Anzucht durch gasdurchlässige Klebefolien der Fa. Abgene sicher verschlossen. Die Bakterien konnten nach ausreichender Inkubationszeit auf dem Schüttler abzentrifugiert (Sigma 3-16KL, Sigma, Osterode) und die Pellets zur späteren Plasmidisolierung eingefroren werden.

2.5.2 ISOLATION VON PLASMID-DNA IM 96ER FORMAT

Plasmid – DNA wurde aus *E. coli* Zellen des Stamms RR1 nach dem Prinzip der alkalischen Lyse verändert nach Birnboim et al. [77] gewonnen. Zur Aufreinigung der Plasmide wurden mindestens 2 h vor Versuchsbeginn je 96er Lochplatte mit Bakterien eine 96er Sephadex™ G-50 Platte vorbereitet. Hierfür wurde mittels einer Lochschablone Sephadex™ G-50 der Körnungsstärke „Medium“ (Amersham, Amersham Großbritannien) in eine MultiScreen™ HV-Filterplatte (Millipore, Billerica, USA) gefüllt und mit 300 µl HPLC Wasser zum Quellen gebracht. Die Bakterienkulturen in den Anzuchtplatten wurden zunächst für 15 min bei 2750 UpM pelletiert und der Überstand zum Autoklavieren abgenommen. Die Pellets wurden in 335 µl Lösung I (für alle Lösungen siehe Punkt 2.8) vollständig resuspendiert, und die Zellen anschließend durch Zugabe von 665 µl frisch angesetzter Lösung II für 10 min auf Eis lysiert. Zur Neutralisierung wurden 500 µl Lösung III zugegeben, die Platten vorsichtig invertiert und für 15 min bei 4°C mit 3500 UpM zentrifugiert. Der Überstand ohne jegliches Präzipitat konnte in einer neuen 96er Deepwell Platte mit 925 µl Isopropanol gemischt, und die darin befindlichen Plasmide durch 45min Zentrifugation bei 18 °C bei 3500 UpM ausgefällt werden. Das Isopropanol wurde zunächst abgeschüttet und durch Anlegen eines Vakuums in einer Vakuumzentrifuge Concentrator 5301 (Eppendorf, Hamburg) bei stehendem Rotor und 30°C vollständig entfernt. Die Plasmide wurden in 50-70 µl HPLC Wasser gelöst. Zur Aufreinigung der Plasmide wurden die oben beschriebenen Sephadex™ Platten zunächst durch Zentrifugieren für 3 min bei 4°C und 2750 UpM getrocknet. Anschließend wurden direkt 150 µl HPLC Wasser hinzugegeben und der eben genannte Zentrifugationsschritt als Waschschrift wiederholt. Die so präparierten Sephadex Platten konnten mit den 50-70µl Plasmidlösung beladen, und die DNA durch einen weiteren Zentrifugationsschritt wie beschrieben aufgereinigt werden. Die Plasmide wurden in diesem letzten Schritt in 96er Microtiter Platten der Fa. Greiner Bio-One eluiert.

2.5.3 HERSTELLEN DER PLASMID-DNA MICROARRAYS DURCH SPOTTEN

Das Spotten der Plasmid DNA Microarrays erfolgte aus „Diamond Ultra“ 384er Lochplatten der Fa. Thermo Fisher Scientific (Waltham, MA USA). Die Plattenbelegung kann im digitalen Anhang D2 eingesehen werden. Vor dem

Umpipettieren aus dem 96er in das 384er Format mussten die Plasmide in den 96er Lochplatten in einer Vakuumzentrifuge bei stehendem Rotor und RT vollständig getrocknet werden. Die Plasmide wurden danach in 30 µl Spotting-Puffer nach Diel et al. [78] (3 X SSC, 1,5 M Betaine) gelöst und in insgesamt 6 Stück 384er Platte transferiert. Das Spotten erfolgte mittels eines Roboters Typ „GeneMachines Omnigrid“ der Fa. Genomic Solutions (Ann Arbor, MI USA) unter Benutzung von 4 Nadeln des Typs „Stealth SMP3™“ der Fa. Telechem (Sunnyvale, CA USA) auf Aminosilan beschichteten Objektträgern des Typs „UltraGAPS“ der Fa. Corning (Lowell, USA). Jede Plasmid DNA Sonde wurde in Triplikaten gespottet, der Sondenabstand auf dem Slide beträgt in horizontaler Richtung 400 µm, in vertikaler Richtung 300 µm. Jede der sechs 384er Platten wird in 4 Subarrays aus je 9 horizontalen und 32 vertikalen Sonden gespottet. Die obere, linke Ecke der Subarrays aus jeder 384er Platte wird durch Spotten von drei Spots aus einem cDNA Gemisch zur Orientierung in den Bilddateien bei der Auswertung markiert. Traten beim Spotten Lücken auf, wurde versucht diese gezielt durch Nachspotten der betroffenen Positionen zu schließen. Die Immobilisierung der Sonden auf den Objektträgern erfolgte durch Bestrahlen der getrockneten Microarrays mit UV Licht in einem Stratalink UV Inkubator der Fa. Stratagene (La Jolla, CA USA) gemäß den Herstellerangaben. Im Rahmen dieser Arbeit wurden 50 Microarrays mit je 6912 Spots aus 2304 Sonden hergestellt. Davon waren 2242 Sonden in pSPORT1 Vektor klonierte EST Fragmente, die restlichen Spots dienten der Kontrolle. Fertig gespottete Microarrays wurden nach der UV Bestrahlung unter Stickstoffatmosphäre bei RT gelagert. Zur späteren Auswertung wurde durch die Software des Spotters aus Listen der Plattenbelegungen eine sog. „gal“ Datei erstellt. Diese ermöglicht mittels eines Rasters bei der späteren Auswertung einzelne Spots in den Bilddateien der gescannten Microarrays den zugehörigen Plasmiden zuzuordnen. Die Datei ist im digitalen Anhang D2 abgelegt.

2.5.4 LINEARE AMPLIFIKATION VON MRNA IN ZWEI AUF EINANDERFOLGENDEN SCHRITTEN

Die vergleichende Hybridisierung sämtlicher Microarrays, sowie die Illumina NGS Sequenzierung der alten MSC, Osteoblasten und Chondrozyten, erfolgten unter Einsatz von doppelt amplifizierter mRNA. Besonders die Microarray Analysen benötigen große RNA Mengen um ausreichend starke Signale erzeugen zu können. Zur Amplifikation des mRNA Anteils der Gesamt-RNA (siehe Punkt 2.2) wurde für

beide aufeinanderfolgende Amplifikationsrunden das „SMART™ mRNA Amplification Kit“ der Fa. Clontech Laboratories (Mountain View, CA USA) genutzt. Das Kit ermöglicht die lineare Amplifikation des mRNA Anteils der Gesamt-RNA ausgehend von einem Oligo-dT Primer, die Herstellerangaben sind im digitalen Anhang D3 aufgeführt. Es wurden stets 3 µl Gesamt-RNA mit Gesamtmengen zwischen 0,378 µg und 4,734 µg RNA als Ausgangsmaterial in die erste Amplifikation eingesetzt. Der mRNA Anteil der gesamt RNA wird zunächst durch eine MMLV reverse Transkriptase in doppelsträngige cDNA umgeschrieben und dabei ein T7 Adapter auf dem Antisense Strang eingefügt. Anschließend erfolgt die Amplifikation des Sense Strangs durch eine in vitro Transkription mittels T7 RNA Polymerase für mindestens 12 h bei 37°C. Es entstehen lange mRNA Moleküle inklusive Poly A Schwanz, der in der zweiten Runde der Amplifikation erneut als Primerbindestelle für den Oligo-dT Primer verwendet werden kann. Das Kit besteht aus zwei separat erhältlichen Komponenten.

Tabelle 2 - 10 Anpassung im Clontech SMART™ Kit Schritt „Primer Extension“

Das gezeigte Pipettierschema wurde in beiden Amplifikationsrunden abweichend von den Herstellerangaben angewendet. Das Programm im Thermocycler entspricht den Herstellerangaben.

Herkunft	Chemikalie	Volumen
Roche	HPLC Wasser	55 µl
Promega	GoTaq PCR Buffer 5 X	20 µl
Promega	MgCl 25 mM	8 µl
Clontech	SMART™ dNTP Mix 50 X	2 µl
Clontech	SMART™ T7 Extension Primer 10 µM	2 µl
Clontech	SMART™ RNase H 10 U / µl	1 µl
GoTaq	GoTaq Polymerase	2 µl
Gesamtvolumen		90 µl

Die laut Herstellerangaben für die cDNA Zweitstrangsynthese („Primer Extension“) zu nutzenden Komponenten waren jedoch nicht vorrätig und wurden wie in nachfolgender Tabelle aufgeführt mit im Labor vorhandenen Materialien wo nötig ersetzt. Dabei wurde sich so eng wie möglich an das vom Hersteller vorgesehene Protokoll gehalten. Für die zweite Runde der Amplifikation werden 20 µl der einfach amplifizierten RNA (zwischen 2,88 µg bis 4,82 µg) mit HPLC Wasser auf 100 µl aufgefüllt, mit 2,5 Volumen reinem Ethanol sowie 1/10 Volumen 3 M Natriumacetat pH 4,1 (Sambrook 1989) versetzt und bei -20°C für mindestens 30 min oder über Nacht gefällt. Durch 45 minütige Zentrifugation bei 4°C konnte die

RNA Pelletiert werden. Das Pellet wurde in 70%igem Ethanol gewaschen und an der Luft getrocknet. Anschließend wurde die mRNA in 3 µl HPLC Wasser gelöst und in die zweite Runde der Amplifikation eingesetzt. Da sich im Rahmen der *in vitro* Transkription der zweiten Runde sehr viel amplifizierte RNA bildet, wird zur Aufreinigung der amplifizierten RNA erneut wie beschrieben gefällt. Der Waschschrift mit 70% Ethanol wird hierbei einmal wiederholt. Die RNA wird abschließend in 100 µl HPLC Wasser gelöst. Die Qualität und Quantität der eingesetzten Gesamt-RNA, sowie der einfach amplifizierten und der doppelt amplifizierten mRNA wird mittels eines Bioanalyser RNA Nano Chips überprüft.

2.5.5 HYBRIDISIERUNG, SCAN UND AUSLESEN DER PLASMID-DNA MICROARRAYS

Die Plasmid – DNA Microarrays wurden mit Alexa Fluor Farbstoffen markierter cDNA, welche aus doppelt amplifizierter RNA synthetisiert wurde, kompetitiv hybridisiert. Dabei wurden stets 2 µg doppelt amplifizierter RNA aus dem Versuchsansatz der alten MSC mit derselben Menge doppelt amplifizierter RNA aus einer der differenzierten Zelllinien gemeinsam auf einen Microarray hybridisiert. Für jede Vergleichsserie undifferenziert/differenziert wurden 10 Microarrays hybridisiert. Aufgrund von Materialknappheit konnten für den Vergleich alte MSC vs. Adipogenese nur 9 Microarrays hybridisiert werden. Um technisch bedingte Fehlerquellen bei der Hybridisierung aufgrund von nicht identischen Fluoreszenzeigenschaften der verwendeten Alexa Fluor Farbstoffe ausschließen zu können [79], wurden die Farbstoffe in der cDNA nach je 5 Experimenten vertauscht (sog. *dye swap*). Die reverse Transkription und Farbstoffmarkierung erfolgte durch das „SuperScript™ Plus Indirect cDNA Labeling System“ der Fa. Invitrogen gemäß Herstellerangaben (siehe Anhang D2). Zur cDNA Synthese wurden sowohl Oligo-dT als auch Hexamer Primer eingesetzt. Die Hybridisierung der Arrays wurde in der „a-Hyb Hybridization Station“ der Fa. Miltenyi Biotec (Bergisch Gladbach) durchgeführt. Die Hybridisierung erfolgte entsprechend der Herstellerangaben gemäß des für die Slides geltenden Protokolls „Application Protocol 07 – V01 – Corning UltraGAPS Oligonucleotide Gene Expression Analysis Microarrays“. Das Protokoll ist im Anhang D3 einsehbar. Die Hybridisierung erfolgte für min. 16 h bei 42°C. Die Microarrays wurden sofort nach Beendigung der Hybridisierung und der sich anschließenden Waschschriffe noch mit dem letzten Waschpuffer benetzt direkt in ein 50 ml Falcon befüllt mit 0,1 X SSC

überführt. Die Trocknung erfolgte so kurzfristig wie irgend möglich durch Abblasen mit reinem Stickstoff aus einer 200 Bar Flasche bei 1-3 Bar Gasdruck. Die streifenfrei getrockneten Microarrays wurden bis zum direkt im Anschluss durchgeführten Scan in einem mit Stickstoff befüllten 50 ml Falcon dunkel gelagert. Das Auslesen der Fluoreszenzsignale erfolgte mittels des „GenTAC™ LS4“ Scanners der Fa. Genomic Solutions und der zugehörigen Software GT LS. Die Microarrays wurden mit der bespotteten Seite vom Laser abgewandt in das Gerät gelegt. Da sich die Anregungswellenlänge für die verwendeten Alexa 555 (555 nm Anregung) und Alexa 647 (650 nm Anregung) Farbstoffe unterscheidet, konnte für jeden Farbkanal die Laserintensität so eingestellt werden, daß die stärksten Signale möglichst knapp unter der Sättigungsgrenze blieben. Gleichzeitig wurde mittels eines PMT Reglers (engl. *photo multiplier tube*) die Signalverstärkung mit dem besten Signal-zu-Hintergrund Verhältnis manuell eingestellt. Erst danach wurden die Fluoreszenzen aller Spots eines Microarrays als Summe von drei aufeinander folgenden Scans gemessen und verrechnet. Es entstand je Kanal eine 16 bit farbtiefe Bilddatei im .tif Format. Gescannt wurde mit einer Auflösung von 10 Pixeln. Die Zuordnung jedes Spots zu den in beiden Farbkanälen gemessenen Fluoreszenzintensitäten erfolgte mittels der Software GenePix™ Pro 6.0 der Fa. Molecular Devices (Silicon Valley, CA USA). Die Software ermöglicht ein Alignment der unter Punkt 2.5.3 erwähnten Rasterdatei mit den Bilddateien aus dem Scanner. Die Spots werden dabei automatisch erkannt (> 1000 Intensität Schwellenwert) und die Intensitäten unter Berücksichtigung jedes individuellen Spot - Umrisses ausgelesen. Die automatische Spoterkennung wurde gegebenenfalls manuell korrigiert, zudem konnten einzelne Spots schlechter Qualität für die spätere Auswertung markiert (engl. „flag“) werden. Das Ausgabeformat von GenePix™ Pro 6.0 („*gene pix results file*“ .gpr) ist eine mit Microsoft Excel kompatible Tabelle.

2.5.6 AUSWERTUNG DER PLASMID-DNA MICROARRAYS

Die Auswertung der Plasmid – DNA Microarrays erfolgte anhand der .gpr Ergebnisdateien in Microsoft Excel. Als Rohdatenwert für die Intensität eines Spots in jedem der beiden Farbkanäle wird die Vordergrundintensität abzüglich der Intensität des lokalen Spothintergrundes verwendet. Da nicht alle Spots gleich guter Qualität sind, bzw. nicht alle EST Moleküle nach Hybridisierung für die weitere Auswertung ausreichend starke und damit für Expressionsanalysen ausreichend

verlässliche Signale liefern, wurden zunächst geeignete Spots von ungeeigneten Spots getrennt. Spots die bei der Bearbeitung der Bilddateien (siehe Punkt 2.5.5) mit einem sog. Flag versehen wurden, wurden ebenso wenig ausgewertet wie solche, deren Signalintensität abzüglich des lokalen Hintergrunds einen Intensitätswert von 1000 in einem der beiden Farbkanäle unterschritten. Da die mittlere Farbintensität der beiden verwendeten Alexa Fluorfarbstoffe chemisch bedingt nicht gleich ist (der grüne Kanal liefert meist stärkere Signale als der rote), wurde eine globale Normalisierung der in die Auswertung eingehenden Messwerte vorgenommen [80]. Hierzu wurde für jeden Spot die Intensität des im Mittel weniger starken Kanals mit dem Quotienten gebildet aus dem Mittelwert der Intensität des stärkeren Kanals geteilt durch den Mittelwert der Intensität des schwächeren Kanals multipliziert. Zur Festlegung differenzieller Genexpression wurde der Quotient der global normalisierten Farbintensitäten jedes Spots zur Basis 2 logarithmiert [81]. Werden Werte von ≤ -1 bzw. ≥ 1 erreicht, gilt das Gen für das entsprechende EST Molekül des Spots kodiert als differenziell exprimiert. Der Expressionsunterschied für eine differenzielle Expression wurde so auf den Faktor 2 festgelegt (\log_2 von 2 = 1 | \log_2 von 0,5 = -1). Das Logarithmieren zur Basis zwei dient hierbei der besseren Skalierbarkeit der Expressionsquotienten. Wenn ein Gen in 8 von 10 hybridisierten Microarrays (7 von 9 im Vergleich MSC alt vs. Adipozyten) unabhängig vom sog. *dye swap* als in dieselbe Richtung differenziell exprimiert bestimmt wurde, galt die differenzielle Genexpression als ausreichend abgesichert.

2.6 GENEXPRESSIONSANALYSE MITTELS ILLUMINA NGS SEQUENZIERUNGEN (RNA – SEQ)

Um Informationen über die Genexpression des kompletten Transkriptoms zu erhalten wurde die RNA der verschiedenen Zelllinien mit der Illumina (San Diego, Californien, USA) Next Generation Technologie sequenziert. Die Arbeiten wurden durch den Sequenzierservice des in Köln ansässigen „Cologne Center for Genomics“ und im eigenen Haus auf einem Illumina GAIIX bzw. HighSeq 2000 Gerät durchgeführt. In beiden Fällen wurde dem Sequenzierservice entweder doppelt amplifiziert mRNA (alte MSC, Osteoblasten, Chondrozyten) bzw. Gesamt-RNA (junge MSC, Adipozyten) übergeben. Sämtliche nachfolgenden Arbeiten wurden vom jeweiligen Sequenzierservice durchgeführt. Die Sequenzierung der Adipozyten RNA ist eine 51 bp sog. „single ended“, die der übrigen RNAs eine 95

bp (alte MSC, Osteoblasten, Chondrozyten) bzw. 101 bp (junge MSC) sog. „paired end“ Sequenzierung. Die RNA der jungen MSC und der Adipozyten wurde in Mainz, die der übrigen Zelllinien in Köln sequenziert. Als Rohdaten wurden Sequenzen im .fastq (Adipozyten) bzw. .qseq (alle übrigen Sequenzierungen) Format zur Verfügung gestellt. Beide Formate enthalten neben der Position des Clusters auf der Flow Cell aus der die Sequenz stammt und der Sequenz an sich auch Qualitätswerte für die Bestimmung jeder einzelnen Base (*phred* Werte, ASCII-64). Das .qseq Format enthält zusätzlich unter anderem den PF Wert (siehe Punkt 2.6.1). Die Qualitätsinformationen wurden in einigen nachfolgenden Analyseschritten berücksichtigt, für manche Anwendungen wurden die Daten in das keine Qualitätsinformationen enthaltene .fasta Format umgewandelt. Um die einzelnen Sequenzen einem Gen zuordnen zu können, wurden die Sequenzen entweder gegen eine annotierte Referenzdatenbank kartiert oder mittels des BLAST Algorithmus abgeglichen. Hierfür kamen sowohl kommerzielle, als auch frei erhältliche Programme und Skripte zum Einsatz. Auch im eigenen Haus durch Diplom Bioinformatiker Benjamin Rieger geschriebene *perl* Skripte fanden Anwendung. Da die Rechenkapazität handelsüblicher PCs für einige der Analysen nicht ausreichend ist, wurden vor allem die BLAST Suchen auf dem Rechencluster „lc2“ des Rechenzentrums der Universität Mainz durchgeführt. Es handelt sich dabei um ein mit Linux Suse Enterprise Server 10 betriebenes Rechencluster, welches als Batchsystem über LSF7 verfügt. Das Cluster basiert auf AMD Opteron 2384 (2,7 GHz) Prozessoren mit 132 Knoten á 8 Kernen. Als Arbeitsspeicher verfügen 102 Knoten über 16 GB, die übrigen Knoten über 32 GB Kapazität. Eine schematische Übersicht der Datenverarbeitung gibt Abbildung 2 - 10 wieder. Auf Ebene der kartierten Sequenzen bzw. Annotierten Sequenzen / Contigs fand eine Normalisierung der Daten statt, die zu nach Expressionsstärke sortierbaren Genlisten führt (normalisierter „Read Count“ Ansatz. Dies ermöglicht den direkten Vergleich der Expressionsstärken von z.B. Markergenen (siehe Punkt 3.7.1). Die Auswertung auf höherer biologischer Ebene, wie Pathway Analysen oder Genontologie Zuordnungen, fand in nachgeordneten Schritten statt. Die Details der einzelnen Schritte werden im jeweiligen Textabschnitt (siehe nachfolgend) erläutert.

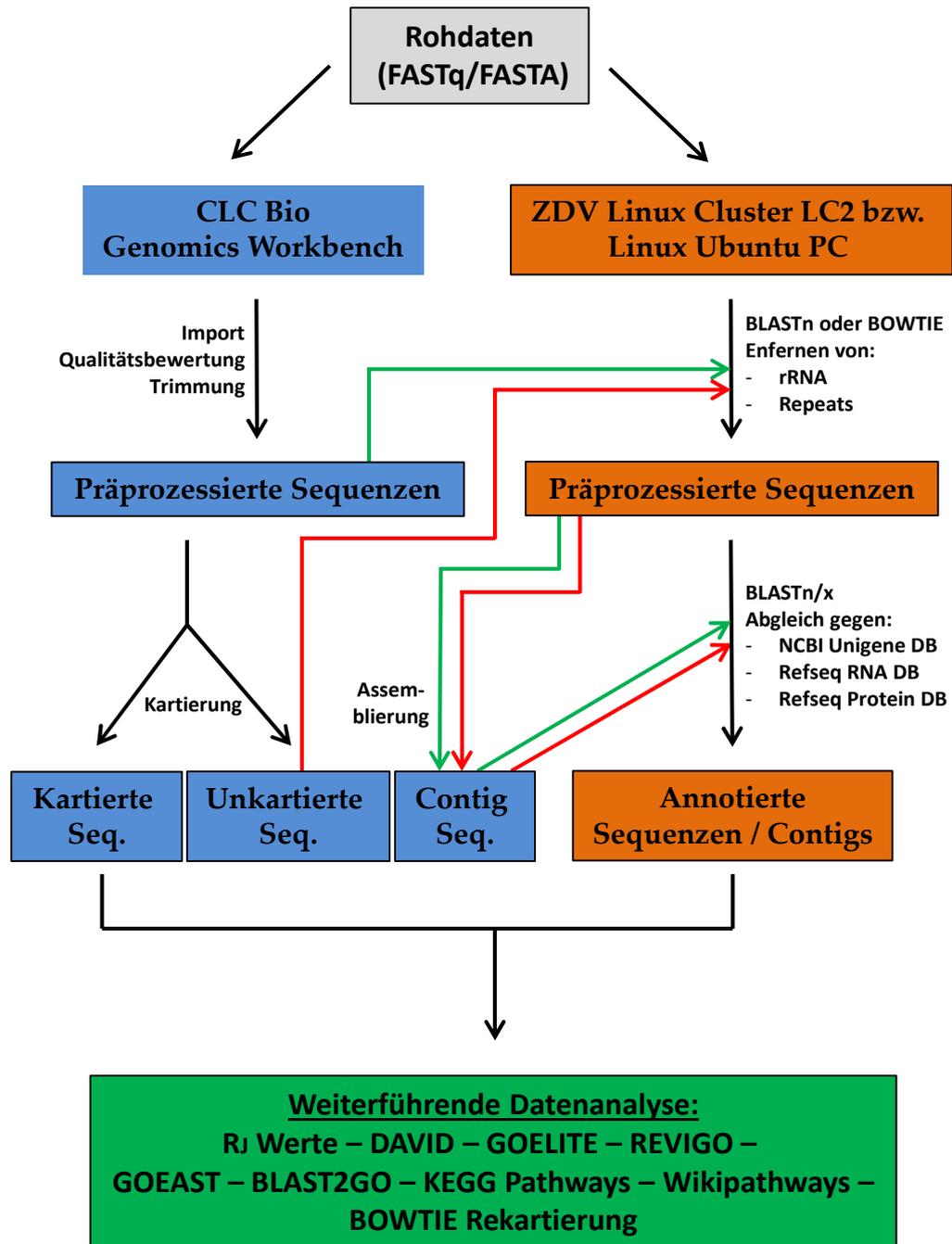


Abbildung 2 - 1 Schematische Darstellung der Datenverarbeitung

Die Abbildung zeigt vereinfacht und nach Plattformen geordnet ein Schema der Datenverarbeitung. In Blau sind mit Hilfe des CLC Workbench Programmes prozessierte Daten angezeigt, in Orange sind durch Anwendung von „open source“ Programmen unter Linux Umgebung prozessierte Arbeitsschritte markiert. Für einige Analysen wurden zwischen den Plattformen Daten ausgetauscht. Dies ist für die Assemblierung der getrimmten, qualitätsbewerteten Rohdaten mit grünen, für die Assemblierung der in der Kartierung nicht zuordenbaren Sequenzen mit roten Pfeilen dargestellt.

Tabelle 2 - 12 listet alle zur Auswertung genutzten Programme und eine kurze Beschreibung derer Funktionen auf. Im Rahmen der NGS Analyse wurden

bisweilen in der Programmiersprache „Perl“ geschriebene Skripte verwendet. Diese gehören zur hauseigenen Skriptesammlung und wurden vom Diplom Bioinformatiker Benjamin Rieger programmiert und zur Verfügung gestellt.

2.6.1 KARTIERUNG VON SEQUENZEN GEGEN ANNOTIERTE BEREICHE DES RINDERGENOMS MIT HILFE DER *CLC GENOMICS WORKBENCH*

Das Programmpaket CLC Genomics Workbench 4.7.2 der Fa. CLC Bio (Aarhus, Dänemark) stellt als kommerzielle Software breite Möglichkeiten zur Analyse von NGS Daten bereit. Die Rohdaten wurden beim Importieren in das Programm getrimmt und nach Qualität bewertet. Im Rahmen der Erfahrung mit NGS Daten die in der Arbeitsgruppe projektübergreifend gesammelt wurden [82], wurden am 5' Ende aller Sequenzen 5 Basen (3 Basen Adipozyten) und am 3' Ende aller Sequenzen 20 Basen (5 Basen Adipozyten) entfernt. An den genannten Basenpositionen wurde in vorangegangenen Arbeiten kritische Qualitätswerte (Phred Werte) bestimmt, die sich auf nachfolgende Analysen negativ auswirken können. Zudem wurden beim Import von Rohdaten im .qseq Format automatisch jene Sequenzen mit einem schlechten Signal Purity Filter Wert (PF – Wert) entfernt. Eine Sequenz erhält einen schlechten PF - Wert, wenn in den ersten 25 Basen mehr als einmal ein „Chastity Wert“ von 0,6 bestimmt wurde (Entfällt für Daten aus der Sequenzierung der Adipozyten, da Rohdaten nur im FASTq Format vorlagen). Der Chastity Wert berechnet sich als Verhältnis der hellsten Intensität einer sequenzierten Base zur Summe der hellsten und zweithellsten Intensität [83]. Der Wert dient der Abschätzung des Signal-zu-Hintergrund Rauschens. Beim Trimmen werden zusätzlich die Bereiche einer Sequenz abgeschnitten, in denen die Irrtumswahrscheinlichkeit für das korrekte Zuweisen (engl. *base calling*) einzelner Basen bei unter 99 % liegt (Phred 20). Sequenzen mit mehr als einer als „N“ (=Nukleotide, unklare Base an dieser Position) bestimmten Base wurden ebenfalls verworfen. Die restliche Sequenz von guter Qualität durften zum Kartieren 15 bp Länge nicht unterschreiten. Die so von Sequenzen schlechter Qualität bereinigten Datensätze bildeten den Ausgangsdatsatz für die Assemblierung zu Contigs und für die Kartierung gegen das Rindergenom. Die Kartierung erfolgte gegen das NCBI Referenzgenom des Rindes in der Version 4.2. Es wurden ausschließlich annotierte Bereiche des Genoms kartiert, erweitert um 2000 bp in 5' und 3' Richtung. Als Grenze für den Abstand der sog. „paired end“ Sequenzen (alle Sequenzierungen außer den Adipozyten) wurden 100 – 500 bp (min / max) angegeben. Für jede

kartierte Sequenz durften nicht mehr als 10 mögliche Treffer im gesamten Genom ermittelt werden, um nicht eindeutig zuordenbare Sequenzen keinen Einfluss auf die Expressionsanalyse nehmen zu lassen. Die Parameter „minimum length fraction“ und „minimum similarity fraction“ wurden auf 0,5 respektive 0,96 gesetzt. Als Maß für die Expressionsstärke wurden die „total exon reads“ (alle in Exons eines Gens kartierte Sequenzen - TER) festgelegt. Eine Normalisierung der Expressionsstärke auf RPKM-Werte [68] wurde, ausgehend von den TER, gemäß nachfolgender Formel berechnet.

$$\text{RPKM} = \frac{\text{Reads}_{\text{kartiert}} \times 10^6}{\text{Transkript (kb)} \times \text{Reads}_{\text{tot}}}$$

Abbildung 2 - 2 Normalisierung durch Berechnen des RPKM Wertes

Der Wert gibt eine Normalisierung der Stärke der Genexpression auf eine Millionen Sequenzen und die Länge eines Transkriptes an („Reads pro Kilobase Transkript pro Millionen kartierter Sequenzen“). Für die Berechnung der RPKM Werte wird die Transkriptlänge der längsten bekannten Splicingvariante eines Gens angenommen.

Die Ergebnisdateien wurden in ein Microsoft Excel kompatibles Format exportiert und sind im digitalen Anhang D4.1 aufgeführt.

2.6.2 BLASTn ABGLEICH DER ILLUMINA SEQUENZEN MIT DER NCBI UNIGENE DATENBANK DES RINDES

Für den BLASTn Abgleich der Rohdaten mit der NCBI Unigene Datenbank wurden Sequenzen im .fasta Format verwendet. Diese wurden aus .fastq Dateien mittels des Perl Skripts „fastq2fasta“, aus .qseq Dateien mittels des Perl Skripts „qseq2fasta“ erzeugt. Die NCBI Unigene Datenbank besteht, getrennt nach Spezies, aus vielen transkribierten Einzelsequenzen (cDNA), welche aufgrund Ihrer Ähnlichkeit untereinander zu Clustern gruppiert gemeinsam einen Unigene Eintrag repräsentieren. Da die von BLAST Suchen benötigten rechnerische Ressourcen über die Kapazität handelsüblicher PCs hinausgehen, wurden sämtliche Analysen auf dem LC2 Rechencluster des Zentrums für Datenverarbeitung der Universität Mainz durchgeführt. Sämtliche BLASTn (und BLASTx) Suchen wurden mittels des Perl Skriptes „Blaster1.2“ hochparallelisiert durchgeführt. Die großen Sequenzmengen machten ein Aufspalten der Daten auf bis zu 1000 Unterpakete und das Verteilen der einzelnen BLAST Analysen auf mehrere Knoten des Rechenclusters nötig

(Batchsystem LSF7). Die Ergebnisse konnten nach Abschluss der Berechnungen wieder zusammengeführt und mittels des Perl Skriptes „Bort2.4“ nach frei festzulegenden Filterkriterien weiter aufbereitet werden. Zur Quantifizierung der Häufigkeit der Zuordnung von Illumina Sequenzen zu einer Sequenz eines Clusters der NCBI Unigene Datenbank wurden die zugehörigen BLASTn Treffer aufaddiert und die Ergebnisse MS Excel kompatibel gespeichert. Um Sequenzen schlechter Qualität sowie nicht genkodierende Sequenzen vor dem Abgleich mit der NCBI Unigene Datenbank zu entfernen, wurden zunächst mittels des Perl Skriptes „fasta_n_filter“ alle Sequenzen mit mehr als 5 als „N“ bestimmten Basen aussortiert. Anschließend wurden sämtliche verbleibenden Sequenzen zunächst mittels des BLASTn Algorithmus gegen die komplette „Repeatmasker“ Datenbank des „Institute for Systems Biology“ (Seattle, Washington USA) abgeglichen um repetitive Sequenzen von der weiteren Analyse ausschließen zu können. Als Grenzwert für die eindeutige Zuordnung zu einer Datenbanksequenz wurde für diese sowie die nachfolgend beschriebenen BLASTn Suchen ein e-value von $1e^{-10}$ festgelegt. Um die Dauer der BLASTn Suchen zu verkürzen und die Größe der Ergebnisdateien zu beschränken, wurde für jede Illumina Sequenz nur der signifikanteste Datenbanktreffer mit einem einzigen Alignment zur zugehörigen Datenbanksequenz abgespeichert. Analog zum Entfernen von Repeats wurden auch für rRNA kodierende Sequenzen durch BLASTn Abgleich mit der „Silva comprehensive ribosomal RNA database“ des MPI Bremen annotiert und von der weiteren Analyse ausgeschlossen. Mit Hilfe der auf den NCBI Datenservern abgelegten Liste „gene2unigene“ wurde eine Zuordnung der Unigene Cluster zu Gennamen in MS Excel möglich. Als Maß für die Expressionsstärke wurden alle Illumina Sequenzen mit BLAST Treffern innerhalb eines Unigene Clusters addiert und zur besseren Vergleichbarkeit zunächst auf RPM Werte (engl. „reads per million“) normalisiert. Für jedes Unigene Cluster ist die Sequenzlänge aller zugehörigen Sequenzen bekannt, die Länge der längsten Sequenz eines Clusters wurde zur Bestimmung des RPKM Wertes für das gesamte Cluster/Gen herangezogen. Die Ergebnisse der BLAST Analysen sind im digitalen Anhang D4.2 einsehbar.

2.6.3 ASSEMBLIERUNG VON ILLUMINA SEQUENZEN UND ANNOTATION DER CONTIGS MITTELS BLASTn ABGLEICH MIT DER REFSEQ RNA DATENBANK

Zur Assemblierung wurden die Daten verwendet, die zuvor vor der Kartierung gegen das Referenzgenom des Rindes getrimmt und von Sequenzen schlechter Qualität bereinigt wurden. Die Assemblierung fand mittels der CLC Genomics Workbench 4.9 statt. Um keine für rRNA oder repetitive Elemente kodierenden Contigs zusammensetzen, wurden vor Beginn der Assemblierung alle Sequenzen durch Kartierung mittels des Programms „Bowtie 1 Version 0.12.7“ [84] mit rRNA und repetitive Elemente beinhaltenden Datenbanken abgeglichen (siehe Punkt 2.6.2) und die Treffer von der weiteren Analyse ausgeschlossen. Die Parameter der Kartierung mittels Bowtie 1 wurden so gesetzt, daß als Treffer nur eine absolute Übereinstimmung von Query und Datenbankeintrag gewertet wurde. Die Anzahl der repetitiven Sequenzen gegen die zur Aufreinigung vor der Assemblierung kartiert werden konnte, wurde im Vergleich zur Aufreinigung der Daten vor der BLASTn Analyse der Einzelsequenzen gegen die Unigene Datenbank noch einmal erweitert. Bereiche, die im Genom des Rindes in den NCBI Genbank Dateien als repetitiv annotiert vorlagen, wurden zusätzlich zu der in der vorangegangenen Analyse verwendeten, repeathaltigen Datenbank zum Abgleich herangezogen. Die Einstellungen zur Assemblierung in der CLC Genomics Workbench wurden so gesetzt, daß nur Contigs mit mindestens 400 bp Länge zugelassen wurden. Die Software benutzt einen De Bruijn Graph zur Assemblierung [65], die k-mer Größen wurden hierbei automatisch festgelegt und betragen für die Sequenzen der jungen MSC 24 bp, für alle übrigen Sequenzierungen 23 bp Länge. Die Sequenzen aus der Adipozyten Zellkultur konnten aufgrund der geringen Länge von maximal 43 bp nicht mit identischen Einstellungen wie die übrigen Sequenzierungen assembliert werden (max. 76 bp junge MSC). Tabelle 2 – 12 gibt die Einstellungen der Assemblierung wieder. Die aus den unterschiedlichen Datensätzen assemblierten Contigs wurden zur Annotation mit der Spezies übergreifenden NCBI Refseq RNA Datenbank durch den BLASTn Algorithmus auf dem LC2 Cluster abgeglichen (siehe Punkt 2.6.2). Für die Refseq RNA Datenbank ist über die FTP Server des NCBI eine Datei, die eine Zuordnung von Gennamen zu den Refseq - Kodes ermöglicht, vorhanden.

Tabelle 2 - 11 Parameter der Assemblierung - CLC Genomics Workbench

Die Tabelle gibt die zur Assemblierung der Sequenzen gesetzten Parameter wieder. Aufgrund des Längenunterschiedes in den Datensätzen konnten zur Assemblierung der Adipozyten Sequenzen nicht dieselben Parameter wie zur Assemblierung der Chondrozyten, Osteoblasten sowie jungen und alten MSC genutzt werden.

Parameter	Adipo (max 43 bp)	Übrige Sequenzierungen (max 70 / 76 bp)
Add conflict annotations	No	No
Conflict resolution	Vote	Vote
Map reads to contigs	Yes	Yes
Update contigs	Yes	Yes
Minimum contig length	400	400
Automatic word size	Yes	Yes
Match mode	random	random
Limit	8	--NA--
Mismatch cost	3	3
Alignment	gapped	--NA--
Insertion cost	3	3
Deletion cost	3	3
Global alignment	No	No
Color space alignment	No	No
Similarity	--NA--	0,98
Length fraction	--NA--	0,98

Nach Abschluss der BLASTn Analyse wurden zunächst alle Contigs, denen Refseq RNA Datenbankeinträge kodierend für dasselbe Gen zugeordnet wurden, bestimmt. Zur Quantifizierung der Expressionsstärke eines Gens diente die Summe aller Sequenzen der Contigs mit demselben Gen als BLASTn Treffer. Um Expressionsstärken besser vergleichbar zu machen, wurden die auf RPM Werte (engl. *reads per million*), also RNA Sequenzen pro eine Millionen sequenzierter Sequenzen normalisiert. Die Contig Sequenzen und weitere Informationen sind im Digitalen Anhang D4.3 verfügbar.

2.6.4 BESTIMMUNG DIFFERENZIELLER GENEXPRESSION

Zur Bestimmung differenziell exprimierter Gene wurden als Maß für die Genexpression die in der Kartierung gegen das Rindergenom bestimmten „Total Exon Read“ Werte zugrunde gelegt (siehe Punkt 2.6.1). Als statistisch abgesicherte Einheit für die Stärke differenzieller Genexpression wurde ausgehend von den TER Werten der sog. R_j – Wert berechnet [85]. Dieser berücksichtigt sowohl die unterschiedliche Anzahl an Gesamtsequenzen in einer Illumina Sequenzierung als

auch die Anzahl der verglichenen Datensätze. Im Rahmen dieser Arbeit wurden stets die TER Werte einer differenzierten Zelllinie mit den TER Werten der Sequenzierung der jungen MSC abgeglichen. Für die jungen MSC wurden die Gene als differenziell exprimiert angenommen, die im Vergleich sowohl zu den Adipozyten und Osteoblasten als hochreguliert bestimmt wurden. Die Formel zur Berechnung des R_j – Wertes als Maß der Signifikanz des Expressionsunterschiedes ist nachfolgend angegeben. Die Berechnung basiert auf einer logarithmierten Likelihood Ratio Statistik R für ein Gen j , welche sich asymptotisch an eine χ^2 -Verteilung annähert.

$$R_j = \sum_{i=1}^m x_{i,j} \log \left(\frac{x_{i,j}}{N_i f_j} \right) \quad f_j = \frac{\sum_{i=1}^m x_{i,j}}{\sum_{i=1}^m N_i}$$

Abbildung 2 - 3 Formeln zur Berechnung des R_j – Wertes

Die Variable m gibt die Anzahl der Datensätze an, $x_{i,j}$ steht für die Anzahl an Sequenzen des Betrachteten Gens j im Datensatz i . N_i repräsentiert die Anzahl aller Sequenzen in Datensatz i und f_j (siehe rechte Formel) stellt ein Maß für die Häufigkeit eines Gens j in sämtlichen Datensätzen dar.

Ab einem R_j – Wert von 8 wurde für ein Gen eine differenzielle Genexpression angenommen. Hierbei gilt, je Größer der R_j – Wert desto stärker der Expressionsunterschied. Da besonders beim Vergleich schwach exprimierter Gene große Signifikanzen im Expressionsunterschied errechnet werden, ohne daß dies mit hoher Wahrscheinlichkeit einen Rückschluss auf eine besondere Funktion der schwach exprimierten Gene zulässt, wurde als Grenze für die Berechnung differenzieller Genexpression ein RPKM Wert von 10 für mindestens eines der verglichenen Gene zugrunde gelegt. Um die Menge als differenziell exprimiert bestimmter Gene zusätzlich auf die signifikantesten Treffer beschränken zu können, wurde zudem ein RPKM basiertes Expressionsverhältnis von mindestens dreifachem Expressionsunterschied voraus gesetzt ($|\log_2(\text{RPKM1} / \text{RPKM2})| \geq 1,5849625$). Die Berechnungen der differenziellen Genexpression sind im digitalen Anhang D5.1 einsehbar.

2.7 EINORDNUNG DER GENEXPRESSIONSPROFILE IN EINEN ÜBERGEORDNETEN BIOLOGISCHEN KONTEXT

Die unter Punkt 2.6 beschriebenen Analysen der Sequenzrohdaten zielen auf die Annotation und anschließende Quantifizierung der Expressionsstärke der

identifizierten Gene, bzw. der Feststellung der Größe des Expressionsunterschiedes eines Gens in unterschiedlichen Datensätzen ab. Um einen biologischen Sinn aus den „Read Count“ basierten Listen ziehen zu können, wurde die Ebene der Betrachtung einzelner Gene verlassen und die Datensätze in einen funktionellen Kontext gesetzt. Hierfür wurden unterschiedliche lokal installierte oder über das Internet frei verfügbare Programme genutzt (siehe Tabelle 2 - 12). Bei der Betrachtung der Platzierung eines Gens in nach Expressionsstärke sortierten Genlisten, ließen sich für die unter Punkt 2.6.1 bis 2.6.3 beschriebenen Methoden im Vergleich keine kritischen Unterschiede erkennen (Ergebnisse siehe Punkt 3.6). Daher wurden für die nachfolgend beschriebenen Auswertungen nur noch die durch Kartierung mittels „CLC Genomics Workbench“ (siehe Punkt 2.6.1) ermittelten Expressionswerte berücksichtigt.

2.7.1 GLOBALER VERGLEICH DER GENEXPRESSIONSPROFILE

Ein Vergleich der Genexpressionsprofile aller Sequenzierungen nach Kartierung untereinander anhand der Expressionsstärken jedes Gens in Form einer sog. „Heatmap“ erfolgte mittels des Programms CLC Genomics Workbench. Als vergleichbares Maß der Genexpression diente der RPKM Wert. Die Expression aller Gene der jungen MSC wird hierbei als Ausgangswerte genutzt zu der die Expression aller Gene in den übrigen Sequenzierungen verglichen wird. Starke Expression (hoher RPKM) eines Gens wird in der Heatmap in Rot, schwächere Expression in Blau dargestellt. Aus den Daten lässt sich ein Stammbaum ableiten, der die Ähnlichkeit der Genexpression in globalem Maßstab nach Art einer Verwandtschaftsbeziehung darstellt. Auch durch Streudiagramme lässt sich die Ähnlichkeit zweier Zelllinien im direkten Vergleich darstellen. Hierfür wurden in Microsoft Excel die zur Basis 10 logarithmierten RPKM Werte aller Gene zweier Zelllinien gegeneinander aufgetragen. Zur besseren Einschätzung der Ergebnisse wurden diagonale Begrenzungslinien die einen Expressionsunterschied um den Faktor ± 3 anzeigen in die Diagramme eingefügt.

2.7.2 ANALYSE DER GENEXPRESSION – FAC DER 100 STÄRKSTEXPRIMIERTEN GENE MITTELS DAVID

Eine erste Charakterisierung der Expressionsprofile unter Verwendung der Read-Count Listen erfolgt anhand der 100 am stärksten exprimierten Gene jeder Zelllinie mittels des Online frei verfügbaren Werkzeugs DAVID [86, 87]. Das Programm ermöglicht eine auf Genontologien [88] basierende funktionelle Gruppierung von umfangreichen Genlisten und berechnet die statistische Signifikanz der Ergebnisse. Als eindeutige Gennummern wurden die sog. „NCBI Entrez Gene ID“ in DAVID eingeladen. DAVID ermöglicht eine große Auswahl an veränderlichen Parametern zu treffen, die eine solche Gruppierung (das sog. „Functional Annotation Clustering - FAC“) für die jeweilige Fragestellung einfach anpassbar macht. Generell sind Genontologien in die Sparten „biological process“, „cellular component“ und „molecular function“ eingeteilt. Ein einzelnes Gen kann in allen drei genannten Gruppen in einem oder auch mehreren sog. Genontologie Oberbegriffen (GO-Terms) annotiert sein. Die Struktur der GO-Terms ist in einem in 5 Ebenen unterteilten Stammbaum organisiert. Die erste Hierarchieebene der mit DAVID abgleichbaren GO-Terms enthält eine Vielzahl Gene die sehr generelle Funktionen abdecken (z.B. „multicellular organismal process“) während die fünfte Ebene sehr spezielle Funktionen beschreibt und entsprechend weniger Gene beinhaltet (z.B. „regulation of actin polymerization or depolymerization“). Funktionell ähnliche GO-Terms haben mit hoher Wahrscheinlichkeit auch eine Überschneidung in einem Teil der darin annotierten Gene. Dies bedingte die Notwendigkeit, die Anzahl und Art der mittels DAVID abgeglichenen Ontologien so zu reduzieren, daß eine überschaubare und gleichzeitig die Funktion der abgeglichenen Gene treffende Analyse der Genlisten möglich wurde. Durch das DAVID FAC werden einander ähnliche GO-Terms in größeren Gruppen zusammengefasst. Im Rahmen der Charakterisierung der 100 am stärksten exprimierten Gene jeder Zelllinie wurden aus den Ontologiegruppen „biological process“ und „molecular function“ die Hierarchieebenen 2 – 5 zum Abgleich heran gezogen. Zusätzlich wurden die Einträge in der „swissprot protein interaction resource keywords“ Datenbank ebenfalls verwendet. Zur statistischen Absicherung wurde für jeden getroffenen GO-Term ein sog. „EASE Score“ (modifizierter Fisher Exakt Test p-Value) automatisch errechnet. Je kleiner der „EASE Score“, desto stärker signifikant ist ein einzelner GO-Term abgedeckt. Um den im FAC gebildeten GO-Term Gruppen als Ganzes ebenfalls ein Signifikanzmaß zuweisen zu können, werden die $-\log_{10}$ Mediane aller „EASE Scores“ der GO-Terms einer Gruppe

gebildet. Der resultierende sog. „Enrichment Score“ gibt mit steigender Größe eine stärker signifikante Anreicherung der GO-Term Gruppe an. Der Abgleich gegen die oben genannten Datenbanken erfolgte mit den DAVID Standardeinstellungen, nur der „*Enrichment Threshold*“ wurde von 1 auf 0,5 „EASE Score“ Einheiten gesenkt. Die Senkung des Signifikanzniveaus erlaubt bei 100 Genen in der Suchanfrage eine größere, aber biologisch aussagekräftigere Auflösung der GO-Term Ergebnisse. Nach Abschluss der Analyse wurde jeder im FAC gebildeten Gruppe Genontologien manuell ein funktionsbeschreibender Oberbegriff zugewiesen. Die Ergebnisse wurden grafisch als Kreisdiagramm in MS Excel dargestellt. Hierbei wurden für jede Zelllinie die 12 am stärksten signifikant abgedeckten GO-Term Gruppen berücksichtigt. Die Rohdaten der Analysen sind im digitalen Anhang D5.2 einsehbar.

2.7.3 ANALYSE DIFFERENZIELL EXPRIMIERTER GENE – ABGLEICH MIT BIOLOGISCHEN PATHWAYS UND ZUORDNEN DER GENONTOLOGIE

Um aus den Listen differenziell exprimierter Gene Rückschlüsse auf tieferschichtige biologischen Zusammenhänge schließen zu können, wurden zunächst sämtliche Gene im Format von *NCBI Entrez Gene ID* Nummern mit biologischen Pathways abgeglichen. Zur Untersuchung der in jungen MSC differenziell hoch exprimierten Gene wurden alle Gene analysiert, die sowohl im Vergleich zu den Adipozyten als auch im Vergleich zu Osteoblasten als differenziell exprimiert bestimmt wurden. Als gut annotierte Pathways wurden die Daten des KEGG (*Kyoto Encyclopedia of Genes and Genomes*) [89] sowie die im Wikipathways Projekt [90] organisierten Pathways zugrunde gelegt. Um eine grafische Darstellung der Treffer im Pathway sowie eine statistische Absicherung der Ergebnisse zu erhalten, wurde zum Abgleich der Listen gegen die KEGG-Pathways erneut das Programmpaket DAVID in Kombination mit dem KEGG eigenen „*KEGG Mapper - Search and color Pathway*“ Werkzeug verwendet. Gegen die Wikipathways wurden mittels des frei erhältlichen Programms GenMAPP-CS [91] kartiert, die statistische Absicherung der Ergebnisse fand in letzterem Fall über das ebenfalls kostenlose GenMAPP-CS Plugin GO-Elite [92] statt. Die 100 am stärksten differenziell exprimierten Gene (250 Gene im Falle der jungen MSC) wurden mittels des online Tools GOEAST [93] einer Go-Term Annotation unterzogen. Um die Ergebnisse in der Menge einzugrenzen und grafisch darstellen zu können, wurden die resultierenden Go-Term Listen (nur Kategorie „*biological process*“) mittels des online Tools REVIGO (**R**educe + **V**isualize

Gene Ontologies) [94] in sog. Treemaps umgerechnet. Hierbei wird erneut der Umstand der Gruppierbarkeit von GO – Terms genutzt (siehe Punkt 2.7.2). Ähnliche Ontologien werden bestimmt und farblich identisch gruppiert. Die Größe der Quadrate die jeder in der Treemap aufgezeigte Go – Term annimmt, ist direkt korreliert mit der Signifikanz seiner statistischen Anreicherung (\log_{10} p-Value). Da der Export der Treemaps wie sie von REVIGO zur Verfügung gestellt werden ausschließlich in Rohdatenform möglich ist, wurden diese mit einer Testversion des Programms Treemap der Fa. Macrofocus (Zürich, Schweiz) graphisch aufbereitet. Die vollständigen Analyseergebnisse sind im digitalen Anhang D5.3 einsehbar.

2.7.4 ASSEMBLIERUNG VON UNKARTIERTEN SEQUENZEN UND CHARAKTERISIERUNG NOCH NICHT ANNOTIERTER TRANSKRIPTTE

In diesem Schritt wurden alle unter Punkt 2.6.1 nicht kartierbaren Sequenzen zunächst analog zum unter Punkt 2.6.3 beschriebenen Vorgehen von repetitiven Elementen und rRNA bereinigt und die verbleibenden Sequenzen in der CLC Genomics Workbench assembliert. Die gebildeten Contigs wurden nach offenen Leserahmen von 100 Aminosäuren Länge beidseitig durchsucht und die Ergebnisse tabellarisch gespeichert. Sämtliche Contigs wurden anschließend auf dem LC2 Rechencluster des ZDV der Universität Mainz mittels des BLASTn/x Algorithmus gegen die NCBI RefSeq RNA- und die nicht redundante Proteindatenbank abgeglichen. Für die BLASTx Suchen wurde hierbei ein E-Value von $1e^{-006}$ als Maß für einen signifikanten Treffer angenommen, BLASTn Treffer mussten weiterhin einen E-Value von $1e^{-010}$ aufweisen. Zudem wurden für jeden Contig 10 Treffer und 10 Alignments in die im .xml Format gehaltene Ergebnisdatei der BLASTx Analyse übernommen. Auch die NCBI gi Nummer jedes Treffers wurde abgespeichert. Diese zusätzlichen Informationen dienen einer erleichterten weiterführenden Annotation der BLASTx Ergebnisse durch die Java Skript basierte Software „BLAST2GO“ [95]. Die BLASTx Ergebnisse sowie die Sequenzdaten der Contigs wurden in BLAST2GO importiert. Das Programm ermöglicht alle 6 Leserahmen jedes Contigs nach Ähnlichkeiten mit bekannten funktionellen Domänen durch eine automatisierte Interpro Suche [96, 97] abzugleichen. Dabei werden die Grundeinstellungen verwendet und alle in Interpro zusammengefassten Datenbanken zum Abgleich heran gezogen. Die Ergebnisse der Assemblierung, der BLASTn/x Analysen und der Interpro Suche wurden für jede Sequenzierung in einer Tabelle zusammengefasst, sämtliche Daten sind im digitalen Anhang D6 einsehbar. Um aus der großen Menge

der aus den 5 Zelllinien assemblierten Contigs potentiell interessante Transkripte zur manuellen Charakterisierung aussuchen zu können, wurden all jene Contigs heran gezogen die

- Über einen offenen Leserahmen von ≥ 100 AS verfügen
- Einen oder mehrere Interpro-Scan Treffer aufweisen
- Keinen BLASTn/x Treffer aufweisen
- Eine Contiglänge von ≥ 500 bp bei ≥ 2000 Sequenzen Abdeckung aufweisen

oder

- Über einen offenen Leserahmen von ≥ 100 AS verfügen
- Einen oder mehr Interpro-Scan Treffer aufweisen
- Sowohl in der BLASTn wie BLASTx Analyse einen nicht charakterisierten (z.B. „*hypothetical protein*“) Treffer aufweisen
- Eine Contigabdeckung von ≥ 1000 Sequenzen aufweisen

oder

- Über einen offenen Leserahmen ≥ 100 AS verfügen
- Drei oder mehr Interpro-Scan Treffer aufweisen
- Keinen BLASTn/x Treffer aufweisen
- Eine Contigabdeckung von ≥ 1000 Sequenzen aufweisen

Die genannten Auswahlkriterien zielen darauf ab, möglichst noch nicht charakterisierte Transkripte mit Ähnlichkeiten zu funktionellen Domänen gut charakterisierter Proteine zu selektieren. Zur weiteren Charakterisierung selektierter Contigs wurden zunächst noch einmal eine manuelle BLAST Suchen gegen das Rindergenom auf der Homepage des NCBI ausgeführt. Ein weiträumiges Stück genomischer DNA beidseitig der Treffer wurden anhand der genomischen Koordinaten aus den Datenbanken des NCBI im annotierten .gbk Format herunter geladen. Sämtliche Sequenzen (qualitätskontrolliert, getrimmt -siehe Punkt 2.6.1) der Zelllinie, aus der der Contig ursprünglich stammt, wurden daraufhin mittels Bowtie 1 erneut gegen das erweiterte Teilstück der genomischen DNA aus der manuellen BLASTn Suche kartiert. Die Sequenzlage dieser Kartierung gegen die annotierte genomische DNA kann nach Konvertierung des genomischen Abschnitts vom .gbk in das .gff3 Format mittels des Programms „Tablet“ [98] grafisch dargestellt werden. Sowohl die Lage des BLASTn Treffers, als auch die Ergebnisse der Kartierung aus Bowtie 1 wurden mit Hilfe des Programms „UniPro UGENE“

[99] in den jeweils bearbeiteten Abschnitt der genomischen DNA annotiert. Die Darstellung der Bowtie 1 Kartierung aus dem Programm Tablet wurde identisch zum annotierten genomischen Bereich aus dem Programm UGENE skaliert (sichtbare Basenpositionen bei identischer Bildschirmauflösung) und die Abbildungen manuell zusammen gelegt.

2.8 ÜBERSICHT ÜBER DIE BIOINFORMATISCHEN HILFSMITTEL

Tabelle 2 - 12 Übersicht über die bioinformatischen Hilfsmittel

Programm / Webseite	Hersteller	Typ	Beschreibung
„Oligonucleotides Properties Calculator“ @ http://www.basic.northwestern.edu/bio-tools/OligoCalc.html	Northwestern University	Online Tool	Hilfsmittel zum Berechnen von Primereigenschaften
„OligoAnalyzer“ @ http://eu.idtdna.com/calc/analyser	Integrated DNA Technologies	Online Tool	Hilfsmittel zum Berechnen von Primereigenschaften
CLC Genomics Workbench 4.7.2	CLC Bio	Kommerzielle NGS Analyse Software	Erlaubt Import, Kartierungen, Assemblierung und statistische Analysen von NGS Daten
Fastq2fasta / qseq2fasta	Hauseigene Skripte von Benjamin Rieger	Perl Skripte	Konvertierung von Sequenzdatei Formaten (fastq oder qseq) in das Fasta Format
Blaster1.2	Hauseigenes Skript von Benjamin Rieger	Perl Skript	Dient der Initialisierung hochparallelisierter BLASTn/x Suchen
Bort2,4	Hauseigenes Skript von Benjamin Rieger	Perl Skript	„Blast output refinement tool“ – dient dem Aufbereiten von Blastergebnissen aus Blaster1.2 nach vom Nutzer definierbaren Kriterien
Fasta_n_filter	Hauseigenes Skript von Benjamin Rieger	Perl Skript	Dient dem Entfernen von Sequenzen mit einer definierbaren Anzahl als „N“ bestimmter Basen aus einer Fasta Datei

Bowtie 1 Version 0.12.7	John Hopkins University	Open Source Programm	Programm zur Kartierung von NGS Daten gegen beliebige Referenzen
"DAVID" @ https://david.ncifcrf.gov/content.jsp?file=about_us.html	National Cancer Institute at Frederick	Online Tool	Bioinformatik, statistische Analysen großer Genlisten
"KEGG Mapper" @ http://www.genome.jp/kegg/tool/map_pathway2.html	Kyoto Encyclopedia of Genes and Genomes (Kanehisa Laboratories)	Online Tool	Abbilden von Genlisten auf biologischen Pathways
GeneMAPP-CS und GoElite als Cytoscape Plugins	Gladstone Institutes	Open Source Programme	Biostatistik und Abbilden von Genlisten auf biologischen Pathways
"GOEAST" @ http://omicslab.genetics.ac.cn/GOEAST/php/batch_genes.php	Institute of Genetics and Developmental Biology, Chinese Academy of Sciences	Online Tool	Errechnen von Genontologie Anreicherungsstatistiken anhand großer Genlisten
REVIGO	Rudjer Boskovic Institute, Croatia	Online Tool	Grafische Darstellung von Genontologien
Macrofocus Treemap	Macrofocus	Kommerzielles Programm, kostenlose Testversion	Generierung von Treemaps
BLAST2Go	BioBam Bioinformatics S.L.	Kostenfreies Programmpaket	Annotation von BLASTn/x Ergebnissen, Proteindomänen Suche
UGENE	Unipro Ugene	Kostenfreies Programmpaket	Anntoation von genomischen Contigs, graphische Darstellung

2.9 PUFFER UND LÖSUNGEN

Aufgeführt werden die Puffer und Lösungen, die im Text des Material und Methoden teils in ihrer Zusammensetzung nicht ausführlich beschrieben werden.

10 x SSC

1,5 M Natriumchlorid; 0,15 M Natriumcitrat in VE-Wasser

10 x TBE

0,89 M TRIS; 0,02 M EDTA; 0,89 M Borsäure in VE-Wasser

Ampicillin-Stammlösung

10 mg/ml Ampicillin in HPLC-Wasser

DNA-Ladepuffer

30 % (w/v) Saccharose; 0,25 % (w/v) Bromphenolblau in VE-Wasser

Ethidiumbromid-Stammlösung

0,5 % (w/v) Ethidiumbromid in 1x TBE-Puffer

LB-Medium (1 X)

2 % (w/v) Trypton; 1 % (w/v) Hefeextrakt; 1 % (w/v) Natriumchlorid in VE-Wasser

Plasmidisolierung im 96er Format **Lösung I** (1000 ml)

10 mM Na-EDTA, pH 8,0, nach Autoklavieren 100 mg RNase A zugeben

Plasmidisolierung im 96er Format **Lösung II** (150 ml)

1,2 g NaOH in 135 ml VE H₂O lösen, Zugabe von 15 ml 20 % SDS Lösung

Plasmidisolierung im 96er Format **Lösung III** (1000 ml)

250 ml 7,5 M Kalium – Acetat mit 115 ml Eisessig versetzen und 635 ml VE H₂O zugeben

3 ERGEBNISSE

3.1 ISOLIERUNG UND *IN VITRO* KULTUR MESENCHYMALER STAMMZELLEN

Bovine Mesenchymale Stammzellen (MSC) wurden aus dem Knochenmark frisch geschlachteter Rinder wie unter Punkt 2.1 beschrieben gewonnen und in Kultur genommen. Die plastikadhärent auf dem Boden der Kulturschalen anwachsenden MSC wiesen einen fibroblastenähnlichen Phänotyp auf.

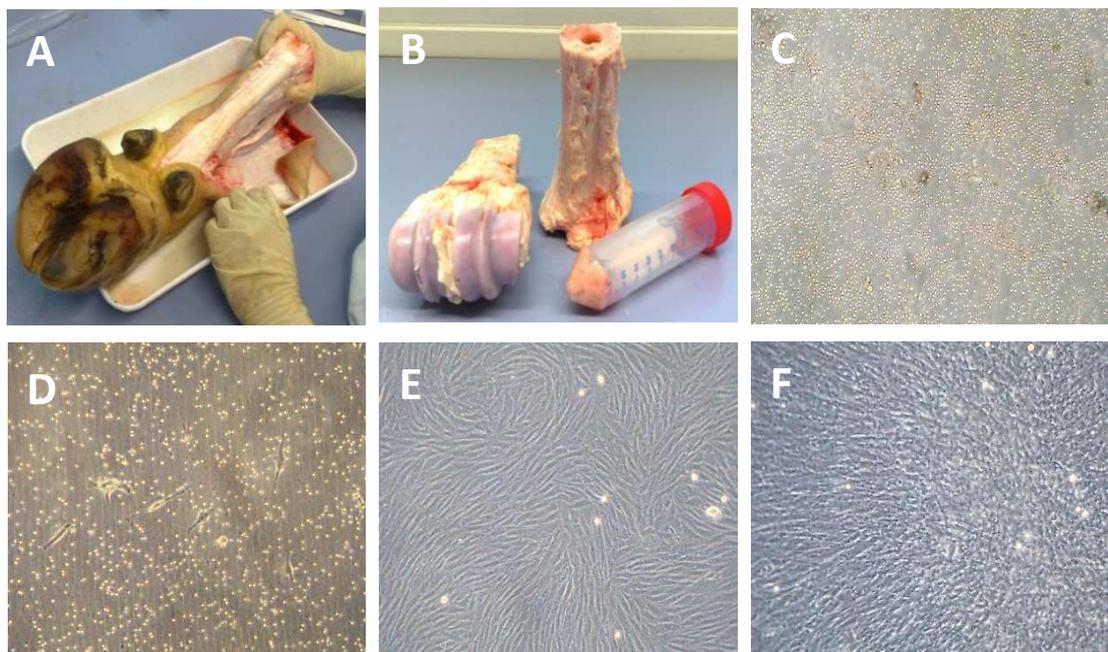


Abbildung 3 - 1 Isolation und Primärkultur von MSC aus bovinem Knochenmark

Schienbeine frisch geschlachteter Rinder werden gekühlt ins Labor transportiert und präpariert (A). Die Tibia wird unter sterilen Bedingungen geöffnet und das Knochenmark ausgeschabt (B). Aus dem Knochenmark gewonnene Zellen werden in Kultur genommen (C – 100X). Nach wenigen Tagen sind adhärent anwachsende, fibroblastenähnliche MSC zu erkennen (D – Tag 6, 100X). Diese bilden bei konfluentem Wachstum die für MSC charakteristische Wirbelstruktur aus (E – Tag 12, 100X). Lässt man die MSC im mehrschichtigen, gewebeähnlichem Verband als Kontrolle für Differenzierungen wachsen, ist diese Struktur nicht mehr erkennbar (F – Multilayer ein Monat nach Passage 1, 100X)

MSC können durch regelmäßige Medienwechsel von nicht adhären wachsenden Zellen anderen Typs aufgereinigt werden [70]. Der Überstand des ersten Medienwechsels nach Isolation enthält erfahrungsgemäß noch nicht angewachsene MSC und wurde noch ein weiteres Mal in Kultur genommen. Je nach Menge des eingesetzten Knochenmarks und in Abhängigkeit von der Grundfläche des Kulturgefäßes (25 oder 75 cm²) kann das Erreichen vollständiger Konfluenz unterschiedlich lange dauern. Das Wachstum der MSC in den Primärkulturen wurde lichtmikroskopisch überprüft und photographisch dokumentiert (siehe

Abbildung 3 - 1). Sobald die Zellen das erste Mal Konfluenz erreichten, wurden sie zur weiteren Proliferation passagiert. Zur nachfolgenden *in vitro* Differenzierung wurden möglichst junge MSC aus konfluenten Flaschen der ersten Passage eingesetzt.

3.2 CHARAKTERISIERUNG MESENCHYMALER STAMMZELLEN

3.2.1 NACHWEIS MSC TYPISCHER ANTIGENE MITTELS IMMUNOHISTOCHEMIE UND TEST AUF MYCOPLASMEN KONTAMINATION

Mesenchymale Stammzellen weisen ein typisches Oberflächenantigen Expressionsprofil auf. Die MSC Primärkulturen wurden kurze Zeit nach der Isolierung im Rahmen der ersten Passage auf die Expression der Antigene CD44 (Positivmarker) und CD45 (Negativmarker) getestet.

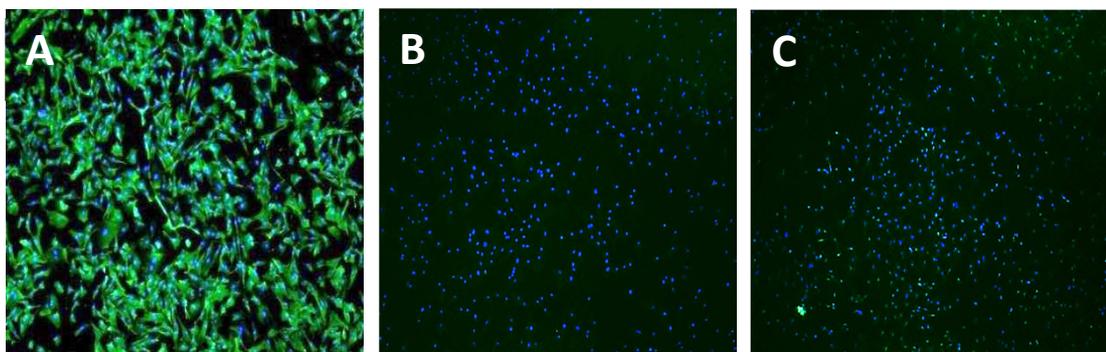


Abbildung 3 - 2 Antikörper – und DAPI Färbungen an undifferenzierten MSC

Mesenchymale Stammzellen zeigen angetarbt mit CD44 Antikörpern ein starkes Signal. Die Kultur erscheint homogen, jedes DAPI Signal aus dem Zellkern weist ein zugehöriges FITC Signal vom Positivmarker CD44 auf (A – 40X). Das FITC Signal fehlt bei Anfärbung mit dem Negativmarker CD45 (B – 40X). Inkubiert man die Zellen nur mit dem FITC gekoppelten, sekundären Antikörper erhält man

ebenfalls kein Signal (C – 40X). In keiner der DAPI Färbungen sind Kontaminationen mit Mycoplasmen zu erkennen.

Beide Antigene sind in der Literatur als speziesübergreifend geeignete Marker zur Charakterisierung von MSC beschrieben worden [71-73]. Zusätzlich zur Färbung mit FITC markierten Antikörpern wird eine DAPI Färbung durchgeführt. Aufgrund der Blaufärbung der Zellkerne durch DAPI kann der Anteil der Signale aus der Antikörperfärbung im Verhältnis zur Zellzahl abgeschätzt werden. Verunreinigungen der MSC Primärkultur mit ebenfalls im Knochenmark vorkommenden Zellen hämatopoietischen Typs (CD45+) können so sichtbar gemacht werden [100]. Die Inkubation mit CD44 Antikörpern lässt auf eine homogene MSC-Kultur ohne Verunreinigung durch andere Zellen schließen. Für jedes Signal aus der DAPI Färbung des Kerns liegt ein Signal des Positivmarkers CD44 vor. Entsprechend ist neben dem DAPI Signal kein Signal aus der Antikörperfärbung bei Inkubation mit CD45 Antikörpern festzustellen. Die DAPI Färbung dient des Weiteren als Test auf eine Mycoplasmen Kontamination in den Primärkulturen. Eine Kontamination mit Mycoplasmen konnte nicht festgestellt werden (Abbildung 3 - 2)

3.2.2 NACHWEIS MSC TYPISCHER TRANSKRIPTS DURCH RT-PCR

Zur weiteren Charakterisierung der Primärkulturen vor Beginn der Differenzierungsexperimente wurden für MSC typische Transkripte nachgewiesen. Aus der Literatur sind speziesübergreifend in undifferenzierten MSC exprimierte Markergene bekannt. Die Expression einiger dieser Marker wurde per RT-PCR vor Beginn der Differenzierungsexperimente überprüft. Als Matrizen-DNA für die RT-PCR diente einzelsträngige cDNA, die aus Gesamt-RNA undifferenzierter MSC verschiedener Isolationen synthetisiert wurde. Die Ergebnisse der PCR wurden auf Agarosegelen sichtbar gemacht (siehe Abbildung 3 - 3). In Tabelle 3 - 1 ist eine Übersicht der Ergebnisse aufgeführt. Wie erwartet ließ sich die Expression aller getesteten MSC-Marker in den undifferenzierten MSC nachweisen. Die Gene *sox2*, *lif* und *oct4* wurden jedoch nur äußerst schwach exprimiert. Die RT-PCR dieser Markergene konnte nicht bei jedem Versuch und nicht in jeder getesteten cDNA erfolgreich durchgeführt werden. Im Falle von *oct4* und *nanog* resultieren die gezeigten Banden im Gel aus einer Reamplifikation mit weiter innen im Amplifikat der ersten PCR bindenden Primern. Markergene, die nicht aus allen getesteten cDNA Synthesen amplifiziert werden konnten, bzw. für deren Amplifikation eine

nested PCR notwendig war, sind in Tabelle 3 - 1 mit einem +/- gekennzeichnet. Die Identität der Banden wurde durch Sequenzierung der PCR Produkte bestätigt.

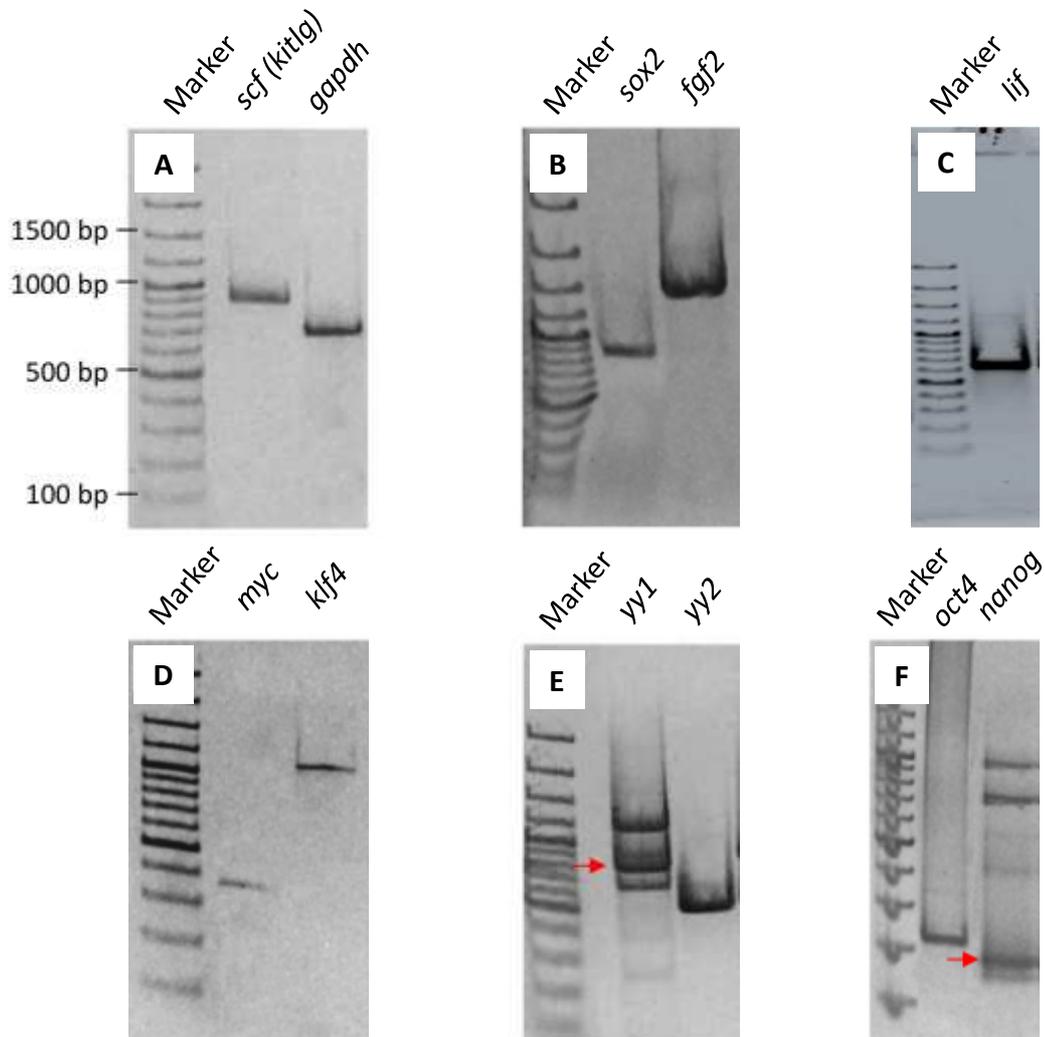


Abbildung 3 - 3 Agarosegele der RT-PCR zum Nachweis von MSC Markergenen

Die Ergebnisse der RT-PCR sind auf 1 % Agarosegelen sichtbar gemacht. Zur Größenabschätzung (siehe A) ist auf jedem Gel derselbe Marker aufgetragen. Für PCR Ansätze, die mehr als ein Produkt aufweisen, ist die Bande in Zielgröße extrahiert worden (siehe roter Pfeil in E und F). Die Identität aller Banden konnte per Sequenzierung bestätigt werden. Abbildung aus der Diplomarbeit von A. Dubberke (IMSB).

Die Expression der im Rahmen der Antikörperfärbung getesteten Oberflächenantigene (siehe 3.2.1) wurde ebenfalls per RT-PCR als Transkript nachgewiesen. Zusätzlich wurde mit CD105 ein weiteres für MSC typisches Antigen-Transkript amplifiziert. Für CD105 war kein rinderspezifischer Antikörper erhältlich, weshalb keine Färbungen durchgeführt werden konnten. Die Expression

der CD Antigene konnte wie erwartet per RT-PCR nachgewiesen werden (siehe Abbildung 3 - 4).

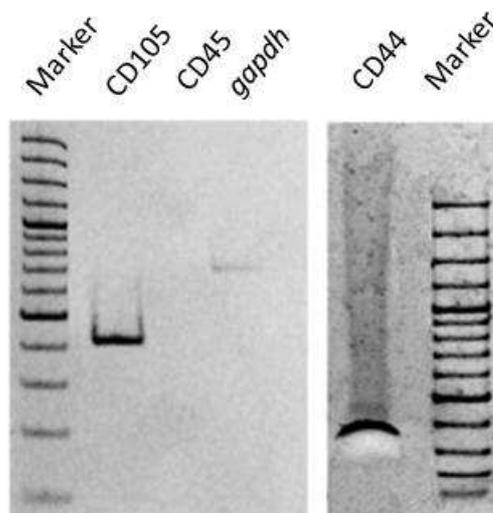


Abbildung 3 - 4 Agarosegele der RT-PCR zum Nachweis von CD Antigenen

Die Ergebnisse der RT-PCR sind auf 1 % Agarosegelen sichtbar gemacht. Die Expression der MSC Positivmarker CD105 und CD44 konnte wie erwartet bestätigt werden. Transkripte des hämatopoetischen Markers CD45 konnten nicht nachgewiesen werden. Abbildung aus der Diplomarbeit von A. Dubberke (IMSB).

Tabelle 3 - 1 Ergebnisse der RT-PCR auf Markergene undifferenzierter MSC

Die Tabelle zeigt eine Übersicht der Ergebnisse der RT-PCR. Für jedes Gen ist eine grobe Funktionsbeschreibung, sowie eine Quellenangabe, die die Relevanz des Gens als Stammzellmarker belegt, aufgeführt. Für Gene, die nicht in jeder getesteten cDNA undifferenzierter MSC nachweisbar waren, bzw. für deren Nachweis eine Reamplifikation vonnöten war, ist in der Ergebnisspalte ein +/- eingetragen.

Markergen	Funktion	Quelle	Ergebnis
<i>sox2</i>	Transkriptionsfaktor, Erhält SC - Status	[32, 101-103]	+/-
<i>fgf2</i>	Wachstumsfaktor, Erhält SC - Status	[32, 104, 105]	+
<i>lif</i>	Cytokin, Differenzierungskontrolle	[32, 106, 107]	+/-
<i>myc</i>	Onkogen, Aktivator	[103, 108, 109]	+
<i>klf4</i>	Transkriptionsfaktor, Erhält SC - Status	[110-112]	+
<i>yy1</i>	Transkriptionsfaktor, Interagiert mit <i>myc</i>	[113, 114]	+
<i>yy2</i>	Transkriptionsfaktor, leitet sich aus <i>yy1</i> ab	[114]	+
<i>oct4</i>	Transkriptionsfaktor, Erhält SC - Status	[115-117]	+/-
<i>nanog</i>	Transkriptionsfaktor, Erhält SC - Status	[101, 116, 117]	+/-
<i>scf (kitlg)</i>	Cytokin, Erhält SC-Status	[106, 118]	+
<i>cd44</i>	MSC - Typisches Oberflächenantigen	[72, 73]	+
<i>cd105</i>	MSC -Typisches Oberflächenantigen	[71, 73]	+
<i>cd45</i>	Oberflächenantigen, negativ in MSC	[71-73]	-

3.3 *IN VITRO* DIFFERENZIERUNG DER MSC

Im Anschluss an die Charakterisierung der undifferenzierten MSC sind Zellen der Passage 1 *in vitro* in Anlehnung an bereits bekannte Protokolle differenziert worden [70]. Für die Chondrogenese wurden MSC der Passage 2 eingesetzt. Die Zellen wurden nach Beginn der Differenzierungen nicht mehr passagiert und wuchsen folglich im mehrschichtigen, gewebeähnlichen Verband weiter (folgend aus dem englischen als „*multilayer*“ bezeichnet). Der Erfolg der Differenzierung wurde durch spezifische, histologische Färbungen überprüft. Als Kontrolle für jede Differenzierung wurden undifferenzierte MSC unter identischen Bedingungen im Standardmedium inkubiert und ebenfalls angefärbt. Auch diese Kontrollen wurden als Multilayer in Kultur gehalten. Sie werden im Folgenden als MSC_{alt} bezeichnet. Eine Übersicht aller Differenzierungen ist im digitalen Anhang D1 aufgeführt.

3.3.1 ADIPOGENE DIFFERENZIERUNG DER MSC

Die adipogene Differenzierung wurde wie unter Punkt 2.3.1 beschrieben durchgeführt. Ausgangspunkt für die Differenzierung waren konfluente Flaschen undifferenzierter MSC der Passage 1. Die MSC wiesen bereits wenige Tage nach Differenzierungsbeginn lichtmikroskopisch gut erkennbare, morphologische Veränderungen auf. Die Zellen begannen die charakteristische wirbelförmige Struktur zu verlieren und im mehrschichtigen, gewebeähnlichen Verband zu wachsen. Ab dem Tag 5 sind deutlich eine große Anzahl kleiner, gelblicher Einschlüsse in den adipogen differenzierten Zellen sichtbar. Dies lässt auf eine Anreicherung von Lipid-Tröpfchen in den Zellen schließen (siehe Abbildung 3 - 5). Gleichaltrige Kontrollen wiesen die in den adipogen differenzierten Flaschen beobachteten gelblichen Einschlüsse nicht auf. Die Differenzierung wurde über eine Dauer von 28 Tagen durchgeführt. Zum Nachweis der Differenzierung wurden adipogen differenzierte Kulturflaschen und Kontrollen fixiert und mit „Oil Red O“ gefärbt [76]. Der Farbstoff bewirkt eine Rotfärbung von Lipiden und wird häufig zum Nachweis adipogener Differenzierung von MSC genutzt [71]. Durch die Fixierung der Zellen im Rahmen der Färbung wurden die Lipideinschlüsse scheinbar ausgewaschen, die rötliche Färbung fällt im Vergleich zur Kontrolle nur unwesentlich stärker aus. Ein eindeutiger Nachweis der Differenzierung konnte

somit durch die Färbung nicht erbracht werden. Die RNA der am Tag 28 geernteten Zellen wurde für nachfolgende Versuche verwendet.

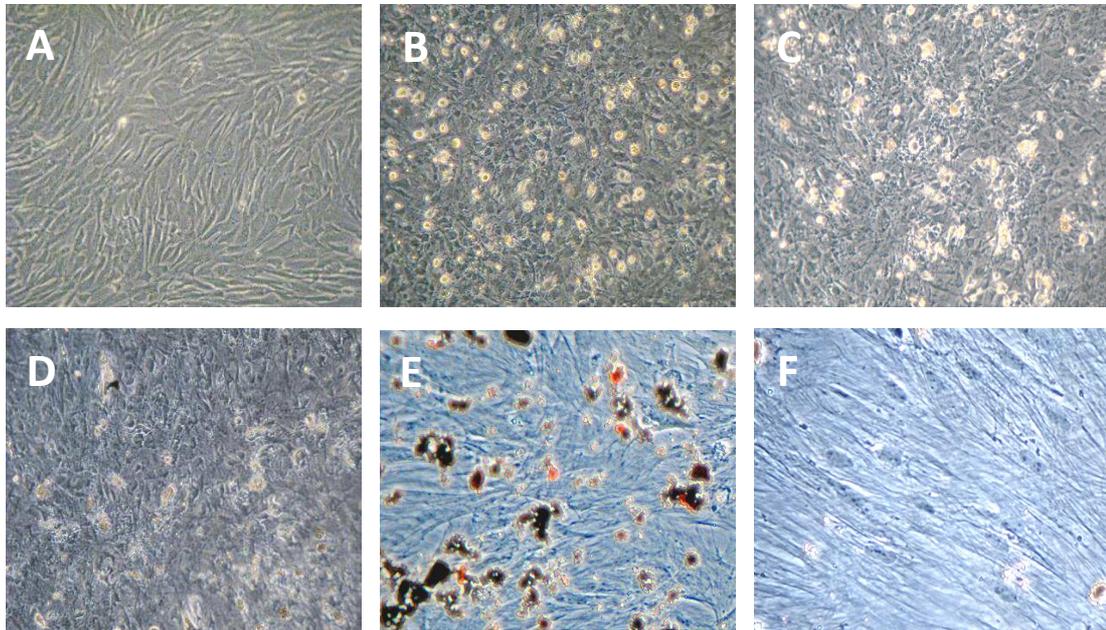


Abbildung 3 - 5 Adipogene Differenzierung Mesenchymaler Stammzellen

Die Differenzierung zu Adipozyten erfolgte über einen Zeitraum von 28 Tagen und wurde mit konfluent wachsenden MSC begonnen (A 100X). Bereits nach 5 Tagen (B 100X) sind viele gelblich erscheinende Einschlüsse in den Zellen zu erkennen, bei denen es sich vermutlich um Lipid-Tröpfchen handelt. Diese sind auch im weiteren Verlauf der Differenzierung am Tag 16 (C 100X) und Tag 28 (D 100X) erkennbar. Eine am Tag 21 durchgeführte „Oil Red O“ Fettfärbung (E 100X) zeigt nur sporadisch und uneindeutig die erwartete rötliche Färbung im Vergleich zur Kontrolle (F 200X). Abbildungsteil E und F aus Diplomarbeit von A. Dubberke (IMSB).

3.3.2 OSTEOGENE DIFFERENZIERUNG DER MSC

Die Osteogene Differenzierung wurde wie unter Punkt 2.3.2 beschrieben durchgeführt. Es wurden undifferenzierte MSC der Passage 1 verwendet. Wie schon bei der Adipogenese beobachtet, traten auch in diesem Versuch nach kurzer Zeit starke morphologische Veränderungen der differenzierten Zellen im Vergleich zur Kontrolle auf. Die osteogen differenzierten Zellen begannen sich zusammenzuziehen, wobei sie trotz der Volumenabnahme eine starke Proliferation erkennen ließen. Zudem wurden körnige Ablagerungen in den Zellen erkennbar. Die als Multilayer wachsenden Zellen verloren teilweise die Fähigkeit plastikadärent zu wachsen und begannen sich im Verlauf der Differenzierung vom Boden des Kulturgefäßes zu lösen. Die abgelösten Bereiche zeigten jedoch kein

lichtmikroskopisch sichtbares Zeichen mangelnder Vitalität. Ein Verlust an Zellmaterial als Folge des Medienwechsels konnte vermieden werden, da sich ablösende Bereiche mit den weiterhin adhärenen Bereichen fest verwachsen blieben. Die frei gewordenen Stellen am Boden der Kulturschale wurden im weiteren Verlauf der Differenzierung wieder bewachsen. Keine der beschriebenen morphologischen Veränderungen der differenzierten Ansätze war in den Kontrollansätzen zu beobachten. Nach 25 Tagen Differenzierung wurden die Zellen abgeerntet. Zur Kontrolle der Differenzierung diente die histologische Färbung mit „Alizarin Red“.

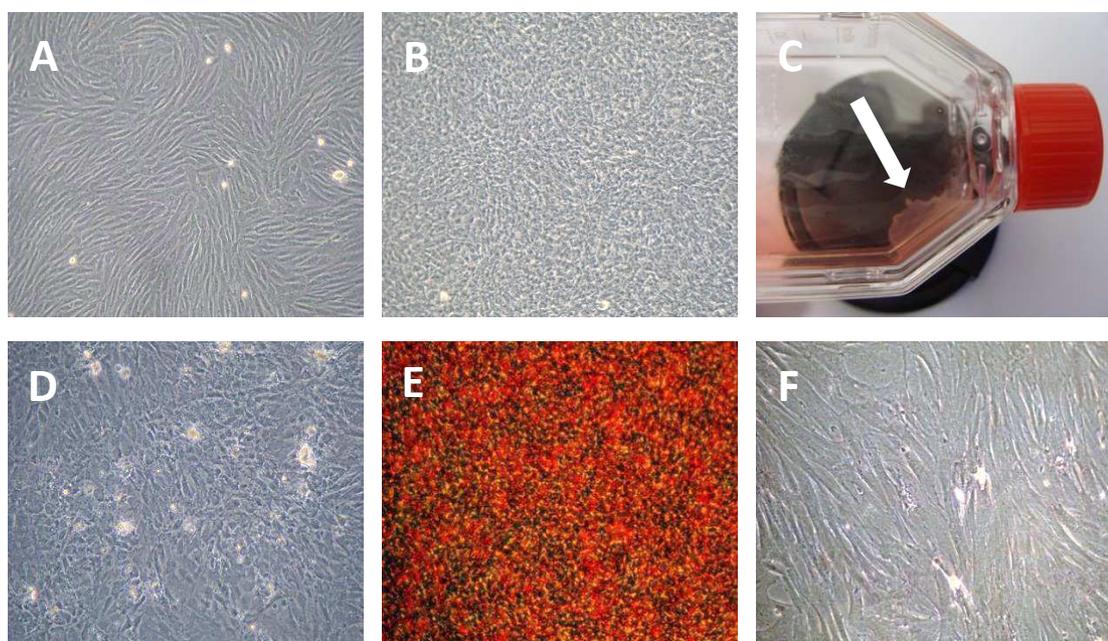


Abbildung 3 - 6 Osteogene Differenzierung Mesenchymaler Stammzellen

Die Differenzierung zu Osteoblasten erfolgte über einen Zeitraum von 25 Tagen und wurde mit konfluent wachsenden MSC begonnen (A 100 X). Nach 5 Tagen im Differenzierungsmedium ziehen sich die Zellen stark zusammen und es sind lichtbrechende Ablagerungen zu erkennen (B 100 X). Am Tag 14 der Differenzierung sind makroskopisch deutliche Ablösungen des Zellrasens in den Flaschen zu erkennen, die allerdings keinen Einfluß auf die Vitalität der Zellen zu haben scheinen (C siehe Pfeil). Die Zellen weisen über den gesamten Zeitraum bis zur Ernte normales Wachstum auf (D 100 X). Eine am Tag 21 durchgeführte Färbung mit Alizarin Red zeigt eindeutig die für Osteoblasten typischen, kalziumhaltigen Ablagerungen auf (E 100 X). Die Kontrolle (F 100 X) der über einen gleich langen Zeitraum gehaltenen undifferenzierten MSC zeigt hingegen keine Rotfärbung. Abbildungsteil E und F aus Diplomarbeit J. Rettich (IMSB).

Der Farbstoff bewirkt eine stark rote Färbung der für Osteoblasten typischen Kalziumablagerungen in der extrazellulären Matrix [119]. Der Vergleich mit der Kontrolle lässt auf eine erfolgreiche osteogene Differenzierung schließen (siehe

Abbildung 3 - 6). Färbungen wurden am Tag 21, 22 und 33 mit stets demselben Ergebnis durchgeführt (Diplomarbeiten A. Dubberke und J. Rettich, IMSB).

3.3.3 CHONDROGENE DIFFERENZIERUNG DER MSC

Die Chondrogene Differenzierung ist wie Punkt 2.3.3 beschrieben durchgeführt worden. Es wurden zunächst konfluent wachsenden MSC der ersten Passage in Standard Zellkulturflaschen verwendet. Schon nach 2 Tagen waren deutliche Veränderungen der Zellstruktur im Vergleich zur Kontrolle erkennbar. Die Zellen begannen zu kondensieren und erschienen im weiteren Verlauf der Differenzierung immer stärker granuliert.

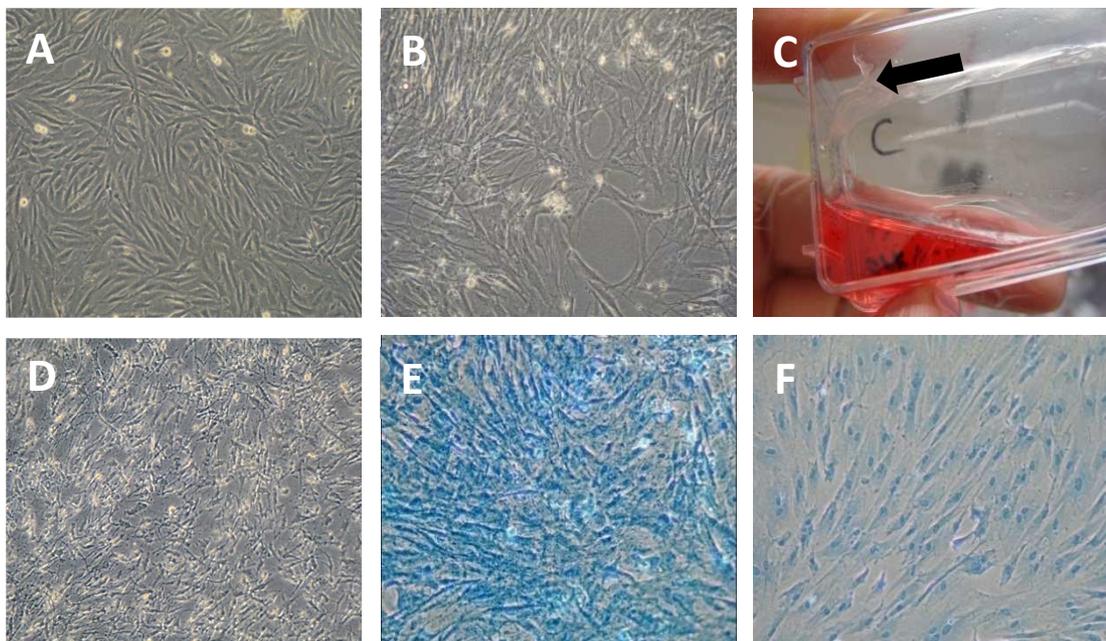


Abbildung 3 - 7 Chondrogene Differenzierung Mesenchymaler Stammzellen

Die Differenzierung zu Chondrozyten erfolgte über einen Zeitraum von 5 Tagen und wurde mit konfluenten Zellen der Passage 2 begonnen (A 100 X). Bereits am Tag 2 der Differenzierung begannen sich die Zellen vom Schalenboden abzulösen und zu kondensieren (B 100 X). Meist ab Tag 7 sind die Ablösungserscheinungen auch mit bloßem Auge gut sichtbar, der Verlust an Zellmaterial durch Medienwechsel wird für nachfolgende Versuche kritisch groß (C). Um ausreichend große Mengen RNA isolieren zu können, wurde die Differenzierung daher nach 5 Tagen abgeschlossen (D 100 X). Die gezeigten Alcian Blue Färbungen wurden im Rahmen der Diplomarbeit von J. Rettich mit 9 Tage lang differenzierten MSC durchgeführt. Sie zeigen die chondrogene Differenzierung (E 100 X) im Vergleich zur Kontrolle (F) an.

Es zeigte sich, daß sich die im Differenzierungsmedium wachsenden Zellen bereits nach wenigen Tagen ablösen und im Verlauf der Medienwechsel ausgewaschen wurden. Ein Wachstum im Multilayer konnte aufgrund der Ablösungen nicht

beobachtet werden. Daher wurden in einem separaten Ansatz undifferenzierte MSC der Passage 1 aus Cryostocks aufgetaut und in speziellen, die Zelladhärenz fördernden Petrischalen der Fa. Greiner in Kultur genommen. Die Zellen zeigten leider auch in diesen Gefäßen die zuvor in den Standard Zellkulturflaschen beobachteten Ablösungserscheinungen im Verlauf der Differenzierung. Die Differenzierung in der vorliegenden Arbeit wurde daher über einen Zeitraum von nur 5 Tagen durchgeführt, um einen zu großen Verlust an Zellmaterial zu vermeiden und in jedem Fall genug Zellen für eine RNA – Präparation gewinnen zu können. Als Färbemittel zum Test auf eine chondrogene Differenzierung wird „Alcian Blue“ Lösung verwendet [119]. Für Chondrozyten typische, saure und sulfatreiche Proteoglycane werden so blau angefärbt (Siehe Abbildung 3 - 7).

3.4 RNA - PRÄPARATION UND RNA - AMPLIFIZIERUNG

Zur Analyse der Genexpression mittels RT-PCR, Microarrays und Hochdurchsatz Sequenzierung muss zunächst die Gesamt-RNA aus den Zellkulturen gewonnen werden. Eine Übersicht aller in dieser Arbeit verwendeten Gesamt-RNA Präparationen ist in Tabelle 3 - 2 aufgelistet. Gesamt-RNA der undifferenzierten MSC der Passage 0 (MSC_neu) wurde im Rahmen der Diplomarbeit von I. Grechowa (IMSB) präpariert.

Tabelle 3 - 2 Übersicht der Gesamt-RNA Präparationen

Die Tabelle zeigt eine Übersicht der Zellkulturen die zur Präparation von Gesamt-RNA in dieser Arbeit verwendet wurden. Die als „alte MSC“ bezeichneten Zellen wurden keinem Differenzierungsmedium ausgesetzt, jedoch als Kontrolle der Differenzierungen für 31 Tage nach Passage 1 im Multilayer (ML) gehalten.

Differenzierung	Kulturdauer gesamt	Zeitraum Differenzierung	Passage	Probename Zellkultur	Wuchsform
Undiff. (msc_neu)	14 Tage	-	0	Präp. I. Grechowa	monolayer
Undiff. (msc_alt)	40 Tage	31 Tage ML.	1	PK 4-2 P1	multilayer
Adipogen	40 Tage	28 Tage	1	SPK 1-2 P1	multilayer
Osteogen	40 Tage	25 Tage	1	SPK 5-3 P1	multilayer
Chondrogen	23 Tage	5 Tage	2	PK 10-1&4 P2	monolayer

Die Quantität und Qualität der RNA wurde durch eine Messung im „Agilent 2000 Bioanalyzer“ bestimmt (siehe Punkt 2.2). Da sowohl für Microarray Experimente (4 µg mRNA/Array), als auch für die Hochdurchsatz Sequenzierung große Mengen RNA benötigt wurden, wurde der mRNA Anteil der Gesamt-RNA einiger Proben

zur Erhöhung der Ausbeute in zwei aufeinander folgenden Schritten linear amplifiziert. Auch der Erfolg der Amplifikationen wurde durch Messung der Proben im Bioanalyser überprüft.

3.4.1 PRÄPARATION VON GESAMT-RNA AUS ZELLKULTUREN

Die Präparation der Gesamt-RNA aus Zellkulturen erfolgte nach Anleitung der Hersteller der Präparationskits (MatMeth). In jedem Falle wurde der in den Kits als optional angebotene DNase Verdau der extrahierten Gesamt-RNA durchgeführt. Als Maß für die Qualität der RNA dient der RIN Wert [120].

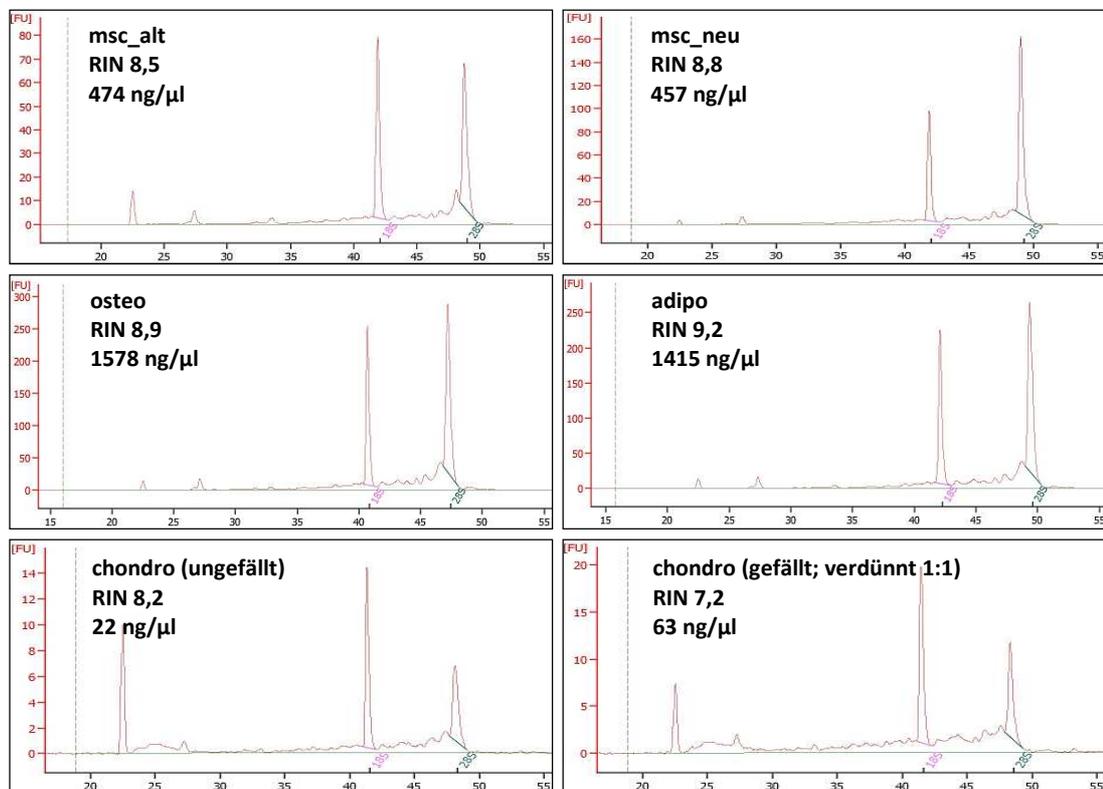


Abbildung 3 - 8 Bioanalyser Messung der Gesamt-RNA

Gezeigt sind die Elektropherogramme und Messwerte der Gesamt-RNA Extraktionen. Die Intensität der Fluoreszenz ist auf der Y -, die Zeit in Sekunden auf der X - Achse abgebildet. Die vier prominenten Peaks in jeder Messung stellen den Marker, die 5 S, 18 S und 28 S rRNA dar (v. l. n. r.). Der RIN Wert dient der Qualitätsbewertung der Proben, der maximal erreichbare Wert ist 10.

Für die Präparation wurden die Zellen aus einer Zellkulturflasche mit 25 cm² Bodenfläche eingesetzt. Im Falle der RNA Isolation aus Chondrozyten wurden die Zellen aus zwei Schalen vor der Präparation zusammen gefügt, um trotz der unter Punkt 3.3.3 beschriebenen Zellverluste während der Differenzierung genügend

Zellen für die Extraktion zur Verfügung zu haben. Anschließend wurde 1 µl der RNA Lösung mit dem Bioanalyser getestet. Bis auf die Gesamt-RNA aus Chondrozyten wiesen alle Proben eine für nachfolgende Versuche geeignete Qualität, Menge und Konzentration auf. Die RNA der Chondrozyten ist mit 22 ng/µl zu niedrig konzentriert und wird gefällt. Danach erfolgt eine erneute Messung im Bioanalyser. Um den mit der Messung verbundenen Probenverlust gering zu halten, wird eine 1:1 verdünnte Probe gemessen. Die Elektropherogramme und Messwerte des Bioanalyzers sind in Abbildung 3 - 8 gezeigt. Durch die Fällung konnte eine RNA Lösung aus Chondrozyten mit einer Konzentration von 126 ng/µl gewonnen werden. Diese Konzentration ist für nachfolgende Versuche ausreichend, auch der durch die Fällung um 1 gesunkene RIN Wert stellt keine kritische Qualitätseinbuße dar.

3.4.2 LINEARE AMPLIFIKATION VON RNA

Die erste Runde der linearen Amplifikation des mRNA Anteils der Gesamt-RNA Proben konnte gemäß der Herstellerangaben mit leichten Anpassungen durchgeführt werden, die zweite Runde wurde in einem komplett angepassten Verfahren durchgeführt (siehe Punkt 2.5.4). Bei jeder Amplifikation wird sog. „sense mRNA“ synthetisiert, so daß die Amplifikate mehr oder weniger exakte Abbilder der *in vivo* mRNA darstellen. Die Resultate der Amplifikationsrunden sind im digitalen Anhang D3 einzusehen. Abbildung 3 - 9 zeigt die Ergebnisse der Messung mittels Bioanalyser aus der zweiten Amplifikationsrunde. Die zwei Runden der Amplifikation konnten für alle RNA-Proben erfolgreich durchgeführt werden. Die verbleibende Kontamination der Proben mit rRNA beträgt nach 2 Amplifikationen maximal 2,4 % in der aRNA der Chondrozyten und ist somit für nachfolgende Anwendungen nicht hinderlich. Anhand des Molekulargewicht Standards lassen sich Aussagen über die Größe der amplifizierten Fragmente und deren Verteilung in der Probe machen. Die aRNA aus Adipozyten, Osteoblasten und den undifferenzierten MSC zeigt die stärkste mittlere Fluoreszenz zwischen 24 und 34 Sekunden Laufzeit. Im Vergleich zum Größenstandard entspricht das einer bei ca. 25 Nukleotiden beginnenden Verteilung der Fragmentgrößen, die ihr Maximum bei ungefähr 200 Nukleotiden erreicht um dann langsam zu Fragmenten von 500 Nukleotiden Länge abzufallen.

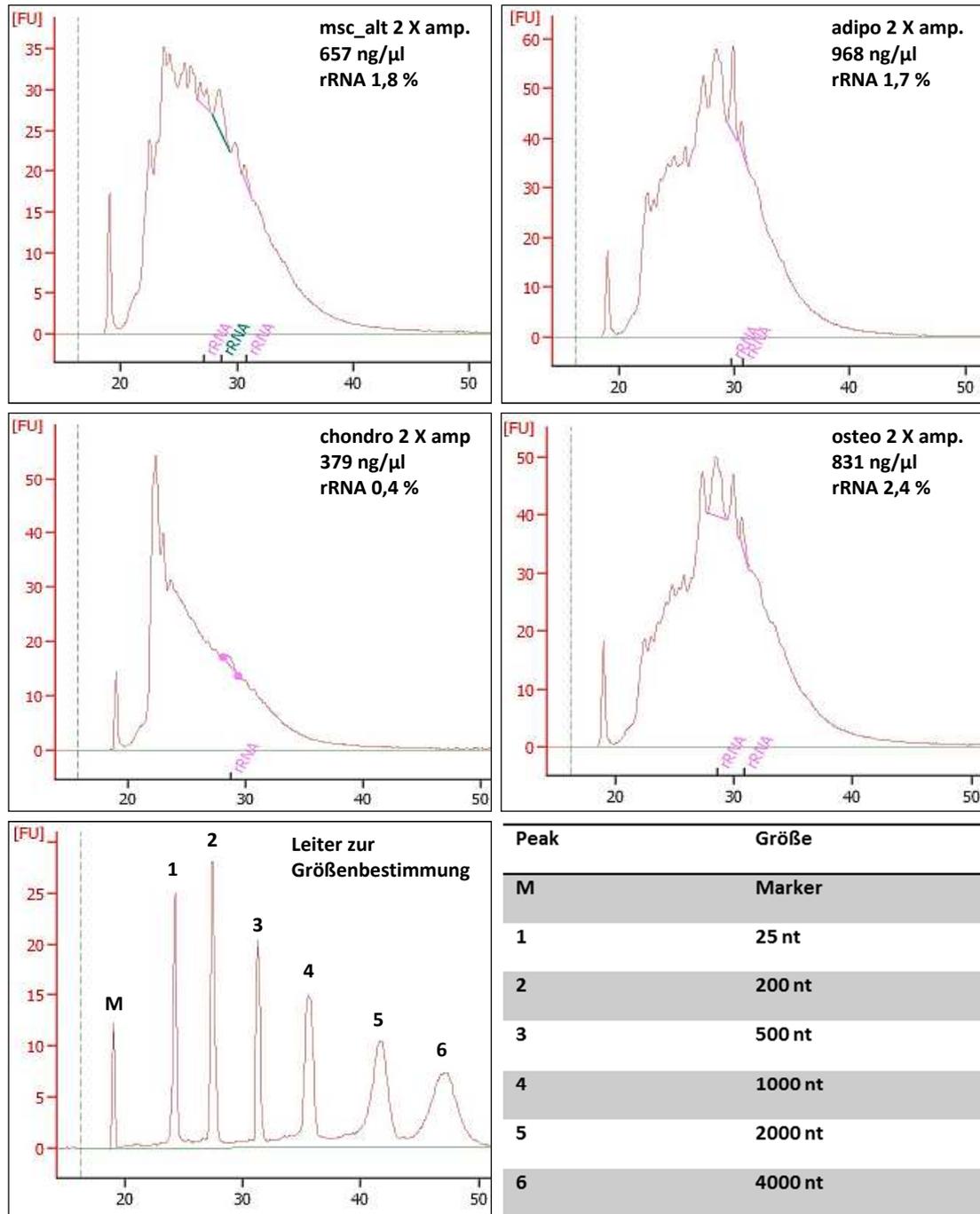


Abbildung 3 - 9 Bioanalyser Messung der 2 X amplifizierten RNA

Gezeigt sind die Elektropherogramme und Messwerte der zweiten Runde der RNA Amplifikation. Die Intensität der Fluoreszenz ist auf der Y -, die Zeit in Sekunden auf der X – Achse abgebildet. Die verbleibende Kontamination der Amplifikate mit rRNA ist in Prozent angegeben. Der prominente Ausschlag bei ca. 18s stellt den Marker dar. Die Leiter und Tabelle dienen der Bestimmung der Größe und Größenverteilung der amplifizierten RNA Fragmente in Nukleotiden anhand der Laufzeit.

Eine Übersicht über die Ausbeute und Effizienz beider Amplifikationsrunden ist nachfolgend in Tabelle 3 - 3 aufgeführt.

Tabelle 3 - 3 Ausbeute und Effizienz der RNA Amplifikationen

RNA	Amplifikation 1 X			Amplifikation 2 X	
	Eingesetzte Menge Gesamt-RNA	Errechnete mRNA Menge (bei 1 % mRNA)	Ausbeute und Effizienz	Eingesetzte Menge 1 X amp. RNA	Ausbeute und Effizienz
msc_alt	1,422 µg	14,22 ng	9,1 µg - 638 X	3,30 µg	76,5 µg - 23 X
osteo	4,734 µg	47,34 ng	13,3 µg - 280 X	4,82 µg	92,1 µg - 19 X
adipo	4,245 µg	42,45 ng	13,2 µg - 311 X	4,80 µg	103,3 µg - 22 X
chondro	0,378 µg	3,78 ng	7,9 µg - 2095 X	2,88 µg	38,8 µg - 13 X

Zur Berechnung der Amplifikationseffizienz der ersten Amplifikationsrunde wird von einem Anteil von 1 % mRNA in den eingesetzten Gesamt-RNA Proben [121] ausgegangen. Auffällig ist, daß die Amplifikation in der ersten Runde umso effizienter abläuft, je weniger Gesamt-RNA eingesetzt wird (siehe Tabelle 3 - 3). So wird in der ersten Amplifikationsrunde der RNA aus Chondrozyten bei einem hypothetischen Einsatz von 3,78 ng mRNA mit einer 2095 fachen Amplifikation mehr als dreimal so effizient amplifiziert, wie dies bei Amplifikation der RNA aus den undifferenzierten MSC mit einem hypothetischen mRNA Anteil von 14,22 ng zu beobachten war. Die erste Runde der Amplifikation der RNA aus Adipozyten und Osteoblasten zeigt bei einem ähnlichen und verhältnismäßig hohen Mengeneinsatz von RNA auch ähnliche und hohe Ausbeuten. Die Effizienz der Amplifikation liegt jedoch ca. um den Faktor 7 unter der für Chondrozyten bestimmten Amplifikationseffizienz. Diese Korrelation aus geringer Menge und hoher Amplifikationseffizienz ist für die zweite Runde der Amplifikation nicht mehr zu beobachten. In die zweite Runde der Amplifikation wird aRNA aus der ersten Runde (statt Gesamt-RNA) eingesetzt, was eine immense Steigerung der Matrizenanzahl zum Beginn der Amplifikation bedeutet. Im Labor zeigte sich bereits direkt nach Abschluss der *in vitro transcription* (IVT) durch ausgefallene Nukleinsäuren die immense Menge an transkribiertem Material an.

3.5 BESTIMMUNG DIFFERENZIELLER GENEXPRESSION MITTELS PLASMID - DNA MICROARRAYS

Bei den in dieser Arbeit verwendeten Microarrays handelt es sich um selbst hergestellte Plasmid DNA-Microarrays. Die Plasmide tragen als genspezifisches Integrat ein cDNA Molekül, welches im Rahmen des EST Projektes der Dissertation von Steffen Rapp [82] aus RNA der fötalen Wachstumsfuge von Rindern

synthetisiert, kloniert und sequenziert wurde. Es wurde stets 2 X amplifizierte cDNA aus undifferenzierten MSC (msc_alt) mit 2 X amplifizierter cDNA aus einer differenzierten Linie gleichzeitig auf den Microarray hybridisiert. Ein Gen wird als differenziell exprimiert betrachtet, wenn das Verhältnis der Expression nach globaler Normalisierung einen zur Basis 2 logarithmierten Wert von ± 1 in mindestens 8 von 10 Microarrays (7 von 9 für Adipozyten vs. MSC Vergleich) überschreitet (siehe 2.5.6). Zur technikenunabhängigen Validierung können den Microarray Ergebnissen die Expressionsverhältnisse, wie sie in der NGS Analyse ermittelt wurden, gegenübergestellt werden (siehe Punkt 4.3). Referenzdaten sind hierbei die Werte des BLASTn Abgleichs der Sequenzen mit der EST beinhaltenden NCBI Unigene Datenbank (siehe Punkt 3.6.2)

3.5.1 HERSTELLUNG DER PLASMID-DNA MICROARRAYS

Von allen im EST Projekt klonierten und sequenzierten Plasmiden war nur ein Teil als Sonde für den Microarray von Interesse. Klone, die als Integrat ribosomale RNA (rRNA), zu kurze Integrate oder nur leeren Vektor enthalten, sind für Genexpressionsanalysen ungeeignet. Auch Klone, die nicht eindeutig sequenziert werden konnten, wurden nicht verwendet. Die übrigen Plasmide wurden zunächst in chemisch kompetente Zellen retransformiert, um die EST Bank als solche erhalten zu können. Dabei wurden von 3936 Klonen des EST Projektes 2242 Klone im 96er Format retransformiert und die Plasmide anschließend isoliert. Der Erfolg der Transformation wurde stichprobenartig mittels Agarosegelelektrophorese überprüft. Hierfür wurden zufällig ausgewählte Plasmide aus einer Platte auf ein Gel aufgetragen (siehe Abbildung 3 - 10 A). Die Konzentration von ebenfalls nach dem Zufallsprinzip ausgewählten Plasmiden wurde zudem photometrisch bestimmt (siehe Abbildung 3 - 10 B). Die Retransformation und Isolation der Sondenplasmide konnte erfolgreich durchgeführt werden.

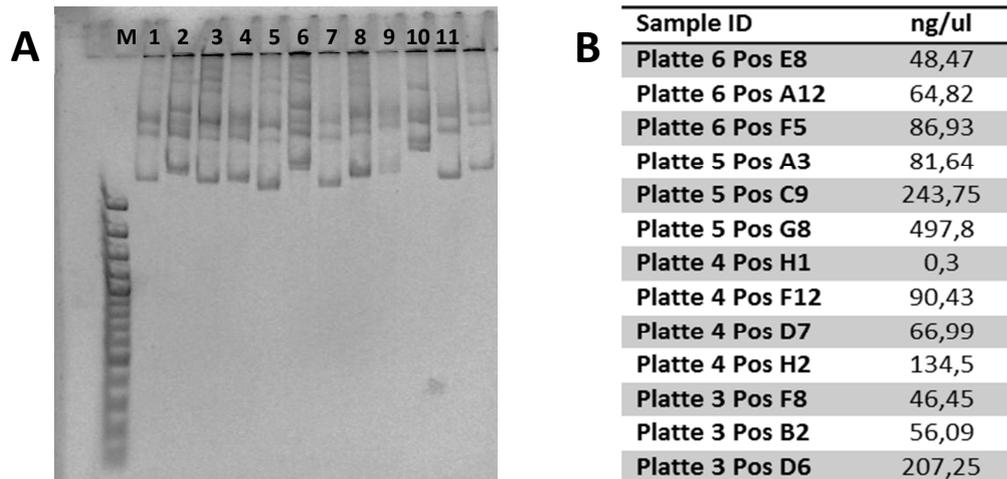


Abbildung 3 - 10 Überprüfung der Sonden zur Herstellung des Microarrays

Auf das 1 % Agarosegel sind 12 zufällig ausgewählte Plasmide der Platte 1 (5 µl aus 70 µl) aufgetragen, die als Sonden auf den Microarray gespottet werden sollen (A). Es zeigen sich die für ringförmige Plasmide typischen Banden im Gel. Verwendet wurde ein 100 bp Marker, die am weitesten oben im Gel laufende Bande hat ein Molekulargewicht von 3000 bp. Photometrische Messungen zeigen, daß die Konzentration der im 96er Format isolierten Plasmide stark schwankt (B), bzw. die Retransformation nicht immer erfolgreich verlief (siehe Teil B Platte 4 Pos H1).

Um die mit diesem Material hergestellten Microarrays vor Beginn der eigentlichen Experimente überprüfen zu können, wurden einzelne Microarrays mit dem sich an DNA anlagernden Fluoreszenzfarbstoff SYBR Green I gefärbt. Qualitative Mängel der Microarrays können so sichtbar gemacht werden (Abbildung 3 - 11). Durch die Färbung ließen sich die allermeisten Sonden im Laserscanner nachweisen.

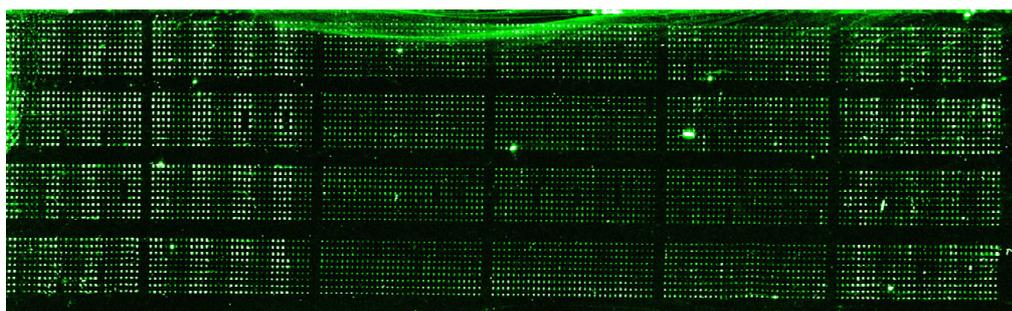


Abbildung 3 - 11 SYBR Green I Färbung eines Microarrays

Die SYBR Green I Färbung zeigt deutlich die Signale der einzelnen Sonden. Auffällig ist, daß besonders die mittleren 8 Subarrays schwächere Signale als die restlichen Subarrays aufweisen.

Ein flächendeckender Ausfall der Retransformation, Plasmidisolierung oder beim Spotten war nicht zu verzeichnen. Unterbrechungen im Muster des Arrays zeigen an, wenn Sonden an einzelnen Stellen nicht erfolgreich gespottet werden konnten. Die Ausfallrate ist jedoch nicht so hoch, daß eine signifikante Anzahl Sonden

betroffen wäre. Zur besseren Orientierung wurden in der fertigen Version der Arrays drei Markerspots vor jeden Subarray hinzugefügt, der gezeigte Microarray stellt also eine noch nicht komplette Erstversion direkt nach dem Spotten dar. Wurden die Lücken durch technische Fehler beim Spotten verursacht, konnten sie durch gezieltes Nachspotten noch geschlossen werden. Auffällig ist, daß die Signale nicht gleichmäßig stark über den Microarray verteilt sind. Besonders die in den mittleren Subarrays befindlichen Sonden lieferten aufgrund geringerer Sondenkonzentration schwächere Signale, als die der an den Rändern gespotteten Subarrays. Die Informationen über die Sonden, die Rohdaten und Auswertungen aller Microarrays sind im digitalen AnhangD2 einzusehen.

3.5.2 BESTIMMUNG DIFFERENZIELLER GENEXPRESSION MITTELS MICROARRAYS – VERGLEICH VON UNDIFFERENZIIERTEN MSC UND OSTEOLASTEN

Durch Microarray Experimente konnten im Vergleich zwischen undifferenzierten und osteogen differenzierten MSC 10 Sonden ermittelt werden, die auf mindestens 8 der 10 hybridisierten Microarrays auf eine differenzielle Expression der repräsentierten 7 Gene schließen lassen. Tabelle 3 - 4 listet alle differenziell exprimierten Gene auf. Von den den 10 Sonden zugrunde liegenden EST Sequenzen kodieren 4 für das Gen *slc25a6*, welches in osteogen differenzierten Zellen differenziell stärker exprimiert wird. Das Gen *enpp1*, welches ebenfalls in Osteoblasten differenziell hoch exprimiert wird, kodiert für eine Pyrophosphatase/Phosphodiesterase. Als ebenfalls in den differenzierten Zellen stark exprimiert konnte das Gen *destrin* bestimmt werden. Das letzte in Osteoblasten differenziell stärker exprimierte Gen kodiert für das ribosomale Protein S2. Das in den undifferenzierten MSC überexprimierte Gen *sema3c* kodiert für ein sekretiertes Signalmolekül, das Transkript des Gens *slc39a1* kodiert ein membranständiges Zink Transporter Protein. Das Gen *gapdh* wurde auch zu Testzwecken, als potentiell Haushaltsgen gespottet. Es scheint in undifferenzierten MSC gegenüber den Osteoblasten differenziell exprimiert zu werden, was für ein Haushaltsgen eher nicht zu erwarten war. Auf eine Normalisierung der Arrayergebnisse anhand der Expression von Haushaltsgenen wurde verzichtet.

Tabelle 3 - 4 Differenzielle Genexpression - Osteoblasten vs. undifferenzierte MSC im Microarray

Die Tabelle listet alle Gene auf, die in mindestens 8 von 10 hybridisierten Microarrays in dieselbe Richtung als differenziell exprimiert ermittelt wurden. Die Sondenzahl gibt an, wie oft die differenzielle Expression mittels unterschiedlicher Sonden auf einem Array gemessen werden konnte.

Sondenzahl	Gen (BLASTn Treffer EST)	Richtung	Log2(R/G)
4	<i>Bos taurus</i> solute carrier family 25 member 6 (SLC25A6)	Differenziert hoch	1,18 - 1,64
1	<i>GAPDH</i>	Undifferenziert hoch	1,15
1	ectonucleotide pyrophosphatase/phosphodiesterase 1 (ENPP1)	Differenziert hoch	1,19
1	<i>Bos taurus</i> chromosome 4 genomic contig similar to semaphorin 3C	Undifferenziert hoch	1,49
1	<i>Bos taurus</i> destrin (actin depolymerizing factor) (DSTN)	Differenziert hoch	1,27
1	<i>Bos taurus</i> ribosomal protein S2	Differenziert hoch	1,1
1	<i>Bos taurus</i> solute carrier family 39 (zinc transporter), member 1 (SLC39A1)	Undifferenziert hoch	1,68

3.5.3 BESTIMMUNG DIFFERENZIELLER GENEXPRESSION MITTELS MICROARRAYS – VERGLEICH VON UNDIFFERENZIIERTEN MSC UND CHONDROZYTEN

Durch Microarray Experimente konnten im Vergleich zwischen undifferenzierten und chondrogen differenzierten MSC 41 Sonden ermittelt werden, die auf mindestens 8 der 10 hybridisierten Microarrays auf eine differenzielle Expression der repräsentierten 11 Gene schließen lassen (siehe Tabelle nachfolgende Seite). Auffällig ist die große Zahl der Sonden, die eine differenzielle Expression anzeigen. Weiterhin ist bemerkenswert, daß eine differenzielle, hohe Genexpression ausschließlich für undifferenzierte MSC bestimmt werden konnte.

Tabelle 3 - 5 Differenzielle Genexpression - Chondrozyten vs. undifferenzierte MSC im Microarray

Die Tabelle listet alle Gene auf, die in mindestens 8 von 10 hybridisierten Microarrays in dieselbe Richtung als differenziell exprimiert ermittelt wurden. Die Sondenzahl gibt an, wie oft die differenzielle Expression mittels unterschiedlicher Sonden auf einem Array gemessen werden konnte.

Sondenzahl	Gen (BLASTn Treffer EST)	Richtung	Log2(R/G)
19	<i>Bos taurus</i> secreted phosphoprotein 1 (<i>SPP1</i>)	Undifferenziert hoch	1,12 -2,56
11	<i>Homo sapiens</i> migration-inducing protein 6 (<i>MIG6</i>)	Undifferenziert hoch	1,38 - 1,86
2	Eukaryotic translation elongation factor 1 alpha 1 (<i>EEF1A1</i>)	Undifferenziert hoch	1,42 - 1,48
2	<i>Bos taurus</i> connective tissue growth factor (<i>CTGF</i>)	Undifferenziert hoch	1,31 - 1,52
1	<i>Bos taurus</i> solute carrier family 39 (zinc transporter), member 1 (<i>SLC39A1</i>)	Undifferenziert hoch	1,28
1	Bovine mRNA 3' fragment for type II collagen pro-alpha 1 chain (<i>COL2A1</i>)	Undifferenziert hoch	1,8
1	<i>Homo sapiens</i> CCAAT/enhancer binding protein (C/EBP), delta (<i>CEBPD</i>)	Undifferenziert hoch	1,29
1	<i>Bos taurus</i> matrix Gla protein (<i>MGP</i>)	Undifferenziert hoch	1,53
1	<i>Homo sapiens</i> DEAD (Asp-Glu-Ala-Asp) box polypeptide 3, X-linked (<i>DDX3X</i>)	Undifferenziert hoch	1,6
1	<i>Bos taurus</i> similar to 60S ribosomal protein L17 (<i>RPL17</i>)	Undifferenziert hoch	1,12
1	<i>Bos taurus</i> ATP synthase, H ⁺ transporting, mitochondrial FO complex (<i>ATP5G3</i>)	Undifferenziert hoch	1,15

3.5.4 BESTIMMUNG DIFFERENZIELLER GENEXPRESSION MITTELS MICROARRAYS – VERGLEICH VON UNDIFFERENZIIERTEN MSC UND ADIPOZYTEN

Durch Microarray Experimente konnten im Vergleich zwischen undifferenzierten und adipogen differenzierten MSC 4 Sonden ermittelt werden, die auf mindestens 7 der 9 hybridisierten Microarrays auf eine differenzielle Expression der repräsentierten 3 Gene schließen lassen (siehe Tabelle 3 - 6)

Tabelle 3 - 6 Differenzielle Genexpression - Adipozyten vs. undifferenzierte MSC im Microarray

Die Tabelle listet alle Gene auf, die in mindestens 7 von 9 hybridisierten Microarrays in dieselbe Richtung als differenziell exprimiert ermittelt wurden. Die Sondenzahl gibt an, wie oft die differenzielle Expression mittels unterschiedlicher Sonden auf einem Array gemessen werden konnte.

Sondenzahl	Gen (BLASTn Treffer EST)	Richtung	Log2(R/G)
1	<i>Bos taurus</i> follistatin-like 1 (<i>FSTL1</i>),	Differenziert hoch	1,94
2	<i>Bos taurus</i> secreted phosphoprotein 1 (<i>SPP1</i>)	Undifferenziert hoch	1,26 - 1,43
1	<i>Bos taurus</i> solute carrier family 25 member 6 (<i>SLC25A6</i>),	Differenziert hoch	1,82

Im Vergleich der Genexpression zwischen MSC und Adipozyten wird der osteogene Marker *spp1* in den undifferenzierten MSC differenziell hoch exprimiert. Das Gen *slc25a6* wird in Adipozyten stärker exprimiert als in undifferenzierten MSC, selbiges gilt für das Gen *fstl1*.

3.6 GENEXPRESSIONSANALYSE MITTELS ILLUMINA HOCHDURCHSATZ SEQUENZIERUNG – PROZESSIEREN DER ROHDATEN

Um einen das komplette Transkriptom umfassenden Einblick in die Genexpression der MSC und ihrer in vitro differenzierten Folgelinien zu erhalten, wurde das von der Fa. Illumina entwickelte Hochdurchsatz-Sequenzierungsverfahren mit der aus den Zellen isolierten RNA durchgeführt [122]. Ein Teil der RNA - Proben wurde im „Cologne Center for Genomics“ (Köln), ein weiterer Teil durch Mitarbeiter des

Nukleinsäure-Kompetenzzentrums der Universität Mainz sequenziert. Die zur Sequenzierung eingesetzte RNA stammt teilweise aus amplifizierten, teilweise aus Gesamt-RNA Proben. Die cDNA Synthese war stets Teil des Sequenzierservice. Die sich aus den verschiedenen eingesetzten Illumina Gerätetypen und Sequenzierungsverfahren ergebenden Unterschiede in Hinblick auf die erhaltenen Rohdaten sind in Tabelle 3 - 7 aufgeführt.

Tabelle 3 - 7 Übersicht der Illumina Sequenzierungen

Sequenzierung	RNA	Plattform / Ort	Paired / Single	Sequenzlänge	Sequenzen
msc_neu	Gesamt-RNA	HiSeq 2000 Mainz	paired end	101 bp	104.774.556
adipo	Gesamt-RNA	HiSeq 2000 Mainz	single	51 bp	64.723.146
msc_alt	2 X amp.	GAllx Köln	paired end	95 bp	92.385.296
osteo	2 X amp.	GAllx Köln	paired end	95 bp	87.599.640
chondro	2 X amp.	GAllx Köln	paired end	95 bp	88.677.062

Die Rohdatenmenge liegt zwischen 3,3 Gigabasen (Adipozyten) und 10,6 Gigabasen (m_sc_neu) RNA. Um Genexpressionsanalysen durchführen zu können, wurden in dieser Arbeit drei unterschiedliche Ansätze zur Ermittlung der Expressionsstärke jedes Gene durchgeführt (siehe Punkt 2.6.).

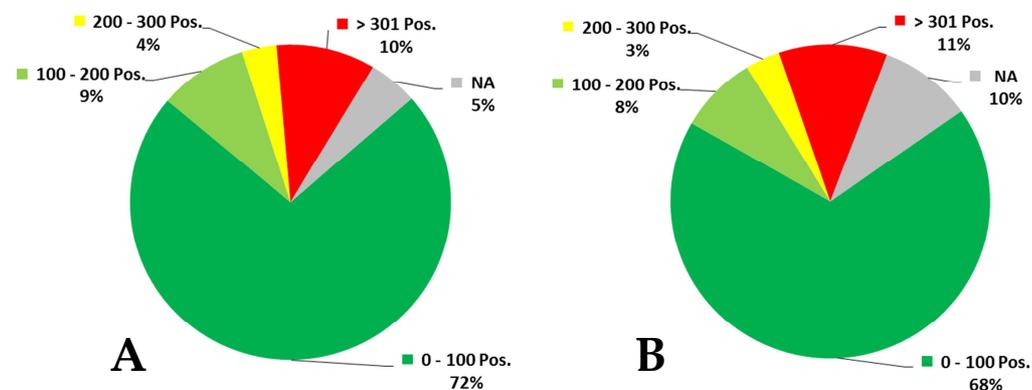


Abbildung 3 - 12 Expressionsstärke der 200 höchstexprimierten Gene in Abhängigkeit zur Bestimmungsmethode

Die Expressionsstärke der Gene wurde für jeden NGS Datensatz durch BLASTn Abgleich der Sequenzen mit der Unigene Datenbank, durch Kartieren der Sequenzen gegen annotierte Genloci, und durch BLASTn Abgleich von Consensussequenzen gegen die Refseq-RNA Datenbank ermittelt. Die Gene wurden für jede Bestimmungsmethode gemäß ihrer Expressionsstärke absteigend sortiert gelistet. Der Listenplatz eines Gens durch Kartierung dient als Referenzwert zum Vergleich mit der Platzierung desselben Gens durch die anderen beiden Bestimmungsmethoden. Für 68 bis 72 % der 200 höchstexprimierten Gene beträgt der Abstand in den genannten Listen weniger als 100 Positionen. Grau unterlegt ist der Anteil an Genen, die nicht in allen drei verglichenen Listen geführt werden (NA). Abbildungsteil A – RNA-Seq Adipozyten B – RNA-Seq junge MSC.

Abbildung 3 - 12 zeigt, daß die drei unterschiedlichen Methoden zur Bestimmung der Expressionsstärke eines Gens vergleichbare Ergebnisse liefern. Die technisch bedingt variierende Länge der Rohdaten (51 bp adipo, 105 bp msc_neu) scheint das Ergebnis der Analyse nicht wesentlich zu beeinflussen.

3.6.1 PROZESSIEREN DER ROHDATEN UND ANNOTATION DER SEQUENZEN – KARTIERUNG GEGEN DAS RINDERGENOM

Für die drei unterschiedlichen Methoden zur Bestimmung der Expressionsstärke wurden unterschiedliche Ansätze zur Prozessierung der Daten vor der eigentlichen Genexpressionsanalyse durchgeführt. Den am wenigsten aufwändigen Ablauf bietet die kommerzielle Software „CLC Genomics Workbench“, die vom Import der Rohdaten bis zum Kartieren der prozessierten Daten mittels RNA-Seq gegen annotierte Bereiche des Genoms alle nötigen Schritte im Programmpaket vorhält. Da nur die Sequenzen in die Auswertung eingehen können, für die bereits eine Annotation in den NCBI Genebank (gbk) Dateien des Rindergenoms vorliegt, müssen nicht genkodierende Sequenzen nicht vor der Analyse entfernt werden. In Tabelle 3 - 8 sind die Ergebnisse der Aufreinigung und Kartierung aller Banken aufgelistet.

Tabelle 3 - 8 Prozessieren und Kartierung der Rohdaten – Kartierung mittels CLC

Gezeigt ist eine Übersicht der Prozessierung der Rohdaten bis zur Kartierung mittels RNA-Seq (Software CLC Genomics Workbench). Je Arbeitsschritt ist die Datenmenge in Sequenzen und auf ganze Prozent der Rohdaten gerundet angegeben. NA – Rohdatenfile der Adipozyten (fastq - Format) enthält keinen „Purity Filter“ Wert.

Prozess	msc_neu		msc_alt		adipo		osteo		chondro	
Rohdaten	104.774.556		92.385.296		64.723.146		87.599.640		88.677.062	
Purity Filter (PF)	3.320.750	3 %	15.848.788	17 %	NA	13.686.354	16 %	14.842.762	17 %	
Trimming	1.782.180	2 %	2.666.171	3 %	6.326.424	10 %	3.173.969	4 %	3.130.544	4 %
Ø Sequenzlänge	70,5 bp		60,3 bp		40,3 bp		59,3 bp		55,5 bp	
RNA-Seq Kartiert	77.479.773	74 %	57.914.880	63 %	43.565.360	67 %	45.196.147	52 %	57.558.231	65 %
Nicht kartierbar	22.191.853	21 %	15.955.457	17 %	14.831.362	23 %	25.543.170	29 %	13.145.525	15 %

Auffällig ist, daß die vom „Cologne Center for Genomics“ durchgeführten Sequenzierungen der 2 X Amplifizierten RNA mit Werten von ca. 16 – 17 % eine große Datenmenge mit geringer Qualität aufweisen (Purity Filter).

3.6.2 PROZESSIEREN DER ROHDATEN UND ANNOTATION DER SEQUENZEN – BLASTn ABGLEICH MIT DER UG-DATENBANK DES RINDES

Der BLASTn Algorithmus benötigt eine wesentlich größere Rechenleistung als dies für Kartierungen der Fall ist. Die Analysen wurden daher auf dem Rechencluster LC2 der Universität Mainz parallelisiert durchgeführt. Ein Trimmen der Sequenzen fand nicht statt, alle Sequenzen wurden mit voller Länge verwendet. Simple repetitive Sequenzen wurden vom verwendeten Programmpaket „blastall“ (NCBI) erkannt und direkt entfernt. Komplexere repetitive Einheiten (z.B. SINES, LINEs usw.) wurden durch BLASTn Abgleich mit einer aus repetitiven Sequenzen bestehenden Datenbank identifiziert und anschließend entfernt. Das gleiche Verfahren wurde zum Entfernen von rRNA Sequenzen angewendet. Alle BLASTn Analysen und die Aufarbeitung der Ergebnisse wurden mit Hilfe von durch Benjamin Rieger geschriebenen Skripten durchgeführt (siehe Punkt 2.6.2). In Tabelle 3 - 9 sind die Ergebnisse der Aufreinigung aller Banken und der nachfolgenden BLASTn Analysen aufgelistet.

Tabelle 3 - 9 Prozessierung und BLASTn Analyse der Rohdaten

Gezeigt ist eine Übersicht der Prozessierung der Rohdaten bis zum BLASTn gegen den bovinen Teil der Unigene Datenbank des NCBI. Je Arbeitsschritt ist die Datenmenge in Sequenzen und in Prozent der Rohdaten angegeben.

Prozess	msc_neu		msc_alt		adipo		osteo		chondro	
Rohdaten	104.774.556		92.385.296		64.723.146		87.599.640		88.677.062	
≥ 5 "N"	320.788	0,31 %	1.510.484	1,63 %	28.243	0,04 %	1.444.478	1,65 %	1.471.034	1,66 %
Repetitiv	501.584	0,48 %	616.129	0,67 %	290.377	0,45 %	362.437	0,41 %	943.484	1,06 %
Ribosomale RNA	8.466.423	8,08 %	20.088.963	21,74 %	660.954	1,02 %	12.625.211	14,41 %	4.607.692	5,2 %
BLASTn Treffer	81.690.710	77,97 %	54.641.867	59,15 %	53.136.349	82,1 %	57.699.165	65,87 %	55.467.650	62,55 %
Nicht zuordenbar	13.795.051	13,17 %	15.527.853	16,81 %	10.607.223	16,39 %	15.468.349	17,66 %	26.187.202	29,53 %

Den größten Anteil an nicht Protein kodierenden Daten machen für ribosomale RNA kodierende Sequenzen aus. Die Spanne ist recht groß, sie reicht von 5,2 % in der Sequenzierung der Chondrozyten bis zu 21,74 % in den alten MSC. Sequenzen mit 5 oder mehr unbestimmbaren Basen („N“) sowie repetitive Sequenzen haben addiert mit maximal 2,72 % in der Sequenzierung der Chondrozyten einen geringen Anteil an den Daten, die nicht zum Abgleich mit der Unigene Datenbank heran gezogen werden. Dennoch fällt auf, daß Sequenzen mit 5 oder mehr als „N“

bestimmten Basen in den Daten des Kölner Anbieters (GAIIX Gerät) mit 1,63% bis 1,66% mindestens 5 mal häufiger vorkommen, als bei den im eigenen Haus durchgeführten Sequenzierungen (HiSeq2000 Gerät). Es konnten durch BLASTn Abgleich mit den Rindereinträgen der UG-Datenbank zwischen 59,15 % der Sequenzen der alten MSC und 82,1 % der Sequenzen der Adipozyten annotiert werden. Der Anteil an potentiell kodierenden Sequenzen, die keinem Cluster der UG – Datenbank zugeordnet werden konnten, liegt zwischen 13,17 % in der Bank der jungen MSC und 29,53 % in der Bank der Chondrozyten.

3.6.3 PROZESSIEREN DER ROHDATEN, ASSEMBLIERUNG UND ANNOTATION DER KONSENSUSEQUENZEN – BLASTN ABGLEICH MIT DER REFSEQ-RNA DATENBANK

Die zum Assemblieren verwendeten Sequenzen aus jeder Zelllinie sind mit den zuvor schon für die RNA-Seq Kartierung prozessierten Daten identisch (siehe Tabelle 3 - 8). Um nach der Assemblierung keine Contigs, die für rRNA oder repetitive Elemente kodieren, zu erhalten, wurden diese Sequenzen zuvor entfernt. Hierfür wurden die Daten gegen eine aus repetitiven Elementen und rRNA bestehende Datenbank mit der Software „Bowtie 1“ kartiert (siehe Punkt 2.6.3). Die Sequenzen, die nicht als Repeat oder rRNA kodierend aussortiert wurden, wurden mit Hilfe der Software „CLC Genomics Workbench“ assembliert. Die Konsensussequenzen wurden anschließend auf dem LC2 Rechencluster gegen die Refseq-RNA Datenbank des NCBI geblastet. Auch bei dieser BLASTn Analyse wurde nur der beste Datenbanktreffer, der mindestens einen E-Value von $1e^{-10}$ aufweist, als Annotation für den Contig angenommen. Die Ergebnisse der Aufreinigung, Assemblierung und der BLASTn Analyse sind in Tabelle 3 - 10 zusammengefasst. Es zeigt sich eine starke Diskrepanz zwischen den Ergebnissen der in Köln sequenzierten, 2 x amplifizierten RNA (Osteo, Chondro, alte MSC) und der im eigenen Haus sequenzierten, nicht amplifizierten RNA (Adipo, junge MSC). Während sich aus den Datensätzen aus Köln zwischen 18,88 % (Osteo) und 23,24 % (alte MSC) der Sequenzen assemblieren lassen, sind 61,17 % (junge MSC) bis 66,29 % (Adipo) der Sequenzen der im eigenen Haus erstellten Datensätze assemblierbar. Die Assemblierungen der Daten aus Adipozyten und jungen MSC weisen zudem mit 1184 bis 1350 bp die im Schnitt längsten Contigsequenzen auf. Die geringe Länge von durchschnittlich 40,3 bp der Sequenzen der Adipozyten-Bank hat keinen negativen Einfluss auf die Länge der daraus gebildeten Contigs. Die

durchschnittliche Länge der Contigs der Sequenzierung der alten MSC, Osteoblasten und Chondrozyten kommt über eine Länge von 785 bp nicht hinaus (alte MSC). Die Anzahl der gebildeten Contigs ist nicht mit der Menge der assemblierten Sequenzen korreliert. Die Sequenzen aus Chondrozyten weisen im Vergleich bei knapp der Hälfte an assemblierten Sequenzen ähnlich viele gebildete Contigs auf, wie die der Adipozyten. Durch BLASTn gegen die Refseq-RNA Datenbank konnten bis zu 92,21% der Contigs (Adipo) annotiert werden. Den niedrigsten prozentualen Anteil an durch die BLASTn Analyse annotierten Contigs weist mit 74% die Bank der jungen MSC auf, wobei dies mit 22.198 Contigs die meisten Sequenzen in absoluten Zahlen sind. Die BLASTn Treffer der übrigen Banken liegen zwischen 79,35 % (Chondro) und 88,37 % (Osteo) der gebildeten Contigs.

Tabelle 3 - 10 Prozessierung der getrimmten Rohdaten, Assemblierung und BLASTn

Als Ausgangsdatensatz werden dieselben Sequenzen wie zur RNA-Seq Analyse verwendet. Das Entfernen von repetitiven Sequenzen und rRNA erfolgt durch Kartierung, die Ergebnisse sind in Sequenzen und Prozent angegeben. Die übrigen Sequenzen wurden assembliert, die Contigs per BLASTn mit der Refseq-RNA Datenbank des NCBI abgeglichen.

Prozess	mssc_neu	mssc_alt	adipo	osteo	chondro
Sequenzen	99.671.626	73.870.337	58.396.722	70.739.317	70.703.756
Ø Sequenzlänge	70,5 bp	60,3 bp	40,3 bp	59,3 bp	55,5 bp
rRNA / Repetitiv (seq)	9.644.780 9,68 %	15.296.153 20,71 %	3.484.005 5,97 %	3.982.851 5,63 %	7.158.689 10,12 %
Assembliert (seq)	60.967.720 61,17 %	17.164.620 23,24 %	38.710.624 66,29 %	13.352.181 18,88 %	14.794.820 20,93 %
Nicht assembliert (seq)	29.059.126 29,15 %	41.409.564 56,06 %	16.202.093 27,74 %	53.404.285 75,49 %	48.750.247 68,95 %
Anzahl Contigs	29.996	21.505	17.219	14.867	17.036
Ø Contiglänge (bp)	1350	785	1.184	753	668
N50 (bp)	828	620	807	604	577
N20 (bp)	1.973	1.021	1.645	974	852
Contig Treffer BLASTn	22.198 74 %	17.193 79,95 %	15.878 92,21 %	13.138 88,37 %	13.518 79,35 %
Nicht zuordenbar	7798 26 %	4312 20,05 %	1341 7,79 %	1729 11,63 %	3518 20,65 %

3.6.4 GLOBALER VERGLEICH DER EXPRESSIONSPROFILE ALLER ZELLINIEN

Um die Genexpression der unterschiedlichen Zelllinien über das gesamte Transkriptom vergleichbar machen zu können, wurde nach Kartierung mittels der

Software CLC Genomics Workbench eine Heatmap erstellt (siehe Abbildung 3 - 13). Die Software verdeutlicht neben der Heatmap durch einen Stammbaum die Ähnlichkeit der Genexpressionsprofile zueinander. Stark exprimierte Gene werden in der Heatmap durch rote Färbung, weniger stark exprimierte durch blaue Färbung angezeigt. Aus der Farbverteilung geht hervor, daß der überwiegende Teil der kartierbaren Gene eher schwach exprimiert wird, die Blaufärbung überwiegt deutlich.

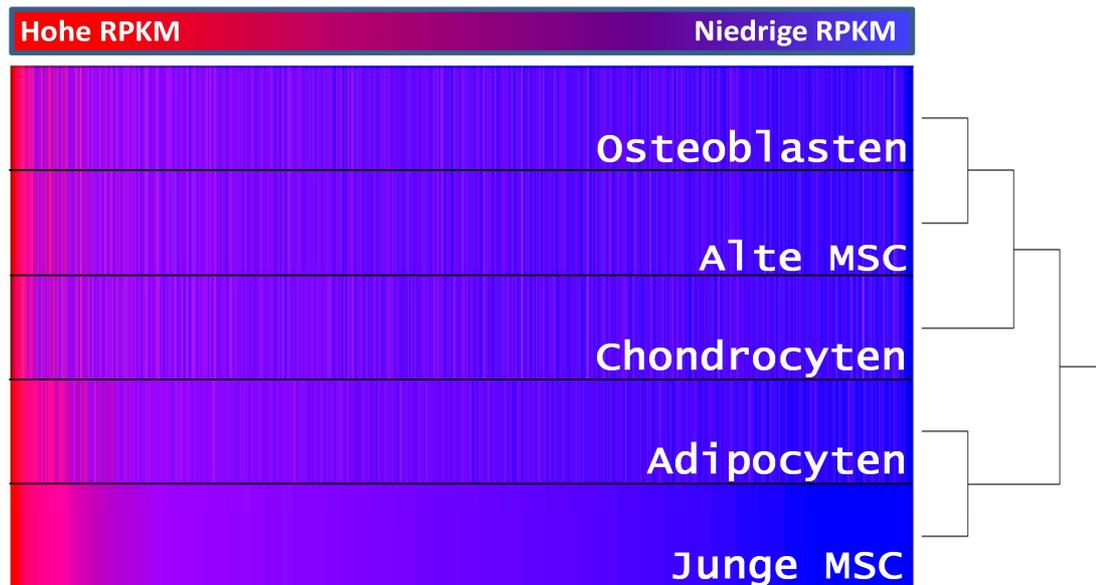


Abbildung 3 - 13 Heatmap der Genexpression nach Kartierung

Die Abbildung gibt für jede Zelllinie die Stärke der Genexpression durch Farbkodierung der RPKM-Werte jedes Gens wieder. Die Sortierung der Gene erfolgt von hohen RPKM-Werten (rot) zu niedrigen RPKM-Werten (blau). Referenz für die Sortierung ist die Sequenzierung der jungen MSC, die zu deren Genexpressionsprofil ähnlichsten Daten der Adipozyten werden direkt darüber abgebildet. Der rechte Teil der Abbildung zeigt das Ähnlichkeitsverhältnis der Genexpression aller Sequenzierungen als Stammbaum (Siehe Punkt 2.7.1).

Das Expressionsprofil junger MSC und das ausdifferenzierter Adipozyten sind sich vergleichsweise ähnlich. Das Expressionsprofil der alten MSC und das der Osteoblasten ähneln sich ebenfalls sehr stark. Einzig die Chondrozyten bilden eine eigene Sparte, die etwas ähnlicher zu den Osteoblasten und alten MSC, als zu den Adipozyten und jungen MSC ist. Auch im Streudiagramm lässt sich die in der Heatmap dargestellte Ähnlichkeit der Sequenzierungen verdeutlichen (Abbildung 3 - 14). Im Falle ähnlicher Expressionsprofile werden mehr Gene um die Mittellinie abgebildet, im Falle unähnlicher Expressionsprofile hingegen außerhalb der beiden äußeren Linien. Gene, deren Datenpunkt über oder unter den äußeren Begrenzungslinien abgebildet ist, werden um den Faktor 3 höher bzw. niedriger

exprimiert. Auffällig ist, daß die Menge dieser differenziell exprimierten Gene auch bei einander ähnlichen Expressionsprofilen ab RPKM-Werten von unter 1 (Skalierung im Graphen 1,00E+00) stark zunimmt. Im Streudiagramm der einander im Expressionsprofil unähnlichsten Zelllinien Osteoblasten und junge MSC liegt ein Großteil der Datenpunkte nicht im Bereich zwischen den äußeren beiden Linien. Der überwiegende Teil der Gene wird demnach stärker als um den Faktor drei differenziell exprimiert.

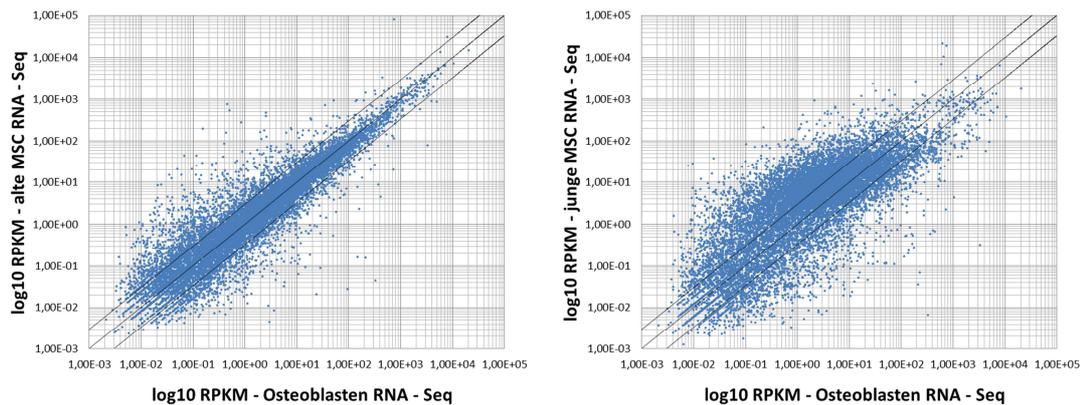


Abbildung 3 - 14 Streudiagramm ähnlicher und unähnlicher Genexpressionsprofile

Gezeigt sind die logarithmisch aufgetragenen RPKM Werte aller Gene der RNA – Seq Kartierung der Osteoblasten gegen eine vom Expressionsprofil her ähnliche (alte MSC) und unähnliche Zelllinie (junge MSC). Auf der mittleren der drei diagonal durch den Graphen verlaufenden Linien liegen alle Gene, deren RPKM Werte in beiden Sequenzierungen identisch sind. Die äußeren Linien markieren die Grenze zu einem Expressionsunterschied um den Faktor drei.

3.7 READ-COUNT BASIERTE ANALYSE DER GENEXPRESSION MITTELS RNA-SEQ

Um die Ergebnisse der Genexpressionsanalyse auf Basis von „Read Counts“ zur Bestimmung der Expressionsstärke einzelner Gene in den Zelllinien funktionell einordnen zu können, muss die quantitative Information in einen biologischen Kontext gebracht werden. Der Vergleich der drei unterschiedlichen Wege zur Bestimmung der Expressionsstärke zeigt, daß prinzipiell alle Methoden vergleichbare Ergebnisse liefern (siehe Abbildung 3 - 12 und Tabelle 3 - 11). Aufgrund der Redundanz der Daten ist eine weitere parallele Analyse der Listen aus den drei genutzten Methoden nicht immer zweckmäßig. Die Ergebnisse der Kartierung durch die Software CLC Genomics Workbench werden im Folgenden dort als Referenzdatensatz stellvertretend weiter analysiert, wo redundante Analysen keinen erkennbaren Mehrwert bieten. Je nach Fragestellung eignen sich

zur weiteren Analyse der Daten unterschiedliche computergestützte Analysewerkzeuge wie DAVID oder der KEGG Pathway Mapper, sowie die Betrachtung der Expression zellinientypisch exprimierter Markergene.

3.7.1 EXPRESSION VON ZELLINIENTYPISCH EXPRIMIERTEN MARKERGENEN

Der Vergleich zellinientypisch exprimierter Markergene dient der Einschätzung des Erfolgs der Differenzierungsexperimente. Auch die Expression von Markergenen für undifferenzierte MSC kann so ergänzend zur RT-PCR überprüft werden (vergleiche Punkt 3.2.2). Die Expression der Markergene ist in Tabelle 3 - 11 zusammengefasst. Für jedes Gen sind aufgrund der unterschiedlichen Methoden zur Bestimmung der Expressionsstärke drei Werte angegeben. Es zeigt sich hierbei, daß die Werte für die Kartierung und die BLASTn Analyse der Einzelsequenzen meist dicht beieinander liegen. Wenn Abweichungen zwischen den drei Methoden zu beobachten sind, betrifft dies meist die Expressionswerte, welche durch Assemblierung und BLASTn Suche in der Refseq-RNA Datenbank bestimmt wurden. Für die Expression von für Adipozyten typischen Markergenen werden in der adipogenen Differenzierung mehrheitlich die höchsten Werte bestimmt. Die Gene *adipoq* (Adiponectin), *acam* (Adipocyte adhesion molecule) und *cebpa* (CCAAT/enhancer-binding protein alpha) weisen die stärkste Expression im Vergleich aller Banken in den Adipozyten auf. Für die Adipozyten-Marker *fabp4* (Fatty acid-binding protein 4, adipocyte) und *pparg* (Peroxisome proliferator-activated receptor gamma) konnten die höchsten Expressionswerte durch genomische Kartierung und BLASTn Abgleich mit der Unigene Datenbank ebenfalls für Adipozyten ermittelt werden. Die Expressionswerte welche durch Assemblierung und BLASTn Abgleich mit der Refseq-RNA Datenbank bestimmt wurden, sind hingegen für *fabp4* und *pparg* in Osteoblasten leicht höher als in Adipozyten. Die osteogenen Marker *ibsp* (Bone sialoprotein II) und *omd* (Osteomodulin) werden erwartungsgemäß am stärksten in Osteoblasten exprimiert.

Tabelle 3 - 11 Expression von zelllinientypischen Markergenen

Farbig hinterlegt ist die höchste Expression, wobei die Werte adipogener Markergene in Gelb, osteogener in Orange, chondrogener in Blau, und jene für undifferenzierte MSC in Lila dargestellt sind. CLC – Kartierung der Sequenzen gegen annotierte Bereiche des Genoms (RPKM Werte). UG – BLASTn der Sequenzen gegen die bovinen Einträge der Unigene Datenbank des NCBI (RPKM Werte). REF – BLASTn der Konsensussequenzen der Contigs gegen die Refseq-RNA Datenbank des NCBI (RPM Werte). NA – Kein Contig mit BLASTn Treffer für das Gen vorhanden, bzw. das Gen war nicht in der Referenz (Genom/Unigene) annotiert.

Marker	Adipozyten			Osteoblasten			Chondrozyten			MSC_neu			MSC_alt		
	CLC	UG	REF	CLC	UG	REF	CLC	UG	REF	CLC	UG	REF	CLC	UG	REF
Adipozyten Markergene															
<i>fabp4</i>	610	98	456	280	68	473	14	6	18	2	1	1	193	65	245
<i>pparg</i>	14	18	31	9	8	52	0	3	NA	9	13	20	4	7	19
<i>adipoq</i>	21	17	57	5	6	34	0	2	NA	0	4	NA	0	3	NA
<i>acam</i>	132	111	423	8	7	101	4	5	51	24	21	82	16	17	202
<i>cebpa</i>	11	11	97	1	1	NA	3	1	9	9	5	25	2	1	11
Osteoblasten Markergene															
Marker	CLC	UG	REF	CLC	UG	REF	CLC	UG	REF	CLC	UG	REF	CLC	UG	REF
<i>ibsp</i>	1	1	NA	47	52	197	0	0	NA	0	0	NA	12	29	70
<i>bglap</i>	0	0	NA	3375	1336	NA	2	1	NA	0	0	NA	78	46	104
<i>runx2</i>	29	NA	36	28	NA	312	4	NA	17	22	NA	17	15	NA	180
<i>spp1</i>	26	18	39	2959	1903	371	92	73	155	106	85	170	2683	2337	158
<i>omd</i>	11	55	208	743	702	300	0	0	NA	2	8	26	19	47	270
Chondrozyten Markergene															
Marker	CLC	UG	REF	CLC	UG	REF	CLC	UG	REF	CLC	UG	REF	CLC	UG	REF
<i>chad</i>	0	1	NA	8	14	87	0	0	NA	0	1	NA	0	1	NA
<i>col2a1</i>	4	0	NA	0	0	NA	0	0	NA	4	0	NA	0	0	NA
<i>comp</i>	29	27	NA	1	0	NA	0	0	NA	1	1	NA	4	4	7
<i>acan</i>	0	0	NA	0	0	NA	0	0	NA	2	2	18	0	0	NA
<i>hapln1</i>	32	28	123	48	42	403	1	1	7	0	0	NA	12	14	176
MSC Markergene															
Marker	CLC	UG	REF	CLC	UG	REF	CLC	UG	REF	CLC	UG	REF	CLC	UG	REF
<i>cd44</i>	21	15	37	86	55	386	57	83	437	215	175	432	56	58	285
<i>cd45</i>	0	0	NA	0	0	NA	0	0	NA	0	0	NA	4	5	46
<i>lif</i>	0	0	NA	0	0	NA	0	1	NA	1	1	2	0	0	NA
<i>myc</i>	26	24	47	12	11	NA	10	12	37	24	24	65	11	14	34
<i>kitlg</i>	20	40	101	1	2	3	0	1	NA	2	4	10	1	1	9
<i>nanog</i>	0	0	0	0	0	NA	0	0	NA	0	0	NA	0	0	NA
<i>oct4</i>	0	0	0	0	0	NA	0	0	NA	0	0	NA	0	0	NA
<i>klf4</i>	25	23	64	5	4	34	43	48	97	20	20	62	25	29	169
<i>fgf2</i>	9	19	136	0	79	NA	1	72	75	0	2	6	0	47	39
<i>sox2</i>	0	0	0	0	0	NA	0	0	NA	0	0	NA	0	0	NA
<i>cd105</i>	80	80	281	19	17	95	20	23	155	63	64	220	27	30	200
<i>yy1</i>	28	24	110	45	37	285	31	35	241	22	23	53	33	42	166

Das Gen *bglap* (Osteocalcin) weist die mit Abstand stärksten RPKM Werte in der Sequenzierung der Osteoblasten auf, auffällig ist hier das Fehlen eines Contigs der dem Gen per BLASTn Abgleich mit der RefSeq-RNA Datenbank zugewiesen werden kann. Der Transkriptionsfaktor *runx2* (Runt-related transcription factor 2) zeigt nach Kartierung in allen Banken eine eher geringe Expression, Adipozyten und Osteoblasten weisen hier ähnlich hohe Expressionswerte auf. In der Osteoblasten-Bank konnten jedoch mit Abstand am meisten Sequenzen zu einem dem Gen zuordenbaren Contig assembliert werden. Das Gen *spp1* (Osteopontin) weist neben einer sehr starken Expression in Osteoblasten eine ähnliche hohe Expression in den alten MSC auf. Die Analyse der Genexpression chondrogener Marker weist deutlich auf eine nicht gelungene chondrogene Differenzierung hin. Als einziges der untersuchten Markergene wird *hapln1* (Hyaluronan and proteoglycan link protein 1, auch *ctrl1* Cartilage-linking protein genannt) in den chondrogen differenzierten Zellen exprimiert. Die Expression ist allerdings mit Werten zwischen 1 (RPKM) und 7 (RPM) allenfalls als basal einzustufen. Für keines der übrigen vier Markergene konnte eine Expression in Chondrozyten gemessen werden, obwohl alle Gene als wichtige Bestandteile der extrazellulären Matrix von Knorpelgewebe gelten. Die Gene *chad* (Chondroadherin) und *hapln1* werden am stärksten in Osteoblasten exprimiert. Die Gene *col2a1* (Alpha-1 type II collagen), *comp* (Cartilage oligomeric matrix protein) und *acan* (Aggrecan core protein) können vornehmlich in den Banken der Adipozyten und jungen MSC nachgewiesen werden. Als molekulare Marker für die undifferenzierten Stammzellen wurden die Expressionswerte der Gene herangezogen, die zuvor schon zur Charakterisierung der Primärzellen per RT-PCR dienten (siehe Punkt 3.2.2). Es zeigt sich, daß mit dem Gen *lif* nur einer der überprüften Marker die höchste Expressionsstärke in den jungen MSC aufweist. Die Expression ist mit RPKM/RPM Werten von 1 bis 2 jedoch sehr schwach. Ebenfalls schwache Expressionswerte wurden in den jungen MSC für die Gene *fgf2* und *kitlg* bestimmt. Lediglich das Oberflächenantigen CD44 kann als stark exprimierter Marker für frische Primärkulturen von MSC gelten, auch wenn hier eine minimal höhere Expression der assemblierten Daten in den Chondrozyten gemessen werden konnte. Die Expression der vor allem als embryonale Stammzellmarker bekannten Gene *nanog*, *oct4* und *sox2* kann in keiner der Banken nachgewiesen werden. Für die Gene *klf4*, *cd105*, *myc* und *yy1* werden in allen Banken mittlere Expressionsstärken bestimmt, ohne daß sich aus den Werten eine besondere Funktion als Marker für MSC interpretieren ließe. Das von MSC laut Literatur nicht exprimierte Oberflächenantigen CD45 wird einzig in den alten MSC

exprimiert, auch wenn die Expression mit RPKM – Werten von 4 - 5 nicht besonders stark ist. Der RPM Wert von 46, der für die BLASTn Suche der assemblierten Daten in den alten MSC ermittelt wurde, weicht von den RPKM Werten der Sequenzkartierung und des BLASTn gegen die UG Datenbank nach oben ab. Insgesamt sind starke Expressionsunterschiede zwischen den jungen und alten MSC zu beobachten. Die Expressionswerte der Markergene nach Sequenzierung lassen sich mit den Ergebnissen der RT-PCR gut vergleichen. Die schwach oder überhaupt nicht exprimierten Gene *lif*, *oct4*, *sox2*, *cd45* und *nanog* ließen sich auch in den RT-PCR Experimenten nicht oder nicht sicher und eindeutig nachweisen.

3.7.2 FUNKTIONELLE ANNOTATION STARK EXPRIMIERTER GENE – JUNGE MSC

Aus der funktionellen Annotation der 100 am stärksten exprimierten Gene lassen sich für jede Zelllinie Rückschlüsse über die transkriptionsaktivsten biologischen Prozesse ziehen. Die Messung der Expressionsstärke eines Gens erfolgt durch Summierung aller Einzelsequenzen, die nach CLC RNA-Seq Kartierung den als Exons eines Gens annotierten DNA Abschnitt zugeordnet werden können (Total Exon Reads, TER). Die Annotation der Genlisten erfolgt mit Hilfe der „Functional Annotation Clustering“ (FAC) Funktion des online Tools DAVID [86, 87]. Die 100 höchstexprimierten Gene werden mit den Gen-Ontologie Datenbanken „*biological process*“, „*molecular function*“ und „*swissprot protein interaction resource keywords*“ abgeglichen. Funktionell ähnliche, statistisch signifikante Treffer aus den drei Datenbanken werden automatisch gruppiert und mit einem sog. „Enrichment Score“ versehen. Ein Gen kann hierbei mehreren Ontologien zugeordnet werden. Je größer der „Enrichment Score“ ist, desto stärker ist die statistische Anreicherung der Gene einer Gruppierung. Jede Gruppe wird manuell mit einem zur Funktion der Datenbanktreffer passenden Oberbegriff versehen. Die 12 Gruppen mit dem höchsten „Enrichment Score“, sowie die 25 am stärksten exprimierten Gene werden im Folgenden abgebildet. Die kompletten Listen der Annotation können im digitalen Anhang D5.2 eingesehen werden. Unter den 100 am stärksten exprimierten Genen werden bislang uncharakterisierte Transkripte manuell annotiert und auf eine mögliche Funktion im Hinblick auf die jeweilige Differenzierung untersucht. Dabei lässt sich auffällig häufig feststellen, daß die hypothetischen Genloci in der ständig aktualisierten NCBI Datenbanken ihre Annotation seitens des NCBI wieder entzogen bekommen haben.

Tabelle 3 - 12 Stark exprimierte Gene - Top 25 der Sequenzierung des Transkriptoms junger MSC

Tabelle 3 - 12 listet die 25 höchstexprimierten Gene auf. Die Spalte „Ontologie“ weist ein Gen als zur entsprechenden Gruppe in Abbildung 3 - 15 zugehörig aus. Durch manuelle Annotation uncharakterisierter Genloci innerhalb der ersten 100 Listenplätze ließen sich keine potentiell interessanten Kandidatengene ermitteln.

Pos.	TER	Symbol	Name	Ontologie Gruppe 1-12
1	7871616	COL1A1	collagen, type I, alpha 1	①③④⑦⑧⑨⑫
2	3721409	COL1A2	collagen, type I, alpha 2	①④⑧⑨
3	1846410	LOC100335409	Annotation zurückgezogen (NCBI)	--Nichtgruppiert--
4	1158627	SPARC	osteonectin	①⑥⑨
5	618393	EEF1A1	eukar. transl. elongation factor 1α1	③⑪
6	547345	SERPINE1	serpin peptidase inhibitor 1	①⑨⑩
7	544949	TNC	tenascin C	⑨
8	503487	ACTG1	Actin, gamma 1	②④⑨
9	503102	THBS1	thrombospondin 1	①③④⑤⑥⑨⑩⑫
10	492265	LOC100335677	Annotation zurückgezogen (NCBI)	--Nichtgruppiert--
11	463000	FLNA	filamin A, alpha	--Nichtgruppiert--
12	447582	VIM	vimentin	⑧
13	435111	LOC100140226	zinc finger protein 160-like	--Nichtgruppiert--
14	424701	FN1	fibronectin 1	①④⑤⑨⑩
15	392386	COL6A1	collagen, type VI, alpha 1	--Nichtgruppiert--
16	378498	LOC781039	BLASTn 1e-75 RPI7 <i>R. norvegicus</i>	--Nichtgruppiert--
17	371049	FBN1	fibrillin 1	①⑥⑧
18	304152	ACTA2	actin, alpha 2	②④⑧⑨
19	293923	MMP2	matrix metalloproteinase 2	①③④⑥⑦⑧⑨
20	271113	COL6A2	collagen, type VI, alpha 2	⑧
21	245924	SERPINH1	serpin peptidase inhibitor 1, (HSP47)	①③⑦⑧
22	229628	TAGLN	transgelin	②④⑨
23	221983	EEF2	eukar. transl. elongation factor 2	③⑪
24	201222	MYH9	myosin, heavy chain 9, non-muscle	③④⑨⑩
25	195008	FBLN2	fibulin 2	--Nichtgruppiert--

Funktionell gruppiert zeigt sich für die jungen MSC eine starke Zugehörigkeit der 100 am stärksten exprimierten Gene zu Prozessen, die mit der extrazellulären Matrix in Verbindung stehen (Abbildung 3 - 15, Ontologie 1, 5, 7, 8, 12). Auch mit der Funktion von Stammzellen zusammenhängende Gengruppen konnten identifiziert werden (Ontologie 4, 9, 10). Einen weiteren übergeordneten Funktionsbereich stellt die Gruppe von mit Transkriptionsprozessen assoziierbaren Genen dar (Ontologie 3, 11).

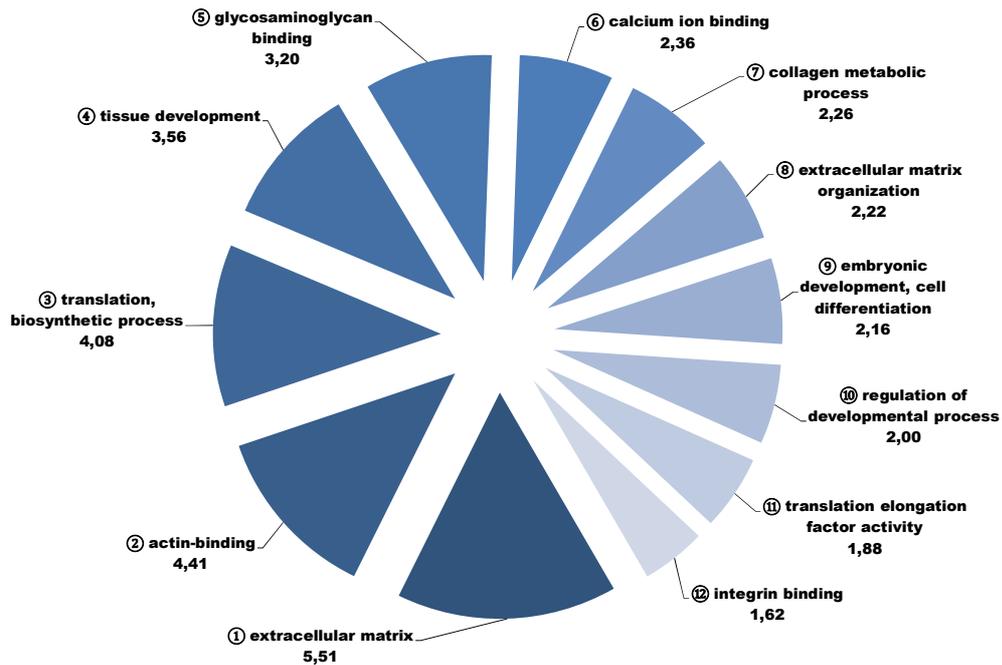


Abbildung 3 - 15 Ontologie der 100 höchstexprimierten Gene – junge MSC

Gezeigt werden die 12 im DAVID FAC am stärksten angereicherten Ontologie-Gruppen. Die Zahlen im Anschluss an den Gruppennamen geben den „Enrichment Score“ als Maß der Anreicherung an.

3.7.3 FUNKTIONELLE ANNOTATION STARK EXPRIMIERTER GENE – ADIPOGENESE

Die Liste der in den Adipozyten stark exprimierten Gene deckt sich auf den ersten 100 Listenplätzen zu 61%, für die in Tabelle 3 - 13 abgebildeten 25 Genen zu 48% (12 Gene) mit den in den jungen MSC stark exprimierten Genen. Diese hohe Ähnlichkeit der Listen steht im Einklang mit den unter Punkt 3.6.4 beschriebenen Ergebnissen des Vergleichs der globalen Genexpressionsprofile aller Zelllinien. Auch in der funktionellen Gruppierung spiegelt sich der hohe Anteil identischer Gene wider. Potentiell interessante, bislang nicht charakterisierte Genloci finden sich in den 100 am stärksten exprimierten Genen nach manueller Annotation keine. Die funktionelle Annotation der 100 am stärksten exprimierten Gene umfasst auch in der Sequenzierung der Adipozyten vornehmlich mit dem Oberbegriff der extrazellulären Matrix assoziierbare Termini (Tabelle 3 - 13, Ontologie 1, 2, 3, 5, 6, 12).

Tabelle 3 - 13 Stark exprimierte Gene - Top 25 der Sequenzierung des Transkriptoms der Adipozyten

Tabelle 3 - 13 listet die 25 höchstexprimierten Gene auf. Die Spalte „Ontologie“ weist ein Gen als zur entsprechenden Gruppe in Abbildung 3 - 16 zugehörig aus. Durch manuelle Annotation uncharakterisierter Genloci innerhalb der ersten 100 Listenplätze ließen sich keine potentiell interessanten Kandidatengene ermitteln.

Pos.	TER	Symbol	Name	Ontologie Gruppe 1-12
1	1865679	COL1A1	collagen, type I, alpha 1	① ② ④ ⑤ ⑦ ⑧ ⑨ ⑩ ⑫
2	1424350	COL1A2	collagen, type I, alpha 2	① ② ⑤ ⑦ ⑧ ⑨ ⑩ ⑫
3	1360372	FN1	fibronectin 1	② ③
4	804252	SPARC	osteonectin	②
5	412272	VIM	vimentin	④
6	333259	EEF1A1	eukar. transl. elongation factor 1α1	④
7	256831	FBN1	fibrillin 1	② ⑫
8	251443	MMP2	matrix metalloproteinase 2	② ④ ⑤ ⑧ ⑨ ⑩ ⑫
9	197365	CRISPLD2	Cysteine-rich secretory protein 11	① ② ③ ⑧
10	173782	COL6A1	collagen, type VI, alpha 1	--Nichtgruppiert--
11	162566	EEF2	eukar. transl. elongation factor 2	④
12	137325	RPLP0	BLOCK 23; ribos. protein, large, P0	④
13	122133	DCN	decorin	② ③ ④ ⑥ ⑪
14	119637	SERPINE2	serpin peptidase inhibitor 2	--Nichtgruppiert--
15	112913	FTH1	ferritin, heavy polypeptide 1	--Nichtgruppiert--
16	104765	LUM	lumican	② ⑥
17	104480	FMOD	fibromodulin	② ⑥
18	103639	AHNAK	AHNAK nucleoprotein	--Nichtgruppiert--
19	102975	TPT1	tumor prot., transl.-controlled 1	④
20	99530	MFGE8	lactadherin	②
21	98560	PKM2	pyruvate kinase, muscle	④ ⑪
22	97024	GJA1	gap junction alpha 1 protein	② ⑦ ⑧ ⑨ ⑩ ⑫
23	94652	COL6A2	collagen, type VI, alpha 2	⑫
24	93915	FLNA	filamin A, alpha	--Nichtgruppiert--
25	92228	FSTL1	follistatin-like 1	② ③

Gene mit Bezug zu den Ontologien Transkription und Energiestoffwechsel werden jeweils einmal (4, 11), mit Entwicklungsprozessen assoziierbare Ontologien (7 bis 10) viermal einem Cluster zugeordnet.

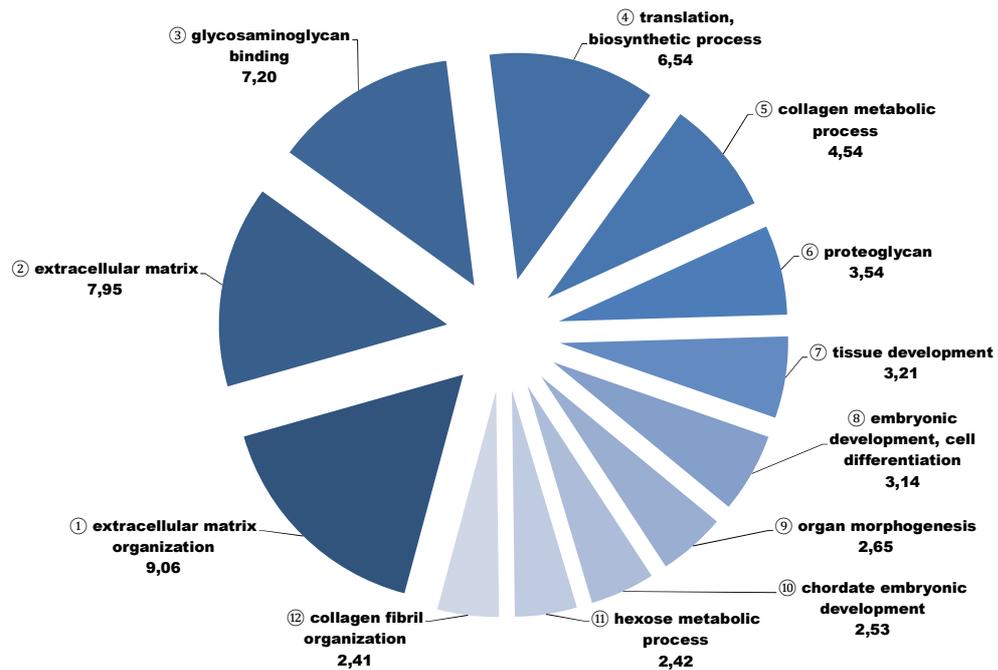


Abbildung 3 - 16 Ontologie der 100 höchstexprimierten Gene – Adipogenese

Gezeigt werden die 12 im DAVID FAC am stärksten angereicherten Ontologie-Gruppen. Die Zahlen im Anschluss an den Gruppennamen geben den „Enrichment Score“ als Maß der Anreicherung an.

3.7.4 FUNKTIONELLE ANNOTATION STARK EXPRIMIERTER GENE – OSTEOGENESE

Unter den 100 am stärksten exprimierten Genen der Osteoblasten befinden sich 44 Gene kodierend für ribosomale Proteine. Die osteogen differenzierten Zellen erscheinen somit sehr stark translationsaktiv, dementsprechend ist die mit Abstand am stärksten angereicherte Ontologie im FAC die der Transkription (siehe Abbildung 3 - 17, Ontologie 1, 2). Wie zuvor bei den jungen MSC und Adipozyten beobachtet, finden sich erneut viele Gene (z.B. Kollagengene) mit Bezug zur extrazellulären Matrix unter den 100 am häufigsten exprimierten Transkripten (Ontologien 3, 4, 11). Die Überschneidung der 100 am stärksten exprimierten Transkripte der Osteoblasten im Vergleich zu denen der jungen MSC liegt mit 28 identischen Genen jedoch weit unter der zwischen MSC und Adipozyten beobachteten Ähnlichkeit. Unter den 25 am stärksten exprimierten Transkripten befindet sich mit 177.976 TER auf dem elften Platz das Gen *spp1*, ein wichtiger Marker für die Biomineralisierung. Dies deckt sich mit den Messungen der Microarray-Experimente und der Analyse der Markergene (vergleiche Punkte 3.5.2 u. 3.7.1).

Tabelle 3 - 14 Stark exprimierte Gene - Top 25 der Sequenzierung des Transkriptoms der Osteoblasten

Tabelle 3 - 14 listet zusätzlich zu den 25 höchstexprimierten Genen ein uncharakterisiertes Protein aus den ersten 100 Listenplätzen auf. Die Spalte „Ontologie“ weist ein Gen als zur entsprechenden Gruppe in Abbildung 3 - 17 zugehörig aus.

Pos.	TER	Symbol	Name	Ontologie Gruppe 1-12
1	958164	LOC781039	BLASTn 1e-75 RPI7 R. norvegicus	--Nichtgruppiert--
2	652047	FSTL1	follistatin-like 1	④ ⑪ ⑫
3	359835	LOC100335677	Annotation zurückgezogen (NCBI)	--Nichtgruppiert--
4	329683	ANXA1	annexin A1	① ③ ④ ⑦ ⑨ ⑫
5	289782	LUM	lumican	④ ⑦
6	288166	EEF1A1	eukar. transl. elongation factor 1 α 1	①
7	262534	PDLIM4	PDZ and LIM domain protein 4	--Nichtgruppiert--
8	247729	LOC783718	ribosomal protein S27a pseudogene	--Nichtgruppiert--
9	244890	VIM	vimentin	--Nichtgruppiert--
10	207434	RPS3A	40S ribosomal protein S3a	①
11	177976	SPP1	secreted phosphoprotein 1	④ ⑤ ⑦ ⑨ ⑩
12	176064	LOC100140226	zinc finger protein 160-like	--Nichtgruppiert--
13	169651	RPLP0	BLOCK 23; ribosomal protein, large, P0	①
14	168011	DCN	decorin	① ④ ⑪
15	167296	RPS18	ribosomal protein S18	① ②
16	165968	RPSA	ribosomal protein SA	① ⑦
17	160225	RPS15	ribosomal protein S15	①
18	153690	ACTG1	Actin, gamma 1	⑦ ⑧
19	150960	RPS2	ribosomal protein S2	① ②
20	147774	RPS3	ribosomal protein S3	① ②
21	143050	CST3	cystatin C	③ ④ ⑥ ⑦
22	141784	RPS29	ribosomal protein S29	①
23	138857	COL1A2	collagen, type I, alpha 2	④ ⑦ ⑨ ⑩
24	136808	LOC787914	ribosomal protein S6	--Nichtgruppiert--
25	134549	COL1A1	collagen, type I, alpha 1	① ④ ⑤ ⑦ ⑨ ⑩
96	42177	C29H11orf10	chromosome 29 orf, human C11orf10	--Nichtgruppiert--

Im FAC der 100 am stärksten exprimierten Gene finden sich Ontologien wie „*biomineralization*“ und „*ossification*“, die auf den osteogenen Charakter der Zellen schließen lassen (Ontologien 7, 10, 12). Des Weiteren sind Ontologien angereichert, die auf der Expression von Strukturproteinen der Genfamilie der Actine beruhen (Ontologien 6, 8). Mit der Ontologie „*tissue development*“ (Ontologie 9) ist eine direkt mit Differenzierungsprozessen assoziierbare Gruppierung vertreten. Durch die große Anzahl ribosomaler Proteine in den 100 am stärksten exprimierten Genen

sind die „Enrichment Scores“ der anderen Ontologien im Vergleich zu denen der jungen MSC und Adipozyten wesentlich erniedrigt.

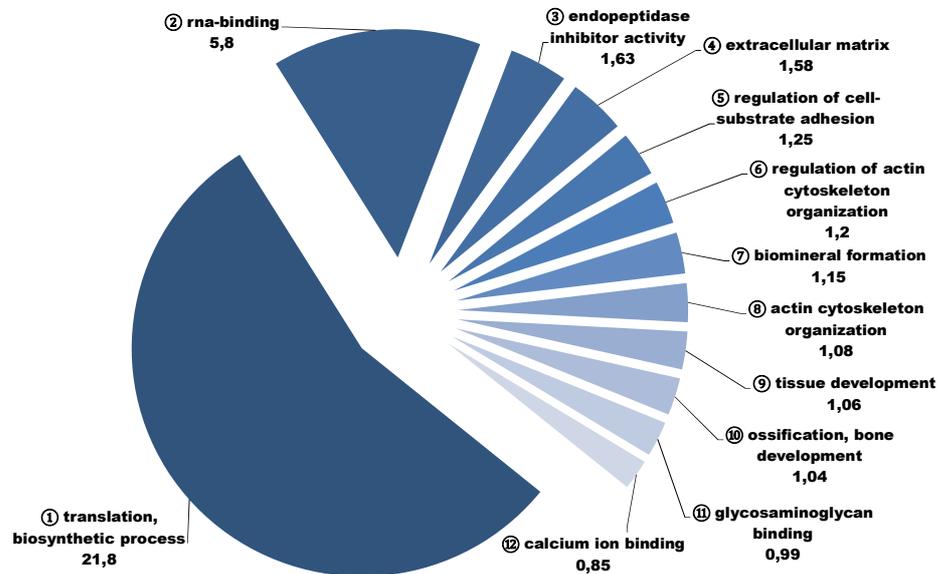


Abbildung 3 - 17 Ontologie der 100 höchstexprimierten Gene – Osteogenese

Gezeigt werden die 12 im DAVID FAC am stärksten angereicherten Ontologie-Gruppen. Die Zahlen im Anschluss an den Gruppennamen geben den „Enrichment Score“ als Maß der Anreicherung an.

Unter den nicht näher charakterisierten Transkripten befindet sich das Gen *C29H11orf10* (oder *TMEM258*) mit 42177 Sequenzen an Position 96. Aufgrund der verhältnismäßig starken Expression des bislang unvollständig beschriebenen Transkriptes erfolgt eine nähere Charakterisierung, um eventuell Rückschlüsse auf eine Funktion des Proteins in der Osteogenese ziehen zu können. Hierfür wird das Protein *C29H11orf10* auf funktionelle Domänen hin untersucht (siehe Abbildung 3 - 18).

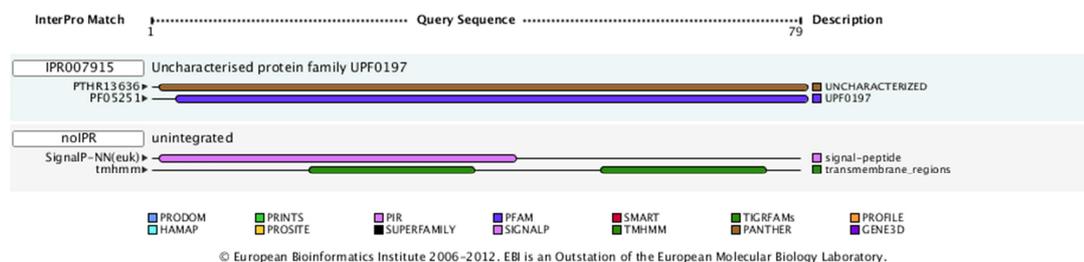


Abbildung 3 - 18 Interpro Analyse der Proteinsequenz des Gens *C29H11orf10*

Für den abgebildeten Interpro-Scan wurde die bovine Proteinsequenz NP_001181992.1 verwendet.

Das Protein von *C29H11orf10* umfasst 79 Aminosäuren, es gehört zur Familie UPF0197 (*uncharacterized protein family*) und enthält eine Multipass-Transmembran-Domäne und ein Singalpeptid.

3.7.5 FUNKTIONELLE ANNOTATION STARK EXPRIMIERTER GENE – ALTE MSC

Die stark exprimierten Gene der alten MSC sind auf den ersten 100 Listenplätzen zu 73% mit denen der Osteoblasten identisch, auf den ersten 25 Plätzen kommen 19 Gene in beiden Listen vor.

Tabelle 3 - 15 Stark exprimierte Gene - Top 25 der Sequenzierung des Transkriptoms der alten MSC

Tabelle 3 - 15 listet die 25 höchstexprimierten Gene auf. Die Spalte „Ontologie“ weist ein Gen als zur entsprechenden Gruppe in Abbildung 3 - 19 zugehörig aus. Durch manuelle Annotation uncharakterisierter Genloci innerhalb der ersten 100 Listenplätze ließen sich keine potentiell interessanten Kandidatengene ermitteln.

Pos.	TER	Symbol	Name	Ontologie Gruppe 1-12
1	5847314	LOC100335409	Annotation zurückgezogen (NCBI)	--Nicht gruppiert--
2	2853793	LOC781039	BLASTn 1e-75 RPI7 R. norvegicus	--Nicht gruppiert--
3	1765796	LOC100335677	Annotation zurückgezogen (NCBI)	--Nicht gruppiert--
4	405081	EEF1A1	eukar. transl. elongation factor 1 α 1	①
5	376504	FTH1	ferritin, heavy polypeptide 1	⑨ ⑩
6	333546	LOC100140226	zinc finger protein 160-like	--Nicht gruppiert--
7	292444	VIM	vimentin	①
8	278232	RPS3A	similar to 40S ribosomal protein S3a	①
9	273574	DCN	decorin	① ③ ④
10	257353	ANXA1	annexin A1; similar to Annexin A1	① ③ ⑥ ⑦ ⑧
11	226451	LOC783718	ribosomal protein S27a pseudogene	--Nicht gruppiert--
12	206829	SPP1	secreted phosphoprotein 1	③ ⑥ ⑦ ⑫
13	201392	RPS18	ribosomal protein S18	① ②
14	201230	RPSA	ribosomal protein SA	① ⑥
15	199712	COL1A1	collagen, type I, alpha 1	① ③ ⑥ ⑦ ⑨
16	190496	RPS3	ribosomal protein S3	① ② ⑨
17	183855	FSTL1	follistatin-like 1	③ ④
18	180095	FN1	fibronectin 1	③ ④ ⑤ ⑨ ⑩ ⑫
19	178317	RPS15	ribosomal protein S15	①
20	170559	LUM	lumican	③
21	162782	MFGE8	milk fat globule-EGF factor 8 protein	③ ⑤ ⑨ ⑫
22	150372	TIMP1	TIMP metalloproteinase inhibitor 1	③ ⑥ ⑧ ⑩
23	148437	COL1A2	collagen, type I, alpha 2	③ ⑥ ⑦ ⑩
24	147708	CCL2	chemokine (C-C motif) ligand 2	③ ⑨ ⑫
25	147005	RPLP0	BLOCK 23; ribosomal protein, large, P0	①

Der Anteil ribosomaler Proteine unter den 100 am stärksten exprimierten Genen ist mit 49 Genen größer als dies in der Sequenzierung der Osteoblasten (44) zu beobachten ist. Auch diese Beobachtung deckt sich mit dem unter Punkt 3.6.4 beschriebenen globalen Expressionsvergleich. Das osteogene Markergen *spp1* wird in den alten MSC wie in den Osteoblasten stark exprimiert. Die Ergebnisse der funktionellen Annotation der 100 am stärksten exprimierten Gene der alten MSC sind denen der Osteoblasten aufgrund der hohen Übereinstimmung der stark exprimierten Gene ähnlich. Auch im FAC der alten MSC ist die Ontologie der „Transkription“ mit Abstand am stärksten angereichert (Ontologie 1, 2), gefolgt von den mit der „extrazellulären Matrix“ assoziierbaren Ontologien (3, 4, 8). Die Ontologie „ossification“ ist im FAC der alten MSC ebenfalls angereichert (Ontologie 6), „biomineral formation“ ist als separate Ontologie hingegen nicht gruppiert. Auf Differenzierungsprozesse und deren Regulierung deuten die Ontologien 7, 9 und 12 hin, während die Ontologien 10 und 11 nicht zelllinienspezifischen Grundprozesse zuzuordnen sind.

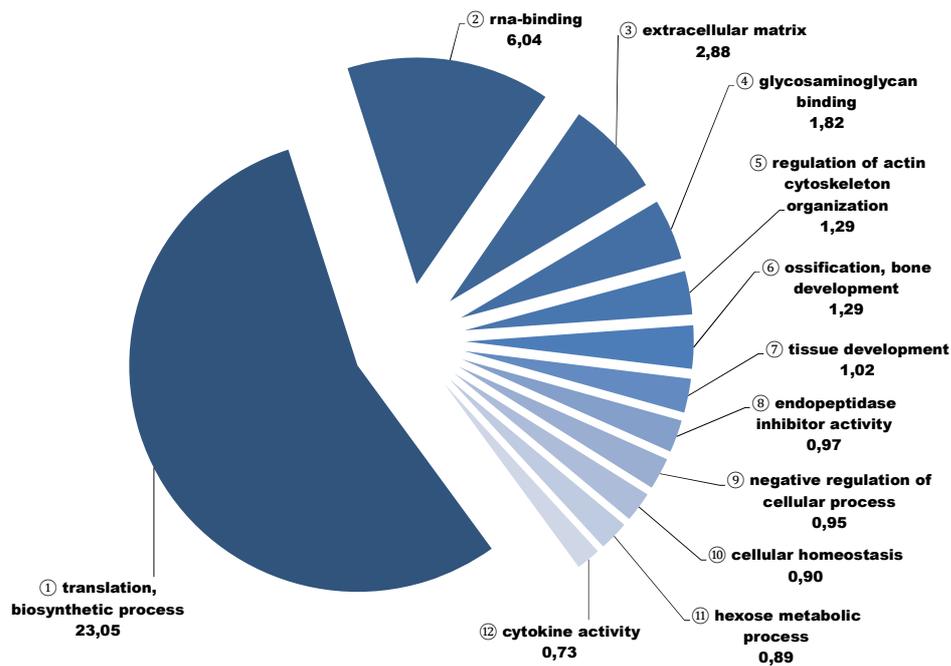


Abbildung 3 - 19 Ontologie der 100 höchstexprimierten Gene – alte MSC

Gezeigt werden die 12 im DAVID FAC am stärksten angereicherten Ontologie-Gruppen. Die Zahlen im Anschluss an den Gruppennamen geben den „Enrichment Score“ als Maß der Anreicherung an.

3.7.6 FUNKTIONELLE ANNOTATION STARK EXPRIMIERTER GENE – CHONDROGENESE

Die 100 am stärksten exprimierten Gene der chondrogen differenzierten Zellen weisen zu den zuvor beschriebenen Expressionsprofilen der Osteoblasten und alten MSC Übereinstimmungen auf. Die größte Ähnlichkeit besteht zu den alten MSC, wo 61 von 100 Genen identisch sind, unter den Top 25 Genen sind es 11 Transkripte (44%).

Tabelle 3 - 16 Stark exprimierte Gene - Top 25 der Sequenzierung des Transkriptoms der Chondrozyten

Tabelle 3 - 16 listet zusätzlich zu den 25 höchstexprimierten Genen ein uncharakterisiertes Protein aus den ersten 100 Listenplätzen auf. Die Spalte „Ontologie“ weist ein Gen als zur entsprechenden Gruppe in Abbildung 3 - 20 zugehörig aus.

Pos.	TER	Symbol	Name	Ontologie Gruppe 1-12
1	1580832	LOC781039	BLASTn 1e-75 RPI7 R. norvegicus	--Nicht gruppiert--
2	1062000	TIMP1	TIMP metallopeptidase inhibitor 1	⑤⑥
3	692485	LOC100335409	Annotation zurückgezogen (NCBI)	--Nicht gruppiert--
4	356200	LOC788801	ferritin-1, chloroplastic-like	--Nicht gruppiert--
5	332200	PDLIM4	PDZ and LIM domain 4	--Nicht gruppiert--
6	327995	RPLP0	BLOCK 23; ribosomal protein, large, P0	①
7	315440	RPS3	ribosomal protein S3	①②③⑦⑧⑨
8	314972	FTH1	ferritin, heavy polypeptide 1	③⑥⑦⑧⑨
9	305280	RPS15	ribosomal protein S15	①
10	302536	PPIA	peptidylprolyl isomerase A (cyclophilin A)	①
11	290412	LOC100140226	zinc finger protein 160-like	--Nicht gruppiert--
12	269624	RPS18	ribosomal protein S18	①②
13	259850	GSTP1	glutathione S-transferase pi 1	--Nicht gruppiert--
14	246937	FTL	ferritin, light polypeptide	--Nicht gruppiert--
15	232323	RPS2	ribosomal protein S2	①②
16	229502	RPSA	ribosomal protein SA	①
17	226986	PRDX1	peroxiredoxin 1	①③⑥⑨
18	225300	LOC783718	ribosomal protein S27a pseudogene	--Nicht gruppiert--
19	224972	RPS27L	ribosomal protein S27-like	①
20	219560	RPL10	ribosomal protein L10	①
21	217857	RPS29	ribosomal protein S29	①
22	211386	TSPO	translocator protein (18kDa)	--Nicht gruppiert--
23	201905	RPS11	ribosomal protein S11	①②
24	197710	RPS5	ribosomal protein S5	①②
25	186514	LOC100295775	ribosomal protein, large, P1 pseudogene	--Nicht gruppiert--
66	86127	C29H11orf10	chromosome 29 orf, human C11orf10	--Nicht gruppiert--

Das bereits unter Punkt 3.7.4 charakterisierte Protein *C29H11orf10* wird von den chondrogen differenzierten Zellen an 66ster Position exprimiert. Am auffälligsten ist, daß die Top 100 exprimierten Gene nur zu 9 Ontologie-Gruppen im FAC gruppierbar sind (siehe Abbildung 3 - 20). Mit Translationsprozessen assoziierbare Ontologien erreichen hierbei die höchsten „Enrichment Score“ Werte (Ontologien 1, 2). Das Vorkommen einer für Chondrozyten typischen extrazellulären Matrix wird nur durch eine Ontologie (Ontologie 5) unterstützt. Auch die in den übrigen Banken häufig vorkommenden, mit Entwicklungsprozessen assoziierbaren Ontologien werden im FAC der Chondrozyten nicht gruppiert. Als einzige Sequenzierung weist die der Chondrozyten klare Hinweisauf, die eine mangelnde Vitalität der Zellen hindeuten (Ontologien 7, 8, 9).

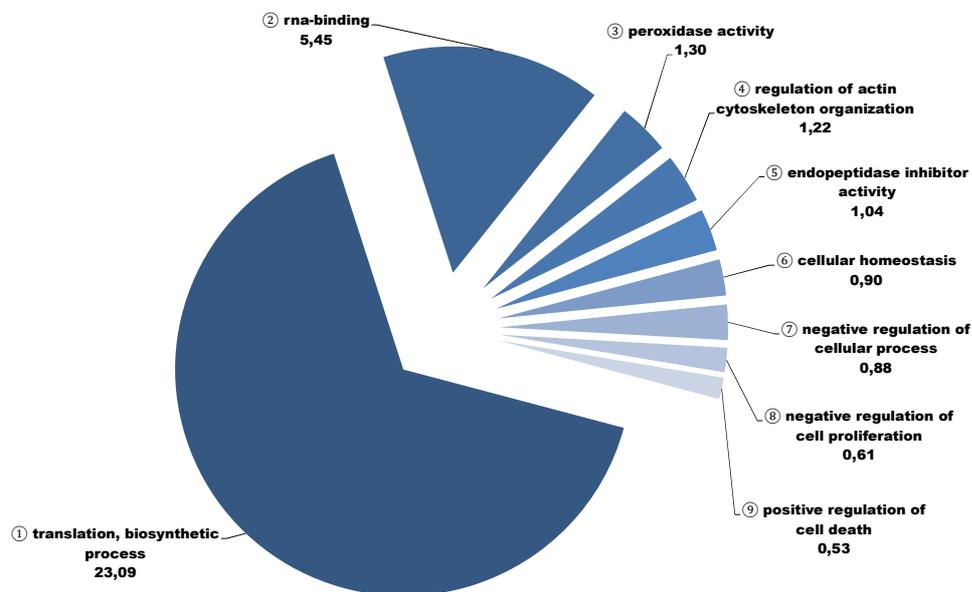


Abbildung 3 - 20 Ontologie der 100 höchstexprimierten Gene – Chondrogenese

Gezeigt werden alle 9 im DAVID FAC angereicherten Ontologie-Gruppen. Die Zahlen im Anschluss an den Gruppennamen geben den „Enrichment Score“ als Maß der Anreicherung an.

3.8 ANALYSE DIFFERENZIELL EXPRIMIERTER GENE MITTELS RNA-SEQ

Die Bestimmung differenzieller Genexpression soll unabhängig von der absoluten Expressionsstärke einzelner Gene einen Einblick in die der Differenzierung der Zellen zugrunde liegenden Mechanismen ermöglichen. Zur Bestimmung

differenziell exprimierter Gene werden die Expressionswerte (Total Exon Reads), die durch Kartieren der Datensätze mittels der Software CLC Genomics Workbench ermittelt wurden, verwendet. Verglichen wurde jeweils die Expression der Gene aus den differenzierten Zelllinien mit der Expression der Gene in den jungen MSC. Als Kriterium, nach dem ein Gen als differenziell exprimiert eingestuft wurde, dient der R_j -Wert [85] welcher größer als 8 sein muss. Um basal exprimierte Gene aus der Analyse ausschließen zu können, musste der RPKM Wert zweier verglichener Gene in zumindest einer der beiden Datenbanken einen Wert von 10 erreichen. Des Weiteren wurde zusätzlich zum R_j -Werte ein dreifacher Expressionsunterschied der RPKM Werte als Kriterium zur Feststellung der differenziellen Expression vorausgesetzt. Eine Übersicht der Analyse der differenziellen Genexpression ist in Tabelle 3 - 17 abgebildet.

Tabelle 3 - 17 Übersicht der Bestimmung differenziell exprimierter Gene

Als Referenz für die Kartierung dient das NCBI Rindergenom (build 4.2) mit 24359 annotierten Genen. Als Maß für die Bestimmung differenzieller Genexpression wird ein R_j -Wert von ≥ 8 , sowie ein mindestens dreifacher Expressionsunterschied ($\log_2 \geq |1,58|$) der RPKM Werte festgelegt. Gene, deren Expression in beiden verglichenen Sequenzierungen einen RPKM Wert von 10 nicht erreicht, werden von der Analyse ausgeschlossen (basale Expression).

Parameter	Junge MSC vs. Adipozyten	Junge MSC vs. Osteoblasten	Junge MSC vs. alte MSC	Junge MSC vs. Chondrozyten
Kartierte Sequenzen (TER) junge MSC	63.743.193			
Kartierte Sequenzen (TER) Differenzierung	33.552.752	26.686.736	39.792.759	38.705.998
Gene > 0 Seq. junge MSC	14.887			
Gene > 0 Seq. Differenzierung	15.113	15.007	15.774	16.870
Gene junge MSC ≥ 10 RPKM	4.608			
Gene Differenzierung ≥ 10 RPKM	5.115	3.122	3.137	3.468
\emptyset Expression junge MSC (RPKM)	27			
\emptyset Expression Differenzierung (RPKM)	25	38	39	42
Diff. Expr. in jungen MSC $R_j \geq 8 ; \log_2 \geq 1,58 $	394	1.989	1.893	1.724
Diff Expr. in Differenzierung $R_j \geq 8 ; \log_2 \geq 1,58 $	480	936	744	1065
Σ Differenziell exprimierter Gene	874	2.925	2.637	2.789

Von den 24359 annotierten Genen im Referenzgenom des Rindes werden zwischen 14887 (junge MSC) und 16870 (Chondrozyten) mit mindestens einer Sequenz

kartiert. Von diesen Genen werden zwischen 3122 (Osteoblasten) und 5115 (Adipozyten) mit einem RPKM von 10 oder höher exprimiert, wobei die durchschnittliche Expression zwischen RPKM Werten von 25 (Adipozyten) bis 42 (Chondrozyten) liegt. Die Anzahl als differenziell exprimiert bestimmter Gene schwankt je nach mit den jungen MSC verglichener Zelllinie sehr stark. Die einander in der globalen Genexpression ähnlichen Zelllinien (vergleiche Punkt 3.6.4) junge MSC und Adipozyten weisen insgesamt 874 differenziell exprimierte Gene auf, während beim Vergleich der jungen MSC mit den Osteoblasten 2925 Gene als differenziell exprimiert bestimmt werden. Mit maximal 21 Genen (alte MSC) werden nur wenige Gene mit einem RPKM von 10 oder größer ausschließlich in einer der verglichenen Zelllinien exprimiert. Zur Bestimmung der in jungen MSC differenziell exprimierten Transkripte lassen sich 250 Gene identifizieren, die sowohl beim Vergleich zwischen jungen MSC und Osteoblasten, als auch beim Vergleich von jungen MSC mit Adipozyten stets differenziell exprimiert werden. Um die differenzielle Genexpression in einen funktionellen Zusammenhang stellen zu können, wurden die vollständigen Genlisten aller differenziell exprimierten Gene eines Vergleichs mittels des online Tools „KEGG Mapper“ mit den dort hinterlegten biologischen Pathways abgeglichen [89]. Zudem werden die Genlisten mittels GenMAPP-CS gegen die Wikipathways [90] kartiert. Die Ergebnisse können so graphisch dargestellt werden.

Tabelle 3 - 18 DAVID/GO-Elite Analyse - Auswahl signifikant angereicherter Pathways

Tabelle 3 - 18 listet für alle Zelllinien eine Auswahl von KEGG bzw. Wiki- Pathways auf, die mit differenziell exprimierten Genen eine signifikante Abdeckung aufweisen.

Pathway	Osteoblasten		alte MSC		Chondrocyten		Adipozyten		junge MSC	
	Gene	p-Value	Gene	p-Value	Gene	p-Value	Gene	p-Value	Gene	p-Value
WNT signaling pathway	-	-	-	-	-	-	-	-	4	3,6E-02
ECM-receptor interaction	-	-	-	-	-	-	-	-	5	3,5E-02
DNA replication	-	-	-	-	-	-	-	-	7	1,6E-05
Cell cycle	-	-	-	-	-	-	-	-	8	4,0E-03
Adipocytokine signaling pathway	-	-	-	-	-	-	6	4,1E-02	-	-
Glycerolipid metabolism	-	-	-	-	-	-	5	3,5E-02	-	-
Focal adhesion	-	-	-	-	-	-	13	8,8E-03	13	3,8E-05
Cytokine-cytokine receptor interaction	-	-	-	-	-	-	14	4,0E-03	-	-
PPAR signaling pathway	-	-	-	-	-	-	8	3,9E-03	-	-
Jak-STAT signaling pathway	-	-	-	-	-	-	9	5,4E-02	-	-
Basal transcription factors	6	4,2E-02	-	-	-	-	-	-	-	-
Ribosome	37	1,3E-23	33	1,3E-21	43	2,1E-28	-	-	-	-
RNA polymerase	7	3,6E-03	6	8,3E-03	8	1,4E-03	-	-	-	-
Spliceosome	22	4,2E-06	15	1,9E-03	25	7,9E-07	-	-	-	-
Oxidative phosphorylation	44	1,0E-22	37	6,8E-19	52	4,4E-28	-	-	-	-
Chemokine signaling pathway	-	-	16	1,6E-02	-	-	-	-	-	-

Eine statistische Absicherung der Kartierung gegen die KEGG Pathways erfolgte durch das online Tool DAVID, bzw. gegen die Wikipathways durch Nutzung des GO-Elite Plugins in GenMAPP-CS [91, 92]. Eine nicht identische Anzahl der kartierten Gene je Pathway zwischen der grafischen Darstellung mittels des KEGG Mappers und der statistischen Analyse mittels DAVID ist durch Unterschiede in der serverseitigen automatischen Generkennung beim Datenimport begründet und unter biologischen Gesichtspunkten nicht von Relevanz. Eine Liste aller differenziell exprimierten Gene und deren Abdeckung der Pathways ist im elektronischen Anhang D5.3 aufgeführt, die am stärksten differenziell exprimierten Gene aller Banken sind am Ende dieser Arbeit auch in gedruckter Form angehängt. Die 100 am stärksten differenziell exprimierten Gene (250 Gene junge MSC) wurden mittels des online Tools GOEAST [93] einer Go-Term Annotation unterzogen. Um die Ergebnisse in der Menge einzugrenzen und grafisch darstellen zu können, wurden die resultierenden Go-Term Listen (nur Kategorie „*biological process*“) mittels des online Tools REVIGO [94] in sog. Treemaps umgerechnet.

3.8.1 DIFFERENZIELLE ERHÖHTE GENEXPRESSION IN JUNGEN MSC

Zum Abgleich der differenziell exprimierten Gene der jungen MSC gegen die KEGG Pathways Datenbank wurden alle 250 Gene ausgewählt, die sowohl im Vergleich zwischen jungen MSC und Osteoblasten, als auch zwischen jungen MSC und Adipozyten in den jungen MSC als differenziell hoch exprimiert bestimmt wurden (siehe hierzu Punkt 4.10.1). Mit dem Wikipathway „*WNT signaling pathway and pluripotency*“ wird ein stammzelltypischer Signaltransduktionsweg mit 4 Genen in statistisch signifikantem Maße abgedeckt. Zählt man zusätzlich zwei in der KEGG Datenbank hinterlegten und ebenfalls dem WNT Signalweg zuordenbaren Gene hinzu, kommt man auf insgesamt 6 Treffer im WNT Signalweg. Dabei handelt es sich um folgende Gene kodieren für die beiden Frizzled Liganden *wnt5a* und *wnt5b* (*wingless-type MMTV integration site family, member 5A/B*), den Rezeptor *ldlr* (*low-density lipoprotein receptor*), das Helix-Loop-Helix Calciumbindeprotein *nkd1* (*naked cuticle homolog 1*), das Proteoglykan *gpc4* (*glypican 4*), das Leucine-Zipper Protein *fosl1* (*fos-related antigen 1*) sowie um das lösliche Protein *sfrp2* (*Secreted frizzled-related protein 2*). Nach Abgleich mit der KEGG Datenbank wird der „*DNA replication*“ Pathway mit 7 Genen als am signifikantesten abgedeckt bestimmt (siehe Tabelle 3 - 18). Im Einklang dazu erfolgt die Kartierung gegen den KEGG Pathway „*cell cycle*“ mit 8 Genen ebenfalls in signifikant starkem Maße. Der KEGG Pathway „*focal*

„adhesion“ weist mit 13 Genen die höchste Anzahl kartierter Gene auf, damit funktionell in Zusammenhang stehend wird auch der KEGG Pathway „ECM-receptor interaction“ mit 5 Genen in signifikantem Maße abgedeckt. Eine Go-Term Analyse aller 250 Gene (Abbildung 3 - 21) zeigt deutlich einen funktionellen Schwerpunkt auf Signaltransduktionswegen („WNT receptor signaling pathway“, „regulation of signaling“) und der Regulation von Entwicklungsprozessen („regulation of growth“, „regulation of developmental process“), sowie der Proliferation und Zellteilung („mitotic cell cycle“, „DNA replication“).



Abbildung 3 - 21 REVI GO Treemap der Go-Term Analyse - junge MSC (250 Gene)

Die Treemap stellt funktionell ähnliche Go-Terms in identischer Farbe gruppiert dar. Die Größe der Rechtecke jedes aufgeführten Go-Terms ist direkt mit dem log10 des p-Value der Anreicherung korreliert. Je mehr Platz ein Go-Term in der Grafik einnimmt, desto signifikanter ist die Anreicherung.

Unter den 250 untersuchten, differenziell exprimierten Genen befinden sich 20 Gene, deren Genprodukt in der zur Kartierung der Sequenzen verwendeten Genomsequenzierung (NCBI *Bos taurus* Build 4.2) noch nicht vollständig charakterisiert war. Nach manueller Annotation konnten 18 der 20 Gene eindeutig zugeordnet, oder aber als fehlerhaft annotierte, hypothetische Loci verworfen werden. Zur genaueren Analyse verbleiben die beiden uncharakterisierten Genloci *LOC100335299* und *C28H1orf198*. Beide Gene werden im Vergleich mit den 4 übrigen Zelllinien als in den jungen MSC differenziell hoch exprimiert bestimmt. Beim Gen *LOC100335299* handelt es sich um ein in Amnioten (NCBI

HomoloGene:124386) stark konserviertes Gen von 99 bis 109 AS Länge. Beim Rind ist *LOC100335299* vorläufig als Modellgen mit 4 Exons annotiert, die resultierende RNA von 1620 bp (XM_005194883) würde aufgrund des kurzen Anteils kodierender Sequenz zu einem nur 101 Aminosäuren langen Protein translatiert. Konservierte Domänen lassen sich *LOC100335299* nicht vorhersagen. Auch *C28H1orf198* weist eine Spezies übergreifende Konservierung auf. Für das Protein wird im Rind als konservierte Domäne die der prokaryotischen DNA polymerase III, Untereinheit gamma und tau als einziger Anhaltspunkt mit mäßiger Signifikanz vorhergesagt (NCBI Conserved Domain Scan, E-Value 8,01E-03).

3.8.2 DIFFERENZIELL ERHÖHTE GENEXPRESSION IN ADIPOZYTEN

Im Vergleich der jungen MSC mit den zu Adipozyten differenzierten Zellen werden 874 Gene als differenziell exprimiert bestimmt. Davon werden 480 Gene gemäß der genannten Kriterien in Adipozyten differenziell stärker exprimiert als in den jungen MSC. Gleicht man die in Adipozyten differenziell exprimierten Gene mit den biologischen Pathways ab, so finden sich mehrere Treffer, die eine statistisch signifikante Abdeckung aufweisen. Der Pathway „*cytokine-cytokine receptor interaction*“ wird mit 15 (14 Gene gemäß DAVID) zu 3 Genen von den in Adipozyten differenziell exprimierten Genen wesentlich stärker als von den in jungen MSC differenziell exprimierten Genen abgedeckt. Die zur Gruppe der gp130 Rezeptoren gehörenden Gene *il6st* (*interleukin 6 signal transducer*) und *osmr* (*oncostatin m receptor*) sind ebenso Teil des „*cytokine-cytokine receptor interaction*“ Pathways, wie die Gene kodierend für die Rezeptoren *ghr* (*growth hormon receptor*) und *il4r* (*interleukin 4 receptor*). Dies ist deshalb von Interesse, da letztgenannte 4 Rezeptoren einen zentralen Bestandteil des mit insgesamt 9 differenziell exprimierten Genen in Adipozyten ebenfalls stark abgedeckten „*jak-stat-signaling pathway*“ darstellen, einem Signalweg, der die gezielte Steuerung der Transkription von Zielgenen durch äußere Einflüsse ermöglicht (siehe Abbildung 3 - 22 A/B). Der Signaltransduktionsweg „*adipocytokine signaling pathway*“ wird mit 6 differenziell stärker exprimierten Genen in den Adipozyten in Einklang mit der Differenzierung der Zellen statistisch signifikant abgedeckt. Selbiges gilt für 8 Gene des „*ppar signaling pathway*“, sowie für den „*glycerolipid metabolism pathway*“, der mit 8 Genen (5 Gene gemäß DAVID) signifikant abgedeckt wird.

Mit insgesamt 15 Genen (13 Gene gemäß DAVID) wird der „focal adhesion“ pathway in den Adipozyten ähnlich stark wie zuvor bei den jungen MSC beobachtet abgedeckt, auch wenn hierbei aufgrund der differentiellen Genexpression keinerlei Überschneidungen der beteiligten Gene zu beobachten sind. Die untere Abbildung zeigt die Ergebnisse der Go-Term Analyse der 100 am stärksten differenziell exprimierten Gene der Adipozyten als Treemap dargestellt.

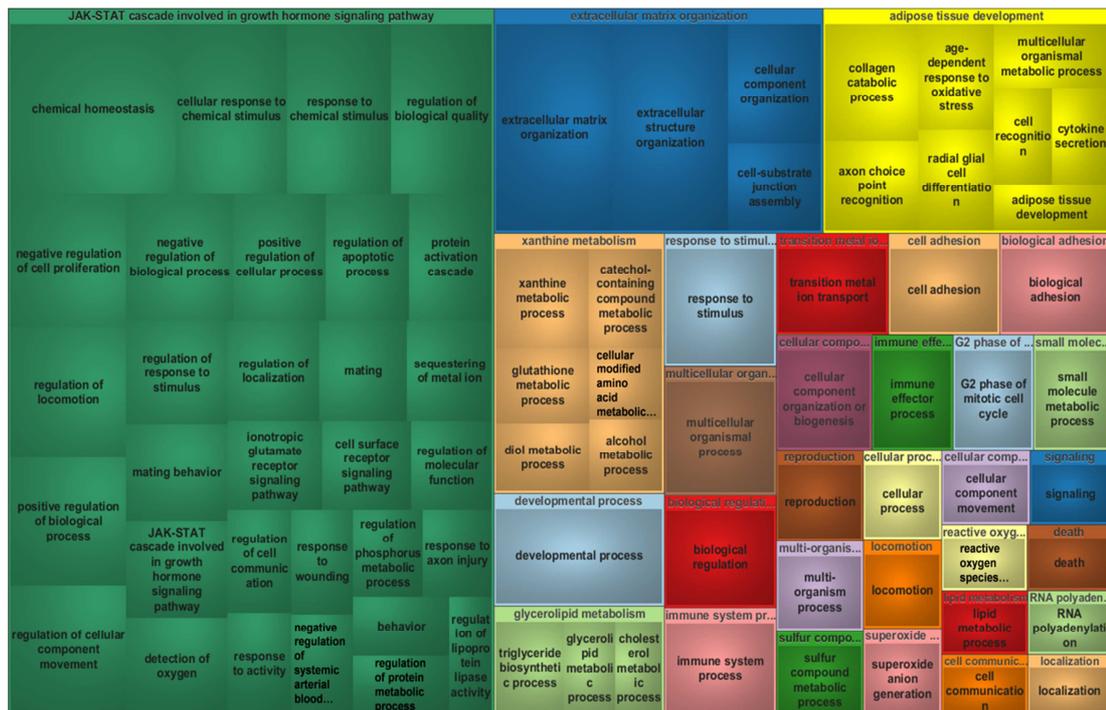


Abbildung 3 - 23 REVIGO Treemap der GO-Term Analyse - Adipozyten (Top 100 Gene)

Die Treemap stellt funktionell ähnliche Go-Terms in identischer Farbe gruppiert dar. Die Größe der Rechtecke jedes aufgeführten Go-Terms ist direkt mit dem log10 des p-Value der Anreicherung korreliert. Je mehr Platz ein Go-Term in der Grafik einnimmt, desto signifikanter ist die Anreicherung.

Unter den 100 im Vergleich zu jungen MSC am stärksten differenziell exprimierten Genen findet sich nach manueller Annotation kein für weitere Analysen interessantes Kandidatengen.

3.8.3 DIFFERENZIELL ERHÖHTE GENEXPRESSION IN OSTEOLASTEN UND ALTEN MSC

Die in alten MSC im Vergleich zu den jungen MSC differenziell hoch exprimierten Gene finden sich zu 73,38% (546 von 744 Genen) auch in der Liste der 936 in Osteoblasten im Vergleich zu den jungen MSC differenziell hoch exprimierten

Gene. Auf eine teilweise redundante, getrennte Betrachtung der differentiellen Genexpression wird daher verzichtet. Die für Osteoblasten typischen Markergene (siehe Punkt 3.7.1) *spp1*, *omd*, *bglap* und *ibsp* werden sowohl in alten MSC, als auch in Osteoblasten als gegenüber den jungen MSC differenziell erhöht exprimiert bestimmt. Beide Zelllinien weisen im Vergleich zu den jungen MSC folglich mit Osteoblasten assoziierbare, differenzielle Genexpressionsprofile auf. Eine GOEAST Analyse der 198 Gene, die nur in alten MSC, nicht jedoch in den Osteoblasten gegenüber den jungen MSC differenziell erhöht exprimiert werden, zeigt unter anderem die GO-Terms „negative regulation of biomineral tissue development“ und „negative regulation of bone mineralization“ als durch die Gene *aspn* (asporin), *srgn* (serglycin) und *ecm1* (extracellular matrix protein 1) statistisch signifikant abgedeckt. Im Unterschied dazu werden durch die 390 exklusiv in den Osteoblasten gegenüber den jungen MSC differenziell erhöht exprimierten Genen die GO-Terms „endochondral ossification“ und „replacement ossification“ durch die Gene *gnas* (guanine nucleotide-binding protein G) und *bmp4* (bone morphogenetic protein 4) kartiert.

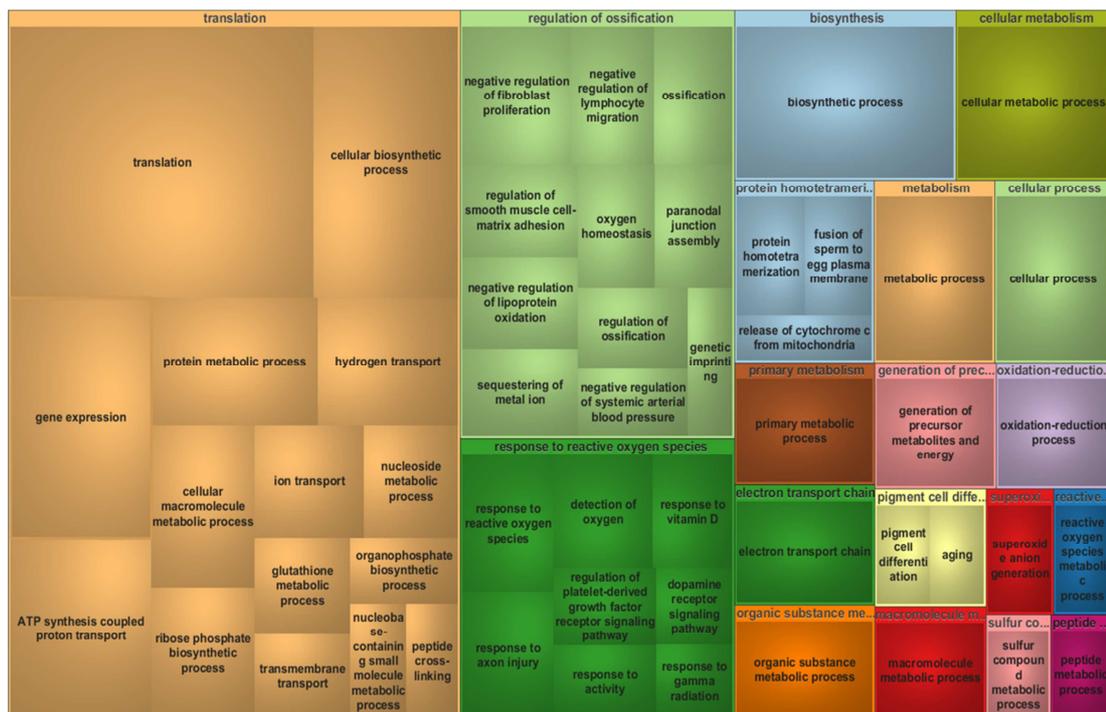


Abbildung 3 - 24 REVIGO Treemap der Go-Term Analyse - Osteoblasten (Top 100 Gene)

Die Treemap stellt funktionell ähnliche Go-Terms in identischer Farbe gruppiert dar. Die Größe der Rechtecke jedes aufgeführten Go-Terms ist direkt mit dem log10 des p-Value der Anreicherung korreliert. Je mehr Platz ein Go-Term in der Grafik einnimmt, desto signifikanter ist die Anreicherung.

Der stark osteogene Charakter der zu Osteoblasten differenzierten MSC lässt sich auch anhand der GO-Term Analyse der 100 im Vergleich zu den jungen MSC am stärksten differenziell exprimierten Gene belegen (siehe Abbildung 3 - 24). Der funktionelle Schwerpunkt der von den Osteoblasten differenziell erhöht exprimierten Gene liegt deutlich auf den Bereichen der Transkription/Translation und der zellulären Energiegewinnung. Es werden nach der DAVID Kartierung 5 signifikant abgedeckte KEGG Pathways bestimmt, die mit den genannten biologischen Prozessen assoziierbar sind (siehe Tabelle 3 - 18, S. 102). Die alten MSC weisen aufgrund des insgesamt sehr ähnlichen Profils der differenziellen Genexpression in 4 der 5 KEGG Pathways ebenfalls eine statistisch signifikante Kartierung auf. Ein auffälliger Unterschied in der Genontologie der differenziell exprimierten Gene zwischen Osteoblasten und alten MSC besteht in den ausschließlich in alten MSC angereicherten GO-Terms für Chemokin- („*chemokine activity*“) und Cytokin Aktivität („*cytokine activity*“). Dieses Ergebnis der GO-Term Analyse findet auch durch den in alten MSC signifikant stark kartierten KEGG Pathway „*chemokine signaling pathway*“ Bestätigung. Unter den 100 am stärksten differenziell exprimierten Genen finden sich nach Vergleich der Osteoblasten mit den jungen MSC 15 uncharakterisierte Transkripte, im Falle der alten MSC 16 uncharakterisierte Transkripte, die in den differenzierten Zellen stärker als in den jungen MSC exprimiert werden. Sieben dieser Gene werden von beiden Zelllinien differenziell exprimiert. Nach manueller Annotation verbleiben 2 sowohl von Osteoblasten als auch von alten MSC differenziell exprimierte Kandidatengene zur weiteren Analyse, darunter der unter Punkt 3.7.4 aufgeführte offene Leserahmen *C29H11orf10*. Beim anderen Gen handelt es sich um den offenen Leserahmen *C5H12ORF57*, ein bei den *Eutelostomi* konserviertes Gen (NCBI HomoloGene: 8402), dessen Transkript ein 122 bis maximal 132 (*Bos taurus*) AS langes Protein ergibt. Das Gen wird alternativ auch als *C10* bezeichnet. Eine Interpro Analyse zeigt eine dem Protein eigene, nicht näher charakterisierte Domäne als Treffer an.

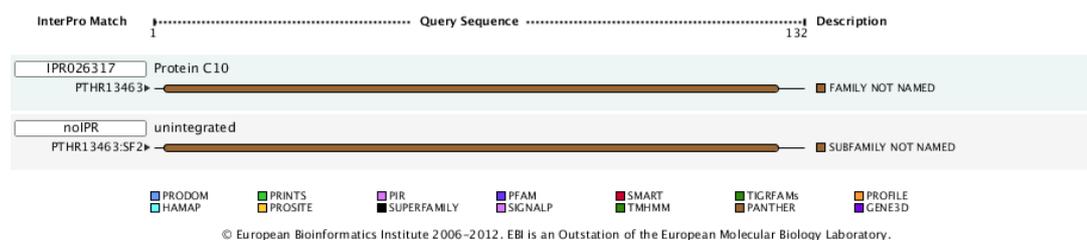


Abbildung 3 - 25 Interpro Analyse der Proteinsequenz des Gens *C5H12ORF57*

Für den abgebildeten Interpro-Scan wurde die bovine Proteinsequenz NP_001068993.1 verwendet.

3.8.4 DIFFERENZIELL ERHÖHTE GENEXPRESSION IN CHONDROZYTEN

Die in Chondrozyten im Vergleich zu den jungen MSC differenziell exprimierten Gene haben nach GO-Term Analyse und der Kartierung gegen die biologischen Pathways keinen nachweisbaren Bezug zu Prozessen des Knochen- oder Knorpelgewebes. Die Kartierung der in Chondrozyten differenziell exprimierten Gene gegen die KEGG Datenbank zeigt zunächst eine auffällige Ähnlichkeit der statistisch signifikant abgedeckten Pathways mit denen, die auch in alten MSC und den Osteoblasten mit dem Schwerpunkt auf Transkription/Translation und Energiegewinnung (siehe Tabelle 3 - 18) kartiert werden. Durch die Analyse der 100 am stärksten differenziell exprimierten Gene mittels des online Tools GOEAST fällt eine hohe Abdeckung von mit Apoptose in Zusammenhang stehenden Ontologien auf (z.B. *regulation of execution phase of apoptosis*). Gene wie *dedd2* (*DNA-binding death effector domain-containing protein 2*), *diablo*, *bad* (*BCL2-associated agonist of cell death*), *pycard* (*Apoptosis-associated speck-like protein containing a CARD domain*), *htra2* (*Serine protease HTRA2, mitochondrial*) und *bok* (*BCL2-related ovarian killer*), die eindeutig proapoptotische Wirkung aufweisen, werden in Chondrozyten differenziell stärker als in den jungen MSC exprimiert. In den jungen MSC finden sich hingegen beim direkten Vergleich mit den Chondrozyten die Apoptose inhibierende Gene wie *survivin* (*apoptosis inhibitor 4 = birc5*) und *api5* (*apoptosis inhibitor 5*) und *ciapin1* (*Cytokine-induced apoptosis inhibitor 1*) unter den differenziell hoch exprimierten Genen. Unter den 100 im Vergleich zu jungen MSC am stärksten differenziell exprimierten Genen findet sich nach manueller Annotation kein für weitere Analysen interessantes Kandidatengen.

3.9 CHARAKTERISIERUNG UNBEKANNTER TRANSKRIPTE

Um im Rind noch nicht annotierte Transkripte finden zu können, wurden alle Sequenzen, die in den CLC RNA-Seq Kartierungen nicht kartiert werden konnten, einer weiteren Analyse unterzogen. Zunächst wurden durch „Bowtie 1“ Kartierungen gegen entsprechende Datenbanken (siehe Punkt 3.6.3) all jene Sequenzen entfernt, die für repetitive Elemente oder ribosomale RNA kodieren. Die verbleibenden Sequenzen wurden mittels der Software „CLC Genomics Workbench“ assembliert und auf offene Leserahmen von mindestens 100 Aminosäuren Länge hin untersucht. Sämtliche Contigs wurden durch BLASTn/x

Abgleich mit der NCBI Refseq-RNA (E-Value 1e-010) bzw. nr-Proteindatenbank (E-Value 1e-006) wenn möglich annotiert. Die Ergebnisse der Assemblierung und BLAST Analysen sind in Tabelle 3 - 19 aufgeführt, die komplette Analyse ist im Digitalen Anhang D6 hinterlegt. Wie zuvor schon bei der Assemblierung der kompletten Datensätze (siehe Punkt 3.6.3) beobachtet, fallen bei der Analyse der Assemblierung der unkartierten Sequenzen ähnliche Unterschiede zwischen den im „Cologne Center for Genomics“ (Köln) sequenzierten, linear amplifizierten Transkriptomen und den im eigenen Haus gewonnenen Daten aus nicht amplifizierten Material auf. Die im eigenen Haus erzeugten Datensätze der jungen MSC und Adipozyten weisen die mit Abstand meisten assemblierten Sequenzen auf. Auch die durchschnittliche Contiglänge ist hier mit mindestens 1063 bp (Adipozyten) am höchsten. Die Länge der einzelnen Sequenzen vor Assemblierung spielt hierbei eine untergeordnete Rolle (Adipozyten Ø 40,81 bp). Die BLASTx Analysen zeigen, daß maximal 34,94 % (junge MSC) der Contigs bereits in den Datenbanken vorhandenen Proteinen zugeordnet werden können. Die Trefferquote beim BLASTn Abgleich der Contigsequenzen mit der refseq-RNA Datenbank liegt zwischen 27% (junge MSC) und 40,32% (Osteoblasten) über der beim BLASTx Abgleich erzielten Quote. Die durch BLAST Analysen nicht zuordenbare Menge an Daten schwankt zwischen 25,11% (Osteoblasten) und 37,46% (junge MSC).

Tabelle 3 - 19 Assemblierung und BLASTn/x Analyse unkartierter Sequenzen

Ergebnisse der Prozessierung der im RNA-Seq Verfahren unkartierten Sequenzen. Repetitive Sequenzen und rRNA werden durch Kartierung entfernt, die verbleibenden Sequenzen assembliert, und die Contigs mittels BLASTn/x Analysen wenn möglich annotiert.

Prozess	msc_neu		msc_alt		adipo		osteo		chondro	
Sequenzen	22.191.853		15.955.457		14.831.362		25.543.170		13.145.525	
Ø Sequenzlänge	67,07 bp		64,27 bp		40,81 bp		64,71 bp		60,24 bp	
rRNA / Repetitiv (seq)	3.210.896	14,47 %	2.700.523	16,93 %	1.085.596	7,32 %	1.085.587	4,25 %	2.213.007	16,83 %
Assembliert (seq)	10.963.490	49,4 %	2.796.944	17,53 %	6.003.822	40,48 %	1.775.475	6,95 %	2.148.983	16,35 %
Nicht assembliert (seq)	8.017.467	36,13 %	10.457.990	65,54 %	7.741.944	52,2 %	22.682.108	88,8 %	8.783.535	66,82 %
Anzahl Contigs	7.192		5.084		3.879		3.118		3.777	
Ø Contiglänge (bp)	1.120		737		1.063		695		627	
N50 (bp)	726		590		733		572		544	
N20 (bp)	1.543		926		1.443		869		775	
BLASTx Treffer (seq)	2.513	34,94 %	1.399	27,52 %	1.776	45,78 %	1.078	34,57 %	952	25,21 %
BLASTn Treffer (seq)	4.498	62,54 %	3.231	63,55 %	3.178	81,93 %	2.335	74,89 %	2.389	63,25 %
Nicht zuordenbar	2.694	37,46 %	1.853	36,45 %	701	18,07 %	783	25,11 %	1.388	36,75 %

Zum Auffinden von funktionellen Domänen wurden die Contigsequenzen zudem mittels Blast2GO [95] automatisch mit den durch Interpro zusammengefassten Datenbanken zur Protein Charakterisierung abgeglichen. Sämtliche so erhaltenen Daten wurden in einer Tabelle kombiniert (nicht gezeigt). Zur weiteren, Analyse wurden solche Contigs herangezogen, die den unter Punkt 2.7.4 angegebenen Bedingungen genügen. Durch die genannten Einschränkungen konnten in den 5 sequenzierten Zelllinien insgesamt 17 potentiell interessante Contigs zur weiteren Analyse ermittelt werden. Aus allen Contigs der jungen MSC wurden 8 Sequenzen, aus denen der Adipozyten 5, den Chondrozyten 3 und den alten MSC eine Sequenz selektiert. Für die aus der Sequenzierung der Osteoblasten stammenden Contigs erfüllte keine Sequenz die Einschlusskriterien. Aus diesen 17 Contigs konnten nach manueller Annotation 4 Sequenzen ausgesucht werden, die potentiell unbekannte Transkripte, oder noch nicht annotierte Bereiche bereits bekannter Transkripte darstellen. Für diese 4 Contigs wurde per BLASTn auf der Homepage des NCBI die Position im Genom des Rindes bestimmt (siehe Tabelle 3 - 20). Für eine erneute Kartierung des Bereichs mit „Bowtie 1“ konnte aus den Datenbanken des NCBI ein zum Blasttreffer passender, genomischer DNA Abschnitt aus dem Rindergenom ausgeschnitten werden. Dieser DNA Abschnitt wurde erneut mit allen NGS Sequenzen aus der Sequenzierung derjenigen Zelllinie kartiert, aus der der zur charakterisierende Contig stammt. Es wurden nur Sequenzen mit 100% Übereinstimmung zur Referenz kartiert.

Tabelle 3 - 20 Übersicht der uncharakterisierten Transkripte

Assemblierungsstatistiken und genomische Lokalisierung der 4 Contigs, die durch manuelle Analysen näher charakterisiert werden sollen. Drei der Contigs entstammen der Assemblierung der uncharakterisierten Sequenzen der Adipozyten Sequenzierung, eine stammt aus der Sequenzierung der jungen MSC.

Name	Länge	Sequenzen	Abdeckung	Genomischer BLASTn	Annotation im Trefferbereich
msc_neu_unm_contig_1245	5940	2223	27,93	Chromosom 10 - Plus/Minus NW_003104280.1 (6,178,227..6,184,127) (6,170,909..6,170,944)	Trifft <i>LOC100847605</i> 5' <i>CSNK1G1</i> 3' <i>PIIB</i>
adipo_unm_contig_1765	2247	6202	113,22	Chromosom 18 - Plus/Minus NW_003104463.1 (7,964,469..7,966,715)	5' <i>KIAA0513</i> 3' <i>FAM92B</i>
adipo_unm_contig_147	2707	1574	23,22	Chromosom 1 - Plus/Minus NW_003103812.1 (8,396,784..8,399,490)	5' <i>LOC100847171</i> 3' <i>BCL6</i>
adipo_unm_contig_3147	3510	3656	43,36	Chromosom 1 - Plus/Plus NW_003103812.1 (7,243,166..7,246,676)	5' <i>LOC100847108</i> 3' <i>LPP</i>

Der Contig 1245 der jungen MSC weist einen zweigeteilten Treffer im Genom des Rindes auf (siehe Abbildung 3 - 26). Eine Intron/Exon Grenze ist anhand der

Kartierung nicht sicher erkennbar. Der bereits annotierte, uncharakterisierte genomische Locus *LOC100847605* wird von Contig 1245 komplett überspannt. Dieser Locus weist laut Annotation der NCBI Datenbank 3 Exons auf. Die Kartierung mit den NGS Sequenzen der jungen MSC Zelllinie ist auch im Bereich der Introns beinahe durchgehend vorhanden, zudem beginnt sie 1014 bp vor dem annotierten Start von *LOC100847605*.

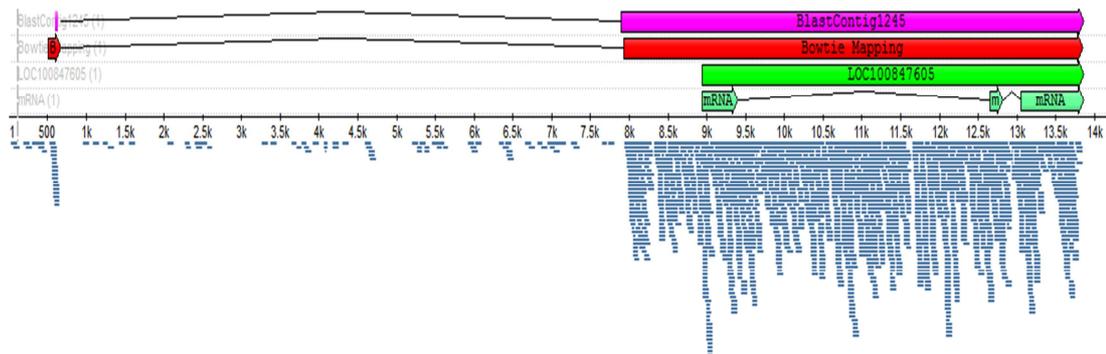


Abbildung 3 - 26 Kartierung der genomischen Region im Bereich des Contig 1245

Abgebildet ist ein ca. 14 kB großes Teilstück der genomischen Region um den zweigeteilten Blasttreffer des Contig 1245 (violett). Die Ergebnisse der Kartierung sind schematisch in Rot („Bowtie_Mapping“) dargestellt, die Abdeckung wird aus dem unteren Teil der Abbildung ersichtlich. Es wurden 2121 Sequenzen kartiert. Der bereits annotierter *LOC100847605* und die zugehörige RNA sind in grüner Farbe abgebildet.

Die Kartierung weist 2 kleine Lücken von 4 und 8 bp ab Base 8296 und Base 11617 des kartierten Referenzabschnitts auf. Der vom Blasttreffer des Contigs 1245 abgedeckte Bereich ist mit dem Kartierungsergebnis nicht deckungsgleich. Der Contig deckt die genomische Referenz am 5' Ende des ersten Treffers erst 100 bp nach Beginn der Kartierung ab. Am 5' Ende des zweiten Treffers beginnt hingegen der vom Contig 1245 abgedeckte Teil der Referenz 34 bp vor den ersten kartierten Sequenzen. Die Interpro Analyse des Contigs 1245 zeigt eine Übereinstimmung mit einem SignalPeptid (SignalP-NN(euk)).

Contig 1765 konnte aus unkartierten Sequenzen der Adipozyten assembliert werden. Im genomischen BLASTn kann die Sequenz auf Chromosom 18 im Abstand von ca. 2 Kb vom in seiner Funktion uncharakterisierten Gen *KIAA0513* lokalisiert werden (siehe Abbildung 3 - 27). Die Interpro Analyse des Contis 1765 zeigt eine seg (SEG) Domäne. Die Kartierung eines um das komplette Gen *KIAA0513* erweiterten genomischen Bereichs mit dem Datensatz der Sequenzierung

der Adipozyten beginnt die Referenz bereits 181 bp vor Exon 1 abzudecken. Danach bestätigt sich die bei NCBI hinterlegte Annotation des Gens bis zum 8ten Exon. Zwischen Exon 8 und 10 werden im Intron fast kontinuierlich Sequenzen mit bis zu 14facher Adeckung an die Referenz kartiert. In diesem Bereich gibt es 4 kleine Lücken zwischen einer und höchstens 17 bp Länge. Das letzte Exon von *KIAA0513* wird bis 4,5 Kb weit über den annotierten Bereich verhältnismäßig stark kartiert. Ab Base 22876 des gezeigten Referenzabschnitts bricht die Abdeckung stark ein und setzt sich bis Base 24204 mit sichtbar geringerer Stärke fort. Auch im kartierten Bereich nach Ende der Annotation des letzten Exons von *KIAA0513* lassen sich kleine Lücken von maximal 8 Basen entdecken.

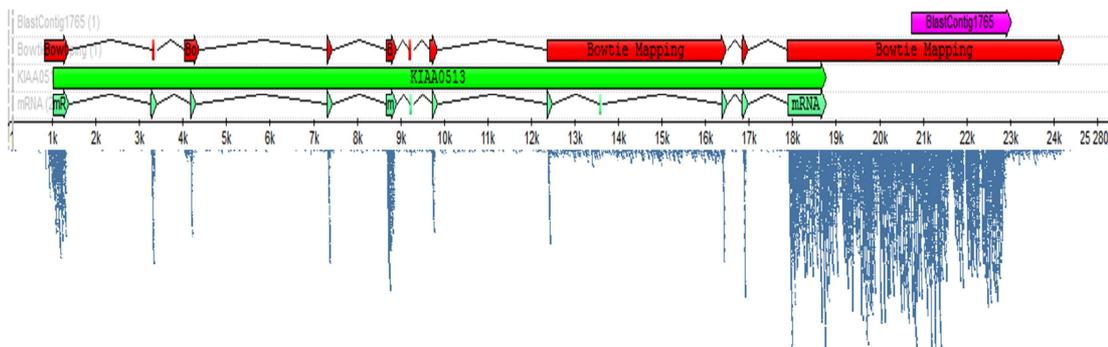


Abbildung 3 - 27 Kartierung der genomischen Region im Bereich des Contig 1765

Kartierung von 17162 Sequenzen an eine 25,8 Kb lange genomische Referenzsequenz im um das Gen *KIAA0513* erweiterten Bereich des BLASTn Treffers des Contig 1765.

Der Contig 147 kann im genomischen BLASTn auf Chromosom 1 zwischen den Genen *BCL6* und dem uncharakterisierten *LOC100847171* lokalisiert werden. Die Kartierung in Abbildung 3 - 28 zeigt eine von wenigen, maximal 30 bp langen Lücken unterbrochene Kartierung der genomischen Referenz zwischen *LOC100847171* und *BCL6*. Die Abdeckung der Referenz mit kartierten Sequenzen beginnt in 549 bp Entfernung vorm ersten Exon von *LOC100847171* und erstreckt sich über 9123 bp mit bis zu 54facher Abdeckung bis ins erste Exon von *BCL6*. Die Kartierung zeigt eine nur basale Expression des gegensätzlich zu *BCL6* orientierten *LOC100847171*. Das Gen *BCL6* wird verhältnismäßig stark exprimiert, die nicht vollständig abgebildete Intron/Exon Struktur wird durch die Kartierung bestätigt.

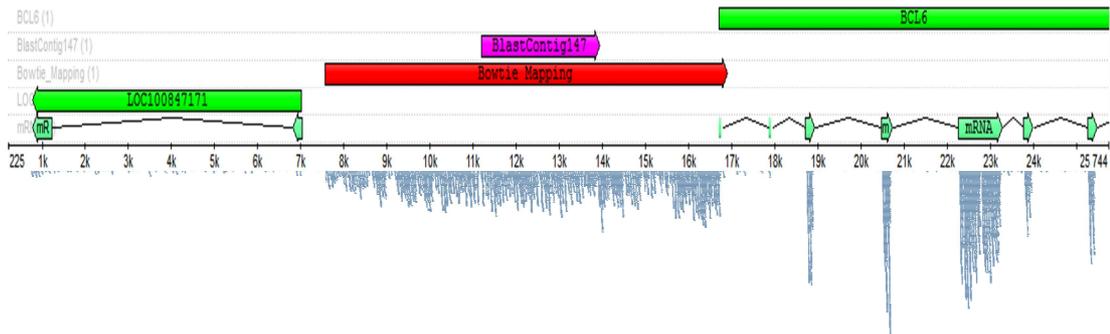


Abbildung 3 - 28 Kartierung der genomischen Region im Bereich des Contig 147

Kartierung von 8302 Sequenzen an eine 25,7 Kb lange genomische Referenzsequenz im Bereich des BLASTn Treffers des Contig 1765. Das Gen BCL6 ist nicht komplett abgebildet, es erstreckt sich noch über weitere 5720 Basen und wird über den gesamten Bereich durch die Kartierung in der annotierten Intron/Exon Struktur unterstützt.

Der Contig 3147 wird durch einen genomischen BLASTn Abgleich auf Chromosom 1 zwischen den in gleicher Richtung orientierten Genen *LPP* und *LOC100847108* lokalisiert (siehe Abbildung 3 - 29). In der „Bowtie 1“ Kartierung zeigt sich eine vom letzten Exon des Gens *LPP* bis zum letzten Exon des *LOC100847108* reichende Abdeckung des Intergenbereichs mit Sequenzen aus der Sequenzierung der Adipozyten RNA. Dieser insgesamt 16.217 bp lange, kartierte Bereich wird von 11 Lücken unterbrochen. Die 8 größten Lücken von maximal 146 bp liegen zwischen 263 und 265 Kb der abgebildeten Referenzsequenz. Die übrigen drei Lücken sind mit wenigen Basenpaaren Länge über den Rest des Kartierungsbereiches verteilt. Auch für die relativ langen Lücken lassen sich keine klaren Intron/Exon Grenzen nach der GT/AG Regel durch die Kartierung feststellen.

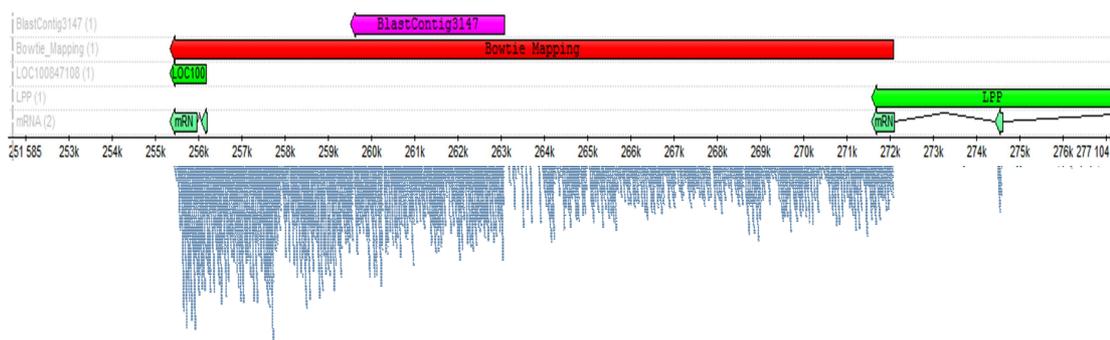


Abbildung 3 - 29 Kartierung der genomischen Region im Bereich des Contig 3147

Abgebildet ist ein ca. 25,5 Kb langes Stück genomischer DNA, welches den BLASTn Treffer des Contig 3147, sowie den kompletten *LOC100847108* und Teile des Gens *LPP* umfasst. Das Gen *LPP* ist auf der genomischen DNA des Rindes mit 683 Kb Länge annotiert.

4 DISKUSSION

Mesenchymalen Stammzellen der höheren Säugetiere kommen vornehmlich im Mark der großen Röhrenknochen vor. Das trilineare Differenzierungspotential zu Fett, Knorpel und Knochenzellen, weist MSC als für adulte Stammzellen vergleichsweise vielseitig aus. Da die Zellen auch nach Abschluss des Längenwachstums im Knochenmark adulter Organismen vorhanden bleiben, liegt eine dauerhaft benötigte Funktion dieser Stammzellen nahe, die über die Gewebebildung während der Wachstumsphase hinausgeht. So nehmen MSC wichtige regenerative Funktionen wahr, und spielen eine unterstützende Rolle bei der Blutbildung. Die Mechanismen der Differenzierung setzen eine komplexe Regulation auf transkriptioneller Ebene voraus. Die Art und Menge der gebildeten Transkripte ist hierbei der Schlüssel zur Steuerung des Differenzierungsprozesses. Ziel dieser Arbeit ist es, anhand eines Zellkulturmodells Einblicke in die Genexpression mesenchymaler Stammzellen und ihrer differenzierten Folgelinien zu erhalten. Als Modellorganismus wurde das Rind (*Bos taurus*) ausgewählt, da aufgrund seiner großen und gut verfügbaren Röhrenknochen native MSC Populationen in ausreichender Menge quasi als Abfallprodukt der Nahrungsmittelindustrie kostenfrei und ohne ethische Bedenken für Forschungszwecke zur Verfügung stehen. Die Gewinnung nicht immortalisierter, humaner MSC Linien aus dem Knochenmark gesunder Personen würde einen erheblichen medizinischen Eingriff darstellen, der nicht zu vertreten ist. Die generelle Übertragbarkeit der am Rind erarbeiteten Ergebnisse auf den Menschen kann aufgrund der hohen genetischen Ähnlichkeit und den vermutlich ebenso ähnlichen Mechanismen der Stammzellendifferenzierung als äußerst wahrscheinlich angenommen werden. Von den ca. 22.000 Genen des Rindes finden sich für 14.345 Gene Orthologe in 7 untersuchten Säugetierarten [69]. Die Genexpressionsanalyse mittels Sequenzierungstechniken der nächsten Generation (NGS = „next generation sequencing“) im Hochdurchsatzverfahren ist derzeit sicher die effektivste Methode, umfassenden Einblick in das Transkriptom von Zielzellen zu erhalten. Neben der

Expressionsstärke einzelner Gene, kann hierbei auch ein funktioneller Zusammenhang durch Abgleich des Expressionsprofils jeder Zelllinie mit biologischen Pathways und Gen-Ontologie Datenbanken erfolgen. Unterschiede zwischen differenzierten und nicht differenzierten Stammzellen können im direkten Vergleich der Expressionsprofile sicht- und ebenfalls funktionell einordenbar gemacht werden. Zusätzlich erlauben die im Rahmen dieser Arbeit durchgeführten Microarray Experimente mit einer zweiten, auf Hybridisierung beruhenden Technik Aussagen über die differenzielle Genexpression im Vergleich von differenzierten und undifferenzierten MSC zu treffen. Die Sequenzierdaten erlauben es ferner, auch noch nicht charakterisierte Transkripte, oder bislang nicht bekannte Abschnitte bereits bekannter Transkripte sichtbar zu machen und zu analysieren.

4.1 PRIMÄRKULTUR UNDIFFERENZierter MSC

Die Aussagekraft der in dieser Arbeit erzielten Ergebnisse hängt wesentlich von der erfolgreichen Etablierung der Primärkulturen potenter Mesenchymaler Stammzellen aus dem Knochenmark des Rindes, sowie deren Differenzierbarkeit *in vitro* ab. Die Isolation von MSC aus dem Knochenmark von Rindern sowie deren Differenzierung wurde bereits von Bosnakowski sowie Mauck und Kollegen beschrieben [70, 131] und die Technik im Rahmen dieser Arbeit in leicht modifizierter Form übernommen. Bereits nach wenigen Tagen lassen sich plastikadherent wachsende, fibroblastenähnliche Zellen am Boden der Kulturschalen lichtmikroskopisch beobachten. Diese bilden mit steigender Zelldichte das für MSC typische, wirbelförmige Wachstumsmuster am Boden der Kulturflaschen aus. Die Antikörperfärbungen weisen junge, undifferenzierten Zellen als CD44 positiv und CD45 negativ aus (siehe Punkt 3.2.1), was sich mit den Spezies übergreifenden Literaturangaben über das Antigenexpressionsprofil von MSC deckt [71-73]. Zudem zeigen die fluoreszenzmikroskopischen Aufnahmen ein homogenes Erscheinungsbild, alle DAPI gefärbten Zellkerne können einem CD44 Signal zugeordnet werden. Der für hämatopoetische Stammzellen typische Marker CD45 zeigt bei klaren DAPI Kernsignalen keinerlei Fluoreszenz. Sein Fehlen ist eines der minimalen Kriterien zur sicheren Identifikation von MSC [71]. Die Ergebnisse der RT-PCR Analysen decken sich mit denen der Fluoreszenzmikroskopie, für CD44 kodierende cDNA lässt sich aus undifferenzierten MSC amplifizieren, CD45 cDNA ist hingegen nicht nachweisbar

(siehe Abbildung 3 - 4). Da ein passender Antikörper gegen bovines CD105 nicht verfügbar war, und eine erhoffte Kreuzreaktion mit einem humanen CD105 Antikörper ausblieb (Daten nicht gezeigt), konnte für CD 105 nur die Expression auf RNA Ebene nachgewiesen werden. In der Literatur finden sich für verschiedene Spezies eine Auswahl an mutmaßlich für MSC typischen Transkripten (siehe Tabelle 3 - 1). Die RT-PCR zur Überprüfung der Expression zeigt für embryonale Stammzellmarker wie *oct4*, *nanog*, *sox2* und *lif* keine eindeutigen Ergebnisse. Zwar konnte für jedes Gen ein PCR Produkt nachgewiesen werden, jedoch galt dies nicht für alle getesteten Kulturen junger MSC, und teilweise war eine sog. „nested“ PCR nötig, um eine klare Bande des Zielgens im Agarosegel zu erhalten (*oct4*, *nanog*). Die Transkriptionsfaktoren *oct4*, *nanog*, *lif* und *sox2* regulieren bei embryonalen Stammzellen die Selbsterneuerung und Differenzierbarkeit, wobei eine direkte Interaktion zwischen *oct4* und *sox2* belegt ist (Heterodimer), und sich alle drei Transkriptionsfaktoren als wahrscheinliche Co-Occupatoren vieler Promotoren herausgestellt haben [132]. Zudem scheint es aufgrund der Arbeit von Pan und Kollegen sehr wahrscheinlich, daß *nanog* ein nachgeschaltetes Gen des *Lif/Stat3* Pathways ist und sich die Expression der letztgenannten 4 Gene in einer Art komplexem regulatorischem Netzwerk gegenseitig beeinflusst [133]. Es ist zunächst naheliegend, daß diese Kerngene der Stammzeleigenschaften dieselben Funktionen in embryonalen und in adulten Stammzellen wahrnehmen. Somit wäre die Expression von ESC Markern ein wichtiges Indiz für die Stammzellcharakteristik der Primärkulturen in dieser Arbeit. Das gezielte Ausschalten der Expression von *oct4* zeigte jedoch für MSC und andere somatische Stammzellen in der Maus in Arbeiten von Lengener und Kollegen keinerlei Einfluß auf deren Differenzierbarkeit und Proliferation [134]. Die Autoren gehen aufgrund der in der Literatur beschriebenen, relativ geringen Mengen nachweisbaren Transkripts im Expressionsvergleich adulter SC/ESC sogar von einer falsch positiven Detektion bzw. der unbeabsichtigten Amplifikation eines der zahlreichen *oct4* Pseudogene in den adulten Stammzellen aus. Somit läge generell keine Funktion des Gens *oct4* für die Pluripotenz in somatischen Stammzellen vor [134]. Daher stellt sich die Frage, ob die Mechanismen der Selbsterneuerung und Differenzierbarkeit von ESC direkt auf adulte Stammzellen übertragbar sind, und welche Aussagekraft der unstenen bzw. schwachen Expression der klassischen ESC Markergene zugestanden werden soll. Die übrigen MSC typischen Transkripte (*fgf2*, *myc*, *klf4*, und *yy1/2*), die ebenfalls mit Stammzeleigenschaften assoziiert sind, konnten per RT-PCR zweifelsfrei nachgewiesen werden. Durch die transkriptomweite Sequenzierung können all

diese Markergene auf die Stärke ihrer Expression hin überprüft werden (siehe Tabelle 3 - 11, Spalte „MSC_neu“). Vergleicht man die Ergebnisse der RT-PCR mit den Expressionsstärken wie sie durch Sequenzierung und normalisierten „read count“ Ansatz ermittelt wurden (siehe Punkt 3.7.1), lassen sich unabhängig von der gewählten Analysemethode der Rohdaten (Kartierung | BLASTn Sequenzen vs. UnigeneDB | BLASTn Contigs vs. RefSeqRNA) keine Widersprüche zum Gelbild feststellen. Für die Gene *oct4*, *nanog* und *sox2* wurden kaum zuordenbaren Sequenzen ermittelt. Für *lif* wurde nur eine äußerst geringe Transkription in den jungen MSC detektiert. Das CD45 Antigen konnte ebenfalls nicht per cDNA Sequenzierung in den jungen MSC, nachgewiesen werden. Der Ansatz der alten MSC weist hingegen überraschend eine schwache Expression des *CD45* Gens auf, hier liegen keine Vergleichsdaten aus der RT-PCR vor. Betrachtet man die Expression der MSC typischen Gene in den differenzierten Zelllinien so fällt auf, daß keines dieser Markergene ausschließlich in undifferenzierten Zellen exprimiert wird. Für alle untersuchten, MSC assoziierten Gene außer *CD44* und *lif* wurde sogar eine höhere Expressionsstärke in den differenzierten als in den undifferenzierten Zellen ermittelt. Ein spezifisches Transkript, das ausschließlich in undifferenzierten MSC vorkommt, und somit einen eindeutigen Marker darstellen würde, ist unter den untersuchten Genen nicht vorhanden. Für zukünftige Versuche bietet sich daher zur Charakterisierung undifferenzierter MSC vor Beginn teurer und zeitaufwändiger Differenzierungen die bereits bewährte immunhistologische Untersuchung der Oberflächenantigene CD44/45 an. Aus den in dieser Arbeit gewonnenen Sequenzdaten lassen sich zwei neue Kandidatengene *nkd1* und *uhrf1* als MSC typische Marker ableiten. Für die Gene wurden per Kartierung gegen das Genom RPKM Werte von 13,37 (*nkd1*) und 12,30 (*uhrf1*) berechnet, wohingegen keine andere der sequenzierten Zelllinien einen Expressionswert von 1 RPKM aufweist. Das Gen *nkd1* ist ein wichtiger Antagonist des kanonischen Teils des WNT Signaltransduktionsweges [135, 136], der bekannt ist für die Steuerung von Differenzierungsprozessen in Stammzellen. Auf die Rolle des WNT Signalweges für die Differenzierung von MSC wird unter Punkt 4.10.1 im Detail eingegangen. Das Gen *uhrf1* ist aktiv an der Methylierung, und damit der Stilllegung von Promotorstrukturen beteiligt. Sen und Kollegen postulierten 2010 für somatische Stammzellen ein epigenetisch fixiertes Methylierungsmuster, welches die Aktivität von Differenzierungsgenen einschränkt und so den Stammzellstatus erhält [137]. Das Ringfingerprotein UHRF1 bindet an die DNA-Methyltransferase-1 (DNMT1) und leitet die Methylierung hemimethylierter Promotorstrukturen durch

Rekrutierung des entsprechenden Protein-Komplexes an die DNA ein. In epidermalen Progenitorzellen konnte durch Sen und Kollegen in RNAi Experimenten mittels Stilllegung von *dnmt1* eine massive Steigerung der spontanen Differenzierung gezeigt werden. Zudem war eine Promotor-Methylierung von 364 aus 1366 mit Differenzierungsprozessen assoziierten Genen in undifferenzierten Zellen nachweisbar. Über 50 % dieser Gene zeigte keine Methylierung mehr in differenzierten Zellen. In den sequenzierten Zelllinien dieser Arbeit liegt die normalisierte Expression von *dnmt1* in undifferenzierten, jungen MSC (RPKM 18) zwischen 1,5 bis 6 mal über der differenzierten Zelllinien (z.B. Adipozyten RPKM 12, alte MSC RPKM 3). Zusammengefasst lassen die Daten auf eine erfolgreiche Etablierung von Primärkulturen mesenchymaler Stammzellen aus den Röhrenknochen des Rindes schließen.

4.2 *IN VITRO* DIFFERENZIERUNG VON MSC UND GENEXPRESSIONS-ANALYSE DIFFERENZIERUNGSSPEZIFISCHER MARKERGENE

Nach Erreichen konfluenten Wachstums wurden die Primärzellen zu Adipozyten, Osteoblasten und Chondrozyten differenziert. Da während der Differenzierungen keine Passagen der Zellen im Differenzierungsmedium durchgeführt wurden, verblieben undifferenzierte Kontrollen ebenfalls unpassagiert für viele Tage im Standardmedium. Mit Ausnahme der Chondrozyten bilden differenzierte wie undifferenzierte Zellen unter solchen Bedingungen einen mehrschichtigen, gewebeähnlichen Zellverband am Boden der Kulturschale (sog. „*multilayer*“) aus. Durch dieses Vorgehen soll gewährleistet werden, daß Zellen mit ähnlicher Kulturzeitdauer und ähnlichen Wachstumsbedingungen verglichen werden können. Für die Färbeversuche zum Nachweis der Differenzierung wurde je ein Kontrollansatz im gleichen Alter wie die differenzierten MSC gefärbt. Die Kontrollen der Differenzierungsansätze tragen nachfolgend die Bezeichnung „alte MSC“. Nach der cDNA Sequenzierung konnte auch die Expression differenzierungsspezifischer Markergene überprüft werden (siehe Punkt 3.7.1). Ideal wäre sicher die cDNA Sequenzierung einer exakt gleich alten Kontrolle („alte MSC“) je Differenzierung gewesen, dies hätte jedoch hohe Kosten verursacht. Bedenkt man beispielweise den zeitlichen Unterschied in der Kulturdauer zwischen osteogener Differenzierung (25 Tage), adipogener Differenzierung (28 Tage) und den alten MSC, deren cDNA sequenziert wurde (31 Tage), ist zudem fraglich, ob

mehr cDNA Sequenzierungen von Kontrollen in diesem Zeitfenster überhaupt einen Unterschied im Expressionsprofil aufgewiesen hätten (Details siehe auch Tabelle 3 - 2). Um auch das Expressionsprofil möglichst ursprünglicher MSC untersuchen zu können, wurde cDNA aus einem gerade konfluent gewachsenem Ansatz am vierzehnten Tag nach der Isolation ebenfalls sequenziert (Bezeichnung „junge MSC“). Für Genexpressionsanalysen ist letzterer Datensatz die Referenz undifferenzierter MSC. Da im Laufe der Arbeit immer mehr signifikante Unterschiede zwischen den alten und den jungen MSC beobachtet werden konnten, bilden diese undifferenzierten Ansätze strikt voneinander zu trennende Kontrollen.

4.2.1 ADIPOGENE DIFFERENZIERUNG VON MSC

Im Falle der adipogenen Differenzierung lassen sich runde, gelblich-lichtbrechende Einschlüsse in den Zellen bereits nach wenigen Tagen erkennen. Es steht zu vermuten, daß es sich um Fetttröpfchen handelt. Leider fielen die Ergebnisse der Färbung mit „Oil Red O“ im Vergleich Differenzierung/Kontrolle nicht so eindeutig aus, als daß von einem zweifelsfreien Nachweis der Differenzierung mittels histologischer Methoden gesprochen werden kann (siehe Punkt 3.3.1). Dies scheint aber eher ein technisches als ein biologisches Problem zu sein. Womöglich hat die Art der Fixierung zum Verlust der Fetttröpfchen durch Extraktion geführt hat, da die entsprechenden Einschlüsse im Laufe der Färbeprozedur sichtbar weniger wurden. Verschiedene Versuche, die Färbetechnik zu modifizieren, lieferten nur wenig bessere Ergebnisse (Diplomarbeit Jan Rettich, Daten nicht gezeigt). Durch eigene und im Rahmen der Diplomarbeit von Anne Dubberke durchgeführte quantitative RT-PCR Messungen zur Ermittlung der Expressionsstärke der adipogenen Markergene *fabp4* und *ppar-γ2* konnte jedoch eine erfolgreiche Differenzierung der MSC zu Adipozyten mit dem verwendeten Differenzierungsprotokoll bewiesen werden (Daten nicht gezeigt). Die cDNA Sequenzierung bestätigt diese Messwerte, alle betrachteten Markergene werden in den zu Adipozyten differenzierten MSC wesentlich stärker als in allen übrigen Zelllinien exprimiert (siehe Tabelle 3 - 11). Lediglich für die Markergene *fabp4* und *pparg* konnte mit einer der drei Methoden (siehe Punkt 3.6) zur Bestimmung der Genexpression ein unwesentlich höherer Wert in Osteoblasten als in Adipozyten bestimmt werden. Aufgrund des nur geringen Expressionsunterschiedes und der Tatsache, daß mittels Kartierung gegen das Genom und BLASTn Abgleich der Einzelsequenzen mit der UniGene Datenbank die stärkste Expression stets bei den

Adipozyten lag, kann dieser Umstand allerdings vernachlässigt werden. Da die „Oil Red O“ Färbung wiederholt Schwierigkeiten verursacht hat, sollte ein Umschwenken auf Sudan III Fettfärbung für zukünftige Versuche in Erwägung gezogen werden [138]. Die Differenzierung der MSC zu Adipozyten konnte im Rahmen dieser Arbeit erfolgreich durchgeführt werden.

4.2.2 OSTEOGENE DIFFERENZIERUNG VON MSC

Auch die *in vitro* Differenzierung der MSC zu Osteoblasten geht mit gut sichtbaren morphologischen Veränderungen einher. Die lokal begrenzten Ablösungen des Zellrasens sind auch mit bloßem Auge zu erkennen. Sie traten jedoch nicht in allen Differenzierungsansätzen auf und sind vermutlich in der Abnahme der Plastikadherenz im Verlauf der Differenzierung begründet. Die Ablösungen scheinen keinen Einfluß auf die Vitalität oder den Verlauf der Differenzierung zu nehmen. Die Färbeversuche mit „Alizarin Red“ zeigen eine deutliche, flächendeckende Mineralisierung der Matrix an, die in den Kontrollen nicht oder nur wesentlich schwächer und sehr vereinzelt auftritt. Dies deckt sich mit den Ergebnissen der Expression osteogener Markergene (siehe Tabelle 3 - 11). Für alle getesteten Marker konnte eine signifikant erhöhte Expression in den Osteoblasten festgestellt werden, die eindeutig auf eine erfolgreiche osteogene Differenzierung der MSC rückschließen lässt.

4.2.3 CHONDROGENE DIFFERENZIERUNG VON MSC

Die chondrogen differenzierten Zellen zeigten bereits kurz nach Beginn der *in vitro* Differenzierung morphologische Veränderungen, eine Kondensation war im Lichtmikroskop deutlich zu erkennen. Am auffälligsten war jedoch die Ablösung der Zellen vom Untergrund der Kulturschale und der damit verbundene Verlust von Zellen durch Auswaschen beim Medienwechsel. Um dem entgegenzuwirken wurde versucht, die Kulturbedingungen für die chondrogene Differenzierung abzuändern. In der Literatur werden für Chondrozyten unterschiedliche, von der Standardzellkultur abweichende Anzucht und Differenzierungsmethoden beschrieben. Leider scheiterten sowohl der Versuch die MSC Anzucht und Differenzierung im Pelletkultur - Verfahren zu etablieren [139], als auch das Einbetten der Zellen in Agarose [131] (Ergebnisse nicht gezeigt). Beide Methoden

zielen darauf ab, für Chondrozyten natürlichere Wachstumsbedingungen bereit zu stellen, als das die Standard Monolayer Zellkultur vermag. Auch der in dieser Arbeit durchgeführte Versuch, der mit dem Beginn der Differenzierung abnehmenden Adhärenz der Zellen durch Collagen beschichtete Zellkulturschalen entgegen zu wirken, führte zu keiner nennenswerten Verringerung der Ablösung der Zellen vom Boden der Kulturgefäße. Ein Wechsel des die Chondrogenese treibenden Proteins TGF- β vom Typ β 1 auf β 3 hatte ebenfalls keine nennenswerte Veränderung des Verhaltens der Zellen unter induktiven Bedingungen zur Folge. Ein Wechsel des Grundmediums sowie des darin enthaltenen Serumgehalts war genauso wenig in der Lage, die Ablösung der Zellen aufzuhalten. Erstaunlicherweise konnte dennoch im Rahmen von Diplomarbeiten eine für Chondrozyten spezifische Färbung mittels Alcian Blue erfolgreich durchgeführt werden (siehe Punkt 3.3.3). Auch die Expression der chondrogenen Markergene *comp*, *biglycan* und *aggrecan* konnte in der Diplomarbeit von Anne Dubberke per quantitativer RT-PCR unter Differenzierungsbedingungen nach 8-9 Tagen nachgewiesen werden. Es wurde daher entschieden, unter Inkaufnahme einer verhältnismäßig kurzen Differenzierungszeit eine Sequenzierung der RNA aus den chondrogen induzierten Zellen durchzuführen. Durch Abbruch der Differenzierung nach 5 Tagen befanden sich die Chondrozyten vermutlich noch nicht im voll ausdifferenzierten Zustand. Eine Untersuchung der Genexpression dieser Zellen wäre im Hinblick auf neue Erkenntnisse zur chondrogenen Differenzierung dennoch interessant. Bosnakowski und Kollegen [70] konnten für bovine MSC einen signifikanten Anstieg der Expression der chondrogenen Markergene *collagen II alpha I*, *aggrecan* und *sox9* ab Tag 3 bis zum Tag 6 der Differenzierung unter ähnlichen Versuchsbedingungen messen. Ein Blick auf die Expression chondrogener Marker der eigenen Zelllinie zeigt leider eindeutig, daß die Zellen keinerlei Kennzeichen chondrogener Differenzierung aufweisen. Dies lässt den Rückschluss zu, daß die Zellen im Rahmen der Differenzierung eine nicht näher zu bestimmende Veränderung der extrazellulären Matrix durchführen, welche einen falsch positiven Nachweis mittels Alcian Blue zur Folge hat. Vorstellbar wäre, daß die Differenzierung zu Chondrozyten möglicherweise unvollständig abläuft, und die extrazelluläre Matrix durchaus anfärbbare, saure Proteoglycane enthält. Ein spontaner Abbruch der Differenzierung zu einem zufälligen Zeitpunkt ist eine Hypothese, mit der sich auch die vormals erfolgreich durchgeführten qRT-PCR Experimente erklären ließen. Denkbar wäre auch ein altersbedingter Effekt auf das Differenzierungspotential der MSC. In Studien an humanen MSC Linien konnte

gezeigt werden, daß besonders das chondrogene Differenzierungspotential der MSC aus dem Knochenmark mit dem Alter des Donors abnimmt [140, 141]. Leider konnte das Alter der Tiere bei Schlachtung im Rahmen der Kooperation mit den Schlachthöfen nicht immer festgehalten werden, da eine exakte Dokumentation dem Personal und Tierärzten im täglichen Arbeitsablauf der Großschlachtereien nicht zuzumuten war. Zwar wurde immer nach Material von frisch geschlachteten Kälbern gefragt, dies stand jedoch nur in Ausnahmefällen zur Verfügung. Somit ist durchaus möglich, daß unterschiedliche Primärkulturen aus unterschiedlichen Tieren unterschiedlich potent differenzierbar sind. Eine Standardisierung der Abläufe und umfangreichere Dokumentation wäre für nachfolgende Versuche sicher wünschenswert. Des Weiteren gilt es zu bedenken, daß MSC vermutlich generell als heterogenes Gemisch unterschiedlich potenter Progenitorzellen im Knochenmark vorkommen (siehe Abbildung 4 - 1), ein Modell, das von Baksh und Kollegen vertreten wird [142].

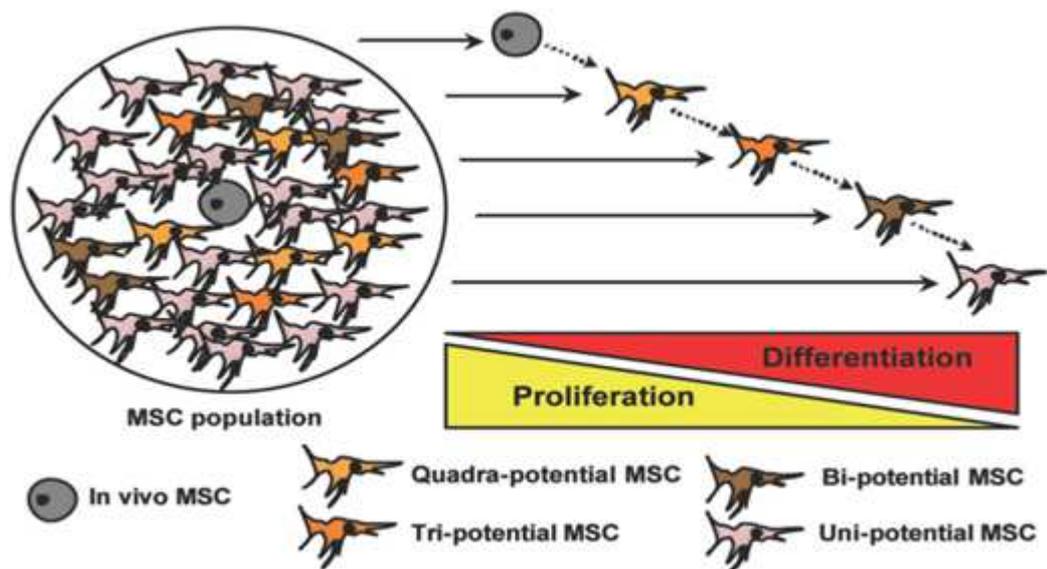


Abbildung 4 - 1 Modell eines heterogenen MSC Gemischs in vivo

Modell zum Vorkommen von hinsichtlich ihres Differenzierungspotentials unterschiedlich potenter MSC Subpopulationen. Eine unbeabsichtigte Selektion von zur Chondrogenese nicht befähigten Progenitorzellen im Rahmen der Isolation wäre denkbar. Auch ein vollständiges Fehlen von tripotenten MSC in den Primärzellen könnte der Grund für die mangelnde Chondrogenese sein. Abbildung verändert nach Baksh et al [142].

Möglicherweise stellten bipotente MSC in den chondrogen differenzierten Primärzellen einen erhöhten Anteil dar, deren Differenzierungspotential auf Adipogenese und Osteogenese beschränkt ist. Die Existenz genau solcher

Subpopulationen wurde von Post und Kollegen bereits beschrieben [143]. Das für MSC klassischerweise geforderte trilineare Differenzierungspotential konnte somit für die in dieser Arbeit isolierten Zellen leider nicht belegt werden.

4.2.4 LANGZEIT-ZELLKULTUR UNDIFFERENZierter MSC IM MULTILAYER

Für MSC sind Veränderungen der Genexpression, der generellen Differenzierbarkeit der Zellen bis hin zu spontanen Differenzierungsprozessen in Zusammenhang mit der Kulturdauer mehrfach beschrieben worden [144, 145]. Die Zellen weisen also eine erhebliche Variabilität unter Zellkulturbedingungen auf, der im Rahmen von Genexpressionsanalysen zwingend Rechnung getragen werden muss. Als Kontrolle für die histologischen Färbungen wurden MSC in Standardkulturmedium unter identischen Wachstumsbedingungen wie die differenzierten Zellen im Multilayer kultiviert. Frisch isolierte Zellen standen nur unmittelbar nach den Schlachthofbesuchen, und auch dann nur in geringen Mengen zur Verfügung. Zudem hätte man etwaige Effekte der verhältnismäßig langen Kulturdauer und des Wachstums im vielschichtigen, gewebeähnlichen Verband auch ohne Einsatz von Differenzierungsmedium mit frisch isolierten Zellen nicht sichtbar machen können. Auch für die Microarrayexperimente und die cDNA Sequenzierung wurden daher die genannten Kontrollen (MSC alt) für die vergleichende Genexpressionsanalyse herangezogen. In den histologischen Färbungen wiesen die Kontrollen keine Ergebnisse auf, die auf eine spontane Differenzierung hindeuten würden. Die Analyse von Markergenen zeigt jedoch eine deutliche Veränderung der Genexpression der Zellen im Vergleich zu den jungen MSC an. Es lassen sich kaum Ähnlichkeiten in der Expression der in Tabelle 3 - 11 aufgeführten Markergene im Vergleich der beiden undifferenzierten Stammzelllinien feststellen. Besonders auffällig ist die verhältnismäßig starke Expression osteogener Marker in den alten MSC (vergleiche auch Punkt 4.5). So wird für die Expression des Matrix Strukturproteines Osteopontin (*spp1* Gen) ein zur Expression in Osteoblasten fast identischer, signifikant erhöhter Wert festgestellt. Eine erhöhte Expression lässt sich auch für die Gene *bglap* und *ibsp* beobachten, auch wenn hier die Expression in den alten MSC nicht an die in den Osteoblasten heran reicht. Für die übrigen Zelllinien werden keine erhöhten Expressionswerte osteogener Marker ermittelt. Eine weitere Auffälligkeit ist der im Vergleich zu den jungen MSC erniedrigte Expressionswert für das CD44 Antigen, eines MSC typischen Hyaluronsäure Rezeptors. Dieser wird in jungen MSC noch

sehr viel stärker als in den übrigen Zelllinien exprimiert, in den alten MSC sinkt das Expressionsniveau auf ein eher mit dem der Osteoblasten vergleichbares Maß ab. Es ließen sich also eine starke Veränderung der Genexpression und eine klare Disposition hin zur spontanen osteogenen Differenzierung in den alten MSC beobachten. Der Grad der Ähnlichkeit zwischen den *in vitro* differenzierten Osteoblasten und den alten MSC bleibt allerdings fraglich, trotz gleich langer Kulturzeiten im Multilayer lässt sich eine mineralisierte Matrix in den alten MSC histologisch nicht nachweisen. Eine vollständige spontane Differenzierung der alten MSC zu Osteoblasten lässt sich anhand der Expression der Markergene und Färbeversuche nicht feststellen. Die prinzipielle Fähigkeit von MSC in speziellen, dreidimensionalen Kulturen spontan osteogen zu differenzieren, wurde jedoch bereits von Tseng und Kollegen [146] beschrieben. Als auslösende Mechanismen, die zur spontanen Differenzierung führten, wurden eine Kondensation des Cytoskelettes und verstärkte Mikrofilament Kontraktionen beschrieben. Wurden diese Vorgänge durch Zugabe von Latrunculin B (verhindert Actin – Polymerisation) und Blebbistatin (Myosin II Blocker) unterbunden, blieb die osteogene Differenzierung aus. Dieser Mechanismus könnte auch in der Langzeit Multilayer Kultur der alten MSC für die beobachteten Veränderungen verantwortlich sein. Allerdings sind die Kulturbedingungen bei Tseng und Kollegen zur Erstellung einer dreidimensionalen Zellkultur wesentlich aufwändiger als ein einfaches Wachstum im mehrschichtigen, gewebeähnlichen Verband. Dies könnte der Grund für die nicht vollständige Differenzierung der alten MSC hin zu Osteoblasten sein.

4.2.5 RNA PRÄPARATION UND RNA AMPLIFIKATION

Sowohl für die cDNA Sequenzierung, als auch für die Microarray Experimente wurde eine große Menge RNA benötigt. Das „Cologne Center for Genomics“ setzte die extrem hohe Menge von 5 µg cDNA zur Sequenzierung voraus, jedes Microarray Experiment benötigte pro Objektträger zwei mal 2 µg mRNA Einsatz. Sowohl differenzierte, als auch undifferenzierte Zellen standen jedoch für eine RNA Extraktion nur in begrenzter Menge zur Verfügung, was eine *in vitro* Amplifikation der vorhandenen RNA zwingend erforderlich machte. Für eine RNA Amplifikation ist als Ausgangsmaterial Gesamt - RNA von sehr guter Qualität in unter Normbedingungen problemlos isolierbarer Menge vonnöten. Die Qualität und Menge der isolierten total - RNA stellte kein Problem dar (siehe Abbildung 3 - 8).

Die beiden Amplifikationsschritte konnte ebenfalls erfolgreich durchgeführt werden (siehe Abbildung 3 - 9 und Tabelle 3 - 3), wobei die erste Runde wesentlich effizienter war als die zweite Runde der Amplifikation. Die große Menge an Matrizen-RNA dürfte zu einer Limitierung der für die IVT benötigten Ressourcen im Verlauf der 12 stündigen Reaktion geführt haben, was zu der im Vergleich zur ersten Runde um mindestens eine Potenz geringeren Amplifikationseffizienz geführt hat. Die aRNA der Chondrozyten zeigt eine zu kleineren Fragmenten verschobene Größenverteilung, die zudem im Bereich der längeren Fragmente im Vergleich zu den anderen Amplifikationen stärker abfällt. Dennoch sind in allen Amplifikationen sowohl ausreichende Mengen, als auch ausreichend große Fragmente für die nachfolgenden Versuche synthetisiert worden. Für die Hochdurchsatz Sequenzierungen wird vom Anbieter eine Größenfraktionierung vor dem Erstellen der Banken durchgeführt, so daß die kleinen Fragmente die Qualität der Daten nicht beeinträchtigen. Bei der Synthese und Markierung von cDNA für Microarray Experimente finden zwei Aufreinigungen der Nucleinsäure mittels Silica-Membranen statt. Kleine DNA-Fragmente (< 100 bp) binden wenig effizient an diese Membranen, so daß auch hier keine negative Beeinträchtigung der Experimente bei einer Hybridisierungstemperatur von (42°C) im geeigneten Hybridisierungspuffersystem zu erwarten ist. Die reverse Transkription mittels RT und OligodT Primer und auch die für Osteoblasten, alte MSC und Chondrozyten zwei Mal nacheinander durchgeführte in vitro Amplifikation von RNA, bergen technisch bedingt das Risiko, keine Vollängen – aRNA zu erzeugen. Es werden so bevorzugt dem 3' Ende näher liegende Teile der Transkripte angereichert. Dies kann für die weiterführende Analyse der Genexpression Auswirkungen haben. Für die Microarray Hybridisierung stünden im ungünstigen Fall vermehrt kurze und 3'UTR haltige cDNA Moleküle zur Verfügung. Hier kommt dem Experiment in dieser Arbeit aber der Umstand zugute, daß die Sonden auf dem Microarray cDNA haltige Plasmide sind - und damit prinzipiell ebenfalls 3'OH nahe Sonden vermehrt vorkommen. Durch die während der EST Bank Herstellung aus polyA+ angereicherter RNA durch Steffen Rapp zusätzlich zum OligodT Primer verwendeten Hexamer Primer, wird die Selektion 3' naher EST allerdings sicher weniger stark ausfallen, als dies bei alleiniger Nutzung eines OligodT Primers der Fall gewesen wäre. Eine etwaige Beeinflussung der Array Ergebnisse bedingt durch die Art der RNA Aufbereitung bleibt daher schwer abschätzbar. Bei der Sequenzierung von 3' lastiger cDNA kann nach Normalisierung auf Transkriptlängen (RPKM – Werte) für solche Gene, die länger sind als das

sequenzierbare Material reicht, ungenaue Werte liefern. Der mathematische Ansatz der RPKM Normalisierung beruht darauf, daß lange cDNA Fragmente bei der Sequenzierung mehr Sequenzen ergeben als eine gleich stark exprimierte, aber kürzere cDNA. Die setzt allerdings die Sequenzierung auch der langen Transkripte in voller Länge voraus.

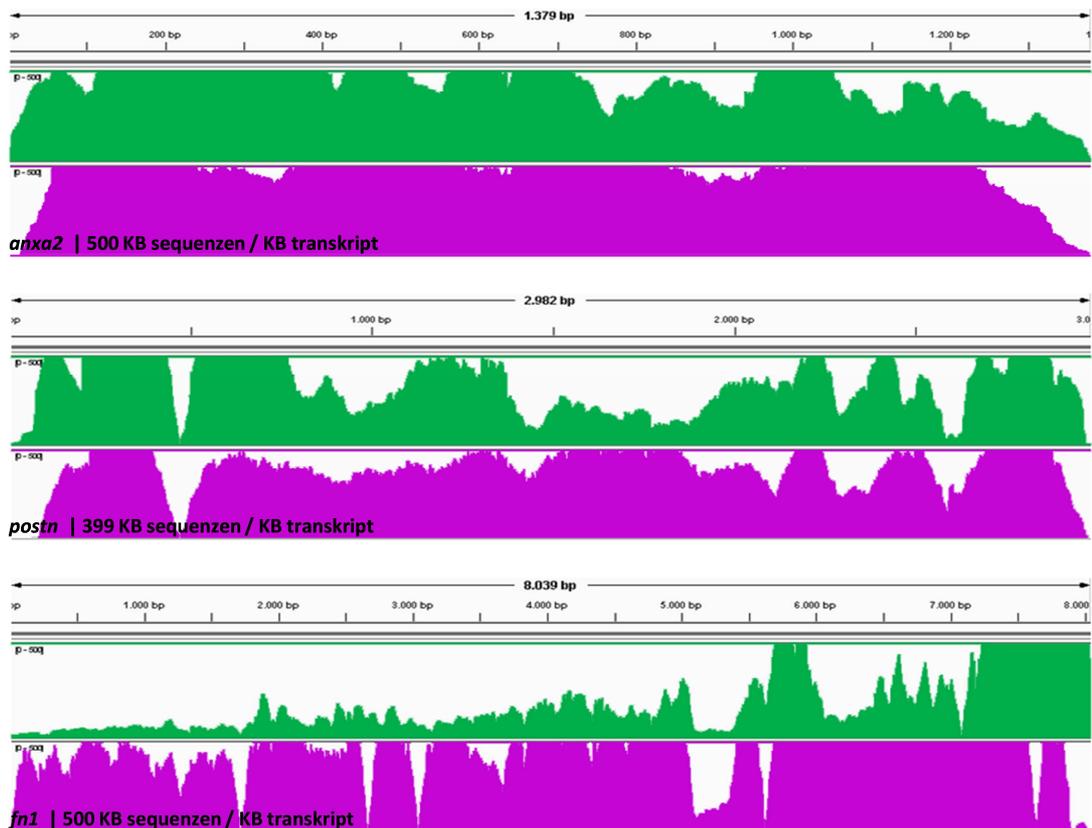


Abbildung 4 - 2 Kartierung zum Vergleich amplifizierter / nicht amplifizierter RNA

Abdeckung unterschiedlich langer cDNA Referenzsequenzen nach Bowtie1 Kartierung mit identischen Transkriptmengen (KBp) amplifizierter (alte MSC, grün) und nicht amplifizierter (junge MSC, lila) RNA. Grafik begrenzt auf maximal 500 fache Abdeckung. Transkripte aufgeführt in 5' nach 3' Richtung.

Für die korrekte Bestimmung der Genexpressionsstärke sind möglichst komplett sequenzierte Transkripte erforderlich, besonders da in dieser Arbeit die Genexpression auch zwischen amplifiziertem und nicht amplifiziertem Material verglichen werden soll. Eine Abschätzung des Effektes der RNA Amplifikation auf die Abdeckung unterschiedlich langer Transkripte ermöglicht die obige Abbildung. Die Transkriptlänge, berechnet anhand der Angaben für den längsten Eintrag jedes Clusters der NCBI Unigene Datenbank, beträgt gemittelt über sämtliche 43489

bovinen Einträge 1,387 KB bei einer Standardabweichung von 1,449 KB. Diese Angaben decken sich mit in der Literatur für Säugetiere publizierten Werten [147]. Daran angelehnt wurde die Abdeckung des Gens *anxa2*, dessen cDNA Länge dem genannten Mittelwert in etwa entspricht, im Vergleich nach Kartierung identischer cDNA Mengen aus Daten von amplifizierter (alte MSC) und nicht amplifizierter RNA (junge MSC) überprüft (siehe Abbildung 4 - 2, erster Teil). Es lassen sich für *anxa2* über die gesamte Transkriptlänge keine gravierenden Unterschiede in der Abdeckung zwischen den Sequenzierungen feststellen. Ähnliches gilt für die Verteilung der kartierten Sequenzen über das *postn* Gen, dessen Länge mit 2982 bp laut Unigene Datenbank in etwa dem Mittelwert zuzüglich der Standardabweichung entspricht. Auch hier werden für beide Sequenzierungen große Mengen Sequenzen über die gesamte Länge kartiert. Über 90 % der Transkripte in der Unigene Datenbank sind kürzer als das *postn* Gen. Einen deutlichen Effekt der doppelten Amplifikation lässt sich jedoch anhand des Beispiels der Kartierung des *fn1* Gens mit seiner verhältnismäßig langen cDNA von über 8000 bp feststellen. Für die Daten aus der Sequenzierung der alten MSC zeigt sich sehr deutlich eine verstärkte Kartierung im 3' Bereich der *fn1* cDNA. Diese Verteilung der Sequenzen über das gesamte Transkript sieht für die jungen MSC Sequenzierungen sehr viel homogener aus. Bei letzterem Zelltyp wurde die Sequenzierung im eigenen Haus auf einem moderneren Illumina Sequenzierer und mit aktualisierten Protokollen zur cDNA Bank Erstellung aus geringen RNA Mengen durchgeführt. Dies machte eine Amplifikation der RNA vor der Sequenzierung unnötig. Für lange Transkripte (> *postn* mit 2982 bp) ist daher eine schwer einschätzbare technische Beeinträchtigung der ermittelten Expressionswerte für Osteoblasten, alte MSC und die Chondrozyten zu befürchten. Dem wurde durch die Bevorzugung der TER (*Total Exon Reads*) statt der RPKM Werte als Maß der Genexpression in vielen Analysen Rechnung getragen. Zudem dürfte der Effekt für die Genexpressionsanalyse insgesamt betrachtet nicht allzu starken Einfluss haben. In den Ergebnissen der BLASTn Analyse der alten MSC gegen die Unigene Datenbank beträgt der Anteil der Cluster mit über 2982 bp Länge und mehr als 1000 zuordenbaren Sequenzen nur 2,9 %.

4.3 BESTIMMUNG DIFFERENZIELLER GENEXPRESSION MITTELS PLASMID - DNA MICROARRAYS

Bei den in dieser Arbeit verwendeten Microarrays kamen als Sonden gespottete Plasmide mit EST Molekülen als genspezifisches Insert zum Einsatz. Ziel ist die Bestimmung differenzieller Genexpression durch kompetitive cDNA Hybridisierung. Plasmid DNA - Microarraytechnik hat sich bereits in vorangegangenen Arbeiten bewährt (Eigene Diplomarbeit 2006, Diplomarbeiten von Christian Pick 2006, Valentin Färber 2008). Grundsätzlich kommen die bereits dort beschriebenen Vor- und Nachteile der Technik auch in dieser Arbeit zum Tragen. Der Umstand, daß beispielsweise im Vergleich zu gespotteten PCR Produkten oder Oligonucleotid Microarrays, nur der Integrierte Anteil der Plasmidsonden DNA ursprünglich auf Transkripte zurück geht, bedeutet potentiell eine limitierte Verfügbarkeit der genspezifischen Sondensequenz während der Hybridisierung. Diesem Effekt kann im Rahmen der DNA-Bindekapazität der Microarrays durch eine möglichst hohe Sondenkonzentration und großflächige Spots in moderatem Abstand entgegengewirkt werden. Die Herstellung der Microarrays kann verhältnismäßig kostengünstig erfolgen, in diesem Fall stand eine sequenzierte EST Bank bereits zur Retransformation aus einem vorangegangenen Projekt bereit [82] – nahezu alle gewünschten Sonden konnten erfolgreich dreifach redundant auf dem Microarray untergebracht werden. Die Anzahl und Art der analysierbaren Gene beschränkt sich natürlich auf jene ca. 900 unterschiedlichen Gene, die in der EST Bank auch enthalten sind. Die Masse jeder Plasmidsonde, die zum Spotten zur Verfügung steht, hängt von der Effizienz der Plasmidisolierung im 96er Format ab. Diese unterliegt in Ihrer Ausbeute einer gewissen Schwankung, wie zufällige Messungen der Sondenkonzentration zeigten (siehe Abbildung 3 - 10 A/B). Die Isolation der Plasmidsonden der in die Mitte des Objektträgers gespotteten Subarrays weist etwas geringere Ausbeuten auf, als dies für die an den beiden Enden des Arrays liegenden Sonden beobachtet werden kann (siehe Abbildung 3 - 11). Dies deutet auf eine systematische Abweichung hin, und führt in den Hybridisierungen zu insgesamt schwächeren Signalen von den betroffenen Sonden. Grund könnte ein Problem mit dem Wachstumsmedium dieser Anzuchtcharge sein. Da das Verhältnis der Signale aus beiden cDNA Markierungen jedoch davon unbeeinträchtigt bleibt, ist dieser qualitative Mangel für die Ermittlung differenzieller Genexpression jedoch nur dann von Belang, wenn die Signalintensität der Sonden sich nur unzureichend stark vom Hintergrund

unterscheiden würde. Die Nachweisgrenze für die Bestimmung differenzieller Genexpression schwach exprimierte Gene wird folglich immer dann zu früh erreicht, wenn die Verfügbarkeit der zur markierten cDNA komplementären Sonde unzureichend ist. Erschwerend kommt hinzu, daß der dynamische Messbereich („*dynamic range*“) des verwendeten Microarray Scanners des Typs GenTAC™ LS4 im Vergleich zu modernen Scannern eher beschränkt ausfällt. Schwache Signale können nicht sicher ausgelesen werden, starke Signale erreichen hingegen leider schnell die Sättigungsgrenze [148]. Daher beschränkt sich die Auswahl analysierbarer Gene auf solche, die moderat bis stark exprimiert werden. Dies wird auch daran deutlich, daß die Anzahl der auswertbaren Spots auf einem Array nach der sehr strikten Qualitätskontrolle auf minimal 55 % fallen konnte (Slide 1 Adipozyten/alte MSC). Neben rein technischen Aspekten spielt hier aber die Tatsache, daß durch das Sonden gebende EST Projekt im Gewebe der fötalen Wachstumsfuge nur unzureichend die Genexpression in Adipozyten wiederspiegelt wird, die entscheidende Rolle. Allein 29,5 % der Spots auf dem Array bleiben unter der geforderten Mindestintensität von 1000 im Farbkanal der Adipozyten RNA. Der Sondenbesatz des Microarrays passt hier nicht ideal zum getesteten Zellkulturmaterial. Die Tatsache, daß Sonden mit unterschiedlichen EST's kodierend für dasselbe Gen mehrfach an unterschiedlichen Positionen des Arrays in unabhängigen Hybridisierungen dieselbe differenzielle Genexpression anzeigen, spricht eindeutig für den technischen Erfolg der Produktion und korrekte Bedingungen bei der Hybridisierung der Microarrays. Die Ergebnisse der Microarray Experimente bieten so eine von NGS Techniken unabhängige Methode zur Ermittlung differenzieller Genexpression. Die Interpretation der Array Ergebnisse wird jedoch durch den Umstand verkompliziert, daß die als undifferenzierte Kontrolle hybridisierte RNA aus doppelt amplifiziertem Material des Ansatzes der „alten MSC“ besteht. Alle Ergebnisse in dieser Arbeit deuten darauf hin, daß gerade dieser Zelltyp sich spontan zu Osteoblasten ähnlichen Zellen differenziert hat. Der Microarray Versuchsaufbau wurde jedoch vor Abschluss der NGS Sequenzierung und Datenanalyse so festgelegt. Ziel war es, etwaige Effekte der relativ langen Kulturzeit im Multilayer in den Differenzierungsansätzen durch unter ähnlichen Bedingungen gehaltene Kontrollen auszuschließen. Für alle differenziell exprimierten Gene wird daher nachfolgend auch ein Abgleich mit den für junge MSC im NGS ermittelten Expressionsunterschieden herangezogen, um einen biologisch relevanten Effekt der Expressionsunterschiede anhand nativer, undifferenzierter MSC sicher belegen zu können. Es dienen die mittels BLASTn

gegen die Unigene Datenbank ermittelten Expressionswerte als Vergleichsgrundlage (siehe Punkt 3.6.2). Diese Datenbank enthält, ähnlich wie die Sonden des Microarrays, EST Sequenzen und ist daher für einen Abgleich am ehesten geeignet. Lassen sich die im Array gemessenen Expressionunterschiede zwischen jungen MSC und den Differenzierungsansätzen ebenfalls bestätigen, kann mit einiger Sicherheit von einem für die ursprüngliche Fragestellung (undifferenziert vs. differenziert) relevanten Effekt ausgegangen werden. Unterschiede in der Stärke der differenziellen Expression im Vergleich der beiden Messmethoden sind hierbei technisch bedingt und bis zu einem gewissen Maß tolerierbar, wichtiger ist eine Konstanz bezüglich der Richtung des Expressionsunterschiedes. Aufgrund des mit insgesamt 4 µg sehr hohen Materialbedarfs konnte durch Hybridisierung mehrerer Arrays nur zuvor doppelt amplifizierte RNA aus alten MSC (Kontrolle) mit ebenfalls doppelt amplifizierter RNA der Differenzierungsansätze verglichen werden.

4.3.1 DIFFERENZIELL EXPRIMIERTER GENE IM VERGLEICH UNDIFFERENZIERTER MSC MIT OSTEOLASTEN

Es handelt sich beim Gen *slc25a6* um eine ADP/ATP Pumpe, die den Austausch von cytoplasmatischem ADP mit durch die Mitochondrien metabolisiertem ATP katalysiert [149]. Eine Funktion von *slc25a6* bei der Osteogenese ist bisher nicht beschrieben. Die Genexpression in den jungen MSC liegt mit einem RPKM von 357 (RPKM: 2831 Osteo | 1577 MSC_alt) sogar noch deutlich unter dem für die alten MSC ermittelten Wert. Eine für Osteoblasten auch biologisch relevante Funktion des Expressionsunterschiedes ist somit durchaus wahrscheinlich, es deutet sich ein erhöhter Energiebedarf der Osteoblasten im Vergleich zu den undifferenzierten Zellen an. Dies deckt sich mit den unter Punkt 3.8.3 aufgeführten Ergebnissen, die offenbar systematische Überexpression von Atmungskettenproteinen wird daher im Rahmen des Punktes 4.10.3 näher diskutiert. Das ribosomale Protein S2 wird in jungen MSC mit einem RPKM von 375 (RPKM: 1075 Osteo | 561 MSC_alt) noch deutlich schwächer exprimiert, als in den alten MSC. Auch hier hat sich im Punkt 3.8.3 ein deutliches Muster an in Osteoblasten überexprimierten, ribosomalen Proteinen gezeigt. Dies spricht für eine verstärkte Proteinbiosynthese und soll ebenfalls unter Punkt 4.10.3 eingehender diskutiert werden. Das Gen *enpp1*, welches in Osteoblasten differenziell hoch exprimiert wird, kodiert für eine Pyrophosphatase/Phosphodiesterase und spielt bekanntermaßen eine wichtige Rolle

bei der Regulation der Ossifikation [150]. Funktionell steht *enpp1* mit der Mineralisierung des Gewebes in Zusammenhang, im Mausmodell bewirkt ein Fehlen des Transkripts einen entsprechend schweren Phänotyp. Es wird ein die Pyrophosphat – Homeostase steuerndes Zusammenspiel mit dem typischen Osteoblasten Markergen Alkaline Phosphatase angenommen. Die beobachteten Expressionsunterschiede sind somit plausibel, sie werden durch die NGS Daten auch für die jungen MSC mit einem RPKM Wert von 12 (RPKM: 541 Osteo | 234 MSC_alt) unterstützt. Für das ebenfalls in den differenzierten Zellen stark exprimierte Gen *destrin* wurde ein positiver Einfluss auf den Erhalt von Knochengewebe in Versuchen mit Mikrogravitations-Modellen in der Maus belegt [151]. Destrin ist ein sehr prozessives Protein und dient dem schnellen Abbau von Actinfilamenten, ein Vorgang der für Restrukturierungsprozesse des Zytoskelettes in Osteoblasten unabdingbar ist. In jungen MSC wurde mit einem RPKM Wert von 693 eine deutlich schwächere Expression als in den auf dem Array hybridisieren Zelllinien festgestellt (RPKM: 2146 Osteo | 1170 MSC_alt). Das in den undifferenzierten MSC differenziell exprimierte Gen *sema3c* kodiert für ein sekretiertes Protein, welches zunächst bei der Axon Wegfindung als Signalmolekül beschrieben wurde. Weiterführende Studien an Endothelzellen belegen eine signifikante Verstärkung der Proliferation und gesteigerte Zelladhäsion durch *sema3c* [152], eine Eigenschaft die mit MSC durchaus überein zu bringen ist. In den jungen MSC wird das Gen mit einem RPKM Wert von 30 weniger stark als in den alten MSC exprimiert (RPKM: 79 MSC_alt | 5 Osteo), was aber immer noch signifikant stärker ist, als dies in Osteoblasten beobachtet wurde. Die Werte legen eine mögliche Funktion des Proteins Sema3c für MSC nahe, der genaue Mechanismus lässt sich aus diesen Ergebnissen jedoch nicht ableiten. Das Gen *slc39a1* kodiert für ein Zink Transporter Protein. Während der Differenzierung von MSC zu Osteoblasten wird *slc39a1* in den Vorläuferzellen zunehmend stark exprimiert, eine künstliche Inhibition der Expression behindert hingegen die Differenzierung [153]. Dies könnte als Hinweis für die präosteogene Charakteristik der MSC in Langzeitkultur gewertet werden. Zumindest tendenziell bestätigten die anhand der NGS Daten ermittelten Expressionswerte bislang stets die Werte der Arrayanalyse. Legt man die Werte der NGS Expressionsbestimmung für *slc39a1* zugrunde (RPKM: 11 MSC_alt | 9 Osteo), kann jedoch keine differenzielle Genexpression festgestellt werden. Es wäre denkbar, daß das EST Fragment, welches als Sonde auf dem Microarray dient, eine Transkriptvariante darstellt, die vornehmlich in den alten MSC exprimiert wird. In den jungen MSC beträgt der

RPKM Wert für *slc39a1* jedoch 77, was mit einer Funktion des Gens im Sinne der frühen Osteogenese nur dann überein zu bringen wäre, wenn auch die jungen MSC zumindest teilweise in diese Richtung zu differenzieren beginnen würden. Ohne eine weitere, unabhängige Messung wie der qRT-PCR kann anhand der vorliegenden Daten kein sicherer Schluss aus den für *slc39a1* widersprüchlichen Expressionswerten getroffen werden. Das Gen *gapdh* wurde bereits in vielen Arbeiten als Haushaltsgen zur Normierung verwendet. Es scheint in undifferenzierten MSC gegenüber den Osteoblasten differenziell exprimiert zu werden, was der des öfteren angezweifelte Theorie der bedingungslos gleichbleibenden Expression von Haushaltsgenen entspricht [154]. In den NGS Daten dieser Arbeit schwankt die Expression von *gapdh* zwischen den Zelllinien bis hin zu maximal 4facher differenzieller Expression, wobei die Werte in Osteoblasten am niedrigsten sind. Dennoch wird das Gen in Osteoblasten mit einem RPKM Wert von 562 weit überdurchschnittlich (RPKM ~19) exprimiert. Ein Engpass im Energiestoffwechsel durch Mangel eines wichtigen Glykolyse - Proteins liegt somit in Osteoblasten vermutlich nicht vor. Trotz der Ähnlichkeit der verglichenen Zelllinien ergeben sich im Vergleich zwischen den alten MSC und den Osteoblasten insgesamt schlüssige Ergebnisse.

4.3.2 DIFFERENZIELL EXPRIMIERTER GENE IM VERGLEICH UNDIFFERENZIERTER MSC MIT CHONDROZYTEN

Der Vergleich der Genexpression zwischen Chondrozyten und alten MSC mittels Microarrays stellt einen Sonderfall dar, da aufgrund des osteogenen Grundcharakters der sog. alten MSC und der fehlgegangenen Differenzierung zu Chondrozyten die ursprüngliche Fragestellung mit den Resultaten schlicht nicht beantwortet werden kann. Auch ein Abgleich mit per NGS Analyse ermittelten Daten für junge MSC schafft an dieser Stelle keinen Mehrwert. Dennoch lassen sich interessante Rückschlüsse aus den Daten der Microarrays ziehen. Es wurden ausschließlich in den alten MSC Gene als differenziell stärker als in den vermeintlichen Chondrozyten exprimiert bestimmt, fast alle Expressionsunterschiede werden durch aus dem Unigene BLASTn Abgleich erhaltenen Daten unterstützt. Die Sonden des Microarrays spiegeln die Genexpression in einem Knochen und Knorpel beinhaltendem Gewebe wieder. Ein solches Ergebnis spricht noch einmal klar gegen eine erfolgreiche chondrogene Differenzierung. Die in den alten MSC differenziell stärker exprimierten Gene sind

vor allem solche, die typischerweise mit der Entwicklung von Knochen und Knorpelgewebe assoziiert sind. So werden beispielsweise die Matrixbestandteile *mgp* und *col2a1* [155, 156] als in alten MSC differenziell stärker exprimiert bestimmt, wobei die Expressionsstärke des Kollagens laut der NGS Daten in beiden Zelllinien getrost vernachlässigt werden kann (\approx RPKM 0). Mit dem Gen *spp1*, welches aufgrund der Messdaten aus 19 Sonden als differenziell exprimiert bestimmt wurde, wird ein bekanntes Markergen für die Mineralisierung von extrazellulärer Matrix in den undifferenzierten MSC sehr stark exprimiert [157]. Auch die Expression des Wachstumsfaktors *ctgf* wird in der Literatur als die Proliferation von Osteoblasten und Chondrozyten fördernd und als die Entwicklung des Skelettsystems unterstützend beschrieben [158]. Die Transkription des differenziell exprimierten Gens *cebpd* wird durch den osteogenen Marker *runx2* initiiert. In primären Osteoblasten erhöht der Transkriptionsfaktor *cebpd* seinerseits die Synthese von IGFI [159]. Das für die Entwicklung von Knochenvorläuferzellen wichtige Gen *slc39a1* wird auch im Vergleich zwischen undifferenzierten MSC und Osteoblasten als in den MSC differenziell hoch exprimiert gemessen (siehe auch Punkt 4.3.1). Erneut gibt die Auswertung der NGS Daten keinen Hinweis darauf, daß eine differenzielle Expression des Gens vorliegt, es wird in den Chondrozyten ähnlich stark wie in den alten MSC und den Osteoblasten exprimiert (RPKM: 11 MSC_alt | 12 Chondro). Das Gen *eef1a* kodiert für einen hoch konservierten Elongationsfaktor, der eine zentrale Rolle in der Proteinbiosynthese spielt [160] und somit auf eine in den MSC stärkere Translationsaktivität hindeutet. Ein weiteres die Translation beeinflussendes Gen, welches als in MSC differenziell exprimiert ermittelt wurde, ist *ddx3x*. Es handelt sich dabei um eine RNA Helikase, die unter anderem die Translation von den Zellzyklus steuernden Genen initiiert [161]. Beim dem durch 11 Sonden als differenziell exprimiert bestimmten Gen *mig6* handelt es sich sehr wahrscheinlich um das ribosomale protein L36a (*rpl36a*). Gleicht man die Sequenz des humanen *mig6* Gens (AY305871.1) mittels BLASTx mit der nicht redundanten Proteindatenbank des NCBI ab, erhält man fast ausschließlich signifikante Treffer für das Gen *rpl36a* in unterschiedlichen Spezies. Die UniProt Proteindatenbank (www.uniprot.org) führt die Gennamen *mig6* und *rpl36a* synonym. Auch das ribosomale Protein RPL17 wurde als in den MSC als differenziell exprimiert bestimmt, was wie die Expression von *rpl36a/mig6* und *eef1a1* sowie *ddx3x* für eine starke Proteinbiosyntheseaktivität spricht. Mit dem Gen *atp5g3* wird in den MSC eine Untereinheit der mitochondriellen ATP Synthase differenziell exprimiert [162], der Energiestoffwechsel scheint in den MSC folglich ebenfalls verstärkt statt zu

finden. Zwar würde dies gut zu der scheinbar höheren Translationsaktivität passen, allerdings wird die differenzielle Expression von *atp5g3* laut NGS Daten nicht unterstützt – hier werden sogar leicht höhere Werte für die Genexpression in den Chondrozyten bestimmt (RPKM: 176 MSC_alt | 199 Chondro). Zusammenfassend lässt sich aus den Ergebnissen ableiten, daß die zu vermeintlichen Chondrozyten differenzierten Zellen leider sicher nichts mit Knorpelzellen gemein haben und sich der tendenziell osteogene Charakter der alten MSC bestätigt.

4.3.3 DIFFERENZIELL EXPRIMIERTER GENE IM VERGLEICH UNDIFFERENZIERTER MSC MIT ADIPOZYTEN

Beim Expressionsvergleich der alten MSC mit den Adipozyten ist die geringe Anzahl der auf dem Microarray als differenziell exprimiert bestimmten Gene auffällig. Zwar standen für diesen Ansatz nur 9 statt 10 Microarrays zur Verfügung, dies, wurde jedoch durch eine Lockerung der Kriterien, in wie vielen unterschiedlichen Hybridisierungen ein Gen als differenziell exprimiert ermittelt werden musste, kompensiert (7 aus 9). Ein Grund liegt vermutlich in der Sondenauswahl gekoppelt mit der für die Berechnung zugrunde gelegten Mindestintensität die ein Messwert erreichen muss, um in die Auswertung ein zu gehen. Die als Sonden verwendeten EST Sequenzen aus der Wachstumsfuge von Rinderembryonen dürften ein für Adipozyten untypisches Kollektiv darstellen. Ein Schwellenwert bei der Mindestintensität eines Spots (>1000) hilft im Falle sehr schwacher Signale in beiden verglichenen Zelllinien die Detektion nicht funktioneller Expressionsunterschiede zu verhindern. Im Falle einer nur basalen Expression in einem Zelltyp, bei gleichzeitig durchschnittlicher Expression im verglichenen Zelltyp könnte sich zwar ein Rückschluss auf eine mögliche Funktion des Gens in der exprimierenden Zelllinie aufdrängen. Es wäre aber schwer abschätzbar, ob nur für Knochen Knorpel Entwicklung typische Gene in Adipozyten des Öfteren weniger stark exprimiert werden, oder ob den unterschiedlich stark exprimierten Genen auch biologische Schlüsselfunktionen innewohnen könnten. Daher werden für jeden Vergleich nur in beiden Zelllinien über der Mindestintensität verordnete Sonden ausgewertet. Das Gen *fstl1* kodiert für ein in die extrazelluläre Matrix sezerniertes Glykoprotein und wird in Adipozyten differenziell stärker exprimiert als in MSC (RPKM: 30 MSC_alt | 247 Adipo). In jungen MSC erreicht das Gen einen RPKM Wert von 194. Es scheint generell an Differenzierungsprozessen von MSC beteiligt zu sein [163], für Adipozyten ist ein

Abfall des Expressionslevels mit einer Reifung von Prä-Adipozyten zu Adipozyten nachgewiesen [164]. Eine differenzielle Expression liegt folglich zwischen jungen MSC und den Adipozyten nicht vor, dennoch scheint das Gen aufgrund der Literaturangaben für beide Zelltypen funktionell relevant zu sein. Die Erfassung des typischen Osteoblasten Markers *spp1* als in den alten MSC differenziell stark exprimiert fügt sich nahtlos in die bisherigen Beobachtungen hinsichtlich des osteogenen Potentials der alten MSC ein. Im Vergleich zu den alten MSC spielt das Gen in den jungen MSC mit einem RPKM von 85 eine eher untergeordnete Rolle. Die Ergebnisse des als in Adipozyten differenziell höher exprimiert bestimmten Gens *slc25a6* sind hingegen weniger kongruent. Lee et al. [165] bestimmten die Expression von *slc25a6* in Fettzellen von Schweinen als sehr stark exprimiert, dies deckt sich auch mit dem umfangreichen EST Profilen des entsprechenden Unigene Datenbankeinträge für Schwein und Mensch. Funktionell handelt es sich beim SLC25A6 Protein um einen ADP/ATP Antiporter der Mitochondrienmembran. Eine Einschätzung der Rolle des Proteins im Rahmen dieser Analyse wird dadurch erschwert, daß sich die Microarray Ergebnisse und die NGS Daten massiv widersprechen (RPKM: 1577 MSC_alt | 251 Adipo). In jungen MSC beträgt der RPKM Wert 357 und liegt damit weit näher an den Werten der Adipozyten, als an den alten MSC. Hier wäre eine Überprüfung der Expression mittels quantitativer RT-PCR zur sicheren Bestimmung der Expressionsverhältnisse das beste Vorgehen. Die vorliegenden Daten erlauben keine Rückschlüsse auf eine mögliche Funktion des Gens im Rahmen der adipogenen Differenzierung.

4.4 GENEXPRESSIONSANALYSE MITTELS ILLUMINA HOCHDURCHSATZ SEQUENZIERUNG – PROZESSIERUNG DER ROHDATEN

Ziel dieses Teils der Arbeit ist eine das komplette Transkriptom umfassende Genexpressionsanalyse durch NGS Sequenzierung. Die hierfür nötige Sequenzmenge ist mit konventionellen Ansätzen wie der Sanger Sequenzierung nicht zu erreichen. Es wurden zwischen 3,3 und 10,6 Gigabasen cDNA sequenziert. Die große Diskrepanz in der Datenmenge liegt an den unterschiedlichen Sequenzierstrategien und NGS – Geräten die zum Einsatz kamen, sie ist somit rein technisch bedingt (siehe Tabelle 3 - 7). Mit 3,3 Gibabasen wurden mit Abstand am wenigsten Daten aus der Sequenzierung der cDNA aus adipogen differenzierten MSC gewonnen. Zudem sind die unidirektional sequenzierten Fragmente von 51 bp

Länge von allen Datensätzen die kürzesten. Legt man für statistische Überlegungen die Werte der Kartierung gegen die als Exon annotierten Bereiche (~ 39 MB) der genomischen DNA des Rindes als Zielregion zugrunde, würden rein rechnerisch eine mittlere Abdeckung von 84 X in den Adipozyten erreicht werden können. Die Frage, wie tief sequenziert für eine Expressionsanalyse mittels NGS tief genug ist, haben Blencowe und Kollegen für menschliche Zelllinien 2009 beantwortet [166]. Laut dieser Studie reichen ca. 90 Mio Sequenzen von 35 bp Länge aus, um eine vollständige Quantifizierung der Genexpression aller Gene zu gewährleisten (siehe Abbildung 4 - 3). Berücksichtigt man die sehr steile Steigung im Graphen (blaue Kurve) und die Tatsache, daß Blencowe et al. von einer Sequenzlänge von nur 35 bp ausgehen, so kann auch für die Sequenzierung der Adipozyten von einer annähernd vollständigen Repräsentanz des Transkriptoms ausgegangen werden. Aus Tabelle 3 - 8 geht hervor, daß 43,565 Mio Sequenzen aus Adipozyten mit durchschnittlich 40 bp Länge kartiert wurden. Dies entspräche auf die Länge umgerechnet ca. 50 Mio Sequenzen, womit laut Abbildung 4 - 3 über 90 % des Transkriptoms analysierbar sind.

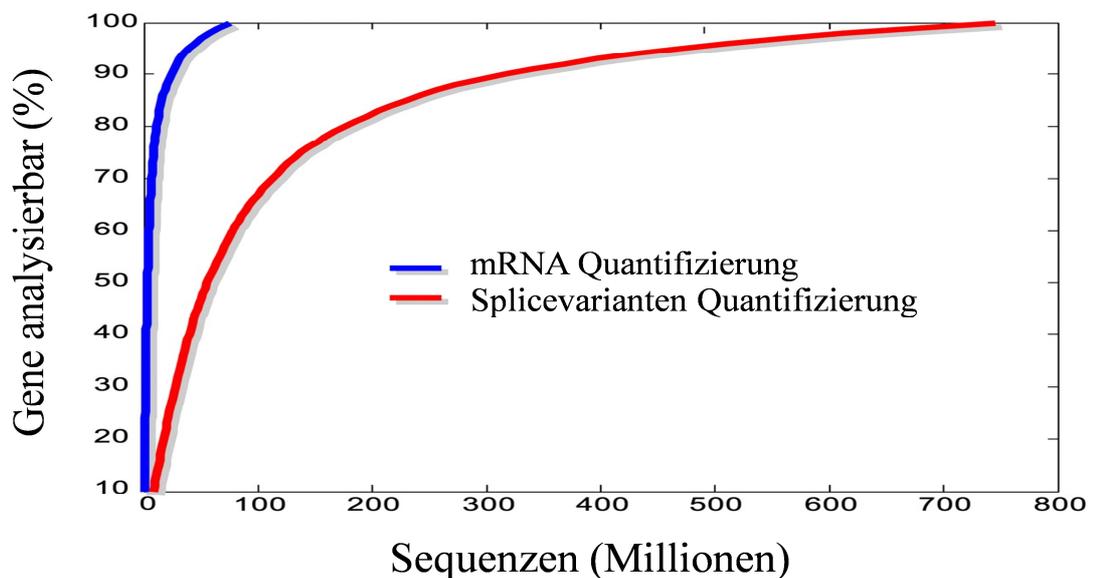


Abbildung 4 - 3 RNA-SEQ: Sequenzen gegen Analysetiefe

Die Abbildung zeigt in Blau wie viele Sequenzen benötigt werden, um für eine bestimmte Prozentzahl aller exprimierten Gene eine sichere Quantifizierung zu ermöglichen. Ausgegangen wird hierbei von einer Sequenzlänge von 35 bp. Die rote Kurve gibt an, daß für eine Analyse auf Splicevarianten Ebene ein Vielfaches an Sequenzen benötigt wird. Verändert nach Blencowe et al. [166].

Auf Ebene der mRNA Quantifizierung ist der Datensatz absolut ausreichend, eine Analyse auf Ebene der Splicevarianten wäre mit den vorhandenen Daten allerdings

nicht in zufriedenstellender Auflösung möglich. Da die übrigen Zelllinien durchweg mehr und zudem längere Sequenzen als die der Adipozyten aufweisen, ist eine völlig ausreichende Datengrundlage gewährleistet. Um die Ergebnisse der Hochdurchsatzsequenzierung überprüfbar machen zu können, wäre ein Abgleich mit den Ergebnissen des Microarrays nur in Einzelfällen möglich. Da anhand des Arrays jedoch nur eine Auswahl an Genen überprüfbar ist und zudem jede EST Sonde meist nur einen Teil eines mehrerer möglicher Transkripte eines Gens darstellen kann, wurden zur Ermittlung der am besten geeigneten Methode der Expressionsquantifizierung drei bioinformatisch unterschiedliche Ansätze durchgeführt. Auf Ebene der einzelnen Sequenzen handelt sich um den Vergleich der Expressionsanalyse mittels Kartierung gegen annotierte Bereiche des Rindergenoms, sowie einer BLASTn Datenbanksuche gegen die bovinen Sequenzcluster der NCBI Unigene Datenbank. Zusätzlich wurden die Sequenzen zu möglichst langen Contigs assembliert und die Anzahl der Sequenzen pro Contig gezählt. Erst nachfolgend erfolgte die Annotation der Konsensussequenzen durch einen BLASTn/x Datenbankabgleich. Obwohl die Methoden sehr unterschiedlich sind, lassen sich aus den Ergebnissen sehr gute Korrelationen zwischen den Auswertungsstrategien feststellen (siehe Abbildung 3 - 12). Die Vergleichbarkeit der mit den unterschiedlichen Methoden generierten Listen nimmt auf höheren Listenplätzen (>200) ab. Die Anzahl besonders stark exprimierter Gene ist vergleichsweise begrenzt, die Reihenfolge der vielen, ähnlich stark exprimierten Gene im mittleren Expressionsbereich ist hingegen beliebig. Der Anteil der den jeweiligen Referenz zuordenbaren Sequenzen schwankt technisch bedingt, wobei die Kartierung und die BLASTn Analyse noch am ehesten vergleichbar sind, da es sich um Analysen auf Einzelsequenzebene handelt (siehe Tabelle 3 - 8 und Tabelle 3 - 9). Aus den statistischen Werten der Prozessierung der Rohdaten lassen sich Rückschlüsse über deren Eignung für die weiterführende Analyse ziehen. Deutlich werden Unterschiede zwischen den amplifizierten, in Köln sequenzierten cDNA Banken und den im eigenen Haus generierten Daten sichtbar. So weist die Statistik der Kartierung mittels des CLC Programmpakets hohe qualitätsbedingte Ausfälle (PF-Filter 16-17 %) in den Rohdaten der in Köln sequenzierten Zelllinien auf. Eine Sequenz wird verworfen, wenn in den ersten 25 Basen mehr als einmal ein Chastity Wert von 0,6 erreicht wird (siehe Punkt 2.6.1). Diese verhältnismäßig hohe Fehlerrate ist vermutlich technisch bedingt, bei zu hoher Clusterdichte überlappen Signale aus unterschiedlichen Clustern und die resultierenden Sequenzen können nicht mehr auseinander gehalten werden. Die im eigenen Haus durchgeführten

Sequenzierungen der RNA aus jungen MSC weist nur einen ca. 5 %igen Anteil an Daten schlechter Qualität auf, hier sind mit 74 % kartierten Sequenzen auch die meisten Treffer im Genom zu verzeichnen. Die insgesamt schlechtere Qualität der in Köln erhobenen Daten wird auch in der BLASTn Analyse deutlich, als „N“ bestimmte Basen kommen in diesen Datensätzen mehr als fünfmal häufiger vor, als dies bei im eigenen Haus sequenzierten Daten beobachtet werden kann. Beim Vergleich der Menge an aussortierten Daten durch Trimmen der Sequenzen fällt auf, daß die Sequenzierung der Adipozyten mit ca. 10 % einen hohen Datenverlust aufweist. Die Trimming-Werte der übrigen Banken sind mit 2 – 4 % sowohl untereinander ähnlich, als auch wesentlich kleiner. Die Länge jeder Sequenz aus der Adipozyten-Bank beträgt in den Rohdaten 51 bp (siehe Tabelle 3 - 7). Beim Trimmen werden die Bereiche einer Sequenz abgeschnitten, in denen die Irrtumswahrscheinlichkeit für das korrekte Zuweisen (engl. *base calling*) einzelner Basen bei unter 99 % liegt. Da besonders am Beginn und am Ende einer Sequenz erfahrungsgemäß technisch bedingte Qualitätsprobleme die Daten stark beeinflussen, wurden zudem die ersten 3 und letzten 5 Basen jeder Sequenz der Adipozyten-Bank entfernt (Übrige Banken: 5' Ende - 5 bp ; 3' Ende - 20 bp). Die restliche Sequenz von guter Qualität darf zum Kartieren 15 bp Länge nicht unterschreiten. Diesen Wert zu unterschreiten ist nach dem Trimmen für die im Schnitt 40,3 Basenpaare langen Sequenzen der Adipozyten-Bank wahrscheinlicher, als für längere Sequenzen aus den übrigen Banken (55,5 – 70,5 bp Länge). Im Unterschied zur Kartierung wurde bei der BLASTn Analyse weniger Augenmerk auf die Qualität der einzelnen Sequenzen gelegt, es wurden zum Abgleich Volllänge-Sequenzen eingesetzt. Um nicht genkodierende Sequenzen vor der Analyse entfernen zu können, wurden diese vor dem Abgleich mit Clustern der Unigene Datenbank ebenfalls mittels BLASTn Analyse gegen repetitive oder rRNA beinhaltende Datensätze entfernt. Hierbei wird deutlich, daß vor allem die alten MSC und die Osteoblasten Zelllinien einen hohen Anteil an ribosomaler RNA aufweisen (21,74 bzw. 14,41% rRNA). In den Daten aus Chondrozyten lässt sich trotz identischer Abläufe in RNA Präparation, Amplifikation und Sequenzierung keine so hohe Durchsetzung mit rRNA feststellen. Die nicht der rRNA zuordenbare Sequenzmenge findet sich scheinbar in den unzuordenbaren Sequenzen wieder, mit 29,53 % der Sequenzen konnte für die Chondrozyten am häufigsten kein BLASTn Treffer erzielt werden. Der Anteil an Sequenzen der Chondrozyten, die nach Kartierung oder BLASTn Analyse eindeutig auf die Referenz passen, liegt methodenunabhängig ähnlich hoch (62,55 – 65 %). Daraus lässt sich schließen, daß

viele der nicht mittels BLASTn identifizierbaren Sequenzen sich aus Qualitätsgründen nicht kartieren ließen. Insgesamt weist die BLASTn Analyse aber eine höhere Toleranz gegen Sequenzierartefakte auf, als dies für die Kartierung beobachtet werden kann. In den Daten aus Adipozyten und Osteoblasten werden deutlich mehr Sequenzen mittels BLASTn Analyse als mittels Kartierung der jeweiligen Referenz zugeordnet. Besonders die Statistik der Daten aus Adipozyten zeigt deutlich, daß aufgrund schlechter Qualität vor der Kartierung aussortierte Sequenzen (10 % Trimming) mittels BLASTn Abgleich der Referenz zuordenbar sind. Grund hierfür ist neben den generellen, technischen Unterschieden zwischen Kartierung (Identitätsabgleich) und BLASTn Analyse (Ähnlichkeit wird toleriert) sicher die Verwendung von Volllänge-Sequenzen. Dies macht selbst bei steigender Fehlerhäufigkeit eine eindeutige Zuordnung zur Referenz wahrscheinlicher, allerdings verbunden mit einer geringeren Stringenz. Nach Assemblierung und anschließender BLASTn Suche der Konsensussequenz gegen die refseq-RNA Datenbank kann auf sehr lange Bruchstücke, teilweise auf komplette Transkripte zurückgegriffen werden. Somit ist die Spezifität der BLASTn Analyse gut gewährleistet. Die Trefferquote liegt für alle Sequenzierungen außer der der jungen MSC im zweistelligen Prozentbereich. In den Statistiken zur Assemblierung kommt der Unterschied zwischen amplifiziertem und nicht amplifiziertem RNA Ausgangsmaterial sehr deutlich zum Tragen. Die durchschnittliche Contiglänge ist selbst bei Assemblierung kurzer Sequenzen wie sie aus den Adipozyten vorliegen ($\bar{\varnothing}$ 40,3 bp) mit 1184 bp wesentlich länger als dies für die Sequenzierung aus amplifizierter RNA bestimmt werden konnte (max. alte MSC $\bar{\varnothing}$ 785 bp). Auch die Anzahl der assemblierten Contigs folgt diesem Trend, die Marke von 20.000 Contigs wird von keiner der Zelllinien mit assemblierter RNA erreicht. Die unter Punkt 4.2.5 beschriebene, technisch bedingte Verschiebung zum 3'OH Ende der Transkripte nach RNA Amplifikation beeinträchtigt die Assemblierbarkeit stark. Die mittlere Länge der Contigs der Sequenzierung der jungen MSC ($\bar{\varnothing}$ 1350 bp) und auch der Adipozyten ($\bar{\varnothing}$ 1184 bp) liegen näherungsweise am unter Punkt 4.2.5 angenommenen mittleren Wert der Transkriptlänge ($\bar{\varnothing}$ 1387 bp), was für die übrigen Sequenzierungen (max 785 bp) nicht zu beobachten ist. Weiterhin auffällig ist der im Vergleich zur Kartierung und BLASTn Analyse niedrige prozentuale Anteil an Sequenzen, die zu Contigs assembliert werden können (max 67 % Adipo). Auch hier wird deutlich, daß amplifiziertes Ausgangsmaterial besonders wenig effizient assembliert werden kann (max 23 % alte MSC). In der Assemblierung ist folglich die Sequenzlänge weniger entscheidend als eine möglichst gleichmäßige

Verteilung über die gesamte Transkriptlänge, die eher kurzen Sequenzen aus den Adipozyten weisen sehr gute statistische Werte auf. Legt man die Aussage der Abbildung 4 - 3 zugrunde, so scheint offensichtlich, daß die Assemblierung als Methode weniger geeignet zur exakten Bestimmung der Genexpressionsstärke ist als Kartierung und BLAST Analyse. Dafür sind durch Assemblierung auch in der Referenz nicht enthaltene oder unvollständige Transkripte auffindbar und aufgrund der im Vergleich zur Einzelsequenz langen Konsensussequenz für die funktionelle Einordnung zugänglich. Zusammenfassend lassen die angewandten Methoden in Kombination eine umfassende Analyse der Genexpression zu.

4.5 UMFASSENDE VERGLEICH DER GENEXPRESSIONSPROFILE ALLER ZELLINIEN

Sowohl die Microarray Ergebnisse, als auch die Analyse der Markergene legen den Schluss nahe, daß alte MSC und Osteoblasten in ihrer Genexpression ähnlich sind, eine Beobachtung die etwas weniger deutlich auch für Adipozyten und jungen MSC zutrifft. Diese Schlussfolgerungen stützen sich auf die ähnliche Expression einiger weniger Gene. Die NGS Daten erlauben hingegen einen das komplette Transkriptom umfassenden Vergleich (siehe Punkt 3.6.4). Die bereits an einzelnen Genen festgemachten Beobachtungen können so bestätigt werden. Wie zuvor die Markergenanalyse gezeigt hat, weisen die alten MSC auch über die Expression aller Gene verglichen eine hohe Ähnlichkeit mit dem Expressionsprofil der Osteoblasten auf. Sowohl nach Ähnlichkeit gruppierte „Heatmaps“ (siehe Abbildung 3 - 13) als auch der direkte Vergleich im Streudiagramm (siehe Abbildung 3 - 14) bestätigt für alte MSC und Osteoblasten sowie für junge MSC und Adipozyten die vermuteten Ähnlichkeiten. Den unter Punkt 4.2.5 bereits diskutierten Hypothesen zur spontanen osteogenen Differenzierung von MSC soll an dieser Stelle nichts hinzugefügt werden. Aus der hohen Ähnlichkeit der Genexpressionsprofile der jungen MSC und der Adipozyten kann der Rückschluss erfolgen, daß die zu Adipozyten differenzierten Zellen möglicherweise trotz der Differenzierung gewisse Eigenschaften mit den jungen MSC teilen. Es kommen mehrere Hypothesen in Frage, die erklären könnten, warum die Genexpressionsprofile gerade dieser Zelllinien sich global betrachtet so stark ähneln. Eine Erklärung wäre, daß die Zellen aus demselben Tier stammen, was allerdings nicht der Fall ist. Des Weiteren wäre ein technischer Grund denkbar, die RNA aus jungen MSC und Adipozyten wurde

im Gegensatz zu den übrigen Zelllinien nicht amplifiziert. Der Effekt der Amplifikation würde allerdings eher die Abdeckung zum 5' Ende eines längeren Transkriptes hin beeinflussen, nicht jedoch die Identität des Gens das transkribiert wird. Gegen ein technisches Artefakt sprechen auch die adipogenen Markergene, die in jungen MSC schlichtweg nicht stark exprimiert werden. Weiterhin denkbar wäre, daß nicht alle MSC unter den die Adipogenese induzierenden Bedingungen eine ihre Genexpression beeinflussende Veränderung durchlaufen. Dies würde bedeuten, daß noch ein gewisser Anteil undifferenzierter Zellen mit den zu Adipozyten differenzierten Zellen eine Mischkultur ausbildet. MSC kommen natürlicherweise in Fettgewebe vor und können daraus auch isoliert und *in vitro* differenziert werden [167]. Fettzellen stellen folglich eine Nische dar, in der MSC *in vivo* gute Bedingungen vorfinden. Dagegen spricht, daß selbst in Standardmedium über einen langen Zeitraum im Multilayer spontane Differenzierungsprozesse zu beobachten sind. Es erscheint daher unwahrscheinlich, daß ausgerechnet in adipogenem Differenzierungsmedium der Stammzellstatus einer großen Menge MSC unverändert erhalten bleibt. Umgekehrt erscheint es wenig wahrscheinlich, daß ausdifferenzierte Adipozyten zufällig ein den undifferenzierten MSC ähnliches Genexpressionsprofil aufweisen. Es ist bekannt, daß Fettzellen ein sehr dynamisches Gewebe darstellen, in dem sterbende Zellen beständig durch neue ersetzt werden. *In vivo* konnte gezeigt werden, daß hierzu zirkulierende Progenitorzellen aus dem Knochenmark rekrutiert werden [168]. Da diese Möglichkeit im Zellkulturversuch ausscheidet, muss eine Dedifferenzierung von Adipozyten zu Progenitorzellen mit MSC ähnlichem Genexpressionsprofil in Betracht gezogen werden. In der Literatur finden sich für de/redifferenzierende Adipozyten vermehrt Hinweise in unterschiedlichen Tiermodellen [169]. Von Fernyhough und Kollegen wurde bewiesen, daß voll ausdifferenzierte Fettzellen *in vitro* dedifferenziert werden können [169, 170]. Es wird angenommen, daß die Zellen auch *in vivo* eine bedeutende Rolle bei der Erneuerung von Fettgewebe spielen. Zusammengenommen könnte dies das den Stammzellen ähnliche Genexpressionsmuster in den Adipozyten erklären.

4.6 FUNKTIONELLE ANNOTATION STARK EXPRIMIERTER GENE - JUNGE MSC

Die 100 in jungen MSC am stärksten exprimierten Gene (TER) haben zum größten Teil mit der Ausbildung der extrazellulären Matrix zu tun (siehe Punkt 3.7.2). Die beiden mit weitem Abstand am stärksten exprimierten Gene *col1a1* und *col1a*, sowie die unter den Top 25 gelisteten Gene *col6a1* und *col6a2* und dem Glycoprotein *fn1* sind sicherlich die prominentesten Vertreter unter den Genen, deren Proteine wichtige Komponenten der extrazellulären Matrix darstellen (siehe Tabelle 3 - 12). In der Literatur ist die Ausbildung einer extracellulären Matrix für undifferenzierte MSC gut dokumentiert. Es ist bekannt, daß sie für die Selbsterhaltung sowie die Differenzierbarkeit eine wichtige Rolle spielt [171, 172] und einen elementaren Teil der Stammzellnische darstellt. Auch die Expression von mit Entwicklungsprozessen assoziierbaren Genen lässt sich nachweisen, wenn auch die Ontologien oft weiter gefasst sind und nicht ausschließlich Stammzellgene im eigentlichen Sinn umfassen (z.B. Actin 1 und 2). Es werden aber auch direkt mit der Erhaltung des Stammzellstatus der MSC bereits beschriebene Gene wie *tnc* unter den 25 am stärksten exprimierten Genen gelistet. Das Gen *tnc* kodiert für ein Protein (Tenascin C) mit vielen EGFR Wiederholungsmotiven und ist ebenfalls Bestandteil der ECM. Es wurde von Rodrigues und Kollegen als [173] als MSC protektiv beschrieben, wobei es die Stammzellen durch Steuerung des EGFR Signalweges und entsprechender Integrinaktivität selbst bei Vorhandensein von Apoptose auslösenden Signalgebern wie *fasl* (Tumor necrosis factor ligand superfamily member 6) vor dem Zelltod schützt. Auch das Gen *tagln* ist in undifferenzierten MSC aus trachealen Aspiraten bereits als exprimiert beschrieben worden und scheint hier auch bei der Differenzierung zu Myofibroblasten eine wichtige Rolle zu spielen [174]. Mit den vorliegenden Daten ist die MSC spezifische Ausprägung dieser Matrix gut dokumentiert. Da die Zellen so früh wie möglich geerntet wurden, besteht die begründete Hoffnung, daß die aufgefundenen ECM Proteine die von den MSC aktiv gebildete *in vivo* Stammzellnische noch weitestgehend unverändert abbilden.

4.7 FUNKTIONELLE ANNOTATION STARK EXPRIMIERTER GENE - ADIPOZYTEN

Die Ähnlichkeit der Listen der 100 am stärksten exprimierten Gene der Adipozyten (siehe Punkt 3.7.3) und jungen MSC ist mit 61 % überraschend hoch, wobei sich dies mit dem unter Punkt 4.5 bereits diskutierten, globalen Expressionsähnlichkeiten gut in Einklang bringen lässt. Es ist nicht verwunderlich, daß sich auch die aus den 100 am stärksten exprimierten Genen angereicherten Ontologien (siehe Abbildung 3 - 16) stark mit denen der jungen MSC ähneln. Es resultieren zwei wichtige Fragen aus diesem Ergebnis – (1) ist die Ähnlichkeit im Expressionsmuster mit Literaturangaben für Fettzellen überein zu bringen, und (2) wie lassen sich die 39 % in jungen MSC nicht unter den Top 100 Genen auffindbaren Gene eingruppiieren. Gene mit direktem Bezug zur ECM spielen offensichtlich auch in den Adipozyten eine große Rolle [175]. Dies deckt sich mit den Beobachtungen von Nakajima und Kollegen, die in bovinen intramuskulären PräAdipozyten während der adipogenen Differenzierung eine starke Expression der Collagene des Typs I-VI sowie des Gens *fn1* feststellen konnten [176]. Sowohl Nakajima als auch Kubo und Kollegen entdeckten, daß sich die ECM im Verlauf der Differenzierung in ihrer Zusammensetzung als veränderlich erweist, wobei von Kubo et. al eine starke *col1a1* Expression erst mit dem Ende der adipogenen Differenzierung detektiert [177]. In jedem Falle stellt die starke Expression von ECM assoziierten Genen auch für Adipozyten und deren Differenzierung eine Grundvoraussetzung dar. Mit dem Gen *timp1* findet sich unter den ECM assoziierten Genen die nur in den Top 100 der Adipozyten, nicht jedoch in der Liste der jungen MSC auftauchen eine Metalloprotease, die im Tierversuch die Adipogenese effektiv unterbindet. Vorhandene Fettzellen vergrößern sich bei TIMP1 Exposition jedoch, es tritt zudem eine Insulinresistenz, systemische Fettsäureüberladung und Fetteinlagerungen in der Leber auf [178]. Ein weiteres Gen in die Liste der exklusiv in den Adipozyten unter den Top 100 vorkommenden Gene ist *epas1*, bei dessen Überexpression in NIH 3T3 Zellen eine starke Akkumulation von Lipiden beobachtet werden konnte. Von Shimba und Kollegen konnte gezeigt werden, das das Gen eine wichtige Rolle in der Regulation der Glukose Aufnahme und Lipid Biogenese durch Steuerung der Transkription der Glukosetransporter *glut1*, *glut4* und des Insulinrezeptors *irs3* einnimmt [179]. Zusammenfassend lässt sich sagen, daß trotz der hohen Ähnlichkeit der stark exprimierten Gene zwischen jungen MSC und Adipozyten klare,

gewebespezifische Unterschiede auszumachen sind und ein für Adipozyten repräsentatives Ergebnis vorliegt.

4.8 FUNKTIONELLE ANNOTATION STARK EXPRIMIERTER GENE – OSTEOBLASTEN UND ALTE MSC

Da die Listen der 100 am stärksten exprimierten Gene in den Osteoblasten und den alten MSC mit 73 % die stärkste Ähnlichkeit im Vergleich aller Zelllinien aufweisen und zudem auch funktionell auf Ebene der Markergene bereits ein starke Ähnlichkeit zwischen alten MSC und Osteoblasten nachgewiesen werden konnte, ist von einer getrennten Diskussion der Ergebnisse kein Mehrwert zu erwarten (siehe Punkte 3.7.4 und 3.7.5). Den ähnlichen Listen der stark exprimierten Gene entsprechend weisen auch die angereicherten Ontologien eine große Überlappung auf (8 von 12). Hierbei ist interessant, daß unter den 12 aufgeführten Ontologien zwei der nur in Osteoblasten angereicherten GO - Terms direkt mit der Knochenentwicklung überein zu bringen sind (*biomineral formation*, *calcium ion binding*). Insgesamt ergibt sich aus der Analyse der stark exprimierten Gene in den Osteoblasten ein für diesen Zelltyp stimmiges Bild, mit den Genen *bglap* (Platz 54), *omd* (Platz 47) und *spp1* (Platz 11) finden sich auch klare osteogene Markergene unter den Top 100 exprimierten Genen in Osteoblasten. Die Gene *omd* und *bglap* finden sich nicht in der entsprechenden Liste der alten MSC. Dies zeigt, daß die alten MSC wahrscheinlich nicht völlig mit *in vitro* differenzierten Osteoblasten identisch sind, sondern nur Differenzierungsmerkmale, die stark in diese Richtung weisen, exprimieren. So ließen sich auch die negativen Ergebnisse der den alten MSC gleichsetzbaren Kontrollen in den Osteoblasten spezifischen Färbeversuchen erklären. Auffällig ist der starke Anstieg von mit Transkription/Translationsprozessen assoziierbaren Genen. Der Anteil der ribosomalen Proteine ist in beiden Zelllinien mit 44 - 49% extrem hoch. Hierzu ließ sich in der Literatur keine entsprechende Angabe zum vermehrten Aufkommen ribosomaler Proteine während der Ossifikation finden. Die Ossifikation ist *in vivo* ein sehr komplexer Prozess, der mit massiven Veränderungen der Ausgangszellen einhergeht [180]. Ein hoher Bedarf an Proteinen des Translationsapparates ist somit durchaus schlüssig. Unter den 100 am stärksten exprimierten Genen befindet sich auch das nicht näher charakterisierte Gen C29H11orf10 (= TMEM258). Das Gen ist in *Coelomata* konserviert (NCBI HomoloGene:40931), es finden sich Homologe

sowohl im Menschen und der Maus, als auch im Zebrafisch und der Fruchtfliege. Die dem Gen *C29H11orf10* zugehörigen bovinen (Bt.42864) und humanen (HS.437779) Unigene Einträge zeigen ein sehr breites EST-Profil, das eine Vielzahl von Geweben umspannt, in den alten MSC wird es jedoch nicht unter den ersten 100 Transkripten gefunden. Leider geben weder die Literatur noch die Analyse der Proteinstruktur verwertbare Ergebnisse zur Einordnung des Gens im Rahmen der Osteogenese. Eine mögliche Rolle bleibt daher ungeklärt. An dieser Stelle könnte nur noch ein funktioneller Assay (knockout/knockdown; Überexpression; Hefe-Zwei-Hybrid-System) Aufschluss bringen. Die breite Expression in unterschiedlichen Geweben lässt jedoch den Rückschluss zu, daß eine ossifikationsspezifische Funktion unwahrscheinlich ist.

4.9 FUNKTIONELLE ANNOTATION STARK EXPRIMIERTER GENE – CHONDROZYTEN

Die Analyse der 100 in den vermeintlichen Chondrozyten am stärksten exprimierten Gene weist erneut nicht auf einen chondrogenen Charakter der Zellen hin (siehe Punkt 3.7.6). Zwar sind die Zellen im Vergleich zu den jungen MSC und den Adipozyten ebenfalls auffällig aktiv in der Expression von Genen des Translationsapparates (ribosomale Proteine), chondrogene Markergene fehlen jedoch in der Liste völlig. Unter den Top 100 Listenplätzen findet sich nicht ein einziges für ein Kollagen kodierendes Gen, ein Befund der einer chondrogenen Differenzierung widerspricht. Dafür lassen sich mit mangelnder Vitalität assoziierbare Ontologien als unter den 100 stärkstexprimierten Genen angereichert ausmachen. So wird beispielsweise mit dem Gen *rps3* ein ribosomales Protein mit beschriebener proapoptotischer Funktion [181] auf Platz 7 der Liste auffallend stark exprimiert. Gleich drei Ontologien (*negative regulation of cell proliferation*, *positive regulation of cell death*, *negative regulation of cellular process*) weisen auf beginnende Apoptose der Zellen hin, ein Ergebnis, was sich auch mit den mikroskopischen Beobachtungen der Zellen während der Differenzierung und dem Ablösen der Zellen im Laufe der Medienwechsel deckt. Zwar ergibt sich kein konkreter Hinweis, welchen negativen Einflüssen die Zellen genau ausgesetzt sind, eine Diskussion im Sinne der möglichen Aufklärung von für Chondrozyten funktionell relevanter Transkriptionsaktivität ist jedoch leider nicht möglich.

4.10 ANALYSE DIFFERENZIELL EXPRIMIERTER GENE MITTELS RNA-SEQ

Die Statistik der Bestimmung differenziell exprimierter Gene zeigt, daß in allen Zelllinien trotz unterschiedlicher TER Anzahl exonisch kartierter Sequenzen (Total Exon Reads) ähnlich viele der 24359 in der Referenz annotierten Gene mit mindestens einer Sequenz kartiert werden (14.887 – 16.870 Gene). Erst ein Filter auf einen RPKM Wert von 10 erlaubt eine Fokussierung auf nicht basal transkribierte Gene. Erneut fällt auf, daß sich klare Unterschiede in der Anzahl der zur weiteren Analyse genutzten Gene zwischen den in Köln sequenzierten Datensätzen aus amplifizierter RNA und den Daten aus dem eigenen Haus ergeben, die ohne eine Amplifikation der RNA vor Sequenzierung errechnet wurden (siehe Tabelle 3 - 17). So werden in den jungen MSC und Adipozyten über 1000 Gene mehr mit einem RPKM Wert von über 10 bestimmt (4608 – 5115 Gene), als dies für die übrigen Zelllinien der Fall ist (3122 – 3468 Gene). Der Grund hierfür kann nicht in der Anzahl der kartierten Sequenzen liegen, da gerade für die Adipozyten mit 5115 Genen > 10 RPKM bei ca. 33 Mio weniger TER verzeichnet werden, als dies für alte MSC und Chondrozyten der Fall ist. Letztere weisen jedoch nur 3137 – 3468 Gene mit > 10 RPKM auf. Dementsprechend werden in jungen MSC und Adipozyten zwar mehr Gene über dem Schwellenwert von 10 RPKM exprimiert, die durchschnittliche Expression ist aber wesentlich geringer (25 – 27 RPKM) als dies für alte MSC, Osteoblasten und Chondrozyten bestimmt wurde (38 – 42 RPKM). Betrachtet man die Anzahl der als differenziell exprimiert bestimmter Gene, zeigt sich ein ähnliches Bild wie es unter Punkt 4.5 diskutiert wurde. Die dort beobachtete Ähnlichkeit zwischen jungen MSC und Adipozyten spiegelt sich in der mit 874 Genen verhältnismäßig geringen Anzahl als differenziell exprimiert bestimmter Gene wider. Im Vergleich der anderen drei Zelllinien mit jungen MSC werden zwischen 2637 und 2925 Gene als differenziell exprimiert ermittelt. Die Zuordnung dieser Gene zu biologischen Pathways ermöglicht eine funktionelle Einordnung der beobachteten Expressionsunterschiede. Als wesentliche Einschränkung bleibt bei der Interpretation der Daten zu beachten, daß gerade die differenzielle Genexpression der differenzierten Zelllinien nur eine Momentaufnahme zum Zeitpunkt der RNA Extraktion darstellen kann. Aufgrund der langen Differenzierungszeiträume können die Zellen als ausdifferenziert betrachtet werden. Ein Monitoring des Differenzierungsverlaufs hätte wesentlich mehr

Entnahmezeitpunkte benötigt, was sowohl aus Kostengründen, als auch aus Gründen der Materialknappheit nicht möglich war.

4.10.1 DIFFERENZIELL ERHÖHTE GENEXPRESSION IN JUNGEN MSC

Zur Bestimmung der stammzellspezifischen, differenziellen Genexpression in jungen MSC beschränkt sich der Abgleich mit den Expressionsprofilen der differenzierten Zelllinien auf die der Adipozyten und Osteoblasten. Für beide genannten differenzierten Linien kann aufgrund der bislang diskutierten Ergebnisse eine erfolgreiche Differenzierung angenommen werden. Das Einbeziehen der Daten aus der Sequenzierung der alten MSC und der Chondrozyten birgt die Gefahr der Verfälschung der Ergebnisse, da diese Linien mit großer Wahrscheinlichkeit als Modell für *in vivo* relevante Differenzierungsprozesse nicht geeignet sind. Zudem erlaubt dieses Vorgehen solche Gene zu selektieren, die unabhängig von der Art der Differenzierung die zum Vergleich heran gezogen wurde, immer in jungen MSC differenziell erhöht exprimiert werden. Die Wahrscheinlichkeit, mit Stammzeleigenschaften assoziierbare Gene zu entdecken, erhöht sich dadurch. Auf diese Weise sind 250 Gene als in den jungen MSC differenziell erhöht exprimiert bestimmt worden. Betrachtet man die biologischen Pathways die durch diese 250 Gene abgedeckt werden, so lassen sie sich funktionell sehr gut mit undifferenzierten MSC übereinbringen. Die verstärkte Expression von Genen zur Replikation von DNA und des Zellzyklus sind für die rasche Proliferation von Stammzellen unabdingbar. Aufgrund der starken Expression von Komponenten der extrazellulären Matrix (siehe Punkt 4.6) ist auch die signifikante Anreicherung einiger ECM assoziierter Gene (*thbs1*, *itga5* und *itga3*, *lama5* und *tnc*) in den biologischen Pathways nicht überraschend. Die MSC förderlichen Eigenschaften von Tenascin C wurden bereits unter Punkt 4.6 aufgezeigt, das Gen ist unter den 25 am stärksten exprimierten Genen in jungen MSC gelistet. Das sekretierte ECM Protein Thbs1 (Thrombospondin-1) wurde nach einer initialen Differenzierung von humanen ES Zellen in zu Oligodendrozyten weiter differenzierbaren Progenitorzellen in erhöhter Menge nachgewiesen [182]. Eine Rolle in anderen adulten Stammzellen wie den jungen MSC ist somit ebenfalls vorstellbar. Interessant ist, daß gerade *lama5* (Laminin 5) laut Studien von Klees und Kollegen [183] in humanen MSC ohne weitere Stimuli eine osteogene Differenzierung auslösen kann, was die partielle osteogene Differenzierung der alten MSC erklären würde. Im Gegensatz dazu konnten Lai und Kollegen zeigen [184], daß besonders

eine ECM mit hohem Laminin Anteil die Stammzeleigenschaften von MSC in Kultur zu erhalten hilft [184]. Daraus kann der Schluss gezogen werden, daß wohl eine einzige Komponente der ECM allein nicht ausschlaggebend für eine Differenzierung oder deren Ausbleiben ist. Bei Integrinen handelt es sich um Transmembranproteine, zusammengesetzt aus unterschiedlichen Untereinheiten (z.B. alpha; beta), die der Signaltransduktion zwischen der ECM und dem Zytoplasma dienen. Eine wichtige Rolle von Integrinen in der Zell-Matrix Kommunikation ist für MSC mehrfach beschrieben worden. Wenig überraschend spielen Integrine sowohl für Differenzierungen, als auch für den Erhalt der Stammzeleigenschaften eine wichtige Rolle. Mittels semiquantitativer RT-PCR konnten in MSC für die Integrine Itga5 und Itga3 durch die Arbeit von Saller und Kollegen [185] stammzellfördernde Eigenschaften bei hoher Expression der Gene belegt werden, die besonders unter physiologischen Sauerstoffkonzentrationen von 2 % zu beobachten waren. Mit der ECM direkt in Zusammenhang steht der biologische Pathway der Fokalen Adhäsion, alle 5 Gene aus dem ECM Pathway werden auch hier eingruppiert. Mit dem Begriff der Fokalen Adhäsion wird eine große Gruppe an Proteinen umrissen, die einerseits die Verankerung der Zellen in der ECM gewährleisten, zusätzlich aber auch direkt an der Signalübertragung (chemisch, mechanisch) zwischen der Matrix und den Zellen beteiligt sind. Das Cytoskelett der MSC ist durch Fokale Adhäsion direkt mit der ECM verbunden. Gerade die Tatsache, daß 13 unterschiedliche Gene der Fokalen Adhäsion sowohl in MSC, als auch in den Adipozyten statistisch signifikant differenziell exprimiert werden, legt eine spezifische Funktion der Transkripte für den jeweiligen Zelltyp nahe. Der Einfluß der Fokalen Adhäsion auf MSC und deren Differenzierung wurde unter anderem von Matthieu und Kollegen diskutiert [186]. Für die undifferenzierten MSC wird dabei vor allem eine Veränderung der mechanischen Stabilität angenommen, aufgrund von Umstrukturierung der Actin Fasern. Hierfür sind besonders die frühen Passagen in Kultur geeignet, Differenzierungsprozesse können in diesem Stadium sogar durch rein mechanische Stimuli eingeleitet werden. Der genaue Mechanismus bleibt leider unklar, allerdings propagieren die Autoren eine Präferenz in der Differenzierungsrichtung anhand der messbaren Steifheit der Zellen und derer äußeren Form vorhersagen zu können. Häufiger untersucht erscheint dagegen der Zusammenhang zwischen Stammzellerhalt, Differenzierung und dem WNT Signalweg [187]. Dieser Signalweg wird ausschließlich in der Analyse der differenziell erhöht exprimierten Gene der jungen MSC als statistisch signifikant angereichert bestimmt. Im Abgleich sowohl mit den

Adipozyten, als auch mit den Osteoblasten werden gegen die Wikipathways 4 Gene (*wnt5a/b*, *ldlr*, *nkd1* und *fosl1*) bestimmt (signifikante Anreicherung), gegen den umfangreicher dokumentierten KEGG WNT Pathway werden 2 weitere Treffer (*sfrp2* und *gpc4*) verzeichnet (hier p-Value > 0,05). Würde man die Genexpression in den Adipozyten außer Acht lassen, werden im Vergleich zu den Osteoblasten allein 27 Gene des WNT Pathways in den jungen MSC als differenziell hoch exprimiert ermittelt. Dies unterstreicht einerseits die Wichtigkeit des Pathways für MSC, andererseits zeigt es erneut die Stammzellnähe der Adipozyten auf. Die Signaltransduktion im WNT Signalweg ist äußerst komplex und nicht vollständig aufgeklärt (siehe Abbildung 4 - 4). Sie kann auf zwei Arten, durch den sog. kanonischen oder über den nicht kanonischen Teil des Pathways erfolgen. Vereinfacht beschrieben resultiert die Aktivierung des kanonischen Teils des Signalwegs aus der Interaktion von WNT Proteinen mit membranständigen „Frizzled“ Rezeptoren in der ECM.

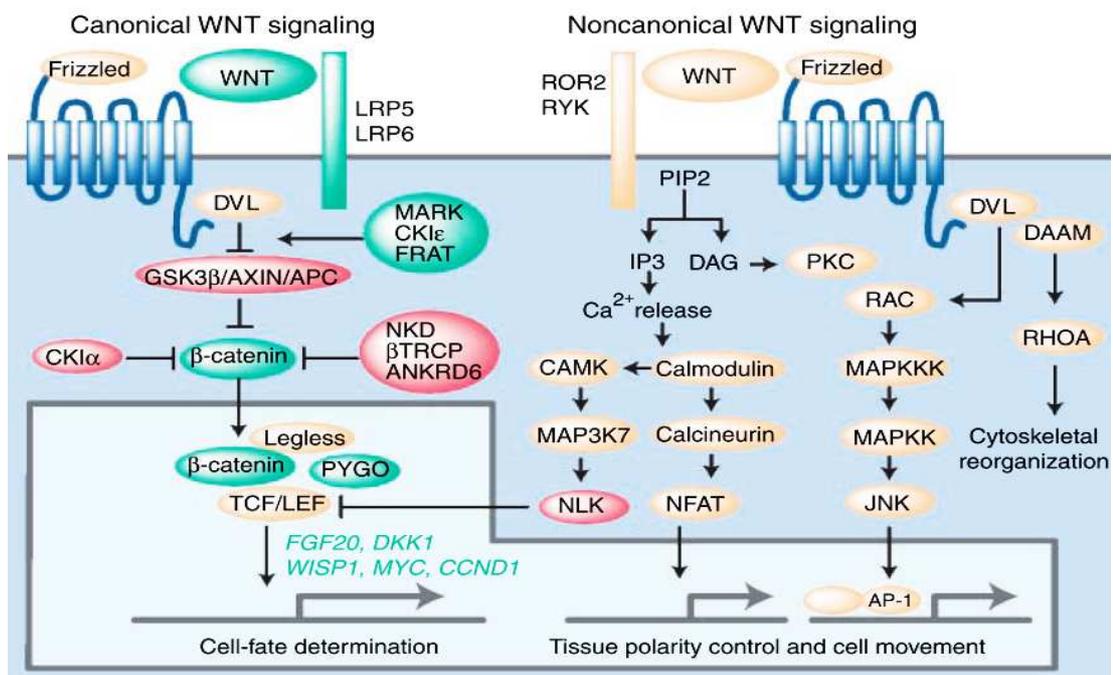


Abbildung 4 - 4 Schematische Darstellung des WNT Pathways

Gezeigt wird die Signalkaskade sowohl des kanonischen, als auch des nicht kanonischen Teil des WNT Signalweges. Der kanonische Teil führt zur Transkription spezifischer Zielgene durch Akkumulation von β -Catening im Zellkern. Der nicht kanonische Teil hemmt den kanonischen Signaltransduktionsweg und führt zur Restrukturierung des Cytoskeletts. Auf Gewebeebene *in vivo* ist eine Festlegung der Polarität und die Migration von Zellen die Folge des nicht kanonischen Teil des Signalwegs, hier spielt JNK/AP-1 regulierte Transkription eine Rolle. Quelle: Katoh et al. 2007 [188].

Über mehrere Zwischenschritte führt dies zum Stopp des Abbaus von β -Catenin durch das Proteasom, so daß sich β -Catenin im Zellkern akkumulieren kann. Im Zellkern werden dadurch Transkriptionsinhibitoren (Histon Deacetylasen = HDACs) verdrängt und Histon Acetylasen rekrutiert. Dies führt zu einer Euchromatisierung und der verstärkten Transkription der WNT Zielgene [187]. Der nicht kanonische Teil des Pathways transduziert Signale unabhängig von β -Catenin, jedoch stehen auch hier Frizzled Rezeptoren am Anfang der Kette. Von dort ausgehend sind die Möglichkeiten der Signalweiterleitung jedoch weit vielfältiger als im kanonischen Teil des WNT Pathways. Die möglichen Kaskaden können als eigene, untergeordnete Signalwege betrachtet werden („*Panar Cell Polarity Pathway*“, „*WNT/Ca²⁺ Pathway*“), die beispielsweise durch assoziierte Korezeptor-Transmembranproteine wie PIP2 und Ror2 mittels Veränderungen im Kalziumspiegel wirken. Zwar bestehen noch Lücken im Verständnis der Signaltransduktion, zusammenfassend lässt sich als Effekt des aktiven, nicht kanonischen WNT Pathways eine Hemmung des kanonischen Teils des WNT Pathways, sowie eine Restrukturierung des Cytoskelettes und die Initiierung von Migrationsprozessen feststellen [187, 188]. Die differenzielle Genexpression in den jungen MSC zeigt klar eine Hemmung des kanonischen Teils des WNT Signalweges (*nkd1* [135, 136]) auf, bei gleichzeitiger Signaltransduktion durch den nicht kanonischen Signalweg (*wnt5a* [189], *wnt5b* [190]). Diverse Studien zeigen für MSC eine direkte Abhängigkeit, sowohl von Differenzierbarkeit als auch von Selbsterneuerungsprozessen, speziell vom kanonischen Teil des WNT Signalweges auf. Hierbei lassen sich durchaus auch direkt widersprüchliche Angaben über den Einfluss des WNT Signalweges finden, so daß ein Einfluß der Signalintensität und eine strenge zeitliche Regulierung der WNT Signale im Rahmen von Differenzierungsprozessen angenommen wird [187]. Brunt und Kollegen konnten zeigen, daß eine geringere Bioverfügbarkeit von β -Catenin direkt mit dem hohen Spenderalter der Patienten korrelierbar ist und die Differenzierbarkeit von MSC negativ beeinflusst. Dies spricht für eine altersbedingt schwache Aktivität des kanonischen Teils des WNT Signalweges, die durch Gabe von LiCl zum Freisetzen von β -Catenin experimentell behoben werden konnte [191]. Interessanter Weise findet sich das in dieser Arbeit in jungen MSC differenziell stark exprimierte Gen *fosl1* in der Arbeit von Brunt und Kollegen unter den in alten Patienten stärker exprimierten Genen - und wäre somit ebenfalls ein potentieller Hemmer des kanonischen Teil des WNT Signalwegs. Überträgt man die Ergebnisse von Brunt und Kollegen auf diese Arbeit, bewirkt eine geringe β -Catenin Bioverfügbarkeit eine

geringe Differenzierung in den jungen MSC. Da das Alter der sog. jungen MSC in dieser Arbeit jedoch keinen Einfluß haben sollte, ist somit ein regulatorischer, dem Verbleib im undifferenzierten Zustand dienender Mechanismus in der starken Expression von Genen des nicht kanonischen Teils des Signalwegs und der Hemmung des kanonischen Teils zu vermuten. Dafür spricht auch, daß mit dem Gen *gpc4* ein bekannter Regulator, sowohl des kanonischen, als auch des nicht kanonischen Teil des WNT Signalweges in jungen MSC differenziell hoch exprimiert wird [192]. Ein weiteres Argument dafür, daß die beobachtete Genexpression im nicht kanonischen Teil des WNT Pathways regulatorisch vermutlich sowohl von Relevanz als auch reversibel ist, lässt sich auch aus der Tatsache ableiten, daß die kanonische WNT Signalkaskade bekanntermaßen notwendig ist für die osteogene Differenzierung von MSC [193]. Da nur ein Teil der Genexpression sich in biologischen Pathways abbilden lässt, ermöglicht eine GO – Term Analyse aller 250 in jungen MSC differenziell hoch exprimierten Gene ein noch differenzierteres Bild von deren Funktion (siehe Abbildung 3 - 1). Deutlich zu erkennen ist die Anreicherung von Prozessen, die mit MSC gut überein zu bringen sind. Die automatische Gruppierung der GO – Terms anhand ihrer familiären Ähnlichkeit spiegelt teilweise die Ergebnisse der Pathway Analyse wieder, mit dem WNT Signalweg assoziierbare Ontologien machen den größten Teil der sog. „*Treemap*“ aus. Des Weiteren überwiegen mit der Zellteilung, Proliferation und Zell Adhäsion und Zellmotilität in Zusammenhang stehende Prozesse, so daß ein MSC spezifisches Muster aus den differenziell stark exprimierten Genen sicher und mit besserer Genauigkeit abgeleitet werden kann, als dies durch die Analyse der 100 am stärksten exprimierten Gene alleine möglich ist (vergleiche Punkt 4.6). Das uncharakterisierte Transkript C28H1orf198 weist im Rind 3 Exons auf und ist unter Amnioten gut konserviert. Auf Proteinebene wird sein Homolog im Menschen (C1orf198) bereits als sicher bestätigt geführt (NCBI), ohne daß sich eine konservierte Domäne darin vorhersagen ließe. Die entfernte Ähnlichkeit zur prokaryotischen DNA polymerase III, Untereinheit gamma und tau als einziger Anhaltspunkt ist daher wohl eher zufälliger Natur. Das Unigene Expressionsprofil lässt keinen Rückschluss auf eine stammzellspezifische Funktion zu, da das Gen in sehr unterschiedlichen Geweben transkribiert wird. Da auch keine Literatur zu diesem Gen vorhanden ist, könnte eine funktionelle Einordnung nur experimentell erfolgen (*knockdown/out*, Überexpression etc.). Für den LOC100335299 ist die Annotation ähnlich dürftig, hier wird für das Rind gar keine konservierte Domäne vorhergesagt, auch existiert kein boviner Unigene Eintrag für diesen Locus. Das

humane Homolog wird unter der Bezeichnung KIAA0040 geführt. Auch der humanen Sequenz ist keine konservierte Domäne zuordenbar, der humane Unigene Eintrag weist im Expressionsprofil fast alle aufgeführten Gewebetypen als ähnlich stark transkribierend auf. Leider ist auch hier ohne experimentelles Vorgehen keine weitere Aussage für eine mögliche Funktion des Gens für MSC zu treffen. In der Literatur wird das humane Homolog zu LOC100335299 in genomweiten Assoziationsstudien mit familiärer Alkoholsucht in Verbindung gebracht [194]. Interessant ist in diesem Zusammenhang, daß die Autoren eine Wechselwirkung mit flankierenden Genen als Hypothese in den Raum stellen. Die flankierenden Gene sind im Menschen und Rind identisch, es handelt sich um die Matrixproteine *tnn* und *tnr* (Tenascin N und R, siehe Tenascin C Punkt 4.6), für welche ein Einfluß auf MSC vorstellbar ist.

4.10.2 DIFFERENZIELL ERHÖHTE GENEXPRESSION IN ADIPOZYTEN

Die Anzahl der zwischen Adipozyten und jungen MSC als differenziell exprimiert bestimmten Gene ist mit 874 im Vergleich zu den beiden übrigen differenzierten Zelllinien gering. Dies ist angesichts der globalen Ähnlichkeit der Expressionsprofile (siehe Punkt 4.5) auch zu erwarten. Somit ist eine funktionelle Relevanz der 480 in Adipozyten differenziell erhöht exprimierten Gene sehr wahrscheinlich, da die ausdifferenzierten Fettzellen phänotypisch nichts mehr mit undifferenzierten MSC gemein haben. Die Ergebnisse der Pathway Analyse weisen direkt mit Fettzellen in Zusammenhang stehende Eigenschaften auf, die klar die erfolgreiche Differenzierung belegen. Der Signaltransduktionsweg „*PPAR Signaling Pathway*“ zeigt eine schwerpunktmäßige Anreicherung von Genen des Lipid Metabolismus und der Adipozyten Differenzierung (nicht gezeigt). Der „*Adipocytokine Signaling Pathway*“, der im Organismus die über Leptin vermittelte Regulation [195] des Energiebedarfs steuert, ist ebenfalls eindeutig mit Fettzellen assoziierbar. Gleiches gilt für den „*Glycerolipid metabolism pathway*“, Glycerolipide stellen die klassische Form eines Energiespeichers in Fettzellen dar. Besonders Auffällig ist die Anreicherung von Genen die der Fokalen Adhäsion zugeordnet werden können, ein Prozess, der schon im Rahmen der differenziellen Genexpression der jungen MSC diskutiert wurde (siehe Punkt 4.10.1). Da aufgrund der differenziellen Expression keine Gene überlappend gelistet werden, liegt hier mit Sicherheit ein für Adipozyten spezifisches Expressionsmuster vor. Legt man die Daten der in Adipozyten hoch exprimierten Gene zugrunde (siehe Punkt 4.7),

macht neben der offensichtlich wichtigen Rolle der ECM in Fettzellen auch die verstärkte Fokale Adhäsion als Verbindung zwischen Cytoskelett und Matrix Sinn. In der Literatur finden sich wenige Hinweise auf eine Rolle der fokalen Adhäsion in der Adipogenese oder für Fettzellen im Allgemeinen [196], eine breiter funktioneller Diskurs findet eher auf Einzelgen - Ebene statt. Eine GO – Term Analyse der 15 Fokalen Adhäsionsgene mittels GOEAST liefert mit der Ontologie „*lipid metabolic process*“ (GO:0006629, p-Value 0.000906) anhand der 4 Gene *pdgfra/b*, *pten* und *pil3r1* einen direkten Bezug Fokaler Adhäsionsgene zu Fettzellen, der sich so auch in der Literatur niederschlägt [197-199]. Dennoch scheint die Fokale Adhäsion insgesamt ein für Adipozyten bislang nicht gut untersuchter Vorgang zu sein. Mit ähnlicher Präsenz (15 Gene) wie die Fokale Adhäsion wird der „*Cytokine-cytokine receptor interaction*“ Signaltransduktionsmechanismus als in den Adipozyten durch differenziell exprimierte Gene signifikant angereichert ermittelt. Zytokine sind bekanntermaßen an der Steuerung von Differenzierungsprozessen, Zellwachstum oder Inflammatorischen Prozessen beteiligt, auch für Fettgewebe ist eine wichtige Rolle vielfach dokumentiert [200-202]. Auffällig ist die Schnittmenge der „*Cytokine-cytokine receptor interaction*“ mit den insgesamt 9 Genen des „*jak-stat-signaling pathway*“ (= *Janus kinase/signal transducers and activators of transcription*). Der letztgenannte Signalweg ist für Adipozyten mehrfach als für die Differenzierung und Homöostase essentiell beschrieben worden [203-205]. Eine Rolle dieses Signalweges in zu Adipozyten differenzierten MSC unterstützt diese Erkenntnisse und zeigt, daß im Rahmen der differenziellen Genexpressionsanalyse mittels NGS spezifische Expressionsmuster, die eine gesicherte biologische Relevanz aufweisen, ermittelt werden konnten. Für die Genontologie Analyse wurden nur die 100 am stärksten differenziell exprimierten Gene im Vergleich Adipozyten gegen junge MSC berücksichtigt. Hierbei wird deutlich, daß der größte Block aus dem Jak-Stat Signalweg zuordenbaren Ontologien besteht, besonders regulatorische Vorgänge werden verstärkt abgebildet. Die Zellen scheinen somit weniger mitten im Differenzierungsprozess, als vielmehr voll ausdifferenziert vorzuliegen. Entwicklungsprozesse und Proliferation werden im Vergleich zu den jungen MSC nur noch untergeordnet abgebildet. Mit Fettzellen assoziierte Ontologien bilden eigene Gruppen aus (*lipid metabolic process*, *glycerolipid metabolism*). Ontologien der Extrazellulären Matrix sind als zweitstärkste Gruppe aufgeführt, was aufgrund der hohen Präsenz unter den am stärksten exprimierten Gene naheliegend ist. Da die gesamte Analyse sich auf einen einzelnen, mit 28 Tagen recht späten Zeitpunkt der Differenzierung stützt, wäre es von Interesse, den Verlauf der Genexpression über

einen längeren Zeitpunkt zu verfolgen. Dies könnte für die im Rahmen dieser Arbeit ermittelten Gene auch kostengünstig mittels qRT PCR erfolgen, um so detailliertere Einblicke in den adipogenen Differenzierungsprozess zu erhalten.

4.10.3 DIFFERENZIELL ERHÖHTE GENEXPRESSION IN OSTEOLASTEN UND ALTEN MSC

Die zuvor beobachtete, geringe Ähnlichkeit in den globalen Expressionsprofilen der Osteoblasten/alten MSC verglichen mit den jungen MSC schlägt sich auch in der Anzahl differenziell exprimierter Gene nieder. Die alten MSC scheinen mit 744 differenziell hoch exprimierten Genen den jungen MSC minimal ähnlicher zu sein scheinen als die Osteoblasten, für die 936 differenziell höher exprimierten Gene beim Abgleich bestimmt wurden. Dieser Unterschied mag zunächst nicht biologisch relevant erscheinen, besonders weil sich alte MSC und Osteoblasten in den gegenüber den jungen MSC differenziell hoch exprimierten Genen zu 73,28 % gleichen. Die Betrachtung der differenziell hoch exprimierten Gene, die im Vergleich zu jungen MSC nur in Osteoblasten, nicht in alten MSC ermittelt werden, zeigt jedoch in der Genontologie Analyse klar signifikante Unterschiede auf (Details siehe Punkt 3.8.3). Die *in vitro* zu Osteoblasten differenzierten Zellen zeigen proosteogene Ontologien angereichert, während die alten MSC bezüglich der knochentypischen Biomineralisierung negative Regulationsmechanismen aufweisen. Diese Beobachtungen decken sich mit den unter Punkt 4.8 diskutierten Unterschieden hinsichtlich der am stärksten exprimierten Gene in beiden Zelllinien. Beide Zelllinien weisen dennoch beim Abgleich mit den biologischen Pathways fast identische Treffer auf, die bis auf eine Ausnahme vollständig unter den Oberbegriffen Transkription und ATP Gewinnung zusammengefasst werden können. Unterschiede bestehen seitens der Osteoblasten in dem nur dort angereicherten KEGG Signalweg „*basal transcription factors*“, der sich nahtlos in die ohnehin erhöhte Transkriptionsaktivität der Zellen einreicht. Betrachtet man die normalisierte Genexpression der Top 25 exprimierten Gene in Osteoblasten und jungen MSC, lässt sich sowohl die offensichtlich erhöhte Transkriptionsaktivität, als auch der verstärkte Energiebedarf erklären. Die Osteoblasten erreichen in der Summe eine um 34.000 RPKM stärkere Expression als die jungen MSC auf den ersten 25 Listenplätzen. Die direkte Korrelation von erhöhtem Energiebedarf bei zunehmender Ossifikation konnte auch in Experimenten von Guntur und Kollegen an unterschiedlich stark ossifizierenden murinen Zelllinien belegt werden [206]. Der

„chemokine signaling pathway“ wird mit 16 differenziell exprimierten Genen nur in alten MSC als statistisch signifikant abgedeckt ermittelt, ein Ergebnis welches sich auch in den entsprechenden GO-Term Analysen der alten MSC niederschlägt (Ergebnisse nicht grafisch dargestellt). Eine Funktion für lange in Kultur befindliche MSC lässt sich hierbei weder aus den differenziell exprimierten Genen, noch mittels einer nachgeschalteten GO-Term Analyse dieser 16 Kandidaten sicher ableiten. Die GO-Term Analyse (GOEAST) zeigt eine sehr breite Streuung vornehmlich regulatorischer Prozesse als Ergebnis, ohne daß mit Bestimmtheit ein Rückschluss auf die Funktion möglich wäre. Die grafische Darstellung der Genontologien der 100 in Osteoblasten am stärksten differenziell hoch exprimierten Gene zeigt hingegen ein stimmiges Bild. Neben Translation und Ossifikation nehmen mit Zellatmung assoziierbare Prozesse den größten Teil der Grafik ein. Bislang nicht charakterisierte, differenziell hoch exprimierte Gene werden im Vergleich zu den jungen MSC in beiden Zelllinien gleichermaßen angetroffen. Die Diskussion des Gens *C29H11orf10* ist unter Punkt 4.8 nachzulesen. Bei *C5H12ORF57* (= *c10*) handelt es sich um ein verhältnismäßig sicher annotiertes Gen, für dessen Homolog (*C12orf57*) beim Menschen ein Nachweis auf Proteinebene besteht. Dem Protein ist eine eigene konservierte Domäne zugewiesen (*Domain of unknown function - DUF4511*), deren Funktion unklar bleibt. Das Gen wird ubiquitär exprimiert, scheint aber im Falle einer Mutation eine Hypoplasie des *Corpus callosum* zur Folge zu haben [207]. Eine Funktion für Osteoblasten lässt sich leider nicht ableiten.

4.10.4 DIFFERENZIELL ERHÖHTE GENEXPRESSION IN CHONDROZYTEN

Die Bestimmung der differentiellen Genexpression im Vergleich der vermeintlich zu Chondrozyten differenzierten Zellen mit jungen MSC bestätigt die schon zuvor angedeuteten Erkenntnisse, daß leider keine chondrogene Differenzierung stattgefunden haben kann. Eine Diskussion der Ergebnisse ist vor diesem Hintergrund schwierig. Die Tatsache, daß die Analyse der biologischen Signalwege vom Profil her den Ergebnissen der Osteoblasten gleicht (steigende Transkriptionsrate und Energiestoffwechsel), kann als Indiz gewertet werden, daß zumindest ein unbekannter Differenzierungsprozess eingeleitet wurde. Das Auffinden proapoptotischer Merkmale in der Genontologie Analyse kann als Hinweis interpretiert werden, daß im weiteren Verlauf massive Probleme die Differenzierung haben abbrechen lassen, ohne daß sich dies näher spezifizieren ließe. Die angewandten Differenzierungsprotokolle sollten auf käuflich erwerbbar,

sicher chondrogen differenzierbare Standardzelllinien übertragen werden, um die Machbarkeit des Assays an sich zu demonstrieren. Mit den vorliegenden Daten ist keine sinnvolle weiterführende Analyse möglich.

4.11 CHARAKTERISIERUNG UNBEKANNTER TRANSKRIPTE

Die Strategie, mit der aus den unkartierten Sequenzen möglicherweise noch nicht beschriebene Transkripte gefiltert wurden, ist sehr stringent und führt auf der einen Seite zu der gewünschten Reduzierung auf wenige Kandidaten (siehe Punkt 2.7.4). Auf der anderen Seite ist die Bedingung minimal 100 Aminosäure lange, offene Leserahmen zu finden für die Entdeckung etwaiger neuer Exons von bekannten Genen potentiell hinderlich, da diese durchaus auch kürzer sein können. Die mittlere Exon Länge im Rind wird in einer groß angelegten Vergleichsstudie mit insgesamt 16 eukaryotischen Genomen von Zhu und Kollegen [208] mit 162,3 bp (~ 54 AS) Länge angegeben. Eine ORF Reduktion um fast die Hälfte hätte leider zu einer unüberschaubar großen Zahl an Kandidaten Contigs geführt. Das zweite einschneidende Filterkriterium stellt die Suche nach Ähnlichkeiten mit bekannten konservierten Domänen dar. Dies soll die Wahrscheinlichkeit eines signifikanten Treffers erhöhen, nur enthält gerade bei unvollständiger Assemblierung nicht jeder letztlich proteinkodierende Contig auch eine funktionelle Domäne des Proteins. Die Strategie ist also insgesamt eher geeignet, neue Gene oder größere Exons ausfindig zu machen, was gerade bei verhältnismäßig gut annotierten Genomen wie dem des Rindes aufgrund der ohnehin breit angelegten *in silico* Prädiktionen nicht übermäßig wahrscheinlich ist. Die vorhergesagten konservierten Domänen in den Contigs sind in der Detailanalyse nicht überzeugend oder entpuppen sich als Zufallstreffer. Aus diesem Grund konnte im Rahmen dieser Arbeit daher auch kein komplett unbekanntes Gen ausfindig gemacht werden, dennoch sind einige interessante Beobachtungen möglich. Betrachtet man die Statistik der Assemblierung ergibt sich ein bereits unter Punkt 4.4 diskutiertes Bild. Sowohl die Anzahl der Contigs, als auch deren Länge scheinen in erster Linie davon ab zu hängen, ob die zugrundeliegende RNA amplifiziert wurde. Die durchschnittliche Sequenzlänge vor Assemblierung spielt erneut eine eher untergeordnete Rolle. Von daher ist es nicht verwunderlich, daß am Ende ausschließlich Kandidaten aus den jungen MSC und den Adipozyten zur weiteren Diskussion verbleiben. Gerade die cDNA - Assemblierung ist technisch bedingt besonders empfindlich im Falle einer

nicht gleichmäßigen und unvollständigen Abdeckung zum 5'Ende der Transkripte hin. Die Rekartierung breiter Bereiche der genomischen DNA um die Lokalisation des BLAST Treffers interessanter Contigs herum ermöglicht auch bei nicht vollständig assemblierten Contigs darüber hinaus gehende Einblicke in die mögliche Struktur des uncharakterisierten Transkriptes. Hier wirkt sich die Kartierung ohne Mismatches (100 % Identität zur Referenz) teilweise negativ aus. Zwar lassen sich sehr große Datensätze verhältnismäßig schnell auf der Referenz abbilden, und auch die große Anzahl kleiner Sequenzfragmente wird mit hoher Wahrscheinlichkeit nicht zufällig falsch kartiert. So kann verhindert werden, daß Transkriptionsaktivität an genomischen Loci suggeriert wird, die auf zufälligen Treffern beruht. Dennoch entstehen dadurch auch Lücken in der Abdeckung sobald nur eine Base einer Sequenz nicht zur Referenz passt. Dies liegt in den meisten Fällen an SNPs, die Individuen spezifisch eine biologisch nicht weiter relevante Abweichung zur Referenz darstellen, und erschwert die exakte genomische Lokalisierung der tatsächlich transkribierten Bereiche, sowie das sichere Auffinden von Intron/Exon Grenzen. Durch der Beschränkung auf relativ stark exprimierte Kandidaten (siehe Punkt 2.7.4) sollte aufgrund der Masse an rekartierten Sequenzen jedoch kein Exon übersehen werden. Dennoch muss auch bedacht werden, daß das Programm „Bowtie1“ in der verwendeten Version Sequenzen nicht über Intron/Exon Grenzen hinweg gesplittet darstellen kann, so daß die Daten hier keine gute Grundlage zur Prädiktion dieser Grenzen bieten.

Mit dem Contig 1245, ermittelt aus den Daten der jungen MSC, wird der bereits bekannte, hypothetische Genlocus *LOC100847605* als genomischer Ursprung ermittelt. Im Gegensatz zur bekannten Lokalisation wird jedoch sowohl durch den BLASTn Treffer des Contigs gegen das Genom, als auch durch die anschließende Rekartierung eine Erweiterung der bereits annotierten Bereichs nötig (siehe Abbildung 3 - 26). Interessant ist hierbei die vollständige Abdeckung mit Sequenzen des als intronisch annotierten Bereichs nach Rekartierung, möglicherweise wurde hier auch unfertig gespleißte Prä-mRNA mit sequenziert. Eine andere Interpretationsmöglichkeit wäre, daß die Intron/Exon Grenzen nicht korrekt vorhergesagt wurden. Der Locus wird seitens des NCBI als nicht mehr länger in der aktuellsten Annotation des Rindergenoms geführt und ist nunmehr als nicht kodierende RNA reklassifiziert worden. Ein Abgleich im UCSC Gene Browser [209] mit bekannten RefSeq RNA Transkripten anderer Species und bekannten EST Sequenzen des Rindes lässt diese jüngste Entfernung des *LOC100847605* aus der

aktuellen Annotation jedoch zumindest zweifelhaft erscheinen (siehe Abbildung 4 - 5). Eine BLASTx Suche des Contig 1245 gegen die RefSeq Protein Datenbank des NCBI zeigt signifikante Treffer zum Protein CSNK1G1 auf, wobei hier allerdings mit maximal 4 % Sequenzabdeckung bei 70% Identität Raum für Spekulationen bleibt. Bei CSNK1G1 handelt es sich um eine Serin/Threonin-Protein Kinase, für das Protein wurde ein evolutionär konservierter Mechanismus der Kopplung von aktivierten WNT Rezeptoren an die cytoplasmatische Signaltransduktion beschrieben [210].

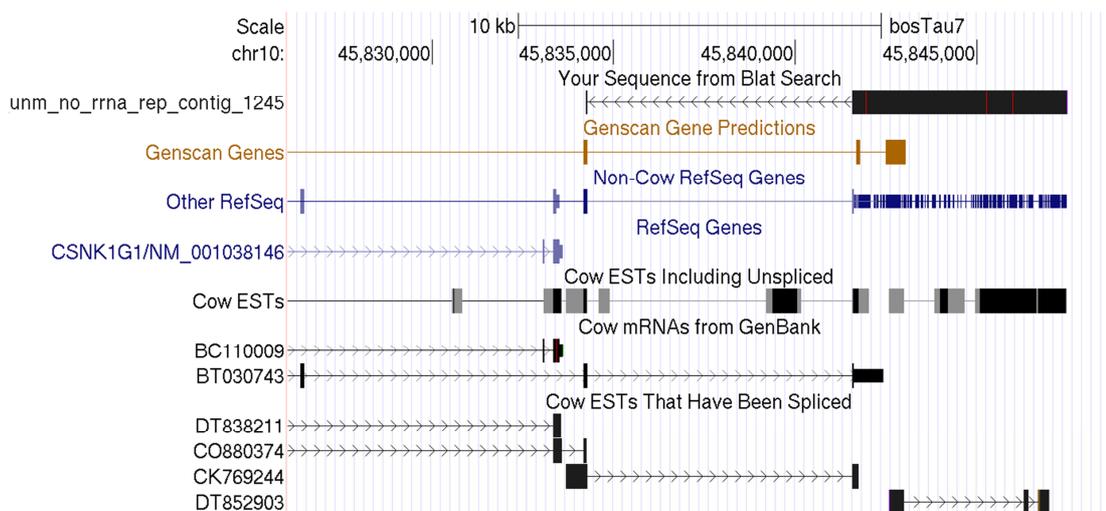


Abbildung 4 - 5 Genomische Lokalisation des Contig 1245 chr10:45,826,026-45,848,706

Gezeigt wird der BLAT (BLAST *like Alignment Tool*) Treffer des Contig 1245 auf der aktuellsten Version des Rindergenoms (Baylor Btau_4.6.1/bosTau7). Die Annotation des *LOC100847605* wird nicht mehr aufgeführt. In einer Vielzahl anderer Spezies (aus Übersichtsgründen nicht ausgeklappt) liegt unter der Bezeichnung „*Other RefSeq*“ das Gen *csnk1g1* an dieser Stelle annotiert vor, etwas weiter in 5' Richtung vom BLAT Treffer des Contig 1245 beginnt die bovine Annotation von *csnk1g1*. Diverse mRNA Sequenzen des Rindes unterstützen die bovine Genexpression an der Position des BLAT Treffers von Contig 1245.

Diese Beteiligung am WNT Signalweg macht das Gen gerade für undifferenzierte Mesenchymale Stammzellen interessant (Vergleiche Punkt 4.10.1). Dennoch bleibt die Frage, ob sich von dem propagierten Transkript ein Protein ableitet zunächst offen. Ohne einen experimentellen Ansatz, beispielsweise ein Nachweis im Westernblot oder durch eine massenspektrometrische Analyse lässt sich dieser Umstand nicht beheben.

Contig 1765 der Assemblierung der unkartierten Sequenzen aus den Adipozyten stellt sich als Verlängerung des unter dem Namen *KIAA0513* annotierten Gens dar. Es stellt sich die Frage, ob in diesem Fall ein verlängertes 3' UTR des Gens

aufgedeckt wurde, dessen biologische Relevanz schwer einschätzbar ist. Es sind regulatorische Einflüsse langer 3'UTR Sequenzen in der Literatur beschrieben [211], deren etwaiger Effekt sich aus den vorliegenden Daten allerdings nicht ableiten lässt. Selbiges gilt für die über das 5' Ende des Gens leicht hinaus reichende Rekartierung am ersten Exon. Eine BLASTx Suche des über das letzte Exon von *KIAA0513* hinaus kartierten Stück DNA ergibt keinen Treffer. Dies lässt eine proteinkodierende Funktion wenig wahrscheinlich erscheinen. Für andere Spezies ist laut den Daten des UCSC Browsers an den über die Annotation nach 3' hinaus gehenden Bereichen ausschließlich nicht kodierende Sequenz annotiert (nicht gezeigt), was das BLASTx Ergebnis untermauert und eher auf keine biologisch relevante Funktion für Contig 1765 rückschließen lässt. Das Gen *KIAA0513* ist im Menschen mit neuroplastischen Vorgängen assoziiert [212].

Auch für den in Adipozyten auffälligen Contig 147, der nach Rekartierung noch über den assemblierten Bereich hinaus im 5' Bereich des Gens *bcl6* (B-cell CLL/lymphoma 6) verortet wird, kann eine noch nicht annotierte Verlängerung des 5'UTR als wahrscheinlichste Erklärung angenommen werden. Zwar kann in dem durch Rekartierung abgedeckten Bereich eine konservierte Domäne ausgemacht werden (PHA03307, transcriptional regulator ICP4), diese ist aber eindeutig viralen Ursprungs. Für den Menschen und auch im Rind sind UTR Bereiche mehrerer *bcl6* zugehöriger Transkripte (RefSeq nicht bovine sowie bovine EST) an der Position des Contig 147 annotiert.

Contig 3147, der ebenfalls den Adipozyten zuzurechnen ist, umfasst nach Rekartierung mit über 16 kb Abdeckung den größten Bereich genomischer DNA. Die Rekartierung verbindet das Gen *lpp* (LIM domain containing preferred translocation partner in lipoma) mit dem nicht länger annotierten Locus *LOC100847108*. Auch in diesem Fall sind in anderen Spezies laut UCSC Annotation ausschließlich UTR Bereiche des *lpp* Gens genomisch annotiert, wobei für *LOC100847108* keinerlei transkribierte Information annotiert ist (Daten nicht gezeigt). Bemerkenswert ist die Größe des genomischen Bereichs auf dem *lpp* im Genom des Rind annotiert ist. Das Gen umfasst fast 684 kb DNA, so daß die 16 kbp durch Rekartierung entdeckten Informationen im Vergleich nicht so stark ins Gewicht fallen, wie sie das bei normal langen Genen täten. Wie für Contig 147 kann auch für Contig 3147 nicht von einer biologischen Relevanz der Ergebnisse ausgegangen werden.

Für alle verlängerten UTR Bereiche gilt jedoch, daß eine mangelnde Annotation eine Verfälschung der Expressionsstärken nach sich zieht. Die Kartierung mittels des CLC Programms berücksichtigt nur wenig über die bestehende Annotation hinaus gehende Bereiche. Dennoch kann nach der Analyse der unbekannteren Transkripte eher von Einzelfällen ausgegangen werden, das Gesamtbild bleibt von solchen kleineren Annotationsfehlern sicherlich unberührt.

5 ZUSAMMENFASSUNG UND AUSBLICK

Mesenchymale Stammzellen (MSC) sind Vertreter der adulten Stammzellen. Sie bergen durch ihre große Plastizität ein immenses Potential für die klinische Nutzung in Form von Stammzelltherapien. Zellen dieses Typs kommen vornehmlich im Knochenmark der großen Röhrenknochen vor und können zu Knochen, Knorpel und Fettzellen differenzieren. MSC leisten einen wichtigen Beitrag im Rahmen regenerativer Prozesse, beispielsweise zur Heilung von Frakturen. Breite Studien demonstrieren bereits jetzt auch bei komplexeren Erkrankungen (z.B. Osteoporose) therapeutisch vielversprechende Einsatzmöglichkeiten. Oft kommen hierbei aus MSC gezielt differenzierte Folgelinien aus Zellkulturen zum Einsatz. Dies bedingt eine kontrollierte Steuerung der Differenzierungsprozesse *in vitro*. Der Differenzierung einer Stammzelle liegt eine komplexe Veränderung ihrer Genexpression zugrunde. Genexpressionsmuster zur Erhaltung und Proliferation der Stammzellen müssen durch solche, die der linienspezifischen Differenzierung dienen, ersetzt werden. Die mit der Differenzierung einhergehende, transkriptomische Neuausrichtung ist für das Verständnis der Prozesse grundlegend und wurde bislang nur unzureichend untersucht. Ziel der vorliegenden Arbeit ist eine transkriptomweite und vergleichende Genexpressionsanalyse Mesenchymaler Stammzellen und deren *in vitro* differenzierten Folgelinien mittels Plasmid - DNA Microarrays und Sequenzieretechniken der nächsten Generation (RNA-Seq, Illumina Plattform). In dieser Arbeit diente das Hausrind (*Bos taurus*) als Modellorganismus, da es genetisch betrachtet eine hohe Ähnlichkeit zum Menschen aufweist und Knochenmark als Quelle von MSC gut verfügbar ist.

Primärkulturen Mesenchymaler Stammzellen konnten aus dem Knochenmark von Rindern erfolgreich isoliert werden. Es wurden *in vitro* Zellkultur - Versuche durchgeführt, um die Zellen zu Osteoblasten, Chondrozyten und Adipozyten zu differenzieren. Zur Genexpressionsanalyse wurde RNA aus jungen MSC und einer

MSC Langzeitkultur („alte MSC“), sowie aus den differenzierten Zelllinien isoliert und für nachfolgende Experimente wo nötig amplifiziert. Der Erfolg der Differenzierungen konnte anhand der Genexpression von spezifischen Markergenen und mittels histologischer Färbungen belegt werden. Hierbei zeigte sich die Differenzierung zu Osteoblasten und Adipozyten erfolgreich, während die Differenzierung zu Chondrozyten trotz diverser Modifikationen am Protokoll nicht erfolgreich durchgeführt werden konnte. Eine vergleichende Hybridisierung zur Bestimmung differenzieller Genexpression (MSC vs. Differenzierung) mittels selbst hergestellter Plasmid - DNA Microarrays ergab für die Osteogenese mit Genen wie *destrin* und *enpp1*, für die undifferenzierten MSC mit dem Gen *sema3c* neue Kandidatengene, deren biologische Funktion aufzuklären in zukünftigen Experimenten vielversprechende Ergebnisse liefern sollte. Die Analyse der transkriptomweiten Genexpression mittels NGS lieferte einen noch umfangreicheren Einblick ins Differenzierungsgeschehen. Es zeigte sich eine hohe Ähnlichkeit im Expressionsprofil von jungen MSC und Adipozyten, sowie zwischen den Profilen der alten MSC (eine Langzeitkultur) und Osteoblasten. Die alten MSC wiesen deutliche Anzeichen für eine spontane Differenzierung in die osteogene Richtung auf. Durch Analyse der 100 am stärksten exprimierten Gene jeder Zelllinie ließen sich für junge MSC und Adipozyten besonders Gene der extrazellulären Matrix (z.B. *col1a1,6* ; *fn1* uvm.) auffinden. Sowohl Osteoblasten, als auch die alten MSC exprimieren hingegen verstärkt Gene mit Bezug zur oxidativen Phosphorylierung, sowie ribosomale Proteine. Eine Betrachtung der differenziellen Genexpression (junge MSC vs. Differenzierung) mit anschließender Pathway Analyse und Genontologie Anreicherungsstatistik unterstützt diese Ergebnisse vor allem bei Osteoblasten, wo nun jedoch zusätzlich auch Gene zur Regulation der Knochenentwicklung und Mineralisierung in den Vordergrund treten. Für Adipozyten konnte mit Genen des „*Jak-STAT signaling pathway*“, der Fokalen Adhäsion, sowie Genen des „*Cytokine-cytokine receptor interaction pathway*“ sehr spannende Einsichten in die Biologie dieses Zelltyps erlangt werden, die sicher weiterer Untersuchungen bedürfen. In undifferenzierten MSC konnte durch differenzielle Genexpressionsanalyse die Rolle des nicht kanonischen Teils des WNT Signalweges als für die Aufrechterhaltung des Stammzellstatus potentiell äußerst einflußreich ermittelt werden.

Die hier diskutierten Ergebnisse zeigen beispielhaft, daß besonders mittels Genexpressionsanalyse im Hochdurchsatzverfahren wertvolle Einblicke in die

komplexe Biologie der Stammzelldifferenzierung möglich sind. Als Grundlage für nachfolgende Arbeiten konnten interessante Gene ermittelt und Hypothesen zu deren Einfluß auf Stammzeleigenschaften und Differenzierungsprozesse aufgestellt werden. Um einen besseren Einblick in den Differenzierungsverlauf zu ermöglichen, könnten künftig NGS Analysen zu unterschiedlichen Differenzierungszeitpunkten durchgeführt werden. Zudem wären weitere Anstrengungen zur erfolgreichen Etablierung der chondrogenen Differenzierung zur vollständigen Analyse der Genexpression des trilinearen Differenzierungspotentials von MSC wünschenswert.

6 ABKÜRZUNGSVERZEICHNIS

3' UTR	Untranslatierte Region 3' Ende eines Gens/mRNA
5' UTR	Untranslatierte Region am 5' Ende eines Gens/mRNA
aRNA	in vitro amplifizierte RNA
AS	Aminosäure
BLAST	<i>basic local alignment search tool. Typen: x, n, p...</i>
BLAT	<i>BLAST like alignment tool</i>
Bp	Basenpaare
bzw.	Beziehungsweise
cDNA	<i>complementary DNA</i>
DNA	<i>deoxyribonucleic acid</i>
DNase	Desoxyribonuklease
dNTP	2'-Deoxynukleosidtriphosphat
E. coli	<i>Escheria coli</i>
EDTA	Ethylendiamintetraacetat
ES	embryonale Stammzelle
EST	<i>expressend sequence tag</i>
FITC	Fluoresceinisothiocyanat
for	<i>forward</i>
gbk	Dateiformat <i>genebankfile</i>
gff3	Dateiformat <i>general feature format</i>
h	Stunde
Kb	Kilobasenpaare
mRNA	<i>messenger RNA</i>
MSC	Mesenchymale Stammzelle
NCBI	<i>National Center for Biotechnology Information</i>
NGS	<i>next generation sequencing</i>
NIH	<i>National Institute of Health</i>
nt	Nukleotid
ORF	<i>open reading frame</i>
Pathway	Zusammenfassend für Signaltransduktions- oder Stoffwechselwege, eingedeutscht
PBS	<i>phosphat buffer saline</i>
PCR	<i>polymerase chain reaction</i>
qRT-PCR	quantitative reverse Transkriptase PCR
rev	<i>reverse</i>
RIN	<i>RNA integrity number</i>
RNA	<i>ribonucleic acid</i>
RNAse	Ribonuklease
RNA-Seq	cDNA Sequenzierung mittels NGS
RPKM	<i>reads per kilobase of exon model per million mapped reads</i>
RPM	<i>reads per million reads sequenced</i>
RT	Raumtemperatur
RT(-PCR)	Reverse Transkriptase (PCR)
TBE	Tris-Borat-EDTA (Puffer)
VE-H2O	Vollentsalztes Wasser
vs.	<i>versus</i>
z.B.	zum Beispiel

7 ABBILDUNGSVERZEICHNIS

Abbildung 1 - 1 Modell der Stammzellteilung	1
Abbildung 1 - 2 Stammzellen in Medizin und Forschung	4
Abbildung 1 - 3 Meilensteine der Stammzelltherapie	5
Abbildung 1 - 4 Modell der Stammzellnische.....	8
Abbildung 1 - 5 Plastizität von MSC	9
Abbildung 1 - 6 Klinische Studien mit MSC gemäß des "US National Institute of Health"	10
Abbildung 1 - 7 Konstruktion und Hybridisierung eines DNA – Microarrays.....	14
Abbildung 1 - 8 Entwicklung der DNA – Sequenzierung – Der Trend ist ungebrochen	15
Abbildung 1 - 9 Ein DNA Fragment aus einer Illumina „Library“	17
Abbildung 1 - 10 Illumina „Bridge Amplification“ und Sequenzierung	18
Abbildung 2 - 1 Schematische Darstellung der Datenverarbeitung	45
Abbildung 2 - 2 Normalisierung durch Berechnen des RPKM Wertes.....	47
Abbildung 2 - 3 Formeln zur Berechnung des Rj – Wertes.....	51
Abbildung 3 - 1 Isolation und Primärkultur von MSC aus bovinem Knochenmark	60
Abbildung 3 - 2 Antikörper – und DAPI Färbungen an undifferenzierten MSC	61
Abbildung 3 - 3 Agarosegele der RT-PCR zum Nachweis von MSC Markergenen .	63
Abbildung 3 - 4 Agarosegele der RT-PCR zum Nachweis von CD Antigenen	64
Abbildung 3 - 5 Adipogene Differenzierung Mesenchymaler Stammzellen	66
Abbildung 3 - 6 Osteogene Differenzierung Mesenchymaler Stammzellen	67
Abbildung 3 - 7 Chondrogene Differenzierung Mesenchymaler Stamzellen	68
Abbildung 3 - 8 Bioanalyser Messung der Gesamt-RNA	70
Abbildung 3 - 9 Bioanalyser Messung der 2 X amplifizierten RNA.....	72
Abbildung 3 - 10 Überprüfung der Sonden zur Herstellung des Microarrays.....	75
Abbildung 3 - 11 SYBR Green I Färbung eines Microarrays.....	75
Abbildung 3 - 12 Expressionsstärke der 200 höchstexprimierten Gene in Abhängigkeit zur Bestimmungsmethode.....	80
Abbildung 3 - 13 Heatmap der Genexpression nach Kartierung.....	85
Abbildung 3 - 14 Streudiagramm ähnlicher und unähnlicher Genexpressionsprofile	86
Abbildung 3 - 15 Ontologie der 100 höchstexprimierten Gene – junge MSC.....	92
Abbildung 3 - 16 Ontologie der 100 höchstexprimierten Gene – Adipogenese.....	94
Abbildung 3 - 17 Ontologie der 100 höchstexprimierten Gene – Osteogenese.....	96

Abbildung 3 - 18 Interpro Analyse der Proteinsequenz des Gens <i>C29H11orf10</i>	96
Abbildung 3 - 19 Ontologie der 100 höchstexprimierten Gene – alte MSC.....	98
Abbildung 3 - 20 Ontologie der 100 höchstexprimierten Gene – Chondrogenese ..	100
Abbildung 3 - 21 REVIGO Treemap der Go-Term Analyse - junge MSC (250 Gene)	104
Abbildung 3 - 22 Pathway Analyse differenziell exprimierter Gene der Adipozyten	106
Abbildung 3 - 23 REVIGO Treemap der GO-Term Analyse - Adipozyten (Top 100 Gene).....	107
Abbildung 3 - 24 REVIGO Treemap der Go-Term Analyse - Osteoblasten (Top 100 Gene).....	108
Abbildung 3 - 25 Interpro Analyse der Proteinsequenz des Gens <i>C5H12ORF57</i>	109
Abbildung 3 - 26 Kartierung der genomischen Region im Bereich des Contig 1245	113
Abbildung 3 - 27 Kartierung der genomischen Region im Bereich des Contig 1765	114
Abbildung 3 - 28 Kartierung der genomischen Region im Bereich des Contig 147	115
Abbildung 3 - 29 Kartierung der genomischen Region im Bereich des Contig 3147	115
Abbildung 4 - 1 Modell eines heterogenen MSC Gemischs in vivo	124
Abbildung 4 - 2 Kartierung zum Vergleich amplifizierter / nicht amplifizierter RNA	128
Abbildung 4 - 3 RNA-SEQ: Sequenzen gegen Analysetiefe	138
Abbildung 4 - 4 Schematische Darstellung des WNT Pathways	151
Abbildung 4 - 5 Genomische Lokalisation des Contig 1245 chr10:45,826,026- 45,848,706	160

8 TABELLENVERZEICHNIS

Tabelle 2 - 1 Primärkultur Anzuchtmedium MSC	26
Tabelle 2 - 2 Übersicht der verwendeten Antikörper.....	27
Tabelle 2 - 3 cDNA Erststrangsynthese.....	28
Tabelle 2 - 4 Primer für die RT - PCR.....	29
Tabelle 2 - 5 Pipettierchema PCR.....	30
Tabelle 2 - 6 Standard PCR-Programm	30
Tabelle 2 - 7 Differenzierungsmedium Adipogenese	32
Tabelle 2 - 8 Differenzierungsmedium Osteogenese	32
Tabelle 2 - 9 Differenzierungsmedium Chondrogenese.....	33
Tabelle 2 - 10 Anpassung im Clontech SMART™ Kit Schritt „Primer Extension“ ...	40
Tabelle 2 - 11 Parameter der Assemblierung - CLC Genomics Workbench.....	50
Tabelle 2 - 12 Übersicht über die bioinformatischen Hilfsmittel.....	57
Tabelle 3 - 1 Ergebnisse der RT-PCR auf Markergene undifferenzierter MSC	64
Tabelle 3 - 2 Übersicht der Gesamt-RNA Präparationen	69
Tabelle 3 - 3 Ausbeute und Effizienz der RNA Amplifikationen	73
Tabelle 3 - 4 Differenzielle Genexpression - Osteoblasten vs. undifferenzierte MSC im Microarray	77
Tabelle 3 - 5 Differenzielle Genexpression - Chondrozyten vs. undifferenzierte MSC im Microarray	78
Tabelle 3 - 6 Differenzielle Genexpression - Adipozyten vs. undifferenzierte MSC im Microarray.....	79
Tabelle 3 - 7 Übersicht der Illumina Sequenzierungen.....	80
Tabelle 3 - 8 Prozessieren und Kartierung der Rohdaten – Kartierung mittels CLC	81
Tabelle 3 - 9 Prozessierung und BLASTn Analyse der Rohdaten.....	82
Tabelle 3 - 10 Prozessierung der getrimmten Rohdaten, Assemblierung und BLASTn	84
Tabelle 3 - 11 Expression von zelllinientypischen Markergenen	88
Tabelle 3 - 12 Stark exprimierte Gene - Top 25 der Sequenzierung des Transkriptoms junger MSC	91
Tabelle 3 - 13 Stark exprimierte Gene - Top 25 der Sequenzierung des Transkriptoms der Adipozyten.....	93
Tabelle 3 - 14 Stark exprimierte Gene - Top 25 der Sequenzierung des Transkriptoms der Osteoblasten.....	95

Tabelle 3 - 15 Stark exprimierte Gene - Top 25 der Sequenzierung des Transkriptoms der alten MSC	97
Tabelle 3 - 16 Stark exprimierte Gene - Top 25 der Sequenzierung des Transkriptoms der Chondrozyten	99
Tabelle 3 - 17 Übersicht der Bestimmung differenziell exprimierter Gene	101
Tabelle 3 - 18 DAVID/GO-Elite Analyse - Auswahl signifikant angereicherter Pathways	102
Tabelle 3 - 19 Assemblierung und BLASTn/x Analyse unkartierter Sequenzen	111
Tabelle 3 - 20 Übersicht der uncharakterisierten Transkripte	112

9 DIGITALER ANHANG

Der digitale Anhang befindet sich in jeder gedruckten Arbeit auf einem Speichermedium des Typs „SD-Karte“ hinterlegt.

Digitaler Anhang D1	Übersicht der Zellkulturen
Digitaler Anhang D2	Übersicht über den <i>Bos taurus</i> EST Plasmid-DNA Microarray
Digitaler Anhang D3	Übersicht über die mRNA Amplifikation
Digitaler Anhang D4.1	Ergebnisse der CLC Kartierung gegen das Rindergenom
Digitaler Anhang D4.2	Ergebnisse des BLASTn Abgleichs mit der Unigene Datenbank
Digitaler Anhang D4.3	Ergebnisse der CLC Assemblierungen und des BLASTn Abgleichs mit der RefSeq RNA Datenbank
Digitaler Anhang D5.1	Ergebnisse der Berechnung differentieller Genexpression
Digitaler Anhang D5.2	DAVID „Functional Annotation Clustering“ der 100 am stärksten exprimierten Gene
Digitaler Anhang D5.3	Ergebnisse der Analyse differentiell exprimierter Gene – Abgleich mit biologischen Pathways und Zuordnen der Genontologie
Digitaler Anhang D6	Ergebnisse der Analyse unkartierter Sequenzen und Charakterisierung noch nicht annotierter Transkripte

10 ANHANG – DIFFERENZIELL EXPRIMIERTE GENE

Im Folgenden sind die am stärksten differenziell exprimierten Gene im Vergleich undifferenzierter MSC (junge MSC) mit den differenzierten Zelllinien in alphabetischer Reihenfolge gelistet (siehe Punkt 3.8). Eine vollständige Übersicht aller Ergebnisse der Analyse differentieller Genexpression ist im Digitalen Anhang D5.3 einsehbar und kann hier aus Kapazitätsgründen nicht abgedruckt werden.

In jungen MSC im Vergleich zu Osteoblasten und Adipozyten differenziell stärker exprimierte Gene (250 Gene)

ACTA2 ACOT7, ACTA2, ACTC1, ACTG2, ACTN1, ALDH1A3, AMOTL2, ANGPTL2, ANKRD1, ANLN, ANXA8, ARF4, ARL4C, ATP8B1, AURKA, AURKB, AXL, BCAM, BCAR1, BYSL, C1QTNF5, C1QTNF6, C28H1orf198, C5H12orf32, CALR, CCNYL1, CDC20, CKAP4, CLEC11A, CNN1, CRIM1, CSPG4, CSRP1, CTH, CUEDC1, DACH1, DDX39, DES, DKKL1, DPYSL3, DRAM1, DSC2, DSC2, DSN1, DTYMK, E2F1, EGR1, EHD3, EMILIN1, EMILIN2, ENC1, EPHB3, ESM1, ESYT1, EZH2, FADS3, FBLN2, FBLN5, FEN1, FLNB, FOS, FOSL1, FOXS1, FSTL3, GALE, GALNTL1, GDF15, GNB1L, GPC4, GTSE1, GYS1, HJURP, HPCAL1, HS3ST2, IDH2, IDH3A, IER3, IER5, IGFBP7, IL1F6, IMPDH1, INCENP, INHBA, ISG12(B), ITGA3, ITGA5, JPH2, JUB, KDELR3, KIF11, KIF20A, KIF22, KIFC1, LAMA5, LDLR, LEPREL1, LHFP, LIG1, LIPG, LMCD1, LMF2, LMNB1, LMNB2, LMOD1, LOC100296944, LOC100297420, LOC100299504, LOC100335299, LOC100335328, LOC100335409, LOC100335829, LOC100336364, LOC513548, LOC520939, LOC521086, LOC538547, LOC540455, LOC617406, LOC618459, LOC781578, LOC782215, LOC784294, LOC785987, LOC788816, LOC789183, LOXL2, LOXL4, LRP4, LRRC59, MAN1A1, MANBAL, MCM2, MCM3, MCM4, MCM5, MEGF6, MEOX1, MGC137027, MGLL, MICAL2, MICAL3, MICALL1, MKI67, MLKL, MLLT11, MMP23B, MRPL15, MTSS1L, MYBL2, MYH10, MYLK, MYO1C, MYO1D, NEDD4L, NES, NGF, NKD1, NOTCH1, NOV, NT5C, NT5E, NTN1, NUA1, OBSL1, ORAI2, PAX6, PCSK1N, PDGFA, PFDN6, PGK2, PHF19, PIGC, PKIG, PLBD1, PLEKHA2, PLK1, PLK2, PLXNA3, POLD1, PQLC1, PRC1, PSMD8, PTK7, PTPRF, PVR, RAI14, RANBP1, RBBP7, RCAN2, RRM2, SDF2L1, SEPX1, SERPINE1, SERPINH1, SFRP2, SFRS2, SHCBP1, SLC17A9, SLC19A1, SLC25A1, SLC2A10, SLC38A5, SLC39A7, SLC6A17, SLC7A5, SMOC1, SMTN, SMURF2, SP7, SPHK1, STC2, STK17A, STX5, TACC3, TAGLN, TBC1D10A, TCAM1, TGFB1I1, THBS1, TIMP3, TJP2, TK1, TMED1, TNC, TNFRSF10A, TNFRSF12A, TOM1, TRIM47, TRIP6, TUBB2B, TUBB2C, TUBB6, TYRO3, UBE2S, UHRF1, VASN, VASP, VAT1, VAX2, VEGFC, VWA1, WDR51A, WHSC1, WISP1, WISP2, WNT5A, WNT5B, Y1F1B, YWHAH, ZYX

In Adipozyten im Vergleich zu jungen MSC differenziell stärker exprimierte Gene (Top 100)

ABCA1, ABI3BP, ACSL1, ADAMTS2, ANGPTL4, ANPEP, AOX1, APP, ARRDC3, ASAM, ASPN, B4GALT1, BICC1, BNIP3L, BTG1, C1R, C1S, CCNG1, CDH11, CHRDL1, CLDN11, COL5A3, CP, CRABP1, CRISPLD2, CXCL12, CXCL14, DCN, DDIT4, DGAT2, DHRS3, ECM1, ECM2, EDIL3, EPAS1, FABP4, FAT4, FBLN1, FKBP5, FN1, FRMD6, FTH1, GLUL, GPAM, GSN, IGFBP5, ITGB5, KLF9, LOC100138434, LOC785446, LOC786490, LUM, MAOA, MASP1, MGST1, MMP13, MVP, MXRA5, MYCBP2, NDRG1, NUDT4, ODZ4, OLFML2B, P4HA3, PCYOX1, PDGFRA, PDGFRB, PLIN1, PLSCR4, POSTN, PRNP, PRSS23, PTEN, PTGFR, PTPN13, PTPRS, PTX3, REV3L, RGL1, RPL12, SCG2, SEMA6D, SEPP1, SERPINE2, SERPINF1, SESN3, SOD2, SPON1, SST, STAT3, SULT1A1, SVIL, TF, TFP12, TGFB1, THY1, TMEM100, TP53INP1, UGP2, XDH

In Osteoblasten im Vergleich zu jungen MSC differenziell stärker exprimierte Gene (Top 100)

AK6, ANXA1, APOD, ARF4, ATOX1, ATP5C1, ATP5H, ATP5I, ATP5J, ATP5L, ATP5O, ATP1F1, BGLAP, C29H11orf10, C5H12ORF57, C7H19orf56, CD63, CD9, COMMD3, COX4, COX5B, COX8A, CST3, CXCL14, DCN, DSTN, EEF1B, EIF3E, EIF3M, FAU, FGF7, FSTL1, FTH1, GNAS, GNG11, H3F3B, ID3, LOC100138915, LOC617040, LOC781039, LOC782233, LOC783718, LOC785691, LOC787914, LOC788801, LUM, MGC148714, MGST3, MP68, MRPS6, NDUFA2, NDUFA3, NDUFA4, NDUFA5, NPM1, OGN, OMD, PDLIM1, PDLIM4, PEBP1, PRDX2, PRSS23, PSMA3, RAB5B, RPL10A, RPL15, RPL22L1, RPL23A, RPL26, RPL27, RPL27A, RPL30, RPL35A, RPL37, RPL37A, RPL39, RPL8, RPS11, RPS15, RPS18, RPS26, RPS27, RPS27L, RPS29, RPS3, RPS3A, RPS5, RSU1, SERF2, SH3BGR3, SKP1, SNRPG, SOD2, SPP1, STRA13, TNFRSF11B, TOMM6, UQCR10, VKORC1, WIF1

In alten MSC im Vergleich zu jungen MSC differenziell stärker exprimierte Gene (Top 100)

ABI3BP, ADM, ANXA1, APOD, ASPN, ATOX1, ATP5H, ATP5I, ATP5J, ATP5L, ATP5O, ATP1F1, BZW1, C29H11orf10, C5H12ORF57, CCL2, CD9, CRABP1, CST3, CST6, CSTB, CTSS, CXCL14, CXCL5, DCN, DEFB4, EBD, EEF1B, EIF3M, FAU, FGF7, FTH1, FTL, GNG11, GPX4, H3F3B, IFI6, LDHA, LGALS3, LIPA, LOC100138915, LOC100296508, LOC100335395, LOC100335409, LOC100335677, LOC617040, LOC781039, LOC782233, LOC783718, LOC785691, LOC788801, LUM, MGC148714, MGP, MMP9, MP68, NDUFA4, PDLIM4, PEBP1, PRDX2, PRSS23, PRSS35, PSMA1, RBM3, RPL10A, RPL23, RPL23A, RPL26, RPL27, RPL30, RPL34, RPL35A, RPL37, RPL37A, RPL39, RPS11, RPS15, RPS15A, RPS18, RPS26, RPS27, RPS27L, RPS29, RPS3, RPS3A, RPS5, RPS6, S100A12, SAA3, SAT1, SH3BGR3, SKP1, SNRPG, SOD2, SPP1, SST, SULT1A1, TIMP1, TSPO, UQCR10

In chondrogen differenzierten Zellen im Vergleich zu jungen MSC differenziell stärker exprimierte Gene (Top 100)

ATOX1, ATP5H, ATP5I, ATP5L, ATP6V0B, BTF3, C29H11orf10, C5H12ORF57, C7H19orf56, CCL2, CD9, CDKN1A, COX4, COX5B, CST3, CST6, CSTB, CUTA, FAU, FTH1, FTL, GDF15, GNG11, GPX4, GSTP1, H3F3B, HINT1, IER3, LOC100295169, LOC100295775, LOC100336324, LOC781039, LOC781724, LOC783718, LOC787914, LOC788801, MMP1, MP68, MT2A, NDUFA3, NDUFA4, NDUFA7, NEURL2, OSTF1, PDLIM4, PERP, PLAUR, PLTP, POLR3K, PPIA, PRDX1, PRDX2, PSENEN, PSMA7, RABAC1, RNASEK, RPL10, RPL10A, RPL15, RPL18, RPL23A, RPL27, RPL27A, RPL29, RPL35A, RPL37, RPL37A, RPL39, RPLP0, RPS11, RPS15, RPS16, RPS18, RPS19, RPS2, RPS20, RPS21, RPS26, RPS27, RPS27L, RPS28, RPS29, RPS3, RPS5, SAT1, SEPW1, SNRNP25, SNRPG, STRA13, SULT1A1, SUPT4H1, TIMP1, TMSB10, TOMM6, TRMT112, TSPO, TUFM, UBA52, UCHL1, UQCR10

11 LITERATURVERZEICHNIS

1. Morrison SJ, Shah NM, Anderson DJ: **Regulatory mechanisms in stem cell biology.** *Cell* 1997, **88**(3):287-298.
2. Haeckel E: **Natürliche Schöpfungsgeschichte : gemeinverständliche wissenschaftliche Vorträge über die Entwicklungslehre im Allgemeinen und diejenige von Darwin, Goethe und Lamarck im Besonderen.** Berlin: Verlag Georg Reimer Berlin; 1868.
3. Haeckel E: **Anthropogenie**, 3 edn. Leipzig: Verlag von Wilhelm Engelmann, Leipzig; 1877.
4. Edwards RG: **Maturation in vitro of human ovarian oocytes.** *Lancet* 1965, **2**(7419):926-929.
5. Edwards RG: **Maturation in vitro of mouse, sheep, cow, pig, rhesus monkey and human ovarian oocytes.** *Nature* 1965, **208**(5008):349-351.
6. Chang MC: **Fertilization of rabbit ova in vitro.** *Nature* 1959, **184**(Suppl 7):466-467.
7. Steptoe PC, Edwards RG: **Birth after the reimplantation of a human embryo.** *Lancet* 1978, **2**(8085):366.
8. Cole RJ, Edwards RG, Paul J: **Cytodifferentiation and embryogenesis in cell colonies and tissue cultures derived from ova and blastocysts of the rabbit.** *Dev Biol* 1966, **13**(3):385-407.
9. Evans MJ, Kaufman MH: **Establishment in culture of pluripotential cells from mouse embryos.** *Nature* 1981, **292**(5819):154-156.
10. Bonnet D, Dick JE: **Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell.** *Nat Med* 1997, **3**(7):730-737.
11. Takahashi K, Ichisaka T, Yamanaka S: **Identification of genes involved in tumor-like properties of embryonic stem cells.** *Methods Mol Biol* 2006, **329**:449-458.
12. Rideout WM, 3rd, Hochedlinger K, Kyba M, Daley GQ, Jaenisch R: **Correction of a genetic defect by nuclear transplantation and combined cell and gene therapy.** *Cell* 2002, **109**(1):17-27.
13. Takahashi K, Yamanaka S: **Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors.** *Cell* 2006, **126**(4):663-676.

14. Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, Yamanaka S: **Induction of pluripotent stem cells from adult human fibroblasts by defined factors.** *Cell* 2007, **131**(5):861-872.
15. Nakagawa M, Koyanagi M, Tanabe K, Takahashi K, Ichisaka T, Aoi T, Okita K, Mochiduki Y, Takizawa N, Yamanaka S: **Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts.** *Nat Biotechnol* 2008, **26**(1):101-106.
16. Park IH, Zhao R, West JA, Yabuuchi A, Huo H, Ince TA, Lerou PH, Lensch MW, Daley GQ: **Reprogramming of human somatic cells to pluripotency with defined factors.** *Nature* 2008, **451**(7175):141-146.
17. Yamanaka S: **A fresh look at iPS cells.** *Cell* 2009, **137**(1):13-17.
18. Woltjen K, Michael IP, Mohseni P, Desai R, Mileikovsky M, Hamalainen R, Cowling R, Wang W, Liu P, Gertsenstein M *et al*: **piggyBac transposition reprograms fibroblasts to induced pluripotent stem cells.** *Nature* 2009, **458**(7239):766-770.
19. Zhou H, Wu S, Joo JY, Zhu S, Han DW, Lin T, Trauger S, Bien G, Yao S, Zhu Y *et al*: **Generation of induced pluripotent stem cells using recombinant proteins.** *Cell Stem Cell* 2009, **4**(5):381-384.
20. Laurent LC, Ulitsky I, Slavin I, Tran H, Schork A, Morey R, Lynch C, Harness JV, Lee S, Barrero MJ *et al*: **Dynamic changes in the copy number of pluripotency and cell proliferation genes in human ESCs and iPSCs during reprogramming and time in culture.** *Cell Stem Cell*, **8**(1):106-118.
21. Larijani B, Esfahani EN, Amini P, Nikbin B, Alimoghaddam K, Amiri S, Malekzadeh R, Yazdi NM, Ghodsi M, Dowlati Y *et al*: **Stem cell therapy in treatment of different diseases.** *Acta Med Iran* 2011, **50**(2):79-96.
22. Ratcliffe E, Glen KE, Naing MW, Williams DJ: **Current status and perspectives on stem cell-based therapies undergoing clinical trials for regenerative medicine: case studies.** *Br Med Bull* 2013, **108**:73-94.
23. Maximov A. A: **Der Lymphozyt als gemeinsame Stammzelle der verschiedenen Blutelemente in der embryonalen Entwicklung und im postfetalen Leben der Säugetiere.** *Demonstrationsvortrag, gehalten in der außerordentlichen Sitzung der Berliner Hämatologischen Gesellschaft am 1 Juni 1909* 1909, **8**(1909):125-134.
24. Becker AJ, Mc CE, Till JE: **Cytological demonstration of the clonal nature of spleen colonies derived from transplanted mouse marrow cells.** *Nature* 1963, **197**:452-454.
25. Friedenstein AJ, Deriglasova UF, Kulagina NN, Panasuk AF, Rudakowa SF, Luria EA, Ruadkow IA: **Precursors for fibroblasts in different populations of hematopoietic cells as detected by the in vitro colony assay method.** *Exp Hematol* 1974, **2**(2):83-92.
26. Pittenger MF, Mackay AM, Beck SC, Jaiswal RK, Douglas R, Mosca JD, Moorman MA, Simonetti DW, Craig S, Marshak DR: **Multilineage potential of adult human mesenchymal stem cells.** *Science* 1999, **284**(5411):143-147.

27. Kuhn NZ, Tuan RS: **Regulation of stemness and stem cell niche of mesenchymal stem cells: implications in tumorigenesis and metastasis.** *J Cell Physiol*, **222**(2):268-277.
28. Gotts JE, Matthay MA: **Mesenchymal stem cells and the stem cell niche: a new chapter.** *Am J Physiol Lung Cell Mol Physiol*, **302**(11):L1147-1149.
29. Battiwalla M, Hematti P: **Mesenchymal stem cells in hematopoietic stem cell transplantation.** *Cytotherapy* 2009, **11**(5):503-515.
30. de Barros AP, Takiya CM, Garzoni LR, Leal-Ferreira ML, Dutra HS, Chiarini LB, Meirelles MN, Borojevic R, Rossi MI: **Osteoblasts and bone marrow mesenchymal stromal cells control hematopoietic stem cell migration and proliferation in 3D in vitro model.** *PLoS One*, **5**(2):e9093.
31. Jing D, Fonseca AV, Alakel N, Fierro FA, Muller K, Bornhauser M, Ehninger G, Corbeil D, Ordemann R: **Hematopoietic stem cells in co-culture with mesenchymal stromal cells--modeling the niche compartments in vitro.** *Haematologica*, **95**(4):542-550.
32. Kolf CM, Cho E, Tuan RS: **Mesenchymal stromal cells. Biology of adult mesenchymal stem cells: regulation of niche, self-renewal and differentiation.** *Arthritis Res Ther* 2007, **9**(1):204.
33. Phinney DG: **Functional heterogeneity of mesenchymal stem cells: implications for cell therapy.** *J Cell Biochem* 2012, **113**(9):2806-2812.
34. Heo JS, Choi SM, Kim HO, Kim EH, You J, Park T, Kim E, Kim HS: **Neural transdifferentiation of human bone marrow mesenchymal stem cells on hydrophobic polymer-modified surface and therapeutic effects in an animal model of ischemic stroke.** *Neuroscience* 2013, **238**:305-318.
35. Catacchio I, Berardi S, Reale A, De Luisi A, Racanelli V, Vacca A, Ria R: **Evidence for bone marrow adult stem cell plasticity: properties, molecular mechanisms, negative aspects, and clinical applications of hematopoietic and mesenchymal stem cells transdifferentiation.** *Stem Cells Int* 2013, **2013**:589139.
36. Song L, Tuan RS: **Transdifferentiation potential of human mesenchymal stem cells derived from bone marrow.** *FASEB J* 2004, **18**(9):980-982.
37. Trounson A, Thakar RG, Lomax G, Gibbons D: **Clinical trials for stem cell therapies.** *BMC Med* 2011, **9**:52.
38. Eslaminejad MB, Fani N, Shahhoseini M: **Epigenetic regulation of osteogenic and chondrogenic differentiation of mesenchymal stem cells in culture.** *Cell J*, **15**(1):1-10.
39. Alwine JC, Kemp DJ, Parker BA, Reiser J, Renart J, Stark GR, Wahl GM: **Detection of specific RNAs or specific fragments of DNA by fractionation in gels and transfer to diazobenzylmethyl paper.** *Methods Enzymol* 1979, **68**:220-242.
40. Delidow BC, Peluso JJ, White BA: **Quantitative measurement of mRNAs by polymerase chain reaction.** *Gene Anal Tech* 1989, **6**(6):120-124.
41. Maxam AM, Gilbert W: **A new method for sequencing DNA.** *Proc Natl Acad Sci U S A* 1977, **74**(2):560-564.

42. Sanger F, Coulson AR: **A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase.** *J Mol Biol* 1975, **94**(3):441-448.
43. Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, Merril CR, Wu A, Olde B, Moreno RF *et al*: **Complementary DNA sequencing: expressed sequence tags and human genome project.** *Science* 1991, **252**(5013):1651-1656.
44. Velculescu VE, Zhang L, Vogelstein B, Kinzler KW: **Serial analysis of gene expression.** *Science* 1995, **270**(5235):484-487.
45. Matsumura H, Urasaki N, Yoshida K, Kruger DH, Kahl G, Terauchi R: **SuperSAGE: powerful serial analysis of gene expression.** *Methods Mol Biol*, **883**:1-17.
46. Schena M, Shalon D, Davis RW, Brown PO: **Quantitative monitoring of gene expression patterns with a complementary DNA microarray.** *Science* 1995, **270**(5235):467-470.
47. Duggan DJ, Bittner M, Chen Y, Meltzer P, Trent JM: **Expression profiling using cDNA microarrays.** *Nat Genet* 1999, **21**(1 Suppl):10-14.
48. Stratton MR, Campbell PJ, Futreal PA: **The cancer genome.** *Nature* 2009, **458**(7239):719-724.
49. Linnarsson S: **Recent advances in DNA sequencing methods - general principles of sample preparation.** *Exp Cell Res*, **316**(8):1339-1343.
50. Dressman D, Yan H, Traverso G, Kinzler KW, Vogelstein B: **Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations.** *Proc Natl Acad Sci U S A* 2003, **100**(15):8817-8822.
51. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z *et al*: **Genome sequencing in microfabricated high-density picolitre reactors.** *Nature* 2005, **437**(7057):376-380.
52. Liu L, Li Y, Li S, Hu N, He Y, Pong R, Lin D, Lu L, Law M: **Comparison of next-generation sequencing systems.** *J Biomed Biotechnol* 2012, **2012**:251364.
53. Fedurco M, Romieu A, Williams S, Lawrence I, Turcatti G: **BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies.** *Nucleic Acids Res* 2006, **34**(3):e22.
54. Adey A, Morrison HG, Asan, Xun X, Kitzman JO, Turner EH, Stackhouse B, MacKenzie AP, Caruccio NC, Zhang X *et al*: **Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition.** *Genome Biol* 2010, **11**(12):R119.
55. Glenn TC: **Field guide to next-generation DNA sequencers (Data for 2014 available online only).** *Mol Ecol Resour* 2011, **11**(5):759-769.
56. Mardis ER: **Next-generation DNA sequencing methods.** *Annu Rev Genomics Hum Genet* 2008, **9**:387-402.
57. Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, Leamon JH, Johnson K, Milgrew MJ, Edwards M *et al*: **An integrated semiconductor device enabling non-optical genome sequencing.** *Nature* 2011, **475**(7356):348-352.

58. Korlach J, Marks PJ, Cicero RL, Gray JJ, Murphy DL, Roitman DB, Pham TT, Otto GA, Foquet M, Turner SW: **Selective aluminum passivation for targeted immobilization of single DNA polymerase molecules in zero-mode waveguide nanostructures.** *Proc Natl Acad Sci U S A* 2008, **105**(4):1176-1181.
59. Liu T, Tsai CH, Lee WB, Chiang JH: **Optimizing information in Next-Generation-Sequencing (NGS) reads for improving de novo genome assembly.** *PLoS One* 2013, **8**(7):e69503.
60. He W, Zhao S, Liu X, Dong S, Lv J, Liu D, Wang J, Meng Z: **ReSeqTools: an integrated toolkit for large-scale next-generation sequencing based resequencing analysis.** *Genet Mol Res* 2013, **12**(4):6275-6283.
61. Zhao W, Zhu QY, Zhang JT, Liu H, Wang LJ, Chen ZQ, Guan LP, Huang XS, Yang L, Yu SY: **Exome sequencing identifies novel compound heterozygous mutations in SPG11 that cause autosomal recessive hereditary spastic paraplegia.** *J Neurol Sci*, **335**(1-2):112-117.
62. Schweiger MR, Kerick M, Timmermann B, Isau M: **The power of NGS technologies to delineate the genome organization in cancer: from mutations to structural variations and epigenetic alterations.** *Cancer Metastasis Rev*, **30**(2):199-210.
63. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215**(3):403-410.
64. de Bruijn NG: **A Combinatorial Problem.** *Koninklijke Nederlandse Akademie v Wetenschappen* 1946, **49**:758-764.
65. Compeau PE, Pevzner PA, Tesler G: **How to apply de Bruijn graphs to genome assembly.** *Nat Biotechnol* 2011, **29**(11):987-991.
66. Zerbino DR, Birney E: **Velvet: algorithms for de novo short read assembly using de Bruijn graphs.** *Genome Res* 2008, **18**(5):821-829.
67. Wang Z, Gerstein M, Snyder M: **RNA-Seq: a revolutionary tool for transcriptomics.** *Nat Rev Genet* 2009, **10**(1):57-63.
68. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B: **Mapping and quantifying mammalian transcriptomes by RNA-Seq.** *Nat Methods* 2008, **5**(7):621-628.
69. Elvik CG, Tellam RL, Worley KC, Gibbs RA, Muzny DM, Weinstock GM, Adelson DL, Eichler EE, Elnitski L, Guigo R *et al*: **The genome sequence of taurine cattle: a window to ruminant biology and evolution.** *Science* 2009, **324**(5926):522-528.
70. Bosnakovski D, Mizuno M, Kim G, Takagi S, Okumura M, Fujinaga T: **Isolation and multilineage differentiation of bovine bone marrow mesenchymal stem cells.** *Cell Tissue Res* 2005, **319**(2):243-253.
71. Dominici M, Le Blanc K, Mueller I, Slaper-Cortenbach I, Marini F, Krause D, Deans R, Keating A, Prockop D, Horwitz E: **Minimal criteria for defining multipotent mesenchymal stromal cells. The International Society for Cellular Therapy position statement.** *Cytotherapy* 2006, **8**(4):315-317.
72. Guo KT, SchAfer R, Paul A, Gerber A, Ziemer G, Wendel HP: **A new technique for the isolation and surface immobilization of mesenchymal**

- stem cells from whole bone marrow using high-specific DNA aptamers. *Stem Cells* 2006, **24**(10):2220-2231.
73. Sung JH, Yang HM, Park JB, Choi GS, Joh JW, Kwon CH, Chun JM, Lee SK, Kim SJ: **Isolation and characterization of mouse mesenchymal stem cells.** *Transplant Proc* 2008, **40**(8):2649-2654.
74. Jung H, Wang SY, Yang IW, Hsueh DW, Yang WJ, Wang TH, Wang HS: **Detection and treatment of mycoplasma contamination in cultured cells.** *Chang Gung Med J* 2003, **26**(4):250-258.
75. Song L, Webb NE, Song Y, Tuan RS: **Identification and functional analysis of candidate genes regulating mesenchymal stem cell self-renewal and multipotency.** *Stem Cells* 2006, **24**(7):1707-1718.
76. Ramirez-Zacarias JL, Castro-Munozledo F, Kuri-Harcuch W: **Quantitation of adipose conversion and triglycerides by staining intracytoplasmic lipids with Oil red O.** *Histochemistry* 1992, **97**(6):493-497.
77. Birnboim HC, Doly J: **A rapid alkaline extraction procedure for screening recombinant plasmid DNA.** *Nucleic Acids Res* 1979, **7**(6):1513-1523.
78. Diehl F, Grahlmann S, Beier M, Hoheisel JD: **Manufacturing DNA microarrays of high spot homogeneity and reduced background signal.** *Nucleic Acids Res* 2001, **29**(7):E38.
79. Rosenzweig BA, Pine PS, Domon OE, Morris SM, Chen JJ, Sistare FD: **Dye bias correction in dual-labeled cDNA microarray gene expression measurements.** *Environ Health Perspect* 2004, **112**(4):480-487.
80. Park T, Yi SG, Kang SH, Lee S, Lee YS, Simon R: **Evaluation of normalization methods for microarray data.** *BMC Bioinformatics* 2003, **4**:33.
81. Yang IV, Chen E, Hasseman JP, Liang W, Frank BC, Wang S, Sharov V, Saeed AI, White J, Li J *et al*: **Within the fold: assessing differential expression measures and reproducibility in microarray assays.** *Genome Biol* 2002, **3**(11):research0062.
82. Rapp S: **Genexpressionsanalyse der fötalen Wachstumsfuge des Rindes.** *Promotion (pH.D.)*. Mainz: Johannes Gutenberg Universität Mainz; 2011.
83. Kircher M, Heyn P, Kelso J: **Addressing challenges in the production and analysis of illumina sequencing data.** *BMC Genomics*, **12**:382.
84. Langmead B, Trapnell C, Pop M, Salzberg SL: **Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.** *Genome Biol* 2009, **10**(3):R25.
85. Stekel DJ, Git Y, Falciani F: **The comparison of gene expression from multiple cDNA libraries.** *Genome Res* 2000, **10**(12):2055-2061.
86. Huang da W, Sherman BT, Lempicki RA: **Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.** *Nat Protoc* 2009, **4**(1):44-57.
87. Huang da W, Sherman BT, Lempicki RA: **Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists.** *Nucleic Acids Res* 2009, **37**(1):1-13.
88. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT *et al*: **Gene ontology: tool for the**

- unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000, 25(1):25-29.
89. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M: **KEGG for integration and interpretation of large-scale molecular data sets.** *Nucleic Acids Res*, 40(Database issue):D109-114.
90. Kelder T, Pico AR, Hanspers K, van Iersel MP, Evelo C, Conklin BR: **Mining biological pathways using WikiPathways web services.** *PLoS One* 2009, 4(7):e6447.
91. Cline MS, Smoot M, Cerami E, Kuchinsky A, Landys N, Workman C, Christmas R, Avila-Campilo I, Creech M, Gross B *et al*: **Integration of biological networks and gene expression data using Cytoscape.** *Nat Protoc* 2007, 2(10):2366-2382.
92. Zambon AC, Gaj S, Ho I, Hanspers K, Vranizan K, Evelo CT, Conklin BR, Pico AR, Salomonis N: **GO-Elite: a flexible solution for pathway and ontology over-representation.** *Bioinformatics* 2012, 28(16):2209-2210.
93. Zheng Q, Wang XJ: **GOEAST: a web-based software toolkit for Gene Ontology enrichment analysis.** *Nucleic Acids Res* 2008, 36(Web Server issue):W358-363.
94. Supek F, Bosnjak M, Skunca N, Smuc T: **REVIGO summarizes and visualizes long lists of gene ontology terms.** *PLoS One* 2011, 6(7):e21800.
95. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M: **Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research.** *Bioinformatics* 2005, 21(18):3674-3676.
96. Hunter S, Jones P, Mitchell A, Apweiler R, Attwood TK, Bateman A, Bernard T, Binns D, Bork P, Burge S *et al*: **InterPro in 2011: new developments in the family and domain prediction database.** *Nucleic Acids Res*, 40(Database issue):D306-312.
97. Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R: **InterProScan: protein domains identifier.** *Nucleic Acids Res* 2005, 33(Web Server issue):W116-120.
98. Milne I, Stephen G, Bayer M, Cock PJ, Pritchard L, Cardle L, Shaw PD, Marshall D: **Using Tablet for visual exploration of second-generation sequencing data.** *Brief Bioinform*, 14(2):193-202.
99. Okonechnikov K, Golosova O, Fursov M: **Unipro UGENE: a unified bioinformatics toolkit.** *Bioinformatics* 2012, 28(8):1166-1167.
100. Zhang L, Chan C: **Isolation and enrichment of rat mesenchymal stem cells (MSCs) and separation of single-colony derived MSCs.** *J Vis Exp* 2010(37).
101. Go MJ, Takenaka C, Ohgushi H: **Forced expression of Sox2 or Nanog in human bone marrow derived mesenchymal stem cells maintains their expansion and differentiation capabilities.** *Exp Cell Res* 2008, 314(5):1147-1154.
102. Greco SJ, Liu K, Rameshwar P: **Functional similarities among genes regulated by OCT4 in human mesenchymal and embryonic stem cells.** *Stem Cells* 2007, 25(12):3143-3154.

103. Park SB, Seo KW, So AY, Seo MS, Yu KR, Kang SK, Kang KS: **SOX2 has a crucial role in the lineage determination and proliferation of mesenchymal stem cells through Dickkopf-1 and c-MYC.** *Cell Death Differ* 2012, **19**(3):534-545.
104. Baddoo M, Hill K, Wilkinson R, Gaupp D, Hughes C, Kopen GC, Phinney DG: **Characterization of mesenchymal stem cells isolated from murine bone marrow by negative selection.** *J Cell Biochem* 2003, **89**(6):1235-1249.
105. Lai WT, Krishnappa V, Phinney DG: **Fibroblast growth factor 2 (Fgf2) inhibits differentiation of mesenchymal stem cells by inducing Twist2 and Spry4, blocking extracellular regulated kinase activation, and altering Fgf receptor expression levels.** *Stem Cells* 2011, **29**(7):1102-1111.
106. Vacanti V, Kong E, Suzuki G, Sato K, Canty JM, Lee T: **Phenotypic changes of adult porcine mesenchymal stem cells induced by prolonged passaging in culture.** *J Cell Physiol* 2005, **205**(2):194-201.
107. Kim DH, Yoo KH, Choi KS, Choi J, Choi SY, Yang SE, Yang YS, Im HJ, Kim KH, Jung HL *et al*: **Gene expression profile of cytokine and growth factor during differentiation of bone marrow-derived mesenchymal stem cell.** *Cytokine* 2005, **31**(2):119-126.
108. Bhandari DR, Seo KW, Jung JW, Kim HS, Yang SR, Kang KS: **The regulatory role of c-MYC on HDAC2 and PcG expression in human multipotent stem cells.** *J Cell Mol Med* 2011, **15**(7):1603-1614.
109. Shimizu T, Ishikawa T, Sugihara E, Kuninaka S, Miyamoto T, Mabuchi Y, Matsuzaki Y, Tsunoda T, Miya F, Morioka H *et al*: **c-MYC overexpression with loss of Ink4a/Arf transforms bone marrow stromal cells into osteosarcoma accompanied by loss of adipogenesis.** *Oncogene* 2010, **29**(42):5687-5699.
110. Jaramillo-Ferrada PA, Wolvetang EJ, Cooper-White JJ: **Differential mesengenic potential and expression of stem cell-fate modulators in mesenchymal stromal cells from human-term placenta and bone marrow.** *J Cell Physiol* 2012, **227**(9):3234-3242.
111. Pacini S, Carnicelli V, Trombi L, Montali M, Fazzi R, Lazzarini E, Giannotti S, Petrini M: **Constitutive expression of pluripotency-associated genes in mesodermal progenitor cells (MPCs).** *PLoS One* 2010, **5**(3):e9861.
112. Saulnier N, Puglisi MA, Lattanzi W, Castellini L, Pani G, Leone G, Alfieri S, Michetti F, Piscaglia AC, Gasbarrini A: **Gene profiling of bone marrow- and adipose tissue-derived stromal cells: a key role of Kruppel-like factor 4 in cell fate regulation.** *Cytotherapy* 2011, **13**(3):329-340.
113. Vella P, Barozzi I, Cuomo A, Bonaldi T, Pasini D: **Yin Yang 1 extends the Myc-related transcription factors network in embryonic stem cells.** *Nucleic Acids Res* 2012, **40**(8):3403-3418.
114. Kim JD, Faulk C, Kim J: **Retroposition and evolution of the DNA-binding motifs of YY1, YY2 and REX1.** *Nucleic Acids Res* 2007, **35**(10):3442-3452.
115. Wei X, Shen CY: **Transcriptional regulation of oct4 in human bone marrow mesenchymal stem cells.** *Stem Cells Dev* 2011, **20**(3):441-449.

116. Liu TM, Wu YN, Guo XM, Hui JH, Lee EH, Lim B: **Effects of ectopic Nanog and Oct4 overexpression on mesenchymal stem cells.** *Stem Cells Dev* 2009, **18**(7):1013-1022.
117. Riekstina U, Cakstina I, Parfejevs V, Hoogduijn M, Jankovskis G, Muiznieks I, Muceniece R, Ancans J: **Embryonic stem cell marker expression pattern in human mesenchymal stem cells derived from bone marrow, adipose tissue, heart and dermis.** *Stem cell reviews* 2009, **5**(4):378-386.
118. Majumdar MK, Thiede MA, Haynesworth SE, Bruder SP, Gerson SL: **Human marrow-derived mesenchymal stem cells (MSCs) express hematopoietic cytokines and support long-term hematopoiesis when differentiated toward stromal and osteogenic lineages.** *J Hematother Stem Cell Res* 2000, **9**(6):841-848.
119. Mok PL, Cheong SK, Leong CF: **In-vitro differentiation study on isolated human mesenchymal stem cells.** *Malays J Pathol* 2008, **30**(1):11-19.
120. Schroeder A, Mueller O, Stocker S, Salowsky R, Leiber M, Gassmann M, Lightfoot S, Menzel W, Granzow M, Ragg T: **The RIN: an RNA integrity number for assigning integrity values to RNA measurements.** *BMC Mol Biol* 2006, **7**:3.
121. Janning W, Knust E: **Genetik:** Thieme Verlag; 2004.
122. Costa V, Angelini C, De Feis I, Ciccodicola A: **Uncovering the complexity of transcriptomes with RNA-Seq.** *J Biomed Biotechnol*, **2010**:853916.
123. Eguchi J, Wada J, Hida K, Zhang H, Matsuoka T, Baba M, Hashimoto I, Shikata K, Ogawa N, Makino H: **Identification of adipocyte adhesion molecule (ACAM), a novel CTX gene family, implicated in adipocyte maturation and development of obesity.** *Biochem J* 2005, **387**(Pt 2):343-353.
124. Yagi M, Suzuki N, Takayama T, Arisue M, Kodama T, Yoda Y, Numasaki H, Otsuka K, Ito K: **Lactoferrin suppress the adipogenic differentiation of MC3T3-G2/PA6 cells.** *J Oral Sci* 2008, **50**(4):419-425.
125. Komori T: **Regulation of bone development and extracellular matrix protein genes by RUNX2.** *Cell Tissue Res*, **339**(1):189-195.
126. Zhu F, Friedman MS, Luo W, Woolf P, Hankenson KD: **The transcription factor osterix (SP7) regulates BMP6-induced human osteoblast differentiation.** *J Cell Physiol* 2011, **227**(6):2677-2685.
127. Shen Z, Gantcheva S, Mansson B, Heinegard D, Sommarin Y: **Chondroadherin expression changes in skeletal development.** *Biochem J* 1998, **330** (Pt 1):549-557.
128. Bosnakovski D, Mizuno M, Kim G, Takagi S, Okumura M, Fujinaga T: **Chondrogenic differentiation of bovine bone marrow mesenchymal stem cells (MSCs) in different hydrogels: influence of collagen type II extracellular matrix on MSC chondrogenesis.** *Biotechnol Bioeng* 2006, **93**(6):1152-1163.
129. Watanabe H, Yamada Y: **Mice lacking link protein develop dwarfism and craniofacial abnormalities.** *Nat Genet* 1999, **21**(2):225-229.
130. Imabuchi R, Ohmiya Y, Kwon HJ, Onodera S, Kitamura N, Kurokawa T, Gong JP, Yasuda K: **Gene expression profile of the cartilage tissue**

- spontaneously regenerated in vivo by using a novel double-network gel: comparisons with the normal articular cartilage. *BMC Musculoskelet Disord*, **12**:213.
131. Mauck RL, Yuan X, Tuan RS: **Chondrogenic differentiation and functional maturation of bovine mesenchymal stem cells in long-term agarose culture.** *Osteoarthritis Cartilage* 2006, **14**(2):179-189.
 132. Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, Zucker JP, Guenther MG, Kumar RM, Murray HL, Jenner RG *et al*: **Core transcriptional regulatory circuitry in human embryonic stem cells.** *Cell* 2005, **122**(6):947-956.
 133. Pan G, Thomson JA: **Nanog and transcriptional networks in embryonic stem cell pluripotency.** *Cell Res* 2007, **17**(1):42-49.
 134. Lengner CJ, Camargo FD, Hochedlinger K, Welstead GG, Zaidi S, Gokhale S, Scholer HR, Tomilin A, Jaenisch R: **Oct4 expression is not required for mouse somatic stem cell self-renewal.** *Cell Stem Cell* 2007, **1**(4):403-415.
 135. Angonin D, Van Raay TJ: **Nkd1 functions as a passive antagonist of Wnt signaling.** *PLoS One*, **8**(8):e74666.
 136. Van Raay TJ, Fortino NJ, Miller BW, Ma H, Lau G, Li C, Franklin JL, Attisano L, Solnica-Krezel L, Coffey RJ: **Naked1 antagonizes Wnt signaling by preventing nuclear accumulation of beta-catenin.** *PLoS One*, **6**(4):e18650.
 137. Sen GL, Reuter JA, Webster DE, Zhu L, Khavari PA: **DNMT1 maintains progenitor function in self-renewing somatic tissue.** *Nature*, **463**(7280):563-567.
 138. Lo Furno D, Graziano AC, Caggia S, Perrotta RE, Tarico MS, Giuffrida R, Cardile V: **Decrease of apoptosis markers during adipogenic differentiation of mesenchymal stem cells from human adipose tissue.** *Apoptosis*, **18**(5):578-588.
 139. Bosnakovski D, Mizuno M, Kim G, Ishiguro T, Okumura M, Iwanaga T, Kadosawa T, Fujinaga T: **Chondrogenic differentiation of bovine bone marrow mesenchymal stem cells in pellet cultural system.** *Exp Hematol* 2004, **32**(5):502-509.
 140. Kanawa M, Igarashi A, Ronald VS, Higashi Y, Kurihara H, Sugiyama M, Saskianti T, Pan H, Kato Y: **Age-dependent decrease in the chondrogenic potential of human bone marrow mesenchymal stromal cells expanded with fibroblast growth factor-2.** *Cytotherapy*, **15**(9):1062-1072.
 141. Kretlow JD, Jin YQ, Liu W, Zhang WJ, Hong TH, Zhou G, Baggett LS, Mikos AG, Cao Y: **Donor age and cell passage affects differentiation potential of murine bone marrow-derived stem cells.** *BMC Cell Biol* 2008, **9**:60.
 142. Baksh D, Song L, Tuan RS: **Adult mesenchymal stem cells: characterization, differentiation, and application in cell and gene therapy.** *J Cell Mol Med* 2004, **8**(3):301-316.
 143. Post S, Abdallah BM, Bentzon JF, Kassem M: **Demonstration of the presence of independent pre-osteoblastic and pre-adipocytic cell populations in bone marrow-derived mesenchymal stem cells.** *Bone* 2008, **43**(1):32-39.

144. Armesilla-Diaz A, Elvira G, Silva A: **p53 regulates the proliferation, differentiation and spontaneous transformation of mesenchymal stem cells.** *Exp Cell Res* 2009, **315**(20):3598-3610.
145. Gou S, Wang C, Liu T, Wu H, Xiong J, Zhou F, Zhao G: **Spontaneous differentiation of murine bone marrow-derived mesenchymal stem cells into adipocytes without malignant transformation after long-term culture.** *Cells Tissues Organs*, **191**(3):185-192.
146. Tseng PC, Young TH, Wang TM, Peng HW, Hou SM, Yen ML: **Spontaneous osteogenesis of MSCs cultured on 3D microcarriers through alteration of cytoskeletal tension.** *Biomaterials*, **33**(2):556-564.
147. Sommer SS, Cohen JE: **The size distributions of proteins, mRNA, and nuclear RNA.** *J Mol Evol* 1980, **15**(1):37-57.
148. Williams A, Thomson EM: **Effects of scanning sensitivity and multiple scan algorithms on microarray data quality.** *BMC Bioinformatics*, **11**:127.
149. Palmieri F: **The mitochondrial transporter family (SLC25): physiological and pathological implications.** *Pflugers Arch* 2004, **447**(5):689-709.
150. Hatch NE, Franceschi RT: **Osteoblast differentiation stage-specific expression of the pyrophosphate-generating enzyme PC-1.** *Cells Tissues Organs* 2009, **189**(1-4):65-69.
151. Shuang F, Sun Y, Yang HH, Shao YC, Li H, Hu W, Zhong J, Zou HX: **Dextrin deletion enhances the bone loss in hindlimb suspended mice.** *Eur J Appl Physiol* 2012.
152. Banu N, Teichman J, Dunlap-Brown M, Villegas G, Tufro A: **Semaphorin 3C regulates endothelial cell function by increasing integrin activity.** *FASEB J* 2006, **20**(12):2150-2152.
153. Tang Z, Sahu SN, Khadeer MA, Bai G, Franklin RB, Gupta A: **Overexpression of the ZIP1 zinc transporter induces an osteogenic phenotype in mesenchymal stem cells.** *Bone* 2006, **38**(2):181-198.
154. Khimani AH, Mhashilkar AM, Mikulskis A, O'Malley M, Liao J, Golenko EE, Mayer P, Chada S, Killian JB, Lott ST: **Housekeeping genes in cancer: normalization of array data.** *Biotechniques* 2005, **38**(5):739-745.
155. Barone LM, Owen TA, Tassinari MS, Bortell R, Stein GS, Lian JB: **Developmental expression and hormonal regulation of the rat matrix Gla protein (MGP) gene in chondrogenesis and osteogenesis.** *J Cell Biochem* 1991, **46**(4):351-365.
156. Zhao Q, Eberspaecher H, Lefebvre V, De Crombrughe B: **Parallel expression of Sox9 and Col2a1 in cells undergoing chondrogenesis.** *Dev Dyn* 1997, **209**(4):377-386.
157. Hu Z, Peel SA, Ho SK, Sandor GK, Clokie CM: **Role of bovine bone morphogenetic proteins in bone matrix protein and osteoblast-related gene expression during rat bone marrow stromal cell differentiation.** *J Craniofac Surg* 2005, **16**(6):1006-1014.
158. Kanaan RA, Aldwaik M, Al-Hanbali OA: **The role of connective tissue growth factor in skeletal growth and development.** *Med Sci Monit* 2006, **12**(12):RA277-281.

159. McCarthy TL, Ji C, Chen Y, Kim KK, Imagawa M, Ito Y, Centrella M: **Runt domain factor (Runx)-dependent effects on CCAAT/ enhancer-binding protein delta expression and activity in osteoblasts.** *J Biol Chem* 2000, **275**(28):21746-21753.
160. Uetsuki T, Naito A, Nagata S, Kaziro Y: **Isolation and characterization of the human chromosomal gene for polypeptide chain elongation factor-1 alpha.** *J Biol Chem* 1989, **264**(10):5791-5798.
161. Lai MC, Chang WC, Shieh SY, Tarn WY: **DDX3 regulates cell growth through translational control of cyclin E1.** *Mol Cell Biol*, **30**(22):5444-5453.
162. Yan WL, Lerner TJ, Haines JL, Gusella JF: **Sequence analysis and mapping of a novel human mitochondrial ATP synthase subunit 9 cDNA (ATP5G3).** *Genomics* 1994, **24**(2):375-377.
163. Genovese JA, Spadaccio C, Rivello HG, Toyoda Y, Patel AN: **Electrostimulated bone marrow human mesenchymal stem cells produce follistatin.** *Cytotherapy* 2009, **11**(4):448-456.
164. Wu Y, Zhou S, Smas CM: **Downregulated expression of the secreted glycoprotein follistatin-like 1 (Fstl1) is a robust hallmark of preadipocyte to adipocyte conversion.** *Mech Dev* 2010, **127**(3-4):183-202.
165. Lee KT, Byun MJ, Lim D, Kang KS, Kim NS, Oh JH, Chung CS, Park HS, Shin Y, Kim TH: **Full-length enriched cDNA library construction from tissues related to energy metabolism in pigs.** *Mol Cells* 2009, **28**(6):529-536.
166. Blencowe BJ, Ahmad S, Lee LJ: **Current-generation high-throughput sequencing: deepening insights into mammalian transcriptomes.** *Genes Dev* 2009, **23**(12):1379-1386.
167. Lee RH, Kim B, Choi I, Kim H, Choi HS, Suh K, Bae YC, Jung JS: **Characterization and expression analysis of mesenchymal stem cells from human bone marrow and adipose tissue.** *Cell Physiol Biochem* 2004, **14**(4-6):311-324.
168. Crossno JT, Jr., Majka SM, Grazia T, Gill RG, Klemm DJ: **Rosiglitazone promotes development of a novel adipocyte population from bone marrow-derived circulating progenitor cells.** *J Clin Invest* 2006, **116**(12):3220-3228.
169. Dodson MV, Hausman GJ, Guan L, Du M, Jiang Z: **Potential impact of mature adipocyte dedifferentiation in terms of cell numbers.** *Int J Stem Cells*, **4**(1):76-78.
170. Fernyhough ME, Vierck JL, Hausman GJ, Mir PS, Okine EK, Dodson MV: **Primary adipocyte culture: adipocyte purification methods may lead to a new understanding of adipose tissue growth and development.** *Cytotechnology* 2004, **46**(2-3):163-172.
171. Mathews S, Bhonde R, Gupta PK, Totey S: **Extracellular matrix protein mediated regulation of the osteoblast differentiation of bone marrow derived human mesenchymal stem cells.** *Differentiation*, **84**(2):185-192.
172. Chen XD: **Extracellular matrix provides an optimal niche for the maintenance and propagation of mesenchymal stem cells.** *Birth Defects Res C Embryo Today*, **90**(1):45-54.

173. Rodrigues M, Yates CC, Nuschke A, Griffith L, Wells A: **The matrikine tenascin-C protects multipotential stromal cells/mesenchymal stem cells from death cytokines such as FasL.** *Tissue Eng Part A*, **19**(17-18):1972-1983.
174. Popova AP, Bozyk PD, Goldsmith AM, Linn MJ, Lei J, Bentley JK, Hershenson MB: **Autocrine production of TGF-beta1 promotes myofibroblastic differentiation of neonatal lung mesenchymal stem cells.** *Am J Physiol Lung Cell Mol Physiol*, **298**(6):L735-743.
175. Ullah M, Sittinger M, Ringe J: **Extracellular matrix of adipogenically differentiated mesenchymal stem cells reveals a network of collagen filaments, mostly interwoven by hexagonal structural units.** *Matrix Biol*, **32**(7-8):452-465.
176. Nakajima I, Muroya S, Tanabe R, Chikuni K: **Extracellular matrix development during differentiation into adipocytes with a unique increase in type V and VI collagen.** *Biol Cell* 2002, **94**(3):197-203.
177. Kubo Y, Kaidzu S, Nakajima I, Takenouchi K, Nakamura F: **Organization of extracellular matrix components during differentiation of adipocytes in long-term culture.** *In Vitro Cell Dev Biol Anim* 2000, **36**(1):38-44.
178. Meissburger B, Stachorski L, Roder E, Rudofsky G, Wolfrum C: **Tissue inhibitor of matrix metalloproteinase 1 (TIMP1) controls adipogenesis in obesity in mice and in humans.** *Diabetologia*, **54**(6):1468-1479.
179. Shimba S, Wada T, Hara S, Tezuka M: **EPAS1 promotes adipose differentiation in 3T3-L1 cells.** *J Biol Chem* 2004, **279**(39):40946-40953.
180. Kronenberg HM: **Developmental regulation of the growth plate.** *Nature* 2003, **423**(6937):332-336.
181. Jang CY, Lee JY, Kim J: **RpS3, a DNA repair endonuclease and ribosomal protein, is involved in apoptosis.** *FEBS Lett* 2004, **560**(1-3):81-85.
182. Chaerkady R, Letzen B, Renuse S, Sahasrabudhe NA, Kumar P, All AH, Thakor NV, Delanghe B, Gearhart JD, Pandey A *et al*: **Quantitative temporal proteomic analysis of human embryonic stem cell differentiation into oligodendrocyte progenitor cells.** *Proteomics*, **11**(20):4007-4020.
183. Klees RF, Salaszyk RM, Vandenberg S, Bennett K, Plopper GE: **Laminin-5 activates extracellular matrix production and osteogenic gene focusing in human mesenchymal stem cells.** *Matrix Biol* 2007, **26**(2):106-114.
184. Lai Y, Sun Y, Skinner CM, Son EL, Lu Z, Tuan RS, Jilka RL, Ling J, Chen XD: **Reconstitution of marrow-derived extracellular matrix ex vivo: a robust culture system for expanding large-scale highly functional human mesenchymal stem cells.** *Stem Cells Dev*, **19**(7):1095-1107.
185. Saller MM, Prall WC, Docheva D, Schonitzer V, Popov T, Anz D, Clausen-Schaumann H, Mutschler W, Volkmer E, Schieker M *et al*: **Increased stemness and migration of human mesenchymal stem cells in hypoxia is associated with altered integrin expression.** *Biochem Biophys Res Commun*, **423**(2):379-385.
186. Mathieu PS, Lobo EG: **Cytoskeletal and focal adhesion influences on mesenchymal stem cell shape, mechanical properties, and differentiation**

- down osteogenic, adipogenic, and chondrogenic pathways. *Tissue Eng Part B Rev*, **18**(6):436-444.
187. Ling L, Nurcombe V, Cool SM: **Wnt signaling controls the fate of mesenchymal stem cells**. *Gene* 2009, **433**(1-2):1-7.
188. Katoh M: **WNT signaling pathway and stem cell signaling network**. *Clin Cancer Res* 2007, **13**(14):4042-4045.
189. Shen YL, Luo Q, Guo YX, Zheng GH, Yu J, Xu YH: **Bone marrow mesenchymal stem cell-derived Wnt5a inhibits leukemia cell progression via activation of the non-canonical Wnt signaling pathway**. *Oncol Lett*, **8**(1):85-90.
190. Hardy KM, Garriock RJ, Yatskievych TA, D'Agostino SL, Antin PB, Krieg PA: **Non-canonical Wnt signaling through Wnt5a/b and a novel Wnt11 gene, Wnt11b, regulates cell migration during avian gastrulation**. *Dev Biol* 2008, **320**(2):391-401.
191. Brunt KR, Zhang Y, Mihic A, Li M, Li SH, Xue P, Zhang W, Basmaji S, Tsang K, Weisel RD *et al*: **Role of WNT/beta-catenin signaling in rejuvenating myogenic differentiation of aged mesenchymal stem cells from cardiac patients**. *Am J Pathol*, **181**(6):2067-2078.
192. Sakane H, Yamamoto H, Matsumoto S, Sato A, Kikuchi A: **Localization of glypican-4 in different membrane microdomains is involved in the regulation of Wnt signaling**. *J Cell Sci*, **125**(Pt 2):449-460.
193. Krause U, Harris S, Green A, Ylostalo J, Zeitouni S, Lee N, Gregory CA: **Pharmaceutical modulation of canonical Wnt signaling in multipotent stromal cells for improved osteoinductive therapy**. *Proc Natl Acad Sci U S A*, **107**(9):4147-4152.
194. Zuo L, Gelernter J, Zhang CK, Zhao H, Lu L, Kranzler HR, Malison RT, Li CS, Wang F, Zhang XY *et al*: **Genome-wide association study of alcohol dependence implicates KIAA0040 on chromosome 1q**. *Neuropsychopharmacology*, **37**(2):557-566.
195. Laharrague P, Truel N, Fontanilles AM, Corberand JX, Penicaud L, Casteilla L: **Regulation by cytokines of leptin expression in human bone marrow adipocytes**. *Horm Metab Res* 2000, **32**(10):381-385.
196. Li JJ, Xie D: **Cleavage of focal adhesion kinase (FAK) is essential in adipocyte differentiation**. *Biochem Biophys Res Commun* 2007, **357**(3):648-654.
197. Fitter S, Vandyke K, Gronthos S, Zannettino AC: **Suppression of PDGF-induced PI3 kinase activity by imatinib promotes adipogenesis and adiponectin secretion**. *J Mol Endocrinol*, **48**(3):229-240.
198. Gorbenko O, Panayotou G, Zhyvoloup A, Volkova D, Gout I, Filonenko V: **Identification of novel PTEN-binding partners: PTEN interaction with fatty acid binding protein FABP4**. *Mol Cell Biochem*, **337**(1-2):299-305.
199. Kim YJ, Kim HJ, Chung KY, Choi I, Kim SH: **Transcriptional activation of PIK3R1 by PPARgamma in adipocytes**. *Mol Biol Rep*, **41**(8):5267-5272.
200. Doerrler W, Feingold KR, Grunfeld C: **Cytokines induce catabolic effects in cultured adipocytes by multiple mechanisms**. *Cytokine* 1994, **6**(5):478-484.

201. Zvonic S, Baugh JE, Jr., Arbour-Reily P, Mynatt RL, Stephens JM: **Cross-talk among gp130 cytokines in adipocytes.** *J Biol Chem* 2005, **280**(40):33856-33863.
202. Shi H, Tzamelis I, Bjorbaek C, Flier JS: **Suppressor of cytokine signaling 3 is a physiological regulator of adipocyte insulin signaling.** *J Biol Chem* 2004, **279**(33):34733-34740.
203. McGillicuddy FC, Chiquoine EH, Hinkle CC, Kim RJ, Shah R, Roche HM, Smyth EM, Reilly MP: **Interferon gamma attenuates insulin signaling, lipid storage, and differentiation in human adipocytes via activation of the JAK/STAT pathway.** *J Biol Chem* 2009, **284**(46):31936-31944.
204. Richard AJ, Stephens JM: **Emerging roles of JAK-STAT signaling pathways in adipocytes.** *Trends Endocrinol Metab*, **22**(8):325-332.
205. Xu D, Yin C, Wang S, Xiao Y: **JAK-STAT in lipid metabolism of adipocytes.** *JAKSTAT*, **2**(4):e27203.
206. Guntur AR, Le PT, Farber CR, Rosen CJ: **Bioenergetics during calvarial osteoblast differentiation reflect strain differences in bone mass.** *Endocrinology*, **155**(5):1589-1595.
207. Akizu N, Shembesh NM, Ben-Omran T, Bastaki L, Al-Tawari A, Zaki MS, Koul R, Spencer E, Rosti RO, Scott E *et al*: **Whole-exome sequencing identifies mutated c12orf57 in recessive corpus callosum hypoplasia.** *Am J Hum Genet*, **92**(3):392-400.
208. Zhu L, Zhang Y, Zhang W, Yang S, Chen JQ, Tian D: **Patterns of exon-intron architecture variation of genes in eukaryotic genomes.** *BMC Genomics* 2009, **10**:47.
209. Karolchik D, Barber GP, Casper J, Clawson H, Cline MS, Diekhans M, Dreszer TR, Fujita PA, Guruvadoo L, Haeussler M *et al*: **The UCSC Genome Browser database: 2014 update.** *Nucleic Acids Res*, **42**(Database issue):D764-770.
210. Davidson G, Wu W, Shen J, Bilic J, Fenger U, Stannek P, Glinka A, Niehrs C: **Casein kinase 1 gamma couples Wnt receptor activation to cytoplasmic signal transduction.** *Nature* 2005, **438**(7069):867-872.
211. Zhao W, Blagev D, Pollack JL, Erle DJ: **Toward a systematic understanding of mRNA 3' untranslated regions.** *Proc Am Thorac Soc*, **8**(2):163-166.
212. Lauriat TL, Dracheva S, Kremerskothen J, Duning K, Haroutunian V, Buxbaum JD, Hyde TM, Kleinman JE, McInnes LA: **Characterization of KIAA0513, a novel signaling molecule that interacts with modulators of neuroplasticity, apoptosis, and the cytoskeleton.** *Brain Res* 2006, **1121**(1):1-11.

12 CURRICULUM VITAE

13 DANKSAGUNG

14 EIDESSTATTLICHE ERKLÄRUNG

Hiermit erkläre ich, die vorliegende Dissertation selbstständig und ohne fremde Hilfe angefertigt zu haben. Weiterhin versichere ich, örtlich übernommene Ausführungen anderer Autoren und an Gedankengänge Anderer anlehrende, eigene Formulierungen entsprechend gekennzeichnet und die Quellen zitiert zu haben.

(Ort, Datum)

(Martin Schulze)