

**Bioinformatisch unterstützte Hochdurchsatz-Analyse
molekularbiologischer und zellbiologischer Prozesse der
Regenerationsvorgänge im Herzen des Molches *Notophthalmus
viridescens***

Inauguraldissertation

zur

Erlangung des Grades

Doktor der Naturwissenschaften

-Dr. rer. nat.-

des Fachbereiches Biologie und Chemie (FB 08)

der Justus-Liebig-Universität Giessen

vorgelegt von

Mario Looso

geb. am 01.10.1979, in Bitburg

Giessen, 2010

Die vorliegende Arbeit wurde am Max-Planck-Institut für Herz- und Lungenforschung in Bad Nauheim in der Zeit April 2007 bis Februar 2010 angefertigt.

Vorgelegt wurde die Arbeit an der Justus-Liebig-Universität Giessen, Fachbereich 08, Biologie und Chemie. Betreut wurde die Arbeit von Prof. Dr. Manfred Kröger.

Dekan:

Prof. Dr. Volkmar Wolters

Institut für Tierökologie und Spezielle Zoologie

Justus-Liebig-Universität Giessen

Heinrich-Buff-Ring 58, 35392 Giessen

Erstgutachter:

Prof. Dr. Manfred Kröger

Institut für Mikro- und Molekularbiologie

Justus-Liebig-Universität Giessen

Heinrich-Buff Ring 26-32, 35392 Giessen

Zweitgutachter:

Prof. Dr. Dr. Thomas Braun

Abteilung Entwicklung und Umbau des Herzens

Max-Planck-Institut für Herz- und Lungenforschung

Parkstraße 1, 61231 Bad Nauheim

Diese Dissertation wird mit Genehmigung des Fachbereichs Biologie und Chemie der Justus-Liebig-Universität Giessen in der vorliegenden Form in der Giessener Elektronischen Bibliothek (<http://geb.uni-giessen.de/geb/>) veröffentlicht.

Teile der Dissertation wurden bereits wie folgt veröffentlicht:

2010

BMC Genomics, Jan 4, 11(1): 4, PMID: 20047682

Analysis of newly established EST databases reveals similarities between heart regeneration in newt and fish

Thilo Borchardt, Mario Looso, Patrick Weis, Julia Kruse and Thomas Braun; Max-Planck-Institut für Herz- und Lungenforschung, Parkstrasse 1, D-61231 Bad Nauheim, Deutschland

Molecular and Cellular Proteomics, Feb 5, doi: 10.1074/mcp.M900426-MCP200

Advanced identification of proteins in uncharacterized proteomes by pulsed in vivo SILAC

Mario Looso, Thilo Borchardt, Marcus Krüger and Thomas Braun; Max-Planck-Institut für Herz- und Lungenforschung, Parkstrasse 1, D-61231 Bad Nauheim, Germany

2009

Posterpräsentation Proteomics Forum 2009, Berlin, Abstract Booklet proteo00170, Nr A043

Indirect non saturated SILAC-based mass spectrometry to identify proteins driving organ regeneration of newts

Mario Looso, Marcus Krüger, Thilo Borchardt, Thomas Braun; Max-Planck-Institut für Herz- und Lungenforschung, Parkstrasse 1, D-61231 Bad Nauheim, Germany

I. Inhaltsverzeichnis

II. Erklärung	ix
III. Danksagung	xi
IV. Einleitung	1
A. Regeneration und regenerative Prozesse	1
B. Regeneratives Potential im zoologischen Vergleich	7
C. <i>Notophthalmus viridescens</i> als Modellorganismus für regenerative Fragestellungen	10
1. Einordnung des grünlichen Wassermolches	11
2. Das Regenerative Potential des <i>Notophthalmus viridescens</i>	14
D. Untersuchungen am <i>N. viridescens</i> mit Hilfe moderner Hochdurchsatzverfahren	22
1. Molekularbiologischer und bioinformatischer Zugang des <i>Notophthalmus</i> <i>viridescens</i>	23
2. Übersicht der Techniken zur quantitativen Transkriptomanalyse	24
3. Übersicht der Techniken zur Proteomanalyse	26
E. Motivation und Ziele der Doktorarbeit.....	30
V. Material und Methoden	33
A. Material.....	33
1. Oligonukleotide	33
2. Lösungen	34
3. Vektoren	35
4. Bakterienstämme	36
5. Kommerzielle Sets	36
6. DNA Größenmarker	36
7. Enzyme	36
8. Chemikalien	37

9.	Rechnerhardware	37
10.	Software	38
B.	Methoden	40
1.	Allgemeine Methoden	40
2.	Spezielle Methoden	42
VI.	Ergebnisse	54
A.	Transkriptomanalyse	54
1.	Konzeption einer cDNA Bibliothek zur Herzregeneration	54
2.	Arraydesign von cDNA Microarrays	55
3.	Expressionsprofil der Herzregeneration	57
4.	Vergleich der regenerativen Zeitpunkte	72
5.	Auswahl der Kandidaten zur DNA Sequenzierung	74
6.	Assemblierung, Annotation und Datenspeicherung	74
7.	Einsatz von Next-Generation-Sequencing Techniken	80
B.	Methodenentwicklung Proteomanalyse für indirekt <i>in vivo</i> SILAC markierte Organismen am Beispiel der Schwanzregeneration	81
1.	Markierung des Molches mit stabilen Isotopen essentieller Aminosäuren	82
2.	Regenerationsprozesse initiieren einen verstärkten Einbau von schweren Isotopen	85
3.	Identifizierung von Peptiden in Proteinsequenzdatenbanken fremder Organismen	87
4.	Verifizierung des bekannten Molchproteoms	96
5.	Verwendung von cDNA Nukleotiddatenbanken zur Proteinidentifizierung	96
6.	Identifizierung von unbekanntem Molchproteinen auf Proteomebene	99
C.	Quantifizierung auf Proteinebene mit partiell markierten SILAC Gewebepools in isolierten Kardiomyozyten	102
D.	Analyse des Molchproteoms während der frühen Herzregeneration	109
1.	Identifizierbares Proteom während der frühen Herzregeneration	112

2.	Identifikation von herzspezifischen, regenerationsspezifischen und bislang unbekanntem Proteinen.....	115
VII.	Diskussion	128
A.	Der Notophthalmus viridescens als Modellorganismus	128
B.	Entwicklung von Hochdurchsatzmethoden für den Molch	132
1.	cDNA Microarrays zur Analyse des Transkriptoms.....	132
2.	Proteom Analyse mittels SILAC und Massenspektrometrie	135
C.	Bewertung der Kandidatenlisten aus biologischen Gesichtspunkten	138
1.	Die frühe Herzregeneration auf Transkriptomebene	139
2.	Die frühe Herzregeneration auf Proteinebene.....	145
VIII.	Ausblick.....	150
IX.	Zusammenfassung	153
X.	Anhang.....	xiii
A.	Abkürzungsverzeichnis.....	xiii
B.	Zusätzliche Tabellen und Abbildungen	xv
C.	Abbildungsverzeichnis	xxvii
D.	Tabellenverzeichnis.....	xxix
E.	Lebenslauf.....	xxx
F.	Publikationen	xxxi
XI.	Literaturverzeichnis	xxxiii

II. Erklärung

Ich erkläre: Ich habe die vorgelegte Dissertation selbständig und ohne unerlaubte fremde Hilfe und nur mit den Hilfen angefertigt, die ich in der Dissertation angegeben habe. Alle Textstellen, die wörtlich oder sinngemäß aus veröffentlichten Schriften entnommen sind, und alle Angaben, die auf mündlichen Auskünften beruhen, sind als solche kenntlich gemacht. Bei den von mir durchgeführten und in der Dissertation erwähnten Untersuchungen habe ich die Grundsätze guter wissenschaftlicher Praxis, wie sie in der „Satzung der Justus-Liebig-Universität Gießen zur Sicherung guter wissenschaftlicher Praxis“ niedergelegt sind, eingehalten.

Bad Nauheim, den

NEMINEM TIME, NEMINEM LAEDE

III. Danksagung

Lieber Leser,

die vorliegende Arbeit stellt die Zusammenfassung meiner wissenschaftlichen Tätigkeiten der letzten drei Jahre am Max-Planck-Institut für Herz- und Lungenforschung in Bad Nauheim dar. Diese in vielfacher Weise sehr lehrreiche und schöne Zeit möchte ich nicht missen. Daher will ich es auch nicht versäumen, einigen Personen für ihre Hilfe und Unterstützung in diesen drei Jahren zu danken.

So gilt mein Dank Herrn Prof. Dr. Dr. Braun, der mir die Möglichkeit gegeben hat, meine Arbeiten im Umfeld eines modernen Forschungsinstitutes durchzuführen und selbige stets unterstützt und gefördert hat. Außerdem gilt mein Dank Herrn Prof. Dr. Kröger für die engagierte Betreuung meiner Arbeit. Weiterhin möchte ich mich bei meinem Forschungsgruppenleiter Dr. Thilo Borchardt für seine jederzeit hilfreiche Beratung und sein allzeit offenes Ohr danken. Insbesondere danke ich für seine Geduld, die er für seinen Bioinformatiker im „Wet Lab“ aufgebracht hat. Daneben danke ich Dr. Marcus Krüger für die Begeisterung für Massenspektrometrie, die er bei mir durch seine überzeugende Art auszulösen vermochte. Weiterhin danke ich meinen Con-Doktoranden Anne, Tanja, Marc, Christian und Jason für Ihre praktische Hilfe bei meiner Arbeit. Allen Kollegen am Institut danke ich für die angenehme Arbeitsatmosphäre.

Um ein, manchmal „nervenaufreibendes“, Forschungsvorhaben erfolgreich abschließen zu können, bedarf es auch des nötigen Ausgleichs und einer entsprechenden Unterstützung außerhalb des Labors. In diesem Zusammenhang möchte ich mich bei meiner Familie insgesamt und ganz besonders bei meinen Eltern Elmar und Marlene bedanken, deren Unterstützung ich mir seit dem Beginn meines Studiums sicher sein durfte. Außerdem danke ich meiner Frau Steffi für ihre Rücksicht und Hilfe, insbesondere in den letzten Monaten vor der Abgabe. Auch gilt ein besonderer Dank meinem Bruder Marcell, der mit seinem Bauvorhaben den nötigen körperlichen Ausgleich am Wochenende für mich schaffen konnte. Abschließend geht mein Dank noch an meine Bundesbrüder bei der Hasso-Rhenania Gießen, deren Aktivitäten und freundschaftliche Runden für die nötige Ablenkung am Abend sorgen konnten.

IV. Einleitung

Im Mittelpunkt dieser Arbeit steht der Grünliche Wassermolch *Notophthalmus viridescens* (*N. viridescens*). Er ist ein Vertreter der Gattung Ostamerikanischer Wassermolch, die im Allgemeinen für ihre besonderen Fähigkeiten zur Regeneration bekannt ist. So lernt man bereits in der Schule, dass der Schwanz des Molches bei Verlust wieder nachwächst. Der Molch genießt besondere Aufmerksamkeit als Modellorganismus für die Regeneration von Organen und Gliedmaßen, weil diese durchaus nicht nur auf den Schwanz beschränkt ist. Fernziel der Untersuchungen zur Regeneration am Molch ist die Übertragung dieser Vorgänge auf Prozesse im Menschen. Voraussetzung dafür ist eine vollständige molekularbiologische Charakterisierung der Regenerationsprozesse im Molch.

Die ungewöhnlichen Fähigkeiten des Molches sind in einem ungewöhnlich großen Genom kodiert (etwa die zehnfache Größe des menschlichen Genomes). Dieses führt zu erheblichen Problemen bei der Interpretation molekularer Daten. Um in die Gesamtproblematik einzuführen wird im Folgenden zunächst die Regeneration als solche definiert und bekannte Formen der Regeneration erläutert. Danach folgt die Einordnung der regenerativen Fähigkeiten verschiedener Organismen durch einen Spezies übergreifenden Vergleich. Nach dieser Einführung zur Regeneration schließt sich zunächst eine allgemeine Beschreibung des Grünlichen Wassermolches an. Diese wird durch die Darstellung der außerordentlichen regenerativen Fähigkeiten und der daraus resultierenden Bedeutung des Molches als Modellorganismus für regenerative Fragestellungen ergänzt. Abschließend werden die Probleme bei der molekularbiologischen Untersuchung des Grünlichen Wassermolches näher erläutert.

A. Regeneration und regenerative Prozesse

Unter Regeneration versteht man im Allgemeinen die Fähigkeit eines Organismus, verletzte oder abgetrennte Körperteile zu reparieren oder zu ersetzen. Diese Fähigkeit ist grundsätzlich bei allen Organismen vorhanden, jedoch variiert der Grad der Ausprägung sehr stark. Das höchstmögliche Regenerative Potential besitzen Pflanzen, die potente Zellen besitzen und sich aus diesen vollständig regenerieren können. Demgegenüber steht bei vielen Tieren allein die Wundheilung als die minimale Form der Regeneration. Stärkere Formen der Regeneration bis hin zur Fähigkeit der Neubildung von ganzen Gliedmaßen ist nur wenigen Arten vorbehalten.

Einleitung

Als Begründer der Regenerationsforschung können Ferchault de Réaumur (1683-1757) und Trembley (1710-1784) angesehen werden. Ferchault de Réaumur beschrieb 1712 erstmals eine Gliedmaßenregeneration bei Krebsen [1], Trembley veröffentlichte 1744 seine Beobachtungen zur Regeneration des Kopfes beim Süßwasserpolypt Hydra [2]. Er beschrieb, dass die Hydra abgetrennte Teile, insbesondere den Kopf, nicht durch Proliferation, sondern durch Umgestaltung bereits vorhandener Zellmasse ersetzt. Im kurzen Zeitabstand dazu wurden 1766 die Planarien (Plattwürmer) und 1768 die Urodelen (Schwanzlurche) als Organismen mit regenerativen Fähigkeiten beschrieben. Die Regeneration von ganzen Gliedmaßen bei Wirbeltieren wurde erstmals 1768 von Spallanzani am Beispiel des Salamanders beschrieben [3]. Es vergingen mehr als zwei Jahrhunderte bis 1975 erstmals das Regenerationspotential der menschlichen Leber gezeigt werden konnte [4]. Planarien und Urodelen stellen bis heute die hauptsächlich untersuchten Organismen bei Fragestellungen zur Regeneration dar. Lediglich der Zebrafisch ist als Modellorganismus für Herz und Flossenregeneration hinzugekommen [5,6].

Die Gruppe der benannten Modellorganismen für Regeneration unter Ausnahme des Zebrafisches zeichnet sich dadurch aus, dass wenige Sequenzinformationen zu Genom, Transkriptom und Proteom vorliegen. Dieser Umstand ist durch die zunächst genetisch, und später molekularbiologisch ausgerichtete biologische Grundlagenforschung bedingt, die Organismen mit kurzer Generationszeit und einfachen Haltungsvoraussetzungen bevorzugte. So zeigt sich heute, dass gemessen an generellen molekularbiologischen Erkenntnissen, das Wissen über regenerative Prozesse verschwindend gering ist. Dieser Umstand wird durch einen Mangel an gut erforschten Modellorganismen und das Fehlen etablierter Methoden zur molekularbiologischen Manipulation verstärkt.

Morgan (1866-1945) und Driesch (1867-1941) teilten die Regeneration bei Metazoen in zwei Gruppen (Abbildung 1):

- Regeneration durch Neuordnung bestehender Zellmasse
- Regeneration durch Zellproliferation

Die erste Form der Regeneration wird als Morphallaxis [7] oder Reparation bezeichnet. Dabei bezeichnet Morphallaxis den Ersatz von fehlenden Körperteilen durch die Umorganisation von bestehenden Reststrukturen. Dazu gehören Umformung, Umdifferenzierung und Umordnung der überlebenden Zellen. Das bekannteste Beispiel für morphallaktische Regeneration ist die bereits erwähnte Regeneration des Kopfes der Hydra. Der Prozess wird durch eine Wundheilung und Wundverschluss ausgehend vom Endoderm hin zum Ektoderm eingeleitet und schließt mit der Ausdifferenzierung von Tentakeln und Hypostoma. In diesem

Zusammenhang konnten zwei Peptide identifiziert werden, der sogenannten Kopfaktivator (HA, 11 Aminosäuren (AS)) [8] und ein weiteres Peptid der Länge 23 AS namens HEADY [9]. HEADY zeigt während der Kopfregeneration eine vorübergehende Verstärkung der Expression (im Folgenden als Aufregulierung/aufreguliert bezeichnet) in der Amputationszone.

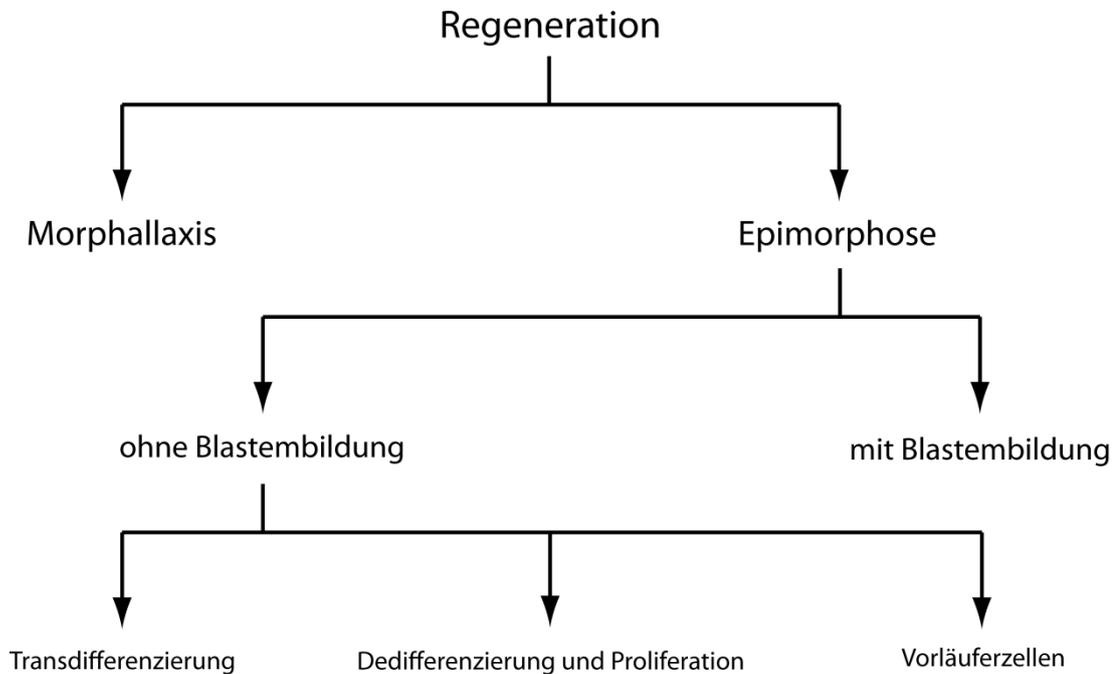


Abbildung 1: Schematische Einteilung der Regeneration

Dargestellt ist die Gliederung der verschiedenen Formen der Regeneration nach den ersten Untersuchungen zur Regeneration von Morgan (1866-1945) und Driesch (1867-1941).

Die zweite erwähnte Form der Regeneration nutzt ausschließlich Zellvermehrung und wird als Epimorphose bezeichnet. Die Epimorphose unterteilt Morgan wieder in zwei Teilbereiche (vergleiche Abbildung 1):

- Regeneration ohne Blastembildung
- Regeneration mit Blastembildung

Unter dem Begriff Blastem wird dabei ein schützendes Konstrukt aus Epithelkappe und darunter akkumulierter dedifferenzierter Zellen an der regenerierenden Wundstelle verstanden. Der Ablauf der Blastembildung sowie der detaillierte Aufbau werden noch einmal separat behandelt (siehe Kapitel IV.C.2.a). Die Regeneration ohne Blastembildung unterteilt sich in drei

Einleitung

weitere Bereiche. Zunächst ist die Transdifferenzierung zu nennen. Hierbei dedifferenzieren vollständig ausdifferenzierte Zellen und proliferieren anschließend. Die Besonderheit der Transdifferenzierung liegt im abschließenden Umdifferenzierungsschritt. Eine dedifferenzierte Zelle gelangt abschließend zu einem unterschiedlichen Zelltyp. Die Transdifferenzierung ist selten, ein Beispiel wäre die Augenlinsenregeneration des Grünlichen Wassermolches (*Notophthalmus viridescens*). Hier entsteht nach Entfernung der Linse aus pigmentierten Zellen der dorsalen Iris neues Linsengewebe und letztendlich eine komplette, voll funktionstüchtige Linse [10,11,12]. Auch bei der Schwanzregeneration des *Ambystoma mexicanum* spielen Mechanismen der Transdifferenzierung eine Rolle. So konnte gezeigt werden, dass Zellen des zentralen Nervensystems aus selbigem migrieren und an der Bildung von Knorpel beteiligt sind [13]. Die selbe Arbeitsgruppe widersprach dieser Transdifferenzierung jedoch kürzlich wieder [14].

Die zweite Form der nicht-blastemgestützten Regeneration ist die Dedifferenzierung mit anschließender Proliferation und Redifferenzierung aus organspezifischen differenzierten Zellen. Das bekannteste Beispiel für diese Art der Regeneration ist die Leberregeneration bei Säugern. Hierbei kann die Regeneration durch eine traumatische (partielle Hepatektomie), chemische oder toxische Schädigung initiiert werden. Einige Faktoren sind in diesem Bereich mittlerweile bekannt, beispielsweise der sekretierte „*hepatocyte growth factor*“ (HGF), ein Molekül, welches in der frühen Leberregeneration initiiert wird. In Kombination mit den weiteren Faktoren „*transforming growth factor alpha*“ (TGF-alpha) und „*acidic fibroblast growth factor*“ (aFGF) werden Hepatozyten *in vitro* zur Replikation angeregt [15]. Die frühesten Signale, die zur Dedifferenzierung führen, sind jedoch unbekannt. Ob die Leberregeneration allein auf dieser Art der Regeneration beruht wird kontrovers diskutiert. Einige Untersuchungen gehen davon aus, dass spezielle, ruhende Zellpopulationen in der Leber ebenfalls eine Rolle bei der Regeneration spielen.

Diese ruhenden Zellen leiten zur dritten Form der nicht blastemgestützten Regeneration über. Die dritte Form läuft auf Basis der Zellen ab, die im Gewebe vorliegen, einen stammzellähnlichen Charakter zeigen und häufig auch als Vorläuferzellen, im Skelettmuskel auch als Satellitenzellen bezeichnet werden. Diese Zellen kommen im Knochenmark, im Gehirn, in der Leber, im Darm und im Herz vor [16,17,18,19,20]. Allgemein bekannt und gut erforscht sind außerdem das blutbildende System, die Erneuerung des Darmepithel und der Haut. Die Regeneration des Muskelgewebes ist erst in den letzten Jahren näher untersucht worden. Da

das Herz des Molches auch zu den Muskeln gehört, soll die Rolle der Satellitenzellen im Muskel im Folgenden näher erläutert werden.

Der Vorgang der Regeneration des adulten Muskels wird als Myoneogenese bezeichnet. Die Satellitenzellen liegen in einer speziellen Umgebung, der sogenannten „*stem cell niche*“, die aus „Extrazellulärer Matrix“ (EZM) besteht. Umschlossen ist diese Matrix von einer Membran, die aus mehreren Zelltypen bestehen kann [21]. Die Zellen liegen an Muskelfasern angelagert im Gewebe vor. Diese ruhenden Vorläuferzellen, die nicht ausdifferenziert sind, werden im Falle einer Schädigung aktiviert und zur Proliferation angeregt. Dabei scheinen die Zellen die Fähigkeit zur unbegrenzten Vervielfältigung zu haben. Dieser Prozess muss dementsprechend genau reguliert werden, damit ein unkontrolliertes Wachstum verhindert wird. Einen wesentlichen Teil dieser Regulation wird wahrscheinlich von der *stem cell niche* geleistet, die Signale selektiv durch die EZM leitet [18]. Die regulatorischen Einflüsse der EZM konnten dadurch gezeigt werden, dass Satellitenzellen nach Isolation *in vitro* nach kurzer Zeit eine Abnahme in ihrer Fähigkeit zur Teilung zeigen [22]. Die Satellitenzellen können das geschädigte Gewebe durch Proliferation, Migration und Ausdifferenzierung reparieren. Dabei bleibt die Anzahl nicht differenzierter, ruhender Vorläuferzellen im Gewebe bestehen, um bei einer weiteren Schädigung wieder die Initiierung der Regeneration über Stammzellteilung starten zu können. Dieser Vorgang wird als *Selbsterneuerung* bezeichnet. Während die Signalwege zur Selbsterneuerung noch unbekannt sind, erfolgt bei Satellitenzellen der Skelettmuskulatur, den myogenen Vorläuferzellen, die Differenzierung über die Faktoren Myf5 und MyoD [23]. Auch bei der Beinregeneration des Salamanders konnte die Existenz und die Beteiligung von Satellitenzellen an regenerativen Prozessen gezeigt werden [24]. Der Umfang dieser Beteiligung ist jedoch noch völlig unklar. Auch hat sich gezeigt, dass in diesem Fall die Satellitenzellen in einer etwas veränderten *stem cell niche* mit doppelter Membran liegen. Satellitenzellen des Salamanders exprimieren, wie die der Maus, Pax7 und sind am Rande von Muskelfasern lokalisiert.

Die epimorphe Regeneration mit Blastembildung wird bei komplexen Gewebedefekten beobachtet. Dazu gehört der Verlust von ganzen Gliedmaßen wie Bein oder Schwanz. In diesem Falle bildet sich an der Wundstelle ein Blastem aus. Dieses besteht aus undifferenzierten Zellen mesenchymalen Ursprungs mit schützender Wundepidermis und erfüllt essentielle Funktionen zur Einleitung der Regeneration [25]. In diesem Blastembereich wird durch einen Prozess der Dedifferenzierung von differenzierten Zellen nahe der Wundstelle, anschließender Proliferation und abschließender Redifferenzierung das amputierte Gewebe narbenfrei ersetzt [25]. Dabei ist das Vorhandensein des Blastems an sich, sowie die Präsenz intakter Nerven [25] und letztendlich

Einleitung

auch Mechanismen der positionellen Identität [26] unabdingbar. Unter positioneller Identität versteht man die Fähigkeit des geschädigten Gewebes, nur Strukturen zu regenerieren, die distal einer erzeugten Amputationsfläche liegen. Diese Fähigkeit ist bei der Beinregeneration vom Schultergelenk bis zu den Zehen vorhanden. Interessanterweise ist die Information der Position im entstehenden Blastem gespeichert und unabhängig vom darunterliegenden Stumpf. Dies konnte gezeigt werden, indem man das Blastem von einem Beinstumpf auf ein anders Gewebe wie beispielsweise den Schwanz transplantierte. Hier entwickelte sich trotz Transplantation ein Teil des Beines [27]. Es konnten einige Proteine identifiziert werden, die bei der positionellen Identität über proximal-distale Gradienten ein „Gedächtnis“ für die aktuelle Position ausbilden. So befähigt der Oberflächenrezeptor Prod 1 im Zusammenspiel mit seinem von Nervenzellen sekretierten Liganden nAG Zellen zur positionellen Orientierung [28]. Dieser Zusammenhang wird detailliert in Abschnitt IV.C.2.a) beschrieben.

Die blastemgestützte Regeneration befähigt Urodele (Schwanzlurche) zur Neubildung von ganzen Körperteilen auch im adulten Entwicklungsstadium. Die zunächst ablaufenden Prozesse, die zur Dedifferenzierung und Blastembildung führen, sind jedoch weitgehend unbekannt. Der zweite Abschnitt der Regeneration, der das Wachstum aus dem Blastembereich, die Musterbildung und Redifferenzierung umfasst, scheint wie bei der Embryonalentwicklung abzulaufen [29].

Agata stellte die Einteilung nach Epimorphose und Morphallaxis in Frage [30]. Er untersuchte regenerative Prozesse bei Amphibien, Süßwasserpolyphen und Planarien. Er kommt zu dem Schluss, dass eine Einteilung von Regeneration vielmehr nach Art und Weise der Vermittlung der positionellen Identität sinnvoll ist und schlägt eine Einteilung nach Distalisation und Interkolation vor. Agata verglich die Regeneration von Extremitäten bei Urodelen mit Körperregeneration bei Planarien und Hydra. Er demonstrierte, dass jeder dieser Regenerationsvorgänge nach Amputation eine schnell einsetzende Distalisation zeigt. Dies bedeutet, dass sich kurz nach Amputation zunächst die distal gelegenen Bereiche des zu regenerierenden Körperteils ausbilden. Der fehlende proximale Bereich wird dann durch Wachstum aus dem ursprünglichen Stumpf heraus gebildet. Diese zweite Wachstumsphase wird als Interkolation bezeichnet.

Zusammenfassend bleibt festzuhalten, dass der Versuch einer Klassifizierung regenerativer Prozesse verschiedener Gewebe, wie Bein, Linsen oder Herzmuskel zumeist schwierig ist und eine dedizierte Einteilung der Regeneration nach blastemgestützt und nicht-blastemgestützt nicht immer gelingt. Häufig spielen mehrere Arten der Regeneration eine Rolle. Der Anteil der

verschiedenen Arten der Regeneration ist dabei wissenschaftlich umstritten und Gegenstand aktueller Forschung.

Die Nutzung allgemeingültiger biologischer Regelmechanismen wie Zelldifferenzierung und Zellteilung durch die Regeneration lässt die starke Verknüpfung mit onkologischen Prozessen vermuten. Erstaunlicherweise scheint die Regeneration jedoch einen diametralen Gegensatz zu onkologischen Prozessen zu bilden. Phänomenologisch fällt auf, dass Urodele, die sich durch hohe regenerative Fähigkeiten auszeichnen, kaum spontane Tumorbildung zeigen. Sogar bei Injektion von kanzerogenen Stoffen in normales oder regenerierendes Gewebe weisen diese keine Beeinträchtigung auf [31,32].

B. Regeneratives Potential im zoologischen Vergleich

Ein Regeneratives Potential ist im Tierreich nahezu bei allen Stämmen vorzufinden. Abbildung 2 zeigt eine Übersicht der Stämme des Tierreiches, für die bereits regenerative Vorgänge beobachtet worden sind. Nur für wenige Stämme ist das Fehlen von regenerativen Fähigkeiten beschrieben. Bei einigen Stämmen liegen noch keine Kenntnisse vor. Genauer untersucht worden sind Vertreter der Nesseltiere (Cnidaria) und Vertreter der Zweiseitentiere (Bilateria), die daher heute auch die Modellorganismen für regenerative Fragestellungen stellen. Bemerkenswert ist die Tatsache, dass in einem Organismus auch verschiedene Arten der Regeneration in verschiedenen Geweben beobachtet werden konnten. So tritt beispielsweise im Molch die Regeneration mit Blastembildung (Bein) und ohne Blastembildung (Linse) auf.

Bei den Nicht-Vertebraten zeigen Vertreter der Nesseltiere und der Plattwürmer das höchste bekannte Regenerative Potential. Wie bereits erwähnt, ist die Hydra als Vertreter der Nesseltiere bezüglich der Regeneration der am besten untersuchte Organismus, für den die Regeneration auch erstmalig beschrieben wurde [33]. Die Hydra zeigt in ihrem Körperbau eine einfache Struktur, die in drei anatomisch unterscheidbare Bereiche unterteilt ist. Dies ist erstens eine Fußstruktur, die dem Tier zum Halt auf dem Untergrund dient. Darüber liegt zweitens eine Körpersäule, die ein Verdauungsorgan beinhaltet. Nach oben hin abgeschlossen liegt drittens eine von Tentakeln umrahmte Maulstruktur, die als Kopfreion bezeichnet wird.

Einleitung

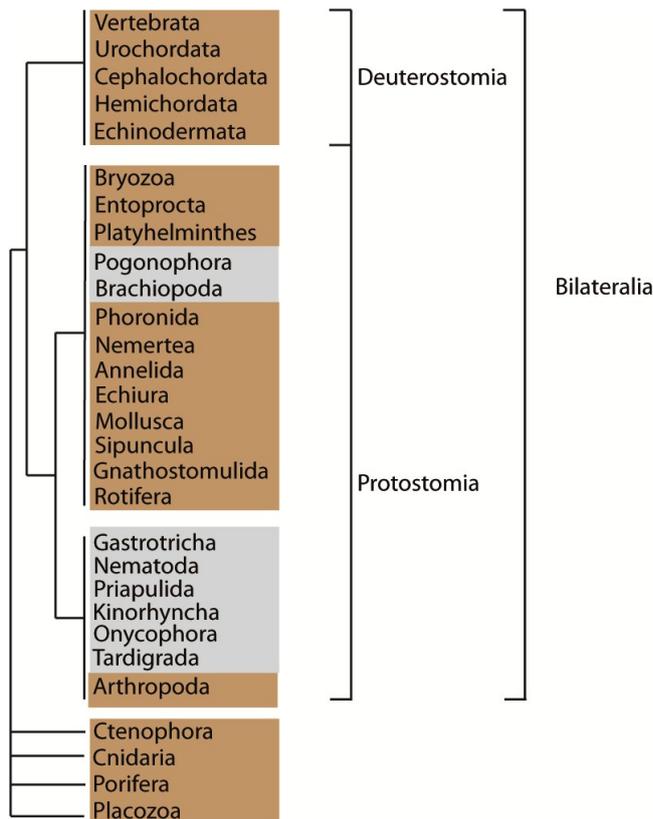


Abbildung 2 nebenstehend: **Regeneratives Potential in mehrzelligen Organismen**

Der dargestellte Baum zeigt die phylogenetische Verteilung der Fähigkeit zur Regeneration in mehrzelligen Organismen. Taxa, die Tiere beherbergen, für die regenerative Prozesse beschrieben sind, wurden braun hinterlegt. Die übrigen Taxa zeigen entweder keine regenerativen Fähigkeiten, oder wurden bislang dahingehend nicht untersucht. Die Abbildung wurde in Anlehnung an [34] erstellt.

Plattwürmer werden seit etwa 100 Jahren als Modellorganismus für Regeneration untersucht. Im Gegensatz zur Hydra bilden Plattwürmer an Amputationsstellen eine Blastemstruktur aus. Diese wird durch Proliferation aus im Gewebe ruhenden somatischen Stammzellen erzeugt, die auch als Neoblasten bezeichnet werden. Die Blastembildung scheint eine konservierte Struktur zur Regeneration innerhalb der Zweiseitentiere (*Bilateria*) zu sein. Darüber hinaus zeigen im kleineren Umfang auch weitere Nicht-Vertebraten Regeneratives Potential. Für Spinnen, Schaben und Milben [35,36] wurde die Regeneration von Körperteilen bereits beschrieben. Auch bei Ringelwürmern konnte die Regeneration nach Dekaptation beobachtet werden. Dargestellt ist der grundlegende Mechanismus in Abbildung 3. Im Stamm der Wirbeltiere wurden verschiedene Tiere beschrieben, die ein hohes Regeneratives Potential haben. Hier sind zunächst zahlreiche Amphibien, aber auch einige Fische zu nennen. Mit deutlichem Abstand bezüglich der regenerativen Fähigkeiten folgen die Säugetiere. Es kann nur vermutet werden, warum sich innerhalb der Wirbeltiere eine so hohe Variabilität bezüglich der regenerativen Fähigkeiten ausgebildet hat. Inwieweit die Genomgröße in diesem Zusammenhang eine Rolle spielt, ist nicht genau zu sagen. Das Genom des Molches wird etwa auf die 10fache Größe des Genomes des Menschen geschätzt [37]. Insgesamt auffällig ist die

bisher beschränkte molekularbiologische Zugänglichkeit nahezu aller Organismen mit hoher regenerativer Kapazität.

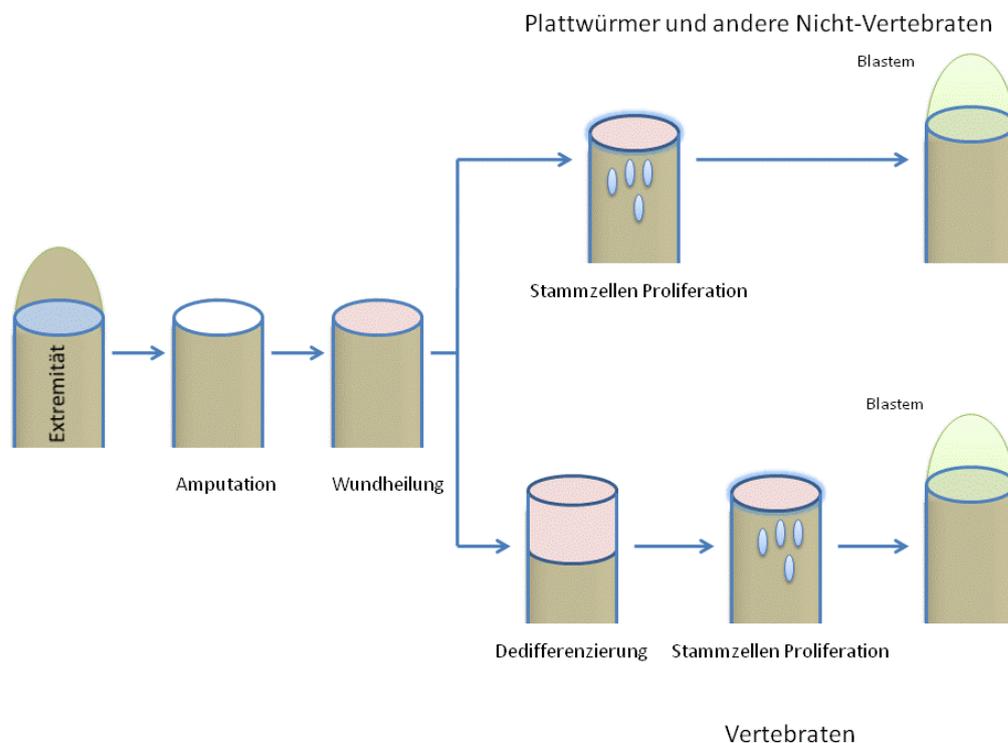


Abbildung 3: Blastemgenerierung nach Beinamputation (schematisch)

Nach Amputation einer Extremität setzt die Wundheilung an der Amputationsfläche ein. Bei Plattwürmern und Nicht-Vertebraten setzt die Regeneration durch Proliferation von zuvor ruhenden Stammzellen ein, die nahe der Amputationsfläche liegen. Bei Vertebraten, wie beispielsweise dem Molch, bildet sich an der Amputationsfläche ein Blastem aus, das die Dedifferenzierung und Proliferation des anliegenden Gewebes anregt. Im Gewebe ruhende Stammzellen spielen eine untergeordnete Rolle. Die Abbildung wurde in Anlehnung an [34] erstellt.

So haben innerhalb der Wirbeltiere urodele Amphibien scheinbar das breiteste Potential zur Geweberegeneration. Molche können Bein-, Schwanz-, Herz-, Linsen-, Zentrales Nerven-, Gehirn-, Kiefer- und Regenbogenhautgewebe regenerieren. Diese Prozesse können auch im adulten Zustand noch stattfinden. *Ambystomatidae* (Salamander) können Bein-, Schwanz- Herz-, Gehirn- und Zentrales Nervengewebe bedingt regenerieren. Beim häufig beschriebenen *Ambystoma mexicanum* zeigt sich ein umgekehrt proportionales Verhältnis zwischen Alter der Tiere und Regenerationsfähigkeit. Während die Larve des *Ambystoma* ein Bein innerhalb ~44 Tagen regeneriert, benötigt der ausgewachsene (Landphase) Salamander bis zu 370 Tage zur vollständigen Regeneration [38]. Auch Frösche, die ebenfalls Bein-, Schwanz-, Regenbogenhaut-

Einleitung

und Linsengewebe regenerieren können, büßen diese Fähigkeit nach der Metamorphose weitestgehend ein [39]. Untersuchungen haben gezeigt, dass die stattfindenden Prozesse nicht eindeutig einem der klassifizierten Regenerationsabläufe zuzuordnen sind. Es wurde beispielsweise bei der Beinregeneration gezeigt, dass neben einer blastemvermittelten Regeneration über die Schritte der Dedifferenzierung, Proliferation und Redifferenzierung auch zuvor ruhende, undifferenzierte Zellen eine Rolle spielen [24].

Bei den Fischen ist primär der Zebrafisch zu nennen. Dieser kann Flossen-, Herz- und zentrales Nervengewebe regenerieren. Bei Säugetieren und insbesondere beim Menschen sind die Leberregeneration und die Regeneration von Fingerspitzen bekannt. Jedoch ist allgemein bekannt, dass auch die Regenerationsfähigkeit des Menschen umgekehrt proportional zum Alter ausgebildet ist.

C. *Notophthalmus viridescens* als Modellorganismus für regenerative Fragestellungen

Molche sind urodele Amphibien, die sich vorwiegend im Wasser aufhalten. Sie gehören zur Familie der Salamander. In der experimentellen Biologie werden der iberische Molch (*Pleurodeles waltl*) und der japanische Molch (*Cynops pyrrhogaster*) eingesetzt. Weitere untersuchte Organismen aus der Klasse der Amphibien sind der Axolotl (*Ambystoma mexicanum*), ein neotener larvaler Salamander, der Tigersalamander (*Ambystoma tigrinum*) und der Krallenfrosch (*Xenopus laevis*).

Der Grünliche Wassermolch *Notophthalmus viridescens* (lat. *viridescens* = grün werden, Abbildung 4) dient dank seiner besonders stark ausgeprägten, regenerativen Fähigkeiten als Modellorganismus für Beinregeneration, Schwanzregeneration, Linsenregeneration des Auges, Rückenmarkregeneration und Herzregeneration. Nachfolgend wird eine phylogenetische Einteilung des *Notophthalmus viridescens* (*N. viridescens*) vorgenommen und danach dessen Regeneratives Potential erläutert. Daran schließt sich die Beschreibung der bereits verhältnismäßig gut erforschten Extremitätenregeneration an, um einen Überblick über bereits bekannte Kandidatengene, Mechanismen und Eigenarten während der Regeneration zu geben. Abschließend wird der Stand der Forschung zur Herzregeneration des Molches dargestellt. Dazu wird die Anatomie des Molchherzens beschrieben und der Ablauf der Herzregeneration deskriptiv dargestellt.



Abbildung 4: Der grünliche Wassermolch *Notophthalmus viridescens*

Notophthalmus viridescens wird bis zu 12 cm lang und bis zu 20 Jahren alt. Das Bild zeigt einen adulten Molch. Der Bauchbereich ist hell orange bis braun mit dunklen Punkten besetzt, der Rücken ist grünlich mit roten Punkten besetzt, die schwarz umrandet sind.

1. Einordnung des grünlichen Wassermolches

Im Folgenden wird der grünliche Wassermolch auf Basis der taxonomischen Systematik eingeordnet und hinsichtlich seines räumlichen Vorkommens als auch anhand seiner markanten Merkmale und Art der Fortpflanzung charakterisiert.

a) Systematik

Um den Vergleich der im Rahmen dieser Arbeit erstellten Sequenzdaten zu anderen Amphibien möglichst genau darstellen zu können, wurde in Tabelle 1 die vollständige Systematik des Molches zusammengestellt.

Einleitung

Tabelle 1: Taxonomie des grünlichen Wassermolches

Reich	Vielzellige Tiere (Metazoa)
Abteilung	Gewebetiere (Eumetazoa)
Unterabteilung	Zweiseitentiere (Bilateria)
Stamm	Chordatiere (Chordata)
Unterstamm	Wirbeltiere (Vertebrata)
Reihe	Landwirbeltiere (Tetrapoda)
Klasse	Amphibien / Lurche (Amphibia)
Unterklasse	Lissamphibia
Ordnung	Schwanzlurche (Caudata)
Familie	Echte Salamander (Salamandridae)
Gattung	Ostamerikanische Wassermolche (Notophthalmus)
Art	Grünlicher Wassermolch (Notophthalmus viridescens)
Unterart	Viridescens

Bekannte Unterarten:

<i>Notophthalmus viridescens viridescens</i>	(Vorkommen in Kanada, Osten und Mitte der USA)
<i>Notophthalmus viridescens louisianensis</i>	(Mitte und tiefer Süden der USA)
<i>Notophthalmus viridescens dorsalis</i>	(Küstenebene von Carolina)
<i>Notophthalmus viridescens piaropicola</i>	(Florida)

b) Vorkommen des *Notophthalmus viridescens*

Der grünliche Wassermolch hat sein Hauptverbreitungsgebiet vom Golf von Mexiko über die ganze Mitte der USA bis hin nach Südkanada. Sein bevorzugter Lebensraum sind sowohl Laub- als auch Nadelwälder. Der Bestand gilt als nicht gefährdet. Dennoch muss aufgrund des Klimawandels und schädigender Umwelteinflüsse ein Rückgang der natürlichen Populationsdichte festgestellt werden [40].

c) *Merkmale und Lebensweise*

Der *N. viridescens* wird zwischen sieben und zwölf Zentimeter lang. Seine Lebenserwartung liegt bei 20 Jahren. Bis zur Geschlechtsreife mit zwei bis drei Jahren lebt er ausschließlich auf dem Land und ist in dieser Zeit meist nachtaktiv. Die Leibfärbung ist während dieser Zeit rot-orange mit roten, schwarz umrahmten Punkten. Mit Erreichen der Geschlechtsreife wechselt *N. viridescens* den Hauptlebensraum vom Land zum Wasser. Dabei ändert er während der Ontogenese auch seine Färbung. Nun trägt er eine olivgrüne Rückenfärbung mit schwarzen Punkten. Der Bauch ist stark gelblich gefärbt. An den Flanken zieht sich eine Reihe kleiner roter Punkte entlang. Die Pupillen sind waagrecht [37]. Ein zweijähriger Molch ist in Abbildung 4 dargestellt.

d) *Fortpflanzung*

Die Paarungszeit liegt im späten Winter und im Frühling. Bevorzugt werden stille Gewässer, die sich meist durch starke Vegetation auszeichnen. Als Balzritual werden Schwanzbewegungen des Männchens eingesetzt [37]. Zur Paarungszeit hat das Männchen einen verbreiterten Schwanzsaum, eine stark gewölbte Kloake und schwarze Brunftschwien an den Vorderbeinen.

Das Männchen umgreift das Weibchen und fächelt ihm Duftstoffe aus der Kloakenregion zu. Ein kugelförmiges Samenpaket wird schließlich auf dem Boden abgesetzt und vom Weibchen mit der Kloake aufgenommen. Das Weibchen legt 200 bis 400 Eier, die es an Blättern von Wasserpflanzen befestigt. Die Larven sind olivgrün und zwischen 5 – 10 mm groß. Sie tragen äußere Kiemenbüschel. Nach fünf Monaten verlassen 4 cm große Jungtiere das Wasser und verbleiben bis zur Geschlechtsreife auf dem Land.

Die Züchtung der Wassermolche in Aquarien stellt sich als recht schwierig dar. Molche benötigen während der Entwicklungsstadien sehr unterschiedliche Umgebungsbedingungen bezüglich der Fütterung, des Vorkommens an Wasser- und Landfläche sowie Temperatur und Feuchtegrad. Züchtungsversuche am Max-Planck-Institut für Herz- und Lungenforschung in Bad Nauheim (MPI-BN) konnten trotz hohem Aufwand zunächst nur geringe Erfolge erzielen. Mittlerweile wurden die optimalen Bedingungen für die Entwicklung von befruchteten Eiern vom betreuenden Personal erarbeitet und die Schlüpftrate der Eier konnte auf 80% gesteigert werden. Die frühe Aufzucht, insbesondere die Art der Fütterung in den ersten Wochen nach dem

Einleitung

Schlüpfen, stellt aber immer noch eine Schwierigkeit dar, die zum Absterben einer großen Zahl an Jungmolchen führt.

2. Das Regenerative Potential des *Notophthalmus viridescens*

Der grünliche Wassermolch zeichnet sich durch ein sehr breites Regeneratives Potential aus. Beschrieben wurde unter anderen die vollständige Regeneration der Retina und Augenlinse [41], des Rückenmarks, für Teile des Herzen [42,43] und von Extremitäten wie Bein und Schwanz [28]. Darüber hinaus zeigt der Molch die Fähigkeit die Haut inklusive der unterliegenden Silberhaut zu regenerieren (eigene Beobachtung). Die Regenerationsfähigkeit der anderen Organe wie Leber, Lunge oder Milz wurden bislang nicht näher untersucht. Ein ähnlich hohes Regeneratives Potential ist jedoch zu vermuten.

a) *Regeneration von Extremitäten*

Die Extremitäten-Regeneration gehört zu den bestuntersuchten Arten der Regeneration und soll daher im Folgenden detailliert dargestellt werden. Generell gilt, dass unabhängig von der Stelle der Amputation exakt der Teil der Extremität nachgebildet wird, der entfernt wurde. Mit der Länge des zu regenerierenden Bereichs steigt die zur vollständigen Regeneration benötigte Zeit, da größeres Längenwachstum mit deutlich mehr Zellteilungen im Blastembereich einhergeht. Beispielhaft für die Extremitätenregeneration ist die Schwanzregeneration in Abbildung 5 dargestellt.

Unmittelbar nach der Amputation wird die Wundstelle von einem Blutpfropfen verschlossen. Danach setzen Mechanismen der Wundheilung ein. Diese zeichnen sich durch ein schnelles Überwachsen der Schnittfläche durch die den Stumpf umrandende Epidermis aus. Dies geschieht innerhalb eines Tages. Nach Überwachsung mit einer dünnschichtigen Epidermis setzt eine Phagozytose ein, die Blutgerinnsel sowie geschädigte und abgetötete Zellreste abbaut. Mitotische Vorgänge kommen während dieser ersten Tage nahezu vollkommen zum Erliegen. Dagegen können apoptotische Vorgänge beobachtet werden. Kleine Blutgefäße sprießen schnell in das überwachsene Epithel ein, was eine gute Blutversorgung gewährleistet. Im nächsten Schritt setzt die Dedifferenzierung im Bereich zwischen schützender Epidermis und mesodermalen Gewebe ein. Dabei gehen die Signale für die Dedifferenzierung von der Epidermis

aus. Dies konnte dadurch gezeigt werden, dass nur originäre Wundepidermis die Fähigkeit besitzt, mesodermales Gewebe zur Dedifferenzierung anzuregen. Bei Versuchen wurde vom Stumpf einer Amputation das Blastem entfernt und normales Hautgewebe darauf transplantiert. Dies führte zu einem Abbruch der Regeneration [44]. Die Zellen im Amputationsbereich lösen sich aus dem jeweiligen Zellverband, beispielsweise Knochenzellen, und verändern ihren zellulären Charakter von differenzierten Zellen hin zu Stammzell-ähnlichen Vorläuferzellen, die zur Proliferation befähigt sind. Dies konnte durch das Erscheinen von Pax7 (Muskelmarker) exprimierenden Zellen während der Regeneration im Wundbereich gezeigt werden. Diese dedifferenzierten Zellen beginnen mit einer starken Proliferation, was zu einer Anreicherung unter der schützenden Epithelschicht führt. Die Epithelschicht entwickelt sich zu einer apikalen epidermalen Kappe (apical epidermal cap, kurz AEC), welche die Richtung der regenerierenden Extremität vorgibt. In der Literatur wurde diese Struktur mit dem aus der Beinentwicklung bekannten *apical ectodermal ridge* (AER) verglichen [45]. Dieser Vergleich wurde sogar auf molekularer Ebene geführt, nachdem gezeigt werden konnte, dass die AEC die gleichen embryonalen Faktoren ausbildet wie auch die AER. Dies sind insbesondere die Familie der Fibroblasten Wachstumsfaktoren (FGF), darunter FGF-1 und FGF-2 [46,47,48].

Durch weiterhin starke Proliferation bildet das Blastem eine konische Struktur in Richtung der zu regenerierenden Extremität aus. Die Zellen in der proximalen Stumpfregeion verlangsamen daraufhin die Proliferationsrate und beginnen durch Redifferenzierung erste Muskel- und Knorpelzellverbände auszubilden. Bei der Beinregeneration geschieht zeitgleich im distalen Bereich die Differenzierung der späteren Zehen aus dem konischen Blastem. Die Redifferenzierung folgt dabei einem fest vorgegebenen Muster. So konnte entgegen früherer Ergebnisse [13] kürzlich gezeigt werden, dass dedifferenzierte Zellen aus Knochen-, Muskel-, Nerven-, und Hautgewebe sowie Schwannsche Zellen nicht zu beliebigen Zellen ausdifferenzieren können. Vielmehr zeigen nur Haut- und Knochenzellen bezüglich der Ausdifferenzierung die Möglichkeit, zu verschiedenen Zelltypen zu differenzieren. Dagegen können ehemalige Muskelzellen nur wieder zu Muskelzellen und Schwannsche Zellen nur wieder zu Schwannsche Zellen auswachsen [14,49].

Zu diesem Zeitpunkt ist die Durchdringung des regenerierenden Bereiches durch nachwachsende Nervenfasern bereits sehr engmaschig. Diese spielen für die Steuerung der Regeneration eine essentielle Rolle. Bereits 1823 konnte durch den englischen Physiker Todd gezeigt werden, dass bei Denervierung des Stumpfes keine Regeneration stattfindet. In diesem Zusammenhang wurden die Oberflächenrezeptoren Prod1 und Meis1/2 entdeckt [50]. Diese

Einleitung

Rezeptoren haben die gemeinsame Eigenschaft, in proximo-distaler Richtung unterschiedliche Expressionsniveau zu zeigen und bei Behandlung mit Retinolsäure (RA) eine veränderte Regulation (im Folgenden als Deregulation bezeichnet) aufzeigen. Es konnte bereits 1983 gezeigt werden, dass die Zugabe von Retinolsäure die positionelle Information im Blastem in Abhängigkeit der Konzentration verändert. Man spricht hier von Proximalisation. Durch Zugabe von niedrigen RA Konzentrationen wurde aus einem Mittelarmblastem nur eine Hand regeneriert, durch Gabe von hohen Konzentrationen wurden ganze Beine regeneriert [51].

Zum Prod1-Protein wurden mittlerweile weitere Erkenntnisse erzielt. Dieser Rezeptor bildet einen Gradienten in proximo-distaler Richtung der Extremität aus [52]. Dieser Gradient spielt bei der positionellen Identifikation bzw. dem positionellen Gedächtnis der Zellen eine Rolle. Der Ligand für diesen Rezeptor begründet aber die eigentliche Abhängigkeit der Regeneration von intakter Innervierung. So wird von den die Nervenfasern umschließenden Schwannschen Zellen der Prod1-Ligand nAG (newt anterior gradient Protein) sekretiert [28]. Dieser Faktor stellt das Bindeglied zwischen der nervabhängigen Regeneration und der RA vermittelten positionellen Identität dar. Die Sekretion dieses Faktors löst in den Drüsenzellen, die im Epithel der Wunddermis liegen, ebenfalls die Sekretion von nAG aus. Die Expression von nAG im Blastembereich fördert die Proliferation der Blastemzellen. Der Prozess der Schwanzregeneration läuft analog zur Beinregeneration ab. Die Regeneration einer 1cm langen Schwanzspitze benötigt insgesamt 60 Tage (Abgebildet sind die Stadien der Regeneration in Abbildung 5).

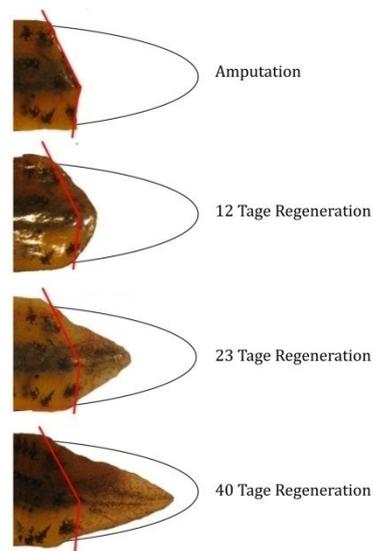


Abbildung 5: Schwanzregeneration des Molches

Eine amputierte Schwanzspitze des Molches regeneriert innerhalb von 2 Monaten. Dargestellt ist der Regenerationsverlauf über die ersten 40 Tage. Zu erkennen ist die Blastembildung innerhalb von 5-12 Tagen nach der Amputation und die AEC 23 Tage nach der Amputation. In schwarz wird schematisch der vollständige Schwanzgröße angezeigt. Bei der Schwanzregeneration wird Haut-, Muskel-, Knochen-, Knorpel- und Zentrales Nervengewebe regeneriert. Auch bei nicht-planarer Schnittführung wird genau der Teil des Schwanzes regeneriert, der distal entfernt wurde. Amputiert wurde ein 1cm langes Stück der Schwanzspitze.

b) *Regeneration des Molchherzen*

Die vorliegende Arbeit beschäftigt sich speziell mit der Regeneration des Molchherzens. Zum besseren Verständnis wird zunächst auf die Anatomie des Molchherzens und anschließend auf den Ablauf der Herzregeneration eingegangen.

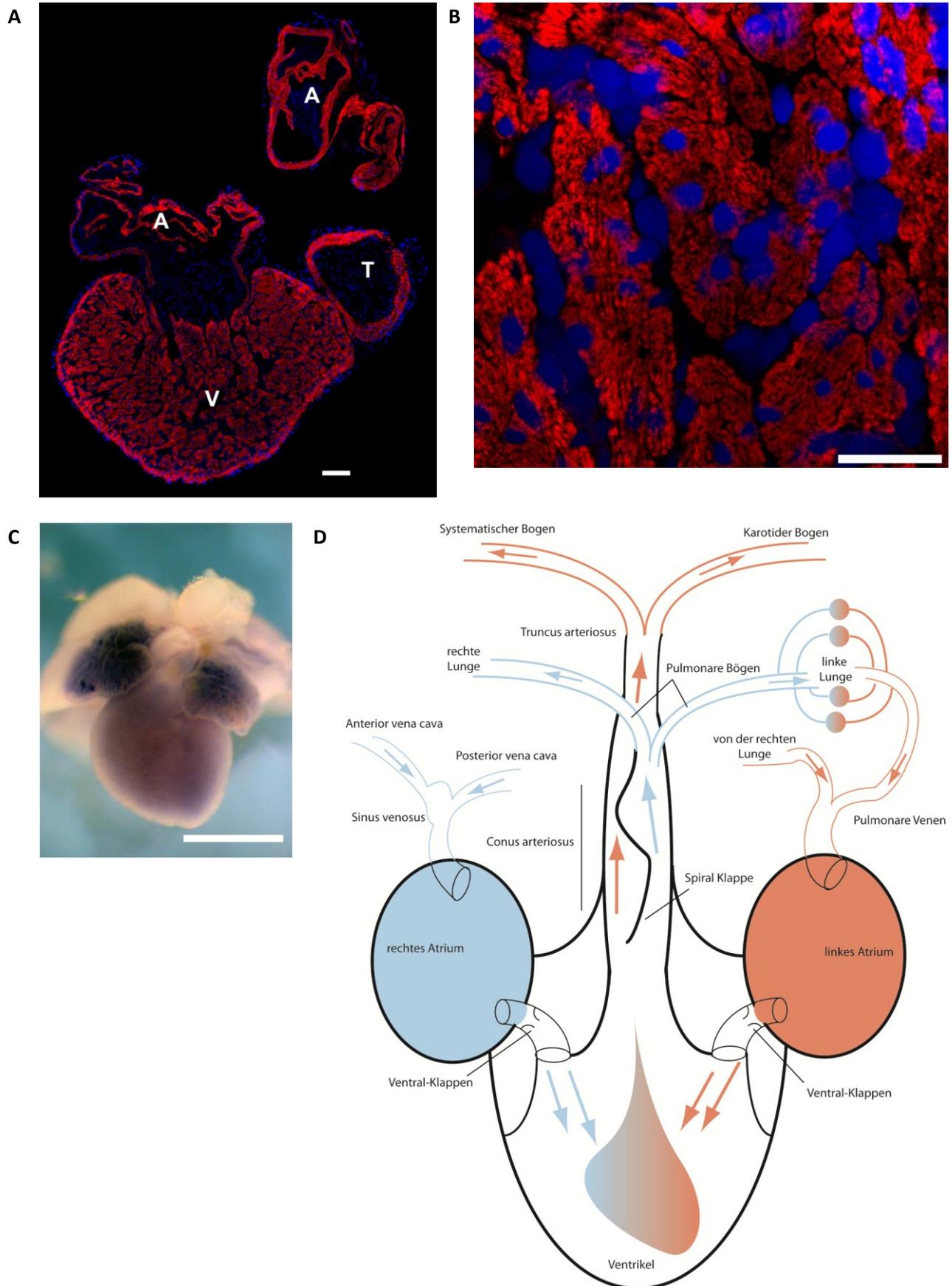
1 Anatomie des Molchherzen

Innerhalb der Amphibien ist die anatomische Struktur des Herzen ähnlich, zeigt aber auch durchaus artspezifische Eigenheiten. Insbesondere die Proportionen und die innere Struktur variieren und zeigen meist eine starke Adaption an äußere Lebensbedingungen. Das Molchherz

Einleitung

setzt sich aus drei Kammern zusammen. Dies sind neben dem Ventrikel das rechte und das linke Atrium. Im Gegensatz zu Säugern ist der Ventrikel jedoch kein kräftiger, dünnwandiger Hohlmuskel mit verhältnismäßig großem Lumen. Vielmehr zeigen sagittale und frontale Schnitte eine schwammähnliche innere Struktur des Ventrikels mit einem sehr kleinen, klar zu bestimmenden Lumen (siehe Abbildung 6 A, B und C). Der Molch besitzt kein Septum für die Trennung zwischen sauerstoffreichem und sauerstoffarmen Blut. Innerhalb des Ventrikels kommt es demnach zu einer partiellen Vermischung der beiden Blutkreisläufe, die jedoch durch die das Atrium abschließenden Ventriklappen gesteuert wird. Der Blutkreislauf stellt sich wie folgt dar:

Im *Sinus venosus* vereinigt sich das sauerstoffarme Blut des Körpers aus *Anterior vena cava* und *Posterior vena cava* und wird durch eine Öffnung in das rechte, meist etwas größere Atrium geleitet. Durch die Ventriklappe fließt das Blut in den schwammähnlichen Ventrikel. Durch Kontraktion des Ventrikels wird das Blut in den vertikal gelegen Teil des *Truncus arteriosus*, den *Conus arteriosus*, gepumpt. Hier regelt eine Spiralklappe den Blutfluss in den Pulmonaren Bogen, der das Blut zur Lunge führt. Nach der Anreicherung mit Sauerstoff wird das Blut über die Pulmonare Vene in das linke Atrium geführt. Von dort durchläuft es abermals den Ventrikel und verlässt über den *Truncus arteriosus* das Herz. Der *Truncus arteriosus* teilt sich in den Karotiden und den Systemischen Bogen. Der Karotide Bogen versorgt den Kopf des Molches, der Systemische Bogen den Körper (schematischer Ablauf des Blutflusses in Abbildung 6 D). Das Herz selbst zeigt keine Versorgung durch Gefäße, sondern bezieht den benötigten Sauerstoff direkt aus dem durchfließenden Blut. Zellulär finden sich im Ventrikel des Molches primär Kardiomyozyten, Fibroblasten, Melanozyten, Granulozyten, Mastzellen, Monozyten und Nervenzellen [42].



Legende der Abbildung auf der nächsten Seite

Abbildung 6: Ungeschädigtes Molchherz

Einleitung

A Dargestellt ist der sagittale Schnitt durch das Herz eines Molches, Ansicht von dorsal. Zu erkennen sind die beiden Vorhöfe (A) und der Ventrikel (V). Außerdem ist der Truncus arteriosus zu erkennen (T). Nicht zu sehen ist der rückwärtig gelegene Sinus venosus und die Aufgliederung des Truncus in die einzelnen Bögen. Das gesamte Herz hat eine Ausdehnung von ca. 1mm^3 . Der Maßstab des Bildes beträgt 0,2mm. Angefärbt wurden Zellkerne mit DAPI (blau), das Protein Aktin mit Phalloidin gekoppeltem FITC (rot) **B** Dargestellt ist ein Ausschnitt aus dem Bereich der Ventrikelmuskulatur aus dem Bild A. Der Maßstab beträgt $50\mu\text{m}$. (Erstellung der Bilder A und B Dipl. Biologin Tanja Piatkowski) **C** Dargestellt ist das eingebettete ungeschädigte Herz eines Molches. Die beiden dunklen Strukturen zeigen die beiden Vorhöfe, dazwischen unterhalb ist der Ventrikel angeordnet. Der Maßstab des Bildes beträgt 1mm (Erstellung Dr. Thilo Borchart) **D** Schematische Darstellung des Blutkreislaufes im Molch. Sauerstoffreiches und sauerstoffarmes Blut mischen sich im Ventrikel. Über eine Spiralklappe im Conus arteriosus wird der Blutfluß Richtung Lungen oder Körper gelenkt. Der Weg des sauerstoffarmen Blutes führt über die Lungen und das linke Atrium wieder in den Ventrikel, sauerstoffreiches Blut findet seinen Weg über den Körper, den Sinus venosus und das rechte Atrium zurück zum Herzen.

2 Herzregeneration des Molches

Bei der Regeneration des Molchherzens handelt es sich nach Einteilung von Morgan und Driesch am ehesten um eine nicht blastemgebundene Epimorphose. Die Abfolge der Regeneration nach partieller Resektion (1/16, 1/8) wurde ursprünglich schon 1974 von Oberpriller et al. [42] beschrieben. Der Herzventrikel besteht aus *Trabeculae*. Dies sind Myozyten, Fibroblasten und Nervenfasern, die von einer dünnen, Einzell-Schicht Epikard umfasst sind. Die entstehende Wunde wird nach Resektion von einem Blutpfropfen verschlossen. Nach etwa zehn Tagen beginnt sich der Pfropfen aufzulösen, es sammeln sich Lymphozyten im Wundrandbereich. Bereits nach dieser Zeit zeigt sich mitotische Aktivität in den Trabeculae. Nach zwei Wochen hat sich eine dünne Epikardwand unterhalb der Restbestände des Pfropfens gebildet, die die Wundstelle gänzlich verschließt. Der Pfropfen wird durch Makrophagen vollständig abgebaut. Nekrotische Prozesse des Gewebes an der Wundstelle werden beendet. Nach drei Wochen bilden sich originäre Gewebestrukturen, mitotische Prozesse in den Trabeculae steigen stark an. Nach vier Wochen ist die Wundheilung abgeschlossen. Jedoch sind nur einzelne funktionelle Myozyten vorhanden, die nur kleine Myofibrillen ausgebildet haben. Oberpriller et al. beendeten an dieser Stelle ihre Beobachtungen und schlossen auf eine teilweise Regeneration beim Molchherzen mit eingeschränkter Funktionalität. Später wurde der Beobachtungszeitraum auf 70 Tage verlängert, was die Entdeckung von stark proliferierenden, partiell dedifferenzierten Kardiomyozyten in der dritten Woche nach Schädigung ermöglichte [53,54,55].

Eine erneute Untersuchung der Herzregeneration beim grünlichen Wassermolch unter veränderter Schädigungsmethode [56] zeigte das vollständige Regenerationspotential des

Herzen. Eine am MPI-BN eigens entwickelte Methode zur Herzmuskelschädigung stellt Schädigungen durch Infarkt oder Entzündung nach. Der Erhalt der äußeren Morphologie des Ventrikels bei gleichzeitiger Schädigung funktioneller Einheiten steht dabei im Mittelpunkt. Die mechanische Zerstörung des Ventrikels wird durch mehrfache Quetschung mit feinen Pinzetten erreicht, so dass das Innere des Herzmuskels eine starke Desorganisation aufzeigt. Eine komplette funktionelle Neuordnung konnte innerhalb von 12-16 Wochen beobachtet werden [57]. Die ursprüngliche trabekulierte Struktur wurde wieder hergestellt. Es bildete sich kein Narbengewebe. Eine Übersicht zu den Stadien der Regeneration zeigt [57] und (Piatkowski, Borchardt, Braun, in Vorbereitung) anhand von sagittalen Schnitten durch das Molchherz.

Der Molch ist hinsichtlich seiner außergewöhnlichen regenerativen Fähigkeiten sehr gut charakterisiert. Jedoch wurde er aufgrund seiner erschwerten Handhabung und vor allem aufgrund von fehlenden Sequenzinformationen und fehlenden etablierten Methoden nur selten als Modellorganismus genutzt. Deshalb drängt sich die Frage auf, warum der Molch überhaupt anderen, besser etablierten Modellorganismen wie beispielsweise dem Zebrafisch vorzuziehen ist. Diese Frage lässt sich wie folgt beantworten:

Obwohl der Molch kein vierkammeriges Herz wie der Mensch besitzt, hat er mit seinem dreikammerigen System durchaus eine höhere Vergleichbarkeit mit dem menschlichen Herzen als etwa die anatomische Struktur des Fischherzens (Siehe Abbildung 7 und Kapitel 3 in [58]), welches einem doppelt geknickten Schlauch ähnelt. Auch ist der Molch genetisch mit Mensch und Maus wesentlich stärker verwandt als der Fisch. Ein weiteres Argument ist das am MPI-BN etablierte Schädigungsmodell des Molchherzens, das einem Infarkt nahe kommt. So wird nicht wie etwa beim Fisch das Herz durch eine partielle Amputation geschädigt, sondern gezielt nur die Struktur des Ventrikels (auch partiell möglich) durch Quetschung geschädigt. Dabei entstehen Ventrikelbereiche, die an sich keine kontraktile Arbeit mehr leisten können und die Gesamtfunktion des Organs beeinträchtigen. Diese Bereiche werden dann durch Apoptose und Proliferation neu aufgebaut. Die eigentliche Morphologie des Herzen wird jedoch nicht verändert. Dieses Modell stellt demnach den menschlichen Herzinfarkt gut nach. Bei einem Infarkt sterben Teilbereiche des Ventrikels durch Unterversorgung mit Sauerstoff ab, beziehungsweise verlieren einen Großteil der funktionstüchtigen Kardiomyozyten. Dadurch verlieren diese Bereiche ebenfalls ihre kontraktile Funktion und beeinträchtigen die Gesamtfunktion des Organes Herz. Solche geschädigten Bereiche im menschlichen Herz bilden dann aber im Gegensatz zum Molch narbige Fibrosen aus, die zu einer dauerhaften Funktionsbeeinträchtigung des Herzen führen. Eine Etablierung einer vergleichbaren

Einleitung

Schädigungsmethode bei Fischen ist bisher nicht bekannt. Dies liegt, neben dem um ein Vielfaches kleineren Herzen, vor allem auch an der kaum vergleichbaren äußeren Struktur der Herzen.

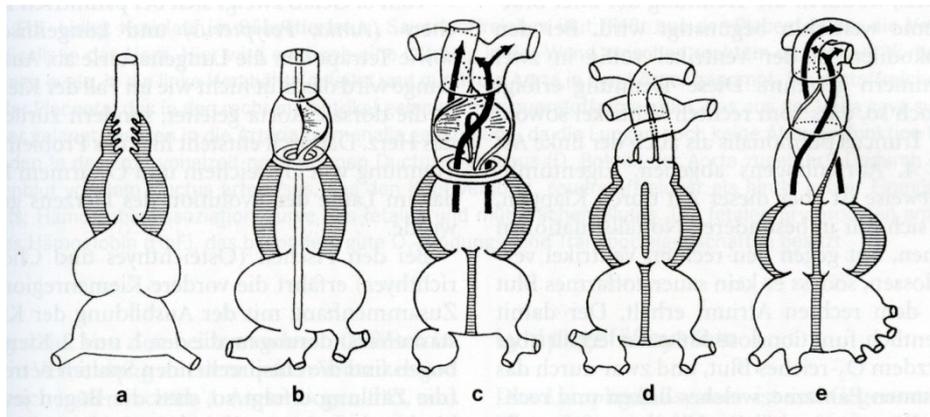


Abbildung 7: Schematische Darstellung der Hauptformen von Wirbeltierherzen

Die einzelnen Herzabschnitte sind hintereinander gezeichnet, Ventrikelwand waagrecht schraffiert, ebenso Septen in Atrium und Ventrikel. Besonders berücksichtigt ist die Torsion des Truncus arteriosus. **a** Fischherz, keine Torsion; **b** Herz eines Lungenfisches (*neoceratodus*), beginnende arterielle Torsion; **c** Zwischenform von Amphibien und Reptilien, Ausbildung eines Vorhofseptums und einer offenen conotruncalen Trennwand. Deutliche Torsion am arteriellen Herzende; **d** Reptilienherz, starke Torsion, drei Schlagadern: Arteria pulmonalis, linke und rechte Aorta, weitergeführte Septierung. Scheidewände im conotruncalen Bereich nicht eingezeichnet; **e** Säugerherz, vollständige Septierung, komplette, 180° betragende Torsion. Das Vogelherz ist ähnlich aufgebaut, nur verläuft hier der Aortenbogen nach rechts. Entnommen aus [59].

D. Untersuchungen am *N. viridescens* mit Hilfe moderner Hochdurchsatzverfahren

Nachdem im vorherigen Kapitel der grünliche Wassermolch als Modellorganismus für regenerative Prozesse vorgestellt wurde, soll im Folgenden dargelegt werden, welche Informationen aus molekularbiologischer Sicht bereits über den Molch vorliegen. Die im Vergleich zu anderen Organismen sehr geringe Menge an vorhandener Information (insbesondere genomische Information) soll die Besonderheiten der Arbeit mit diesem Organismus betonen. Wie aus den folgenden Kapiteln hervorgeht, kann die Analyse von Genom, Transkriptom und Proteom als wissenschaftliche Herausforderung betrachtet werden, da

vergleichbare Arbeiten zum Molch zumeist fehlen und sich bereits allein aus der Größe des Molchgenomes besondere Schwierigkeiten ableiten.

1. Molekularbiologischer und bioinformatischer Zugang des *Notophthalmus viridescens*

Mit Ausnahme des Zebrafisches haben viele regenerative Modellorganismen gemein, dass nur unzureichende Informationen in Form von Nukleotidsequenzen (Genom oder Transkriptom), nicht kodierenden RNA oder dem Proteom vorliegen. Insbesondere fehlen auch für verschiedene Molekülklassen, wie beispielsweise kleine RNA (siRNA, miRNA), öffentlich zugängliche Informationen. Zudem wird das Genom des Molches mit $2,5 \times 10^{10}$ bp auf etwa die zehnfache Größe des menschlichen Genoms geschätzt, was eine komplette Entschlüsselung trotz hochmoderner Sequenzierungsmethoden bisher ausschließt, da eine Assemblierung ohne Matrize erst bei entsprechend hoher Abdeckung erfolgen kann. Um eine solche hohe Abdeckung zu erreichen, würden die Kosten denen einer herkömmlichen, Sanger basierenden Sequenzierungsmethode entsprechen. Das Transkriptom des grünlichen Wassermolches ist bislang nur am MPI-BN genauer untersucht worden und bietet anhand einer ersten cDNA Bank und der Sequenzierung von 11520 isolierten cDNAs (expressed sequence tags, EST) einen ersten Überblick [60]. Öffentliche Proteindatenbanken enthalten ohne diese EST Sequenzen weniger als 100 nicht redundante annotierte Aminosäuresequenzen. Aufgrund der geringen Sequenzinformation gibt es für den Molch keine kommerziellen Produkte für Hochdurchsatzanalysen, wie beispielsweise Microarrays. Will man solche Screening-Techniken verwenden, müssen entsprechende Arrays selbst hergestellt werden. Darüber hinaus existieren bislang nur wenige etablierte molekularbiologische Werkzeuge wie Vektor-, Viren- oder Zellkultursysteme, die die gezielte genetische Manipulation oder Untersuchung des Organismus *N.viridescens* ermöglichen. Auch gibt es nur sehr wenige, gegen Molchproteine gerichtete, kommerziell verfügbare Antikörper, die beispielsweise Immunohistologische Färbungen (IHC) oder Western Blots zulassen. Hier ist man auf Antikörper, die gegen homologe Strukturen der Proteine anderer Organismen gerichtet sind, angewiesen. Auch DNA-Sonden wurden bislang kaum verwendet.

Diese aus molekularbiologischer Sicht beschränkten Kenntnisse zum grünlichen Wassermolch ziehen einen erschwerten bioinformatischen Zugang nach sich. Das Fehlen von entsprechenden öffentlichen Datenbanken erfordert zunächst eine komplett eigene

Einleitung

Datenhaltung, Assemblierung und Datenannotation. Bei der Annotation sind Schwierigkeiten wie der Umgang mit Assemblierungsinformationen oder der fehlenden Möglichkeit von 1:1 Zuordnungen zu beachten. Auch fallen rein bioinformatische Problemstellungen an, wie etwa die Programmierung von und der Zugriff auf Datenbanken mit mehreren 100.000 Einträgen.

2. Übersicht der Techniken zur quantitativen Transkriptomanalyse

Neben der Reversen Transkriptase Polymerase Ketten Reaktion (RT-PCR), die als gängige Methode zum Nachweis einzelner Transkripte dient, hat sich im Laufe der letzten Jahre die Array-Technologie als Hochdurchsatz-Screening-Verfahren des Transkriptoms etabliert. Dabei ermöglicht die Microarray-Technologie die parallele Analyse der Expression von vielen tausend Genen. Außerdem zeichnet sich diese Methode durch die Wiederverwendbarkeit der erzeugten Daten aus. So können einmal erstellte Datensätze unter verschiedenen Gesichtspunkten analysiert werden. In diesem Zusammenhang sind die ArrayExpress [61] und GEO [62] Datenbank zu nennen, in der öffentlich zugänglich Array-Daten abgelegt sind. Grundsätzlich unterscheidet man zwischen cDNA und Oligonukleotid Arrays, zum Beispiel kommerziellen *Affymetrix* Arrays. Bei den cDNA Arrays unterscheidet man zusätzlich je nach Untergrund, auf die die cDNA Sequenzen aufgetragen werden, zwischen Glas- und Membran-Arrays. Da für diese Arbeit cDNA Glas-Arrays hergestellt und verwendet wurden, sollen nur diese im Folgenden genauer erklärt werden. Der erstmalige Einsatz der Glasarrays erfolgte 1995 an der Stanford University (Kalifornien, USA). Das Prinzip beruht darauf, dass cDNA Sequenzen (500 – 1500 bp) aus einer PCR Reaktion mittels „Spotter“ auf einen Glaschip aufgebracht werden. Die Zuordnung cDNA zu Position auf dem Array wird im Spotting Pattern festgehalten. Neuere Arrays fassen bis zu 50.000 Punkte. Im Unterschied dazu werden bei Oligonukleotid Arrays nur kurze Sequenzen bis zu einer Länge von 30 bp aufgebracht.

cDNA Arrays ermöglichen die Analyse der Genexpression eines Zelltyps oder eines Gewebes zu einem speziellen Zeitpunkt. Dabei kann nicht direkt quantitativ auf die mRNA Menge zurückgeschlossen werden, sondern nur relativ im Verhältnis zu einem Vergleichszelltyp oder Vergleichsgewebe. Ein typischer Versuch stellt sich so dar, dass die Genexpression relativ zwischen einem geschädigten oder krankhaften Gewebe und seinem gesunden Pendant analysiert wird. Dazu wird Gesamt-RNA sowohl aus dem geschädigten als auch gesunden Gewebe isoliert. Anschließend wird die mRNA mittels reverser Transkription in komplementär-DNA (cDNA) umgewandelt und mit jeweils unterschiedlichen Fluoreszenzfarbstoffen markiert.

Abschließend wird die markierte cDNA des gesunden Gewebes und des geschädigten Gewebes im Verhältnis 1:1 gemischt und dieses Gemisch auf das Array aufgebracht. Die markierte Ziel (*Target*) cDNA hybridisiert mit der gespotteten (*Probe*) cDNA auf dem Array. Überschüssige und ungebundene cDNA wird in einem zweiten Schritt abgewaschen. Durch Detektion der Signale durch einen geeigneten hochauflösenden Scanner können zwei Bilder des Arrays erstellt werden. Dabei wird das Array nach Anregung der Fluoreszenzfarbstoffe jeweils einmal mit der Emissionswellenlänge der beiden Fluoreszenzfarbstoffe gescannt. Die beiden Bilder werden anschließend verglichen. In einem als *Gridding* bezeichnetem Arbeitsschritt wird dabei zunächst das Bild in Bildabschnitte unterteilt, die jeweils nur noch einen Spot des Arrays enthalten. Im folgenden Arbeitsschritt, der als *Segmentation* bezeichnet wird, wird innerhalb dieser Bildabschnitte jeweils über Algorithmen der Bildverarbeitung der Bereich des eigentlich hybridisierten Spots errechnet und umgebende Bereiche als Hintergrund deklariert. Abschließend wird beim Schritt der *Extraktion* ein Repräsentationswert (Intensität) für jeden Spot errechnet. Durch Bildung des Verhältnisses zwischen den beiden gemessenen Intensitäten lässt sich ein Rückschluss auf die Richtung der Veränderung einer Genexpression als auch auf die relative Änderung ziehen. Um Nebeneffekte, die durch Hintergrundstrahlung und Verschmutzungen entstehen können, auszuschließen, müssen die Intensitäten zuvor durch geeignete Methoden angeglichen werden. Das bevorzugte Einsatzgebiet der cDNA Arrays stellt sich folgendermaßen dar:

- *Tissue profiling*: welche Gene sind in einem Gewebe überhaupt exprimiert?
- Vergleich zwischen krankhaftem, beispielsweise Tumorgewebe, und gesundem Gewebe
- Beobachtung dynamischer Prozesse durch mehrere zeitliche Momentaufnahmen

Kommerzielle Microarrays sind für den grünlichen Wassermolch nicht verfügbar. Dies liegt sowohl an der geringen Sequenzinformation, die für diesen Organismus vorhanden ist, als auch der verhältnismäßig kleinen Forschungsgemeinde, die mit diesem Modellorganismus arbeitet.

In letzter Zeit hat sich neben der Array-Technologie durch neue technische Möglichkeiten eine weitere Hochdurchsatzmethode etabliert, das *Deep Sequencing*. Der Ansatz des *Deep Sequencings* wird durch neue Sequenzierungsmethoden von Geräten der sogenannten zweiten Generation verfolgt. Hier kommt nicht mehr die traditionelle parallele Sanger Methode durch Kettenabbruch zum Einsatz. Vielmehr wird eine DNA-Polymerase während der DNA-Gegenstrangsynthese durch hochauflösende Kameras live beobachtet. Dabei ist der Einbau eines neuen Nukleotides in den Gegenstrang an eine enzymatische Lichtgenerierung via Luziferase

Einleitung

gekoppelt, der durch ausschließliche Zugabe einer Art von Nukleotiden zugeordnet werden kann. Über die Lichtstärke kann auch der Einbau von Poly-N Sequenzen abgebildet werden. Diese Art der Sequenzierung wird als Pyrosequenzierung bezeichnet. Marktreife Anwendungen dieser Sequenzierungsart finden sich im 454 FLX Sequencer (Roche). Diese Systeme erlauben die Generierung von bis zu 5Giga Basen /Woche/Gerät bei einer maximalen Leselänge von 400bp.

Weitere etablierte Hochdurchsatztechniken sind der SOLiD 3 Plus Sequencer (ABI), der auf dem Prinzip des *sequencing by ligation* beruht. Der SOLiD erzeugt Leselängen von etwa 35 bp bei insgesamt 6 Giga Basen pro Lauf. Daneben gibt es weiterhin noch den Genome Analyser Sequencer (Illumina), der auf dem Prinzip des *sequencing by synthesis* beruht. Dieses Gerät erreicht 5,2 Giga Basen/Woche/Gerät bei einer maximalen Leselänge von 72 bp.

3. Übersicht der Techniken zur Proteomanalyse

Die Untersuchung des Proteoms mit Hochdurchsatzverfahren hat sich in den letzten Jahren als eigenständige Disziplin unter Nutzung der Massenspektrometrie entwickelt. Dabei ist neben der anfänglichen reinen Detektion von Proteinen und Peptiden die Detektion von modifizierten Peptiden, insbesondere phosphorylierten Peptiden in den Vordergrund gerückt. Darüber hinaus wird seit 2003 intensiv an der Entwicklung von quantitativen Methoden gearbeitet. Im Blickpunkt der quantitativen Proteomics stehen Modifikationen an Peptiden oder Proteinen, die zu definierten Unterschieden in der Masse führen. Diese Modifikationen erlauben die Messung von zwei oder mehr interessierenden Zuständen in einem gemeinsamen Messgang, was die Quantifizierung aus einem Spektrum heraus zulässt. Zu unterscheiden sind:

- Metabolische Markierung
- Chemische Markierung
- Verwendung interner Standards

Darüber hinaus wird versucht, auch ohne Modifikationen quantitative Aussagen treffen zu können. Dies verlangt jedoch den *in silico* Vergleich zwischen mehreren separaten Messungen und das Wiederfinden von Spektren einzelner Proteine bzw. Peptide. Bislang konnte diese Technik nur für ungenaue Expressionsaussagen genutzt werden. Die chemische und metabolische Markierung verspricht dagegen sehr viel sensitivere Analysen.

Bei dieser metabolischen Markierung werden stabile Isotope in das Proteom des Versuchsorganismus (oder Zellkulturen) eingebracht. Hierzu zählt die Verwendung von schweren N^{15} Isotopen, die über die Nahrung eingebracht werden. Erstmals wurde dies für quantitative Aussagen 1998 von Langen et al. verwendet, die über eine N^{15} substituierte Fütterung Mikroben markierten [63]. Heute werden partielle N^{15} Markierungen auch bei Säugern, beispielweise Mäusen durchgeführt [64]. Ein jedoch nicht unwesentlicher Aspekt bleibt die hohe Komplexität der N^{15} markierten Spektren. Durch den nicht klar definierten Abstand zwischen detektierten Peptiden und korrespondierenden N^{15} Derivaten ist die Auffindung von Isotopencluster Paaren sehr komplex. Darüber hinaus wirkt sich die hohe Anzahl an verwendeten N^{15} , auch in kurzen Peptiden deutlich auf die Breite der Isotopencluster aus, was häufig zu Überlagerungen führt.

Neben der metabolischen N^{15} Markierung hat sich seit 2002 die Verwendung von stabilen Isotopen in Aminosäuren etabliert [65]. Häufig verwendet werden die Aminosäuren Lysin und Arginin, bei denen alle C^{12} Atome durch schweres C^{13} bzw. N^{14} durch N^{15} Isotope ersetzt sind. Durch Kombination der benannten stabilen Isotope wurden definiert schwere Aminosäuren mit einem Massenunterschied von 6, 8, 10 oder 12 Da konstruiert. Da es sich bei Lysin um eine essentielle Aminosäure handelt, kann durch eine geeignete Diät die Anreicherung der schweren Aminosäureisotope in zu untersuchenden Zellpopulationen erreicht werden. Wählt man eine geeignete Markierungszeit (abhängig vom zu untersuchenden Organismus/Zellart), kann das zu untersuchende Proteom komplett mit schweren essentiellen Aminosäuren markiert werden. Durch die Verwendung von geeigneten Enzymen bei der Aufbereitung der Proteinproben, beispielsweise Lys-C (schneidet C-terminal nach Lysin), sind alle im vollständigen Verdau entstehenden Peptide C-terminal mit einer schweren Aminosäure markiert. Führt man diese Probenaufbereitung für das Experiment und die schwere Kontrolle gleichzeitig durch, finden sich im Spektrum zwei Isotopencluster zu einem Peptid, die sich im Masse/Ladungsverhältnis genau um die Differenz der verwendeten schweren und leichten Aminosäure unterscheiden.

Die als *stable isotope labeling with amino acids in cell culture* (SILAC) bezeichnete Methode wurde zunächst in Zellkulturen verwandt [66]. Später konnte diese Art der Markierung auch für adulte Mäuse *in vivo* genutzt werden [67]. Hiermit war es erstmalig möglich, gut reproduzierbare quantitative Aussagen *in vivo* tätigen zu können. Ein weiterer Vorteil der SILAC Methode ist die Nutzbarkeit der leicht-schweren Isotopencluster Paare für die Identifizierung in entsprechenden Proteinsequenz Datenbanken. Diese Eigenschaft der SILAC Methode wird in dieser Arbeit häufig genutzt und daher später noch genau beschrieben. Die Methode liefert

Einleitung

generell die höchste Genauigkeit im Bezug auf quantitative Aussagen zum Proteom. Im Gegenzug stellt sich die Datenverarbeitung auf Peptidbasis schwierig dar. So bedarf es geeigneter Software, um Peptide nach deren Identifizierung in einer Proteinsequenzdatenbank zunächst wieder zu ganzen Proteinen und bei mehreren Zuordnungsmöglichkeiten zu Proteingruppen zusammenzufassen. Dieses Problem tritt insbesondere immer dann verstärkt auf, wenn zugrundeliegende Proteindatenbanken partiell oder stark redundant sind oder aber auch Isoformen von einzelnen Proteinen enthalten sind.

Unter chemischen Modifikationen versteht man die nachträgliche Markierung von Proteinen oder Peptiden nach entsprechender Vorbereitung. Dabei werden N terminal (N-hydroxysuccinimide(NHS)) oder an Lysin-Seitenketten chemische Modifikationen mit definierter Masse angebracht. Gängige Modifikationen sind im einzelnen *isotope coded protein labels* (ICPL) [68], *isotope tags for relative and absolute quantification* (iTRAQ) [69], *tandem mass tags* (TMT) [70], und *Acetic/succinic Anhydride* [71]. Die wesentliche Herausforderung bei der Entwicklung von geeigneten chemischen Modifikationen ist die notwendige, nahezu gleichzeitige Elution (von der LC Säule) der unterschiedlich markierten Proteine/Peptide. Wenn durch eine Modifikation eines Proteins/Peptids eine deutliche Verzögerung des Retentionszeitpunktes geschieht, führt dies zu erheblichen Schwierigkeiten bei der nachfolgenden Datenanalyse, da ein leicht-schweres Protein/Peptid Paar nicht mehr im gleichen Spektrum erscheint. Diese Problematik ist bei stabilen Isotopen in Aminosäuren nicht gegeben. Weiterhin ist bei der chemischen Modifikation eine Aufbereitung der Proteine vor der Markierung nötig. Dieser Arbeitsschritt wird für jede Probe separat durchgeführt und kann daher zu Fehlern führen, die durch unterschiedliche Prozessierung verursacht werden.

Die Nutzung interner Standards ist die älteste Methode bei quantitativen Fragestellungen und wird gemeinhin als *absolute quantification of proteins* (AQUA) bezeichnet [72]. Bei dieser Methode werden bekannte Mengen eines mit stabilen Isotopen markierten Peptides zur Probe hinzugefügt und dann anhand des Intensitätenniveaus die Menge des originär leichten Peptids errechnet. Diese Methode ist sehr genau, empfiehlt sich aber nur bei der gezielten Untersuchung von einem oder ein paar speziellen, zuvor bekannten Proteinen. Dieser Ansatz ist beliebt für die Analyse und Validierung von speziellen Peptidmodifikationen [73].

Bei der Wahl der Quantifizierungsmethode ist generell darauf zu achten, welche Genauigkeit verlangt ist, welche Art der biologischen Proben vorliegt und an welcher Stelle die Prozessierung

der Proben Einschränkungen vorgibt. Wie Abbildung 8 zeigt, können die unterschiedlichen Methoden der Modifikation zu unterschiedlichen Zeitpunkten der Proben-Prozessierung eingesetzt werden.

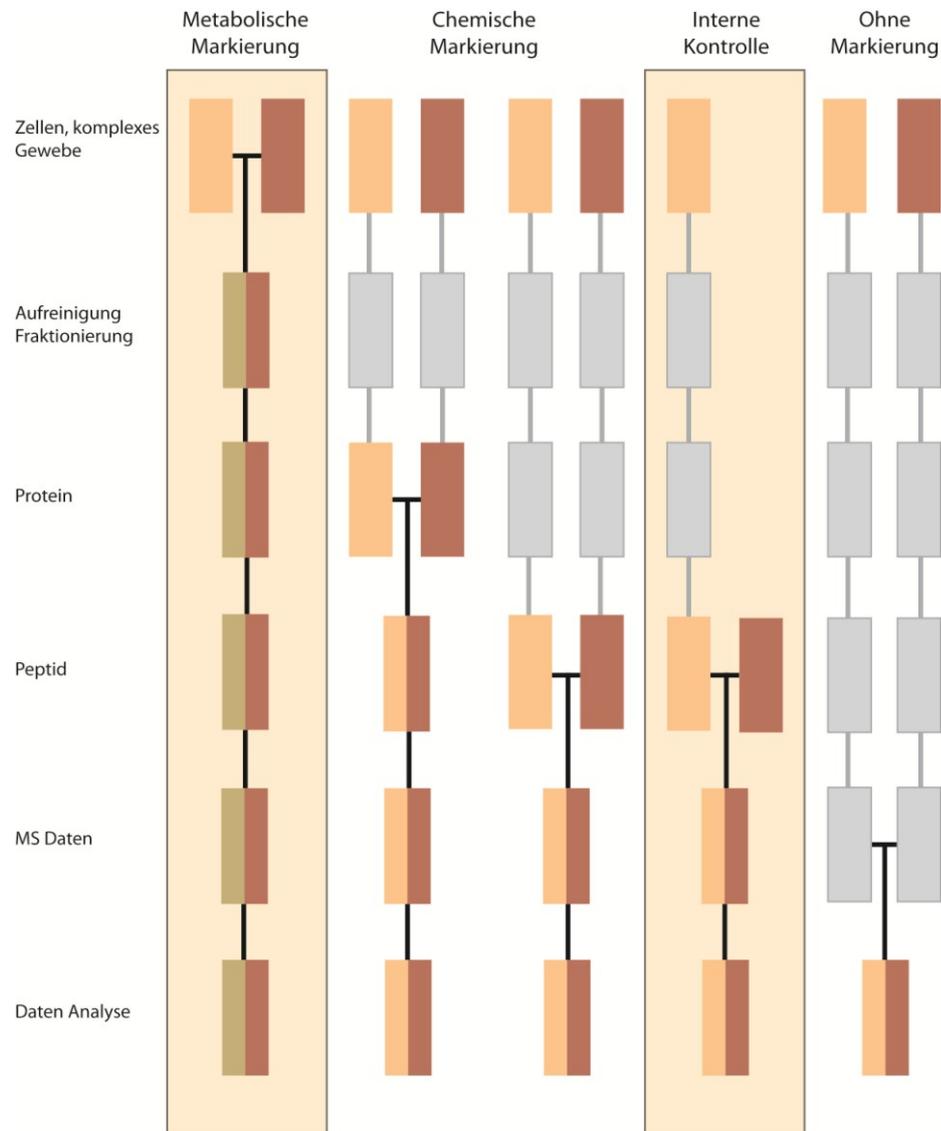


Abbildung 8: Ablauf einer markierten Massenspektrometrischen Analyse

Die braunen und orangen Rechtecke repräsentieren zwei unterschiedliche experimentelle Stadien. Die Spalten von links (metabolische Markierung) nach rechts(ohne Markierung) zeigen den Arbeitsablauf für verschiedene Markierungsmethoden. Die horizontalen Ebenen zeigen den sequenziellen Ablauf einer Massenspektrometrischen Analyse vom komplexen Gewebe/Zellverband über die Proteinaufbereitung bis zur Datenanalyse. Die schwarzen Linien zeigen den Zeitpunkt an, zu dem die experimentellen Stadien gemischt werden. Die grauen Kästchen zeigen Arbeitsschritte, in denen durch separierte Arbeitsschritte probenabhängige Prozessierungsfehler auftreten können. Je früher die Mischung der Proben stattfindet, desto geringer sind die prozessierungsbedingten Fehler. Die Abbildung wurde in Anlehnung an [74] erstellt.

Einleitung

Je mehr Arbeitsschritte separat ausgeführt werden, desto häufiger können Unterschiede in der Prozessierung Einfluss auf die Quantifizierung nehmen. Abbildung 8 soll außerdem verdeutlichen, dass nur die metabolische Markierung eine komplette gemeinsame Prozessierung zweier Proben erlaubt und somit die o.g. Art von Fehlern eliminiert.

E. Motivation und Ziele der Doktorarbeit

Der grünliche Wassermolch *N. viridescens* ist einer der wenigen Organismen, die im adulten Zustand die Fähigkeit zur vollständigen Regeneration erhalten haben. Dabei kommen verschiedene regenerative Prozesse (nach Morgan und Driesch) zum Einsatz und führen nach Schädigung zur narbenfreien Rekonstruktion. Die Tatsache, dass *N. viridescens* eine engere genetische Verwandtschaft zu Säugetieren besitzt als Fische, die ebenfalls regenerative Fähigkeiten besitzen, macht ihn zu einem hochinteressanten Modellorganismus. Unter anderem besteht bei Organaufbau, Organfunktion und dem Größenverhältnis der Organe eine höhere Ähnlichkeit. Gemessen an seiner enormen Bedeutung für regenerative Fragestellungen sind kaum Informationen zum *N. viridescens* vorhanden. So gibt es wenige verfügbare Daten bezüglich Aufzucht und Haltung, Lebensbedingungen und Anatomie. Auch aus molekularbiologischer Sicht ist der Wissensstand, wie oben beschrieben, ebenfalls sehr gering, was eine bioinformatische Herausforderung bei der Arbeit mit diesem Organismus darstellt. Daneben sind regenerative Prozesse an sich bislang wenig verstanden. Einzelne Vorgänge und Faktoren, die eine wesentliche Rolle spielen konnten schon identifiziert werden, jedoch ist man von einem umfassenden Bild der molekularen Vorgänge, die eine Geweberegeneration vermitteln, weit entfernt. Schwerpunkte der bislang erfolgten Untersuchungen liegen auf Bein- und Linsenregeneration. Die Untersuchung der vollständigen Regeneration des Herzmuskels ist dagegen noch ein weitgehend unbekanntes Gebiet. Das grundlegende Verständnis der regenerativen Vorgänge im Herzen ist hinsichtlich der Tatsache, dass Herzerkrankungen den größten Teil an menschlichen Erkrankungen darstellen, die zum vorzeitigen Tod in den Industriestaaten führen, von höchstem Interesse.

Ziel dieser Arbeit ist demnach die Mitgestaltung eines generellen Ansatzes zum Verständnis der regenerativen Vorgänge am Herzen des Molches. Dazu soll durch gezielten Einsatz von Hochdurchsatzverfahren und bioinformatischen Methoden ein möglichst gesamtheitlicher Datensatz erzeugt werden, der die Regeneration des Herzen beschreibt.

Eckpunkte dieses Ansatzes sind:

1. die Erstellung einer umfassenden cDNA Bibliothek zum Transkriptom des Herzens in Regeneration (Projekt in vierköpfiger Arbeitsgruppe),
2. die Erstellung eines cDNA Microarrays zur Transkriptomanalyse (in Kooperation mit einer Microarray-Spotting Arbeitsgruppe am MPI für molekulare Genetik, Berlin),
3. die gezielte Sequenzierung signifikant deregulierter als auch im Vorkommen nachgewiesener cDNAs (Sequenzierung am MPI für molekulare Genetik, Berlin),
4. die Entwicklung einer umfassenden Annotationsroutine für ESTs durch Homologiesuche,
5. die Entwicklung einer Methode zum *in vivo* Einbau von stabilen Isotopen in Proteine des Molches,
6. die Proteomanalysen von regenerierendem Molchgewebe,
7. die Ausnutzung einer Isotopenmarkierung zur Identifizierung von Proteinen in organismusfremden Proteindatenbanken,
8. die Entwicklung einer Methode zur Nutzbarmachung von EST Sequenzbanken zur Proteomanalyse
9. die Identifizierung von bislang unbekanntem Proteinen, die charakteristisch für den Molch und die Regeneration sind.

Im Einzelnen sollte im Rahmen dieser Arbeit die Erstellung und methodische Auswertung einer auf cDNA Microarrays erstellten Zeitreihe zur Herzregeneration durchgeführt werden. Dazu müssen in sämtlichen Arbeitsschritten geeignete Algorithmen entwickelt werden, da nicht auf fertig annotierte Proteinlisten oder annotierte EST Banken zurückgegriffen werden kann. Ein Schwerpunkt liegt auf dem Einsatz von Algorithmen zur Homologiesuche, Alignmenterstellung und zur automatischen Annotation. Das Fehlen von Datenbanken verlangt die Entwicklung einer kompletten eigenen Datenorganisation und Datenhaltung. Durch geeignete bioinformatische Methoden sollen essentielle Expressionsmuster während der Herzregeneration identifiziert werden.

Darüber hinaus sollte eine Methode zur Proteinidentifizierung mithilfe der Massenspektrometrie entwickelt werden. In diesem Zusammenhang soll über die Markierung mit stabilen Isotopen die Nutzung von Proteinbanken fremder Organismen erreicht werden. Außerdem soll eine Methode zur Nutzung von EST Sequenzbanken entwickelt werden, welche die Verwendung der erstellten cDNA Sequenzen zur Proteinidentifizierung erlaubt. In diesem massenspektrometrischen Bereich liegt der Schwerpunkt dieser Arbeit. Bislang wurden

Einleitung

Organismen ohne zugrundeliegende Sequenzinformation nicht im Hochdurchsatz auf Proteinebene untersucht.

Weiterhin wird durch den kombinierten Einsatz dieser Methoden zur Transkription- und Proteomanalyse gezeigt, dass es eine bislang unbekannte Gruppe an Proteinen gibt, die spezifisch für den Molch ist. Dazu werden gezielt Transkripte herausgearbeitet, die zwar nicht annotiert werden können, für die dennoch das Vorkommen von zugehörigen Aminosäuresequenzen gezeigt werden kann.

Insgesamt stellt diese Arbeit unter Berücksichtigung der Eigenheiten des Modellorganismus *N. viridescens* wesentliche Schritte zur Erstellung, Bereitstellung und Verwendbarkeit von Hochdurchsatzdaten auf Transkriptom- und Proteomebene in den Mittelpunkt. Diese werden auf globaler Ebene abgeglichen und daraus Aussagen zu potentiell interessanten Kandidaten-Untergruppen abgeleitet. Insbesondere werden in Ihrer Zusammensetzung bislang unbekannte Kandidaten-Gruppen aufgezeigt, aus denen sich weiterer intensiver Forschungsbedarf ableitet.

V. Material und Methoden

A. Material

1. Oligonukleotide

Tabelle 2: Sequenzen der Sense und Antisense Primer

Transkript	Primer 5' - 3'	Primer 3' - 5'
Contig 1148	151 sense TATTGTTCCGAACTTCGACTG	387 antisense AAAGGTCCACGTGCATTATTC
Contig 1556	77 sense GGAGATATCGAGTGTGCTTC	333 antisense CACACAGCTGCGTTGAAAG
Contig 2460	189 sense TGTGTTCCGAAACGGAATAAG	426 antisense ACCATCGTTCTTGTTCAGG
Fibronectin	220 sense GATATCAGGTCGAGCCTTCTG	423 antisense TGCCTTGTGCCCTTAATAC
ATPase Na/K	196 sense GTCAACGATGTGGAGGATACC	427 antisense GGCAATATGAGAGGAAAGCAG
ATP2A2	987 sense degenerate AAATSGGYAARGTRTATCGAC	1213 antisense degenerate TTVACAGCHCGRGGGTCWG
HSP27 1	406 sense degenerate AACTKATGGTGAAGACCAAGG	718 antisense CACACAGGGTCTTTGTGCTAC
Laminin alpha 2	297 sense ACTTTGATGGAAGTGGATTCG	542 antisense ATTCCCATCACACAACTTC
Slc25a4	125 sense GCAGTAGCTGGTCTTGTTC	407 antisense GTTACAAAACCCATCCGAGAC
similar to human ribosomal protein S21	sense AAGTAACCATGCAGAACGATG	antisense GGCTAACCGAAGGATAGATC

Material und Methoden

2. Lösungen

Alle aufgeführten Lösungen wurden mit bidestilliertem H₂O aus einer Reinwasseranlage angesetzt, das vor Gebrauch autoklaviert wurde. Weitere Lösungen, die nur im Zusammenhang mit einzelnen Anwendungen hergestellt wurden, werden bei den entsprechenden Protokollen separat erwähnt.

Tabelle 3: Lösungen

Chloramphenicol-Stammlösung: 30mg/ml in Ethanol
10 x DNA Ladepuffer: 5g Ficoll 400; 5ml 0,5 M EDTA (pH 8); 5ml 20% SDS; 0,0625g Bromphenolblau; 0,0625g Xylencyanol; mit H ₂ O auf 25 ml
50 x TAE Puffer: 2M Tris-HCl; 57,1 ml konzentrierte Essigsäure; 0,05M EDTA, pH 8.5
Narkosebad: 0,1g Ethyl 3-Aminobenzoessäure, Methansulfonsäure 98%; 100ml H ₂ O
Desinfektionsbad: 5g Sulfamerazin; 1l H ₂ O
Freezing Medium: 65% Glycerol, 0,1 M MgSO ₄ , 25 mM Tris.Cl pH8
Erststrang Puffer Array Hybridisierung: 250mM Tris.Cl (pH 8,3), 375mM KCl, 15mM MgCl ₂
TE Puffer: 10mM Tris pH 8,0, 1mM EDTA
Lysepuffer: 4% SDS, 100mM Tris/HCl pH 7.6, 0.1M DTT
Amonium Bikarbonat (ABC): 0.05 M NH ₄ HCO ₃ in H ₂ O
Lösung UA: 8 M urea, 0.1 M Tris/HCl pH 8.5

IAA Lösung: 0.05 M Iodoacetamid in Lösung UA

Ladepuffer für Proteingele: 3ml 1M Tris-HCl, pH 6.8, 14ml Glycerol 87%, 10ml SDS 10%, 18ml ddH₂O, 2.368mIDTT 1M

Stammlösung 20x SSC Puffer: tri-Natriumcitrat 88,23 g/L (0,3 M), Natriumchlorid 175,32 g/L (3 M), pH 7 (HCl)

Spottingpuffer: 3xSSC, 1,5 M Betain

3. Vektoren

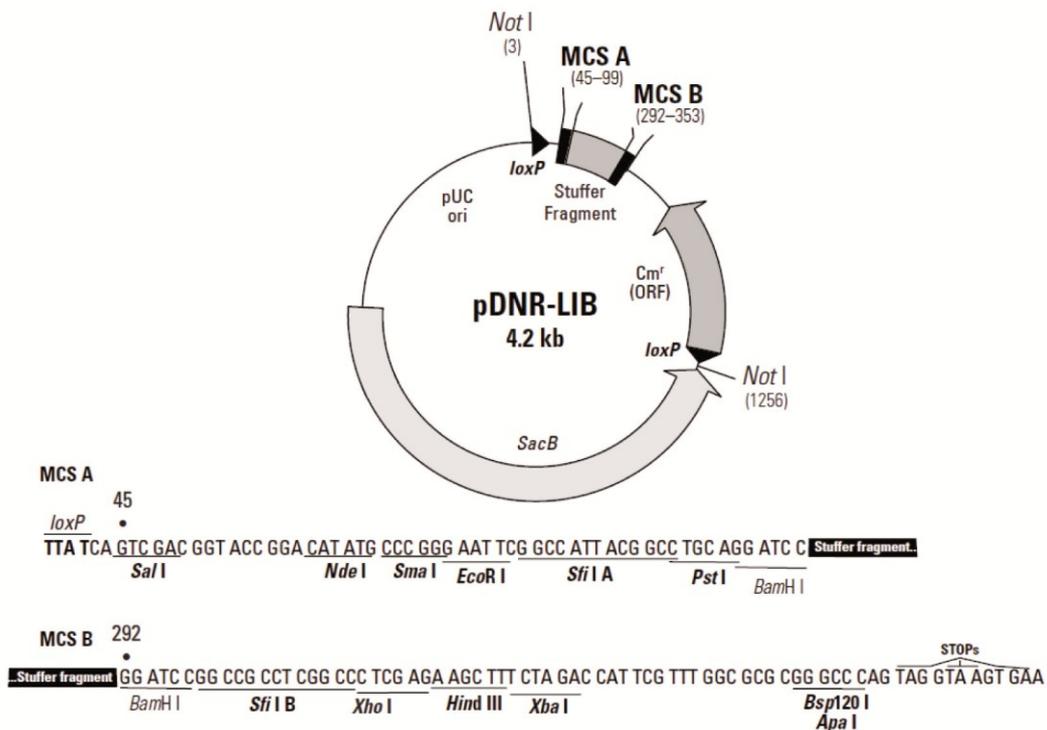


Abbildung 9: pDNR-LIB Vector von BD Biosciences, San Jose, California

Der Vektor wurde für die Erstellung der richtungsabhängigen cDNA Bibliothek verwendet. Es wurden die degenerierten Sfi 1 A und Sfi 1 B Schnittstellen aus der Multiple Cloning Site (MCS) genutzt [75,76].

Material und Methoden

4. Bakterienstämme

XL1-blue: Stratagene, Cedar Creek, Texas, recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F' proAB lacI^qΔM15 Tn10 (Tet^r)]

5. Kommerzielle Sets

ProteoExtract, Calbiochem, Cat. No. 539790

Message Amp II aRNA Amplifikation Kit Cat. No. AM1751

6. DNA Größenmarker

Lambda DNA: EcoRI/HindIII verdaut

pUC 18: Sau3A verdaut

7. Enzyme

Tabelle 4: Verwendete Enzyme

LysC	Wako, Richmond
Proteinkinase K	Boehringer, Mannheim
Restriktionsenzym Sfi1	New England Biolabs, Ipswich
Reverse Transkriptase	Boehringer Mannheim
RNase A	Promega, Madison
T7 RNA Polymerase	Promega, Madison

Taq Polymerase	Eppendorf, Hamburg
Trypsin	Promega, Madison

8. Chemikalien

Soweit im Einzelnen nicht anders erwähnt, wurden alle Chemikalien wie in nachfolgender Tabelle benannt bezogen.

Tabelle 5: Häufig genutzte Chemikalien

LB Medium	Roth, Karlsruhe
LB Agar	Roth, Karlsruhe
Sulfamerazine	Sigma Aldrich, Hamburg
Trizol Reagens	Invitrogen, Karlsruhe

9. Rechnerhardware

Alle Berechnungen wurden auf Hardware von Dell durchgeführt. Verwendet wurden Optiplex GX620, Precision T3400, Precision M6440, Poweredge 1950 und Poweredge 2950 Systeme. Durchgängig waren alle Systeme mit Pentium 4, Core 2 Quad und Centrino 2 Technologie von Intel ausgestattet.

10. Software

Alle verwendeten Systeme wurden mit den Betriebssystemen *Server 2003*, *XP Professional* und *Vista Professional* vom Microsoft betrieben. Darüber hinaus kamen für bioinformatische Fragestellungen die in Tabelle 6 benannten, kommerziellen oder frei verfügbaren Softwarepakete zum Einsatz:

Tabelle 6: Software

Name	Version	Beschreibung
Acuity http://www.moleculardevices.com/pages/software/gn_acuity.html	4.0	Auswertung von Microarraydaten
BioGPS http://biogps.gnf.org	Webbasiert, jeweils aktualisiert	Übergreifende, allgemeine Genannotationen
Cytoscape/BinGO http://www.cytoscape.org/	2.6.1/2.3	GO Annotierung und GO Analysen
Excalibur http://www.thermo.com/com/cda/product/detail/1,,1000001009250,00.html	2.0.5	Darstellung von Massenspektrometrischen Messdateien
GenePix http://www.moleculardevices.com/pages/software/gngenepixpro.html	6.1	Auswertung von Microarray Scans
Lasergene Paket http://www.dnastar.com/	7.0.0	Assemblierung und Visualisierung von EST Sequenzen
MASCOT/MASCOT Daemon http://www.matrixscience.com/	2.2.02	Identifikation von Peptiden über Massen und Fasta Datenbanken
MaxQuant http://www.maxquant.org/	13.09	Detektion und Zusammenfassung von SILAC markierten Peptiden/Peptidpaaren
Primer3 http://frodo.wi.mit.edu/primer3/	Webbasiert, jeweils aktualisiert	Primerdesign auf cDNA Sequenzen
Venny http://bioinfogp.cnb.csic.es/tools/venny/index.html	Webbasiert, jeweils aktualisiert	Venn Diagramme auf Basis beliebiger IDs

B. Methoden

Das Kapitel Methoden ist in die Bereiche allgemeine Methoden und spezielle Methoden unterteilt. Im ersten Teil sind allgemeine molekularbiologische Methoden kurz beschrieben, die im Rahmen dieser Arbeit angewendet wurden. Im zweiten Teil sind die speziellen Methoden, die teilweise nicht auf generell zugängliche Methoden sondern auf Entwicklungen in der Arbeitsgruppe oder Eigenentwicklungen beruhen, ausführlicher dargestellt.

1. Allgemeine Methoden

a) *Sterilisation von Lösungen und Geräten*

Alle zur Verwendung gekommenen Labormaterialien wie Behälter, Probengefäße, Pipettenspitzen etc. als auch alle verwendeten Lösungen wurden sterilisiert. Dazu kam ein Autoklav der Firma Sauter (Sulgen) und Fedegari (Albuzzano) zum Einsatz (30 min, 2.2 bar, 134°C). Verwendetes Operationsbesteck wurde ebenfalls autoklaviert. Während der Nutzung erfolgte die Desinfektion zwischen den einzelnen Operationsschritten über 70% Ethanolbad und Abflammen des Instrumentes.

b) *Extraktion von Nukleinsäuren aus Geweben*

Zur Gewinnung der RNA wurden die Gewebeproben zuerst mechanisch in einer Kugelmühle der Firma Retsch, Haan zerkleinert. Die gleichzeitige Lyse der Zellen erfolgte durch das Reagenz *Trizol*. Aus dem Zellextrakt wurden durch Zentrifugation die unlöslichen Zelltrümmer entfernt. Um die Proteine aus dem gereinigten Zellextrakt zu fällen, wurde Chloroform zugesetzt. Die Nukleinsäuren blieben in der wässrigen Phase. Durch Zentrifugation konnten die beiden Phasen getrennt werden. Die Proteine sammelten sich in der Interphase. Die RNA befand sich in der oberen wässrigen Phase. Mit Isopropanol wurde die RNA nach einer Überführung in ein neues Gefäß gefällt und nach Waschen in 70% Ethanol in einer geeigneten Menge RNase freiem H₂O aufgenommen.

c) *DNA-Sequenzierung*

Die DNA Sequenzierung wurde auf Basis der zuvor erstellten cDNA Bibliothek durchgeführt. Als Ausgangsmaterial wurden Bakterienklone in 384 well Format verwendet, die den pDNR-Lib Vektor mit entsprechenden Molchtranskripten tragen. Selbige wurden maschinell in neue 384 well Platten umgesetzt und für die Sequenzierung in 190µl 2YT Medium und 30mg/l Chloramphenicol 16h bei 37°C und 1100rpm kultiviert. Die Plasmid DNA wurde nach einem Protokoll des MPI für Molekulare Genetik mit einem automatisierten Pipetiersystem aufgereinigt. Dabei werden die Bakterien alkalisch lysiert und die Plasmide in einem Polyethylen-Glycol-2-Propanol-Gemisch präzipitiert. Gewaschen wurde mit 70% Ethanol. Aus dieser Behandlung resultiert genügend Material für eine Sequenzierung. Die Sequenzierung erfolgte auf einem ABI 3730xl Kapillarsequencer von Applied Biosystems (Foster City), die nach der Sanger Ketten Abbruch Methode arbeiten.

d) *Amplifikation von DNA-Fragmenten mit Hilfe der Polymerase-Kettenreaktion (PCR)*

Die PCR wurde in einer Biorad (München) Tetrad Cyclyer mit vier 96 well PCR Köpfen durchgeführt. Auf Eis wurde folgender Vorlagepuffer erstellt und dann jeweils 100µl in jedes well der 96 well Platten vorgelegt.

7828µl H₂O bidest

980µl 10xPCR Puffer

392µl 50mM MgCl₂

200µl 10mM dNTP-Mix

200µl 10µM pDNR-Lib Primer Mix (Leserichtung und Revers)

200µl Taq-Polymerase (am MPI hergestellt)

In die Vorlage wurden 3µl der Bakterienkultur pipettiert.

Die PCR wurde mit folgenden Einstellungen für 100.000 Bakterienklone durchgeführt:

96°C 5min, (94°C 30sec, 57°C 45 sec, 72°C 150sec) x 37 Zyklen, 72°C 5min, 4°C

Material und Methoden

e) *Gelelektrophorese*

Die Gelelektrophorese wurde für alle PCR Produkte aus der cDNA Bibliothek durchgeführt. Verwendet wurden 400ml (384 Spuren) Agarosegele (1%) mit 4µl Ethidiumbromid. Jeweils für 4 96 well Platten wurden 10µl der PCR Reaktion mit 2µl DNA Ladepuffer vermischt und aufgetragen. Zur späteren farblichen Unterscheidung der Platten kamen die DNA Ladepuffer Xylene Cyanol und Bromophenol Blau zum Einsatz. Die Laufzeit wurde bei 120 Volt mit ~30 Minuten gewählt. Dokumentiert wurden die Gele mit einem Biometra (Göttingen) BioDocAnalyzer System.

2. Spezielle Methoden

a) *Konstruktion einer normalisierten cDNA-Bibliothek*

Für die cDNA Bibliothek wurden Herzen zu 13 Regenerationszeitpunkten (ungeschädigt, scheinoperiert, 30 min, 2h, 6h, 12h, 1d, 2d, 4d, 7d, 14d, 21d und 35d) verwendet. Aus diesen Pools wurde auch die aRNA erstellt, die für die spätere Array Hybridisierung verwendet wurde. Es wurden zu jedem der dreizehn Zeitpunkte jeweils vier Pools erstellt, die aus jeweils vier Herzen bestanden. Aus den jeweiligen Pools wurde die Gesamt-RNA nach der Trizol Methode isoliert. Für die cDNA-Synthese wurde ein Gemisch aus jeweils gleichen Mengen Gesamt-RNA aus den Einzelpools eingesetzt. Das Protokoll für die Erstellung doppelsträngiger cDNA erfolgte nach der SMART Technologie. Die anschließende Normalisierung erfolgte über die DSN-Normalisierungsmethode und wurde bei der Firma Evrogen (Moskau, Russland) durchgeführt. Zunächst wurde für die Erststrangsynthese eine MMLV-basierte Reverse Transkription durchgeführt. Diese Reverse Transkriptase synthetisiert vom 3' Ende beginnend unter Einsatz eines Poly dT Primers mit 3' Zusatz den Doppelstrang. Der Zusatz des poly dT Primers trägt die Annealing Sequenz für die spätere PCR. Bei Erreichen des 5' Endes der mRNA fügt die Transkriptase noch eine freistehende kurze poly C Sequenz an. An diese freistehende poly C Sequenz hybridisiert der 3' Bereich eines sogenannten PlugOligos, der die Annealing Sequenz für die spätere PCR im 5' Bereich trägt (siehe Abbildung 10 A). Damit ist die Erststrangsynthese abgeschlossen. Anschließend wird die Doppelsträngige cDNA durch PCR mit den endspezifischen Oligonukleotiden hergestellt, wodurch die Doppelstränge beidseitig ein definiertes Sequenzmotiv tragen. Die anschließende Normalisierung erfolgt durch die DSN Methode, indem zunächst die Doppelstränge durch Erhitzen getrennt werden. Im anschließenden

Abkühlungsschritt hybridisieren die komplementären Einzelstränge wieder. Je nach Grad der Abundanz eines Transkriptes finden Einzelstränge unterschiedlich schnell einen komplementären Partnerstrang. Durch Zugabe einer Doppelstrang spezifischen DNase (DSN aus der Kamchatka Krabbe) werden nur Doppelstränge abgebaut. Statistisch gesehen wurden somit alle hochabundanten Transkripte in ihrer Anzahl wesentlich reduziert. Die DNase Reaktion wurde mittels Inhibitor abgestoppt (siehe Abbildung 10 B). An dieser Stelle folgte die PCR mit spezifischen Primern, die an den zuvor eingebrachten 5' und 3' Adaptern binden und Schnittstellen für das Restriktionsenzym Sfi1 enthalten. Transkripte, die bei der Erststrangsynthese nicht bis zum 5' Ende synthetisiert wurden, werden während der PCR nicht amplifiziert.

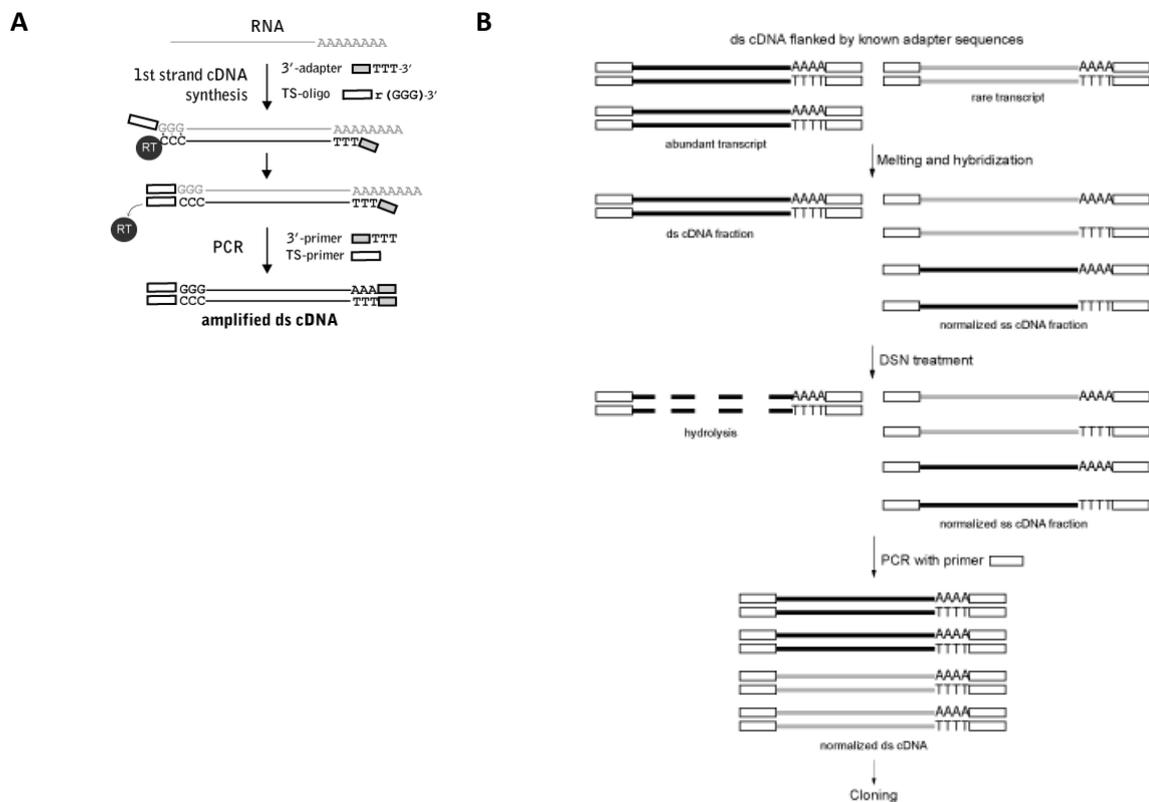


Abbildung 10: SMART und DSN Technologie

A Dargestellt ist die SMART Technologie zur selektiven Erststrangsynthese von mRNA und definierter Anbringung von Adaptersequenzen. **B** Normalisierung von cDNA Banken nach der Duplex-Specific nuclease (DSN) Technik. Die Bilder wurden von <http://www.evrogen.com/services> entnommen.

Die PCR Produkte wurden nach der Amplifizierung mit dem Restriktionsenzym Sfi1 mit der degenerierten Erkennungssequenz GGCCNNNNGGCC geschnitten und in das Plasmid pDNR-Lib,

Material und Methoden

welches passende Restriktionsschnittstellen für Sfi1 besitzt, in gerichteter Orientierung kloniert. Anschließend wurden die Vektoren durch Elektroporation in *E. coli* XL1-Blue transformiert. Der pDNR-Lib Vektor trägt eine Resistenz gegen das Antibiotikum Chloramphenicol. Zwei unabhängige Bibliotheken wurden erstellt, die laut Herstellerangaben aus mindestens 5×10^5 individuellen Klonen bestehen sollten.

Die originäre Bakterienkultur wurde 50µl/1,8ml LB Medium verdünnt und auf Agarplatten (24cm x 24cm Platte, 200ml Volumen, 12,5µg/ml Chloramphenicol) ausplattiert. Kultiviert wurde für 12h bei 37°C. Danach wurden einzelne Bakterienkolonien in 96 well Platten gepickt. In der 96 well Platte wurde jeder Bakterienklon einen Tag bei 37°C in 100µl LB Medium unter Zusatz von 20µg/ml Chloramphenicol inkubiert. Um Verdunstung des Nährmediums zu vermeiden wurden die Platten mit Parafilm abgedeckt. Mittels PCR (3µl Bakterienkultur als Matrize) und Gelelektrophorese wurden die vereinzelt Klone auf das Tragen eines Inserts überprüft. Die PCR wurde im 96 well Format durchgeführt. Anschließend wurden jeweils vier 96 well Platten in 384 well Platten überführt und unter Zugabe von 20µl Freezing Medium in zweifacher Ausführung bei -80°C gelagert. Die zuvor entstandenen PCR Produkte wurden zunächst mittels Isopropanol Fällung mit anschließendem Waschschrift in 70% Ethanol präzipitiert, in einem entsprechenden Volumen Spottingpuffer aufgenommen und in 384 well Platten umgesetzt und bei -80°C bis zum Arrayspotting verwahrt.

b) *RNA Amplifizierung*

Für die spätere Hybridisierung wurde Gesamt-RNA aus den betrachteten Schädigungszeitpunkten des Herzen jeweils amplifiziert. Dazu wurde das *Message Amp II aRNA Amplifikation Kit* (Ambion, Ca. No. AM1751) verwendet. Die Amplifikation wurde gemäß der Herstelleranleitung durchgeführt. Eingesetzt wurde jeweils 1µg Gesamt-RNA. Die Inkubationszeit wurde mit 14 Stunden gewählt. Die beschriebene optionale zweite Amplifizierungsrunde wurde nicht durchgeführt. Das Kit basiert auf der cDNA Synthese mithilfe von T7 Oligo(dT) Primern, am Ende der Amplifizierung liegt *antisense*-RNA (aRNA) vor.

c) *Herstellung und Fluoreszenzmarkierungen von cDNA-Proben für die Chiphybridisierungen*

Zur Markierung wurden Zufallsprimer verwendet. Dazu wurden 2µg amplifizierter mRNA und 4µl der Zufallsprimer Mixtur (Random hexamer mix, Promega, Madison) vermischt. Die Mischung wurde mit ddH₂O auf 15,4µl aufgefüllt und für 10 min auf 65°C erhitzt. Dann wurde das erstellte Gemisch jeweils zu folgendem Cy3/Cy5 Reaktionsgemisch hinzugefügt:

6µl	5x Erststrang Puffer
3µl	0,1M DTT
0,6µl	Unmarkiertes dNTP Gemisch
3µl	Cy3 oder Cy5 (25nM dCTP)
1µl	RNAsin (RNase Inhibitor)
2µl	Superscript III (Invitrogen, Karlsruhe, Reverse Transkriptase)

Unmarkiertes dNTP Gemisch:

5µl	dATP 100mM
5µl	dCTP 100mM
5µl	dGTP 100mM
5µl	dTTP 100mM

Das Gemisch wurde für 5 min bei 25°C inkubiert. Danach folgt eine Inkubation für 2h bei 50°C. Es folgte das Abstoppen der Reaktion durch Zugabe von 30µl 0,1N NaOH, 2mM EDTA und Erhitzung auf 65°C für 10min. Die Neutralisation erfolgte durch 30µl 0,1M HCl. Im nächsten Schritt wurden die beiden zu vergleichenden Cy3/Cy5 Proben auf eine Microcon YM-30 Säule (Millipore) aufgetragen. Diese wurde zuvor mit 260µl TE äquilibriert. Es folgte der erste Waschschrift durch Zentrifugation für 12min bei 14000g und 20°C. Im zweiten Waschschrift wurde zunächst der Durchfluss aus dem ersten Waschschrift entfernt und die Säule mit 450µl TE für 12min bei 14000g erneut zentrifugiert. Im dritten Waschschrift wurde der Durchfluss abermals entfernt und ein Gemisch aus 450µl TE, 20µl humaner Cot1 DNA (1mg/ml) und 2µl einer aus Hefe stammender PolyA RNA (10mg/ml) aufgetragen. Es wurde abermals bei 14000g und 20°C für 12min zentrifugiert. Die Microcon wurde auf ein neues Gefäß gewechselt und durch Zentrifugation herausgelöst. Durch Zugabe von 30µl DEPC H₂O (1 Minute Inkubation) wurde die Probe eluiert. Das Gesamtvolumen wurde in einer Speed Vac auf 5µl reduziert. Es wurden 160µl Nexterion Hybridisierungspuffer hinzugefügt (Peqlab 39-1066075). Anschließend

Material und Methoden

wurden die gelösten Proben für 5 min bei 95°C denaturiert und zum Entfernen nicht gelöster Partikel bei 14000g für 15 min zentrifugiert.

d) *Hybridisierung der cDNA-Microarrays mit fluoreszenzmarkierten cDNA-Proben*

Die Microarrays wurden mit Nexterion Block E (Schott, Mainz, Cat. No. 1066069) für 15min bei 50°C geblockt. Danach wurden sie mit destilliertem H₂O gewaschen und bis zur vollständigen Trocknung zentrifugiert. 80µl des fluoreszenzmarkierten cDNA-Gemisches wurden auf die vorgewärmten Microarrays pipettiert und mit Deckgläschen abgedeckt. Die Microarrays wurden in eine wasserdichte Hybridisierungskammer eingelegt, die ein Pufferreservoir von 40µl Hybridisierungspuffer enthielten, um ein Austrocknen der hybridisierten Proben zu verhindern. Die Kammer wurde verschlossen und über Nacht unter Lichtausschluss bei 42°C im Wasserbad hybridisiert. Am nächsten Tag wurden die hybridisierten Microarrays gewaschen. Dies geschah in 200ml Gefäßen. Dazu wurde zunächst die Hybridisierungskammer komplett getrocknet und geöffnet. Die Deckgläschen wurden entfernt. Die nachfolgenden Waschschriffe geschahen unter Lichtauschluss, um ein Ausbleichen der hybridisierten Proben zu verhindern. Im ersten Schritt für 10min in einem 2xSSC, 0,1%SDS Gemisch gewaschen. Danach folgte für 10min eine Waschung in 2xSSC. Der dritte Waschritt erfolgte für 5min in 1xSSC, der letzte Waschschritt für 2min in 0,2xSSC. Danach wurden die Microarrays bei 100g für 5 min getrocknet. Anschließend wurden die Microarrays in lichtundurchlässigen Kammern aufbewahrt und innerhalb von 24 Stunden auf einem GenePix 4000B Scanner System analysiert.

e) *Aufschluss von Gewebe*

Komplexes Gewebe wurde nach Schockgefrierung in flüssigem Stickstoff zur weiteren Verarbeitung in geeignetem Lysepuffer aufgetaut und mittels Ultraturrax (IKA, T10 basic) auf höchster Stufe 2x30 sec. zerkleinert.

f) *Proteinkonzentrationsbestimmung*

Die Bestimmung der Proteinkonzentration erfolgte nach Bradford mit dem *Dc Protein Assay Kit* (Biorad München, Kat. Nr. 5000113, 5000114) und durch Absorptionsmessung mit einem *FluoStar Galaxy* (BMG Lab Technologies, Offenburg). Dabei wurden zu jeder Messung aus einer Eichlösung BSA 2µg/µl drei Eichkurven erstellt. Jede zu messende einzelne Probe wurde dreimal angesetzt. Nach Messung der Absorption wurden die Proben gemittelt.

g) *Auftrennung von Proteinen im Gel, Fixierung und Färbung*

Zur Auftrennung der isolierten Proteinfractionen wurden Gradientengele verwendet (Novex Bis-Tris Gel Systeme, Invitrogen Karlsruhe, 4-12% NuPAGE, 1mmX10well). Zum Auftragen ins Gel kam 4xLP/DTT Laufpuffer zum Einsatz. Die Geltaschen wurden mit 40µl Proteinlösung geladen, die Proben danach bei 70V für 15min in der Geltasche gesammelt, abschließend bei 120V über ~3h aufgetrennt.

Nach der Auftrennung erfolgte die Fixierung der Proteine im Gel. Dazu wurde das Gel 10min mit einer Fixierlösung (25ml H₂O, 20ml Methanol, 5ml Essigsäure) überdeckt. Anschließend erfolgte die Färbung über Nacht mit einem Colloidal Blue Commassie Staining Kit (Novex, 27.5ml H₂O, 10ml Methanol, 10ml Novex Colloidal Blue Stain Solution A, 2.5ml Novex Colloidal Blue Staining Solution B). Am nächsten Tag erfolgte die Entfärbung durch zweistündiges Waschen in MilliQ Wasser und dreimaligem Wasserwechsel.

h) *Proteinverdau im Gel und Aufbereitung*

Das gefärbte Gel wurde fotografiert und mit einer Schnittschablone versehen. Das Gel wurde nach Schnittschablone in entsprechende Gelscheiben zerschnitten, die jeweils zu einer Spur entsprechende Banden enthielten. Diese Gelscheiben wurden in geeignete Reaktionsgefäße vereinzelt und weiterbehandelt:

- Manuelle Zerkleinerung der Gelscheibe in 1-2 mm³ Würfel
- Überschichtung mit 50mM Ammonium Bikarbonat (ABC) / 50% Ethanol
- Inkubation 20min bei Raumtemperatur
- Überstand abnehmen und werfen
- Wiederholung der vorherigen drei Schritte

Material und Methoden

- Überschichtung mit 100% abs. Ethanol
- Inkubation 20min bei Raumtemperatur
- Überstand abnehmen und verwerfen
- Eintrocknen der Gelstücke mittels Vakuum Zentrifuge
- Überschichten mit 10mM DTT in 50mM ABC
- Inkubation 40min bei 56°C
- Überstand abnehmen und verwerfen
- Überschichten mit 55mM Iodacetamid in 50mM ABC
- Inkubation 30min bei Raumtemperatur, Dunkelheit
- Überstand abnehmen und verwerfen
- Überschichtung mit 50mM ABC
- Inkubation 10min bei Raumtemperatur
- Überstand abnehmen und verwerfen
- Überschichtung mit 100% Ethanol
- Inkubation 10min bei Raumtemperatur
- Überstand abnehmen und verwerfen
- Überschichtung mit 50mM ABC
- Inkubation 10min bei Raumtemperatur
- Überstand abnehmen und verwerfen
- Überschichtung mit 100% Ethanol
- Inkubation 10min bei Raumtemperatur
- Überstand abnehmen und verwerfen
- Wiederholung der vorherigen drei Schritte
- Eintrocknen mit Vakuum Zentrifuge
- Zusatz von 12µg Protease Lys-C, gelöst in 50mM ABC
- Inkubation 15min
- Überschichtung mit 50mM ABC
- Inkubation über Nacht bei 37°C
- Überstand abnehmen und sammeln
- Überschichtung mit 30% Acetonitril / 3% Trifluoressigsäure (TFA)
- Inkubation 30min bei Raumtemperatur
- Überstand abnehmen und sammeln
- Überschichtung mit 70% Acetonitril
- Inkubation 30min bei 22°C
- Wiederholung der letzten drei Schritte
- Überstand abnehmen und sammeln
- Überschichtung mit 100% Acetonitril (ACN)
- Inkubation 30min bei Raumtemperatur
- Wiederholung der letzten drei Schritte
- Überstand abnehmen und sammeln
- Einengen des Gesamtvolumen auf 80µl mittels Vakuum Zentrifuge
- Aquilibrierung der Pipettenspitze mit C18 Material

- Proteinprobe 1:1 mit 5%ACN / 1%TFA mischen und vortexen
- Proteinprobe in Pipettenspitze füllen
- Abzentrifugieren mit 2600 rpm, 4min
- Spitze waschen mit 5% ACN
- Abzentrifugieren mit 2600 rpm, 4min
- Aufbewahrung der Pipettenspitze bis zur Messung bei 4°C

i) *Filter assoziierte Proben Präparation (FASP)*

Zur Probenpräparation wurden insgesamt mindestens 200µg Protein benötigt. Zunächst wurden 30µl des Zelllysates mit 200µl UA vermischt und auf einen 30k Filter (Millipore Billerica, Microcon Centrifugal Filter Devices, Ultracel YM-30) aufgetragen. Der Filter wurde für 15 min. bei 14000g abzentrifugiert. Dieser Schritt wurde mit weiteren 20µl UA wiederholt. Anschließend wurden 100µl IAA Lösung auf den Filter aufgebracht und bei 600 rpm im Thermomixer (95°C) für 1 min erhitzt, dann im Dunkeln für 20 min inkubiert. Anschließend wurde bei 14000 g für 10 min abzentrifugiert. Es wurden 100µl UB auf den Filter aufgetragen und bei 14000 g 15 min abzentrifugiert. Dieser Schritt wurde zweimal wiederholt. Dann wurden 60 µl ABC, 20µl UB und 2µl LysC auf den Filter aufgetragen und bei 22°C für 4 h inkubiert. Dann wurden weitere 2µl LysC hinzugefügt und über Nacht bei 22°C inkubiert. Am nächsten Tag wurde der Verdau durch den Filter bei 14000 g in ein neues Gefäß zentrifugiert. Anschließend wurde dreimal mit 50µl 0.5 M NaCl bei 14000 g für jeweils 10 min gewaschen. Abschließend wurde der Filter für die anschließende Auftrennung (OffGel oder Gel) entsalzt.

j) *OffGel Probenpräparation*

Bei der OffGel Elektrophorese (Agilent Technologies, Santa Clara, 3100 OffGel Fractionator Kit) werden die Peptide nach isoelektrischem Punkt getrennt. Die Trennung erfolgt in einem Kammer getrennten pH Gradientengel, wobei die Peptide in Lösung gehalten werden. Nutzung der OffGel Apparatur nach Herstelleranleitung.

k) *HPLC und Massenspektrometrie*

Eine Agilent 1200 nanoflow Flüssigchromatograph (LC) System (Agilent Technologies, Santa Clara) wurde über eine Nanoelektrospray Quelle (Proxeon Biosystems, Odense) mit einem LTQ

Material und Methoden

Orbitrap XL Massenspektrometer-Instrument (Thermo Fisher Scientific, Waltham) verbunden. Die Chromatographische Auftrennung der Peptide erfolgte auf *reverse-phase* Säulen (ReproSil-Pur C18-AQ 3 µm resin (Dr. Maisch GmbH, Ammerbuch-Entringen)). LysC verdaute Protein Proben wurden bei einer Flussrate von 0.5 µl/min und einem linearen Gradienten bei einer Flußrate von 0.2 µl/min aufgetrennt. Die Aufnahmen der LTQ-FT Full Scan MS Spektren (m/z Verhältnis 350 bis 1750) geschahen mit einer Auflösung von R=60000 bei m/z=400. Für jeden Full Scan wurden die 5 Ionen der stärksten Intensität isoliert und in der linearen Ionenfalle durch Kollision (CID/HCD Methode) fragmentiert (MS²)

l) Analyse der LC-MS/MS Daten

Die Rohdaten wurden mithilfe der Software MaxQuant [77] in ein MASCOT konformes Format umgewandelt. Die MASCOT Version 2.2.02 wurde für die Datenbanksuche und Proteinidentifikation verwendet. Als Suchparameter wurde für alle MASCOT Suchen LysC Verdau und Carbamidomethylierung des Cysteins als feste Modifikation eingestellt sowie Oxidation des Methionin als variable Modifikation gewählt. Die maximal erlaubte Massenabweichung für MS und MS² Scans wurde jeweils mit 10ppm und 0.5 Da eingestellt. Für Peptididentifikationen wurden neben eigenen Molch EST Banken die *International Protein Index* (IPI, <http://www.ebi.ac.uk/IPI/IPIhelp.html>) Banken zu Maus, Mensch und Zebrafisch (jeweils in Version 3.37) und die *National Center for Biotechnology Information* (NCBI) Proteinsequenz Banken zu *Xenopus*, *Ambystoma* und *N. viridescens* eingesetzt. Alle Banken wurden als DECOY Ziel Proteinbanken angelegt [78]. Eine minimale Peptidlänge von 6 Aminosäuren und das Vorhandensein von mindestens 2 Peptiden, wovon ein Peptid als beschreibendes (unique) Peptid vorliegen muss, wurden als Qualitätsparameter für Proteinidentifikationen zu Grunde gelegt. Die Falsch-Positiv-Rate wurde auf Basis der reversen Sequenz-Treffer aus den DECOY Banken bestimmt und mit <1% festgelegt.

m) Protein Klassifizierung mit Gene Ontology

Vorliegende assemblierte Contig Sequenzen wurden soweit möglich mit Gene Ontology (GO, <http://www.geneontology.org/>) Termen annotiert. Dazu wurden von der Datenbank Ensembl (<http://www.ensembl.org/index.html>) zu den Organismen *Danio rerio* (DR), *Bos taurus*, *Homo*

sapiens, *Mus musculus*, *Xenopus tropicalis* und *Gallus gallus* cDNA Sequenzen heruntergeladen, die sowohl eine Ensembl Identifikation als auch mindestens eine GO Annotation haben. Die Selektion erfolgte über das Tool Biomart (www.biomart.org). Die zu jedem Organismus vorliegende Sequenzdatenbank im Fasta-Format wurde jeweils mittels BLAST nach hoher Homologie zu den Molchsequenzen durchsucht (E-value Cutoff bei e^{-20}). Zu jedem Molch Contig wurde nur der beste Treffer in der jeweiligen Organismusdatenbank verwendet.

n) *Haltungsbedingungen der Molche*

1 Generelles

Alle verwendeten Tiere werden aus den Vereinigten Staaten, *Sullivan newt farm*, Tennessee, Nashville bezogen. Alle verwendeten Tiere sind 3-4 Jahre alt. Die Tiere werden in Gruppen von bis zu 15 Tieren in einer Aquariumsanlage mit Landgang am MPI gehalten. Die konstante Wassertemperatur liegt bei 18-20°C. Die Fütterung adulter Tiere geschieht im Drei-Tage-Rhythmus mit roten Fliegenlarven oder Artemien.

2 Fütterung mit Mausleber

Die Fütterung mit Mausleber von voll markierten Lysin-6 Mäusen dient dem Einbau dieser *schweren* Aminosäure in neu synthetisierte Proteine des Molches. Dazu werden Tiere aus der Aquariumsanlage in Einzelbecken umgesetzt und für einen Zeitraum von vier Wochen mit einer reinen Mausleber Diät von *leichten* Mäusen an das neue Futter gewöhnt. Nach diesem Zeitraum wird kontrolliert mit zwei Gramm im Drei-Tage-Rhythmus schwere Mausleber gefüttert. Die komplette Aufnahme des Futters wurde kontrolliert. Die Gesamtfütterungszeit mit schwerer Mausleber betrug 60 Tage.

3 Narkose

Vor sämtlichen operativen Eingriffen wurden die Molche in einer Lösung mit Betäubungsmittel (Sigma Aldrich, Hamburg), 0,1g/100ml) anästhesiert. Die Molche stellen bei dieser Betäubung nach fünf Minuten jegliche Bewegungen ein. Die Betäubungszeit wurde mit 20 min gewählt.

Material und Methoden

4 Wunddesinfektion

Vor und nach operativen Eingriffen am Schwanz oder Herz des Molches wurden die Molche für mindestens drei Stunden in einem Desinfektionsbad (Sulfamerazine, Sigma Aldrich, Hamburg) (5g/l) gehalten.

o) *Operative Eingriffe am Molch*

Eingesetztes Operationsbesteck: Gerade Pinzette, gebogene Pinzette, Dekapsulationsschere, Feinschere, Wattestäbchen, Cellulosetücher, Desinfektion nach jeder Verwendung durch Abflammen und eintauchen in 70% Ethanol, Ablage auf steriler Unterlage.

Schwanzamputation: Die narkotisierten Molche werden aus dem Betäubungsbad genommen, mit Leitungswasser dreifach ab gespült und auf einem mit Alkohol getränktem Tuch platziert. Mittels Feinschere wird ein 1 cm großes Stück der Schwanzspitze amputiert. Operierte Molche werden für zehn Minuten auf dem Tuch belassen. In dieser Zeit verschließt ein sich bildender Blutpfropfen die Wundstelle. Danach werden die Molche für drei Stunden ins Desinfektionsbad gesetzt. Die amputierte Gewebeprobe wird sofort in flüssigen Stickstoff überführt.

Herzschädigung: Die narkotisierten Molche werden aus dem Betäubungsbad genommen, mit Leitungswasser dreifach ab gespült und auf einem mit Alkohol getränktem Tuch rückwärtig platziert. Der Eingriff erfolgt von ventral oberhalb der vorderen Extremitätenansätze. Die Einschnittstelle wird mit 70% Ethanol desinfiziert. Mittels Feinschere wird die Oberhaut und die unterliegende Silberhaut im Bereich des *Truncus arteriosus* durchtrennt. Mit einer gebogenen Pinzette wird das Herz am *Truncus arteriosus* fixiert und mit einer geraden Pinzette durch zweimal 15faches leichtes Quetschen längs und quer des Ventrikels geschädigt. Die Einschnittstelle wird sterilem Wattestab abgetrocknet und mit Gewebekleber (B Braun, Melsungen, Histoacryl) verschlossen. Operierte Molche werden für 10 min auf dem Tuch belassen. Danach werden die Molche für drei Stunden ins Desinfektionsbad gesetzt.

Herzentnahme: Die narkotisierten Molche werden aus dem Betäubungsbad genommen, mit Leitungswasser dreifach ab gespült und auf einem mit Alkohol getränktem Tuch rückwärtig platziert. Der Eingriff erfolgt von ventral oberhalb der vorderen Extremitätenansätze. Die

Einschnittstelle wird mit 70% Ethanol desinfiziert. Mittels Feinschere wird die Oberhaut und die unterliegende Silberhaut im Bereich des *Truncus arteriosus* durchtrennt. Mittels Dekaptionsschere erfolgt die Dekaption. Mit einer gebogenen Pinzette wird das Herz am *Truncus arteriosus* fixiert und am *Sinus venosus* mit der Feinschere abgetrennt. Das im Herzen verbliebene Blut wird mit einem sterilen Wattestab abgesaugt und das Herz sofort in flüssigen Stickstoff überführt. Gewebeproben werden bis zur Weiterverwendung bei -80°C gelagert.

VI. Ergebnisse

Der Ergebnisteil gliedert sich in die beiden in dieser Arbeit hauptsächlich bearbeiteten Teilgebiete der Transkriptom- und Proteomanalyse. Zunächst wird auf die Transkriptomanalyse eingegangen, die mithilfe eines in der Arbeitsgruppe selbst erstellten Microarrays durchgeführt wurde. Beispielhaft wird die frühe Wundreaktion näher beschrieben. Daran schließt sich der Schwerpunkt dieser Arbeit an, der sich mit der Methodenentwicklung und Durchführung einer Proteomanalyse in komplexen regenerierenden Geweben des Molches beschäftigt.

A. Transkriptomanalyse

Die Transkriptomanalyse stellt den Arbeitsgang ausgehend von einer Konzeption einer cDNA Bibliothek und eines cDNA Microarrays bis hin zur Durchführung und Auswertung einer kompletten Zeitreihe der Herzregeneration dar. Detailliert werden die frühen Zeitpunkte der Herzregeneration auf Transkriptomebene beschrieben. Diese frühen Zeitpunkte werden im Kapitel zur Proteomanalyse wieder aufgegriffen und dienen als Vergleich. Außerdem wird die gezielt ausgewählte Sequenzierung und Strategie zur Annotation der cDNA Bibliothek dargestellt. Abschließend werden Ergebnisse aus einem Sequenzierungsprojekt auf modernen Hochdurchsatz Sequenzierungsgeräten dargestellt.

1. Konzeption einer cDNA Bibliothek zur Herzregeneration

Für die Erstellung der cDNA Herzbibliothek wurden insgesamt 104 Molchherzen verwendet. Um einen möglichst vollständigen Überblick über alle im Herzgewebe vorkommenden, insbesondere regenerationspezifische, Transkripte zu erhalten, wurde die Bibliothek dabei nicht nur aus einem einzelnen Regenerationszeitpunkt hergestellt. Vorangegangene Versuche haben gezeigt, dass zu spezifischen Zeitpunkten der Regeneration viele Transkripte stark dereguliert werden [60]. Diese Deregulation reicht annähernd bis zum Abstellen der Transkription für einzelne Gene. Ebenso hat sich bei vorangegangenen Untersuchungen gezeigt, dass solche Transkripte mit niedriger Kopienzahl kaum oder gar nicht mehr in einer cDNA Bibliothek vorkommen [60]. Daher wurden für die Herstellung dieser cDNA Bank Herzen aus 13 Zeitpunkten verwendet (13

Zeitpunkte x 2 verwendete Pools x 4 Herzen = 104 Herzen, Zeitpunkte: ungeschädigt, scheinoperiert, 30min, 2h, 6h, 12h, 1d, 2d, 4d, 7d, 14d, 21d und 35d) und in einem Pool zusammengefasst. Aus diesem Pool wurde die Gesamt-RNA verwendet. Eine weitere Verbesserung hinsichtlich der Vollständigkeit der cDNA Bibliothek konnte dahingehend erreicht werden, dass die Bibliothek normalisiert wurde (Siehe Methoden V.B.2.a). Anschließend wurde die komplexe, normalisierte Bibliothek in Bakterien transfiziert, ausplattiert, kultiviert und vereinzelt. Eine Überprüfung der Redundanz innerhalb der normierten cDNA Bibliothek wurde auf zwei Arten durchgeführt. Zunächst wurden 149 Klone ausgewählt, die im Gel eine einheitliche Bandengröße aufzeigten. Diese Klone wurden sequenziert und die erhaltenen Sequenzen assembliert. Diese vorselektierte Auswahl ergab 78 unterschiedliche Contigs. In einem zweiten Test wurde eine komplette 384 *well* Platte sequenziert. Die Platte enthielt nach Assemblierung 382 unterschiedliche Contigs (beinhalten Singeltons und Contigs). Die qualitative Überprüfung der beiden vorangegangenen Assemblierungen wurde über die Annotation der einzelnen Contigs geführt. Ein Vergleich der Contig Annotationen zeigte durchgehend unterschiedliche Annotationen für sämtliche annotierbaren Contigs.

Nach dem Nachweis einer geringen Redundanz innerhalb der Bibliothek wurden die Kulturen ausplattiert und die Bakterienklone von Hand vereinzelt. Zu jedem Bakterienklon erfolgte mittels PCR die Amplifikation des enthaltenen Vektorinserts, welches anschließend für eine spätere Aufbringung auf den Glasträger eines Microarrays aufbereitet wurde (Siehe Methoden V.B.2.a). Dieser Vorgang erfolgte für insgesamt 100.000 Bakterienklone. Für ~98% der Klone konnte eine spezifische PCR Amplifikation anhand einer entsprechenden Gelbande gezeigt werden. Die Bakterienklone sowie die zugehörigen PCR Produkte wurden in zweifacher Kopie bei -80°C bis zur weiteren Verwendung eingelagert.

2. Arraydesign von cDNA Microarrays

Für das Spotting der Microarrays war ein Arraydesign zu erstellen, das eine spätere Auswertung der Microarrays unter verschiedenen Gesichtspunkten ermöglichen sollte. Dazu waren zunächst in jedem der 48 Blocks des Microarrays neben jeweils zwei Leerspots auch sogenannte *Housekeeping* Spots einzubringen. Diese werden von Normalisierungsalgorithmen als konstante Werte für unveränderte Expression verwendet. Dazu werden Transkripte verwendet, die in nahezu allen Zellen mit hoher Kopienzahl hergestellt werden. Ein klassisches *Housekeeping*-Gen auf Microarrays ist das Aktin Gen. Untersuchungen haben jedoch gezeigt, dass bei Regeneration

Ergebnisse

am Molchbein beta-Aktin als Normalisierungsfaktor ungeeignet ist, da es während der Regeneration stark dereguliert exprimiert wird [79]. Daher wurde GAPDH als Normalisierungsfaktor gewählt [79]. In der Arbeitsgruppe am MPI durchgeführte Untersuchungen haben jedoch gezeigt, dass auch GAPDH zu manchen Regenerationszeitpunkten eine deregulierte Transkription aufweist. Daher wurde auf eine hauseigene Transkriptionsanalyse (Borchardt, Looso, Braun, in Vorbereitung) zurückgegriffen und zwei Sequenzen mit der Bezeichnung Contig 167 und Contig 242 als *Housekeeping* Gene gewählt. Die beiden Sequenzen wurden in jeden der 48 Blocks des Arrays mit fünf Kopien über die Fläche des Blocks verteilt gespottet. Dabei kodiert Contig 167 für *expressed in non-metastatic cells 2 protein*, und Contig 242 für das *ribosomale Protein S21 (RPS21)*. Beide Sequenzen zeigten während der Herzregeneration eine konstant hohe Intensität, waren also zu keinem Zeitpunkt dereguliert.

In jedem Block der Microarrays wurden zehn weitere Spots untergebracht, die einen Teil der Vektorsequenz des pDNR-Lib Klonierungsvektors tragen. Diese Vektorsequenzen dienen der Berechnung des Hintergrundsignals (Rauschen) für Signale niedriger Intensität (Spiked in Controls). Dabei wurde die Annahme zu Grunde gelegt, dass Transkripte mit niedriger Kopienzahl auch nach der Markierung mit Cy3/Cy5 Fluoreszenzfarbstoffen nur mit einer geringen Kopienanzahl vorliegen und nach der Hybridisierung auf dem Microarray ein schwaches Signal ergeben. Die Kernfrage in diesem Zusammenhang ist, insbesondere in Hinsicht auf die Quantifizierung, die Unterscheidung niedriger robuster Signale vor dem Hintergrund. Es hat sich gezeigt, dass sich Vektorsequenzen auf Microarrays besonders gut dazu eignen, Unterscheidungen zwischen realen Signalen und Hintergrundrauschen bzw. unspezifischen Bindungen zu treffen. Dazu wird eine Häufigkeitsfunktion der Intensitäten zu den bekannten Vektorspots berechnet. Der Schwellenwert wird dann dynamisch für jede Wellenlänge eines jeden Arrays gewählt, unter dem 95% der Intensitäten aller bekannten Vektor Spots liegen. Dieser Schwellenwert zur Unterscheidung von realen Signalen zu Hintergrundsignalen liefert vergleichbare Schwellenwerte zu allgemein angewendeten Algorithmen wie der *BSD* Methode oder der *Accumulated Distribution* Methode [80]. Insgesamt konnte gezeigt werden, dass einige Spots geringer Intensität in die Auswertung mit aufgenommen werden können, weil sie ein reales Signal zeigen, welches bei Standardalgorithmen aus der Auswertung herausgefallen wäre [81].

3. Expressionsprofil der Herzregeneration

Die hybridisierten cDNA Microarrays wurden elektronisch eingelesen und die Intensitäten der Spots berechnet. Zur weiteren systematischen Verarbeitung wurden die Intensitäten in einer Datenbank abgelegt. Zu den Microarrays der zehn ausgewählt untersuchten Zeitpunkte (2h, 6h, 1d, 2d, 4d, 7d, 14d, 21d, 35d und 1d scheinoperiert) wurde jeweils eine Blockweise Lowess Normalisierung [82] auf Basis der mittleren Spotintensität abzüglich der zirkulär ermittelten mittleren Hintergrundintensität durchgeführt. Alle Spots der verwendeten Microarrays wurden auf Intensität, Konformität und Identität überprüft. Dazu wurden die Parameter wie folgt gewählt:

- Gültige Spots dürfen in beiden gemessenen Wellenlängen nur in maximal 3% aller zugehörigen Pixel eine Helligkeitssättigung vorweisen
- Eventuell automatisch gesetzte Marker müssen den Spot als gültig identifizieren
- Das Verhältnis Signal zu Hintergrund muss mindestens in einer der beiden gemessenen Wellenlängen größer/gleich drei sein

Für alle folgenden Rechenoperationen werden nur Spots weiterverwendet, die diese Anforderungen erfüllen. Zu jedem experimentellem Zeitpunkt wurden vier biologischen Replikate (siehe V.B.2.a) miteinander verrechnet. Außerdem wurden zu jedem Pool die Ratios der Dye-Swap Experimente (2x4) erstellt. Mit diesem Ansatz können einseitige Effekte der detektierten Spot Ratios aufgrund von Unterschieden in der Messung der beiden verwendeten Wellenlängen und Farbstoffen ausgeschlossen werden. Zu den untersuchten zehn Zeitpunkten der Herzregeneration liegen somit jeweils acht Microarrays zur Verrechnung als Replikate vor, jeweils 2 Arrays (normal und Dye-Swap) pro biologischem Replikat. Zu jedem Experiment wurden die Spots als signifikant dereguliert angesehen, die ein zur Basis 2 logarithmiertes Verhältnis >1 in 6/8 Microarrays aufzeigen und einen nach Standard Benjamini-Hochberg [83] korrigiertem p-Wert < 0.05 (One Sample t-Test) haben. In Tabelle 7 wurden zu allen Microarrays die zusammengefassten Ergebnisse dargestellt.

Ergebnisse

Tabelle 7: Deregulierte Microarray Spots im Verlauf der Herzregeneration

Dargestellt sind die Gesamtanzahl der deregulierten Spots zu den zehn Regenerationszeitpunkten des Herzens (2 Stunden bis 35 Tage). Außerdem ist ein Kontrollzeitpunkt Scheinoperation (sham) aufgenommen. Die dargestellten Werte beziehen sich auf die absolute Anzahl der deregulierten Spots nach im Text beschriebenen Qualitätskriterien. Die Spalten 5 und 6 zeigen die Anzahlen für ein Verhältnis (Betrag) der Intensitäten geschädigtes-/ungeschädigtes Gewebe größer 1. Die Spalten 7 und 8 zeigen die Anzahlen der Spots für einen Betrag des Verhältnisses größer 1,5. Die Spalte 3 (Teil) bezieht sich auf den ersten und zweiten Teil des Microarrays, die jeweils 50.000 Spots tragen. Rot markiert sind die Spotanzahlen des scheinoperierten Zustandes sowie der Regenerationszustände 2 Stunden und 6 Stunden, die im Text näher analysiert werden.

Zeitpunkt	P Wert	Teil	Spots insgesamt	R>1 6/8	R<-1 6/8	R>1.5 6/8	R<-1.5 6/8
2 Stunden	<0.05	1	35371	389	78	183	6
6 Stunden	<0.05	1	25249	31	10	6	0
1 Tag	<0.05	1	35075	215	27	33	5
2 Tage	<0.05	1	25427	107	25	26	0
4 Tage	<0.05	1	31715	232	64	63	2
7 Tage	<0.05	1	23868	258	119	80	0
14 Tage	<0.05	1	37219	315	26	69	3
21 Tage	<0.05	1	30584	68	4	14	0
35 Tage	<0.05	1	25054	102	15	13	0
Sham	<0.05	1	27277	58	12	0	3
2 Stunden	<0.05	2	26874	325	43	101	1
6 Stunden	<0.05	2	25252	23	7	1	0
1 Tag	<0.05	2	24129	75	5	9	1
2 Tage	<0.05	2	24506	119	36	36	4
4 Tage	<0.05	2	26524	130	37	22	4
7 Tage	<0.05	2	20328	105	67	32	2
14 Tage	<0.05	2	28225	239	17	78	0
21 Tage	<0.05	2	28189	79	1	18	0
35 Tage	<0.05	2	23278	66	8	6	0
sham	<0.05	2	24402	14	13	0	6

a) Exemplarische Auswertung der Zeitpunkte ScheinOP, 2 Stunden und 6 Stunden Regeneration

Da die Gesamtmenge der gesammelten Daten sehr umfangreich ist, soll aus praktischen Gründen an dieser Stelle exemplarisch anhand der Zeitpunkte Scheinoperation, zwei stündige- und sechsstündige Regeneration der Grad des molekularbiologischen Einblickes gezeigt werden, der mit einem derartigen Hochdurchsatz-Ansatz erreicht werden kann. Insbesondere kann der Vergleich zwischen scheinoperierten Tieren und den frühen Regenerationszeitpunkten zwei

Stunden und sechs Stunden helfen, zwischen einer allgemeinen Wundantwort und inflammatorischen Reaktion (primär im Blut zu detektieren) und einer frühen, die Regeneration einleitenden Antwort zu unterscheiden. Zu Zwecken der Visualisierung des Expressionsverlaufes einzelner Kandidaten wurden unter anderem SOM Cluster (*self organizing maps*) erzeugt.

1 Scheinoperierte Herzen

Der Datensatz der scheinoperierten Herzen zeigt insgesamt im Vergleich zu allen anderen Zeitpunkten die geringste Anzahl an signifikant deregulierten Spots. Die Zahlen in Klammern geben im Folgenden jeweils die Anzahl aller Transkripte inklusive ribosomaler Sequenzen an. Diese wurden aus der Gesamtbetrachtung herausgenommen. So sind auf dem Gesamtarray (Teil 1 und Teil 2 gemeinsam) insgesamt 86 (97) Spots dereguliert, davon 64 (72) Spots mit Ratio > 1 in 6/8 Microarrays, p Wert <0.05 aufreguliert und 22 (25) Spots mit Ratio <-1 in 6/8 Microarrays und p Wert <0.05 abreguliert. Insgesamt bilden die 86 (97) Spots 46 unterschiedliche Contigs ab. Die 64 (72) aufregulierten Spots gehören nach Assemblierung zu 31 unterschiedlichen Contigs. Die 22 (25) abregulierten Spots gehören nach Assemblierung zu 15 unterschiedlichen Contigs.

Zu denselben 46 Contigs konnten 25 mit einer signifikanten Sequenzähnlichkeit annotiert werden. Von diesen 25 Contigs zeigten jedoch 10 Contigs ausschließlich eine Homologie in den NCBI EST Banken, die jeweils nur sehr unvollständig annotiert sind. Diese Zuordnungen belegen allerdings, dass es sich um transkribierte Sequenzen handelt, die auch in anderen Organismen vorkommen. Die EST Bank Zuordnungen kommen aus den Organismen *Xenopus*, *Ambystoma mexicanum*, *Ambystoma tigrinum* und *Cynops pyrrhogaster*. Die übrigen 21 zeigten keinerlei Sequenzähnlichkeit.

Es finden sich zu den genannten 46 deregulierten Contigs zum scheinoperiertem Zeitpunkt in Bezug auf das Gesamtexperiment (neun weitere Zeitpunkte) sechs auffällige Expressionsmuster. Diese werden im Folgenden, soweit annotiertbar, unter Angabe des getroffenen Organismus, des zugehörigen Datenbank Identifiers und des E-values angegeben. So ist zunächst eine Gruppe an Contigs auffällig, die nahezu gleichbleibend über alle Zeitpunkte stark ab- oder aufreguliert ist. Dies sind im einzelnen bei den Aufregulierten *Tissue factor pathway inhibitor 2* (*Rattus norvegicus*, emb|CAD22046.1, 6e-049), *Transmembrane 4 I6 family member, tm4sf1* (*Homo sapiens*, P30408, 3e-076), *Heat shock protein 25* (*Taeniopygia guttata*, gb|ACH45320.1, 4e-029), *Myosin heavy chain 3* (*Canis lupus familiaris*, ref|NP_001107184.1, ~0), *Potassium voltage gated channel, isk related family, member 4* (*Oryctolagus cuniculus*, ref|NP_001082783.1, 2e-042) und

Ergebnisse

Actin alpha cardiac muscle 1 (*Xenopus tropicalis*, Q6P640, 1e-126). Bei den Abregulierten konnten *Complement factor B* (*Rattus norvegicus*, ref|NP_997631.2, 2e-061), *Complement component 1 q subcomp.* (*Rattus norvegicus*, ref|NP_001008524.1, 5e-030) und *Ribonucleotide reductase M2* (*Xenopus tropicalis*, ref|NP_001007890.1, ~0) annotiert werden. Dargestellt ist dieser Cluster in Abbildung 11 A. Wie oben erwähnt konnten einige Contigs nicht über Sequenzhomologie annotiert werden. Unter diesen ist auch ein abreguliertes Contig, welches insgesamt die meisten ESTs (4 ESTs, Contig 5116, DB Id 5085) beinhaltet.

Neben diesem ersten Cluster 1 finden sich vier Contigs, die zwar bei scheinoperierten Herzen aufreguliert sind, im zeitlichen Verlauf aber bei zwei Stunden nach Herzschädigung eine wesentlich stärkere Aufregulation zeigen (mittlerer Fold Change (FC) = 2.1, Cluster 2) und zu diesem extrem frühen Zeitpunkt ihre maximale Expression erreichen (Peak). Im weiteren zeitlichen Verlauf sinkt deren Expressionsniveau dann kontinuierlich ab. Nach zwei Tagen ist das mittlere Expressionsniveau bereits um mehr als die Hälfte gesunken (mittlerer FC=0,7), und erreicht nach 14 Tagen nahezu Normalniveau (mittlerer FC= 0,05). Diese Contigs gehören folglich zu der frühen Antwort auf Stress- und Schädigungsereignisse, die bislang im Zusammenhang mit Regenerationsabläufen nicht in Verbindung gebracht wurden, da solch frühe Stadien bislang nicht untersucht wurden (Abbildung 11 B). In diesem Cluster 2 findet sich zu zwei Contigs keinerlei Sequenzähnlichkeit. Dazu gehört auch das Contig mit der maximalen Anzahl an ESTs (26, Contig 4000/DBid 5591). Zu einem Contig findet sich eine homologe nicht annotierte Sequenz im Frosch (*Xenopus laevis*, gb|AAI08888.1, 8e-085). Einzig das Protein *Translocase of outer mitochondrial membrane 22* (Tomm22) kann verlässlich annotiert werden (*Xenopus tropicalis*, gb|AAI60473.1, 8e-22).

Ein weiteres Cluster 3 aus sechs Contigs zeigt im Verlauf der Regeneration ähnlich dem vorhergehenden Cluster 2 eine massive Aufregulierung. Jedoch ist diese nicht nur zu einem Zeitpunkt zu beobachten, sondern erstreckt sich über eine längere Zeitperiode. Die Contigs im Cluster 3 erfahren zum Zeitpunkt ein Tag nach Schädigung ein massive Aufregulation. Diese hohe Expression wird bis zum Regenerationszeitpunkt 14 Tage nahezu unverändert beibehalten und sinkt dann rapide auf das normale Niveau ab. In dieser Gruppe konnte nur ein Contig über Homologiesuche annotiert werden. Es handelt sich um den *Tissue inhibitor of metalloproteinase 1* (*N.viridescens*, gb|ABB88702.1, 1e-125). Dargestellt ist dieser Cluster 3 in Abbildung 11 C.

Das Cluster 4, bestehend aus zwei Contigs, zeigt ebenfalls einen Expressionsverlauf, der nur zu einem Zeitpunkt einen deutlichen Peak der Expression zeigt. Diese Transkripte konnten als *Tissue factor pathway inhibitor 2* (*Bos taurus*, ref|NP_877589.1, 6e-49) und *AMBP Protein*

(*Rattus norvegicus*, sp|Q64240, 2e-022) annotiert werden und weisen im beobachteten Verlauf ein mittleres Expressionsmaximum von $FC=2,1$ bei einem Tag nach Schädigung auf. Dargestellt ist dieser Cluster 4 in Abbildung 11 D.

Das Cluster 5 enthält vier Contigs, von denen ein Contig mit *Tropoelastin 1* (*Xenopus tropicalis*, SP_TREM:Q0Q5Z2_XENTR, 4e-76) annotiert werden konnte. Zwei weitere zeigen eine Homologie zu nicht annotierten EST Sequenzen des *Ambystoma*, ein weiteres Contig keinerlei Homologie. Die Spots dieses Cluster 5 sind im scheinoperiertem Gewebe aufreguliert, zeigen aber dann im Verlauf der Regeneration zwischen zwei und vier Tagen eine massive Abregulierung. Nach 14 Tagen ist die Transkription wieder stark aufreguliert um zu späten Zeitpunkten hin kontinuierlich abzunehmen. Dargestellt ist dieser Cluster 5 in Abbildung 11 E. *Tropoelastin 1* ist ein lösliches Protein, welches für den Bau von Elastin benötigt wird. Elastin gehört zur Extrazellulären Matrix und ist primär in Gefäßen (Arterien) und anderen Gebilden, die nach Kontraktion oder Dilatation in ihre ursprüngliche Form zurückgeführt werden müssen, zu finden.

Neben Cluster 5 findet sich auch noch ein umgekehrt exprimierter Cluster 6 aus drei Contigs, von denen nur ein Contig annotiert werden konnte. Dieses Transkript kodiert für *Avidin* (*Xenopus tropicalis*, SP_TREM:A7YYL1_XENTR, 3e-34). In diesem Cluster zeigt sich eine massive Abregulierung im scheinoperierten Gewebe. Im regenerativen Verlauf des Herzens zeigt sich hingegen eine etwas schwächere Abregulierung in den Zeitpunkten zwei Stunden bis 24 Stunden. Dann steigt die Expression kurz und stark an, erreicht ihr Maximum mit einem mittleren $FC= 1,1$ bei sieben Tagen, und fällt dann wieder in eine Abregulation zurück, die nochmals bei 21 Tagen nach Schädigung den Grad der Abregulation wie im scheinoperiertem Gewebe erreicht. Dargestellt ist dieser Cluster in Abbildung 11 F.

Zusammenfassend lässt sich festhalten, dass unter allen deregulierten Contigs im scheinoperiertem Zustand kein Spot enthalten ist, der ausschließlich im scheinoperiertem Zustand eine Expressionsänderung zeigt. Dies deutet darauf hin, dass die bei der Scheinoperation ausgelöste Wundantwort, die im Herz detektierbar ist (humorale Immunantwort, Minimalschädigungen des Herzgewebes und des herzumgebenden Gewebes), in ähnlicher Weise auch während der Regeneration abläuft, bzw. diese Mechanismen zu einer regenerativen Antwort dazugehören.

Ergebnisse

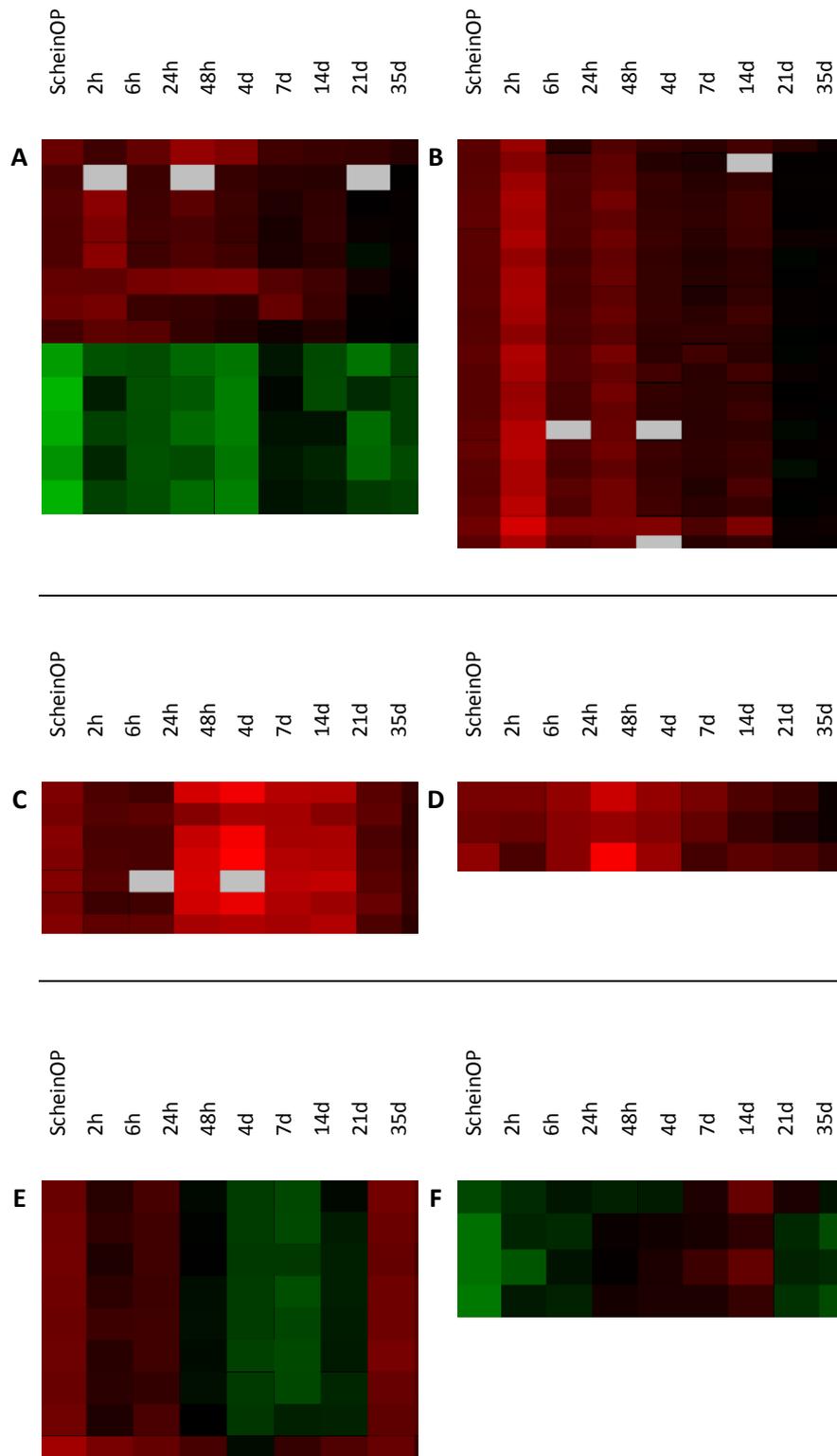


Abbildung 11: SOM Cluster zum scheinoperierten Herzen

Legende auf der nächsten Seite

Dargestellt sind sechs SOM Cluster, die Spalten von links nach rechts zeigen jeweils den scheinoperierten Zustand, dann die regenerativen Zustände von zwei Stunden bis 35 Tagen. Jede Zeile repräsentiert eine Microarray-Koordinate. Die Werte für die einzelnen Zeitpunkte wurden jeweils als Mittelwert der acht durchgeführten Experimente gewählt. Die Farbwerte wurden entsprechend diesem errechneten Mittelwert auf einer Farbskala abgebildet (grün abreguliert, rot aufreguliert, weiß ungültiger Wert). **A** Das Cluster 1 zeigt Microarray Spots zu Transkripten, die im scheinoperiertem Gewebe dereguliert sind. Das Cluster zeigt keinen definierten Verlauf der Expression über die Regeneration hinweg. Vielmehr handelt es sich um generell auf- oder abregulierte Transkripte. Lediglich eine Abnahme der Intensität der Deregulation ist im späten Verlauf der Regeneration erkennbar. **B** Das Cluster 2 zeigt Microarray Spots zu Transkripten, die im scheinoperiertem Gewebe aufreguliert sind. Das Cluster zeigt einen definierten Verlauf der Expression über die Regeneration hinweg. Das Expressionsniveau erreicht schon nach zwei Stunden im Regenerationsverlauf seine maximale Expressionshöhe und schwächt sich dann über den Verlauf der Regeneration stark ab. Nach 48h hat sich die maximale Expressionshöhe halbiert. Nach sieben Tagen ist nahezu das Normalniveau erreicht. Zu den Zeitpunkten 14 Tage bis 35 Tage Regeneration ist keine Deregulation mehr festzustellen. **C** Das Cluster 3 zeigt Microarray Spots zu Transkripten, die im scheinoperiertem Gewebe aufreguliert sind. Das Cluster zeigt einen definierten Verlauf der Expression über die Regeneration hinweg. Das Expressionsniveau erreicht im Bereich nach 24 Stunden bis sieben Tage im Regenerationsverlauf sein maximales Expressionsniveau und schwächt sich dann über den Verlauf der Regeneration stark ab. Nach 14d hat sich die maximale Expressionshöhe halbiert. Nach 21 Tagen ist nahezu das Normalniveau erreicht. **D** Das Cluster 4 zeigt Microarray Spots zu Transkripten, die im scheinoperiertem Gewebe aufreguliert sind. Das Cluster zeigt einen definierten Verlauf der Expression über die Regeneration hinweg. Das Expressionsniveau erreicht zum Zeitpunkt 24h im Regenerationsverlauf sein Maximum und schwächt sich dann über den Verlauf der Regeneration wieder ab. Nach 14d Tagen ist nahezu das Normalniveau erreicht. **E** Das Cluster 5 zeigt Microarray Spots zu Transkripten, die im scheinoperiertem Gewebe aufreguliert sind. Das Cluster zeigt einen bemerkenswerten Verlauf der Expression über die Regeneration hinweg. Das Expressionsniveau schwächt sich nach anfänglicher Aufregulation nach sechs Stunden wieder ab und erreicht zum Zeitpunkt 48 Stunden bis sieben Tage eine Abregulation. Nach 14 Tagen sind die Transkripte jedoch wieder aufreguliert. Im weiteren Verlauf schwächt sich diese Aufregulation wieder ab. **F** Das Cluster 6 zeigt Microarray Spots zu Transkripten, die im scheinoperiertem Gewebe abreguliert sind. Das Cluster zeigt einen definierten Verlauf der Expression über die Regeneration hinweg. Das Expressionsniveau zeigt anfänglich eine leichte Abregulierung und erfährt ab Zeitpunkt vier Tage eine Aufregulierung. Diese erreicht nach 14 Tagen ihr Maximum. Zum Zeitpunkt 21 Tage ist wiederum eine leichte Abregulierung zu erkennen, die ihre stärkste Ausprägung nach 21 Tagen zeigt, dann aber wieder zur normalen Expression zurückkehrt.

Unter den insgesamt 15 abregulierten Transkripten im scheinoperierten Zustand sind auffälligerweise vier Contigs mit Teilen des Komplementsystems annotiert, so *Faktor B*, *Komponente C1q C Chain* und *Komponente C1q B Chain* (redundante Contigs für *Komponente C1q C Chain*).

Faktor B ist ein essentielles Mitglied bei der Auslösung des alternativen Weges der Komplementsystem-Aktivierung. Hier zeigt sich eine massive Abregulation zum Zeitpunkt 24 Stunden. Die *Komponente C1* spielt ausschließlich im ersten Schritt bei der Aktivierung des

Ergebnisse

klassischen Aktivierungsweges des Komplementsystems eine Rolle und ist ebenfalls abreguliert. Diese starke Abregulierung der beiden Aktivierungswege des Komplementsystems an übergeordneter Stelle ist ungewöhnlich, da grundsätzlich eine inflammatorische Reaktion des Organismus auf einen operativen Eingriff zu erwarten wäre.

Weiterhin kodieren zwei Contigs für *Avidin*. Ein Contig kodiert für das *Rrm2-prov protein*. Avidin ist ein Glycoprotein, welches im Hühnerei detektiert wurde. Es hat die Fähigkeit, Biotin zu binden. Diese Bindefunktion wird häufig in der Molekularbiologie für Affinitätschromatographien genutzt. So kommt die Biotinylierung und Bindung an Avidin (Streptavidin) beispielsweise bei der seriellen Analyse der Genexpression (SAGE) und auch der 454 Sequenzierungstechnik zum Einsatz. Die biologische Funktion von Avidin ist unbekannt. Avidin wurde neben einigen Vögeln auch im Frosch detektiert. Das Rrm2-prov protein konnte im Zusammenhang mit Zellproliferation gesetzt werden. So reduziert die siRNA gebundene Inhibition von Rrm-2-provoprotein die Zellproliferationsrate *in vivo* und *in vitro* [84,85]. Die restlichen Contigs sind unbekannt oder haben nur eine EST Annotation.

2 Herzen nach zweistündiger Regeneration

Der Datensatz des zwei Stunden regenerativen Herzgewebes zeigt insgesamt im Vergleich zu allen anderen Zeitpunkten die größte Anzahl an signifikant deregulierten Spots. So sind auf den beiden Teilarrays insgesamt 722 (835) Spots dereguliert. Die Zahlen in Klammern geben hier und im Folgenden jeweils die Anzahl aller Transkripte inklusive ribosomaler Sequenzen an. Diese wurden aus der Gesamtbetrachtung herausgenommen. Davon sind 621 (714) Spots stark aufreguliert (Ratio >1 in 6/8 Chips, p Wert <0.05) und 101 (121) stark abreguliert (Ratio <1 in 6/8 Chips, p Wert <0.05). Diese Spotanzahlen spiegeln jedoch nur bedingt die absolute Anzahl an endgültigen Transkripten wider, da innerhalb der Aufregulierten einige Transkripte hoch abundant sind. So beinhalten die aufregulierten Spots 399 Spots, die alle für das Contig 4000 (DB Id) kodieren. Insgesamt kodieren die 621 (714) aufregulierten Spots für 164 aufregulierte Contigs. Daneben kodieren die 101 (121) abregulierten Spots für 29 abregulierte Contigs.

Zunächst fällt im Gegensatz zum ScheinOP-Datensatz auf, dass Transkripte existieren, die ausschließlich nach zwei Stunden Regeneration eine Deregulation aufweisen. Diese sind also spezifisch für den Zeitpunkt zwei Stunden Regeneration dereguliert.

Aus der Gesamtmenge der zu zwei Stunden Regeneration deregulierten Spots lassen sich drei markante Cluster mit unterschiedlichem Expressionsverlauf identifizieren. Diese werden im Folgenden, soweit annotierbar, unter Angabe des getroffenen Organismus, des zugehörigen Datenbank Identifiers und des E-values angegeben.

Das erste Cluster (Cluster 7) zum Zeitpunkt zwei Stunden Regeneration, beschreibt eine Gruppe an Transkripten, die spezifisch für diesen Schädigungszeitpunkt erscheint. Die enthaltenen Transkripte zeigen eine massive Aufregulation ausschließlich zu diesem Zeitpunkt. Alle anderen beobachteten Zeitpunkte sind kaum oder gar nicht dereguliert. In dieser Gruppe lassen sich neun unterschiedliche Contigs annotieren. Dies sind *piggyBac transposase Uribo2* (*Xenopus tropicalis*, dbj|BAF82026.1, 1e-101), *Myosin light chain 1av* (African clawed frog, SP_TREM:Q90ZP1_XENLA, 4e-091), *troponin C type 1 (slow)* (*Bos taurus*, ref|NP_001029523.1, 5e-046), *Eukaryotic translation elongation factor 1 delta* (*Homo Sapiens*, SP_TREM:B2RAR6_HUMAN, 2e-013), *transducin (beta)-like 3* (*Xenopus tropicalis*, gb|AAI35428.1, 2e-058), *satellite CYP-St-BamB1* (*Cynops pyrrhogaster*, dbj|AB238601.1, 3e-019), *satellite 2 DNA* (*N. viridescens*, emb|X04478.1, 1e-125), *satellite 2 repetitive DNA element cps2-2hhc* (*Cynops pyrrhogaster*, gb|U37317.1, 1e-21) und *myosin heavy chain 6* (*Canis lupus familiaris*, ref|NP_001107183.1, ~0). Darüber hinaus kodieren 16 Contigs für Sequenzen, die nur mit ESTs anderer Organismen annotiert werden konnten. Dargestellt ist dieser Cluster in Abbildung 12 A. Außerdem zeigten 45 weitere Contigs keinerlei signifikante Homologie zu bekannten Sequenzen.

Der zweite Cluster (Cluster 8) zum Zeitpunkt zwei Stunden beschreibt eine Gruppe von Transkripten, die ebenfalls stark aufreguliert ist, deren Abschwächung der erhöhten Expression jedoch etwas langsamer über den Regenerationsverlauf geschieht. Bei den hier vertretenen Transkripten erreicht das Expressionsniveau erst nach vier bis sieben Tagen wieder das normale Maß. Im weiteren Verlauf der Regeneration sind diese Transkripte vollkommen unauffällig.

Ergebnisse

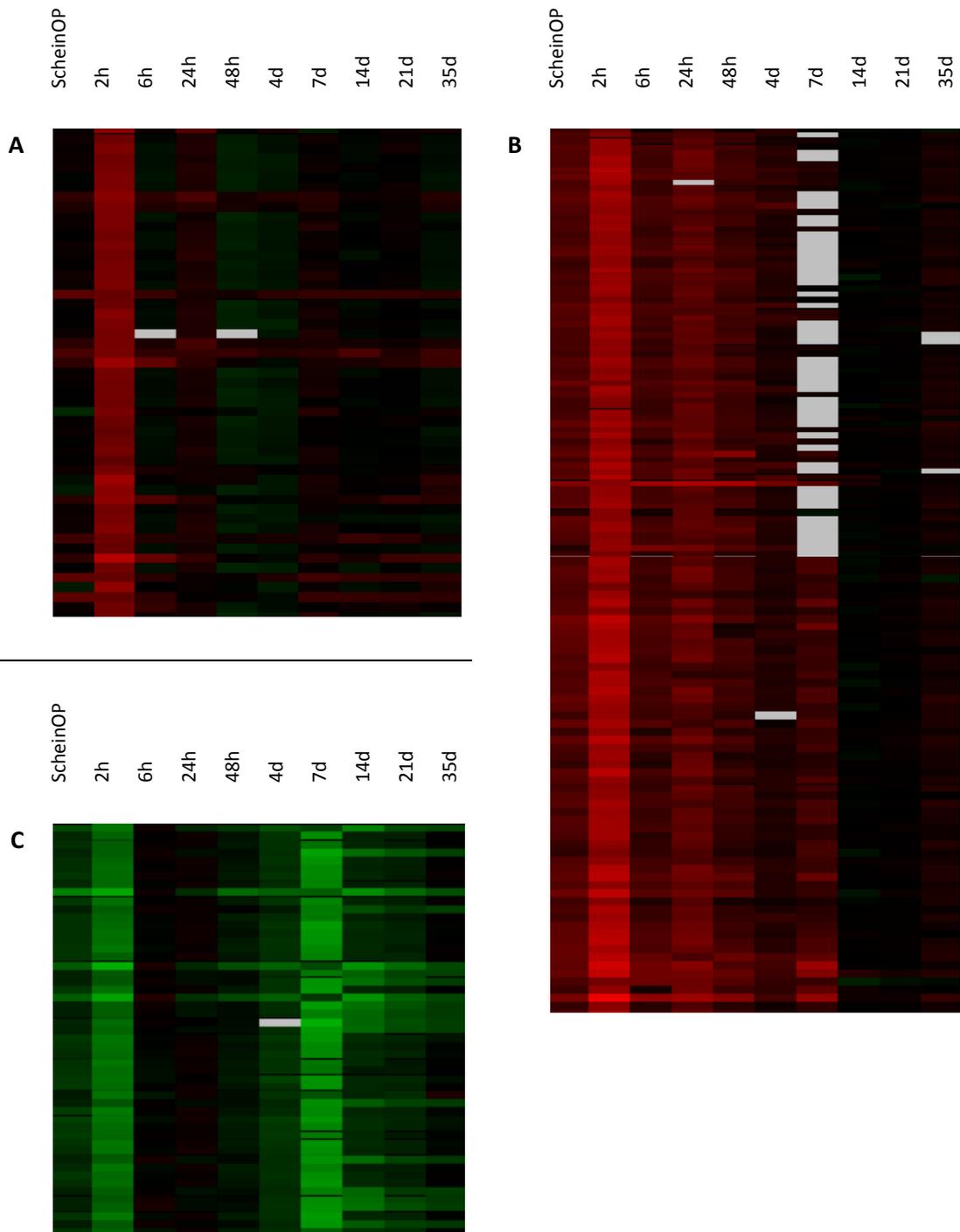


Abbildung 12: SOM Cluster, 2 Stunden Regeneration des Herzen

Legende auf der nächsten Seite

Dargestellt sind drei SOM Cluster, die Spalten von links nach rechts zeigen jeweils den scheinoperierten Zustand, dann die regenerativen Zustände von zwei Stunden bis 35 Tagen. Jede Zeile repräsentiert eine Microarray-Koordinate. Die Werte für die einzelnen Zeitpunkte wurden jeweils als Mittelwert der acht durchgeführten Experimente gewählt. Die Farbwerte wurden entsprechend diesem errechneten Mittelwert auf einer Farbskala abgebildet (grün abreguliert, rot aufreguliert, weiß ungültiger Wert). **A** Das Cluster 7 zeigt Microarray Spots zu Transkripten, die ausschließlich im zwei Stunden regenerierendem Gewebe aufreguliert sind. Das Cluster zeigt einen definierten Verlauf der Expression über die Regeneration hinweg. Das Expressionsniveau zeigt weder im scheinoperiertem noch zu einem anderen Zeitpunkt als zwei Stunden nach Schädigung eine signifikante Expressionsänderung. **B** Das Cluster 8 zeigt Microarray Spots zu Transkripten, die stark im zwei Stunden regenerierendem Gewebe aufreguliert sind und dann bis zum Zeitpunkt vier Tage kontinuierlich wieder schwächer exprimiert werden. Das Expressionsniveau zeigt auch im scheinoperiertem Zustand eine leichte Aufregulation. **C** Das Cluster 9 zeigt Microarray Spots zu Transkripten, die im zwei Stunden regenerierendem Gewebe abreguliert sind. Das Cluster zeigt einen definierten Verlauf der Expression über die Regeneration hinweg. Das Expressionsniveau steigt zunächst bis sechs Stunden nach Schädigung wieder an, so dass keine Expressionsänderung mehr detektiert werden. Dieses normale Expressionsniveau wird bis zum Zeitpunkt 48 Stunden nach Schädigung gehalten, danach sinkt das Expressionsniveau wieder stark ab. Die niedrige Expression steigt abermals langsam wieder an, hat aber zum Ende der Beobachtung mit dem Zeitpunkt 35 Tage nach Schädigung das ursprüngliche Expressionsniveau noch nicht wieder erreicht. Im scheinoperiertem Gewebe kann eine leichte Abregulation gezeigt werden.

In dieser Gruppe ließen sich die Transkripte *ADAM metalloproteinase with thrombospondin type 1 motif, 1* (*Xenopus laevis*, dbj|BAE94917.1, 1e-115), *Tomm22 protein* (*Xenopus tropicalis*, gb|AAI60473.1, 8e-022), *insulin induced gene 2* (*Sus scrofa*, ref|NP_001123440.1, 2e-091), *tissue factor pathway inhibitor 2* (*Bos Taurus*, ref|NP_877589.1, 1e-090), *signal sequence receptor, gamma* (*Xenopus tropicalis*, emb|CAJ81804.1, 2e-086), *Chain A, Crystal Structure Analysis Of Human Serine Dehydratase* (*Homo sapiens*, pdb|1P5J, 2e-036) und *interleukin 8* (*Gallus gallus*, ref|NP_990349.1, 5e-023) annotieren. Weitere drei Contigs zeigten nur eine Homologie zu anderen EST Banken aus *Cynops pyrrhogaster* und dem *Ambystoma*. Außerdem sind vier Contigs enthalten, die keinerlei Sequenzhomologie zu anderen Organismen zeigen. Darunter auch das in hoher Kopienzahl vorliegende Contig 4000. Dargestellt ist dieser Cluster in Abbildung 12 B.

Der dritte Cluster (Cluster 9) zeigt eine Gruppe an Transkripten, die im zwei Stunden regenerativen Gewebe signifikant abreguliert sind. Diese Abregulation ist jedoch nach sechs Stunden Regeneration nicht mehr zu detektieren und bleibt nahezu auf unverändertem Niveau bis zu 48 Stunden nach der Schädigung. Dann setzt nach vier Tagen abermals eine leichte Abregulation ein, die nach sieben Tagen ihr Maximum erreicht. Im weiteren Verlauf der Regeneration normalisiert sich das Expressionsniveau der Transkripte wieder und erreicht zum Beobachtungszeitpunkt 35 Tage nach Schädigung nahezu ein zum ungeschädigten Herz

Ergebnisse

unverändertes Signal. Der scheinoperierte Zustand ist bei diesen Transkripten nicht signifikant dereguliert.

In dieser Gruppe konnte zu allen Transkripten eine homologe Sequenz in anderen Organismen gefunden werden. Dargestellt ist dieser Cluster in Abbildung 12 C. Bis auf eine Sequenz (*unnamed protein product, Macaca fascicularis*, dbj|BAE89198.1, 6e-055) konnten alle weiteren entsprechend als *Peripheral-type benzodiazepine receptor (Oncorhynchus mykiss*, SP_TREM:Q7ZZV7, 1e-060, 5 unterschiedliche Contigs), *Hemoglobin subunit alpha-2 (Pleurodeles waltlii* , SW:HBA2_PLEWA, 5e-061, 3 unterschiedliche Contigs), *larval alpha-globin (Hynobius retardatus*, dbj|BAA86389.1, 9e-029), *Interferon-inducible protein Gig2 (Carassius auratus*, SP_TREM:Q7T2N0_CARAU, 6e-038), *Peroxiredoxin 2 (Rattus norvegicus*, gb|AAH58481.1, 7e-089), *ATP-binding cassette sub-family B member 8 (Xenopus tropicalis*, SW:ABCB8_XENTR, 7e-021) und *Importin alpha 4 protein (Xenopus laevis*, gb|AAI69966.1, 1e-078) annotiert werden.

3 Herzen nach sechsständiger Regeneration

Der Datensatz des 6 Stunden regenerativen Herzgewebes zeigt im Vergleich zu allen anderen Zeitpunkten nur eine relativ kleine Anzahl an signifikant deregulierten Spots. So sind auf den beiden Teilarrays 1 und 2 insgesamt 58 (71) Spots dereguliert. Die Zahlen in Klammern geben hier und im Folgenden jeweils die Anzahl aller Transkripte inklusive ribosomaler Sequenzen an. Diese wurden aus der Gesamtbetrachtung herausgenommen. Davon sind 42 (54) Spots stark aufreguliert (Ratio >1 in 6/8 Chips, p Wert <0.05) und 16 Spots stark abreguliert (Ratio <1 in 6/8 Chips, p Wert <0.05). Die aufregulierten Spots kodieren für 31 Contigs, 23 davon konnten annotiert werden.

Daneben kodieren die 16 abregulierten Spots für elf abregulierte Contigs, wovon fünf annotiert werden konnten. Die Contigs werden im Folgenden, soweit annotierbar, unter Angabe des getroffenen Organismus, des zugehörigen Datenbank Identifiers und des E-values angegeben.

Von den fünf abregulierten Contigs konnte *Dyskerin (Bos taurus*, ref|NP_001098865.1, 6e-090, Contig5683) zum Zeitpunkt sechs Stunden Regeneration im Vergleich zu den anderen Regenerationszeitpunkten als am stärksten abreguliert detektiert werden. Einen ähnlichen Verlauf über die beobachteten regenerativen Zeitpunkte zeigt *MT1 gene for metallothionein 1 (Gallus gallus*, dbj|AB258231.1, 2e-015, Contig15161). Das Contig29355 kodiert für ein

Transkript, welches nur im *Cynops pyrrhogaster* (dbj|FS291488.1, 1e-021) über ein EST annotiert werden konnte. Dieses Contig zeigt zusammen mit drei weiteren, nicht annotierten Contigs (Contig5116, Contig6120, Contig7365) im Verlauf der Regeneration eine starke Abregulierung, die erst nach 48 Stunden ihr Minimum erreicht, dann im Expressionsniveau wieder leicht ansteigt, um dann im Zeitraum 14 Tage nach Schädigung erneut massiv abreguliert zu werden. Die übrigen zwei annotierten Contigs hatten ebenfalls nur Ähnlichkeit zu ESTs im *Ambystoma* (Contig4846 und Contig5387). Neben diesen beiden Contigs finden sich zwei weitere, nicht annotierte Contigs, die über den Verlauf der Regeneration das gleiche Expressionsmuster zeigen (Contig6837 und Contig7082). Dieses Expressionsmuster zeichnet sich dadurch aus, dass zu Anfang der Regeneration bis zum Zeitpunkt sechs Stunden eine starke Abregulation erfolgt, dann aber bereits zum Zeitpunkt ein Tag Regeneration eine leichte, nach zwei Tagen eine deutliche Aufregulierung stattfindet. Diese schwächt sich im weiteren Verlauf über die Zeitpunkte vier und sieben Tage nochmals ab, und wiederholt sich dann noch einmal zum Zeitpunkt 14 Tage Regeneration. (siehe Abbildung 13)

Die 31 aufregulierten Contigs gliedern sich im Wesentlichen in vier auffällige Gruppen. Zum Ersten sind dies sieben Contigs (Contig6821, Contig22966, Contig4371, Contig20101, Contig15136, Contig4540, Contig18263) die eine maximale Aufregulierung zum Zeitpunkt sechs Stunden zeigen, danach aber wieder rasch in ihrer Überexpression zurück gehen und dann zwischen zwei und vier Tagen Regeneration das normale Expressionsniveau erreichen. In dieser Gruppe lassen sich sechs Contigs als *core promoter element binding protein* (*Xenopus laevis*, gb|BC044713.1|, 3e-057), *Deoxyribonuclease I-like 3, like* (*Danio rerio*, emb|CAX14380.1, 7e-028), *B cell translocation protein 2* (*Sus scrofa*, gb|ABW83761.1, 4e-053), *Fos-related antigen 2, transcript variant 2* (FOSL2) (*Bos taurus*, ref|XM_864138.2, 2e-013), Ras like G protein Rad (*Cynops pyrrhogaster*, dbj|BAB18863.1, 1e-154) und *Dual specificity phosphatase 5* (*Xenopus tropicalis*, Q5CZZ4, 8e-015) annotieren. Eine zweite Gruppe von 13 Contigs (Contig22954, Contig15157, Contig5607, Contig14464, Contig6439, Contig15122, Contig16085, Contig15190, Contig5152, Contig6460, Contig7515, Contig8475, Contig17490) zeigt ein ähnliches Expressionsmuster im regenerativen Verlauf. Der Unterschied besteht darin, dass diese Transkripte zwar auch bei sechs Stunden Regeneration ihr maximales Expressionsniveau erreichen, diese hohe Expression dann aber erst im Verlauf bis zum Zeitpunkt 14 Tage Regeneration auf das normale Expressionsniveau zurückfällt. In dieser Gruppe lassen sich acht Contigs als *glycoprotein IX* (platelet) (*Sus scrofa*, ref|NP_001135461.1 9e-012), *kruppel-like factor 6* (*Sus scrofa*, ref|NP_001127825.1, 9e-042), *tissue factor pathway inhibitor 2* (*Rattus norvegicus*, ref|NP_775164.1, 6e-040), *protein phosphatase 1, regulatory (inhibitor) subunit 15A*

Ergebnisse

(*Bos taurus*, ref|NP_001039643.1, 1e-011), *tissue factor pathway inhibitor 2, isoform CRA_b* (*Rattus norvegicus*, gb|EDL84391.1, 4e-046), *ADAM metallopeptidase with thrombospondin type 1 motif, 1* (*Taeniopygia guttata*, ref|XM_002189365.1, 1e-011), *alpha-1-microglobulin/bikunin precursor, like precursor* (*Danio rerio*, ref|NP_957412.1, 2e-022) und *solute carrier family 7 (cationic amino acid transporter, y+ system), member 3* (*Xenopus tropicalis*, ref|NP_001116958.1, 2e-028) annotieren. Drei weitere Contigs können nur durch ESTs annotiert werden, die übrigen zwei zeigen keine Homologie.

Die dritte Gruppe fasst Contigs zusammen, die ihr maximales Expressionsniveau nicht zum Zeitpunkt sechs Stunden Regeneration aufzeigen, sondern sich zu diesem Beobachtungszeitpunkt gerade in einer ansteigenden oder abfallenden Phase der Expression befinden. In diese Gruppe gehören vier Contigs (Contig6959, Contig7112, Contig23290, Contig6498), von denen neben einer EST Annotation nur das *Transmembrane 4 L6 family member 1* (*Homo sapiens*, P30408, 3e-076) Protein annotiert werden kann. Die letzte Gruppe zeigt kein einheitliches Muster im Verlauf der Regeneration. Ihr ist jedoch gemeinsam, dass es jeweils zu starken Richtungsänderungen der Expression im Zeitverlauf der Regeneration kommt. In dieser Gruppe sind sieben Contigs (Contig27170, Contig29318, Contig22974, Contig24929, Contig5476, Contig22962, Contig8936) enthalten. Davon können *peptidyl-Lys metalloendopeptidase* (*Gallus gallus*, ref|XM_422045.2, 6e-024), *highly repetitive sequence* (*Cynops pyrrhogaster*, dbj|AB122113.1, 5e-014), *ADAM metallopeptidase with thrombospondin type 1 motif* (*Xenopus laevis*, dbj|BAE94917.1, 1e-115) und *cysteine-rich, angiogenic inducer, 61* (*Gallus gallus*, ref|NP_001026734.1, 7e-080) annotiert werden. Die übrigen drei zeigen keine Homologie.

Da zum Zeitpunkt sechs Stunden Regeneration insgesamt nur wenige unterschiedliche Contigs eine Deregulation aufzeigen, sind die zugehörigen Spots in Abbildung 13 gemeinsam in Form eines Hierarchischen Clusters dargestellt. Die im Text genannten Contignummern sind in der Abbildung eingefügt, um den beschriebenen Verlauf der Expression nachverfolgen zu können.

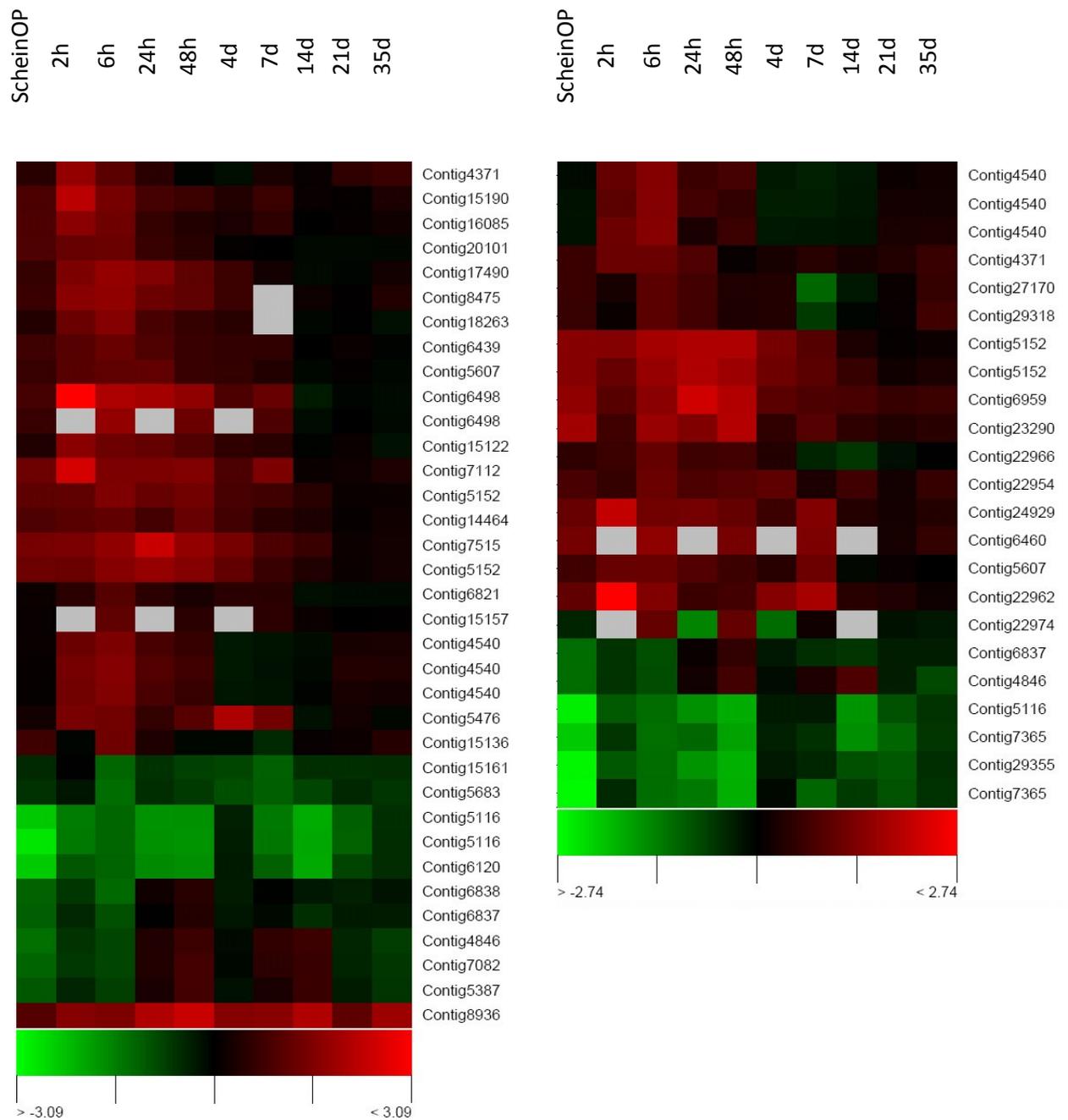


Abbildung 13: Hierarchischer Cluster der deregulierten Spots nach 6 Stunden Regeneration

Die linke Spalte zeigt das hierarchische Cluster für den Microarray Teil 1, die rechte Spalte das Cluster für den Microarray Teil 2. Die Spalten zeigen die einzelnen beobachteten Zeitpunkte. Die Annotation ist über die Contignummer angefügt. Es lassen sich die Hauptgruppen „durchgängig stark abreguliert“, „bis zu 14 Tagen stark aufreguliert“ und „bis zu 2 Tagen stark aufreguliert“ unterscheiden. Dazu kommt noch die Gruppe, die eine „Abregulierung mit zwischenzeitlicher temporärer Aufregulierung“ und eine „Aufregulierung mit zwischenzeitlicher temporärer Abregulierung“ zeigen.

4. Vergleich der regenerativen Zeitpunkte

Der Vergleich der drei Zeitpunkte ScheinOP, zwei Stunden und sechs Stunden Regeneration führt zu der in Abbildung 14 dargestellten Contig Verteilung. Während 29 Contigs ausschließlich im scheinoperiertem Gewebe als signifikant dereguliert vorkommen, sind es im zwei Stunden regenerativen Gewebe 172 unterschiedliche Contigs, im sechs Stunden regenerativen Gewebe 24 unterschiedliche Contigs. In allen drei Zuständen gemeinsam sind vier Contigs dereguliert, darunter annotierbar *tissue factor pathway inhibitor 2* und *alpha-1-microglobulin/bikunin precursor, like precursor*. Die beiden übrigen zeigen keine Homologie.

Daneben finden sich in der Gruppe der ausschließlich in ScheinOP und zwei Stunden Regeneration exprimierten Contigs neben drei nicht annotierbaren Contigs die Proteine *Tomm22 protein (Xenopus tropicalis, gb|AAI60473.1, 8e-022)*, *myosin, heavy chain 3, skeletal muscle, embryonic (Canis lupus familiaris, ref|NP_001107184.1, ~0)*, *rgs2 protein (Xenopus tropicalis, ref|NP_001025622.1, 2e-082)*, *putative heat shock protein 25 variant 1 (Taeniopygia guttata, gb|ACH45320.1, 4e-029)* und *complement component 1, q subcomponent, C chain (Rattus norvegicus, ref|NP_001008524.1, 6e-031)*. Alle Transkripte sind jeweils in beiden Beobachtungen in die gleiche Richtung dereguliert.

Zu den neun Contigs, die ausschließlich im regenerierendem Herz vorkommen (zwei und sechs Stunden), konnten die Proteine *phosphatase 1, regulatory (inhibitor) subunit 15A, solute carrier family 7 (cationic amino acid transporter, γ^+ system), member 3, cysteine-rich, angiogenic inducer, 61, ADAM metallopeptidase with thrombospondin type 1 motif, 1, B-cell translocation gene 2, Ras like G protein Rad*, nochmals *ADAM metallopeptidase with thrombospondin type 1 motif, 1* und 2 Contigs zu ESTs annotiert werden (*Ambystoma mexicanum, gb|CN037492.15e-011* und *Rattus norvegicus, B2BKZ7, 9e-012*). Diese Transkripte können als regenerationspezifisch angesehen werden. Eine Sequenzanalyse der beiden Contigs, die eine Homologie zu den ESTs zeigen, erbrachten für das *Ambystoma*-Homolog zwei konservierte Zinkfingerdomänen (Zinc knuckle, Interpro IPR001878, Position 1-11 und 21-33).

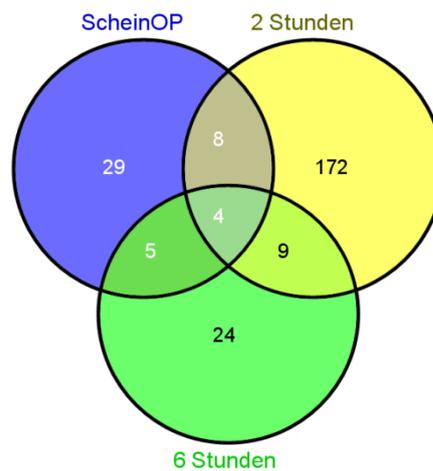


Abbildung 14: Vergleich der deregulierten Contigs zu den Zeitpunkten ScheinOP, 2 Stunden und 6 Stunden Regeneration als Venndiagramm

Dargestellt ist der Abgleich der drei Zustände Scheinoperiert, zwei Stunden Regeneration und sechs Stunden Regeneration. Zu jedem Zustand des Gewebes sind jeweils die signifikant deregulierten Contigs vermerkt. Der Überschneidung zu allen drei Zeitpunkten liegt bei vier Transkripten. Deutlich hebt sich der Zeitpunkt zwei Stunden Regeneration heraus, der deutlich mehr deregulierte Transkripte als die anderen beiden Zustände aufweist.

In der Gruppe der Transkripte, die ausschließlich im Regenerativen Gewebe zwei Stunden signifikant dereguliert sind, finden sich einige bekannte Vertreter der frühen Wundantwort. So sind beispielsweise die beiden Interaktionspartner *nuclear receptor subfamily 4, group A, member 1* (nr4A1) und *Jun d proto-oncogene* stark aufreguliert. Beide zeigen einen deutlichen Peak zum Zeitpunkt zwei Stunden Regeneration, wobei das hohe Expressionsniveau rasch innerhalb von einem Tag wieder auf normale Expressionsniveau zurücksinkt und danach unauffällig bleibt. Diese beiden Interaktionspartner spielen bei der Verhinderung von Apoptose eine Rolle [86,87].

Zusammenfassend kann festgehalten werden, dass die Analyse der frühen Regenerationszeitpunkte zwei Stunden und sechs Stunden im Vergleich zum scheinoperierten Gewebe eine Reihe an potentiellen Kandidaten für weitergehende biologische Funktionsanalysen hervorgebracht hat.

5. Auswahl der Kandidaten zur DNA Sequenzierung

Eine komplette Sequenzierung der 100000 Klone umfassenden cDNA Bibliothek war aus Kostengründen nicht sinnvoll. Es sollte daher eine geeignete repräsentative Auswahl an Sequenzen aus der Gesamtbibliothek so ausgewählt werden, dass möglichst viele unterschiedliche Sequenzen enthalten sind. Dazu wurde zunächst eine weitere Microarrayanalyse auf allen Experimenten durchgeführt. In dieser Analyse wurden nach gleichen Parametern, wie zuvor beschrieben, signifikant deregulierte Spots herausgesucht. So wurde eine Gesamtanzahl von 2994 potentiell deregulierten Spots sequenziert. Darüber hinaus wurde eine zweite Gruppe an Spots aus den Microarrays erstellt, die zwar keine Deregulation in den untersuchten Zeitpunkten vorweisen, aber über ihre Bindungsintensität ein deutliches Vorhandensein generell und zusätzlich ein gelungenes Spotting auf dem Microarray zeigten. Diese Liste umfasste 44544 Spots. Darüber hinaus wurden einige weitere Spots zur Auffüllung ganzer 96 well Sequenzierungsplatten ausgewählt, damit der Aufwand für *Re-Arraying* vor der Sequenzierung in einem überschaubaren Rahmen blieb. Insgesamt wurden 52243 Koordinaten der Bibliothek ausgewählt und sequenziert. Die im vorherigen Kapitel genannten Annotationen sind das Ergebnis dieser ausgewählten Sequenzierung.

6. Assemblierung, Annotation und Datenspeicherung

Nach der Sequenzierung wurden die erstellten Sequenzen in Zusammenarbeit mit dem Max-Planck-Institut für molekulare Genetik in Berlin weiterverarbeitet. Die insgesamt 52243 erstellten Sequenzen (52% der gesamten cDNA Bibliothek) wurden zunächst mit dem Programm Phred [88] auf Lesqualität der Sequenz hin überprüft. Mit dem Programm Lucy [89] wurden flankierende Vektorsequenzen entfernt. Als Schwellenwert für die Sequenzlänge wurde ein Minimum von 50 Basenpaaren gesetzt. 1178 Sequenzen fielen unterhalb dieses Schwellenwertes. 42 weitere Sequenzen enthielten nur Bestandteile der flankierenden Vektorsequenz. Außerdem mussten 1908 Sequenzen aufgrund zu hohen Grundrauschens (Tracedata) ausgeschlossen werden. 78 Sequenzen bestanden nur aus Klonierungsrelikten. Weitere 500 Sequenzen stellten rRNA Sequenzen dar und wurden entfernt (megablast mit Schwellenwert $1e-10$, siehe Bemerkungen in VI.A.3.a). Somit wurden 48537 Sequenzen mit einer mittleren Leslänge von 591 Basenpaaren für die Assemblierung übernommen. Dazu wurde das Programm wcd [90] verwendet. Zunächst wurden 5510 initiale Sequenz Cluster aus 33750 ESTs erstellt. Aus diesen Clustern wurden 6155 Contigs assembliert. Das Contig mit den meisten ESTs

enthielt 610 ESTs, das längste Contig erreichte eine assemblierte Länge von 2683 Basenpaaren, die mittlere Contiglänge liegt bei 831 Basenpaaren. Zu den 6155 Contigs kommen 20439 sogenannte Singletons, die keinem Contig zugeordnet werden konnten. Somit lag abschließend eine Gesamtzahl von 26594 hochqualitativen, unterschiedlichen Sequenzen vor. Im Folgenden wird aus Gründen der Vereinfachung generell der Begriff Contig sowohl für zusammengesetzte Contigs als auch für Transkripte verwendet, die nur aus einem EST bestehen.

Im nächsten Schritt wurden die 26594 neuen Molch-Sequenzen soweit wie möglich annotiert, indem nach hohen Sequenzähnlichkeiten zu bereits bekannten Sequenzen in öffentlich zugänglichen Sequenzdatenbanken gesucht wurde. Hierzu kamen der BLAST Algorithmus (basic local alignment search tool [91]) in den Varianten BLASTn, BLASTx und tBLASTx zum Einsatz.

Unter Verwendung des *BLASTc/3* Programmes (<http://www.ncbi.nlm.nih.gov/staff/tao/URLAPI/netblast.html>), das den Start einer Ähnlichkeitssuche zu einer Sequenz in den Sequenzdatenbanken der NCBI (National Center for Biotechnology Information) aus der Konsole ermöglicht, wurde eine umfassende Annotationssuche für die neu erstellten Molch-Sequenzen durchgeführt. Mithilfe einer automatisierten Programmabfolge wurden alle Contig/Singleton Sequenzen jeweils mit den Parametern

- BLASTn gegen NCBI NR (Nukleotide)
- BLASTn gegen NCBI EST (Nukleotide)
- tBLASTx gegen NCBI EST (Nukleotide)
- tBLASTx gegen NCBI NR (Nukleotide)
- BLASTn gegen NCBI HTGS (Nukleotide)
- BLASTx gegen NCBI NR (Proteine)

gestartet. Die Ergebnisse aller 159564 BLAST Aufrufe wurden als einzelne Textdateien abgelegt. Für potentielle Ergebnisse wurde der Wahrscheinlichkeitswert für das zufällige Vorkommen des gefundenen Alignment mit $1e-10$ als Schwellenwert gesetzt. Alle Alignments mit niedrigerer zufälliger Wahrscheinlichkeit wurden für die weitere Verarbeitung mit Herkunftsdatenbank, Parametersatz, Alignment und weiteren elementaren Kennwerten gespeichert und in einer hausinternen MySQL Datenbank abgelegt. Aus der Gesamtheit aller Treffer in den entsprechenden Datenbanken konnte die in Abbildung 15 aufgeführte Verteilung erstellt werden. In dieser Verteilung werden für jeden der genannten BLAST-Datenbank Varianten und

Ergebnisse

zu jedem annotierten Contig die beste Annotation auf den Organismus der Quelle hin überprüft und nach dem häufigsten Vorkommen von einzelnen Organismen hin sortiert.

In diesem Zusammenhang sind zwei wesentliche Effekte zu beachten. Zum einen hat der Inhalt der verschiedenen Datenbanken selbst einen starken Effekt auf das Vorkommen der gefundenen Organismen mit den am besten bewerteten Homologien. So sind beispielsweise in den EST Datenbanken ein Vielfaches mehr *Ambystoma* und *Xenopus* Sequenzen hinterlegt, als in der Proteindatenbank NR. Daneben fällt besonders ins Auge, dass sich die resultierenden sechs Tabellen in Abbildung 15 zu den BLAST-Datenbank Varianten in ihrer Quell-Organismus Zusammensetzung kaum unterscheiden. Insbesondere treten die am besten untersuchten Organismen Mensch, Maus und Ratte kaum in Erscheinung, obwohl eine Mehrheit der in den Datenbanken vorliegenden Sequenzen auf diese Organismen verweisen.

Aufgrund dieser Beobachtung sollte überprüft werden, ob sich anhand der höchstwertigen Annotationen dieser signifikanten Verteilung ein Rückschluss auf Verwandtschaftsverhältnisse ziehen lässt. Um diese Frage zu beantworten, wurde der etablierte Cytochrom C Stammbaum herangezogen, der in Abbildung 16 gezeigt ist. Als Vertreter der Amphibien wurde in diesem Stammbaum der Ochsenfrosch (Bullfrog) gewählt. Augenscheinig ist beim Vergleich dieses Stammbaumes mit den in Abbildung 15 gegebenen Treffern das hohe Maß der Übereinstimmung bezüglich der daraus abgeleiteten Nähe der Verwandtschaft und der Anzahl der höchstwertig getroffenen Homologien. So finden sich stets, wie erwartet, die nahen verwandten Amphibien (*Xenopus* und *Ambystoma*) immer auf den ersten Plätzen, sobald eine signifikante Anzahl an Sequenzen von *Xenopus* und/oder *Ambystoma* in der jeweilig betrachteten Bank enthalten ist.

In vier von sechs BLAST Läufen finden sich die Vögel Huhn (*Gallus gallus*) und Zebrafink (*Taeniopygia guttata*) vor anderen hohen Wirbeltieren in den Trefferlisten. Nach dem Cytochrom C Stammbaum sind Vögel den Amphibien wesentlich näher verwandt als beispielsweise hochentwickelte Säuger. Auch bestätigt sich mit der erhöhten Trefferquote der Säuger Pferd (*Equus caballus*), Schwein (*Sus scrofa*), Rind (*Bos taurus*), Hund (*Canis*) und Makake (*Macaca*) die Homologie der gefundenen Annotationen mit dem Cytochrom C Stammbaum, der diese Tiere den Amphibien näher verwandt zuordnet als dem Menschen oder der Maus.

Auf Basis aller BLAST Ergebnisse wurde eine Annotationsdatei für alle sequenzierten Plattenkoordinaten erstellt. Vermerkt wurden, sortiert nach E-value, die besten fünf Treffer zu homologen Sequenzen. Um eine bestmögliche Annotation zu gewährleisten, wurde neben dem

E-value nach den durchsuchten Datenbanken sortiert. So wurde zunächst überprüft, ob ein signifikanter Treffer in der NR Protein oder NR Nukleotid Datenbank vorlag. War dies der Fall, so wurden diese Treffer herangezogen. Nur wenn dies nicht der Fall war, wurden Treffer aus den EST und HTGS Datenbanken herangezogen. Waren auch keine Treffer in diesen Banken vorhanden, wurden die Sequenzen mit dem Attribut „no Hit“ versehen. Somit gliedert sich die gesamte Annotationsdatei in vier Hauptgruppen, die hochwertig annotierten, die nur EST annotierten, die nicht annotierten und die nicht sequenzierten Koordinaten. Diese Annotationsdatei stellt die Basis für die Auswertung der Microarray und später folgenden Proteom Experimente dar, da nur durch diese Datei eine Zuordnung der auf den Microarrays gemessenen Expressionsdaten auf Koordinatenbasis zu den einzelnen annotierten Transkripten möglich ist.

Ergebnisse

Organismus	blastn nr	Organismus	blastx nr	Organismus	blastn est
Xenopus	479	Xenopus	1513	Ambystoma	4111
Taeniopygia guttata	441	Gallus gallus	578	Xenopus	499
Gallus gallus	439	Taeniopygia guttata	444	Gallus gallus	448
Equus caballus	142	Danio rerio	236	Taeniopygia guttata	175
Canis	132	Equus caballus	224	Sus scrofa	139
Bos Taurus	129	Bos taurus	189	Bos taurus	126
Pan troglodytes	106	Pan troglodytes	168	Rattus norvegicus	90
Macaca	94	Canis	157	Danio rerio	77
Sus scrofa	91	Rattus norvegicus	152	Canis	65
Organismus	tblastx est	Organismus	tblastx nr	Organismus	blastn htgs
Ambystoma	1847	Xenopus	1478	Bos taurus	609
Xenopus	828	Gallus gallus	636	Rattus norvegicus	602
Gallus gallus	408	Taeniopygia guttata	440	Sus scrofa	450
Bos Taurus	217	Bos taurus	201	Macaca	100
Danio rerio	114	Danio rerio	177	Gallus gallus	81
Sus scrofa	114	Canis	166	Danio rerio	66
Rattus norvegicus	108	Equus caballus	164	Pan troglodytes	56
Taeniopygia guttata	95	Macaca	153	Canis	32
Macaca	92	Pan troglodytes	139	Pongo	25

Abbildung 15: Zuordnung Organismus zu bester gefundener Annotation (e-value<e-10)

Die sechs Tabellen zeigen die durchgeführten Kombinationen aus Homologiesuche-Algorithmen (BLAST) und Protein- und Nukleotiddatenbanken. Aufgeführt sind jeweils die neun Organismen mit den meisten Treffern während der Homologiesuche in der jeweiligen Datenbank. In die Auswertung wurde für jedes Transkript jeweils der Organismus des besten Treffers aufgenommen. Die rechte Spalte zeigt somit die Anzahl der Transkripte, die ihren besten Homologietreffer zu dem in der linken Spalte genannten Organismus haben. Die Hintergrundfarben zeigen die Ordnung an. Entsprechende Farben wurden im Cytochrom C Baum in Abbildung 16 hinterlegt.

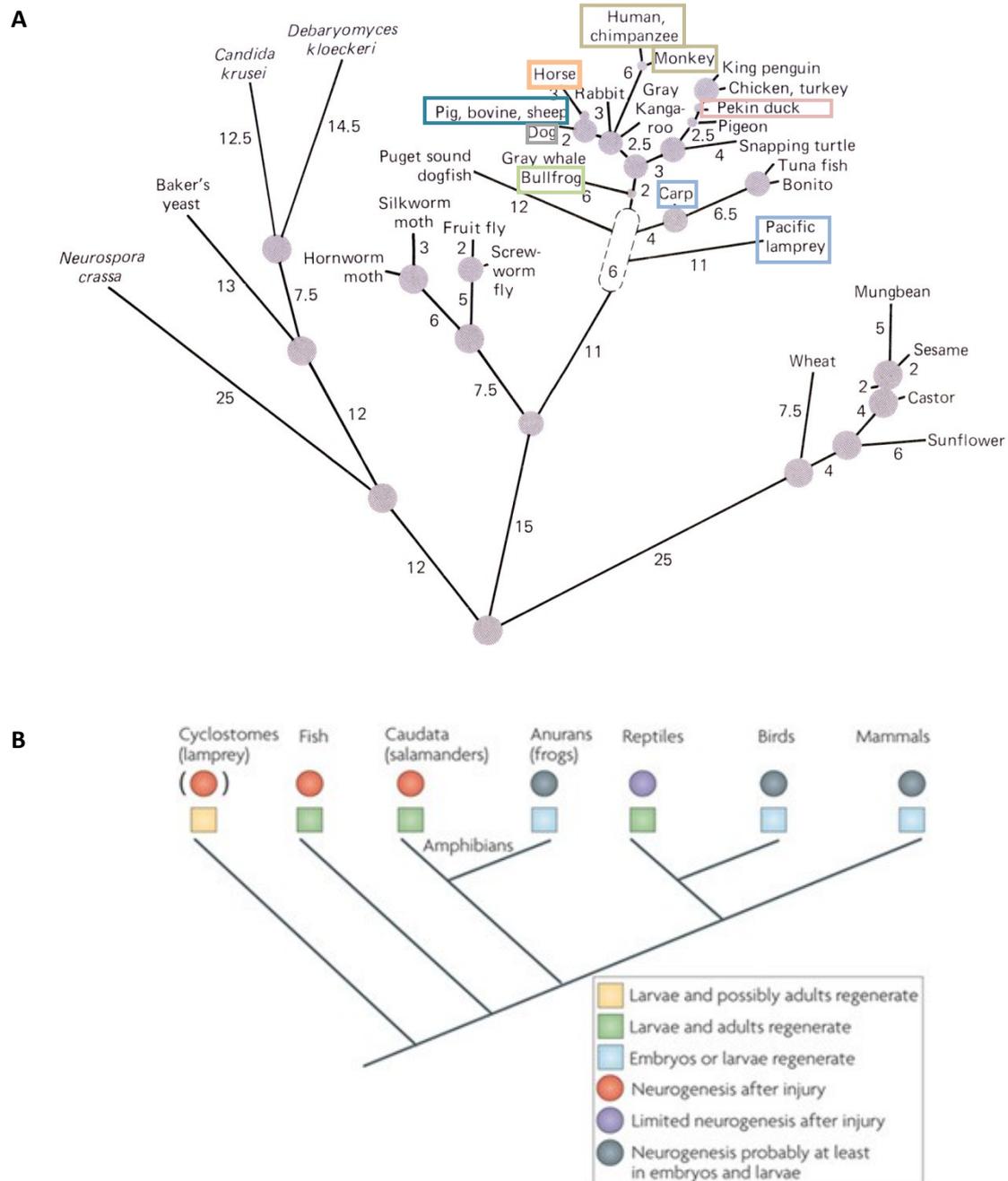


Abbildung 16: Phylogenetische Stammbäume

A Phylogenetischer Baum gebildet auf Basis von der Sequenz von Cytochrom C (biochemisch) in verschiedenen Organismen. Die Anzahl von Aminosäure-Austauschen entlang der Äste ist angegeben (Abbildung übernommen aus [92]). Der Vertreter der Amphibien (Bullfrog) zeigt auf Cytochrom C Basis eine nähere Verwandtschaft zu Reptilien, Vögeln und mehreren Säugern (Pferd, Schwein etc.) als zu Mensch und Maus. Die markierten Organismen sind Abbildung 15 zu entnehmen. **B** Staumbaum der Regenerationsfähigkeit über die Spezies hinweg. Es wurde die Fähigkeit zur Regeneration in den Entwicklungsstadien unter den Organismen in Beziehung zueinander gesetzt (Abbildung übernommen aus [93]).

7. Einsatz von Next-Generation-Sequencing Techniken

Das klassische Sequenzierungsverfahren nach der Sanger Kettenabbruch Methode ist in größeren Projekten sehr kostenintensiv. Daher wurde experimentell überprüft, ob auch das sogenannte Pyrosequenzierungsverfahren für eine Sequenzierung der cDNA Bibliothek in Frage kommt. Dazu wurde in Kooperation mit dem Max-Planck-Institut für molekulare Genetik in Berlin ein Testlauf auf einem Roche 454 GS FLX Sequenzer durchgeführt. Die zentrale Fragestellung in diesem Zusammenhang ist, ob eine *de-novo* Assemblierung von Transkripten ohne Matrize gelingen kann. Für Bakteriengenome gibt es bereits einige *de-novo* Sequenzierungen und Assemblierungen [94,95,96]. Besonders interessant in diesem Zusammenhang war außerdem der Versuch der *de-novo* Transkriptom Sequenzierung bei Wespen (*Polistes metricus*) [97] und dem *Axolotl* (*Ambystoma mexicanum*) [98]. In beiden Fällen wurden in einem 454 GS Lauf mehrere 100.000 Reads generiert. Die mittlere Leselänge lag bei der Wespen Sequenzierung bei 120 Basenpaaren, beim *Axolotl* bei 215 Basenpaaren. Die unterschiedliche mittlere Leselänge hängt von der Verwendung unterschiedlicher Sequenzierungsadapter ab. Selbige werden in regelmäßigen Abständen verbessert und von Roche kommerziell vertrieben. Für Ende 2009 wurde eine Verbesserung angekündigt, die mittlere Leselängen von 400 Basenpaaren verspricht. In keiner der beiden Publikationen [97] und [98] wurden genauere Angaben zu den Ergebnissen bezüglich einer reinen *de-novo* Transkriptom Assemblierung gemacht. In beiden Fällen wurde für die Assemblierung auf bereits bestehende DNA-Sequenz Banken zurückgegriffen.

Im Fall der Wespe wurde aufgrund der kaum vorhandenen Sequenzinformation auf den Sequenzpool der nah verwandten Honigbiene zurückgegriffen. Aus den knapp 400.000 erstellten Reads konnten 3017 Contigs assembliert werden, die zu Bienengenomen ortholog waren. Darüber hinaus gehende, nicht orthologe Contigs wurden nicht veröffentlicht [97].

Im Fall des *Axolotls* wurden die 429.000 erstellten Reads mit den bereits vorhandenen 34683 EST Sequenzen aus klassischer Sanger Kettenabbruch Sequenzierung zusammen zu 61127 Contigs assembliert. Davon hatten jedoch 55626 Contigs eine Gesamtlänge unter 500 Basenpaaren (also quasi nicht assembliert). Eine Sequenzhomologie konnte nur für weniger als 25446 Contigs (40%) gefunden werden.

Im Rahmen dieser Arbeit konnte für alle 454 GS FLX Sequenzen (14644 Reads, Verwendung von Reaktionsrahmen) eine mittlere Leselänge von 226 Basenpaaren erreicht werden. Für diese Sequenzen konnten nach Clipping und Assemblierung 1144 Contigs und 10514 Singletons erstellt

werden. Diese haben eine mittlere Länge von 275 Basenpaaren. Das Contig mit der höchsten Abdeckung beinhaltet 13 ESTs und kodiert für ein ribosomales Protein (*ribosomal protein L10, Xenopus laevis*, E-value $9e-39$).

Somit kann zusammenfassend festgehalten werden, dass sich moderne Hochdurchsatz-Sequenzierungstechniken bisher nicht effizient für Organismen mit großen Genomen und Transkriptomem anwenden lassen. Entsprechende Sequenzierungsläufe produzieren zwar große Datenmengen, diese führen jedoch ohne Matrize zu einem sehr geringen Informationsgewinn. Wenn man die Kosten für einen solchen Hochdurchsatz Sequenzierungslauf mit einbezieht, sind die Kosten für eine klassische Sanger Kettenabbruch Sequenzierung (bei wesentlich höherer Leselänge) bei gleichem Informationsgewinn (Anzahl neuer Transkripte mit Annotation) immer noch deutlich geringer. Um aber eine auf Sanger Kettenabbruch Basis erstellte EST Bank auf Lesefehler und Single Nucleotide Permutations (SNP) hin zu verifizieren, empfehlen sich diese neuen Techniken sehr wohl. Eine Verbesserung der Leselänge bei der Pyrosequenzierung wird in der Zukunft jedoch mit großer Wahrscheinlichkeit zu einer deutlichen Effizienzsteigerung der Hochdurchsatzanalysen führen.

B. Methodenentwicklung Proteomanalyse für indirekt *in vivo* SILAC markierte Organismen am Beispiel der Schwanzregeneration

Die Regenerationsfähigkeit des Molches wurde bislang mit klassischen Methoden der Molekularbiologie auf Proteinebene untersucht (Erstellung von spezifischen Antikörpern durch Immunisierung). Alle öffentlich verfügbaren Sequenzen stammen aus diesen Versuchen. Aufgrund der beschriebenen Einschränkungen auf Sequenzinformationsebene und Limitierungen in der Probenprozessierung (Zellkulturen), wurden bislang keine Hochdurchsatz-Untersuchungsverfahren auf Proteinebene durchgeführt. Im Rahmen dieser Arbeit wurde in Anlehnung an die etablierte SILAC Methode zur Quantifizierung von Proteinen [67] eine neue Methode entwickelt, sequenztechnisch unerforschte Organismen wie den Molch überhaupt im Hochdurchsatz *in vivo* zu untersuchen. Dabei wird der Einbau der schweren Isotope in Aminosäuren in dreierlei Hinsicht genutzt. Erstens dienen die leicht/schweren Peptidpaare (im Folgenden H/L Ratios) zur Annotationsfindung in organismusfremden Datenbanken und hierbei insbesondere zur verbesserten Peptididentifizierung. Zweitens werden die leicht/schweren Peptidpaare zur Identifizierung von Peptiden genutzt, die in einer cDNA Bibliothek gefunden werden können. Drittens soll die partielle Markierung zu einer Methodenentwicklung für

Ergebnisse

quantitative Aussagen verwendet werden. Diese Methode hat zum Ziel, schweres Gewebe als umfassenden normierenden Standard zum Abgleich mit Regenerationszeitpunkten des Molches zu verwenden. Im Folgenden wird die Methodenentwicklung zur grundsätzlichen Proteomanalyse eines genetisch unbekanntem Organismus und insbesondere die Annotation der Proteine am Beispiel des Molches im Einzelnen beschrieben.

Für die Untersuchungen zur Methodenentwicklung sind verschiedene Gewebe und Zeitpunkte der Regeneration zur Verwendung gekommen. Dies begründet sich darin, dass je nach Fragestellung die Wahl der Proben und Gewebe auf höchste Verfügbarkeit und schnellste Überprüfbarkeit hin getroffen wurde. So wurden grundsätzliche Fragen wie „die Möglichkeit zur indirekten *in vivo* Markierung mit Lysin beim Molch oder Annotation von Peptiden durch Nutzung von Datenbanken fremder Organismen“ am regenerierenden Schwanzgewebe des Molches überprüft. Dies hat primär den Grund, dass das Schwanzgewebe des Molches schnell regeneriert, die Probenpräparation wesentlich einfacher ist als beim Herzen und entsprechende Proben in größerem Umfang erzeugt und entnommen werden können.

In den anschließenden Kapiteln VI.C und VI.D werden die Erfahrungen aus den Untersuchungen des Schwanzgewebes auf das Herzgewebe und das regenerierende Herzgewebe des Molches angewandt. Dies liegt primär daran, dass die Gewinnung von regenerierendem Herzgewebe in ausreichender Menge sehr aufwändig ist. Im Folgenden wird jeweils erläutert, an welchem Organismus, welchem Gewebe und welchem Schädigungszeitpunkt die einzelnen Messungen durchgeführt wurden.

1. Markierung des Molches mit stabilen Isotopen essentieller Aminosäuren

Die SILAC Methode basiert auf dem metabolischen Einbau von essentiellen Aminosäuren in Zellkulturen, deren Nährmedien ausgesuchte Aminosäuren enthalten, deren Kohlenstoff und/oder Stickstoffatome mit stabilen Isotopen (^{13}C bzw. ^{15}N) modifiziert sind. Dies führt im Fall des in dieser Arbeit verwendeten Lysin-6 zu einer Zunahme der Molmasse von 6 Dalton (Vergleiche Abbildung 17). Im Rahmen dieser Arbeit wird untersucht, ob stabile Isotope in Anlehnung an die SILAC-Maus [67] auch zur Markierung von Amphibien *in vivo* genutzt werden können. Dazu wurden Molche mit einer Diät aus Lebergewebe von SILAC-Mäusen gefüttert. Nach mehreren Zeitpunkten wurde die Einbaurate von schweren Aminosäuren überprüft. Damit die SILAC Methode verwendet werden kann, ist zunächst zu zeigen, dass die hier verwendete

schwere Aminosäure Lysin, wie bei der Maus, auch für den Molch eine essentielle Aminosäure ist. Dieser Nachweis wurde indirekt *in vitro* geführt. Da eine komplette Markierung des adulten Molches *in vivo* sehr schwierig ist (Dauer bis zur Geschlechtsreife und Reproduktion mind. 2 Jahre), wurden Erkenntnisse aus folgendem Nebenversuch herangezogen.

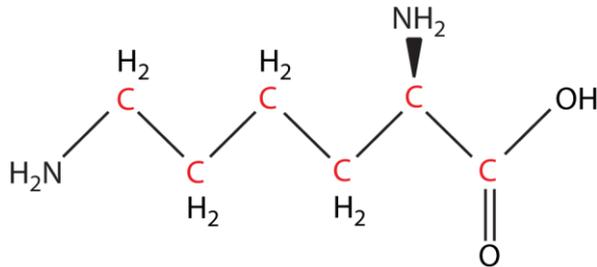


Abbildung 17: Schweres [^{13}C] Lysin

Die Aminosäure Lysin enthält an allen sechs Kohlenstoffatom Positionen (rot) ein schweres Isotop [^{13}C], was im Gegensatz zu normalen [^{12}C] Kohlenstoffatomen zu einer Massenänderung von sechs Dalton führt.

Zellkulturen des Molches aus A1 Zellen (Zellpopulation aus dem Blastem eines regenerierenden Molchbeines, Durchführung der Zellkultivierung und Zellmarkierung Biologe Jia Chen, MPI Bad Nauheim) zeigen bei Einsetzung eines SILAC Nährmediums mit schwerem Lysin nach einer Woche eine durchschnittlich 82% schwere Lysin Markierungsrate, nach zwei Wochen eine durchschnittlich 90% Markierungsrate, nach drei Wochen und nach vier Wochen eine durchschnittliche Markierungsrate von 92%. Vergleicht man die Kinetik dieser Markierung mit publizierten SILAC Ansätzen [67], kann davon ausgegangen werden, das Lysin auch für den Molch eine essentielle Aminosäure darstellt. Andernfalls sollte eine deutliche Menge unmarkierten Lysins in den Proteinen der Zellkultur vorkommen.

Die Methodenentwicklung der *in vivo* SILAC Markierung des Molches wurde nach dem in Abbildung 18 gezeigten Schema geplant und durchgeführt. Ausschließlich mit schwerem [^{13}C] Lysin Gewebe gefütterte Molche werden Gewebeproben ohne Regeneration des Schwanzes und im späten Regenerationszustand „40 Tage“ abgenommen und massenspektrometrisch untersucht. Zur Peptididentifizierung werden die daraus resultierenden Spektren gegen mehrere Proteinsequenzdatenbanken abgeglichen. Aus den Ergebnissen für die beiden untersuchten Gewebestände und durch Vergleich mit verschiedenen Proteinsequenzbanken lassen sich die in Kapitel VI.B.2 bis VI.B.6 dargestellten Rückschlüsse ziehen.

Ergebnisse

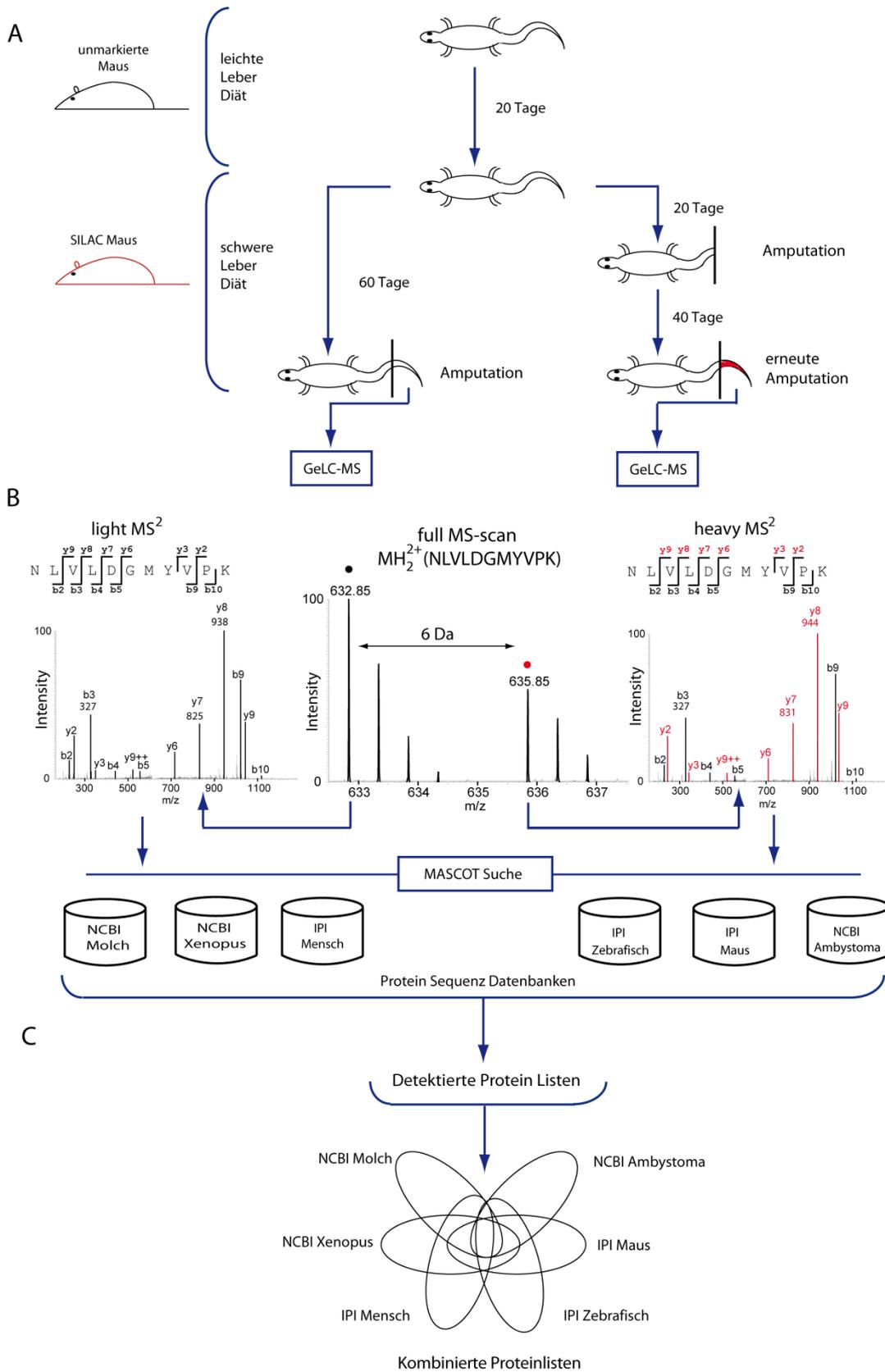


Abbildung 18: Fütterungs- und Auswertungsschema des SILAC Molches: Legende auf der nächsten Seite

A Lebergewebe einer vollständig markierten SILAC Maus wird im drei-Tage Rhythmus als leichte Leber Diät mit [¹²C]-Lysin und als schwere Leber Diät mit [¹³C]-Lysin an Molche verfüttert. Der Gesamtfütterungszeitraum beträgt 80 Tage, der Fütterungszeitraum mit schwerem Lebergewebe 60 Tage. Innerhalb dieser Zeitspanne wurde der Hälfte der Molche nach 20 Tagen Fütterungszeit die Schwanzspitze amputiert. Die Schwanzregeneration fand 40 Tage statt. Anschließend erfolgte die erneute Amputation und massenspektrometrische Untersuchung. **B** Aus den Fullscan Spektren, die Peaks mit einen Massenunterschied von 6 Dalton aufweisen (in der Mitte dargestellt, Ladung 2) werden einzelne Peaks selektiert und fragmentiert (MS²). Diese können sowohl zum leichten Isotopencluster oder dem [¹³C]-Lysin Isotopencluster gehören (leichtes MS² Spektrum links, schweres MS² Spektrum rechts). Eine Vorauswahl der SILAC Peak Paare erhöht die Peptid Identifizierungsrate. Die Identifizierung der Fragmentspektren wird aufgrund des Fehlens von annotierten Datenbanken zum Molch auf mehreren organismusfremden Proteinsequenzdatenbanken ausgeführt. **C** Die in den genutzten Banken identifizierten Peptide werden überlappt und auf Basis der Peptidsequenzen zu Proteingruppen zusammengefasst. Alle identifizierten Proteingruppen bilden Proteinlisten. Einzelne Proteingruppen enthalten Proteinidentifikatoren aus verschiedenen Proteinsequenzdatenbanken (Abbildung verändert nach [99]).

2. Regenerationsprozesse initiieren einen verstärkten Einbau von schweren Isotopen

Nach einer Gesamtfütterungszeit von 60 Tagen liegt die mittlere Markierungsrate von Schwanzgewebe aller detektierten Proteine mit einer H/L Ratio bei 11,5%. Im zugehörigen geschlossenen symmetrischen Intervall der Standardabweichung um den Mittelwert liegen 93,4% aller detektierten Proteine (vergleiche Abbildung 19, schwarze Verteilung). Das Intervall reicht von 0% bis 24% (Im Vergleich dazu liegt im komplexen Herzgewebe nach 60 Tagen eine mittlere Einbaurrate von 15,5% mit einer Standardabweichung von SD=7,6 vor). Wird innerhalb der Fütterungszeit ein regenerativer Prozess initiiert (Schwanzamputation), steigt die Einbaurrate der schweren Aminosäuren massiv an. Hier zeigt sich nach 40 Tagen der Regeneration eine mittlere Markierungsrate aller detektierten Proteine von 46,7%. Das geschlossene symmetrische Intervall der Standardabweichung um den Mittelwert umfasst 86,1% aller detektierten Proteine und reicht von 35,4% bis 58% (vergleiche Abbildung 19, graue Verteilung). Diese Steigerung der Einbaurrate ist mit der verstärkten mitotischen Aktivität des Gewebes zu erklären. Bei der Neubildung von Schwanzgewebe ist der Bedarf an essentiellen Lysin wesentlich höher als bei „ruhendem“ Gewebe. Dieser Bedarf kann durch die Wiederverwendung von bereits im Gewebe eingebautem leichtem Lysin, welches durch katabolische Prozesse bereitgestellt wird, nicht mehr gedeckt werden. Der Bedarf an Lysin wird in metabolischen Prozessen zu 80% aus anderweitig ablaufenden katabolischen Prozessen und zu 20% aus dem Organismus neu hinzugefügtem Lysin gedeckt. Die maximal erreichbare Markierungsrate ist dabei jeweils über die relative Isotopen Häufigkeit (relative isotope abundance (RIA)) limitiert [100]. Abbildung 19

Ergebnisse

zeigt die Verschiebung der mittleren Markierungsrate. Die Abbildung beruht auf den 603 Proteinen, die in der IPI Maus Version 3.37 detektiert wurden. Die übrigen verwendeten Datenbanken Zebrafisch (*Danio rerio*), *Xenopus* und *Ambystoma* zeigen eine vergleichbare Verschiebung der Gesamtmarkierungsrate.

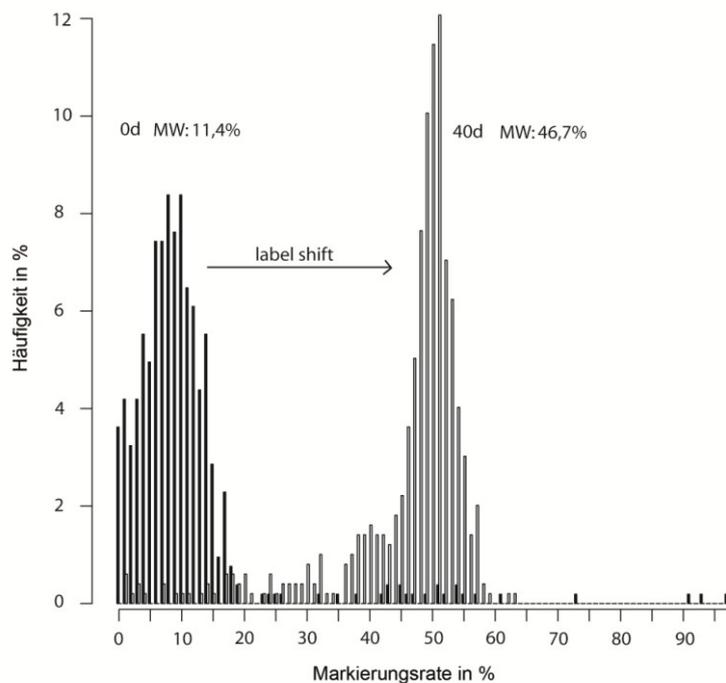


Abbildung 19: Veränderung der mittleren Markierungsrate im Schwanz

Prozentuale Häufigkeitsverteilung der Proteine des nicht regenerierenden Gewebes sind in schwarz, die Proteine des 40 Tage regenerierenden Gewebes sind in grau abgebildet. Während der Regeneration wird der verstärkte Einbau von schweren Isotopen initiiert (Verschiebung der mittleren Markierungsrate). Dargestellt ist auf der x Achse die prozentuale Markierungsrate. Auf der y Achse ist die prozentuale Häufigkeit der jeweiligen Markierungsrate aufgetragen. Während ohne Schädigung nach 60 Tagen Fütterungszeit die mittlere Markierungsrate bei 11,4% liegt, wird sie bei gleicher Gesamtfütterungszeit durch eine Regeneration von 40 Tagen auf die mittlere Markierungsrate von 46,7% angehoben. Insgesamt liegen der Graphik 603 identifizierte Proteine in der IPI Maus Datenbank (Version 3.37) zugrunde.

3. Identifizierung von Peptiden in Proteinsequenzdatenbanken fremder Organismen

Da nur wenige Proteinsequenzen des Molches in öffentlichen Datenbanken zur Verfügung stehen, wurde überprüft, ob Proteindatenbanken aus anderen Organismen (hier organismusfremde Datenbanken genannt) zur Identifizierung von Peptiden genutzt werden können. Dem lag die Idee zu Grunde, dass wie bei der Homologiesuche zwischen Nukleotidsequenzen mittels des BLAST Algorithmus auch massenspektrometrisch aufgefundene, sequenziell identische Peptide eines Proteins dazu genutzt werden können, Proteine zu detektieren und zu identifizieren. Da der Molch bei der Homologiesuche auf Nukleotidbasis starke Ähnlichkeit zu verschiedenen Organismen zeigte, wurde auch auf Proteombasis ein möglichst breiter Ansatz zur Peptid-Homologiesuche verfolgt. Zunächst wurden aus den öffentlich verfügbaren Proteinsequenzen der nächsten Verwandten des Molches wie Salamandern und Fröschen jeweils eine Organismus spezifische Proteindatenbank erstellt. Diese Banken wurden in DECOY [78] Datenbanken transformiert und in ein für die Suchsoftware MASCOT konformes Format gebracht. Da für Maus und Mensch die umfassendsten Proteindatenbanken vorliegen (insbesondere hinsichtlich der weitgehenden sekundären Annotation), wurden diese ebenfalls für die Homologiesuche im Molch genutzt. Hier wurden International Protein Index (IPI) Banken verwendet. Da auch der Zebrafisch als Modellorganismus für Regenerationsforschung gilt, wurde auch eine IPI Proteindatenbank für diesen Organismus verwendet. Für diese Untersuchung kamen die IPI Banken der Version 3.37 zum Einsatz. Um einen quantitativen Vergleich zwischen den Datenbanken, aber auch über die Spezifität der homologen Proteine innerhalb der Treffer in einer Proteinbank anstellen zu können, wurde SILAC markiertes Schwanzgewebe des Molches aus nicht regenerierendem und regenerierendem Schwanzgewebe verwendet.

Die meisten individuellen Proteine wurden in der menschlichen Proteindatenbank gefunden. Zu 838 Proteinen konnten zwei oder mehr homologe Peptide identifiziert werden. Davon zeigten 756 eine H/L Ratio. Dies bedeutet, für diese Peptide konnte eine leichte und eine schwere Form (Abstand 6 Da) detektiert und dieses Paar anhand der Intensität quantifiziert werden. Insgesamt konnten in den verschiedenen Datenbanken 2994 Proteine identifiziert werden. Nach Entfernung redundanter Proteine konnten 1107 unterschiedliche Proteine detektiert werden (Abbildung 20 C). Im Verhältnis zur überschaubar kleinen Anzahl an bekannten Molchproteinen konnte durch diesen Ansatz die Anzahl an detektierbaren Proteinen für den Molch (73 nicht redundante Proteine) um den Faktor 15 erweitert werden (Abbildung 20 A). Die weitere

Ergebnisse

Aufschlüsselung der einzelnen verwendeten Datenbanken sowie weitere Werte zur Anzahl der detektierten Peptide und einzigartig identifizierenden Peptide (als unique Peptide bezeichnet) sind in Tabelle 8 dargestellt.

Tabelle 8: Peptide und Proteine in den organismusfremden Datenbanken

Angegeben sind zu den verwendeten Datenbanken (Spalte 1) jeweils die Gesamtanzahl aller darin identifizierten Peptide (Spalte 2), die Untermenge der identifizierenden (Unique) Peptide (Spalte 3) und die Anzahl der zugehörigen Proteine (Spalte 4). Für Proteine mit mehr als zwei Unique Peptiden wurde eine H/L Ratio gemessen (Spalte 5)

Datenbank	Anzahl Peptide sequenziert gesamt	Anzahl Unique Peptide sequenziert	Anzahl Proteine	Anzahl Proteine mit H/L Ratio und ≥ 2 Unique Peptide
IPI Mensch	4156	2909	838	756
IPI Maus	2626	1938	603	566
IPI Zebrafisch	1966	1299	486	432
NCBI Ambystoma	5284	4618	393	100
NCBI Xenopus	2280	1716	602	546
NCBI Molch	1320	1032	72	42

Weiterhin wurden für jede verwendete Datenbank die jeweils detektierten Proteine in die Gruppen "nur ungeschädigt", "nur in Regeneration" und "in beiden Zuständen" unterteilt. Mit Hilfe dieser Aufteilung wurde untersucht, ob die Detektion von Proteinen in nur einem der untersuchten Zeitpunkte als eine Anreicherung dieser Proteine gedeutet werden kann und gegebenenfalls daher auf eine starke Expressionsveränderung geschlossen werden kann. Die Stärke der einzelnen Gruppen kann der Abbildung 20 B entnommen werden. Die SILAC

Markierung der einzelnen detektierten Peptide ist exemplarisch in Abbildung 21 gezeigt. Zu den Datenbanken Maus, Zebrafisch und *Xenopus* ist jeweils 1 Peptidpaar zu den Zuständen ungeschädigt und in Regeneration dargestellt. Die ausgewählten Proteine *Rab27B*, *Eif4a1b* und *Tagln2* sind für den Molch nicht bekannt.

Bei einer entsprechenden Betrachtung der insgesamt 1107 detektierten, nicht redundanten Proteine (Abbildung 20 C) zeigen 116 Proteine ein ausschließliches Vorkommen im ungeschädigten Zustand, 79 ein ausschließliches Vorkommen im regenerativen Zustand des Gewebes. An dieser Stelle sollte untersucht werden, ob diese Proteingruppen eindeutig einen Schluss auf das Ursprungsgewebe ermöglichen und somit auch von einer massiven Deregulation der Proteine im vorliegenden Versuch ausgegangen werden kann. Dazu wurde auf Basis der beiden Proteingruppen jeweils eine *Gene Ontology* (GO) Analyse durchgeführt. Zu 98 der 116 Proteine der ungeschädigten Gruppe konnte mindestens eine GO Annotation gefunden werden, außerdem zu 63 der 79 Proteine aus der regenerativen Gruppe. Zu jeder Gruppe wurde jeweils eine Analyse auf Anreicherung einzelner GO Terme aus der *biological process* (P) Ontologie mit Hilfe des GO Werkzeuges BINGO [101] durchgeführt. Dazu wurden in beiden Gruppen die ausgewählten Proteine den GO Termen der P Ontologie zugeordnet. Aus den 98 Proteinen der Gruppe „Ungeschädigt mit GO Term Annotation“ hatten 68 eine P Annotation. Aus den 63 Proteinen der Gruppe „Regeneration mit GO Term Annotation“ hatten 41 eine P Annotation. Auf diesen annotierten Termen wurde mit der Anzahl der enthaltenen Proteine ein Hypergeometrischer Test auf Überexpression einzelner Terme im Vergleich zur gesamten bekannten GO Annotationsliste durchgeführt. Die resultierenden GO Terme aus den Proteingruppen „Ungeschädigt“ und „Regeneration“ wurden abermals gegeneinander abgeglichen. Dabei gehörten die 68 Proteine der „Ungeschädigten Gruppe“ zu 79 unterschiedlichen GO Termen. Die 41 Proteine der „Regeneration Gruppe“ bildeten 33 unterschiedliche GO Terme ab. Augenscheinlich ist der in dieser Gruppe überrepräsentierte Term *wound healing*, mit einem p Wert von 0.0037. Diesem gehören drei Proteine an. Dies sind im Einzelnen *Fibronectin 1, Isoform 1 of Alpha-1-Antitrypsin* und *Isoform Gamma-B of Fibronogen Gamma Chain*. 17 GO Termen wurden Proteine aus beiden Gruppen zugeordnet.

Ergebnisse

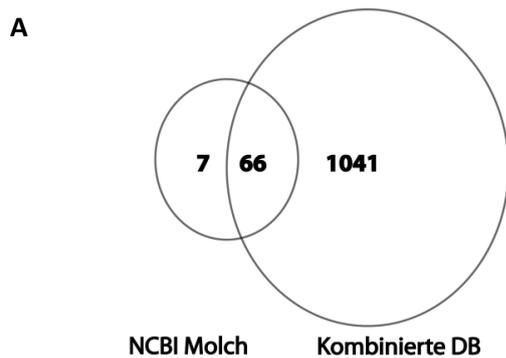
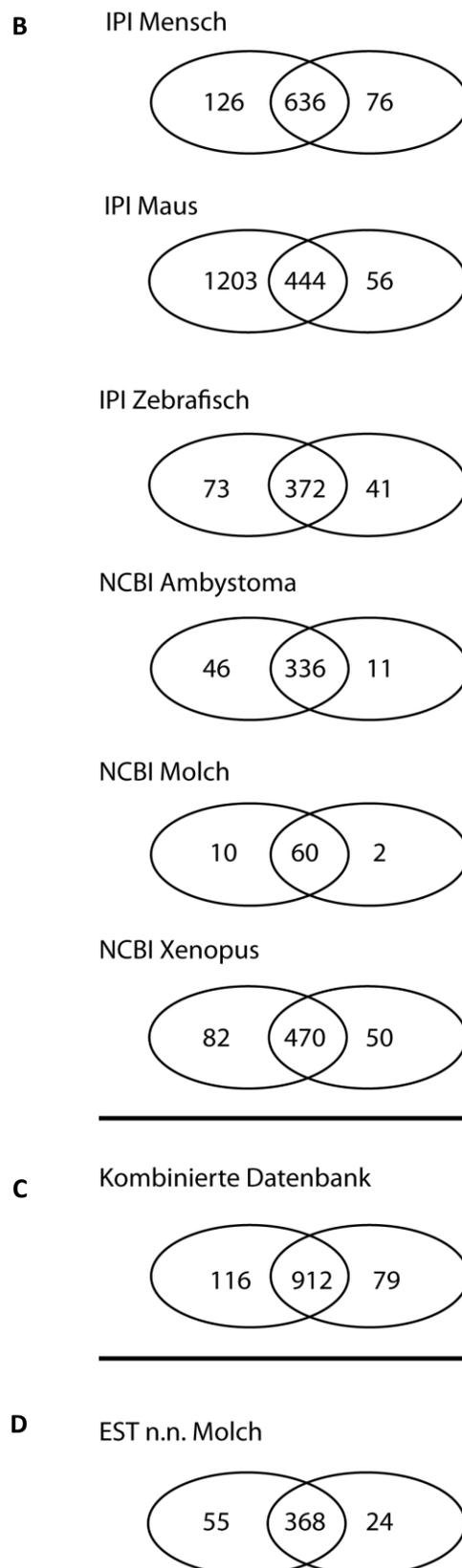


Abbildung 20 über- und nebenstehend: Erweitertes detektierbares Molchproteom

A Das bekannte Molchproteom umfasst 73 nicht redundante Proteine, die eine Mindestlänge von 6 AS haben und mindestens ein Lysin beinhalten. Durch den kombinierten Einsatz von organismusfremden Datenbanken konnten insgesamt 1107 unterschiedliche Proteine identifiziert werden. Peptide zu 7 der insgesamt 73 Proteine des Molches konnten in diesem Versuch nicht identifiziert werden. **B** Dargestellt ist die Anzahl der detektierten Proteine in den Datenbanken Mensch, Maus, Fisch, Ambystoma, Molch und Xenopus. Die Venndiagramme geben jeweils die Anzahl der detektierten Proteine an, die links "nur im ungeschädigten Gewebe", Mitte "in beiden Geweben" und rechts "nur im regenerierendem Gewebe" vorkommen. Die angegebenen Werte beziehen sich jeweils auf die Anzahl der Proteine in der einzelnen Datenbank.

C Das Venndiagramm bezieht sich auf die Kombination aller Datenbanken aus B und enthält keine redundanten Proteine mehr. Die linke Menge gibt die Anzahl der Proteine ausschließlich im ungeschädigten Gewebe, die rechte Menge die Anzahl der Proteine ausschließlich im regenerierenden Gewebe nach 40 Tagen an. **D** Das Venndiagramm zeigt die Anzahl der detektierten Proteine in der hauseigenen nicht normalisierten (n.n.) cDNA Bank. Die linke Menge gibt die Anzahl der Proteine ausschließlich im ungeschädigten Gewebe, die rechte Menge die Anzahl der Proteine ausschließlich im regenerierenden Gewebe nach 40 Tagen an.



Um nun einen Rückschluss auf die tatsächliche Expression von Proteinen ziehen zu können, die nur in einem Gewebeszustand detektiert wurden, wurde eine Reverse-Transkriptase-PCR (RT-PCR) durchgeführt. Als Basis für die Primererstellung diente die hauseigene cDNA Bank. Die Primer wurden mit dem Werkzeug Primer3 [102] entworfen.

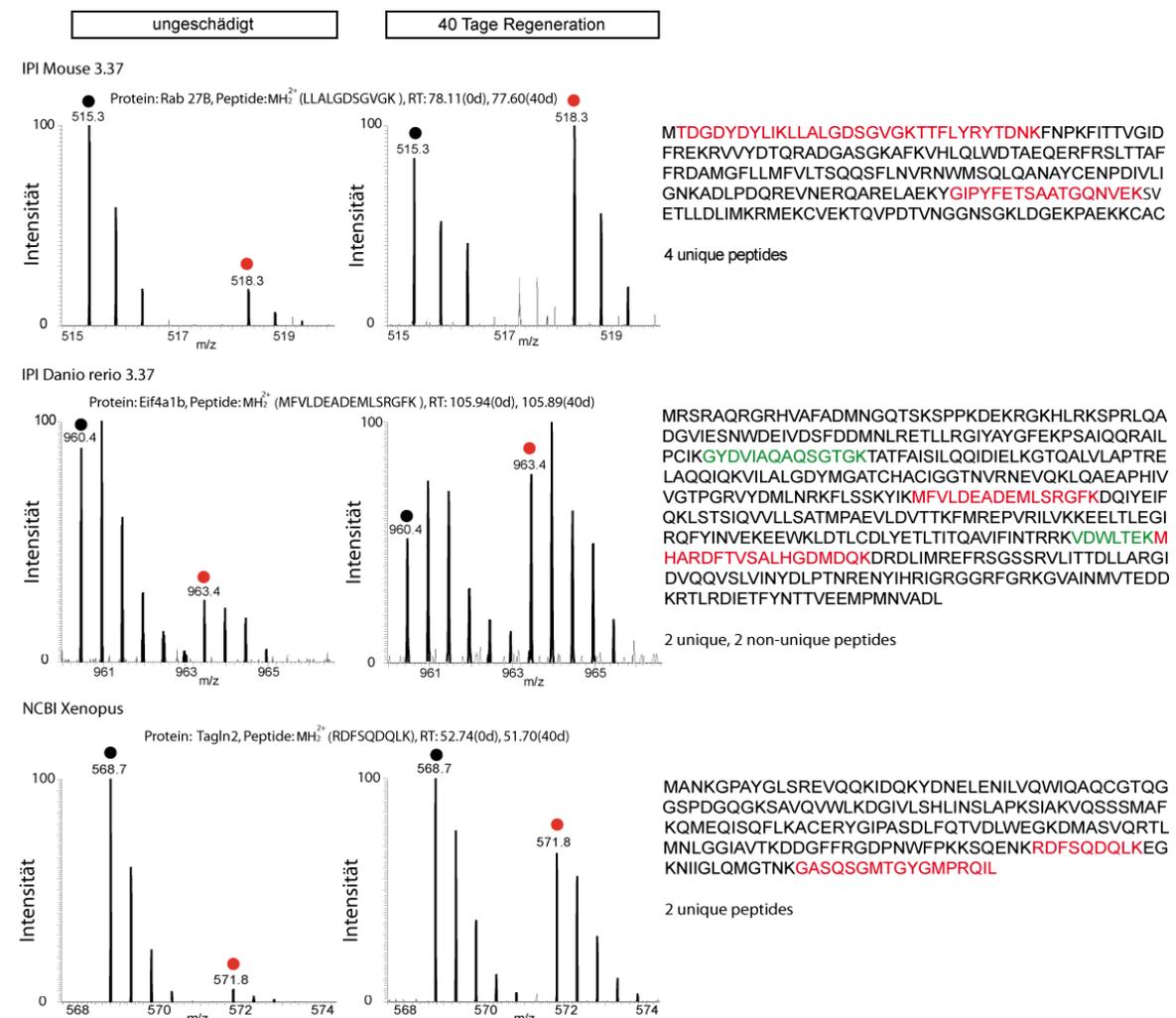


Abbildung 21: SILAC Peptidspektren zu organismusfremden Datenbanken

In der linken Spalte sind die Spektren zu 3 Peptiden im Gewebeszustand ungeschädigt abgebildet. In der mittleren Spalte sind die Spektren der gleichen Peptide im Gewebeszustand Regeneration dargestellt. Rechts ist die zugehörige Proteinsequenz mit farblicher Markierung der detektierten Peptidsequenzen abgebildet. Rote Sequenzmarkierungen zeigen einzigartige Peptide, grüne Sequenzmarkierungen einfache Peptide, die zu mehreren Proteinen passen. In den Spektren ist jeweils der monoisotopische Peak mit einem Punkt markiert. Der schwarze Punkt markiert dabei die leichte Peptidvariante ohne schweres Lysin, der mit einem roten Punkt um 6 Da verschobene Peak das gleiche Peptid mit einem eingebauten schweren Lysin.

Ergebnisse

Zu fünf Proteinen der „Ungeschädigten Gruppe“ und einem Protein der „Regeneration Gruppe“ lagen in der Bibliothek entsprechende EST Sequenzen vor. Der Vertreter der „Regeneration Gruppe“ ist *Fibronectin*, ein Protein das eine Rolle bei der Zelladhäsion spielt. Die übrigen fünf Proteine spielen entweder bei der Regulation oder der Kontraktion der gestreiften Muskulatur eine Rolle. Im Einzelnen handelt es sich um *ATPase Na⁺/K⁺ alpha 1*, ein Membranprotein, das den elektrochemischen Gradienten von Na⁺ und K⁺ Ionen über die Plasmamembran aufrechterhält [103], oder um *ATP2A2*, ein intrazellulärer Ca⁺⁺ Transporter, der zu den *SERCA Ca⁺⁺ ATPasen* zählt [104]. Des Weiteren *Laminin alpha 2*, ein extrazelluläres Protein, das bei der Regulation von Cholin eine Rolle spielt [105]. Das *Hitzeschock Protein 27kDa protein 1* (HSP 27) spielt bei Stressantworten und der Aktin Organisation eine Rolle [106]. Das *mitochondrial carrier, adenine nucleotide translocator, member 4* (Slc25a4) gehört zur Familie der *Solute carrier family 25* und ist eine ADP/ATP Translokase [107].

Die RT-PCR wurde auf isolierter cDNA zu den Regenerationszeitpunkten des Schwanzes 0, 4, 14 und 40 Tage durchgeführt (Durchführung Dr. Thilo Borchardt, MPI-BN). Für die 5 Proteine aus der „Ungeschädigten Gruppe“ konnte jeweils eine Abregulierung von mindestens 50% auf Transkriptomebene gezeigt werden, wobei die maximale Abregulation beim Zeitpunkt 14 Tage nach Schädigung festzustellen war. Zum Zeitpunkt 40 Tage stieg die Menge der Transkripte bereits wieder an. Der Vertreter der „Regeneration Gruppe“, *Fibronectin 1*, zeigte eine 4,5 fache Aufregulierung. Die entsprechenden Banden sind in der Abbildung 22 dargestellt.

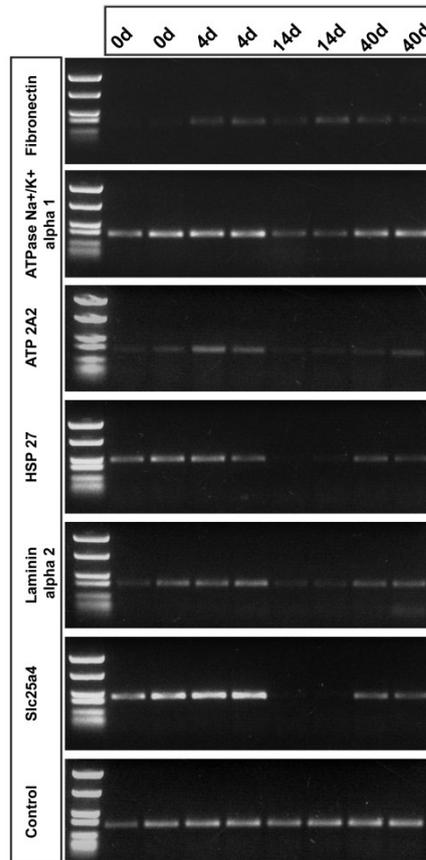


Abbildung 22: RT-PCR zu deregulierten Proteinen aus GO-Analyse

Dargestellt sind jeweils 2 RT-PCR Reaktionen zu den Zeitpunkten 0, 4, 14 und 40 Tage Regeneration des Schwanzes (Jeweils doppelte Ausführung). Die oberste RT-PCR zeigt die Banden für Fibronectin, ein Protein welches nur im regenerierenden Schwanz detektiert wurde. Die übrigen 5 RT-PCR Abbildungen zeigen Banden zu Proteinen aus der Gruppe, die nur in ungeschädigtem Schwanzgewebe identifiziert wurden. Die Primer zu den 6 dargestellten RT-PCRs wurden mithilfe der hausinternen EST Bibliothek erstellt. Fibronectin zeigt bei niedriger Gesamtintensität eine starke Aufregulierung. Die 5 übrigen (ATPase Na^+/K^+ alpha 1, ATP2A2, HSP27, Laminin alpha 2, Slc25a4) zeigen über den Zeitverlauf eine starke Abregulierung. Die Kontrolle wurde mit Primern zum ribosomalen Protein S21 Transkript durchgeführt, welches auf den Microarrays auch als internes Haushaltsgen verwendet wurde.

Somit kann abschließend vermerkt werden, dass, durch GO Analyse und RT-PCR Ergebnisse bestärkt, ein Rückschluss auf die Expression gezogen werden kann, wenn Proteine in einem Versuch nur in einzelnen Stadien detektiert werden. Um hier eine statistisch verlässliche Aussage treffen zu können, müsste der Stichprobenumfang jedoch erweitert werden.

Ergebnisse

Indirekt ist bereits belegt, dass die Verwendung von mehreren organismusfremden Datenbanken zu einer Erhöhung der Gesamtzahl der Proteinidentifizierungen führt. Dieser Sachverhalt soll an dieser Stelle noch einmal näher erläutert werden. Die Identifizierung von Peptiden in organismusfremden Datenbanken ist in der Regel sehr limitiert, da bei strengen Suchparametern nur solche Peptide identifiziert werden können, die nahezu unverändert im Molch und in der Fremddatenbank vorliegen. Publierte Ansätze von Proteomuntersuchungen mithilfe der Massenspektrometrie schlagen daher die Verwendung relaxierter Suchparameter [108] oder von *de-novo* Peptididentifizierung nachgestellten Datenbanksuchen (MS-BLAST) bei sehr nahe verwandten Organismen vor [109]. Da bei der Suche mit relaxierten Parametern die Falsch-Positiv-Rate drastisch steigt und da es für den Molch keine gut annotierten Datenbanken naher Verwandter gibt, wurde in dieser Arbeit ein alternativer Ansatz durch Verwendung einer Kombination mehrerer verschiedener Datenbanken genutzt. Diesem Vorgehen liegt die Annahme zugrunde, dass Proteine partiell konserviert zwischen Molch und jeweilig betrachtetem fremden Organismus vorliegen, und sich diese partiell konservierten Bereiche von Organismus zu Organismus unterscheiden.

Diese Annahme konnte durch eine Analyse der in den einzelnen Banken identifizierten Peptide bestätigt werden. Stellvertretend für drei Proteinsequenzbanken (IPI Mensch, Maus und Zebrafisch) wurden jeweils alle identifizierten Peptide auf Basis ihrer Sequenz gegeneinander abgeglichen. Wie in Abbildung 23 gezeigt, kann durch die Verwendung mehrerer Banken die Anzahl der identifizierten Peptide deutlich gesteigert werden. Beispielsweise führt die Verwendung der drei benannten IPI Datenbanken im Vergleich zur alleinigen Verwendung der IPI Zebrafischdatenbank zu einer Steigerung der Peptid Identifizierungsrate von ~45% (2232 Peptide in IPI Fisch auf 4233 Peptide in allen 3 IPI Banken).

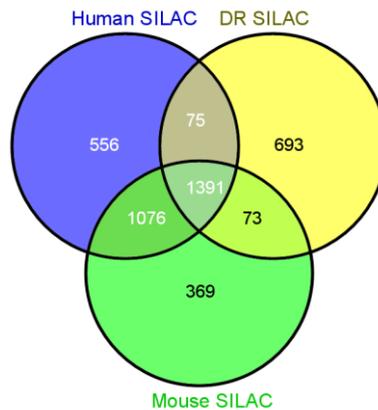


Abbildung 23: Abgleich identifizierter Peptide in drei IPI Proteinsequenzbanken

Dargestellt ist die Überlappung von identifizierten Peptidsequenzen in den IPI Datenbanken Mensch, Maus und Zebrafisch (*Danio rerio*, DR). Durch die Verwendung mehrerer Proteinsequenzbanken ist eine Erhöhung der Gesamtanzahl der identifizierten Peptide zu erreichen. Identische Peptide zwischen Molch und anderen Organismen variieren je nach Datenbank.

In den einzelnen Banken liegt die Identifizierungsrate aller erstellten MS² Spektren, die in den einzelnen IPI Banken gesucht werden, zwischen 3% und 10%. Dabei sind bei der Schwanzexperimenten insgesamt 210041 *Full-MS* Scans erstellt worden. Zu diesen MS Scans wurden 610032 MS² Scans erstellt, von denen 554284 als qualitativ hochwertig für Datenbanksuchen verwendet wurden. In der Proteinsequenzdatenbank IPI Maus wurden beispielsweise 8131 Peptide zu SILAC Peak Paaren und 10998 Peptide zu nicht gepaarten Isotopenclustern identifiziert. Damit ergeben sich für selektierte SILAC Paare eine Identifizierungsrate von 5.68% und für nicht SILAC Paare eine Identifizierungsrate von 3.2%. Für andere Sequenzbanken zeigt sich eine ähnliche gesteigerte Identifizierungsrate für SILAC Isotopencluster Paare. Damit konnte gezeigt werden, dass SILAC-Paare in der Regel um mehr als 1/3 häufiger zu Peptididentifizierungen führen.

4. Verifizierung des bekannten Molchproteoms

Das öffentlich zugängliche Proteom des *N.viridescens* ist mit etwas mehr als 100 bekannten Sequenzen sehr klein. Die bekannten Proteine wurden aus den NCBI Datenbanken ausgelesen und in ein MASCOT konformes Format gebracht. Außerdem wurde die Bank auf Redundanzen überprüft. Nach Überprüfung der Sequenzen konnten 73 unterschiedliche Proteine festgestellt werden, die eine Mindestlänge von 7 AS plus 1 Lysin vorweisen (nach manueller Assemblierung). Zu dieser Bank konnten in der Messung Schwanzregeneration 66 unterschiedliche Proteine detektiert werden (nach manueller Assemblierung, vergleiche auch Abbildung 20 A), die sich auf Peptidebene als 72 unterschiedliche Proteingruppen darstellten (Peptid Assemblierung, vergleiche auch Abbildung 20 B). Zu diesen 72 Proteingruppen zeigten 42 eine H/L Ratio. Somit konnten durch eine Messung gleichzeitig über 90% aller bekannten Molchproteine in ihrem Vorkommen detektiert werden. Die einzelnen detektierten Proteine sind in der Tabelle 15 (siehe Kapitel X.B) hinterlegt.

5. Verwendung von cDNA Nukleotidbanken zur Proteinidentifizierung

Die in Kapitel VI.A.1 erläuterte normalisierte EST Bibliothek und eine weitere, ältere, hauseigene, allerdings nicht normalisierten EST Bibliothek [60] des regenerierenden Molch-Herzens wurden ab dieser Stelle für weitere Protein Annotationen genutzt. Die ältere, nicht normalisierte EST Bibliothek besteht aus ursprünglich 11520 Sequenzen. Nach der Qualitätsanalyse blieben 9698 hochwertige Sequenzen. Diese ergaben nach Assemblierung 2894 Contigs. Diese Contigs wurden wie in Kapitel VI.A.6 beschrieben annotiert. Über 1600 Contigs konnten hochwertig annotiert werden.

Aus der in Kapitel VI.A.1 beschriebenen, neuen normalisierten Bank wurden 52243 Sequenzen sequenziert. Die Auswahl der Sequenzen aus der kompletten Bank wird in Kapitel VI.A.5 beschrieben. Nach Qualitätsanalyse der Sequenzen lag eine Gesamtanzahl von 48537 hochqualitativen Sequenzen mit einer mittleren Leselänge von 591 Basenpaaren vor. Zu diesen ESTs wurde eine Assemblierung, wie in Kapitel VI.A.6 beschrieben, durchgeführt.

Die assemblierten Contigs beider EST Bibliotheken konnten jedoch für Proteomanalysen nicht verwendet werden. Bei der Assemblierung werden bei fehlender Übereinstimmung von Nukleotiden an entsprechender Stelle Platzhalter in die Sequenz eingefügt. Bei mehreren ESTs im Alignment werden je nach Wahrscheinlichkeit für ein Purin, Pyrimidin oder andere Nukleotid-

Kombinationen fest definierte Platzhalter gesetzt. Ist das Nukleotid vollkommen zufällig, wird ein N gesetzt. Grund dafür sind Lesefehler in der Sequenz und SNPs. Bei der Erstellung einer Proteinsequenz Datenbank auf Basis der assemblierten Contigs führen diese Platzhalter zu nicht definierbaren Aminosäure Sequenzen. Mit einer rekursiven Routine wurde versucht zu allen Platzhaltern die wahrscheinlich reale Nukleotidsequenz herzuleiten. Die hohe Anzahl an Platzhalter Nukleotiden führte jedoch bei rekursivem Aufruf zu einer Proteindatenbank mit mehreren Millionen Einträgen, die mit vorliegenden Algorithmen zur Massenidentifizierung (MASCOT) nicht mehr verarbeitet werden konnte.

Daher wurde in einem zweiten Ansatz zur Erstellung der Proteombank auf die vorherige Assemblierung verzichtet. Bei diesem Ansatz soll das Zusammenfügen von zusammengehörigen Transkriptsequenzen nach der Peptididentifizierung anhand selbiger stattfinden. Der schematische Ablauf dieses Ansatzes ist in Abbildung 24 dargestellt. Da das Leseraster der ESTs unbekannt ist, wurden zunächst alle EST Sequenzen in den drei Vorwärts-Leserastern in Proteinsequenzen übersetzt. Da es sich bei den vorliegenden Sequenzen um *Single Reads*, also einzelne Sequenzierungen ohne Fehlerkorrektur, handelt, sind häufig Lesefehler bzw. Leseraster Verschiebungen zu beobachten. Um die Sequenzen dennoch nutzen zu können, wurden für jeden der drei Vorwärts-Leseraster die einzelnen gefundenen offenen Leseraster (*open reading frames, ORFs*) voneinander getrennt und als individuelle Sequenz durchnummeriert. ORFs, die kein Lysin enthalten oder nach *in silico* LysC Verdau eine Länge weniger als sieben Aminosäuren (AS) aufweisen, wurden entfernt. Für die übrigen ORFs wurde aus EST Bezeichnung, Leseraster und laufender ORF Nummer ein Schlüssel generiert, der den Ursprung jedes Peptides eindeutig rückverfolgen läßt. Über diesen Ansatz konnten Peptide, zu denen es eine EST Sequenz ohne Lesefehler gibt, detektiert werden. Die Annotation der einzelnen ORFs, die auf den Contigs beruht, kann über die EST Identifikation und die zugehörige Assemblierungsinformation geschehen. Um statistisch die Generierung vieler nicht wirklich existierender ORFs und deren potentielle Identifizierung abschätzen zu können, wurde auf Basis der gesamten EST ORF Bank wieder eine DECOY Bank erstellt. Diese wurde in ein entsprechendes MASCOT Format gebracht und zur Peptididentifizierung bei Molchgewebe verwendet.

Ergebnisse

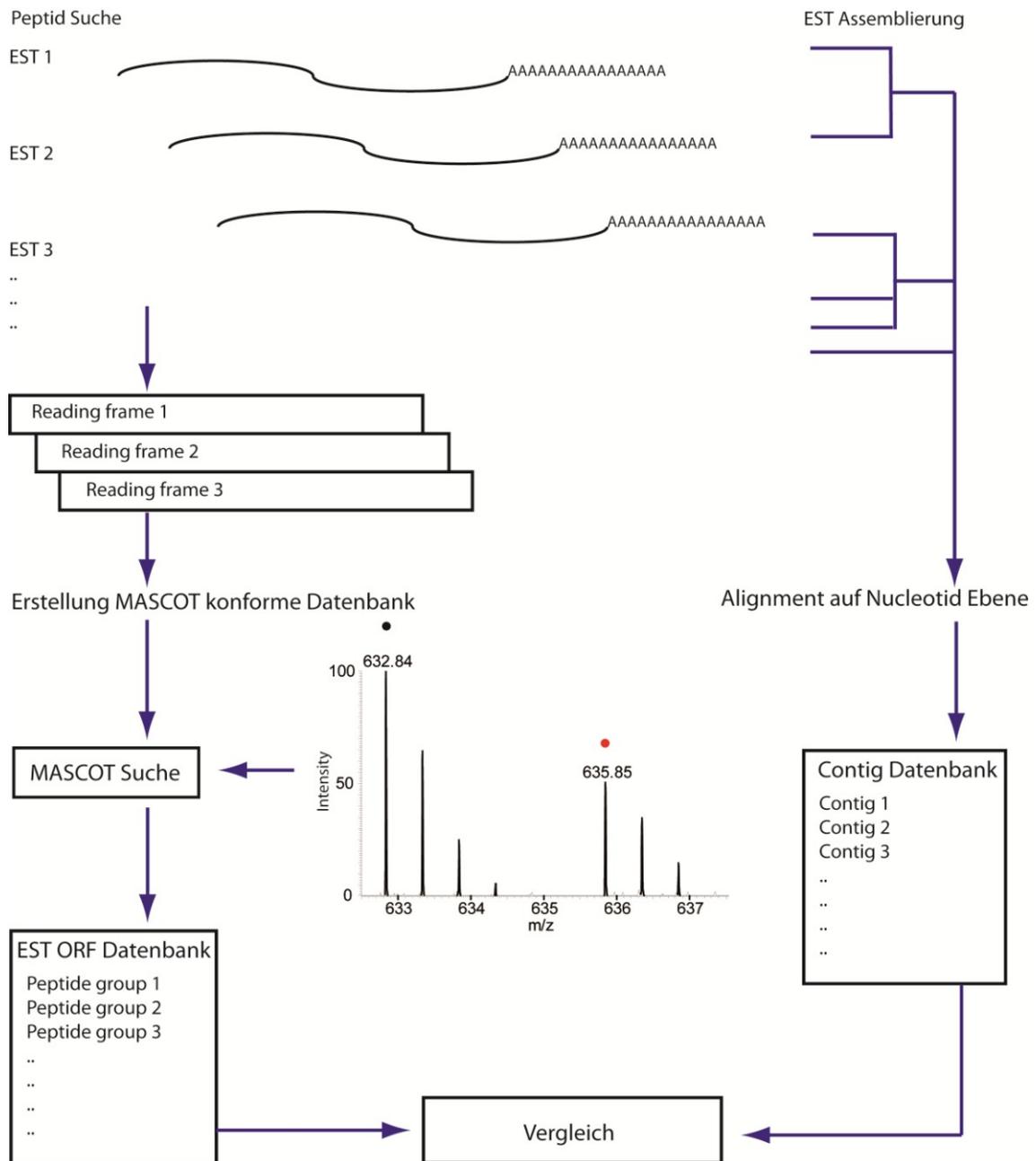


Abbildung 24: Schematischer Abgleich zwischen Nukleotid und Peptid Assemblierung

Die EST Sequenzen durchlaufen wie im Schema dargestellt zwei Wege der Assemblierung. LINKS: Zunächst werden sie in die drei möglichen Vorwärts-Leserahmen übersetzt und aus diesen Peptidsequenzen eine MASCOT konforme DECOY Datenbank erstellt. In dieser Bank werden Peptide identifiziert und dann zu Peptidgruppen zusammengefasst. RECHTS: Die EST Sequenzen werden auf Nucleotidebene assembliert und zu daraus resultierenden Contigs zusammengefasst. Abschließend werden die Peptidgruppen mit den assemblierten Contigs auf deren Zusammensetzung hin verglichen. Die Überschneidung zwischen auf Peptidebene und Nucleotidebene gruppierten ESTs beträgt über 90%.

Die zuvor schon verwendete Beispielmessung Schwanzregeneration wurde erneut genutzt, um die hier dargestellte, zunächst exemplarische Peptidsuche durchzuführen. Für diese Darstellung kam die zu MASCOT konform erstellte Datenbank zur nicht normalisierten, älteren cDNA Bank des Molchherzens zum Einsatz. Diese Bank umfasst ORFs zu 9698 ESTs. Insgesamt wurden bei dieser Suche 447 unterschiedliche Peptidgruppen erstellt, die aufgrund überschneidender Peptidsequenzen assembliert wurden. Zu diesen assemblierten Peptidgruppen wurden über die zuvor erstellten künstlichen Schlüssel die zugehörigen EST Sequenzen ermittelt. Zu jeder einzelnen Peptidgruppe wurden dann die zugehörigen Assemblierungs-Identifikatoren (Contigs) herausgesucht. Im Idealfall sollte dabei zu verschiedenen EST Schlüsseln die gleiche Contignummer gefunden werden.

Von 447 Peptidgruppen bilden 412 Peptidgruppen Peptide aus nur einem auf Nukleotidbasis assemblierten Contig ab. 22 Peptidgruppen bilden zwei Contigs ab. Somit verbleiben 13 Gruppen, die auf mehr als zwei unterschiedliche Contigs abbilden. Eine der größten Peptidgruppen, mit 627 getroffenen Peptiden, bildet die Maximalanzahl von sechs zugehörigen Contigs ab. Bei einer Überprüfung der zugeordneten Annotationen zu diesen sechs Contigs zeigt sich, dass alle für *beta-Globin* kodieren, jedoch aufgrund der kaum überschneidenden Sequenzen nicht zu einem Contig assembliert wurden. Weitere Peptidgruppen, die Peptide für mehrere Contigs enthalten, zeigen Annotationen für das mitochondriale Transkriptom. Insgesamt werden 382 unterschiedliche Contigs durch die 447 Peptidgruppen getroffen.

Zusammenfassend bleibt festzuhalten, dass die Verwendung von alignierten und assemblierten cDNA Contig Datenbanken mit Platzhaltern für Permutationen oder Lesefehlern für Peptididentifikationen nicht möglich ist. Lesefehler, insbesondere bei Contigs mit wenigen zugehörigen ESTs können nicht detektiert oder auch nur verarbeitet werden. Jedoch können cDNA EST Sequenzen nach hier vorgestellter Aufbereitungsmethode erfolgreich für Peptididentifikationen nach massenspektrometrischer Untersuchung verwendet werden. Weiterhin wird durch eine Zusammenfassung von gefundenen überschneidenden Peptiden die Assemblierung auf Nukleotidbasis gut abgebildet (412/447).

6. Identifizierung von unbekanntem Molchproteinen auf Proteomebene

Die Annotation von cDNA Sequenzen erfolgte wie in Kapitel VI.A.6 beschrieben. Dabei konnte für die nicht normalisierte, ältere Bank zu ~1200 Sequenzen keine Homologie detektiert werden.

Ergebnisse

Eine Überprüfung auf enthaltene offene Leserahmen mit der Mindestlänge 22 AS führte zu einer Anzahl von 174 potentiellen unbekanntem Proteinen, die nur im Molch vorkommen oder noch nicht bei anderen Organismen detektiert worden sind. Ob es sich aber um real kodierende Sequenzen handelt, die auch translatiert werden, ist auf cDNA Ebene nicht festzustellen. Durch die Verwendung der cDNA EST Bank für Peptididentifizierung in beschriebener Form konnte beim Schwanzregenerationsexperiment für neun *in silico* translatierte ESTs ohne Annotation mindestens ein qualitativ hochwertiges Peptid detektiert werden. Exemplarisch sind in Abbildung 25 drei dieser EST Sequenzen jeweils mit Peptidspektrum und RT-PCR Banden dargestellt. Die identifizierten Peptide zeigen jeweils das C-terminale LysC Schnittmuster. Die RT-PCR Banden zeigen das Vorkommen des Transkriptes im Schwanzgewebe, wenn auch mit einer niedrigeren Kopienzahl als im Herzgewebe (vergleiche ebenfalls Abbildung 25). Das ausgewählte Contig 1556 zeigt außerdem auf Transkriptomebene im Herzgewebe eine starke Aufregulierung in der späteren Regenerationsphase ab 21 Tage (Array Experiment). Weiterhin zeigen die EST Sequenzen alle einen Poly A Schwanz und ein Polyadenylierungssignal.

Zusammengefasst kann davon ausgegangen werden, dass es sich bei diesen Kandidaten um reale, neue bzw. bislang in keinem anderen Organismen identifizierte Proteine handelt, obwohl für nahe Verwandte des Molches wie beispielsweise *Xenopus* und *Ambystoma* bereits sehr umfassende Nukleotid Datenbanken vorliegen (z.B. zum *Xenopus* NCBI 1971412 ESTs). Auch in diesen wurden keine homologen Sequenzen detektiert. Weitere Untersuchungen müssen zeigen, ob hier sogar von *N. viridescens* spezifischen Proteinen ausgegangen werden kann.

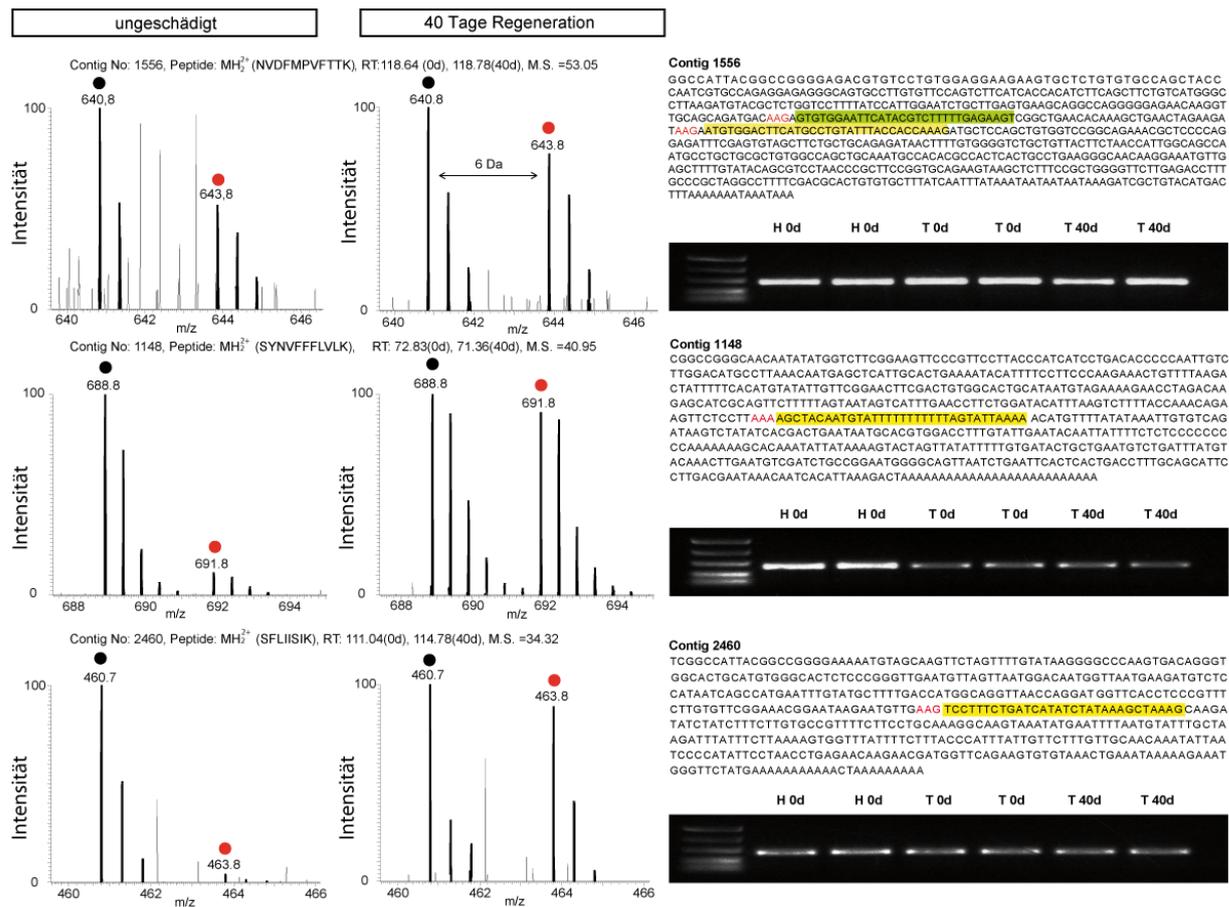


Abbildung 25: Identifizierung unbekannter Proteine aus EST Sequenzen

Dargestellt sind H/L Peptidpaare dreier bislang unbekannter Proteine zu den Zeitpunkten ungeschädigt (links) und 40 (rechts) Tage Regeneration. Durch einen schwarzen Punkt sind die monoisotopischen Peaks der leichten Peptide gekennzeichnet, durch einen roten Punkt die monoisotopischen Peaks der Peptide mit schwerem Lysin Einbau. Die Isotopencluster sind entsprechend um 6 Da verschoben. Rechts sind die zugehörigen EST Sequenzen abgebildet. Das identifizierte Peptid ist gelb hinterlegt, in rot das C-terminale LysC Schnittmuster, in grün weitere identifizierte Peptide. Alle Sequenzen zeigen einen Poly A Schwanz und Polyadenylierungssignal. Außerdem sind jeweils rechts unter der Sequenz die entsprechenden Banden aus einer RT-PCR abgebildet (Agarose Gel, Ethidiumbromid Färbung). H 0d zeigt die RT-PCR Bande in ungeschädigtem Herzgewebe, T 0d und T 40d jeweils die Banden aus ungeschädigtem und regenerierendem Schwanzgewebe. Zu allen Geweben wurden RT-PCR Reaktionen aus zwei individuellen Gewebepools durchgeführt.

An dieser Stelle waren die Vorversuche am Schwanzgewebe, die primär der Methodenentwicklung dienten, abgeschlossen. Im nächsten Schritt sollten die erarbeiteten Methoden bei Untersuchungen am Herzgewebe, insbesondere am regenerierenden Herzgewebe verwendet werden. Für diese Untersuchungen kam dann auch die im Rahmen dieser Arbeit mitgestaltete, neue normalisierte cDNA Bibliothek für Peptididentifizierungen zum Einsatz, die

Ergebnisse

bereits aufgrund ihrer Mehrzahl an Transkripten eine größere Anzahl an Proteinidentifizierungen und insbesondere dadurch auch eine größere Anzahl an potentiell neuen Proteinen erwarten lässt. Abschließend sei vermerkt, dass eine Suche nach den detektierten Peptidmassen aus dem Experiment der Schwanzregeneration auf der neuen normalisierten Datenbank insgesamt 2449 Proteingruppen identifizieren konnte (1776 mit mind. 2 Peptiden), von denen 34 Stück auf Transkripte verwiesen, die keine Annotation zeigten und somit wahrscheinlich für neue, bislang unbekannte Proteine kodieren. Diese und weitere Kandidaten werden im Zusammenhang mit den detektierten neuen Proteinen aus einem Herzregenerationsdatensatz in Abschnitt VI.D.2 genauer dargestellt.

C. Quantifizierung auf Proteinebene mit partiell markierten SILAC Gewebepools in isolierten Kardiomyozyten

Das Quantifizieren von Proteingemischen mit der SILAC Methode wurde bisher nur in markierten Zellkulturen und *in vivo* nur in der vollständig markierten SILAC Maus durchgeführt [67]. Die Möglichkeit der Quantifizierung mit teilmarkierten Proteingemischen ist neu und soll daher in dieser Arbeit zunächst für homogene Zellkulturen gezeigt werden. Die Entwicklung dieser veränderten Methodik zur Quantifizierung wurde in Zusammenarbeit mit Frau Dipl. Biologin Anne Konzer (MPI Bad Nauheim) durchgeführt. Die grundsätzliche Frage zur Möglichkeit einer Quantifizierung mit teilmarkierten SILAC Protein Gemischen wurde an Kardiomyozyten der Maus untersucht, da hier ausreichend Material zur Verfügung steht und bei der Methodenentwicklung zur Quantifizierung auf entsprechende Expertise im Zusammenhang mit der Aufbereitung von Mausgewebeprobe zurückgegriffen werden konnte. Auch konnte durch die hohe Annotationsdichte von Mausproteinen eine größere Proteinanzahl hinsichtlich ihres Expressionsverhaltens überprüft werden. Und letztendlich kann durch die Verwendung des Mausgewebes bei der Methodenentwicklung zur Quantifizierung mit teilmarkierten SILAC Protein Gemischen auch auf definiert hergestellte SILAC Gemische zurückgegriffen werden, die sich aus der vollständig markierten Maus erstellen lassen. Das komplette Versuchsschema kann der Abbildung 26 entnommen werden. Zur Bewertung der Quantifizierung kamen Kardiomyozyten (Herzmuskelzellen) aus normalen, vollständig SILAC-markierten Mäusen und Kardiomyozyten aus „alten“, nicht markierten Wildtypmäusen zum Einsatz. Die daraus abgeleitete vergleichende Quantifizierung wird als Beleg für die technische Durchführbarkeit von Quantifizierungsaussagen auf Basis von teilweise markierten SILAC-Pools herangezogen.

Zur Durchführung der quantitativen Vergleiche wurden jeweils aus Kulturen von isolierten Kardiomyozyten der vollständig markierten SILAC Maus und einer nicht markierten Wildtyp Maus je ein Proteinextrakt erstellt. Die Konzentration der komplexen Proteingemische wurde jeweils nach Bradford bestimmt. Aus den beiden Proteingemischen wurden anhand der bestimmten Konzentrationen jeweils drei Pools mit der SILAC-Markierungskonzentration 100%, 70%, 50% und 30% durch Mischen erstellt (vergleiche Abbildung 26). Die erstellten Pools haben somit für jedes enthaltene Protein die gleiche, definierte H/L Ratio.

Von diesen zwölf Pools (4 Konzentrationen, dreifache Bestimmung) wurde abermals jeweils die Proteinkonzentration nach Bradford bestimmt. Im nächsten Schritt wurde jeder der zwölf Pools im Verhältnis 1:1 mit einer Proteinaufbereitung einer Kultur von isolierten „alten“ Kardiomyozyten einer nicht markierten Maus gemischt. Deren Konzentration wurde ebenfalls nach Bradford bestimmt. „Alte“ Kardiomyozyten bedeutet dabei eine Isolation der Zellen aus einer „alten“ Maus (>24 Monate). Die zwölf Proben wurden jeweils massenspektrometrisch gemessen.

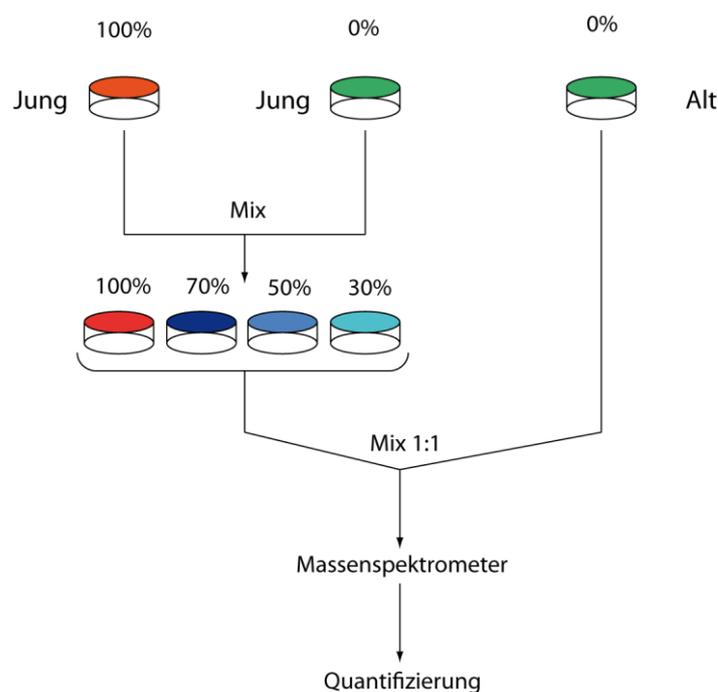


Abbildung 26: Versuchsschema partiell markierte SILAC Pools

Im Versuch wurde ein SILAC Pool 100% markierter junger Kardiomyozyten und ein Pool unmarkierter junger Kardiomyozyten (0%) zur Herstellung von definierten SILAC Pools der Konzentration 100%, 70%, 50% und 30% verwendet. Diese Pools wurden jeweils mit nicht markierten „alten“ Kardiomyozyten im Verhältnis 1:1 gemischt, gemessen und dann auf Basis des schweren Pools quantifiziert.

Ergebnisse

Die in Tabelle 9 dargestellten Ergebnisse dieser Untersuchung zeigen, dass unabhängig von der prozentualen Markierungsrate des verwendeten schweren SILAC-markierten Proteingemisches eine reproduzierbare Quantifizierung erreicht werden kann. Dabei wird die Quantifizierung mit dem 100% markierten Pool, was einer etablierten Methode entspricht [67], als Referenzwert angenommen. Zunächst wurden zu jedem Gemisch die drei technischen Replikate miteinander verrechnet und daraus die Gruppe der dreifach eindeutig detektierten und identifizierten Proteine bestimmt. Insgesamt wurden Peptide zu 357 unterschiedlichen Proteinen detektiert. Nach Qualitätsanalyse der identifizierten Peptide konnten insgesamt 260 unterschiedliche Proteine detektiert werden. Der 100% SILAC-markierte Pool umfasste 180 Proteine, der 70% SILAC-markierte Pool umfasste ebenfalls 180 Proteine, der 50% SILAC-markierte Pool umfasste 184 Proteine und der 30% SILAC-markierte Pool 200 Proteine.

Tabelle 9: Mittelwerte und normalisierte Mittelwerte der SILAC-Markierung in verdünnten Proteingemischen

Die Tabelle zeigt den Mittelwert der gemessenen SILAC-Markierung aller Proteine in den definiert erstellten, komplexen Proteingemischen reiner SILAC Pool, auf 70%, 50% und 30% verdünnter SILAC Pool. Die tatsächlich errechneten Werte zeigen eine deutliche Kongruenz mit den jeweiligen Erwartungswerten, die bei 1, 0.7, 0.5 und 0.3 liegen. Die unteren Zeilen zeigen den Mittelwert der detektierten Proteingruppen vor und nach Normalisierung an.

Mittelwert	Pool 100%	Pool 70%	Pool 50%	Pool 30%
Nicht normalisiert	1.083	0.567	0.46	0.333
normalisiert	1.179	1.12	1.093	1.012

Zu diesen Proteingruppen wurde jeweils die mittlere H/L Ratio bestimmt. Hier lag der Erwartungswert für die einzelnen Gruppen jeweils bei der eingesetzten SILAC Markierungsrate der Proteingemische, also bei 1, 0.7, 0.5 und 0.3. Die gut korrelierenden, tatsächlich gemessenen Werte sind der Tabelle 9 zu entnehmen. Eine graphische Darstellung der Verteilung der tatsächlich gemessenen H/L Ratios zeigt Abbildung 27. Um eine Vergleichbarkeit der Quantifizierungskoeffizienten der einzelnen Proteingruppen herzustellen, wurde anschließend

die Schnittmenge der in allen Gruppen detektierten Proteine erstellt. Diese Gruppe umfasst 149 unterschiedliche Proteine (Abbildung 28). Alle weiteren Berechnungen wurden innerhalb dieser Schnittmenge von Proteinen durchgeführt und die auf die Poolkonzentration normalisierten H/L Ratio Werte verwendet. Der Mittelwert der normalisierten H/L Ratios der Proteingruppen kann ebenfalls aus Tabelle 9 entnommen werden.

Abbildung 27: Häufigkeitsfunktion „verdünnte SILAC Proteingemische“

Dargestellt ist die interpolierte Häufigkeitsfunktion über die absolute Anzahl an Proteinen in einer H/L Ratio Gruppe. Übereinandergelegt sind die Messungen aus den jeweiligen definiert hergestellten schweren Pools von 30-100% Lysin im 1:1 Gemisch mit leichten, „alten“ Kardiomyozyten. Bezogen auf die H/L Ratio zeigt der 100% Pool die breiteste Streuung, der 30% Pool die kleinste Streuung. Der Mittelwert jeder Kurve liegt nahe am errechneten Mittelwert, der das Mischungsverhältnis beschreibt. Blau: 100% Pool, Rot: 70% Pool, Grün: 50% Pool, Lila 30% Pool.

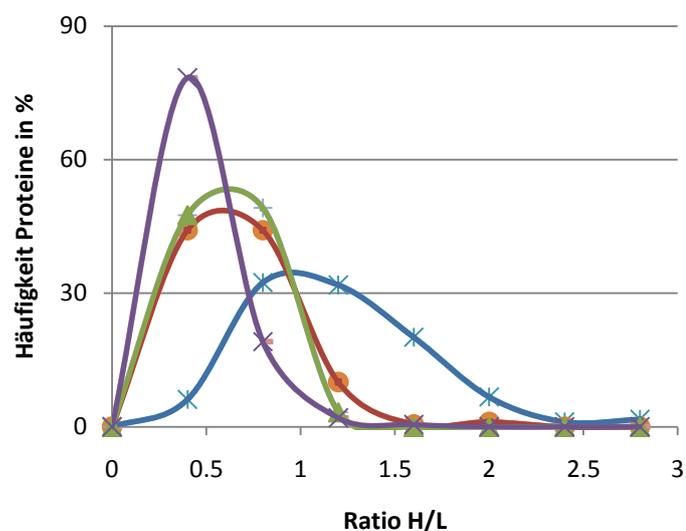
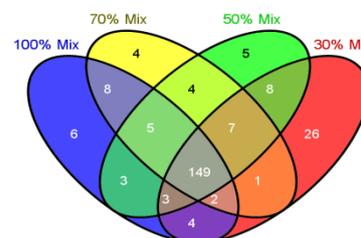


Abbildung 28: Venn-Diagramm der detektierten Proteine aus „verdünnte SILAC Proteingemische“

Um eine Vergleichbarkeit der erstellten verdünnten SILAC-Pools zu ermöglichen, wurde die Gesamtanzahl der detektierten Proteine in den verschiedenen Pools auf die Schnittmenge reduziert (149 Proteine), die in allen Pools detektiert wurden.



Um die Qualität der Quantifizierung mit niedrigen SILAC markierten Proteingemischen bestimmen zu können, wurde zunächst der Korrelationskoeffizient zwischen den einzelnen Pools

Ergebnisse

berechnet. Diese können Tabelle 10 entnommen werden. Auffällig ist die hohe Korrelation zwischen den weitest entfernten SILAC Pools von 100% und 30% SILAC Markierung, der bei $r = 0.97$ liegt (sehr hohe Korrelation). Dargestellt ist diese Korrelation in Abbildung 29. Neben der generellen Korrelation zwischen den Proteingruppen soll die Reproduzierbarkeit von stark deregulierten Proteinen mit den verschiedenen stark markierten Pools gezeigt werden. Dieser Schritt wurde durchgeführt, da gute Korrelationswerte häufig auf die Masse der nicht deregulierten Kandidaten aus Hochdurchsatzexperimenten zurückzuführen ist. Dazu wurden jeweils die Untergruppen der in diesem Experiment aufregulierten Proteine mit einer H/L Ratio >1.5 bestimmt (nicht logarithmiert). Die Anzahl der zu diesem statischen Schwellenwert aufregulierten Proteine sind in Tabelle 11 dargestellt. Diese Gruppen deregulierter Proteine wurden im nächsten Arbeitsschritt gegeneinander abgeglichen. Von maximal 13 aufregulierten Proteinen waren zwölf in allen Gruppen vertreten. Weiterhin waren drei Proteine jeweils in drei Gruppen vorhanden. Diese Untersuchung wurde mit den abregulierten Proteinen wiederholt. Hier wurde der nicht logarithmierte Schwellenwert für Abregulation auf 0.66 gesetzt. Die gemeinsame Gruppe der abregulierten Proteine umfasste 44 von maximal 46 Proteinen. Dazu kommen neun Proteine, die noch in drei Gruppen einheitlich dereguliert waren. Das zugehörige Venn-Diagramm in Abbildung 30 verdeutlicht die genannten Zahlenverhältnisse.

Tabelle 10: Korrelation der verdünnten SILAC Proteingemische

Die definiert erstellten SILAC-Protein-Gemische wurden jeweilig auf deren Korrelation überprüft. Dazu wurden basierend auf den überall detektierten Proteinen (149) und der normalisierten H/L Ratios jeweils die Korrelationskoeffizienten berechnet, die alle deutlich über 0.9 lagen. Den Vergleich mit der höchsten SILAC-Verdünnung stellt die Kombination 100% Pool zu 30% Pool dar, der dennoch eine vergleichbar hohe Korrelation mit den anderen weniger stark verdünnten Pools zeigt.

	Vergleich 100%:70%	Vergleich 100%:50%	Vergleich 100%:30%	Vergleich 70%:50%	Vergleich 70%:30%	Vergleich 50%:30%
Korrelations- koeffizient	0.990804	0.991256	0.968371	0.997319	0.947863	0.9547

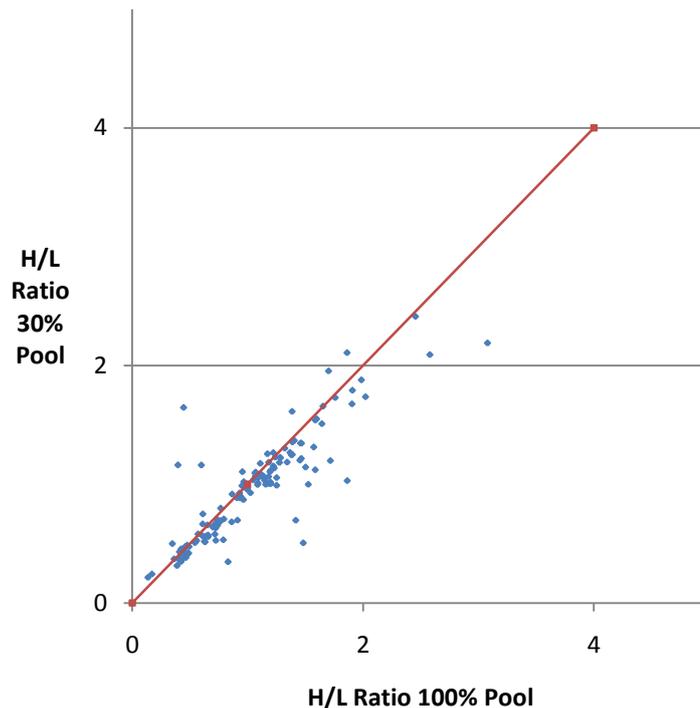


Abbildung 29: Korrelation der SILAC Proteingemische 30% und 100%

Dargestellt sind die normalisierten H/L Ratios der Pools 100% SILAC Proteine (x-Achse) und des 30% verdünnten SILAC Protein Pool (y-Achse). In rot ist die „ideale Korrelation“ dargestellt (Korrelationskoeffizient 1). Der berechnete Korrelationskoeffizient liegt bei 0.97, es handelt sich also um eine sehr hohe Korrelation.

Zusammenfassend kann festgehalten werden, dass ausgehend von einem voll SILAC markiertem Proteingemisch die prozentuale Isotopenmarkierung von 100% bis 30% nahezu identische quantitative Aussagen erlaubt. Dies kann doppelt genutzt werden.

Zum einen kann bei der Arbeit mit 100% SILAC markierten Proteinen als Quantifizierungsstandard auch eine geringere SILAC Konzentration verwendet werden. Dies hat unmittelbar einen hohen ökonomischen Wert, da die Markierung mit stabilen Isotopen *in vivo* zu 100% mit deutlich höherem finanziellem Aufwand verbunden ist. Durch Verdünnung der Ausgangsmaterialien können bis zu dreimal mehr Messungen mit der gleichen Menge SILAC Standard Gewebe erfolgen als bisher.

Zum anderen ist belegt, dass die Teilmarkierung von Proteingemischen dazu genutzt werden kann, teilmarkierte Zellkulturen oder Organismen als SILAC Standard zur Quantifizierung von Proteinen zu verwenden. Dies ist besonders interessant bei Organismen, für die keine 100%

Ergebnisse

Markierung erreicht werden kann. Damit ist die in dieser Arbeit entwickelte indirekte Markierungsmethode vor allem für die Regenerationsforschung relevanten Organismen, wie Molch oder Zebrafisch, anwendbar. Durch Mischung von unmarkierten, im Mittelpunkt des Interesses stehenden Gewebes (z.B. regenerierendes Herzgewebe), mit markierten Proteinpools kann neben der bereits dargestellten verbesserten qualitativen Proteindetektion zusätzlich eine Quantifizierungsmethode abgeleitet werden. Dabei ist der Umweg über eine separate Messung des schweren Pools notwendig, deren Ziel es ist, für möglichst viele Proteine eine Standardisierungs H/L Ratio zu detektieren, die dann zu Quantifizierungszwecken herangezogen werden kann. Beispielhaft ist dieser Ansatz in Zusammenhang mit der Herzregeneration in Kapitel D.2 beschrieben.

Tabelle 11: Deregulierte Proteine in verdünnten SILAC Proteingemischen

Die Tabelle zeigt die absolute Anzahl der als dereguliert angenommenen Proteine (normalisierte H/L Ratio > 1.5 oder < 0.66, nicht logarithmiert) in den untersuchten Proteinpools, die durch 1:1 Mischung von Proteinextrakten aus „alten“ Kardiomyozyten und zuvor erstellten, definiert verdünnten SILAC Proteinpools erstellt wurden.

Regulation	100% normalisiert	70% normalisiert	50% normalisiert	30% normalisiert
Aufreguliert > 1.5	21	13	16	16
Abreguliert < 0.66	46	54	51	54

Ergebnisse

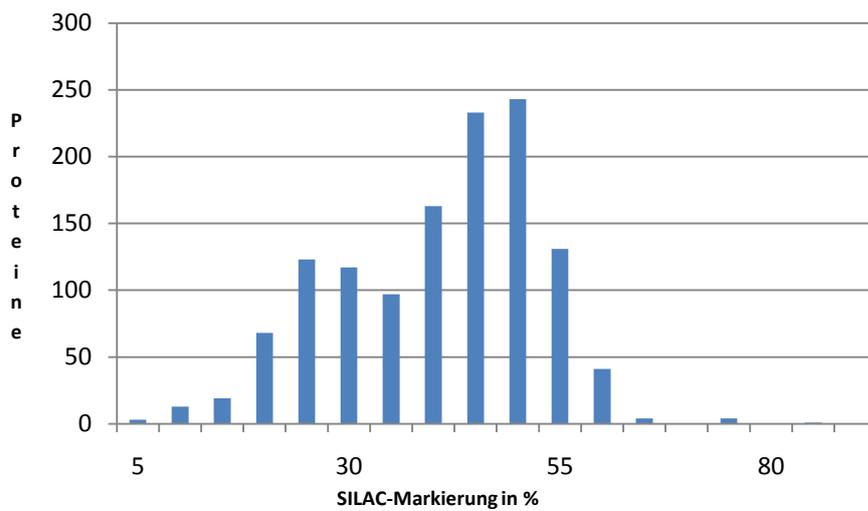


Abbildung 31: Verteilung des SILAC-markierten Proteinpools

Die x Achse zeigt die prozentuale Markierung von detektierten Proteingruppen mit [$^{13}\text{C}_6$] Lysin, jeweils zu Gruppen von 5% zusammengefasst (bins). Die y Achse zeigt die absolute Anzahl der detektierten Proteine in der jeweiligen Gruppe. Die mittlere Markierungsrate liegt bei 38,1%. Es liegen 1260 Proteine zu Grunde, die in der normalisierten Molch cDNA Bank detektiert wurden.

Dieser schwere komplexe Proteinpool wurde nun als Standard verwendet, indem er im Verhältnis 1:1 zu leichtem, regenerierendem Herzgewebe hinzugefügt wurde. Als zu untersuchende Zeitpunkte wurden normal gefütterte Molche mit ungeschädigtem Herzen sowie normal gefütterte Molche nach sechsständiger Herzregeneration gewählt. Das Versuchsschema ist aus Abbildung 32 zu entnehmen.

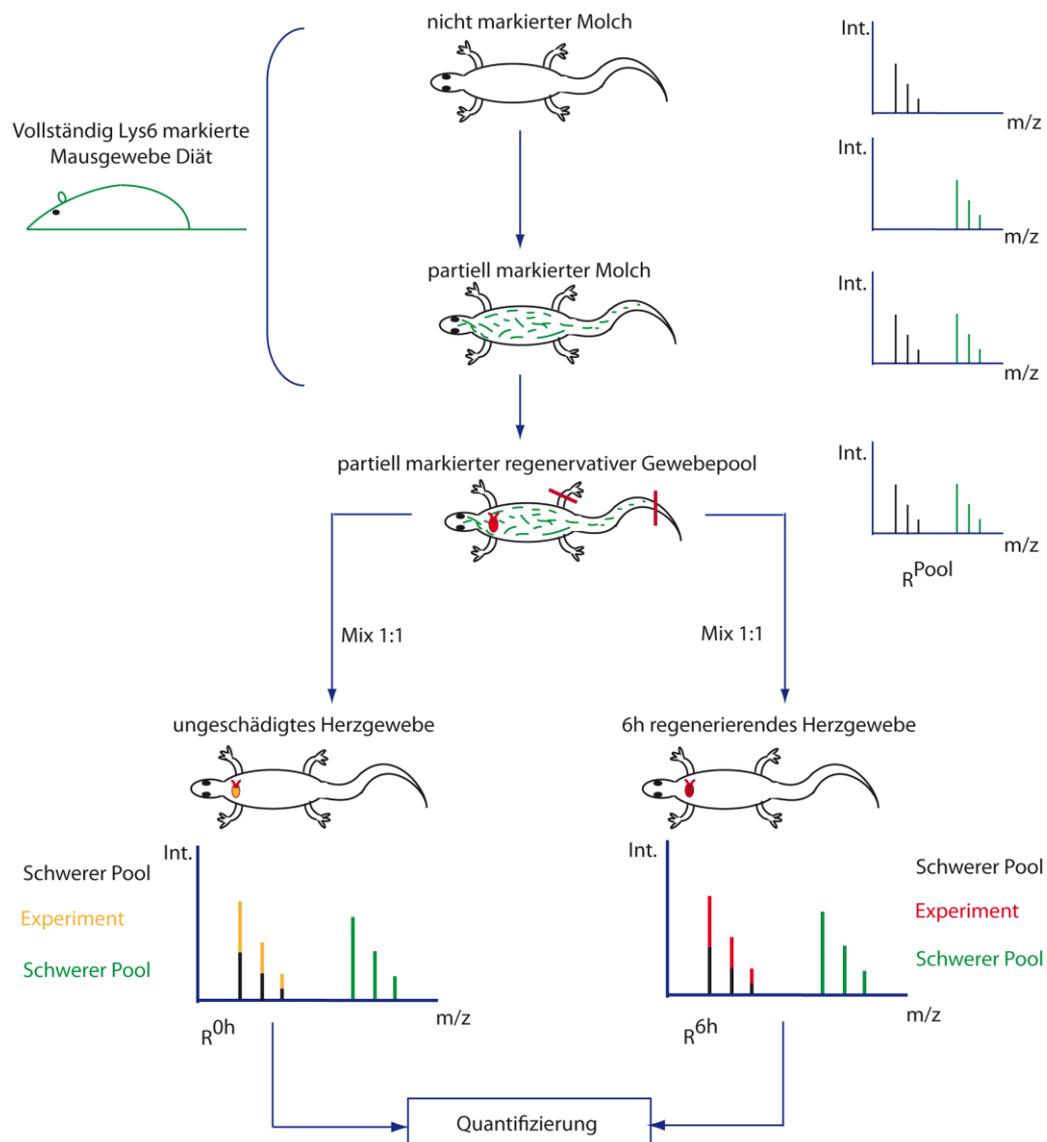


Abbildung 32: Versuchsschema Quantifizierung des Proteoms des regenerierenden Molchherzen

Dargestellt ist das Versuchsschema zur Einmischung von Lys6 markiertem schweren Standardgewebe in unmarkiertes, regenerierendes Herzgewebe. Molche werden dazu zunächst definiert mit schwerem Mausgewebe gefüttert. Aus den daraus erzeugten partiell markierten Molchen wird regenerierendes Gewebe des Herzens, des Schwanzes und des Beines gesammelt und zum schweren Proteinstandard Pool gemischt. Dieser Pool wird jeweils im Verhältnis 1:1 mit nicht regenerierendem Herzgewebe und sechs Stunden regenerierendem Herzgewebe gemischt. Die entstandenen Gemische werden jeweils massenspektrometrisch gemessen und die H/L Ratios erstellt. Über den in beide Proben einheitlich eingebrachten schweren Proteinstandard kann zwischen den beiden Zuständen des Herzen quantifiziert werden.

Ergebnisse

1. Identifizierbares Proteom während der frühen Herzregeneration

Zunächst wurde nach Messung der gemischten Pools „schwerer Standard/Herz normal“ und „schwerer Standard/Herz regenerierend sechs Stunden“ eine Suche auf der normalisierten cDNA Bank des Molches durchgeführt. Insgesamt wurden 3039 unterschiedliche Proteingruppen detektiert. Davon zeigten 2265 Proteingruppen mindestens zwei Peptide. Weitere 358 Proteine basieren auf einem einzelnen Peptid, welches mit der *Higher Energy Collision Dissociation* (HCD) Methode gemessen wurden [110], die eine zuverlässige Proteinidentifikationen anhand eines einzelnen Unique Peptides zulässt (hohe Genauigkeit von 0.5 ppm). Damit konnten insgesamt 2623 unterschiedliche, hochwertig detektierte Proteingruppen im Herzgewebe gezeigt werden. Die detektierten Proteine der drei Messungen teilen sich wie in Abbildung 33 A dargestellt in die Gruppen „exklusiv im schweren Pool“, „exklusiv im ungeschädigten Gemisch“, „exklusiv im 6 Stunden regenerierenden Gemisch“ und die jeweiligen Schnittgruppen auf. Der gemeinsame Überschneidung der drei Messungen liegt nahe 90%. Nach entsprechenden Qualitätsparametern zeigen 1707 Proteine eine H/L Ratio in allen drei gemessenen Proteingemischen. Nach weiterer Einschränkung der zugelassenen maximalen Abweichung der H/L Ratios der einzelnen Peptide in einer Proteingruppe auf 50% blieben 1353 unterschiedliche Proteingruppen.

Führt man die Proteinidentifizierung auf der IPI Mensch Datenbank aus (Version 3.54), finden sich 1011 unterschiedliche Proteingruppen mit mindestens zwei Peptiden. Davon zeigen 788 Proteingruppen eine H/L Ratio. Die Verteilung der Proteinidentifikationen ist in Abbildung 33 B abgebildet. Eine entsprechende Suche auf der IPI Maus Datenbank (Version 3.54) erbringt 1067 unterschiedliche Proteingruppen mit mindestens zwei identifizierten Peptiden (vergleiche in Abbildung 33 C). Eine Suche auf der IPI Zebrafisch Datenbank (Version 3.54) erbringt 757 unterschiedliche Proteingruppen. Die Verteilung der Proteinidentifikationen ist in Abbildung 33 D dargestellt.

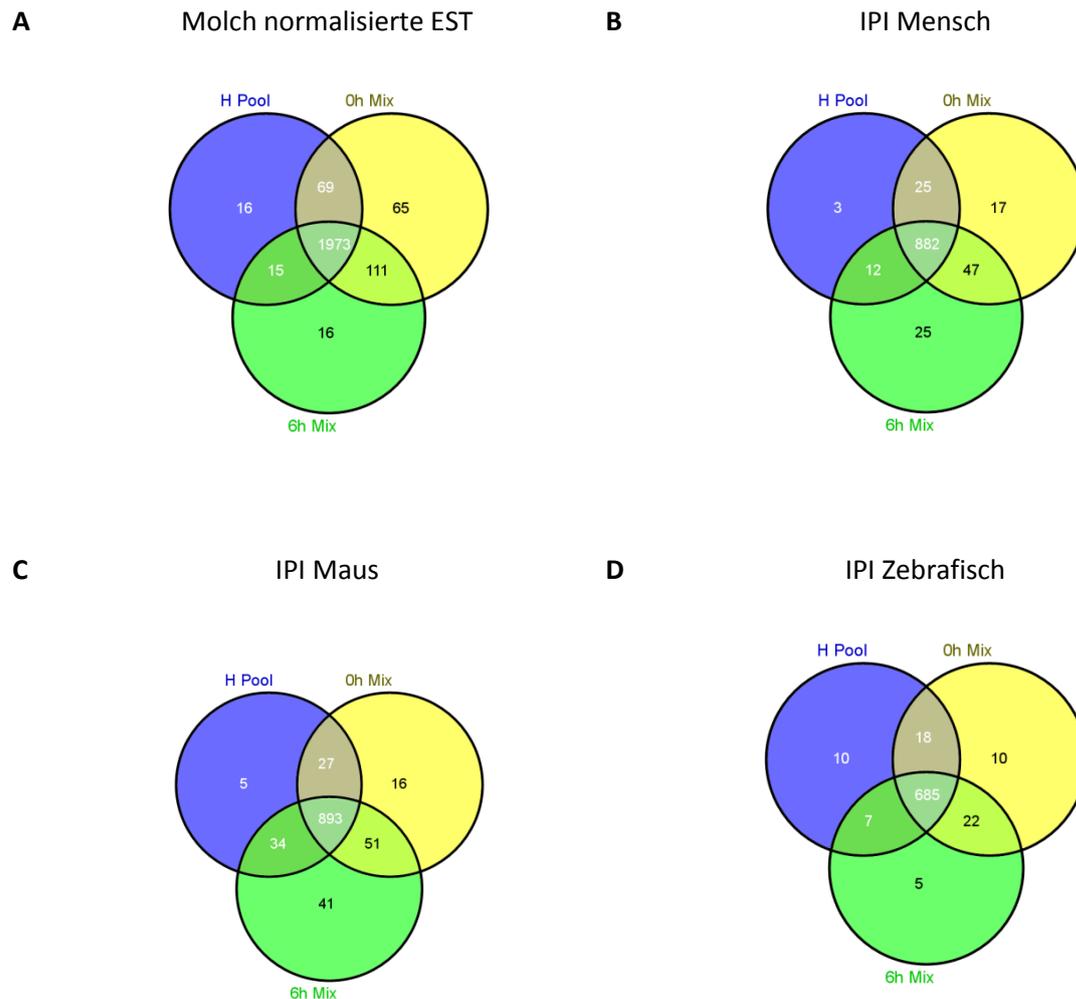


Abbildung 33: Detektierte Proteine im regenerierenden Herzgewebe, dargestellt als Venn-Diagramm

A Dargestellt sind die gemessenen Gruppen komplexer, schwerer Aminosäurepool (H Pool), ungeschädigtes Herzgewebe in 1:1 Mischung mit schwerem Pool(Oh Mix) und 6 Stunden regenerierendes Herzgewebe in 1:1 Mischung mit schwerem Pool (6h Mix). Die Peptididentifizierung wurde auf der normalisierten EST Molch Datenbank durchgeführt. Der Überschneidung der Messungen liegt bei ~87%. **B** Peptididentifizierung auf der IPI Mensch Datenbank. Die Überschneidung der Messungen liegt ebenso bei ~87%. **C** Peptididentifizierung auf der IPI Maus Datenbank. Die Überschneidung der Messungen liegt bei ~83% **D** Peptididentifizierung in der IPI Zebrafisch Datenbank. Die Überschneidung der Messungen liegt bei ~90%

Abschließend wurde eine mehrere Organismen übergreifende Datenbankanalyse durchgeführt, ähnlich der Identifikation von Proteinen im Schwanzgewebe. Abermals wurden dazu die identifizierten Peptide aus den einzelnen Banken IPI Mensch, IPI Maus, IPI Zebrafisch, NCBI *Xenopus*, NCBI *Ambystoma*, NCBI *N. viridescens*, nicht normalisierte, hausinterne Molch

Ergebnisse

EST Bank und normalisierte, hausinterne Molch EST Bank herangezogen. Nach Überlappung der entsprechenden Sequenzen konnten insgesamt 4230 Proteingruppen mit mindestens zwei Peptiden oder einem Peptid (HCD Methode, jedoch jeweils mindestens ein unique Peptid) identifiziert werden.

In den folgenden Kapiteln dieser Arbeit werden die identifizierten Proteine, die als spezifisch für den Molch und die Regeneration eingeordnet werden, näher untersucht. Um auch einen Gesamtüberblick über alle identifizierten Proteine zu geben, wurde an dieser Stelle zu den identifizierten Proteinen eine Überexpressionsanalyse mittels „Gene Ontology“ durchgeführt. Für die Ontologien „Molecular Function“ und „Biological Process“ wurde ein Abgleich gegen die menschliche „Gene Ontology“ Annotation durchgeführt. Entsprechende Graphen sind in der Abbildung 39 A und B (im Anhang) dargestellt. Aufgrund der großen Menge an Proteinen wurden nur die Terme als überexprimiert angesehen, die einen nach Benjamini und Hochberg korrigierten p-Wert kleiner als 0.0001 aufzeigten. In der Menge der identifizierten Proteine (Molch Est Datenbank) ließen sich bezüglich der molekularen Funktion verstärkt Proteine zu sechs Termen identifizieren. Diese Proteine wurden zunächst der Funktion der NADH-Dehydrogenase Aktivität (GO:0008137, 7.6255E-9, 15/42, NADH dehydrogenase (ubiquinone) activity| GO:0003954, 7.6255E-9, 15/42, NADH dehydrogenase activity| GO:00016651, 2.5703E-7, 17/70, oxidoreductase activity, acting on NADH or NADPH) zugeordnet. Außerdem waren die Terme Proteinbindung (GO:0005515, 2.2178E-7, 370/7033, protein binding) RNA Bindung (GO:0003723, 7.6255E-9, 68/692, RNA binding) und Translationsregulatorische Aktivität (GO:00045182, 1.3004E-6, 20/108, translation regulator activity) sowie Isomerase Aktivität (GO:0016853, 1.3004E-6, 24/152, isomerase activity) und Cofaktorbindung (GO:0048037, 2.3723E-6, 29/220, cofactor binding) stark überexprimiert.

Hinsichtlich der Ontologie Biologische Prozesse wurden Proteine in metabolischen Prozessen (katabolische Prozesse), die insbesondere zur Glykolyse gehören, sehr stark überexprimiert (GO:0008152, 3.0218E-14, 414/7116, metabolic process| GO:0009056, 3.5972E-12, 83/764, catabolic process| GO:0006096, 2.4340E-8, 15/44, glycolysis). Außerdem sind die Terme Translation (GO:0006416, 9.9281E-12, 60/464, translation) und Translation Elongation (GO:0006455, 1.7871E-14, 30/102, translational elongation) sehr stark vertreten. Weiterhin fallen die Terme Proteinfaltung (GO:0006457, 1.7807E-6, 27/182, protein folding), Elektronentransportkette am Mitochondrium (GO:0006120, 6.1840E-8, 14/40, mitochondrial electron transport, NADH to ubiquinone) und Gestreifte Muskel Kontraktion (GO:0006941, 7.7627E-6, 12/42, striated muscle contraction) auf.

Da diese GO Annotationen der identifizierten Proteine über die menschliche GO Annotation erfolgte, darf davon ausgegangen werden, dass die dargestellten Proteine, Prozesse und Funktionen besonders stark zwischen Mensch und Molch konserviert sind, da aufgrund der gewählten GO Methode vor der Zuweisung eines GO Termes zu einem Molchprotein zunächst die Zuordnung eines menschlichen Proteins zu einer Molchsequenz erfolgen musste. Da es sich verstärkt um generelle metabolische (katabolische) Prozesse, Prozesse der Energieübertragung am Mitochondrium oder Funktionen der Bindung (Proteine, RNA, Cofaktoren) handelte, ist eine starke Konservierung der Proteine zu erwarten. Dies ist der Fall, da es sich dabei um grundlegende Mechanismen in der Zelle handelt, die alle Organismen gemein haben.

2. Identifikation von herzspezifischen, regenerationspezifischen und bislang unbekanntem Proteinen

Nach der Identifizierung von Proteingruppen in der normalisierten EST Bank des Molches, teilt sich die Gesamtgruppe der detektierten und identifizierten Proteine in drei Hauptgruppen. Diese Gruppen sowie der Ablauf der Suche nach bislang unbekanntem kodierenden Sequenzen sind in Abbildung 34 graphisch hinterlegt. Zunächst lässt sich die Gruppe benennen, die auf Basis von Homologiesuchen, wie in Kapitel VI.A.6 beschrieben, eine klare Zuordnung zu einer bereits bekannten Sequenz erlaubt. In dieser Gruppe sind für die Suche auf den NR Nukleotid und NR Proteindatenbanken zusammengefasst 12936 annotierte, theoretisch unterschiedliche Contigs enthalten.

Daneben findet sich eine Gruppe an Sequenzen, die zwar eine signifikante Homologie zu einer bereits bekannten Sequenz in einem anderen Organismus aufzeigt, diese homologe Sequenz jedoch selber noch nicht annotiert ist. Dies sind insbesondere all diejenigen Sequenzen, die lediglich einen Treffer in den NCBI EST Banken haben. Hierbei fällt besonders die verhältnismäßig große Gruppe ins Auge, die ausschließlich eine EST Homologie aufweist, die sich auf die nahen Verwandten des Molches (*Xenopus*, *Ambystoma*, *Cynops* etc.) bezieht. Insgesamt zeigen 2149 Contigs eine Ähnlichkeit nur in den EST Banken oder der HTGS Bank. Von diesen, rein auf EST Basis annotierten Contigs weisen 45 Contigs eine exklusive Homologie zu fernen Verwandten wie Zebrafisch, Huhn und Zebrafink auf. 77 weitere Contigs zeigen Homologien zu benannten fernen Verwandten und nahen Verwandten. Der Großteil mit 2027 Contigs aber zeigt ausschließlich Homologien zu den nahen Verwandten. In dieser Liste wurden Treffer zu *Xenopus* Unterarten, *Ambystoma* Unterarten, *Cynops* Unterarten und dem Molch selbst gezählt. Gleich

Ergebnisse

man diese Contigs mit den detektierten Proteingruppen ab, so finden sich im Herzgewebe zwölf Sequenzen, die für elf Proteingruppen kodieren. Bei Hinzunahme des zuvor gemessenen Schwanzgewebes kommen weitere drei kodierende Sequenzen hinzu. Von den elf herzspezifischen Proteingruppen sind beim Vergleich mit der Gruppe aller auf Transkriptebene (Microarray Analyse) zu irgendeinem Zeitpunkt deregulierten Spots zwei Spots dereguliert.

Die dritte Gruppe letztendlich indiziert kodierende Transkripte, für die keine signifikante Sequenzhomologie gezeigt werden konnte. Für das beschriebene Herzexperiment finden sich hier 75 unterschiedliche Proteingruppen zu 140 Einzelsequenzen. Bei Erweiterung um das anfänglich beschriebene Schwanzexperiment erweitert sich diese Gruppe um 20 Proteingruppen (37 Einzelsequenzen) auf insgesamt 95 Proteingruppen. Gleicht man die im Herzen vorkommenden unbekannt 75 Proteingruppen mit der Gruppe aller auf Transkriptebene (Microarray Analyse) zu irgendeinem Zeitpunkt deregulierten Spots ab, so finden sich 47 deregulierte Spots ohne Homologie, die für zwölf unterschiedliche Proteingruppen kodieren.

Fasst man diese letzten beiden beschriebenen Gruppen zusammen, deuten sie auf eine ganze Gruppe an Proteinen hin, die bislang unbekannt sind und spezifisch für Amphibien zu sein scheinen. Diese Annahme begründet sich darin, dass diese ESTs zunächst bei einem selektiv auf PolyA endende mRNA Sequenzen hin amplifiziert und kloniert wurden. Außerdem wurden diese ESTs nach der Sequenzierung über eine Qualitätsanalyse als hochwertige Sequenzen angenommen (Kriterien waren Länge, keine Kontaminationen, keine rRNA). Auf dem erstellten Microarray zeigten diese ESTs jeweils ein Signal, einige sind sogar innerhalb des Experimentes zur Herzregeneration zu einem beobachteten Zeitpunkt dereguliert. Dass es sich tatsächlich um kodierende Transkripte handelt, ist abschließend mit der Detektion von zugehörigen Peptiden belegt. Diese zusammengefasste Gruppe umfasst für das Herzexperiment 86 Proteingruppen (75 ohne Annotation und 11 nur mit EST Annotation), durch das Schwanzexperiment kommen weitere 23 (20 ohne Annotation und drei nur mit EST Annotation) Proteingruppen hinzu. Nimmt man die Liste der deregulierten Spots aus dem Microarray Experiment hinzu, so sind 14 detektierte Proteingruppen auf Transkriptebene während des Experimentes zur Herzregeneration in einem Zeitpunkt dereguliert.

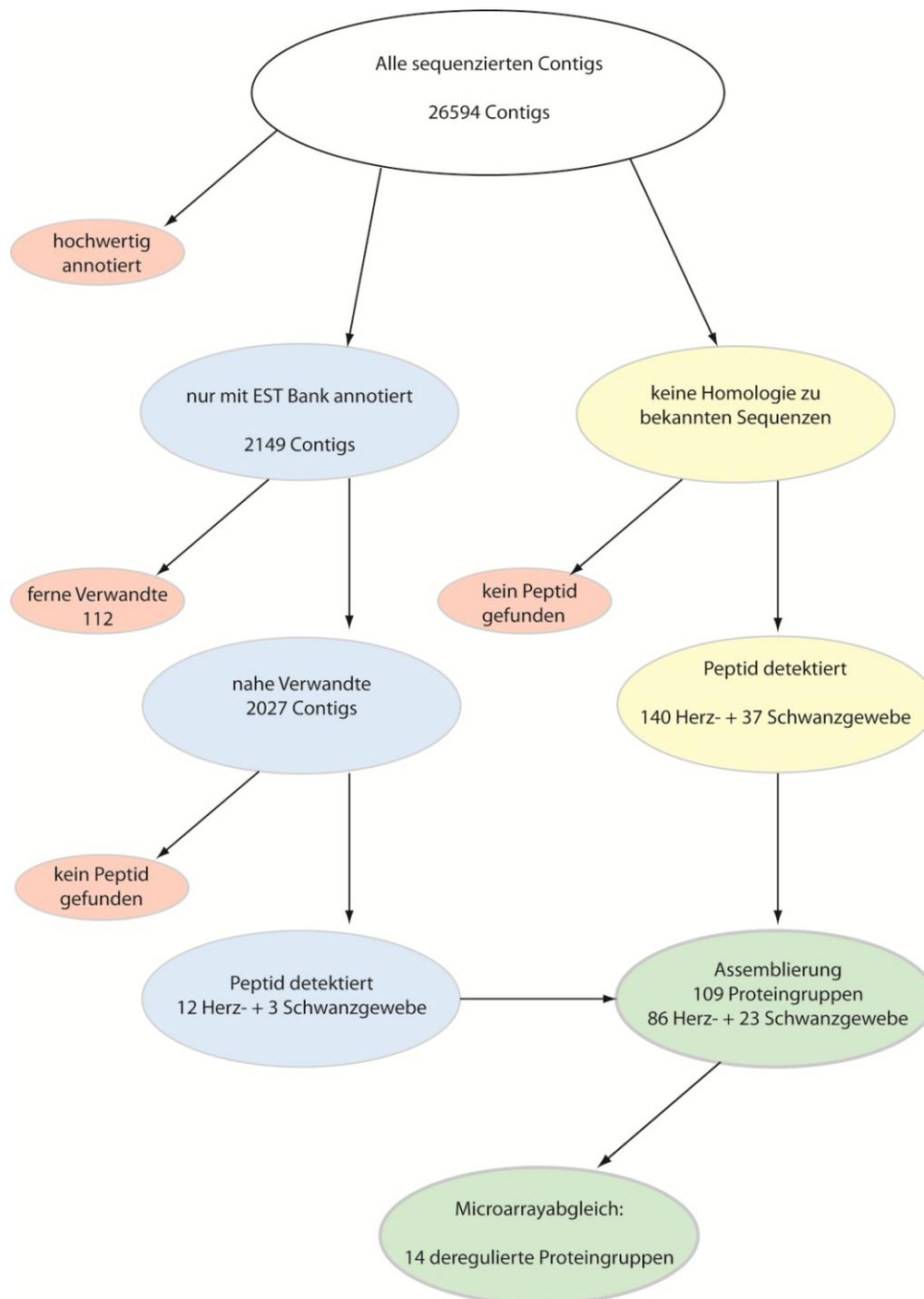


Abbildung 34: Ablaufdiagramm zur Suche von unbekanntem, kodierenden ESTs

Dargestellt ist die Schrittabfolge ausgehend vom Gesamtpool aller erstellten Contigs (Nukleotidebene) zur Identifizierung von unbekanntem und in der Regeneration dereguliertem unbekanntem Proteinen. Die Gesamtgruppe aller assemblierten Contigs teilt sich in die drei Gruppen ANNOTIERT (links), EST ANNOTIERT (Mitte) und NICHT ANNOTIERT (rechts). Der weitere EST ANNOTIERT Strang (blau) zeigt die Abspaltung der Contigs, die nur eine Homologie in fernen Verwandten wie Vögeln und Fischen zeigen. Die übrigen Contigs werden auf das Vorkommen von detektierten Peptiden hin untersucht. Der gelbe Strang NICHT ANNOTIERT rechts zeigt die direkte Suche nach Peptiden in der Gruppe der Contigs ohne Annotation. Schließlich werden die beiden Stränge auf Proteinebene zu Gruppen

Ergebnisse

zusammengefasst. Ebenfalls vermerkt ist die Anzahl der Proteine, die im Herzgewebe und zusätzlich im Schwanzgewebe detektiert wurden. Abschließend wurde wie unten dargestellt durch einen Abgleich mit den Ergebnissen der Arrayanalyse nach denjenigen Proteingruppen gesucht, die auf Transkriptebene dereguliert sind.

Die final detektierten 109, potentiell amphibien-spezifischen Proteingruppen, basieren auf einer Assemblierung auf Peptidebene. Diese Assemblierung zeigt ein hohes Maß an Ähnlichkeit zu einer Assemblierung auf Nukleotidebene. Eine Assemblierung der zu Grunde liegenden einzelnen EST Sequenzen erzeugt ein Nukleotid *Assembly* mit 110 Contigs.

Um die detektierten Nukleotidsequenzen etwas genauer charakterisieren zu können, wurden zu den jeweiligen EST Sequenzen mittels Leseraster und laufender ORF Nummer die entsprechenden detektierten Peptidsequenzen herausgesucht. Zu diesen Peptid-Sequenzen wurde auf Basis der *Pfam* Datenbank [111] eine Suche nach konservierten Proteindomänen bzw. Proteinfamilien durchgeführt. Bei dieser Suche konnten insgesamt 17 signifikant konservierte Strukturen aufgezeigt werden. Dies sind fünf unterschiedliche Pfam B Domänen, drei Proteinfamilien und neun Proteindomänen. Die einzelnen konservierten Bereiche sind in Tabelle 12 dargestellt.

Tabelle 12: Konservierte Proteinstrukturen

Zu 109 Sequenzen ohne signifikante Homologie konnten 17 unterschiedliche konservierte Strukturen in der Pfam Datenbank [111] detektiert werden. Vermerkt ist jeweils die Sequenzkoordinate, der Name der konservierten Struktur, der Typ der Struktur und der Irrtums-Wahrscheinlichkeitswert.

Sequenz ID	Name	Typ	E-value
Plate117_K6_2#SeqID-Frame#ORF-4	Apo-CII	Family	1.50E-07
Plate186_F7_1#SeqID-Frame#ORF-0	C1-set	Domain	3.60E-10
Plate165_E1_1#SeqID-Frame#ORF-0	Cathelicidins	Family	9.30E-07
Plate113_M19_3#SeqID-Frame#ORF-0	DAN	Domain	1.40E-05
Plate050_J14_2#SeqID-Frame#ORF-0	DAN	Domain	1.80E-05
Plate167_G6_3#SeqID-Frame#ORF-0	efhand	Domain	3.60E-05
Plate182_A15_3#SeqID-Frame#ORF-0	Gelsolin	Domain	3.30E-09
Plate020_N12_2#SeqID-Frame#ORF-0	LEM	Domain	5.60E-14
Plate096_J24_3#SeqID-Frame#ORF-0	Pfam-B_1037	Pfam-B	2.90E-07
Plate030_O1_1#SeqID-Frame#ORF-0	Pfam-B_2534	Pfam-B	1.10E-14
Plate030_O1_1#SeqID-Frame#ORF-0	Pfam-B_3408	Pfam-B	1.80E-07
Plate096_J24_3#SeqID-Frame#ORF-0	Pfam-B_6525	Pfam-B	7.60E-05
Plate096_J24_3#SeqID-Frame#ORF-0	Pfam-B_843	Pfam-B	8.70E-09
Plate158_O20_2#SeqID-Frame#ORF-0	Pkinase	Domain	2.10E-07
Plate117_M24_1#SeqID-Frame#ORF-0	RIG-I_C-RD	Family	5.90E-06
Plate007_H10_2#SeqID-Frame#ORF-0	TIL	Domain	8.50E-13
Plate028_P23_2#SeqID-Frame#ORF-1	ubiquitin	Domain	7.80E-07
Plate158_M18_3#SeqID-Frame#ORF-4	ubiquitin	Domain	7.80E-07
Plate183_C18_1#SeqID-Frame#ORF-1	ubiquitin	Domain	7.80E-07
Plate187_K14_1#SeqID-Frame#ORF-2	ubiquitin	Domain	1.00E-06
Plate197_K15_3#SeqID-Frame#ORF-2	ubiquitin	Domain	1.00E-06
Plate015_L11_1#SeqID-Frame#ORF-1	ubiquitin	Domain	1.90E-06
Plate133_H14_1#SeqID-Frame#ORF-2	ubiquitin	Domain	2.10E-06
Plate103_H24_1#SeqID-Frame#ORF-0	ubiquitin	Domain	2.80E-06
Plate118_I6_3#SeqID-Frame#ORF-0	V-set	Domain	3.10E-05
Plate120_O5_1#SeqID-Frame#ORF-0	V-set	Domain	4.70E-05

Aus dieser hier dargestellten Gruppe an bislang nicht beschriebenen Proteinen mit einer konservierten Proteindomäne soll beispielhaft die Sequenz mit der DAN Domäne näher erläutert werden. Diese ist aufgrund ihrer potentiellen Funktion als Regulator während der Morphogenese von besonderem Interesse für die Regeneration. Für die beiden Sequenzen mit der DAN Domäne wird zunächst das identisch gefundene Alignment für die Sequenzen 113_M19 und 50_J14 gezeigt (Abbildung 35).

Ergebnisse

```
#HMM      ssCaPkkfelvkvTLncpeekkkrvkkvelveeckCesvse
#MATCH    +C+P+k+++v ++ +c++ e +kv+ v++c C+++s+
#PP       58*****999..89*****9875
#SEQ      RTCMPskTHTVPLKfQCEGGVEI-SRKVMWVQNCLCKETS I
```

Abbildung 35: Alignment der DAN Domäne

HMM markiert die hinterlegte Sequenz der Domäne in der Pfam Datenbank [111] und MATCH zeigt die real überdeckende Sequenz an. PP gibt die Wahrscheinlichkeit für jede einzelne AS der EST Treffersequenz an, dass sie zur Domäne gehört. Das Sternsymbol steht für höchste Wahrscheinlichkeit. SEQ markiert die treffende Sequenz des ESTs, dargestellt im Farbencode basierend auf dem Wahrscheinlichkeitswert PP. Grün stellt dabei einen sehr hohen Wahrscheinlichkeitswert dar.

Die DAN Domäne besteht aus neun konservierten Cysteinen und liegt extrazellulär vor. Die Cysteine bilden häufig Schwefelbrücken aus. Die DAN Familie ist nach dem ersten Protein benannt, das die Familie begründet hat. Neben dem Protein DAN gehören noch die beiden Proteine *Cerberus* und *Gremlin* zur Familie. Das Gremlin Protein ist ein Antagonist während der Knochenmorphogenese. Es wurde vermutet, dass alle Mitglieder der DAN Familie Gegenspieler zu *TGF beta* Liganden sind [112]. Untersuchungen am Frosch haben jedoch auch gezeigt, dass das DAN Protein mit *Growth/differentiation factor 5* (GDF-5) interagiert und auf diesem Wege über *GDF-5/6/7* die *Bone Morphogenic Proteins* (BMP) Signalwege reguliert [113]. Die DAN Familie selbst gehört wieder zur *Cystine-knot Cytokine* Superfamilie, die aus neun Proteinfamilien besteht. All diese Proteine haben gemeinsam, dass jeweils sechs Cysteine drei Disulphidbrücken ausbilden. Die neun Proteinfamilien sind im einzelnen *Coagulin*, *Cys knot*, *Hormone 6*, *NGF*, *Noggin*, *PDGF*, *Sclerostin*, *TGF beta* und *DAN*.

Für den Sequenzabschnitt bp 5 bis bp 139 finden sich für die vorliegende Sequenz nicht signifikante, jedoch hohe Ähnlichkeiten zum *WNT1 inducible signaling pathway protein 1* (WISP1) mehrerer Organismen (25 unterschiedliche), wie beispielhaft für drei Organismen in Abbildung 36 dargestellt. Jedoch ist dieser konservierte Sequenzbereich bei der Molchsequenz N terminal, bei den getroffenen WISP1 Proteinen durchgängig im mittleren Bereich des Proteins liegend.

>ref|XP_002189669.1| PREDICTED: WNT1 inducible signaling pathway protein 1 [Taeniopygia guttata]

Length=359

GENE ID: 100220694 LOC100220694 | WNT1 inducible signaling pathway protein 1 [Taeniopygia guttata]

Score = 60.1 bits (144), Expect = 8e-08

Identities = 21/45 (46%), Positives = 29/45 (64%), Gaps = 0/45 (0%)

Frame = +2

```

Query 5      PNYCGYCSNRRTCMPSTHTVPLKFCQEGGVEISRKVMWVQNCLC 139
          P YCG C++ R C P K+ T+ ++FQC G E S K+MW+ C C
Sbjct 290   PKYCGVCTDNRCTPYKSKTIEVRFQC PDGTEFSWKIMWINACFC 334

```

>ref|NP_001019750.1| **WNT1 inducible signaling pathway protein 1 [Gallus gallus]**

gb|AAI21159.1| WNT1 inducible signaling pathway protein 1 [Gallus gallus]

Length=359

GENE ID: 420322 WISP1 | WNT1 inducible signaling pathway protein 1

[Gallus gallus]

Score = 59.3 bits (142), Expect = 1e-07

Identities = 21/45 (46%), Positives = 29/45 (64%), Gaps = 0/45 (0%)

Frame = +2

```

Query 5      PNYCGYCSNRRTCMPSTHTVPLKFCQEGGVEISRKVMWVQNCLC 139
          P YCG C++ R C P K+ T ++F+C G EIS K+MW+ C C
Sbjct 290   PKYCGICTDNRCTPYKSKTTEVRFEC PDGTEISWKIMWINACFC 334

```

>ref|NP_001108294.1| **hypothetical protein LOC100137691 [Xenopus laevis]**

gb|AAI57731.1| LOC100137691 protein [Xenopus laevis]

Length=385

GENE ID: 100137691 LOC100137691 | hypothetical protein LOC100137691

[Xenopus laevis] (10 or fewer PubMed links)

Score = 58.9 bits (141), Expect = 2e-07

Identities = 23/45 (51%), Positives = 29/45 (64%), Gaps = 0/45 (0%)

Frame = +2

```

Query 5      PNYCGYCSNRRTCMPSTHTVPLKFCQEGGVEISRKVMWVQNCLC 139
          P YCG C++ R C PSKT T+ ++F CE G + VMWVQ C C
Sbjct 315   PLYCGSCTDGRCTPSKTRTMRVRFHCEGDGTFQKSMWVQKCKC 359

```

Abbildung 36: Ergebnisse einer BLAST Homologie Suche in der NCBI NR Datenbank

Dargestellt sind die Alignments der detektierten unbekanntenen Molchsequenzen zu den Organismen Zebrafisch, Huhn und Frosch. Der homologe Bereich ist jeweils identisch von AS Position 5 bis 139. In diesem Bereich wurden auch entsprechende Peptide identifiziert. Die drei Sequenzbereiche zeigen gleichwertige Scores.

Ergebnisse

Analysiert man die zugrundeliegenden Sequenzen des Molches genauer, so finden sich zunächst im zugehörigen Contig (Contig 3735) zu den detektierten ESTs 113_M19 und 50_J14 zwei weitere, etwas kürzere, aber identische ESTs (174 B9 und 177 M18). Diese beiden Sequenzen sind ~100 Nukleotide kürzer. Interessanterweise liegen jedoch beide Peptide, die zu diesem Contig detektiert wurden, im ersten Sequenzabschnitt, der bei den beiden weiteren Sequenzen fehlt. Die Gesamtsequenz des Contigs zeigt ohne qualitative Berücksichtigung des Chromatogramms eine Gesamtlänge von ~720 Nukleotiden. An der Stelle 716 beginnt eine Poly A Sequenz. Führt man auf Basis des Chromatogramms eine Qualitätsanalyse durch (*Phred* [88]), verkürzt sich die Gesamtsequenz auf 685 Nukleotide. Die detektierten Peptide FQCEGGVEISRK und VMWVQNCLCK finden sich jeweils im ersten ORF der beiden ESTs:

(113_M19, Leseraster 1, ORF 0)

IPNYCGYCSNRRTCMPSTHTVPLK**FQCEGGVEISRK**VMWVQNCLCKETSIIKKKPKGLQQKKRAQGASR
EYLE

(50_J14, Leseraster 2, ORF 0)

PNYCGYCSNRRTCMPSTHTVPLK**FQCEGGVEISRK**VMWVQNCLCKETSIIKKKPKGLQQKKRAQGASRE
YLEX

Die Peptide finden sich auch im ersten ORF des Contigs 3735, welches sich aus Überlagerung dieser beiden ESTs zusammensetzt. Analysiert man diesen ersten ORF der Gesamtsequenz hinsichtlich seines Nachweises auf Peptidebene, so erkennt man, dass zu diesem ORF keine weiteren Peptide über LysC Verdau mehr detektiert werden können, da durch die hohe Lysin Dichte im ORF keine weiteren Peptide ausreichender Länge mehr entstehen, die im Massenspektrometer beachtet, bzw. in der späteren Peptididentifizierung weiterverwendet werden (zu kurz, nicht K endständig). Im gleichen Leseraster findet sich im weiteren Verlauf des Contigs im direkten Anschluss ein weiterer ORF der Länge 51 AS und dann mit 114 Nukleotiden Abstand zwei weitere, direkt aufeinander folgende ORFs der Länge 28 AS und 34 AS. Diese ORFs zeichnen sich jedoch durch eine Lysin Seltenheit aus, so dass insgesamt nur ein weiteres theoretisches Peptid entstehen könnte, welches bei der Massenspektrometrischen Messung nach LysC Verdau detektiert werden könnte. Die Variabilität der H/L Ratio zwischen den Messungen der beiden Peptide ist als gering einzustufen. Sie ergeben sich für den Zeitpunkt ungeschädigtes Herzgewebe mit 24.003 Variabilität und sechs Stunden regenerierendes Herzgewebe mit 14.87 Variabilität (jeweils prozentuelle Variabilität).

Um die Contigsequenz auf Vollständigkeit am 5' Ende hin zu überprüfen, wurde eine weitere Assemblierung mit den bereits assemblierten und alignierten Sequenzen des Molches durchgeführt. Die Erkenntnis, dass eine Sequenzähnlichkeit zum WISP1 Protein nur im ersten Sequenzabschnitt vorliegt, legte einen modifizierten Assemblierungsansatz nahe, bei dem nur der Bereich für eine Assemblierung berücksichtigt wird, der die Sequenzähnlichkeit zum WISP1 Protein zeigt. Eine solche modifizierte Assemblierung ermöglicht die Auffindung von kürzeren überlappenden Bereichen, ohne die strengen Assemblierungsparameter aufzuweichen. Bei dieser Suche findet sich eine weitere Sequenz der Länge 400bp, die mit einer Überlappung von ~100 bp genau den nachgewiesenen kodierenden Bereich trifft. Daher wurde nach Kontrolle der Chromatogramme eine neue übergreifende Sequenz erstellt, die nun eine Gesamtlänge 990 bp aufzeigt. Nach Translation dieser Sequenz vergrößert sich der erste Leserahmen der ursprünglichen Sequenz von 73 AS auf 161 AS. Eine erneute Homologiesuche mit der verlängerten Sequenz erbringt gleichwertige Treffer zu den Sequenzen von *secreted cysteine-rich protein cyr61* (Score = 126 bits (316), Expect = 6e-27) und *WNT1 inducible signaling pathway protein 1* (Score = 112 bits (279), Expect = 1e-23).

Beide getroffenen Proteine sind Mitglieder der CCN Wachstumsfaktoren Familie und werden auch als CCN1 (Cyr61) und CCN4 (WISP1) bezeichnet. Eine eindeutige Zuordnung ist auf Basis dieser Homologien nicht möglich. Es scheint aber wahrscheinlich, dass es sich an dieser Stelle um einen Vertreter dieser Familie handelt. Innerhalb der CCN Familie gibt es sechs Familienmitglieder (CCN1 – CCN6, jeweils verschiedene Synonyme). Die CCN Familie trägt unter anderem die bereits oben beschriebene *Cystine knot like* Domäne C-terminal. Zusätzlich zeigen Mitglieder der Familie eine *Thrombospondin type 1 domain* (TSP 1), die in der verlängerten Sequenz n terminal mit einem E-value von 3.1e-06 detektiert werden kann (Position 4-49, vergleiche Abbildung 37). Außerdem findet sich bei den Mitgliedern der CCN Familie n terminal eine *insulin-like growth factor binding protein (IGFBP)* Domäne. Diese kann jedoch in der vorliegenden Sequenz nicht mehr nachgewiesen werden, da das N-terminale Ende zu dieser Sequenz entweder anders aufgebaut, oder aber nicht vollständig ist.

Ergebnisse

```
#HMM      seWspCSvTCgkGirvRqrtcksnqkggepCtgdvqeteaCkmqkC
#MATCH    +eW+ CS+TCg Gi++R +  ++          +Ct  +q  ++C+ ++C
#PP       8*****99995.....6****98.7*****
#SEQ      TEWTHCSKTCGTGISTRILTWSNA-----SCTPRAQ-RRICMIRPC
```

Abbildung 37: Alignment der TSP 1 Domäne

HMM markiert die hinterlegte Sequenz der Domäne und MATCH zeigt die real überdeckende Sequenz an. PP gibt die Wahrscheinlichkeit für jede einzelne AS an, SEQ markiert die Suchsequenz, dargestellt im Farbencode basierend auf dem Wahrscheinlichkeitswert PP. Grün stellt dabei einen sehr hohen Wahrscheinlichkeitswert dar.

Betrachtet man nun neben der Peptid- und Proteinidentifizierung zu dieser Sequenz noch einmal die zugehörigen Daten aus dem Microarray Experiment, ergeben sich für dieses Contig die in Tabelle 13 dargestellten Expressionsdaten. In der Tabelle sind nur drei der vier EST Koordinaten aufgeführt, die vierte ist ein Array-Spotting Ausfall. Aus den dargestellten Daten ist zunächst zu ersehen, dass sich die drei Koordinaten über den Zeitverlauf der Regeneration homolog verhalten. Innerhalb der Zeitreihe fällt auf, dass das Transkript im scheinoperierten Gewebe nicht dereguliert ist, jedoch ab zwei Stunden Regeneration bis sechs Stunden Regeneration aufreguliert ist. 24 Stunden nach der Schädigung sinkt das Niveau der Aufregulierung leicht ab, steigt dann aber wieder bis zum Zeitpunkt vier Tage nach Schädigung stark an und erreicht das Maximum. Nach sieben Tagen ist immer noch eine Aufregulierung zu detektieren, nach 14 Tagen Regeneration ist das Transkript dann wieder unauffällig reguliert.

Um die auf Transkriptebene in der Regeneration deregulierte Sequenz auch auf Proteinebene auf Deregulation zu überprüfen, wäre eine klassische Westernblot Quantifizierung möglich. Der dazu nötige Antikörper ist aber nicht verfügbar. Ein erster Hinweis, ob an dieser Stelle nach sechs Stunden der Herzregeneration auch potentiell ein dereguliertes Protein vorliegt, kann aber rechnerisch ermittelt werden. Dazu werden die jeweiligen H/L Ratios aus den massenspektrometrischen Untersuchungen des Herzens und des schweren Standards herangezogen.

Tabelle 13: Expressionsdaten des Contigs 3735 (CCN Homolog)

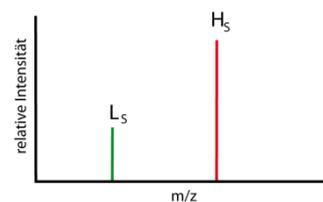
Dargestellt sind drei zugehörige EST Koordinaten des Contigs (Zeilen). Von links nach rechts sind die Beobachtungszeiträume von Scheinoperiert über zwei Stunden Regeneration bis 35 Tage Regeneration dargestellt. Die Expressionswerte sind als Logarithmus zur Basis 2 abgebildet. Werte größer 0 zeigen eine relative Aufregulation in der Schädigung, Werte kleiner 0 eine relative Abregulation in der Schädigung. Jeder Expressionswert stellt ein über den Mittelwert verrechnetes Experiment von acht Microarrays dar. Information über die Variabilität des Wertes sowie die Signifikanz bzgl. der Abweichung des Wertes vom Null Wert zeigt jeweils der rechts neben den Expressionsdaten angeführte p Wert (One Sample t-Test - Benjamini-Hochberg).

ID	sham	Sham p Wert	2 St.	2 St. p Wert	6 St.	6 St. p Wert	24 St.	24 St. p Wert	48 St.	48 St. p Wert
Plate113_M19	0.254	1	1.475	0.138203	1.367	0.00946	0.629	0.156869	1.056	0.165498
Plate050_J14	0.222	1	1.512	0.140147	1.255	0.012649	0.672	0.279882	0.944	0.196417
Plate174_B09	-0.074	1	0.825	0.579039	1.138	0.166611	0.336	1	0.659	0.615909

ID	4 Tage	4 Tage p Wert	7 Tage	7 Tage p Wert	14 Tage	14 Tage p Wert	21 Tage	21 Tage p Wert	35 Tage	35 Tage p Wert
Plate113_M19	2.127	0.000792	1.375	0.051669	-0.194	0.756791	0.219	0.441777	-0.072	1
Plate050_J14	1.866	0.007027	1.383	0.084354	-0.179	0.86344	0.171	0.541349	-0.079	1
Plate174_B09	1.748	0.027434	0.977	0.396653	0.04	1	0.117	0.58021	-0.158	1

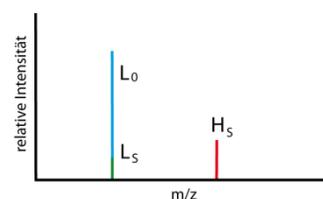
Zunächst wird die Markierungsrate der detektierten Peptide zum Contig 3735 im schweren SILAC Pool bestimmt. Diese Ratio wird als R_p bezeichnet. P steht in diesem Zusammenhang für Standard Pool. Somit ist R_S wie folgt definiert:

$$Ratio R_p = \frac{Heavy_p}{Light_p}$$



Dazu kommt die Markierungsrate aus der massenspektrometrischen Messung des 1:1 Gemisches „schwerer Pool/ungeschädigtes Herzgewebe“. Benannt wird diese H/L Ratio mit R_0 . Die 0 steht in diesem Zusammenhang für Zeitpunkt 0 Tage Regeneration, also ungeschädigt. Bei dieser H/L Ratio besteht das leichte Peptid aus der addierten Menge des im schweren Standard enthaltenen leichten Peptides und aus der ungeschädigten Herzprobe. Definiert ist R_0 als:

$$Ratio R_0 = \frac{Heavy_p}{Light_p + Light_0}$$

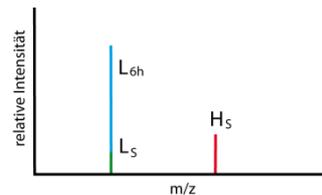


Ergebnisse

Abschließend wird die Markierungsrate aus der massenspektrometrischen Messung des 1:1 Gemisches „schwerer Pool/6 Stunden regenerierendes Herzgewebe“ bestimmt. Benannt wird diese H/L Ratio mit R_{6h} . Das 6h steht in diesem Zusammenhang für Zeitpunkt sechs Stunden Regeneration. In dieser Ratio besteht das leichte Peptid aus der addierten Menge des im Standard enthaltenen leichten Peptides und aus den Peptiden des Regenerationszeitpunktes sechs Stunden.

Definiert ist R_{6h} als:

$$\text{Ratio } R_{6h} = \frac{\text{Heavy}_P}{\text{Light}_P + \text{Light}_{6h}}$$



Aus rechnerischen Gründen werden die Ratios R_P , R_0 und R_{6h} jeweils invertiert und als R_P^* , R_0^* und R_{6h}^* bezeichnet. Über die Ratio R_P^* des schweren SILAC Pools kann nun das Protein normalisiert [114] und dann R_0^* gegen R_{6h}^* quantifiziert werden. Dies geschieht im Einzelnen wie folgt:

Für R_0^* und R_{6h}^* interessiert jeweils die auf R_P^* normalisierte H/L Ratio R_{N0}^* , beziehungsweise R_{N6h}^* . Exemplarisch soll die Rechnung für R_{N0}^* gezeigt werden. R_{N0}^* wird definiert als Signalstärke der Peptide aus dem ungeschädigten Gewebe, geteilt durch die Summe der Signale der Peptide aus dem schweren Pool:

$$\text{Ratio } R_{N0}^* = \frac{\text{light}_0}{\text{light}_P + \text{heavy}_P}$$

Dies führt mit:

$$R_0^* - R_P^* = \frac{\text{light}_0 + \text{light}_P}{\text{heavy}_P} - \frac{\text{light}_P}{\text{heavy}_P} = \frac{\text{light}_0}{\text{heavy}_P}$$

Aufgelöst nach:

$$\text{light}_0 = (R_0^* - R_P^*) * \text{heavy}_P$$

Zur Ratio:

$$R_{N0}^* = \frac{\text{light}_0}{\text{light}_P + \text{heavy}_P} = \frac{(R_0^* - R_P^*) * \text{heavy}_P}{\text{light}_P + \text{heavy}_P} = \frac{(R_0^* - R_P^*)}{\frac{\text{light}_P}{\text{heavy}_P} + 1} =$$

$$\frac{(R_0^* - R_P^*)}{R_P^* + 1}$$

Die beiden so errechneten Ratios R_{N0}^* und R_{N6h}^* können relativ zueinander ins Verhältnis gesetzt werden, was einen Rückschluss auf die relative Menge des Proteins zulässt.

$$R_{EXP}^* = \log_2 \left(\frac{\frac{(R_{6h}^* - R_P^*)}{R_P^* + 1}}{\frac{(R_0^* - R_P^*)}{R_P^* + 1}} \right) = \log_2 \left(\frac{R_{6h}^* - R_P^*}{R_0^* - R_P^*} \right)$$

Mit:

$$f(R_{EXP}^*) = \begin{cases} \text{aufreguliert, } R_{EXP}^* > 0 \\ \text{abreguliert, } R_{EXP}^* < 0 \end{cases}$$

Setzt man die entsprechenden Werte für die zum Contig 3735 zugehörigen H/L Ratios in diese Formel ein, so erhält man die in Tabelle 14 aufgeführten Werte.

Tabelle 14: Berechnungsschritte zur Quantifizierung des Contigs 3735

Quantifizierung auf Proteinebene auf Basis einer Einmischung eines SILAC markierten schweren Pools im Verhältnis 1:1. 1/R Pool definiert die invertierte H/L Ratio im schweren Pool, der den Messstandard für das Contig 3735 angibt. 1/R Mix0 Stunden gibt die invertierte H/L Ratio im gemischten Pool mit ungeschädigtem Herzgewebe an. 1/R Mix6 Stunden gibt die invertierte H/L Ratio im gemischten Pool mit ungeschädigtem Herzgewebe an. R Norm Mix0 Stunden gibt den normierten Wert für die Mischung mit ungeschädigten Gewebe an, R Norm Mix6 Stunden schließlich gibt den normierten Wert für die Mischung mit ungeschädigten Gewebe an. Die letzte Spalte Rnorm Mix6/R Norm Mix0 gibt den zur Basis 2 logarithmierten Quotienten der beiden normierten Ratios zu 6 Stunden regenerierendem Gewebe und ungeschädigtem Gewebe an.

1/R Pool	1/R Mix0 Stunden	1/R Mix6 Stunden	R Norm Mix0 Stunden	R Norm Mix 6 Stunden	Rnorm Mix6/R Norm Mix0
0.874891	0.979144	1.123659	0.055605	0.132684	1.254704

Der mit „Rnorm Mix6/R Norm Mix0“ bezeichnete Wert gibt also rechnerisch eine relative Aufregulierung von ~ 1.25 an. Dieser Wert korreliert sehr gut mit den auf Transkriptebene gefundenen Werten. Eine manuelle Analyse der Spektren zeigt ebenfalls eine starke Aufregulierung im regenerativen Molchherzen. Die beiden Spektren (Full MS, MS², Peptid mit detektierten Ionen) für das Peptid VMWVQNCLCK sind in Abbildung 38 (siehe X.B) dargestellt.

Um diese Sequenz nun genauer zu analysieren sind weitere Untersuchungen nötig. Zunächst wären weitere massenspektrometrische Messungen mit anderen Proteasen sinnvoll, um ein verändertes Schnittmuster zu erzeugen, damit weitere ORFs des Contigs 3735 gezeigt werden könnten. Insbesondere könnte auf diesem Wege untersucht werden, ob die Sequenz über mehrere kodierende Leseraster oder ORFs verfügt. Daneben sollte die detektierte Deregulation mit RT-PCRs kontrolliert werden. Auf diesem Wege könnten auch nähere Informationen bezüglich des Vorkommens der zwei Transkriptvarianten, die sich durch ~ 100 bp unterscheiden, gewonnen werden (Einsatz verschiedener 5' Primer). Auch die Lokalisation des Transkriptes und zugehörigen Proteins spielt wie bereits bei anderen bekannten Hauptakteuren der Regeneration [28] eine entscheidende Rolle. Schließlich sollte die Deregulation auf Proteinebene durch einen klassischen Westernblot verifiziert werden. Zum Zeitpunkt dieser Ausarbeitung ist jedoch noch kein entsprechender Antikörper verfügbar.

VII. Diskussion

Die außergewöhnlichen Fähigkeiten des Molches, zahlreiche Gewebearten schnell und vollständig zu regenerieren sind schon seit Jahrzehnten bekannt. Deshalb verwundert es nicht, dass der Molch als Modellorganismus für die Molekularbiologie der Regeneration ausgewählt wurde. Das Fehlen etablierter Methoden erschwert allerdings die Arbeit mit diesem Organismus erheblich. Das größte Hindernis ist dabei die enorme Größe des Molchgenoms. Außerdem ist die Gattung Molch mit ihren regenerativen Fähigkeiten offenbar taxonomisch so weit von anderen Organismen entfernt, dass sich andere Gattungen nicht ohne weiteres zu Vergleichen heranziehen lassen. Deshalb war es notwendig in der vorliegenden Arbeit zahlreiche Methoden abzuwandeln bzw. neu zu entwickeln. Die Vor- und Nachteile dieser Methoden, die Möglichkeiten der Anwendung auf anderen Organismen und die Grenzen der Aussagekraft sollen an dieser Stelle diskutiert werden.

A. Der *Notophthalmus viridescens* als Modellorganismus

Mit der Wahl des Molches *Notophthalmus viridescens* als Modellorganismus für regenerative Fragestellungen wurde bewusst ein wenig erforschter Organismus in den Mittelpunkt dieser Arbeit gestellt. Daher geht mit dieser Wahl gleichzeitig die erschwerte molekularbiologische Handhabung einher. Dem gegenüber stehen aber die außerordentlichen regenerativen Fähigkeiten dieses Molches, welche in diesem hohen Grad der Ausprägung bei anderen Wirbeltieren bislang nicht bekannt sind (vergleiche IV.C).

In diesem Gegenpiel zwischen Handhabbarkeit und exklusiver Fähigkeiten gilt es genau abzuwägen, für welche Fragestellungen die Arbeit mit dem Molch sinnvoll ist, und welche Fragestellungen eher mit anderen, besser erforschten Organismen durchzuführen sind. Um diese Abwägung durchführen zu können, ist zunächst ein Grundstock an Informationen und Techniken zum Modellorganismus Molch zu erarbeiten, der die Abschätzung von Aufwand und Durchführbarkeit von Fragestellungen erlaubt. Diese Arbeit befasst sich in einem Schwerpunkt mit der Sammlung, Organisation und Entwicklung dieser grundlegenden Informationen.

Heutzutage stehen Methoden zur Verfügung, große Datenmengen durch massive parallele DNA-Sequenzierung relativ rasch zu generieren. Man kann neben genomischen Daten auch

Daten auf Ebene der Genexpression, dem sogenannten Transkriptom, oder auf Proteinebene des Proteoms erhalten. All diese Daten sind aber erst interpretierbar, wenn ausreichend Vergleichsdaten zur Verfügung stehen, um diese bereits bekannten Phänomenen zuzuordnen, oder aber als neue unbekannte Kandidaten klassifizieren zu können.

Es konnte gezeigt werden, dass mit vertretbarem Aufwand durch Anpassung gängiger Techniken ein übergreifender molekularbiologischer Überblick für diesen Modellorganismus generiert werden konnte. So bietet die Erstellung einer normalisierten cDNA Bank einen ausreichenden Grundstock an genetischer Information, um darauf basierend Hochdurchsatzmethoden zu entwickeln. Die Durchführung des Microarray-Experimentes auf dieser cDNA Basis zeigt die Möglichkeit auf, das Transkriptomanalysen, auch als zeitliche enge Beobachtungsreihen, nach einmaliger Erstellung der Microarrays leicht möglich sind und die daraus resultierenden Daten einen tiefen Einblick in Expressionsmuster gewähren. Eine solche Microarray-Zeitreihe erlaubt die Darstellung des Transkriptionsniveaus vieler einzelner Transkripte über den regenerativen Verlauf hinweg. Die Wahl zur älteren Technologie der cDNA Microarrays [115] ist an dieser Stelle mit zwei Punkten zu begründen. Zum Ersten war zum Zeitpunkt des Arraydesigns noch keine ausreichende Sequenzinformation für die Erstellung von heute modernen Oligo Arrays verfügbar, diese entstand erst nach der Auswertung der cDNA Arrays. Zum zweiten konnte durch die Kooperation mit einer in dieser Technik erfahrenen Arbeitsgruppe (AG Kreuzberger, Max-Planck-Institut für molekulare Genetik, Berlin) eine vergleichbar hohe Spotdichte von 50.000 Spots/Glasträger erreicht werden. Daneben sollten die auf den Microarray aufgebrachten langen cDNA Stränge für weitergehende Analysen bezüglich der untranslatierten und kodierenden Bereiche genutzt werden. Die erstellten Microarrays können aufgrund ihres Designs auch für die Beobachtung der Regeneration in anderen Geweben verwendet werden (Linse, Schwanz, Bein, Gelenkknorpel, Daten nicht gezeigt).

Ein vergleichbares Experiment zur Herzregeneration, beispielsweise an dem vollständig sequenzierten Modellorganismus Maus, ist schwer durchführbar. Hier fehlt zunächst die grundsätzliche Fähigkeit zur vollständigen Regeneration und im speziellen zur Herzregeneration. Und selbst, wenn man diese Fähigkeit induzieren könnte und würde, wäre auch hier die Durchführung einer Microarray Zeitreihe ähnlich aufwändig und kostenintensiv wie beim Molch. Außerdem bietet eine zugrunde liegende cDNA Bank, die explizit aus regenerierendem Herzgewebe erstellt wurde, den spezialisierten und kompletten Umfang an Transkripten, die während der Regeneration des Herzens eine Rolle spielen. Eine ähnliche Schwerpunktsetzung im Design ist bei der Verwendung kommerzieller Arrays nicht gegeben, da hier beim Design

Diskussion

grundsätzlich Wert auf eine breit gefächerte Verwendung gelegt wird. Insbesondere sind nicht bekannte, bzw. als nicht kodierende klassifizierte Elemente kaum auf einem kommerziellen Array zu finden. Um also beispielsweise auch im Modellorganismus Maus eine vergleichbar gezielte, herzspezifische Untersuchung machen zu können, müsste auch hier ein eigener angepasster Array erstellt werden. Im Gegenzug hält eine standardisierte Lösung auf Basis von kommerziellen Arrays den Vorteil für sich, bei der Auswertung und Analyse auf gut erprobte und leicht zugängliche Algorithmen und Anwendungen, aber auch eine weit ausgebildete Expertise zurückgreifen zu können. Dies ist bei cDNA Arrays nicht gegeben.

Wechselt man von der Transkriptomanalyse zur Proteomanalyse, so gelten für die Wahl des Modellorganismus Molch nahezu die gleichen Argumente wie bei der Transkriptomanalyse. Es konnte jedoch gezeigt werden, dass man unter der kombinierten Verwendung von eigenen EST Banken und organismusfremden Proteindatenbanken zur Peptididentifizierung vergleichbare Ergebnisse bezüglich der Anzahl der detektierbaren Proteingruppen für den Molch erzielen kann, wie in etablierten Versuchen im Modellorganismus Maus. Insbesondere konnte hier gezeigt werden, dass durch die Verwendung von gezielt erstellten Datenbanken (Sequenzen ohne Annotation) bislang völlig unbekannte Proteine und damit kodierende Transkripte identifiziert werden konnten, die bei der Verwendung von unspezifischen Datenbanken, die sich auf die bereits erzielten Erkenntnisse stützen, nicht möglich gewesen wären.

Letztendlich können also moderne Methoden der Transkriptomanalyse und Proteomanalyse auch für einen nahezu unbekanntem Organismus genutzt werden. Daher richtet sich die Wahl zu einem unüblichen Modellorganismen wie dem Molch letztendlich alleine daran aus, wie exklusiv die Fähigkeiten, die es zu untersuchen gilt, für den einzelnen Organismus sind.

Jedoch soll an dieser Stelle nicht die aus bioinformatischer Sicht größte Schwierigkeit und Ungenauigkeit bei der Arbeit mit einem unüblichen Modellorganismus verschwiegen werden. Die größte Schwierigkeit nach Erstellung eines Sequenzpools, unabhängig von der verwendeten Technik, stellt die immer wiederkehrende Frage der Assemblierung und damit verbundenen Annotation dar. So stellt sich die Frage, ob eine primäre redundante Sequenzbank (ESTs) zunächst assembliert und dann mittels Homologiesuche annotiert wird, oder ob nicht vielmehr eine vorherige Homologiesuche und Assemblierung auf Basis der gefundenen Annotationen sinnvoll ist. Hier hat sich ersterer Weg als der allgemein anerkannte durchgesetzt. Jedoch ist die primäre Annotation der erzeugten Transkripte (Contigs) keine 1:1 Beziehung, da es in verschiedenen Organismen und in verschiedenen Datenbanken immer mehrere, durchaus gleichwertige, homologe Sequenzen gibt. Setzt man bei hoher Sequenzähnlichkeit eines

Molchtranskriptes zu einem Transkript eines fremden annotierten Organismus gleiche Funktion voraus, so bleibt immer noch die Verwaltung von mehreren gleichwertigen Treffern in mehreren Organismen, die dann durchaus auch unterschiedlich annotiert sein können.

Dieser Problematik kann im ersten Schritt entweder durch die Reduzierung auf den statistisch „besten“ Treffer oder die Mitführung mehrerer Annotationen begegnet werden. Bei der Reduzierung auf einen Treffer verliert man jedoch wichtige Informationen, insbesondere immer dann, wenn der beste Treffer selbst eine nicht annotierte Sequenz darstellt. Bei der Mitführung mehrerer Annotationen treten die Probleme spätestens im nächsten Schritt auf, wenn es darum geht, GeneOntology Annotationen, Pathway Informationen oder generell Verwandtschaftsbeziehungen und funktionelle Annotationen herleiten zu wollen. Hier bedeuten mehrere primäre Annotationen auch immer sehr viele sekundäre funktionelle Annotationen, was letztendlich zu einer starken Aufweichung von nachfolgenden Analysen führt, da vieles mit vielem „irgendwie“ in Beziehung steht. Somit ist klar, dass bei der Arbeit mit ungewöhnlichen Modellorganismen zunächst ein erhöhter Aufwand darauf verwendet werden muss, die vorliegenden Sequenzen in eine handhabbare Form zu bringen. Diese Organisation muss vor allem möglichst genau und fehlerfrei geschehen, da durch die, im Fall des Molches, kleine potentielle Nutzergruppe Fehlannotationen und falsche Datenablagen kaum kontrolliert und verbessert werden (bzw. werden können). Bei großen, häufig genutzten Datenbankprojekten werden jedwede Zuordnungen durch große Anwenderzahlen validiert. In diesen Punkten liegt der aus bioinformatischer Sicht größte Vorteil in der Verwendung von gemeinen Modellorganismen. Hier stellen umfangreiche Datenbanken vom Genom über das Transkript bis hin zum Protein nebst regulatorischen Einheiten alles umfangreich, komplett und auch in vielen (Datei) Formaten zur Verfügung.

Aus rein biologischer Sicht leiten sich noch einige andere, zumeist handwerkliche Schwierigkeiten ab. So fehlt nahezu jedwede althergebrachte und evolutionierte Erfahrung mit dem Molch. Damit fehlen auch Protokolle, Plasmide, Konstrukte und andere Werkzeuge, um den Molch genetisch zu manipulieren. Auch können viele traditionelle Nachweisverfahren wie Western Blots und Immunohistologische Färbungen nur unter erhöhtem Aufwand durchgeführt werden, da sich alleine der Bezug bzw. die Herstellung von entsprechenden Antikörpern aufwändig ist.

Zusammenfassend ist festzuhalten, dass die Verwendung eines genetisch wenig untersuchten Organismus wie dem *N. viridescens* zwar Einschränkungen bei der Handhabbarkeit mit sich bringen, diese aber mit entsprechendem Aufwand, auch in kleineren Arbeitsgruppen,

Diskussion

überbrückt werden können. Die größte Schwierigkeit besteht in der in diesem Fall eigenverantwortlichen Generierung und Haltung von Sequenzinformationen und entsprechenden Annotationen sowie daraus abgeleiteten Daten wie beispielsweise Expressionsdaten. Dieser Aufwand ist jedoch je nach untersuchter Fragestellung gerechtfertigt, insbesondere dann, wenn es sich um auf diesen Organismus beschränkte, exklusive Fragestellungen handelt. Eine Fragestellung ist jedoch immer einzeln genau auf diese Exklusivität hin zu überprüfen.

B. Entwicklung von Hochdurchsatzmethoden für den Molch

Im Rahmen dieser Arbeit wurden für den Molch entsprechende Methoden zur Durchführung von Hochdurchsatzanalysen auf Transkriptom- und Proteomebene adaptiert. Zur Analyse des Transkriptoms kamen selbsterstellte cDNA Microarrays zum Einsatz, für die Proteomanalyse wurde eine Gel/Offgel Proteintrennung mit anschließender Flüssigchromatographie, gekoppelt mit einer Elektrosprayionisation und einem Ionenfallen/Orbitrap Massenspektrometer verwendet. Die Entwicklung der beiden Methoden wird im Folgenden diskutiert.

1. cDNA Microarrays zur Analyse des Transkriptoms

cDNA Microarrays stellen eine ausgereifte Technologie dar, um das Vorkommen von Transkripten in einer Zelle im Hochdurchsatz zu untersuchen. Das gilt auch im Vergleich zu neueren Array-Technologien, die einen höheren Durchsatz, eine höhere Genauigkeit und einen verringerten Aufwand in der Probenaufbereitung ermöglichen.

So übertreffen die neueren Microarray Technologien wie die Oligo Arrays die hier verwendete cDNA Array Technologie in der Anzahl der gleichzeitig zu untersuchenden Spots um den Faktor 60 (cDNA Array max. 50.000 Probes, Affymetrix Mouse Exon Array bis zu 1.200.000 Probes). Dies hat dazu geführt, dass cDNA Microarrays im Laufe der letzten Jahre in ihrer Verbreitung und Nutzung stark zurückgegangen sind. Dieser Effekt wurde noch dadurch verstärkt, dass sich für die Oligo Arrays eine einheitliche Auswertungssoftware etabliert hat, die die komplizierte Auswertung von Oligoarrays in einem *BlackBox* Prinzip ermöglichen. Dieser Umstand, gepaart mit der vergleichsweise aufwändigen Auswertung von cDNA Microarrays hat zu einer enormen Verbreitung und Anwendung von Oligo Microarrays geführt. Jedoch kann man

diese Oligoarrays fast ausschließlich als kommerziell vorgefertigte Produkte erwerben, sie sind also nicht frei gestaltbar. Erst seit kurzem gibt es kommerzielle Oligoarrays mit individuellem Design, welche verstärkt zur Analyse von miRNAs eingesetzt werden. Diese Punkte haben vier wesentliche Einschränkungen zur Folge, die den Einsatz von Oligoarrays nur bedingt empfehlen und den Einsatz von cDNA-Arrays trotz der Vorteile der Oligoarrays rechtfertigen:

Zunächst wäre an dieser Stelle die Annotation der einzelnen Probesets zu nennen. Diese stammen aus entsprechenden Annotationsdateien der Microarray Hersteller. Sämtliche Statistiken und Ergebnislisten sind auf Basis dieser Annotationsdateien aufgebaut. Diese Dateien werden in unregelmäßigen Abständen als *Updates* herausgebracht. Betrachtet man aber diese Dateien etwas genauer, so stellt man fest, dass sich selbst für die gut untersuchten Modellorganismen immer wieder starke Verschiebungen in den Annotationsdateien ergeben. Dies hat zur Folge, dass einmal ausgewertete Arrays später, mit einer neueren Annotationsdatei ausgewertet, andere Ergebnisse hervorbringen. Dies liegt unter anderem daran, dass zwar der physikalisch zu Grunde liegende Sequenzpool, der auf dem Array hinterlegt ist, gleich bleibt, sich jedoch die Interpretation dieser Sequenzen ständig ändert. So kommen selbst zu gut erforschten Modellorganismen in einem Annotationsupdate nicht nur Annotationen hinzu, indem beispielsweise vorhergesagte, potentielle Transkripte (z. B. RIKEN Klone) einen Namen bekommen. Ebenso lassen sich manche bereits fest annotierte Transkripte nachträglich nicht reproduzieren, da durchaus auch fest annotierte Transkript-Namen in ein *Potential-* oder *Predicted-Transcript* umgewandelt werden. Auch können Probesets komplett aus der Auswertung herausfallen, weil aufgrund der verwendeten Sequenzen keine eindeutige Zuordnung zu einem Transkript mehr möglich ist. Solche Fehler im Design treten häufig erst lange nach dem primären Arraydesign auf. Diese starken Veränderungen in den Annotationsdateien haben ihren primären Ursprung in der Technik der Arrays selbst. So führen kurze Oligonukleotid-Sequenzen von 20-25 nt zu häufigen Kreuzhybridisierungen, deren Verrechnung (innerhalb eines Probesets) und Berechnung (Intensität der Kreuzhybridisierung selbst) sich schwierig darstellt. Auch Isoformen bzw. mehrere Spleißvarianten aus einem Gen stellen für Oligoarrays eine Herausforderung dar [116]. Diese Problematik besteht für cDNA Microarrays nicht. Durch die Verwendung von langen cDNA Sequenzen (mehrere 100 bp) haben Kreuzhybridisierungen nahezu keinen nennenswerten Effekt. Und durch die Nutzung der langen Sequenzen sind auch Spleißvarianten zu einem Gen leicht auseinander zu halten. Daher muss eine Annotationsdatei für den Microarray nur einmal erstellt werden. Starke Veränderungen darin entstehen nur noch durch zusätzliche Informationen, da die Abhängigkeit der verwendeten Sequenzen untereinander minimal, bzw. die Diversität der einzelnen Sequenzen maximal ist.

Diskussion

Daneben zeigen Oligoarrays bezüglich ihres Designs einen unspezifischen Charakter hinsichtlich eines biologischen Experimentes. Obwohl bei den Arrays der aktuellen Generation von Transkriptomarrays gesprochen wird, die alle existierenden Transkripte beherbergen [117], bleibt diese Aussage eine unbestätigte Behauptung. Der Nachweis dazu kann nicht erbracht werden, da potentiell „vergessene“ Transkripte entweder keinen Hybridisierungspartner auf dem Array finden oder aber durch bereits beschriebene Kreuzhybridisierungen zu Signalen führen, die nicht interpretiert oder nicht isoliert werden können.

Hier spielen cDNA-Microarrays ihre größte Stärke aus. Dadurch, dass vor der Erstellung des cDNA Arrays ein entsprechender mRNA Pool isoliert und daraus eine cDNA Bank durch Umschreibung erzeugt wird, ist die Experimentsspezifität eines cDNA Mikroarrays beliebig skalierbar. So konnte in Kapitel VI.A.1 gezeigt werden, dass durch die gezielte, herzspezifische Erstellung des cDNA Pools eine große Anzahl an Transkripten in den Pool eingebracht werden konnte, deren Existenz bislang gar nicht bekannt war. Die gezeigten Microarray-Analysen (vergleiche Kapitel VI.A.3.a) ergaben außerdem spezifische Cluster mit sich stark verändernden Expressionsraten, zu denen die Sequenzen nur zum Teil annotierbar waren. Die Anzahl von ~100 dieser Transkripte, die weiterhin schon als kodierende Transkripte bestätigt werden konnten (entsprechende Peptididentifizierung) verstärkt die Annahme, dass durch spezifische Transkriptpools immer noch neue Transkripte identifiziert werden können, die in ihrer Sequenz bislang noch vollkommen unbekannt sind.

Im Fall des Molches spielt das gewählte Versuchsdesign eine entscheidende Rolle bei der Wahl für cDNA-Microarrays. So lassen cDNA-Arrays immer nur relative Aussagen zwischen zwei zu beobachtenden experimentellen Stadien zu. Bei einer Zeitreihe zur Herzregeneration normiert diese Technologie also bereits auf das ungeschädigte Gewebe. Oligoarrays hingegen liefern absolute Expressionswerte nur eines experimentellen Stadiums. Da im Fall der Herzregeneration eine zeitliche Beobachtungsreihe über den Regenerationsverlauf erstellt wurde, empfahl sich die cDNA-Array-Technologie aus wirtschaftlichen Gesichtspunkten. So wurden für die Beobachtung eines Regenerationszeitpunktes acht Microarrays angefertigt, die sowohl jeweils zwei technische Replikate mit DyeSwap und vier biologische Replikate nach vorherigem poolen umfassten. Um eine vergleichbare Abdeckung mit der Oligoarray Technologie zu erreichen, wären 16 Arrays nötig, da jede Probe einzeln gemessen hätte werden müssen. Die Normalisierung auf den ungeschädigten Zeitpunkt muss außerdem bei den Oligoarrays auf rein rechnerischem Wege erfolgen.

Letztendlich spielt auch in diesem Bezug wieder das Fehlen von kommerziell beziehbaren Produkten für einen „Rand-Modellorganismus“ wie den Molch eine entscheidende Rolle bei der Wahl der Transkriptom-Screening Technologie. Die Konzentration der kommerziellen Anbieter ist auch hier auf Standard Organismen beschränkt.

2. Proteom Analyse mittels SILAC und Massenspektrometrie

Die Untersuchung von Proteomen verschiedener Organismen *in vivo* mithilfe der Massenspektrometrie stellt immer noch eine Herausforderung dar [67,99,118]. Zwar haben sich für Untersuchungen an Zellkulturen und einzelligen Organismen standardisierte Verfahren entwickelt, die auch die Behandlung von genetisch unbekanntem Zellen erlauben, die Hochdurchsatz-Analyse vielzelliger Organismen beschränkt sich aber zumeist immer noch auf Standard-Modellorganismen wie Mensch, Maus, Ratte, Drosophila und Zebrafisch [119]. Für genau diese Organismen liegen auch entsprechende, gut beschriebene und funktionell annotierte Proteindatenbanken vor, insbesondere die *international protein index* (IPI) Datenbanken von EMBL-EBI. Bei diesen Proteinsequenzdatenbanken konnte über stringentes Abgleichen zwischen eben diesen Banken (und den daraus errechneten Massen) mit im Versuch experimentell erzeugten Peptidspektren und Peptidfragmentspektren eine hohe Qualität bezüglich der Genauigkeit und der Spezifität bei Datenbanksuchen erreicht werden. Moderne Hochdurchsatzanordnungen, in denen Auftrennungsverfahren (beispielsweise HPLC) *online* mit Ionisierungsverfahren (beispielsweise ESI) und anschließendem Massenanalysator gekoppelt sind, erlauben die Auflösung von bis zu mehreren tausend Proteinen [120] in einer Versuchsanordnung. Daneben gibt es vielfältige Ansätze, Proteome ohne entsprechende umfangreiche Hintergrundinformation zu untersuchen. In diesem Bereich haben sich drei wesentliche Entwicklungsstränge ausgebildet.

Zunächst ist an dieser Stelle die massenspektrometrische *de novo* Sequenzierung von Proteinen/Peptiden zu nennen. Diese Techniken beruhen darauf, aus MS² Daten die Sequenz von Peptiden abzulesen (ohne Datenbankabgleich). Dieser Vorgang ist sehr stark theoretisch und informatisch geprägt und Gegenstand aktueller Forschung [121,122,123]. Diese Methode wird bereits sehr erfolgreich für Organismen mit weitreichenden genomischen Informationen verwendet, auf deren Basis dann kodierende Bereiche detektiert werden können. Für den Einsatz bei Organismen vollkommen ohne genomischen Hintergrund eignen sich diese Techniken aber noch nicht, obwohl die Identifizierung entsprechender einzelner Peptide bereits *de novo*

Diskussion

gut funktioniert. Das Mapping auf ein funktionelles Protein und insbesondere die Anordnung und Zusammengehörigkeit von Peptiden erweist sich jedoch als schwierig. Insbesondere der Nachweis von detektierten unbekanntem Peptiden auf Transkriptebene, was in der Regel über degenerierte Primer und PCR Reaktionen erreicht wird, schließt diese Techniken für Hochdurchsatzuntersuchungen aus.

Daneben hat es eine Entwicklung gegeben, die zwar die klassische Datenbanksuche einschließt, aber gezielt in einem genomisch unbekanntem Organismus wie dem Molch von evolutionär veränderten Proteinen ausgeht, die mit einer Veränderung der Primärstruktur der Proteine einhergeht. Bei diesem Ansatz wird zunächst auf umfassenden Proteindatenbanken nach zu MS² Spektren passenden Peptiden gesucht. Diese und weitere Hintergrundspektren [124] werden aus der Gesamtmasse aller Spektren entfernt und nur für die übrigen eine *de novo* Sequenzbestimmung durchgeführt. An die daraus resultierenden Peptidsequenzen schließt sich eine Homologiesuche auf Basis des BLAST Algorithmus an, der in der hier verwendeten Variante als MS-BLAST bezeichnet wird [109,124]. Dieser modifizierte Algorithmus ordnet die Gesamtmenge aller *de novo* detektierten Peptide Proteinen zu, wobei Lücken bei der Zuordnung einzelner Peptide nicht erlaubt sind, zwischen den detektierten Peptiden aber beliebig lange Lücken eingefügt werden können. Diese Strategie adressiert bereits die Schwierigkeit des *Mappings* von unbekanntem Peptiden auf gesamte Proteine, indem sie den Schritt der Assemblierung mit der Homologiesuche und somit der Annotation vereinigt. Bei dieser Strategie werden jedoch an zwei Stellen Annahmen durch Annäherung getroffen, die abschließend die Unterscheidung zwischen tatsächlich vorliegenden Peptiden/Proteinen und falschen Interpretationen erschweren. So wird zunächst auf Basis eines MS² Spektrums eine unbekannte „theoretische“ Sequenz abgeleitet, die dann durch Homologiesuche nochmals „aufgeweicht“ wird [109]. Somit kann man bei der Detektion nur eines Peptides zu einem Protein und dann auf Basis dessen Grad der Ähnlichkeit nur schwer entscheiden, ob es sich nun real um dieses Peptid oder um eine Ansammlung von Ungenauigkeiten handelt. In einer Studie zur Einzelligen Grünalge (*Dunaliella salina*) konnten mit diesem Ansatz insgesamt etwa 50 Proteine detektiert werden, deren Sequenz zuvor nicht bekannt war [125].

Der in dieser Arbeit dargestellte Ansatz versucht die Problemstellung der Hochdurchsatz-Analyse von Proteinen in genetisch unbekanntem komplexen Organismen durch eine neue Variation einzelner Bearbeitungsschritte zu begegnen. Dabei sind die Stärken der hier entwickelten Methode vor allem in dem Bereich der sicheren Identifikation und damit auch Annotation einzelner gemessener Peptide zu finden.

Zum einen werden nur Peptide als detektiert und identifiziert angesehen, die strengen Qualitätskriterien/Parametern genügen (hohe Auflösung, hohe Massengenauigkeit, keine variablen Modifikationen). Durch Verwendung der [¹³C] substituierten Aminosäuren kann bei der Identifikation von Peptiden in den Spektren außerdem auf die Möglichkeit der SILAC Isotopencluster-Paarfindung zurückgegriffen werden. Die Annotation zum Peptid erfolgt dann aber nicht, wie im MS BLAST Ansatz, durch einen weiteren Schritt der Ähnlichkeitssuche nach der Detektion. Vielmehr ist die Annotation bereits zuvor auf Basis der zu durchsuchenden Sequenzdatenbank geschehen. In Grenzfällen wird dabei die Annotation durch Mitführung mehrerer gleichwertiger Ähnlichkeiten verifiziert (Mehrfachzuordnung von Peptiden zu Proteinen). Die Möglichkeit, evolutionär veränderte Proteine auf diesem Wege zu detektieren, wird dabei durch zwei parallele Maßnahmen realisiert. Zum einen werden Proteindatenbanken aus verschiedenen Organismen, von sehr nah verwandt (*Ambystoma*, *Xenopus*) bis sehr fern verwandt (*Homo sapiens*), also einer breiten Streuung der Organismen innerhalb der Wirbeltiere, verwendet. Durch den gezielten Einsatz verschiedener Banken sind zu vielen Proteinen diverse „Versionen“ verfügbar (Homologe), in denen erwartungsgemäß einzelne Peptide mit der Proteinsequenz des Molches übereinstimmen.

Zum anderen wird durch den gezielten Einsatz von EST Banken zusätzlich die Detektion von evolutionär stark veränderten Peptiden möglich, die so nur im Molch vorliegen. Somit können Peptide aus der Molch EST Bank konservierte Peptide aus den übrigen verwendeten Banken zu hochwertigen Proteinidentifikationen aufwerten. Auch ist auf diesem Wege eine *quasi de novo* Detektion von Peptiden möglich. Die Auffindung von Peptiden zu EST Datenbanksequenzen, die auf Nukleotidniveau vollkommen ohne Annotation vorliegen, können interessante und bislang unbekannte kodierende Sequenzen enthalten.

Wie diese Arbeit zeigen kann, ist durch die Kombination der beschriebenen Ansätze ein Gesamtergebnis erreichbar, dass grundsätzlich die Hochdurchsatz-Analyse des Molchproteoms zulässt, ein Ziel, das vorher nicht erreicht werden konnte (Insgesamt über 4000 Proteinidentifizierungen). Weiterhin konnte daraus sogar ein quantitativer Ansatz entwickelt werden. Sofern sich dieser quantitative Ansatz final entwickeln lässt, unterscheidet sich eine solche Proteomanalyse im Umfang nicht mehr wesentlich von Untersuchungen in Standard-Modellorganismen wie Maus oder Zebrafisch.

Vergleicht man die neuen Möglichkeiten zur Proteomanalyse mit den bislang verwendeten, klassischen Techniken wie beispielsweise der Detektion von Proteinen durch Bindung entsprechender Antikörper (Immunohistologische Färbung, Westernblot), verdeutlichen sich die

Diskussion

Vorteile der Methode. So konnte in dieser Arbeit gezeigt werden, dass durch eine einzige Anwendung mehr als 90% aller bislang bekannten Proteine des Molches detektiert und identifiziert werden konnten. Bedenkt man die zeitlich und experimentell aufwändige Generierung von „eindeutigen“ Antikörpern und letztendlich auch die Kosten einer solchen Herstellung, stellt die hier entwickelte Methode der partiellen Markierung von Proteinen *in vivo* mit stabilen Isotopen einen enorme Verbesserung im Zusammenhang mit der Analyse von Proteomen im Molch dar. Jedoch ersetzt diese Methode nicht die klassischen Techniken, da beispielsweise die Lokalisation oder die differentielle Expression innerhalb des untersuchten Gewebes nicht dargestellt werden kann.

C. Bewertung der Kandidatenlisten aus biologischen Gesichtspunkten

Wie alle Screening Verfahren sind auch Sequenzierungen, Microarrays und Proteomuntersuchungen mit Focus auf einen höchstmöglichen Durchsatz bezüglich der Anzahl an zu untersuchten Kandidaten nur ein erster Schritt zu einer tiefgehenden biologischen Funktionsanalyse. Jedoch können diese Screeningverfahren einen guten, systembiologischen Überblick ermöglichen, auf dessen Basis sich durch geeignete Methoden hochwertige Kandidatenlisten erzeugen lassen, die sich für eine tiefgehende Analyse empfehlen. Je nach Fragestellung können einmal erstellte Datensätze auch nach verschiedenen Gesichtspunkten ausgewertet werden.

In dieser Arbeit wurden aus erstellten Versuchsreihen exemplarisch die frühen Zeitpunkte der Herzregeneration herausgegriffen. Diese frühe Phase ist durch klassische Prozesse der frühen Wundantwort und Reaktionen des Immunsystems geprägt. Jedoch liegt die Vermutung nahe, dass gerade zu diesen frühen Zeitpunkten schon Signalprozesse eingeleitet werden, die sich in irgendeiner Form von klassischen Prozessen der Wund- und Immunantwort unterscheiden. Genau diese Unterschiede sollten dann auch die besondere Fähigkeit des Molches ausmachen, die anstatt zu der Bildung von Narbengewebe zu der Dedifferenzierung, Proliferation und Regeneration des geschädigten Gewebes überleiten. Umfangreiche Screenings auf Transkriptom- und Proteomebene haben dabei die Möglichkeit zu einem übergreifenden Vergleich zwischen bekannten molekularbiologischen Abläufen und im Molch veränderten Abläufen ermöglicht.

1. Die frühe Herzregeneration auf Transkriptomebene

Aus der Transkriptomanalyse, die in dieser Arbeit für die frühen Zeitpunkte der Regeneration zwei Stunden und sechs Stunden nach Schädigung durchgeführt wurde, sollen an dieser Stelle einige Kandidatengene näher diskutiert werden, die über einen subtraktiven Ansatz zum scheinoperierten Gewebe aufgefallen sind. Im Einzelnen sind dies die Transkripte zu sechs annotierbaren Sequenzen (Vergleiche Kapitel VI.A.3.a). So fällt zunächst die Aufregulation der *phosphatase 1, regulatory (inhibitor) subunit 15A* auf. Diese Phosphatase wird auch als *Growth arrest and DNA damage-inducible protein* (GADD34) bezeichnet. GADD34 ist ein Gen, welches als Antwort auf Stress von Zellen exprimiert wird. Es wird davon ausgegangen, dass GADD34 über Interaktion mit dem Protein p53 [126], einem bekannten Kandidaten bei der Vermittlung von DNA Reparaturprozessen, die DNA Reparatur in gestressten Zellen einleitet. Dies geschieht über eine Aufregulierung von GADD34. Das Protein p53, ein Transkriptionsfaktor, regelt dabei über Aktivierung von p21 die Blockierung des Zellzyklus (Verlangsamung der Mitose) und die Einleitung der DNA Reparaturmechanismen. Interessanterweise konnte GADD34 im normalen menschlichen Herzgewebe auf hohem Expressionsniveau im Cytoplasma von Kardiomyozyten nachgewiesen werden. Tritt jedoch eine ischämische Schädigung auf, wird GADD34 beim Mensch im geschädigten Bereich abreguliert [127]. Diese Abregulierung wurde im Zeitraum ein Tag bis sieben Tage nach Schädigung beobachtet. Ungeschädigte Bereiche halten jedoch ein hohes Expressionsniveau an GADD34. Im Molchherz ist GADD34 hingegen zum frühen Zeitpunkt 2 Stunden Regeneration bereits stark aufreguliert, wobei dieser Zustand der Aufregulation über den Verlauf der Regeneration bis etwa 14 Tage nach Schädigung anhält. Erst danach fällt GADD34 auf ein normales Expressionsniveau zurück. An dieser Stelle wäre eine Aufklärung der Lokalisation (Histologie) von GADD34 im Molchherz von besonderem Interesse. Auf dem Microarray läßt sich nur eine relative Aufregulierung bezogen auf das gesamte Molchherz zeigen. Ob es aber zu unterschiedlicher Expression zwischen geschädigten und ungeschädigten Bereichen des Herzens kommt, kann mit dieser Technik nicht aufgeklärt werden. Jedoch ist dieses Transkript ein interessanter Kandidat für weitere Untersuchungen, da hier ein potentieller Unterschied zwischen dem Verhalten von Molchkardiomyozyten, die nach Schädigung zur Proliferation übergehen, und menschlichen Kardiomyozyten, die nach Schädigung zur Hyperthrophie übergehen, vorliegt.

Auch für das *B-cell translocation gene 2* (BTG2) liegt beim Molch zum Zeitpunkt zwei Stunden Regeneration eine massive Aufregulation, für den Zeitpunkt 6 Stunden Regeneration eine mittlere Aufregulation vor. Bereits zum Zeitpunkt 1 Tag Regeneration und im folgenden

Diskussion

Regenerationsverlauf ist BTG2 unauffällig reguliert. BTG2 wurde als frühes Antwortsignal auf Wachstumssignale identifiziert [128]. Dies erscheint sehr interessant, da gezeigt werden konnte, dass BTG2 ebenfalls in Interaktion mit p53 steht [129] und somit eine Verkettung mit Prozessen der Verzögerung von Zellwachstum besteht. Dabei stellt BTG2 ein Effektormolekül von p53 dar, das durch aktiviertes p53 verstärkt transkribiert wird. Es wurde gezeigt, dass eine Reduktion von p53 immer auch eine direkte Reduktion der BTG2 Expression nach sich zieht. Daraus resultierend wurde eine erhöhte Anfälligkeit von Zellen (Fibroblasten) für krankhafte, krebsartige Veränderungen abgeleitet [130]. Eine entsprechende onkogene Transformation der Fibroblasten wurde bei diesen Untersuchungen durch Ras ausgelöst. Die künstliche Repression von BTG2 führt zur Aufregulierung von Cyclin D1 und Cyclin E1 (Anregung der Zellteilung), sowie einer Hyperphosphorylierung des Retinoblastoma Proteins (Rb). Rb inhibiert die Zellteilung in Muskelzellen und wird durch Phosphorylierung inaktiviert. Daneben konnte für BTG2 eine entwicklungspezifische Funktion gezeigt werden. In BTG2 KO Mäusen wurde nachgewiesen, dass BTG2 über Smad1 und Smad8 die BMP Signalkaskaden reguliert und das Fehlen von BTG2 zu Missbildungen führt. Bei Fröschen wurde außerdem bei translationaler Hemmung von BTG2 eine Missbildung der *Chorda vertebralis* nachgewiesen. Im Fall der Herzregeneration des Molches bedarf die Expression von BTG2 einer genaueren Untersuchung. So ist besonders auffällig, dass BTG2 nur sehr kurz aufreguliert ist, bereits nach wenigen Stunden die Aufregulation abebbt. Wenn BTG2 auch beim Molch eine Hemmfunktion bezüglich des Zellwachstums und der Zellteilung (Hemmung von Cyclin D1 und E1 sowie Aktivierung von Rb) nach Stress hat, scheint es nach anfänglicher Aktivierung des selbigen eine aktive Abregulation bzw. Hemmung von BTG2 zu geben, was auf eine anschließende verringerte Hemmung der Zellproliferation schließen lassen könnte.

Als weiterer, regenerationsspezifischer Kandidat ist die *ADAM metalloproteinase with thrombospondin type 1 motif 1* (ADAMTS1) aufgefallen. ADAMTS1 ist ein Enzym, welches die Extrazelluläre Matrix degradiert und gehört zu den Matrix Metalloproteasen (MMPs). Diese spielen bei der Reparatur von durch Infarkt geschädigtem Gewebe eine Rolle [131], indem sie einen Einfluss auf die Remodellierung des Ventrikels nehmen [132,133]. Auch wurde eine Funktion als Angiogenese Inhibitor nachgewiesen [134]. Es konnte gezeigt werden, dass ADAMTS1 in ungeschädigtem Endothelgewebe auf sehr niedrigem Niveau exprimiert wird, dagegen im Myokard keine Expression von ADAMTS1 nachzuweisen ist [135]. Eine Aufregulierung von ADAMTS1 konnte aber nach experimentellem Koronararterienverschluss bei Ratten bereits nach 3 Stunden detektiert werden. Über in situ Hybridisierungen konnte gezeigt werden, dass sowohl das Endothel als auch das Myokard ADAMTS1 exprimieren, dabei das

Infarktgewebe selbst den Schwerpunkt bildet [135]. Eine spätere Untersuchung zu den Zeitpunkten 2, 7, 14 und 28 Tagen nach Infarktsetzung zeigte keine Deregulation von ADAMTS1 im Myokard und Endothel mehr [135]. Diese frühe und schnelle Aufregulation unterscheidet sich wesentlich vom Expressionsmuster anderer MMPs [131]. Im regenerierenden Molchherzen zeigt sich eine starke Aufregulation bereits nach 2 Stunden, die jedoch nach 6 Stunden Regeneration das Maximum erreicht und erst nach 14 Tagen Regeneration wieder auf das normale Expressionsniveau abfällt. Dieser im Vergleich zur Ratte veränderte Expressionsverlauf, insbesondere hinsichtlich der zeitlichen Komponente der erhöhten Expression läßt auf eine erhöhte Aktivität im regenerierenden Molchherzen schließen. Dabei ist zu vermuten, dass Prozesse der Um- und Neustrukturierung einen massiven Umbau der bestehenden Zellen im Molchventrikel und der sie einbettenden Umgebung (Extrazelluläre Matrix) notwendig ist. Die bei Molchen nicht einsetzende Narbenbildung durch verstärkten Kollagenaufbau könnte auch auf eine verstärkte bzw. einen verlängerten Einsatz von entsprechenden Kollagen abbauenden Enzymen wie den MMPs sein.

Das *cysteine-rich angiogenic inducer 61* (CCN1) Protein gehört zur CCN Familie, die sechs Mitglieder beinhaltet. Diese Familie trägt charakteristisch die vier Domänen *insulin like growth factor binding protein*, den *von Willebrand factor type C*, die *thrombospondin type 1 domain* und die *carboxy-terminal domain*. Auch ein Protein aus der massenspektrometrischen Analyse wie in VI.D.2 beschrieben scheint nach einer Sequenzanalyse dieser Familie zugehörig zu sein. Die Mitglieder der CCN Familie werden sekretiert und sind in der Extrazellulären Matrix und an die Zellmembran gebunden (Integrine) vorzufinden [136]. Das CCN1 Protein bindet über die 4 Domänen an IGF, mehrere Integrine ($\alpha_v\beta_3$, $\alpha_v\beta_5$, $\alpha_6\beta_1$, $\alpha_{11b}\beta_3$, $\alpha_M\beta_2$) und über den Corezeptor *Heparan Sulphate ProteoGlycans* (HSPG) [137]. Je nachdem, welches Integrin gebunden wird, löst CCN1 unterschiedliche Prozesse in der Zelle aus. Dabei werden die Signalwege Wnt, NF- κ B, Tyrosin Kinase und Akt verwendet (In Abhängigkeit vom Zelltyp). Es wurde gezeigt, dass je nach Zelltyp gebundenes CCN1 bei der Embryogenese, Tumorgenese, Angiogenese und Apoptose mitwirkt.

CCN1 ist während der Embryogenese in Maus und Frosch genau beschrieben worden. Dabei konnte gezeigt werden, das CCN1 bei Fröschen mit dem Wnt Signalweg interagiert und bei Suppression von CCN1 die Sekundärachsenentwicklung über Wnt8 während der Entwicklung gehemmt wird [138]. Bei Mäusen ist das Fehlen von CCN1 im Embryonalstadium letal, da es zu massiven Fehlentwicklungen bei der Vaskularisierung der Plazenta kommt [139]. Neben diesen Entwicklungsstudien wurde CCN1 primär in Untersuchungen zu Brust- und Prostatakrebs beschrieben [140,141]. So führt die Überexpression von CCN1 in Brustzellen (MCF-12A) nach

Diskussion

Induktion in Nacktmäuse zu Tumorentwicklungen [142]. Dagegen zeigen Prostatakarzinome eine Abregulation von CCN1 [143].

Eine Studie, die sich mit einer der wenigen regenerativen Fähigkeiten der Säuger auseinandersetzt, zeigt zudem die Funktion von CCN1 bei Frakturen. Hier konnte gezeigt werden, dass CCN1 bei der Angiogenese während der Frakturheilung eine Rolle spielt. CCN1 tritt dabei mit dem *Vascular Endothelial Growth Factor* (VEGF) in Interaktion. Es konnte eine parakrine Kopplung von VEGF auf CCN1 gezeigt werden, bei der die Gabe von VEGF direkt eine massive Aufregulierung von CCN1 in Osteoblasten bewirkt [144]. Das erhöhte Expressionsniveau von CCN1 in der extrazellulären Matrix verursacht eine Migration von Endothelzellen. Diese CCN1 vermittelte Zellmigration konnte *in vivo* als primärer Stimulus für Angiogenese gezeigt werden [144], die für die Frakturregeneration essentiell ist. Letztendlich konnte auch gezeigt werden, dass durch die Suppression von CCN1 die Frakturregeneration zum Zeitpunkt 14 Tage stark verzögert war [144]. In dieser Arbeit konnte gezeigt werden, dass während der Herzregeneration des Molches die Expression von CCN1 ihr Maximum schon zum Zeitpunkt der zweistündigen Regeneration erreicht. Dann fällt das Expressionsniveau bis zum Zeitpunkt 24 Stunden Regeneration leicht ab (log Ratio:0,6), um dann abermals noch einmal kontinuierlich bis zum Zeitpunkt 7 Tagen Regeneration stark anzusteigen und sich dann im Verlauf bis zum Zeitpunkt 14 Tage wieder zu normalisieren. Dieses Expressionsmuster unterscheidet sich von dem des in der Proteomanalyse detektierten CCN Familienmitgliedes (siehe VI.D.2), das nur in den sehr frühen Zeitpunkten bis hin zu einem Tag Regeneration eine Aufregulation zeigt. Zum Expressionsverhalten von CCN1 während der Frakturregeneration liegen in der beschriebenen Studie keine Erkenntnisse vor. Jedoch impliziert das Expressionsverhalten von CCN1 im Molch eine interessante Regulierung desselbigen, welche genauer untersucht werden sollte.

Das *Ras like G protein Rad* (nRad) zeigt wie die bereits beschriebenen Kandidaten eine rasche Aufregulierung in den frühen Zeitpunkten zwei und sechs Stunden Regeneration, die zum Zeitpunkt sechs Stunden ihr maximales Expressionsniveau erreicht und dann innerhalb der ersten beiden Tage wieder auf das normale Expressionsniveau zurück sinkt. Die weitere Regeneration verläuft unauffällig. Das nRad wurde bezüglich der Regeneration bereits im japanischen Molch *Cynops pyrrhogaster* charakterisiert [145] und auch mit der *Cynops pyrrhogaster* Sequenz annotiert. nRad zeigte eine Aufregulation während der Beinregeneration des Molches. Die Überexpression erfolgte beim Bein bereits zum Zeitpunkt 4 Stunden nach Amputation und sank dann zwischen 11 und 21 Tagen wieder auf das Normalniveau. In ungeschädigtem Gewebe konnte keine nRad Expression gezeigt werden. Über *in situ* Hybridisierung konnte nRad im Umfeld von Zellkernen der Muskelzellen in Nähe der

Amputationsfläche lokalisiert werden. In distaler Richtung nahm die Expressionsstärke ab, so dass an der AEC keine Expression von nRad mehr via *in situ* Hybridisierung mehr gezeigt werden konnte. Mittels einer RT-PCR konnte jedoch im Blastem eine geringfügige nRad Expression gezeigt werden. Die Zugabe von Retinolsäure, die eine verstärkte Dedifferenzierung im regenerierenden Bein nach sich zieht [146], erhöhte gleichzeitig die Expression von nRad in proximal gelegenen Beingewebe und verlängerte außerdem die Phase der nRad Überexpression über den zuvor beschriebenen Zeitraum hinaus. Aus diesen Ergebnissen wurde eine Korrelation zwischen der Stärke der Dedifferenzierung und der Expression von nRad abgeleitet. Außerdem wurde aus den Ergebnissen abgeleitet, dass Rb zum Zeitpunkt der Regeneration phosporiliert und damit inaktiviert vorliegt, was eine Proliferation von Zellen voraus setzt. Rb wird normalerweise konstant in Muskelzellen exprimiert [145].

Der letzte Kandidat ist das *solute carrier family 7 member 3* Transkript, welches auch als *Cationic amino acid transporter 3* (CAT3) bezeichnet wird. Dabei ist die Bezeichnung *Cationic amino acid transporter/glycoprotein-associated* die Familienbezeichnung, die aus 14 Mitgliedern besteht. Die Expression von CAT3 wurde für nahezu alle Gewebstypen in der embryonalen Maus gezeigt, wobei ein Schwerpunkt der Expression im Hirn festzustellen ist [147]. Über die genaue interagierende Funktion von CAT3 im zellulären Zusammenhang ist bislang wenig bekannt, wenn auch die Funktion des Arginin Transportes bereits tiefgehend beschrieben wurde [148]. Bei der Herzregeneration des Molches ist CAT3 bereits zum Zeitpunkt zwei Stunden Regeneration aufreguliert und erreicht zum Zeitpunkt sechs Stunden Regeneration das Expressionsmaximum im regenerativen Verlauf. Nach vier Tagen ist nur noch eine sehr leichte Aufregulierung festzustellen, die sich dann bis zum siebten Tag der Regeneration normalisiert.

Neben den hier beschriebenen deregulierten Transkripten wurden weitere Kandidatengene erwartet, die bei der Schädigung von Säugerherzen eine Rolle bei der frühen Wundantwort spielen. Es wurde erwartet, dass die Expression dieser Transkripte mithilfe des Zeitreihen-Microarrayexperimentes aufgezeigt werden kann. Beispiele für diese Kandidaten sind die *Toll-like receptor* (TLR) kontrollierten Signalwege, die Signalwege der Komplementkaskade und Signalwege, die über *reactive oxygen species* (ROS) gesteuert werden. Dabei wirken alle diese inflammatorischen Signalwege über die Aktivierung des *nuclear factor* (NF) κ B Systems.

So ist beispielsweise bekannt, dass innerhalb der TLR Familie (12 Mitglieder) der TLR2 und TLR4 im Herzen von Mäusen und Ratten exprimiert sind. Dabei wird TLR4 bei einem Infarkt bei Maus und Ratte verstärkt exprimiert [149]. Mäuse ohne TLR4 hingegen zeigten bei Initiierung eines Infarktes eine reduzierte Infarktgröße und eine abgeschwächte Inflammatorische Antwort [150].

Diskussion

Dagegen zeigten TLR2 Mäuse keine Reduktion der Infarktgröße, aber ein reduziertes Maß an Fibrose im vom Infarkt angrenzenden Gewebe des Herzens [151].

Auch spielen beim Verlauf der Regeneration beispielsweise die Wnt Signalwege über *beta-catenin* [152] und über BMP [153,154] aktivierte TGF-beta Signalwege eine essentielle Rolle.

Jedoch konnten nur sehr wenige dieser angesprochenen sekretierten Signalproteine und Rezeptoren überhaupt auf den verwendeten Microarrays detektiert werden. So finden sich beispielsweise in der Gesamtmenge aller annotierten Molchsequenzen, die auf dem Array hinterlegt sind, nur zwei Toll-like Rezeptoren (TLR5 und TLR22), die jedoch keine Deregulation während des gesamten Regenerationsverlaufes zeigen. Auch finden sich zu den Genen der Komplementkaskade und der Wnt und TGF-beta nur sehr wenige Transkripte. So konnte anhand der Microarrays *beta-catenin* annotiert (ref|NP_445809.2 , *Rattus norvegicus*, e-value: 1e-179) und eine schwache Aufregulierung im späten Verlauf der Reperation gezeigt werden (7-35 Tage, Aufregulation 0.8).

Das insgesamt nur so auffallend wenige der erwarteten Kandidaten in der sequenzierten cDNA Bibliothek vorkommen, könnte unter anderem durch drei Punkte der angewandten Arbeitsfolge bedingt sein.

So wurden zunächst bei der Erstellung der cDNA Bank sehr kurze und sehr niedrig abundante Transkripte benachteiligt. Dies geschah dadurch, dass sehr kurze cDNA (nicht im Bereich 1000 - 1200 nt) in einem Aufreinigungsschritt entfernt wurden und sehr seltene RNAs nach dem Mischen der Gewebe aus den Regenerationszeitpunkten durch viele verschiedene, in den einzelnen Zeitpunkten hoch abundanten Transkripten, überlagert wurden. Diesem bekannten Phänomen sollte durch die Normalisierung der cDNA Bank entgegen gewirkt werden.

Daneben war die Auswahl der zu sequenzierenden Bibliothekskoordinaten (wie in VI.A.5 beschrieben) sehr selektiv für Transkripte, die auf dem Microarray eine signifikante Deregulation, oder aber eine signifikante Menge an in der zu untersuchenden Probe vorkommendem Transkript zeigten. Diese Art der Auswahl zur Sequenzierung versprach zwar eine Anreicherung an Transkripten, die während des Verlaufes der Regeneration überhaupt eine Deregulation erfahren oder zumindest generell stark exprimiert werden, benachteiligt aber all die Transkripte, die entweder nicht deutlich dereguliert waren oder aber nur sehr niedrig abundant auf dem Array oder innerhalb des untersuchten Gewebes vorkamen. Die Stärke dieses Effektes ist nur schwer abzuschätzen. Betrachtet man jedoch die große Anzahl an Sequenzen innerhalb der cDNA Bibliothek, die nur als einzelner Read vorliegen, so darf davon ausgegangen werden, dass noch ein großer Teil an interessanten und regenerationsrelevanten Transkripten im nicht sequenzierten Teil der Bibliothek liegt. Dies lässt sich auch dadurch erklären, dass ein

Transkript, das nur wenige Male oder sogar nur einmal auf dem Array vorkommt viel seltener ein signifikantes Deregulationssignal über die Wiederholungen der Experimente hinweg ergibt, als ein Transkript, das in mehreren Kopien auf dem Array vorkommt. Für einzeln vorkommende Transkripte spielt letztendlich auch die Qualität des Microarrayspottings eine essentielle Rolle. Zuletzt spielt neben den beiden erwähnten Aspekten auch noch der, der Sequenzierung nachgelagerte Prozess, der Annotationsfindung eine Rolle. Wie in dieser Arbeit beschrieben, kann nur für etwas mehr als die Hälfte aller Transkripte überhaupt eine Annotation gefunden werden. Dies bedeutet aber nicht, dass die übrigen Transkripte nicht auch für bereits bekannte Proteine kodieren. Es darf angenommen werden, dass durchaus einige Sequenzen des Molches weniger stark konserviert sind und so über Homologiesuchen nicht identifiziert werden können.

2. Die frühe Herzregeneration auf Proteinebene

Die Erstellung eines Herzproteomes oder eines Proteomes einer speziellen Zellpopulation des Herzens ist kürzlich in der Literatur, insbesondere mit Hochdurchsatzverfahren, mehrfach beschrieben worden [155,156,157,158]. Dabei wurden generell quantitative Ansätze, die Herzentwicklung oder Hypertrophie im Herzen betrachtet. Als Modellorganismen kamen dabei Hühner [155] (Embryonen) Mäuse [156,157] oder Ratten [158] zum Einsatz. Ein Proteom eines sich regenerierenden Herzens wurde bislang nicht veröffentlicht. Die in dieser Arbeit entwickelte Methode zur Proteomanalyse des Molchherzens stellt erstmals einen Hochdurchsatz-Datensatz zu diesem biologischen Vorgang zur Verfügung. Dabei hat sich gezeigt, dass durch entsprechende Verfahren ein Durchsatz erreicht werden kann, der mit anderen Untersuchungen in Standardmodellorganismen in diesem Bereich vergleichbar ist. Neben der rein quantitativen Beschreibung des erstellten Datensatzes zur frühen Wundantwort des Molchherzens zu Anfang der Herzregeneration konnten über einen differenziellen Ansatz auch interessante Kandidaten für eine weitergehende Funktionsanalyse aufgezeigt werden. Unter differenziellem Ansatz ist dabei die Gruppierung der Proteine in „exklusiv im ungeschädigten Herz“ oder „exklusiv nur im sechs Stunden regenerierenden Herz“ zu verstehen. Eine genaue Quantifizierung aller identifizierten Proteine durch Ausnutzung der SILAC Markierung fehlt aber noch. Da diese Arbeit verstärkt auf die Identifizierung neuer Proteine und Regeneration spezifischer Proteine abzielt, sollen die Kandidaten der zweiten genannten Gruppe im Weiteren diskutiert werden. Diese nähere molekularbiologische Betrachtung einzelner Kandidaten dient der Einordnung der Proteine in potentiell während der Regeneration ablaufende Prozesse. Dabei werden jedoch Kandidaten außer Acht gelassen, die bekanntermaßen eine Abregulation während der Regeneration zeigen

Diskussion

müssen. Dies sind vor allem die Proteine des Sarkomers, die im Verlauf der Dedifferenzierung stark abreguliert werden. Dass diese Proteine aber trotz starker Abregulation nicht in der Gruppe der „exklusiv im ungeschädigten Herzen“ vorkommenden Proteine enthalten sind, liegt wahrscheinlich an zwei Gründen. So sind die Proteine des Sarkomers grundsätzlich im Muskelgewebe sehr hoch abundant vertreten, so dass eine Detektion immer an der Sättigungsgrenze liegt. Dadurch kann eine Deregulation dieser Proteine wahrscheinlich nicht detektiert werden. Daneben ist das getestete regenerative Gewebe immer auch mit ungeschädigtem Herzgewebe kontaminiert, da eine vollständige Schädigung des Herzmuskels, aufgrund des notwendigen minimalen Funktionserhaltes zum Überleben des Molches, nicht möglich ist.

Betrachtet man die erwähnte Gruppe der exklusiv im regenerierenden Gewebe identifizierten Proteine (Auf Basis der normalisierten, translatierten Molch EST Sequenzbank), so findet man 16 hoch qualitativ detektierte Proteine. Erweitert man diese Gruppe um die Proteine, die nur mit einem *unique* Peptid identifiziert wurden, bzw. neben dem regenerierenden Herzen auch noch im schweren regenerationslastigen Pool detektiert wurden, finden sich 90 unterschiedliche Proteine. Diese Gruppe kann sinnvollerweise zusammengefasst werden, da der schwere Pool zu einem großen Teil aus regenerierenden Geweben des Molches zusammengestellt ist. Diese Gruppe darf jedoch nicht als ausschließlich nur im regenerierenden Herzen vorkommende Proteine angesehen werden, jedoch ist die Wahrscheinlichkeit dazu erhöht. Eine Untersuchung dazu, wie in VI.B.3 dargestellt, konnte diesen Sachverhalt zeigen. Eine biologische Interpretation der Kandidaten unterstützt diese These. Analysiert man die genannte Gruppe auf die Funktionalität und Interaktion der beteiligten Proteine, so findet sich zunächst eine Gruppe von Proteinen, die mit Endoplasmatischem Retikulum Stress (ER Stress) in Verbindung gebracht werden. Dies sind unter anderem *Wolframin* [159], *Hsp40* [160], *SEC13-like 1* [161], *cysteine-rich with EGF-like domains 2* (CRELD2) und *UBX domain-containing protein 4* [162]. Das Endoplasmatische Retikulum (ER) ist ein wichtiges Zellorganell bezüglich der Faltung und Modifikation von Transmembran- und sekretierten Proteinen [163]. In Ausnahmefällen führen Fehlfunktionen des ER zu Ansammlungen von falsch, oder nicht gefalteten Proteinen, welche als ER Stress bezeichnet werden. ER Stress führt zu Reaktionen entsprechender Sensoren, die als PERK [164], IRE [165], und ATF6 [166] bezeichnet werden. Dabei liegt ATF6 als Transmembranprotein des ER vor. Im Falle von ER Stress wandert ATF1 zum Golgiapparat und wird dort von Proteasen aktiviert. Nach dieser Aktivierung dient ATF selbst als Aktivator für mehrere Gene, insbesondere Chaperone. Es konnte gezeigt werden, dass ATF1 ebenfalls ein Aktivator für das CRELD2 Protein darstellt [167]. Das CRELD2 Protein ist ausschließlich im

schweren Pool und im regenerierenden Herzen vorzufinden. Betrachtet man zusätzlich das Arrayexperiment, so kann eine leichte aber signifikante Aufregulierung zwischen 24 Stunden und vier Tagen Regeneration festgestellt werden. Für Wolframin zeigt sich eine Aufregulierung zum Zeitpunkt sieben Tage Regeneration. Betrachtet man die Funktion des ER, so kann mit dem Wissen über die Aufregulation mehrerer sekretierter Proteine (Transkriptionsanalyse) und der generell gesteigerten Proteinsynthese zu Beginn der Herzregeneration von verstärkter Aktivität des ER ausgegangen werden. Insbesondere die Phase der Umprogrammierung der Zellen (Dedifferenzierung), die durch das Auftauchen spezifischer Marker gekennzeichnet ist [56], geht hier mit der Aktivierung von ER verbundenen Proteinen gleich zu Beginn der Regeneration einher. Es kann davon ausgegangen werden, dass insbesondere der Prozess der Dedifferenzierung auch eine umfassende Neusynthese von Proteinen, wie auch eine enorme katabolische Aktivität mit sich bringt.

Bei weiterer Betrachtung fällt außerdem das *Small muscle protein X-linked (SmpX)* in der Gruppe der potentiell regenerationspezifischen Proteine auf. Dieses Protein konnte wiederum nur im schweren Pool und im regenerierenden Herz detektiert werden. Betrachtet man zusätzlich die Array Ergebnisse, so kann man eine leichte Aufregulierung zum Zeitpunkt zwei Stunden Regeneration feststellen, die sofort wieder zurückgeht, und später zum Zeitpunkt sieben Tage Regeneration eine leichte Abregulierung erkennen. Das SmpX Protein wird in der Literatur neben der Detektion in der Skelettmuskulatur und den Pulmonarvenen häufig mit dem Vorkommen in der Herzmuskulatur verknüpft. Es konnte gezeigt werden, dass SmpX im Zusammenhang mit dem Homeobox Gen *Nkx2-5* steht und bei dessen Fehlen nachweislich abreguliert wird [168]. *Nkx2-5* wurde ursprünglich in *Drosophila* entdeckt und als essentiell für die Herzentwicklung eingestuft [169]. Außerdem wirkt sich das vollkommene Fehlen von *Nkx2-5* nachweislich auf die Regulation einer Reihe von Transkriptionsfaktoren aus, unter anderem *N-myc* [170]. Im Mausherz wird SmpX ausschließlich im arbeitenden Myokard der Vorhöfe und des Ventrikels exprimiert. Obwohl das Fehlen von SmpX in Mäusen keinen Phänotyp erzeugt, zeigte die Überexpression *in vitro* eine Beeinträchtigung der Fähigkeit zur Zellfusion bei Myoblasten. Außerdem wurden die Aktivität der Transkriptionsfaktoren *nuclear factor of activated T cells (NFAT)* und *myocyte enhancer-binding factor (MEF)* beeinträchtigt. Das Protein NFAT konnte ebenfalls auf Transkriptionsebene detektiert werden. Es zeigt über den Regenerationsverlauf aber keine Deregulierung.

Das Protein *N-myc (and STAT) Interactor (NMI)* wurde ebenfalls in der Gruppe der Proteine, die nur im schweren Pool und im regenerierenden Herzen vorkommen, detektiert. Obwohl der

Diskussion

Namen NMI dieses impliziert, handelt es sich nicht um einen ausschließlichen Interaktionspartner von n-Myc, sondern wird häufig auch als aktiver Partner von *interferon-induced protein 35 (IFP35)* [171] geführt. Daneben wurden Interaktionen mit c-Myc [172], Sox-10 [173], c-fos [174] und allen Stats außer Stat2 [175] gezeigt. Im Molchherzen wurde NMI nur im schweren Pool und im 6 stündigen Regenerationszeitpunkt des Herzens detektiert. Auf Array Basis zeigt NMI keine Deregulation. Besonders gut ist die Funktion von NMI in Brustkrebszellen beschrieben. Hier wurde gezeigt, dass NMI im umgekehrt proportionalen Verhältnis zur Malignität der Zellreihe steht. Dabei ist das Expressionsniveau von NMI direkt über IFN- γ reguliert [175]. Auch in malignen Brustkrebs Zelllinien mit reduzierter NMI Expression konnte über IFN- γ Zugabe eine Aufregulierung von NMI erreicht werden. Weiterhin konnte gezeigt werden, dass NMI selbst in seiner Form als Transkriptions-Cofaktor die Expression von Dickkopf 1 (Dkk1) steuert [176]. Dies ist dahingehend interessant, da Dkk1 ein bekannter Inhibitor des Wnt/beta-catenin Signalweges ist.

Ein weiterer Kandidat ist das *cysteine and glycine-rich protein 2 (CSR2)*, welches auch ausschließlich im sechs Stunden regenerierenden Herzgewebe detektiert wurde. CSR2 gehört zur CSR Familie, die eine Gruppe von Proteinen mit LIM Domäne umfasst. Dabei ist die LIM Domäne eine Zink Finger Domäne, die die Bindung des LIM-Proteines an DNA, RNA, Proteinen oder Lipidsubstraten erlaubt. Dadurch wirken Proteine mit LIM Domäne bei der Genexpression, Zytoskelettorganisation und der Entwicklung. Das CSR2 Protein selbst trägt zwei LIM Domänen mit Zink Bindungskapazität. Ein verstärktes Vorkommen von CSR2 wurde in glatter Muskulatur gezeigt [177,178,179]. Dabei war eine Expression im Herzen von Ratten und Menschen nicht nachweisbar (Methode Northern Blot, [179]) Es wurde gezeigt, dass bei Ratten nach Verletzung der karotiden Arterie die glatten Muskelzellen einen phänotypischen Wechsel von differenzierten Zellen hin zu dedifferenzierten, proliferierenden Zellen vollziehen. Dabei erreicht die Proliferationsrate in Ratten zwischen den Zeitpunkten 2 Tage und 4 Tage nach Schädigung ihr Maximum [180]. Gleichzeitig sinkt bei Schädigung der karotiden Arterie bei Ratten die Expressionsrate von CSR2 zwei Tage nach Schädigung um 60% und bleibt über den achten Tag nach Schädigung hinaus abreguliert. Daraus wurde ein direkter Zusammenhang zwischen Proliferation von glatten Muskelzellen und der Abregulierung von CSR2 abgeleitet [179]. In diesem Zusammenhang steht die ausschließliche Detektion von CSR2 im frühen regenerierenden Herzgewebe des Molches im Gegensatz zu bereits bekannten Informationen zu CSR2. Dabei muss beachtet werden, dass das Molchherz zum größten Teil aus Myozyten besteht und ein Sauerstoff versorgendes Koronargefäßsystem fehlt. Dennoch existiert eine Auskleidung des Ventrikels, der Vorhöfe und des Truncus mit einer Endokard ähnlichen

Zellschicht. Vergleicht man die gegebenen Expressionsdaten von Ratten nun mit den Expressionsdaten auf Transkriptionsebene des Molchherzen ist die entsprechende mRNA ab dem Beobachtungszeitpunkt 6 Stunden bis zum Beobachtungszeitpunkt 7 Tage aufregliert. Vergleicht man an dieser Stelle die Expression weiterer typischer Glattmuskelzellproteine wie *Transgelin* oder *smooth muscle actin*, so zeigen auch diese Proteine im Regenerationsverlauf des Herzens auf Transkriptionsebene eine starke Aufregulierung in der Regeneration des Herzens zwischen sieben und 14 Tagen nach Schädigung (Persönliche Mitteilung Dr. Thilo Borchardt, MPI-BN). Dieses Verhalten lässt darauf schließen, dass entweder Glattmuskelzellen im Herzen selbst vorliegen und während der Regeneration ein Expressionsmuster zeigen, welches sich von dem, wie es in anderen Organismen wie der Ratte vorliegt, unterscheidet. Andernfalls könnten diese typischen Glattmuskelzellproteine im Molchherzen aber auch nicht ausschließlich für diese Zellen spezifisch sein und könnten daher eine gesonderte Rolle bei der Regeneration spielen. Bekannt ist aber auch, dass Kardiomyozyten unter Stress durchaus andere Glattmuskelmarker wie *alpha-smooth muscle actin* exprimieren [181] und eine Überexpression von CSRP2 daher rühren könnte.

Der letzte exklusiv im 6 Stunden regenerierendem Gewebe vorkommende Kandidat ist das *RAB22A member RAS oncogene family Protein* (RAB22A). Das Protein RAB22A gehört zur Gruppe der kleinen GTPasen. Bislang ist über dieses Protein sehr wenig bekannt. Es wurde bislang in der Skelettmuskulatur nachgewiesen [182]. Darüber hinaus wurde es über Microarrayanalysen (Affymetrix) in vielen Geweben als exprimiert eingestuft, wobei sich hier eine besonders hohe Expression für den *Prefontalen Cortex*, NK Zellen und T Zellen zeigt (BioGPS). Auf den Transkriptionsarrays zum Molchherzen zeigt RAB 22A keine Deregulation. Dieser Umstand bedeutet jedoch nicht, dass eine verstärkte Bildung des Proteins ausgeschlossen ist. So könnten Mechanismen der posttranskriptionellen Regulation, wie endogene siRNAs oder miRNAs, an dieser Stelle eine Rolle spielen. Diese könnten bei gleichbleibender RAB22A mRNA Menge durch Abregulation der entsprechenden, inhibierend wirkenden, siRNA oder miRNA zur Produktion des RAB22A Proteins führen, die zuvor trotz vorhandener mRNA verhindert wurde.

VIII. Ausblick

Die Arbeiten am Modellorganismus grünlicher Wassermolch werden mittlerweile am Max-Planck-Institut Bad Nauheim seit fünf Jahren durchgeführt, in deren Verlauf die ersten schwierigen Schritte bei der Arbeit mit einem „neuen“ Modellorganismus erfolgten. Neben der Generierung von primären Sequenzinformationen stand die allgemeine molekularbiologische Methodenentwicklung im Mittelpunkt (Zellkulturen, Immunohistologische Färbungen, Westernblots, Werkzeuge zur genetischen Manipulation (Viren, Vektorsysteme)). Die in dieser Arbeit dargestellte Methodenentwicklung zur Proteomanalyse stellt den Anfang eines neuen Untersuchungsfeldes der Regeneration allgemein und im speziellen im Herzen dar. Aus den Erkenntnissen, die aus den ersten Messungen im regenerierenden Schwanzgewebe und Herzgewebe gezogen wurden, sollen i) erste Kandidaten isoliert werden, die für eine weitergehende, klassisch biologische tiefgehende Analyse herangezogen werden und ii) unter Verwendung der SILAC Markierung eine Methode zur Hochdurchsatzquantifizierung auf Proteinebene etabliert werden.

Dabei zeigt sich durch die in dieser Arbeit dargestellten Ergebnisse, dass bereits aus den Daten der Methodenentwicklung hochwertige Kandidaten mit enger Verbindung zum Regenerationsablauf identifiziert werden können. Insbesondere die hohe Anzahl an potentiell amphibien-spezifischen Proteinen stellt hier einen interessanten Pool dar. Daneben lässt sich wie beispielhaft für das Contig 3735 gezeigt, zumindest rechnerisch, bereits eine Quantifizierung auf Proteinebene durchführen. Dabei dient das Einmischen eines schweren [¹³C]-Lysines als Standard, über den eine relative Quantifizierung ermöglicht wird. Dieser erste Ansatz zur Quantifizierung soll weiter verfolgt werden und auf Basis eines gesamten Datensatzes mit klassischen Methoden (Westernblots) verifiziert werden. Mit dieser Methode soll dann ein quantifizierender Datensatz zum Molchherz in Regeneration erstellt werden, der sich zeitlich mit den Beobachtungszeiträumen im Transkriptom deckt. Aus dieser kombinierten Beobachtung über den Regenerationsverlauf auf Transkriptom- und Proteomebene verspricht man sich am Max-Planck-Institut Bad Nauheim der Aufklärung der grundsätzlichen Fähigkeit des Molches zur Regeneration näher zu kommen. Vergleichend werden insbesondere hierzu bereits Transkriptuntersuchungen auf Microarraybasis in Kooperation mit anderen Arbeitsgruppen durchgeführt, die ihren Fokus auf Beinregeneration, Linsenregeneration oder Gelenkregeneration gesetzt haben. Hier können durch geeignete Abgleiche identische

Expressionsmuster, aber auch Unterschiede während der allgemeinen Regeneration herausgearbeitet werden.

Grundsätzlich hat sich die SILAC Markierung *in vivo* als vielfach nützliches Werkzeug herausgestellt. Am Max-Planck-Institut Bad Nauheim ist der Versuch, die Effizienz der SILAC Markierung zu verbessern und die Methoden zur Quantifizierung zu verfeinern ein aktuelles Arbeitsgebiet. Dabei steht neben der Entwicklung einer speziellen Diät aus 100% SILAC markiertem Mausegewebe zur Fütterung von Zebrafischen und Molchen auch der Ansatz zur Markierung spezieller Zellpopulationen (A1 Zellen aus regenerierendem Beinblastem) *in vitro*, die dann als Standard zur Quantifizierung eingesetzt werden sollen.

Die vergleichsweise umfangreichen Sequenzinformationen, die durch die Erstellung der hier beschriebenen cDNA Bank gesammelt wurden, sollen im Rahmen einer weiteren Arbeit zu diesem Thema in eine weiterentwickelte, relationale Datenbank abgelegt werden, die dann dem Fachpublikum zur Verfügung gestellt werden soll. In diesem Zusammenhang steht vor allem die funktionelle Annotation von primär annotierten Transkripten im Mittelpunkt, beispielsweise durch Zuhilfenahme von *Gene Ontology* Termen und *Interpro* [183] Zuordnungen.

Die Verwendung moderner *Next Generation Sequencing* Techniken steht seit längerer Zeit im Fokus der Molch Arbeitsgruppe am Max-Planck-Institut Bad Nauheim. Durch Verwendung dieser Techniken verspricht man sich, der Problematik der fehlenden genomischen Sequenzinformation zu begegnen. Jedoch hat sich gezeigt, dass die neuen Techniken bislang noch nicht für diesen Zweck herangezogen werden können. Dies liegt primär an der im Verhältnis zur klassischen Sanger Sequenzierung kurzen Leselänge dieser neuen Techniken. Ohne eine zu Grunde liegende Referenzsequenz stellt sich die Assemblierung noch als unüberwindbares bioinformatisches Problem dar. Mit dem Trend zu höheren Leselängen (Roche 454 Technologie von 100 bp im FLX zu 400-500 bp im Titanium Paket) könnte sich diese Problematik aber auflösen. Dennoch wurden die aktuellen Sequenzierungstechniken bereits für den Modellorganismus Molch genutzt. Nach einer Größenselektion von RNA der Länge 20 bis 70 bp wurde mit Hilfe eines Roche 454 FLX Systems eine ~300.000 Sequenzen umfassende *small RNA* Bank zu mehreren Gewebeklassen des Molches erstellt. Die Klassifizierung und Annotierung dieser Bank ist Gegenstand eines aktuellen Projektes. Dabei steht neben der Identifizierung von bekannten und neuen miRNAs und endogenen siRNAs die Klassifizierung von weiteren kurzen RNA Klassen im Mittelpunkt.

Ausblick

Zusammenfassend soll festgehalten werden, dass der breit gefächerte Ansatz zur Aufklärung der Fähigkeit des Molches zur Regeneration am örtlichen Institut in einem systembiologischen Ansatz verfolgt wird. Die Verarbeitung und Handhabung der dabei entstehenden Datensätze stellt eine primär bioinformatische Herausforderung dar. Ziel der bioinformatischen Analysen ist dabei aber stets die Selektion besonders vielversprechender Kandidaten, die dann mithilfe klassischer molekularbiologischer Verfahren weiter analysiert werden müssen.

IX. Zusammenfassung

Der Molch *N. viridescens* ist ein Organismus mit außergewöhnlichen Fähigkeiten im Bereich der Regeneration. Gemessen an seiner enormen Bedeutung für die Regenerationsforschung ist jedoch sehr wenig über ihn bekannt. Es bestehen kaum öffentlich verfügbare Informationen zu Genom, Transkriptom und Proteom, es liegen kaum angepasste molekularbiologische Methoden vor, und ebenso wenig gibt es verfügbare Erkenntnisse zu Haltung, Aufzucht und Lebensraum. Dennoch stellt der Molch aufgrund seiner außergewöhnlichen Fähigkeiten, insbesondere im Bereich der Herzregeneration, einen unverzichtbaren Modellorganismus dar, der die Aufklärung einzigartiger Fragestellungen zur Regeneration ermöglicht.

In dieser Arbeit konnte gezeigt werden, dass trotz dieser benannten Einschränkungen der Einsatz von modernen Hochdurchsatzmethoden möglich ist. So wurden neben der Anpassung von Methoden zur Transkriptionsanalyse auf cDNA Microarray Basis eine neue Methode zur Proteomanalyse unter Verwendung von SILAC markierten Peptididentifizierungen entwickelt. Mit diesen Methoden konnten wesentliche Anteile der transkriptionellen Antwort auf gezielte Schädigungen entschlüsselt werden. Eine wesentliche Grundlage für diese beiden Hochdurchsatztechniken stellt die ebenfalls beschriebene EST Bibliothek dar. Der normalisierte Datensatz, bestehend aus ~ 100.000 Einzelklonen, dazu ~50.000 Einzelsequenzen (ESTs), die nach Assemblierung zu ~26.000 sequenziell unterschiedlichen Transkripten (Contigs) zusammengefasst werden konnten, beschreibt die Herzregeneration bereits sehr tiefgehend. Dieser Sequenzdatensatz wurde unter Verwendung der Algorithmen blastn, blastx und tblastx jeweils gegen die Nukleotid Sequenzsammlung, die Proteinsequenzsammlung und EST Sequenzsammlung von NCBI abgeglichen und sämtliche signifikante Zuordnungen in einer internen relationalen Datenbank abgelegt. Zur Annotation wurden neben statistischen Beschreibungen auch die Alignments selbst, als auch die Ursprungsorganismen der Alignments geeignet abgelegt.

Auf der Basis dieses annotierten Sequenzdatensatzes wurde ein cDNA Microarray erstellt, mit dessen Hilfe die Herzregeneration in einer umfangreichen Beobachtungszeitreihe untersucht werden konnte. Es wurde exemplarisch gezeigt, dass insbesondere die frühe Wundantwort sehr detailliert die Abläufe der einsetzenden Regeneration beschreibt. Insbesondere das Vorkommen vieler bekannter *Immediate early genes* wie beispielsweise *fos* und *jun*, verschiedener Transkriptionsfaktoren, Metalloproteasen und Mitglieder der *CCN* Familie zeigen, dass auch im Molchherzen grundsätzliche, vielfach bekannte Mechanismen einsetzen, die so auch eine

Zusammenfassung

Wundheilung bei höheren Organismen einleiten. Einige Besonderheiten, die die Wundantwort des Molches von der höherer Organismen unterscheiden, wie beispielsweise die Abregulation des primären und sekundären Aktivierungsweges des Komplementsystems oder die sehr kurzfristige Expression von BTG2, sind erste Hinweise auf potentielle Kandidaten, die eine veränderte Wundantwort hin zu einer „regenerativen Antwort“ mitgestalten könnten. Diese über den differenzierten Ansatz detektierten Kandidaten empfehlen sich zur tiefgehenden Analyse, an deren Ziel das bessere Verständnis der totalen Regeneration steht.

Neben der Transkriptomanalyse auf Basis von cDNA Microarrays wurde mit der massenspektrometrischen Untersuchung von regenerierendem Molchgewebe eine bisher für diesen Organismus noch nicht genutzte Methode eingesetzt. In diesem Bereich wurden schwerpunktmäßig die Felder i) Proteinmarkierung mit stabilen Isotopen, ii) Peptididentifizierung in organismusfremden Datenbanken, iii) Erweiterung des identifizierten Peptidspektrums durch die Verwendung mehrerer Datenbanken, iv) Nutzbarmachung einer EST Sequenzbank, v) unterschiedliche Proteinidentifizierungen in unterschiedlichen Regenerationszeitpunkten und vi) die Identifizierung von bislang unbekanntem kodierenden Sequenzen bearbeitet. Im Einzelnen konnte zunächst in einem Experiment zur Schwanzregeneration gezeigt werden, dass sich Molche durch Fütterung mit schweren essentiellen Aminosäuren in Form von Lebergewebe innerhalb von 60 Tagen markieren lassen. Die Markierungsrate steigt dabei in regenerierendem Gewebe stark an. Des Weiteren lassen sich mit dem Proteomics Ansatz und Massenspektrometrie ~90% aller bislang bekannten Molchproteine in einer Messung nachweisen. Weiterhin konnten durch den Einsatz mehrerer organismusfremder Datenbanken zur Peptididentifizierung im Schwanz über 1000 unterschiedliche Proteine identifiziert werden. Der Einsatz einer vorläufigen EST Bank erbrachte ~450 unterschiedliche Proteine.

Eine umfangreichere Versuchsreihe zur frühen Herzregeneration auf der neu erstellten cDNA Bank und unter Ausnutzung der neu erarbeiteten Methode ermöglichte über 3000 Proteinidentifizierungen in der normalisierten EST Sequenzdatenbank. Der Einsatz der Lysin markierten Proteine diente dabei der verbesserten Peptididentifizierung. Der Einsatz einer kombinierten Proteindatenbanksuche erbrachte über 4000 Proteinidentifikationen. Weiterhin konnte durch die SILAC Markierung die Grundlage für eine spätere Quantifizierung der Proteine gelegt werden.

Innerhalb der über 3000 detektierten Proteine der EST Bank konnte eine Gruppe von 109 hochwertig detektierten Proteingruppen isoliert werden, die ausschließlich eine EST Homologie

zu den nahen Verwandten wie *Xenopus* und *Ambystoma*, oder gar keine Sequenzähnlichkeit zeigen. Somit ist auf diesem Wege das Vorkommen der Sequenzen im Transkriptom (cDNA), eine relative Quantifizierung auf Transkriptomebene (cDNA Microarray) und der Nachweis dafür erbracht, dass diese Sequenzen für Proteine kodieren (Peptididentifizierung). Im Abgleich mit den relativen Expressionsdaten auf Transkriptebene aus den Microarrayexperimenten konnten weiterhin 14 neu entdeckte kodierende Transkripte als dereguliert eingruppiert werden. Zusammengefasst kann bei dieser Protein Gruppe von amphibien-spezifischen Proteinen ausgegangen werden. Innerhalb dieser Gruppe an potentiell unbekannt Proteinen konnten 17 signifikant konservierte Strukturen über eine entsprechende Mustersuche nachgewiesen werden. Exemplarisch wurde aus dieser Gruppe das Transkript, welches die DAN Domäne beinhaltet, einer weitergehenden Sequenzanalyse unterzogen. Die detektierten Kandidaten, insbesondere die deregulierten, empfehlen sich für eine nähere, traditionelle molekularbiologische Analyse. Insbesondere die genaue Lokalisation im regenerierenden Herzen, die Art der exprimierenden Zellen und die funktionelle Abhängigkeit voneinander, sollten bei weiteren Analysen im Mittelpunkt stehen.

Zusammenfassung

X. Anhang

A. Abkürzungsverzeichnis

ABC	Ammonium Bikarbonat
ACN	Acetonitril
AEC	apical epidermal cap
AER	apical ectodermal ridge
aRNA	antisense RNA
AS	Aminosäuren
bp	Basenpaar/-re
BrdU	Bromodesoxyuridin
BSA	Rinderserumalbumin
°C	Grad Celsius
cDNA	zur mRNA komplementäre DNA die durch Reverse Transkription entsteht
CID	Collision-induced dissociation
Cy5	Indocarbocyanin 5
Cy3	Indocarbocyanin 3
bzw.	beziehungsweise
d	Tag/-e
d.h.	das heißt
ddH ₂ O	zweifach destilliertes H ₂ O
DNA	Desoxyribonukleinsäure
DNase	Desoxyribonuklease
dNTP	Desoxynukleosidtriphosphat
DR	<i>danio rerio</i>
DSN	duplex-specific nuclease
DTT	1,4-Dithiothreit
EDTA	Ethylendiamintetraacetat
ESI	Elektrospray Ionisation
EST	Expressed Sequence Tags
et al.	et altera
EtOH	Ethanol
EZM	Extrazelluläre Matrix
FC	Fold Change
FGF	Fibroblast-Growth-factor
GO	<i>Gene Ontology</i>
GeLC-MS	SDS-Gel Auftrennung-Flüssigchromatographie-Massenspektrometrie
h	Stunde
HCD	Higher Energy Collision Dissociation
HPLC	High Pressure Liquid Chromatography
<i>in silico</i>	Prozessierung am Computer
<i>in vitro</i>	in der Zellkultur; außerhalb des Organismus
<i>in vivo</i>	im lebenden Organismus

Anhang

IPI	International Protein Index
KO	<i>knockout</i>
LB	Luria Bertani Kulturmedium
LP/DTT	Laufpuffer mit DTT
LC	Flüssigchromatographie
M	molar
mM	milli molar
μ	Mikro ($1 \cdot 10^{-6}$)
MeOH	Methanol
MPI	Max-Planck-Institut
MPI-BN	Max-Planck-Institut Bad Nauheim
MS	Massenspektrum
MS ²	Fragmentspektrum
NCBI	<i>National Center for Biotechnology Information</i>
N.v.	<i>Notophthalmus viridescens</i>
ORF	open reading frame
PCR	Polymerase-Ketten-Reaktion
RNase	Ribonuklease
RNA	Ribonukleinsäure
rpm	rounds per minute (Umdrehungen pro Minute)
RT	Raumtemperatur
SDS-PAGE	Natrium-dodecylsulfat-Polyacrylamide-Gelelektrophorese
SILAC	stable isotope labeling with amino acids in cell culture
SNP	Single Nucleotide Permutations
SSC	standard sodium citrate
Taq	<i>Thermus aquaticus</i>
TFA	Trifluoressigsäure

B. Zusätzliche Tabellen und Abbildungen

Tabelle 15: Im Massenspektrometer detektierte Proteine des Molches *N. viridescens* aus der NCBI Datenbank

Die Tabelle zeigt die 72 in der massenspektrometrischen Messung des regenerierenden Schwanzgewebes identifizierten Proteine aus der öffentlich verfügbaren NCBI *N. viridescens* Proteinsequenzdatenbank. Aufgeführt sind unter Protein Beschreibung die einzelnen NCBI NR Identifikationsnummern und deren Beschreibung. Unter „Unique Peptides ungeschädigt“ und „40 Tage“ sind jeweils die Anzahl der in diesem Regenerationszustand identifizierten einzigartigen Peptide vermerkt. Außerdem unter „Unique und Razor Sequence Coverage“ die prozentuale Abdeckung des detektierten Proteines durch die einzelnen Unique und Razor Peptide. Abschließend ist noch das Molekulargewicht des Proteines vermerkt.

Protein Beschreibung	Unique Peptides (seq) ungeschädigt	Unique Peptides (seq) 40 Tage	Unique + Razor Sequence Coverage [%]	Mol. Weight [kDa]
>gi 82659076 gb ABB88702.1 tissue inhibitor of metalloproteinase 1	15	11	100	23.959
>gi 10799285 gb AAG21827.1 keratin 17	14	12	100	24.366
>gi 2231239 gb AAB61990.1 NvTbox4	9	5	100	10.374
>gi 2231237 gb AAB61989.1 NvTbox3	7	6	100	10.302
>gi 545155 gb AAB29802.1 homeobox gene NvHBox-2.7 product (type 1 homeodomain) [Notophthalmus viridescens=newts, tail, Peptide Partial, 60 aa]	6	1	100	7.3925
>gi 90184290 gb ABD91086.1 rhodopsin	5	4	100	12.169
>gi 704358 gb AAA86685.1 homeoprotein	4	4	100	4.346
>gi 1004367 gb AAA77685.1 NvPax-6	4	2	100	5.8507

Anhang

>gi 64116 emb CAA34745.1 NuHox-1 protein C-terminal fragment (104 AA)	0	1	99	12.302
>gi 151349589 gb ABS01492.1 anterior gradient protein 2	20	12	97.6	18.727
>gi 13561421 gb AAK30303.1 AF296616_1 NADH dehydrogenase subunit 2	12	8	96.5	38.485
>gi 400006 sp P31262.1 HXC5_NOTVI Homeobox protein Hox-C5 (NvHbox-3.4);>gi 555494 gb AAA49397.1 homeobox protein	4	2	95.8	8.9792
>gi 1143523 emb CAA62492.1 newt skeletal myosin heavy chain	4	3	92.7	6.3679
>gi 90185040 gb ABD91461.1 histone H3a	8	10	86.2	12.294
>gi 704352 gb AAA86682.1 homeoprotein	3	1	86.1	4.462
>gi 1095176 prf 2107332A Msx-1 gene;>gi 565628 emb CAA57791.1 Msx-1	15	6	85	30.965
>gi 6706301 emb CAB65965.1 connective tissue growth factor	21	13	83.3	38.098
>gi 28372363 gb AAO38044.1 complement component C5	20	16	82.9	49.083
>gi 27734466 sp Q9YHB3.1 RFNG_NOTVI Beta-1,3-N-acetylglucosaminyltransferase radical fringe (nrFng) (O-fucosylpeptide 3-beta-N-acetylglucosaminyltransferase);>gi 4204693 gb AAD10827.1 radical fringe protein	11	9	80.6	45.047
>gi 28372361 gb AAO38043.1 complement component C3	38	27	79.5	45.486

>gi 54610074 gb AAV35135.1 recombination activating protein 1	17	29	79.1	58.361
>gi 476727 gb AAA49394.1 fibroblast growth factor receptor 1	20	17	77.6	43.834
>gi 545156 gb AAB29803.1 homeobox gene NvHBox-3 product (type 1 homeodomain) [Notophthalmus viridescens=newts, tail, Peptide Partial, 60 aa]	2	2	76.7	7.728
>gi 213396 gb AAA49396.1 homeobox protein;>gi 399984 sp P31261.1 HXA2_NOTVI Homeobox protein Hox-A2 (NvHbox-2.8)	2	2	75.6	10.764
>gi 109255184 gb ABG27017.1 transcription factor Oct4	2	3	75.5	11.25
>gi 476731 gb AAA49398.1 keratinocyte growth factor receptor	7	0	75.1	82.07
>gi 829621 gb AAA80227.1 MCM3	46	28	73.6	90.356
>gi 2370337 emb CAA04656.1 cytokeratin 8	33	26	73	55.152
>gi 1173170 sp P18516.2 RARG_NOTVI Retinoic acid receptor gamma (RAR-gamma) (Nuclear receptor subfamily 1 group B member 3) (Retinoic acid receptor delta) (RAR-delta);>gi 64130 emb CAA49564.1 retinoic acid receptor delta 1a	5	4	71.1	56.629
>gi 94957952 gb ABF47211.1 Tolloid-like 1 protein	4	4	70.1	22.251
>gi 12659082 gb AAK01180.1 GLI3 protein	20	11	68.4	31.964
>gi 577785 emb CAA58043.1 myogenic regulatory factor; transcription factor	4	3	68	24.148

Anhang

>gi 6466795 gb AAF13029.1 AF070733_1 transcription factor Prox1	16	10	67.6	40.518
>gi 2231235 gb AAB61988.1 NvTbox2	4	4	66.3	10.249
>gi 1666661 emb CAA70428.1 retinoblastoma protein	37	44	65.3	102.99
>gi 64124 emb CAA78621.1 retinoic acid receptor alpha 2 >gi 133488 sp P18514.1 RARA_NOTVI Retinoic acid receptor alpha (RAR-alpha) (Nuclear receptor subfamily 1 group B member 1);>gi 226520 prf 1516344A retinoic acid receptor alpha;>gi 64126 emb CAA35602.1 retinoic acid receptor alpha	12	13	65.3	50.612
>gi 60207666 gb AAX14807.1 matrix metalloproteinase 3/10b	26	12	64.6	53.05
>gi 60207592 gb AAX14804.1 matrix metalloproteinase 3/10a	15	16	62.8	54.599
>gi 60207645 gb AAX14806.1 collagenase	22	21	62.5	53.233
>gi 3182939 sp Q91145.1 COCA1_NOTVI Collagen alpha-1(XII) chain;>gi 632648 gb AAA80217.1 type XII collagen alpha-1 chain	24	31	61.8	101.65
>gi 60207620 gb AAX14805.1 matrix metalloproteinase 9	20	16	61.6	75.945
>gi 2497974 sp Q91400.1 FINC_NOTVI Fibronectin (FN);>gi 914306 gb AAB34250.1 fibronectin; matrix 4 antigen; MT4 antigen	2	1	61.6	21.96
>gi 545159 gb AAB29806.1 homeobox gene NvHBox-6 product type 1 homeodomain [Notophthalmus viridescens=newts, tail, Peptide Partial, 56 aa]	0	3	58.9	6.7256

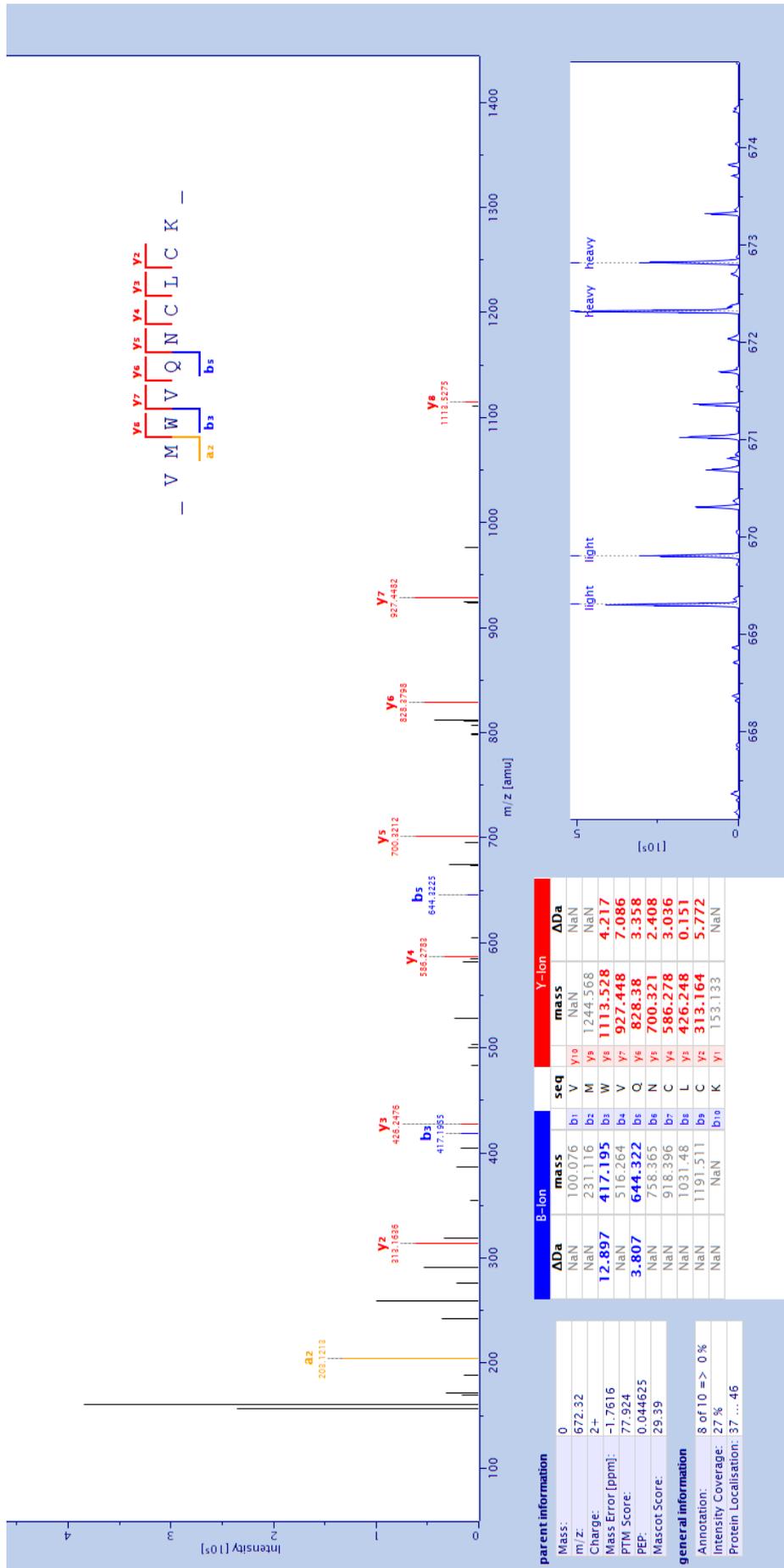
>gi 84181316 gb ABC55062.1 growth arrest-specific 6	14	15	58.7	74.726
>gi 1705494 sp P53771.1 BOX5_NOTVI Homeobox protein box-5 (NvHbox-5);>gi 433820 emb CAA45095.1 NvHBox-5 protein	9	4	57.4	28.346
>gi 2896035 gb AAC03108.1 hedgehog segment polarity homolog	11	8	55.7	45.072
>gi 23507736 gb AAB62743.2 regeneration blastema forelimb-specific Tbx	25	21	53.9	57.862
>gi 467691 emb CAA55125.1 B24 protein	29	21	53	83.382
>gi 94957950 gb ABF47210.1 snail protein	1	2	52.7	8.5579
>gi 226278 prf 1504304A retinoic acid receptor	4	3	50.8	45.105
>gi 94957954 gb ABF47212.1 Sox2 transcription factor	5	4	47.1	13.234
>gi 3024695 sp Q91060.1 TBA_NOTVI Tubulin alpha chain;>gi 468544 emb CAA83457.1 alpha-tubulin	15	15	46.9	50.093
>gi 2370339 emb CAA04655.1 cytokeratin type II	18	11	45.6	50.638
>gi 13561422 gb AAK30304.1 AF296616_2 NADH dehydrogenase subunit 1	2	0	44.4	14.229
>gi 64133 emb CAA49566.1 retinoic acid receptor delta 2	2	1	42.4	7.2122
>gi 157280811 gb ABV29331.1 Prod 1	3	1	38.7	11.854
>gi 400022 sp P31263.1 HXD11_NOTVI Homeobox protein Hox-D11 (NvHox-2)	10	8	37.9	31.782

Anhang

>gi 15826776 pdb 1FMM S Chain S, Solution Structure Of Nfgf-1;>gi 55583918 sp Q7SIF8.1 FGF1_NOTVI Heparin-binding growth factor 1 (HBGF-1) (Fibroblast growth factor 1) (FGF-1) (Acidic fibroblast growth factor) (aFGF)	2	0	37.9	15.015
>gi 23452506 gb AAN33052.1 T-domain transcription factor Tbx4	8	2	37.1	47.384
>gi 577787 emb CAA58044.1 transcription factor	3	1	29.2	12.732
>gi 4090567 gb AAC98812.1 putative fibroblast growth factor-4	7	10	29.1	22.033
>gi 32307647 gb AAP79220.1 retinoic acid receptor beta-2	0	1	29.1	13.344
>gi 57164672 gb AAW34271.1 retinoic acid receptor beta-2	8	9	27.5	50.484
>gi 58202527 gb AAW67315.1 NADH dehydrogenase subunit 4	5	1	26.1	25.538
>gi 1705493 sp P53770.1 DLX3_NOTVI Homeobox protein DLX-3 (Box-4) (NvHbox-4);>gi 432378 emb CAA45094.1 NvHBox-4 protein	2	4	24.5	30.654
>gi 109255186 gb ABG27018.1 transcription factor BMP2/4	2	0	20.3	34.666
>gi 3914106 sp Q91154.1 MYF5_NOTVI Myogenic factor 5 (Myf-5);>gi 577789 emb CAA58045.1 myogenic regulatory factor; transcription factor	6	4	18.4	28.834
>gi 123242 sp P14858.1 HXC6_NOTVI Homeobox protein Hox-C6 (NvHox-1) (FH-2);>gi 64118 emb CAA32139.1 unnamed protein product	1	0	15.4	26.901
>gi 133495 sp P18515.1 RARB_NOTVI Retinoic acid receptor	1	1	13.9	18.204

beta (RAR-beta) (Nuclear receptor subfamily 1 group B member 2)				
>gi 545158 gb AAB29805.1 homeobox gene NvHBox-5 product type 1 homeodomain [Notophthalmus viridescens=newts, tail, Peptide Partial, 60 aa]	1	0	13.3	7.1613
>gi 1843450 emb CAA49562.1 largest of the predicted retinoic acid receptor delta 1 isoforms employing a CUG initiation site ;>gi 226521 prf 1516344B retinoic acid receptor delta;>gi 64131 emb CAA49565.1 retinoic acid receptor delta 1b ;>gi 64135 emb CAA35603.1 retinoic acid receptor delta	1	2	3.5	70.185
>gi 476729 gb AAA49395.1 fibroblast growth factor receptor 2;>gi 82109401 sp Q91147.1 FGFR2_NOTV1 Fibroblast growth factor receptor 2 precursor (FGFR-2)	1	0	3	81.824

A



Anhang

Abbildung 38: Spektren zum Contig 3735

Dargestellt sind die Full MS sowie Fragmentspektren zum Peptid VMWVQNCLCK nach Detektion im ungeschädigten Molchherzen (A) und im geschädigten Molchherzen nach sechs Stunden (B). Die detektierten Fragmentationen zum Parention (y, b, a) sind farblich gekennzeichnet und jeweils mit ihren statistischen Werten dargestellt. Darüber hinaus zeigt das FullMS Spektrum deutlich die Aufregulierung im regenerierenden Zustand. So ist der leichte Peak im FullMS des sechs Stunden Regenerierenden Herzen (B) deutlich höher als der korrespondierende schwere Peak. Im ungeschädigten Herzgewebe (A) liegt die umgekehrte Situation vor.

Anhang

Abbildung 39: Überexpression von Proteinen nach Gene Ontology Analyse

Dargestellt sind die GO-Graphen nach der humanen GO-Annotation. Es liegen mit humanen Proteinen annotierte Molchproteine zu Grunde. Die Wurzeln der beiden Bäume bilden jeweils die Wurzel der Ontologie ab. Terme sind als farblich markierte Kreise dargestellt. Weiß markierte Terme deuten nicht überexprimierte Terme an, die zur Vollständigen Aufstellung des nötig waren. Der Farbverlauf hellgelb bis orange visualisiert den steigenden Grad der Überexpression. Ausgehend vom Wurzelterm zeigen die Verbindungen zu untergeordneten Termen eine Spezialisierung an, im Umkehrschluss von einem speziellem Term zurück eine Verallgemeinerung. Der Durchmesser der einzelnen Terme beschreibt die absolute Gesamtanzahl der zugrundeliegenden Proteine. **A:** Die Ontologie Biological Process zeigt eine Überexpression im kompletten Teilgraphen der metabolischen und katabolischen Prozesse (linke Seite). Außerdem fallen die Terme Translation, Proteinfaltung und Muskelkontraktion auf. **B:** Die Ontologie Molecular Function zeigt eine starke Überexpression im Bereich Protein-, Coenzym- und RNA-Bindung sowie bei Proteinen, die mit der NADH-Dehydrogenase Aktivität in Verbindung gebracht werden.

C. Abbildungsverzeichnis

Abbildung 1: Schematische Einteilung der Regeneration	3
Abbildung 2 nebenstehend: Regeneratives Potential in mehrzelligen Organismen	8
Abbildung 3: Blastemgenerierung nach Beinamputation (schematisch)	9
Abbildung 4: Der grünliche Wassermolch <i>Notophthalmus viridescens</i>	11
Abbildung 5: Schwanzregeneration des Molches	17
Abbildung 6: Ungeschädigtes Molchherz	19
Abbildung 7: Schematische Darstellung der Hauptformen von Wirbeltierherzen	22
Abbildung 8: Ablauf einer markierten Massenspektrometrischen Analyse	29
Abbildung 9: pDNR-LIB Vector von BD Biosciences, San Jose, California	35
Abbildung 10: SMART und DSN Technologie	43
Abbildung 11: SOM Cluster zum scheinoperierten Herzen	62
Abbildung 12: SOM Cluster, 2 Stunden Regeneration des Herzen	66
Abbildung 13: Hierarchischer Cluster der deregulierten Spots nach 6 Stunden Regeneration	71
Abbildung 14: Vergleich der deregulierten Contigs zu den Zeitpunkten ScheinOP, 2 Stunden und 6 Stunden Regeneration als Venn-Diagramm	73
Abbildung 15: Zuordnung Organismus zu bester gefundener Annotation ($e\text{-value} < e^{-10}$)	78
Abbildung 16: Phylogenetische Stammbäume	79
Abbildung 17: Schweres [^{13}C] Lysin	83
Abbildung 18: Fütterungs- und Auswertungsschema des SILAC Molches: Legende auf der nächsten Seite .	84
Abbildung 19: Veränderung der mittleren Markierungsrate im Schwanz	86
Abbildung 20 über- und nebenstehend: Erweitertes detektierbares Molchproteom	90
Abbildung 21: SILAC Peptidspektren zu organismusfremden Datenbanken	91
Abbildung 22: RT-PCR zu deregulierten Proteinen aus GO-Analyse	93
Abbildung 23: Abgleich identifizierter Peptide in drei IPI Proteinsequenzbanken	95
Abbildung 24: Schematischer Abgleich zwischen Nukleotid und Peptid Assemblierung	98
Abbildung 25: Identifizierung unbekannter Proteine aus EST Sequenzen	101
Abbildung 26: Versuchsschema partiell markierte SILAC Pools	103
Abbildung 27: Häufigkeitsfunktion „verdünnte SILAC Proteingemische“	105
Abbildung 28: Venn-Diagramm der detektierten Proteine aus „verdünnte SILAC Proteingemische“	105
Abbildung 29: Korrelation der SILAC Proteingemische 30% und 100%	107
Abbildung 30: Venn-Diagramm der deregulierten Proteine aus verdünnten SILAC Proteingemischen	109
Abbildung 31: Verteilung des SILAC-markierten Proteinpools	110
Abbildung 32: Versuchsschema Quantifizierung des Proteoms des regenerierenden Molchherzen	111
Abbildung 33: Detektierte Proteine im regenerierenden Herzgewebe, dargestellt als Venn-Diagramm	113

Anhang

<i>Abbildung 34: Ablaufdiagramm zur Suche von unbekanntem, kodierenden ESTs</i>	<i>117</i>
<i>Abbildung 35: Alignment der DAN Domäne</i>	<i>120</i>
<i>Abbildung 36: Ergebnisse einer BLAST Homologie Suche in der NCBI NR Datenbank</i>	<i>121</i>
<i>Abbildung 37: Alignment der TSP 1 Domäne.....</i>	<i>124</i>
<i>Abbildung 38: Spektren zum Contig 3735.....</i>	<i>xxiv</i>
<i>Abbildung 39: Überexpression von Proteinen nach Gene Ontology Analyse.....</i>	<i>xxvi</i>

D. Tabellenverzeichnis

<i>Tabelle 1: Taxonomie des grünlichen Wassermolches</i>	12
<i>Tabelle 2: Sequenzen der Sense und Antisense Primer</i>	33
<i>Tabelle 3: Lösungen</i>	34
<i>Tabelle 4: Verwendete Enzyme</i>	36
<i>Tabelle 5: Häufig genutzte Chemikalien</i>	37
<i>Tabelle 6: Software</i>	39
<i>Tabelle 7: Deregulierte Microarray Spots im Verlauf der Herzregeneration</i>	58
<i>Tabelle 8: Peptide und Proteine in den organismusfremden Datenbanken</i>	88
<i>Tabelle 9: Mittelwerte und normalisierte Mittelwerte der SILAC-Markierung in verdünnten Proteingemischen</i>	104
<i>Tabelle 10: Korrelation der verdünnten SILAC Proteingemische</i>	106
<i>Tabelle 11: Deregulierte Proteine in verdünnten SILAC Proteingemischen</i>	108
<i>Tabelle 12: Konservierte Proteinstrukturen</i>	119
<i>Tabelle 13: Expressionsdaten des Contigs 3735 (CCN Homolog)</i>	125
<i>Tabelle 14: Berechnungsschritte zur Quantifizierung des Contigs 3735</i>	127
<i>Tabelle 15: Im Massenspektrometer detektierte Proteine des Molches N.viridescens aus der NCBI Datenbank</i>	xv

**Der Lebenslauf wurde aus der elektronischen
Version der Arbeit entfernt.**

**The curriculum vitae was removed from the
electronic version of the paper.**

F. Publikationen

Posterpräsentation auf dem "Regeneration Meeting 2006" in Ascona (September 2006)

Gene expression profiling of regenerating hearts and limbs in the newt *Notophthalmus viridescens*

Thilo Borchardt¹, Mario Looso¹, Patrick Weiss¹, Julia Kruse¹, Henning Witt², Patricia Ruiz², Thomas Braun¹

¹Max-Planck-Institute for Heart and Lung Research, Department of Cardiac Development and Remodelling, Parkstrasse 1, D-61231 Bad Nauheim

²Center for Cardiovascular Research, Hessische Straße 3-4, 10115 D-Berlin

Posterpräsentation auf der "Proteomics Forum" in Berlin (April 2009)

Indirect non-saturated SILAC-based mass spectrometry to identify proteins driving organ regeneration of newts

Mario Looso¹, Marcus Krüger¹, Thilo Borchardt¹, Thomas Braun¹

¹ Max-Planck-Institute for Heart and Lung Research, Department for Cardiac Development and Remodelling, Bad Nauheim, Germany

Posterpräsentation auf der „German Conference on Bioinformatics“ in Berlin (September 2009)

A Gene Ontology based analysis of high-throughput data via a multivariate approach

Marc Bruckskotten¹, Mario Looso¹, Anne Konzer¹, Franz Cemic² and Thomas Braun¹

¹ Max-Planck-Institute for Heart and Lung Research, Department for Cardiac Development and Remodelling, Bad Nauheim, Germany

² Fachhochschule Gießen Friedberg, Institut für Biochemische Verfahren und Analysen (IBVA)

Angenommen bei BMC Bioinformatics

PCA2GO: Find highly expressed GO-Terms with a new multivariate statistics based method

Marc Bruckskotten¹, Mario Looso¹, Franz Cemic², Anne Konzer¹, Jürgen Hemberger², Markus Krüger¹ and Thomas Braun¹

¹ Max-Planck-Institute for Heart and Lung Research, Department for Cardiac Development and Remodelling, Bad Nauheim, Germany

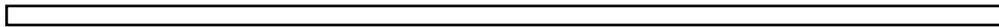
² Fachhochschule Gießen Friedberg, Institut für Biochemische Verfahren und Analysen (IBVA)

Publiziert bei BMC Genomics

Analysis of newly established EST databases reveals similarities between heart regeneration in newt and fish

Thilo Borchardt, Mario Looso, Patrick Weis, Julia Kruse and Thomas Braun

Max-Planck-Institut für Herz- und Lungenforschung, Parkstrasse 1, D-61231 Bad Nauheim, Germany



Publiziert bei Molecular and Cellular Proteomics (MCP)

Advanced identification of proteins in uncharacterized proteomes by pulsed in vivo SILAC

Mario Looso, Thilo Borchardt, Marcus Krüger and Thomas Braun

Max-Planck-Institut für Herz- und Lungenforschung, Parkstrasse 1, D-61231 Bad Nauheim, Germany

XI. Literaturverzeichnis

1. Reaumur R-AFd (1712) Sur les diverses reproductions qui se font dans sel Ecrevisse, les Omars, les Crabes et entr'autres sur celles de leurs Jambes et de leurs Ecailles.
2. Trembley A (1744) Memoires pour servir a l'histoire d'un genre de polypes d'eau douce.
3. Spallanzani L (1768) Prodomo di un opera da imprimersi sopra la riproduzioni animali.
4. Widmann JJ, Fahimi HD (1975) Proliferation of mononuclear phagocytes (Kupffer cells) and endothelial cells in regenerating rat liver. A light and electron microscopic cytochemical study. *Am J Pathol* 80: 349-366.
5. Poss KD, Keating MT, Nechiporuk A (2003) Tales of regeneration in zebrafish. *Dev Dyn* 226: 202-210.
6. Poss KD, Wilson LG, Keating MT (2002) Heart regeneration in zebrafish. *Science* 298: 2188-2190.
7. Morgan TH (1901) Regeneration And Liability To Injury. *Science* 14: 235-248.
8. Bodenmuller H, Schaller HC (1981) Conserved amino acid sequence of a neuropeptide, the head activator, from coelenterates to humans. *Nature* 293: 579-580.
9. Lohmann JU, Bosch TC (2000) The novel peptide HEADY specifies apical fate in a simple radially symmetric metazoan. *Genes Dev* 14: 2771-2777.
10. Tsonis PA, Del Rio-Tsonis K (2004) Lens and retina regeneration: transdifferentiation, stem cells and clinical applications. *Exp Eye Res* 78: 161-172.
11. Del Rio-Tsonis K, Tsonis PA (2003) Eye regeneration at the molecular age. *Dev Dyn* 226: 211-224.
12. Tsonis PA (2000) Regeneration of the lens in amphibians. *Results Probl Cell Differ* 31: 179-196.
13. Echeverri K, Tanaka EM (2002) Ectoderm to mesoderm lineage switching during axolotl tail regeneration. *Science* 298: 1993-1996.
14. Kragl M, Knapp D, Nacu E, Khattak S, Maden M, et al. (2009) Cells keep a memory of their tissue origin during axolotl limb regeneration. *Nature* 460: 60-65.
15. Thorgeirsson SS (1996) Hepatic stem cells in liver regeneration. *FASEB J* 10: 1249-1256.
16. Alvarez-Buylla A, Lim DA (2004) For the long run: maintaining germinal niches in the adult brain. *Neuron* 41: 683-686.

Literaturverzeichnis

17. Bjerknes M, Cheng H (2005) Gastrointestinal stem cells. II. Intestinal stem cells. *Am J Physiol Gastrointest Liver Physiol* 289: G381-387.
18. Theise ND (2006) Gastrointestinal stem cells. III. Emergent themes of liver stem cell biology: niche, quiescence, self-renewal, and plasticity. *Am J Physiol Gastrointest Liver Physiol* 290: G189-193.
19. Wilson A, Trumpp A (2006) Bone-marrow haematopoietic-stem-cell niches. *Nat Rev Immunol* 6: 93-106.
20. Leri A, Kajstura J, Anversa P, Frishman WH (2008) Myocardial regeneration and stem cell repair. *Curr Probl Cardiol* 33: 91-153.
21. Schofield R (1978) The relationship between the spleen colony-forming cell and the haemopoietic stem cell. *Blood Cells* 4: 7-25.
22. Machida S, Spangenburg EE, Booth FW (2004) Primary rat muscle progenitor cells have decreased proliferation and myotube formation during passages. *Cell Prolif* 37: 267-277.
23. Ustanina S, Carvajal J, Rigby P, Braun T (2007) The myogenic factor Myf5 supports efficient skeletal muscle regeneration by enabling transient myoblast amplification. *Stem Cells* 25: 2006-2016.
24. Morrison JI, Loof S, He P, Simon A (2006) Salamander limb regeneration involves the activation of a multipotent skeletal muscle satellite cell population. *J Cell Biol* 172: 433-440.
25. Yokoyama H (2008) Initiation of limb regeneration: the critical steps for regenerative capacity. *Dev Growth Differ* 50: 13-22.
26. Brockes JP (1996) Retinoid signalling and retinoid receptors in amphibian limb regeneration. *Biochem Soc Symp* 62: 137-142.
27. Stocum DL (1984) The urodele limb regeneration blastema. Determination and organization of the morphogenetic field. *Differentiation* 27: 13-28.
28. Kumar A, Godwin JW, Gates PB, Garza-Garcia AA, Brockes JP (2007) Molecular basis for the nerve dependence of limb regeneration in an adult vertebrate. *Science* 318: 772-777.
29. Gardiner DM, Carlson MR, Roy S (1999) Towards a functional analysis of limb regeneration. *Semin Cell Dev Biol* 10: 385-393.
30. Agata K, Saito Y, Nakajima E (2007) Unifying principles of regeneration I: Epimorphosis versus morphallaxis. *Dev Growth Differ* 49: 73-78.
31. Zilakos NP, Tsonis PA, Del Rio-Tsonis K, Parchment RE (1992) Newt squamous carcinoma proves phylogenetic conservation of tumors as caricatures of tissue renewal. *Cancer Res* 52: 4858-4865.
32. Tsonis PA, Eguchi G (1981) Carcinogens on regeneration. Effects of N-methyl-N'-nitro-N-nitrosoguanidine and 4-nitroquinoline-1-oxide on limb regeneration in adult newts. *Differentiation* 20: 52-60.

33. Lenhoff SGLHM (1986) *Hydra and the Birth of Experimental Biology, 1744: Abraham Trembley's Memoirs Concerning the Natural History of a Type of Freshwater Polyp with Arms Shaped Like Horns*. Boxwood Press, Pacific Grove.
34. Sanchez Alvarado A, Tsonis PA (2006) Bridging the regeneration gap: genetic insights from diverse animal models. *Nat Rev Genet* 7: 873-884.
35. French V (1976) Leg regeneration in the cockroach, *Blattella germanica*. II. Regeneration from a non-congruent tibial graft/host junction. *J Embryol Exp Morphol* 35: 267-301.
36. Haupt J, Coineau Y (1978) Moulting and morphogenesis of sensilla in a prostigmatid mite (acari, actinotrichida, actinedida: caeculidae) I. Mechanoreceptive bristles. *Cell Tissue Res* 186: 63-79.
37. William E. Duellman LT, John Hopkins (1994) *Biology of Amphibians*.
38. Young HE, Bailey CF, Dalley BK (1983) Gross morphological analysis of limb regeneration in postmetamorphic adult *Ambystoma*. *Anat Rec* 206: 295-306.
39. Pearl EJ, Barker D, Day RC, Beck CW (2008) Identification of genes associated with regenerative success of *Xenopus laevis* hindlimbs. *BMC Dev Biol* 8: 66.
40. Rohr JR, Madison DM (2003) Dryness increases predation risk in efts: support for an amphibian decline hypothesis. *Oecologia* 135: 657-664.
41. Del Rio-Tsonis K, Washabaugh CH, Tsonis PA (1995) Expression of pax-6 during urodele eye development and lens regeneration. *Proc Natl Acad Sci U S A* 92: 5092-5096.
42. Oberpriller JO, Oberpriller JC (1974) Response of the adult newt ventricle to injury. *J Exp Zool* 187: 249-253.
43. Borchardt T, Braun T (2007) Cardiovascular regeneration in non-mammalian model systems: what are the differences between newts and man? *Thromb Haemost* 98: 311-318.
44. Mescher AL (1976) Effects on adult newt limb regeneration of partial and complete skin flaps over the amputation surface. *J Exp Zool* 195: 117-128.
45. Christensen RN, Tassava RA (2000) Apical epithelial cap morphology and fibronectin gene expression in regenerating axolotl limbs. *Dev Dyn* 217: 216-224.
46. Christensen RN, Weinstein M, Tassava RA (2002) Expression of fibroblast growth factors 4, 8, and 10 in limbs, flanks, and blastemas of *Ambystoma*. *Dev Dyn* 223: 193-203.
47. Dungan KM, Wei TY, Nace JD, Poulin ML, Chiu IM, et al. (2002) Expression and biological effect of urodele fibroblast growth factor 1: relationship to limb regeneration. *J Exp Zool* 292: 540-554.
48. Mullen LM, Bryant SV, Torok MA, Blumberg B, Gardiner DM (1996) Nerve dependency of regeneration: the role of Distal-less and FGF signaling in amphibian limb regeneration. *Development* 122: 3487-3497.

Literaturverzeichnis

49. Kragl M, Knapp D, Nacu E, Khattak S, Schnapp E, et al. (2008) Novel insights into the flexibility of cell and positional identity during urodele limb regeneration. *Cold Spring Harb Symp Quant Biol* 73: 583-592.
50. da Silva SM, Gates PB, Brockes JP (2002) The newt ortholog of CD59 is implicated in proximodistal identity during amphibian limb regeneration. *Dev Cell* 3: 547-555.
51. Maden M (1983) The effect of vitamin A on the regenerating axolotl limb. *J Embryol Exp Morphol* 77: 273-295.
52. Kumar A, Gates PB, Brockes JP (2007) Positional identity of adult stem cells in salamander limb regeneration. *C R Biol* 330: 485-490.
53. Bader D, Oberpriller J (1979) Autoradiographic and electron microscopic studies of minced cardiac muscle regeneration in the adult newt, *Notophthalmus viridescens*. *J Exp Zool* 208: 177-193.
54. Bader D, Oberpriller JO (1978) Repair and reorganization of minced cardiac muscle in the adult newt (*Notophthalmus viridescens*). *J Morphol* 155: 349-357.
55. Oberpriller JO, Oberpriller JC, Arefyeva AM, Mitashov VI, Carlson BM (1988) Nuclear characteristics of cardiac myocytes following the proliferative response to mincing of the myocardium in the adult newt, *Notophthalmus viridescens*. *Cell Tissue Res* 253: 619-624.
56. Laube F, Heister M, Scholz C, Borchardt T, Braun T (2006) Re-programming of newt cardiomyocytes is induced by tissue regeneration. *J Cell Sci* 119: 4719-4729.
57. Kruse J (2008) Analyse der Herzregeneration nach mechanischer Schädigung des Ventrikels des adulten Molches *Notophthalmus viridescens*. Dissertation, Universität Halle-Wittenberg.
58. Volker Storch UW (2004) Kurzes Lehrbuch der Zoologie. Elsevier Spektrum Akademischer Verlag 8. Auflage.
59. Welsch SVU (2004) Kurzes Lehrbuch der Zoologie. Elsevier Spektrum Akademischer Verlag 8. Auflage.
60. Borchardt T, Looso M, Bruckskotten M, Weis P, Kruse J, et al. (2010) Analysis of newly established EST databases reveals similarities between heart regeneration in newt and fish. *BMC Genomics* 11: 4.
61. Kapushesky M, Emam I, Holloway E, Kurnosov P, Zorin A, et al. Gene expression atlas at the European bioinformatics institute. *Nucleic Acids Res* 38: D690-698.
62. Barrett T, Troup DB, Wilhite SE, Ledoux P, Rudnev D, et al. (2009) NCBI GEO: archive for high-throughput functional genomic data. *Nucleic Acids Res* 37: D885-890.
63. Lahm HW, Langen H (2000) Mass spectrometry: a tool for the identification of proteins separated by gels. *Electrophoresis* 21: 2105-2114.
64. Wu CC, MacCoss MJ, Howell KE, Matthews DE, Yates JR, 3rd (2004) Metabolic labeling of mammalian organisms with stable isotopes for quantitative proteomic analysis. *Anal Chem* 76: 4951-4959.

65. Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, et al. (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* 1: 376-386.
66. Ong SE, Kratchmarova I, Mann M (2003) Properties of ¹³C-substituted arginine in stable isotope labeling by amino acids in cell culture (SILAC). *J Proteome Res* 2: 173-181.
67. Kruger M, Moser M, Ussar S, Thievensen I, Lubber CA, et al. (2008) SILAC mouse for quantitative proteomics uncovers kindlin-3 as an essential factor for red blood cell function. *Cell* 134: 353-364.
68. Schmidt A, Kellermann J, Lottspeich F (2005) A novel strategy for quantitative proteomics using isotope-coded protein labels. *Proteomics* 5: 4-15.
69. Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, et al. (2004) Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics* 3: 1154-1169.
70. Thompson A, Schafer J, Kuhn K, Kienle S, Schwarz J, et al. (2003) Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem* 75: 1895-1904.
71. Ji J, Chakraborty A, Geng M, Zhang X, Amini A, et al. (2000) Strategy for qualitative and quantitative analysis in proteomics based on signature peptides. *J Chromatogr B Biomed Sci Appl* 745: 197-210.
72. Gerber SA, Rush J, Stemman O, Kirschner MW, Gygi SP (2003) Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc Natl Acad Sci U S A* 100: 6940-6945.
73. Kirkpatrick DS, Gerber SA, Gygi SP (2005) The absolute quantification strategy: a general procedure for the quantification of proteins and post-translational modifications. *Methods* 35: 265-273.
74. Ong SE, Mann M (2005) Mass spectrometry-based proteomics turns quantitative. *Nat Chem Biol* 1: 252-262.
75. Sauer B (1994) Site-specific recombination: developments and applications. *Curr Opin Biotechnol* 5: 521-527.
76. Abremski K, Hoess R (1984) Bacteriophage P1 site-specific recombination. Purification and properties of the Cre recombinase protein. *J Biol Chem* 259: 1509-1514.
77. Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* 26: 1367-1372.
78. Elias JE, Gygi SP (2007) Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods* 4: 207-214.
79. Vascotto SG, Beug S, Liversage RA, Tsilfidis C (2005) Nvbeta-actin and NvGAPDH as normalization factors for gene expression analysis in limb regenerates and cultured blastema cells of the adult newt, *Notophthalmus viridescens*. *Int J Dev Biol* 49: 833-842.

Literaturverzeichnis

80. Gao X, Fu X, Li T, Zi J, Luo Y, et al. (2003) Determining a detectable threshold of signal intensity in cDNA microarray based on accumulated distribution. *J Biochem Mol Biol* 36: 558-564.
81. Looso M (2006) Filterung und Verifizierung von Signalen niedriger Intensität auf cDNA Microarrays. Diplomarbeit FH Gießen-Friedberg.
82. Argyropoulos C, Chatziioannou AA, Nikiforidis G, Moustakas A, Kollias G, et al. (2006) Operational criteria for selecting a cDNA microarray data normalization algorithm. *Oncol Rep* 15 Spec no.: 983-996.
83. Benjamini Y, Hochberg, T. (1995) Controlling the False Discovery Rate: a practical and powerful approach to multiple testing. *J Royal Stat Soc B* 85: 289-300.
84. Heidel JD, Liu JY, Yen Y, Zhou B, Heale BS, et al. (2007) Potent siRNA inhibitors of ribonucleotide reductase subunit RRM2 reduce cell proliferation in vitro and in vivo. *Clin Cancer Res* 13: 2207-2215.
85. Avolio TM, Lee Y, Feng N, Xiong K, Jin H, et al. (2007) RNA interference targeting the R2 subunit of ribonucleotide reductase inhibits growth of tumor cells in vitro and in vivo. *Anticancer Drugs* 18: 377-388.
86. Mullican SE, Zhang S, Konopleva M, Ruvolo V, Andreeff M, et al. (2007) Abrogation of nuclear receptors Nr4a3 and Nr4a1 leads to development of acute myeloid leukemia. *Nat Med* 13: 730-735.
87. Li QX, Ke N, Sundaram R, Wong-Staal F (2006) NR4A1, 2, 3--an orphan nuclear hormone receptor family involved in cell apoptosis and carcinogenesis. *Histol Histopathol* 21: 533-540.
88. Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 8: 175-185.
89. Chou HH, Holmes MH (2001) DNA sequence quality trimming and vector removal. *Bioinformatics* 17: 1093-1104.
90. Hazelhurst S, Hide W, Liptak Z, Nogueira R, Starfield R (2008) An overview of the wcd EST clustering tool. *Bioinformatics* 24: 1542-1546.
91. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215: 403-410.
92. Creighton TE (1984) *Proteins, Structures and Molecular Properties*. ISBN 0-7167-1566-X.
93. Tanaka EM, Ferretti P (2009) Considering the evolution of regeneration in the central nervous system. *Nat Rev Neurosci* 10: 713-723.
94. Yang JC, Madupu R, Durkin AS, Ekborg NA, Pedomallu CS, et al. (2009) The complete genome of *Teredinibacter turnerae* T7901: an intracellular endosymbiont of marine wood-boring bivalves (shipworms). *PLoS One* 4: e6085.

95. Kataeva IA, Yang SJ, Dam P, Poole FL, 2nd, Yin Y, et al. (2009) Genome sequence of the anaerobic, thermophilic, and cellulolytic bacterium "Anaerocellum thermophilum" DSM 6725. *J Bacteriol* 191: 3760-3761.
96. Zhang Z, Liu C, Zhu Y, Zhong Y, Zheng H, et al. (2009) Complete genome sequence of *Lactobacillus plantarum* JDM1. *J Bacteriol*: 5020-5021.
97. Toth AL, Varala K, Newman TC, Miguez FE, Hutchison SK, et al. (2007) Wasp gene expression supports an evolutionary link between maternal behavior and eusociality. *Science* 318: 441-444.
98. Monaghan JR, Epp LG, Putta S, Page RB, Walker JA, et al. (2009) Microarray and cDNA sequence analysis of transcription during nerve-dependent limb regeneration. *BMC Biol* 7: 1.
99. Looso M, Borchardt T, Krueger M, Braun T (2010) Advanced identification of proteins in uncharacterized proteomes by pulsed in vivo SILAC. *Mol Cell Proteomics*.
100. Doherty MK, Hammond DE, Clague MJ, Gaskell SJ, Beynon RJ (2009) Turnover of the human proteome: determination of protein intracellular stability by dynamic SILAC. *J Proteome Res* 8: 104-112.
101. Maere S, Heymans K, Kuiper M (2005) BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* 21: 3448-3449.
102. Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* 132: 365-386.
103. Peters WH, Swarts HG, de Pont JJ, Schuurmans Stekhoven FM, Bonting SL (1981) (Na⁺ + K⁺)ATPase has one functioning phosphorylation site per alpha subunit. *Nature* 290: 338-339.
104. Ver Heyen M, Reed TD, Blough RI, Baker DL, Zilberman A, et al. (2000) Structure and organization of the mouse *Atp2a2* gene encoding the sarco(endo)plasmic reticulum Ca²⁺-ATPase 2 (SERCA2) isoforms. *Mamm Genome* 11: 159-163.
105. Tisi D, Talts JF, Timpl R, Hohenester E (2000) Structure of the C-terminal laminin G-like domain pair of the laminin alpha2 chain harbouring binding sites for alpha-dystroglycan and heparin. *EMBO J* 19: 1432-1440.
106. Song H, Ethier SP, Dziubinski ML, Lin J (2004) Stat3 modulates heat shock 27kDa protein expression in breast epithelial cells. *Biochem Biophys Res Commun* 314: 143-150.
107. Vieira HL, Haouzi D, El Hamel C, Jacotot E, Belzacq AS, et al. (2000) Permeabilization of the mitochondrial inner membrane during apoptosis: impact of the adenine nucleotide translocator. *Cell Death Differ* 7: 1146-1154.
108. Pham TK, Wright PC (2007) Proteomic analysis of *Saccharomyces cerevisiae*. *Expert Rev Proteomics* 4: 793-813.

Literaturverzeichnis

109. Habermann B, Oegema J, Sunyaev S, Shevchenko A (2004) The power and the limitations of cross-species protein identification by mass spectrometry-driven sequence similarity searches. *Mol Cell Proteomics* 3: 238-249.
110. Zhang Y, Ficarro SB, Li S, Marto JA (2009) Optimized Orbitrap HCD for quantitative analysis of phosphopeptides. *J Am Soc Mass Spectrom* 20: 1425-1434.
111. Finn RD, Tate J, Mistry J, Coghill PC, Sammut SJ, et al. (2008) The Pfam protein families database. *Nucleic Acids Res* 36: D281-288.
112. Hsu DR, Economides AN, Wang X, Eimon PM, Harland RM (1998) The *Xenopus* dorsalizing factor Gremlin identifies a novel family of secreted proteins that antagonize BMP activities. *Mol Cell* 1: 673-683.
113. Dionne MS, Skarnes WC, Harland RM (2001) Mutation and analysis of Dan, the founding member of the Dan family of transforming growth factor beta antagonists. *Mol Cell Biol* 21: 636-643.
114. Liao L, Park SK, Xu T, Vanderklish P, Yates JR, 3rd (2008) Quantitative proteomic analysis of primary neurons reveals diverse changes in synaptic protein content in *fmr1* knockout mice. *Proc Natl Acad Sci U S A* 105: 15281-15286.
115. Schena M, Shalon D, Davis RW, Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270: 467-470.
116. Laajala E, Aittokallio T, Lahesmaa R, Elo LL (2009) Probe-level estimation improves the detection of differential splicing in Affymetrix exon array studies. *Genome Biol* 10: R77.
117. Kadota K, Nakai Y, Shimizu K (2009) Ranking differentially expressed genes from Affymetrix gene expression data: methods with reproducibility, sensitivity, and specificity. *Algorithms Mol Biol* 4: 7.
118. Boettger T, Beetz N, Kostin S, Schneider J, Kruger M, et al. (2009) Acquisition of the contractile phenotype by murine arterial smooth muscle cells depends on the *Mir143/145* gene cluster. *J Clin Invest* 119: 2634-2647.
119. Liska AJ, Shevchenko A (2003) Expanding the organismal scope of proteomics: cross-species protein identification by mass spectrometry and its implications. *Proteomics* 3: 19-28.
120. Oeljeklaus S, Meyer HE, Warscheid B (2009) New dimensions in the study of protein complexes using quantitative mass spectrometry. *FEBS Lett* 583: 1674-1683.
121. Mo L, Dutta D, Wan Y, Chen T (2007) MSNovo: a dynamic programming algorithm for de novo peptide sequencing via tandem mass spectrometry. *Anal Chem* 79: 4870-4878.
122. Frank AM (2009) A ranking-based scoring function for peptide-spectrum matches. *J Proteome Res* 8: 2241-2252.
123. Kim S, Gupta N, Bandeira N, Pevzner PA (2009) Spectral dictionaries: Integrating de novo peptide sequencing with database search of tandem mass spectra. *Mol Cell Proteomics* 8: 53-69.

124. Shevchenko A, Sunyaev S, Liska A, Bork P (2003) Nanoelectrospray tandem mass spectrometry and sequence similarity searching for identification of proteins from organisms with unknown genomes. *Methods Mol Biol* 211: 221-234.
125. Katz A, Waridel P, Shevchenko A, Pick U (2007) Salt-induced changes in the plasma membrane proteome of the halotolerant alga *Dunaliella salina* as revealed by blue native gel electrophoresis and nano-LC-MS/MS analysis. *Mol Cell Proteomics* 6: 1459-1472.
126. Su ZZ, Emdad L, Sarkar D, Randolph A, Valerie K, et al. (2005) Potential molecular mechanism for rodent tumorigenesis: mutational generation of Progression Elevated Gene-3 (PEG-3). *Oncogene* 24: 2247-2255.
127. Morton E, Macrae IM, McCabe C, Brown SM, White F (2006) Identification of the growth arrest and DNA damage protein GADD34 in the normal human heart and demonstration of alterations in expression following myocardial ischaemia. *Int J Cardiol* 107: 126-129.
128. Bradbury A, Possenti R, Shooter EM, Tirone F (1991) Molecular cloning of PC3, a putatively secreted protein whose mRNA is induced by nerve growth factor and depolarization. *Proc Natl Acad Sci U S A* 88: 3353-3357.
129. Rouault JP, Falette N, Guehenneux F, Guillot C, Rimokh R, et al. (1996) Identification of BTG2, an antiproliferative p53-dependent component of the DNA damage cellular response pathway. *Nat Genet* 14: 482-486.
130. Boiko AD, Porteous S, Razorenova OV, Krivokrysenko VI, Williams BR, et al. (2006) A systematic search for downstream mediators of tumor suppressor function of p53 reveals a major role of BTG2 in suppression of Ras-induced transformation. *Genes Dev* 20: 236-252.
131. Cleutjens JP, Kandala JC, Guarda E, Guntaka RV, Weber KT (1995) Regulation of collagen degradation in the rat myocardium after infarction. *J Mol Cell Cardiol* 27: 1281-1292.
132. Yamamoto K, Kusachi S, Ninomiya Y, Murakami M, Doi M, et al. (1998) Increase in the expression of biglycan mRNA expression Co-localized closely with that of type I collagen mRNA in the infarct zone after experimentally-induced myocardial infarction in rats. *J Mol Cell Cardiol* 30: 1749-1756.
133. Wollert KC, Studer R, Doerfer K, Schieffer E, Holubarsch C, et al. (1997) Differential effects of kinins on cardiomyocyte hypertrophy and interstitial collagen matrix in the surviving myocardium after myocardial infarction in the rat. *Circulation* 95: 1910-1917.
134. Vazquez F, Hastings G, Ortega MA, Lane TF, Oikemus S, et al. (1999) METH-1, a human ortholog of ADAMTS-1, and METH-2 are members of a new family of proteins with angio-inhibitory activity. *J Biol Chem* 274: 23349-23357.
135. Nakamura K, Hirohata S, Murakami T, Miyoshi T, Demircan K, et al. (2004) Dynamic induction of ADAMTS1 gene in the early phase of acute myocardial infarction. *J Biochem* 136: 439-446.
136. Lau LF, Lam SC (1999) The CCN family of angiogenic regulators: the integrin connection. *Exp Cell Res* 248: 44-57.

Literaturverzeichnis

137. Pi L, Ding X, Jorgensen M, Pan JJ, Oh SH, et al. (2008) Connective tissue growth factor with a novel fibronectin binding site promotes cell adhesion and migration during rat oval cell activation. *Hepatology* 47: 996-1004.
138. Latinkic BV, Mercurio S, Bennett B, Hirst EM, Xu Q, et al. (2003) *Xenopus* Cyr61 regulates gastrulation movements and modulates Wnt signalling. *Development* 130: 2429-2441.
139. Mo FE, Muntean AG, Chen CC, Stolz DB, Watkins SC, et al. (2002) CYR61 (CCN1) is essential for placental development and vascular integrity. *Mol Cell Biol* 22: 8709-8720.
140. Lv H, Fan E, Sun S, Ma X, Zhang X, et al. (2009) Cyr61 is up-regulated in prostate cancer and associated with the p53 gene status. *J Cell Biochem* 106: 738-744.
141. Hirschfeld M, zur Hausen A, Bettendorf H, Jager M, Stickeler E (2009) Alternative splicing of Cyr61 is regulated by hypoxia and significantly changed in breast cancer. *Cancer Res* 69: 2082-2090.
142. Tsai MS, Bogart DF, Li P, Mehmi I, Lupu R (2002) Expression and regulation of Cyr61 in human breast cancer cell lines. *Oncogene* 21: 964-973.
143. Croci S, Landuzzi L, Astolfi A, Nicoletti G, Rosolen A, et al. (2004) Inhibition of connective tissue growth factor (CTGF/CCN2) expression decreases the survival and myogenic differentiation of human rhabdomyosarcoma cells. *Cancer Res* 64: 1730-1736.
144. Athanasopoulos AN, Schneider D, Keiper T, Alt V, Pendurthi UR, et al. (2007) Vascular endothelial growth factor (VEGF)-induced up-regulation of CCN1 in osteoblasts mediates proangiogenic activities in endothelial cells and promotes fracture healing. *J Biol Chem* 282: 26746-26753.
145. Shimizu-Nishikawa K, Tsuji S, Yoshizato K (2001) Identification and characterization of newt rad (ras associated with diabetes), a gene specifically expressed in regenerating limb muscle. *Dev Dyn* 220: 74-86.
146. Kim WS, Stocum DL (1986) Retinoic acid modifies positional memory in the anteroposterior axis of regenerating axolotl limbs. *Dev Biol* 114: 170-179.
147. Closs EI, Boissel JP, Habermeier A, Rotmann A (2006) Structure and function of cationic amino acid transporters (CATs). *J Membr Biol* 213: 67-77.
148. Rotmann A, Vekony N, Gassner D, Niegisch G, Strand D, et al. (2006) Activation of classical protein kinase C reduces the expression of human cationic amino acid transporter 3 (hCAT-3) in the plasma membrane. *Biochem J* 395: 117-123.
149. Frantz S, Kobzik L, Kim YD, Fukazawa R, Medzhitov R, et al. (1999) Toll4 (TLR4) expression in cardiac myocytes in normal and failing myocardium. *J Clin Invest* 104: 271-280.
150. Oyama J, Blais C, Jr., Liu X, Pu M, Kobzik L, et al. (2004) Reduced myocardial ischemia-reperfusion injury in toll-like receptor 4-deficient mice. *Circulation* 109: 784-789.
151. Shishido T, Nozaki N, Yamaguchi S, Shibata Y, Nitobe J, et al. (2003) Toll-like receptor-2 modulates ventricular remodeling after myocardial infarction. *Circulation* 108: 2905-2910.

152. Kawakami Y, Rodriguez Esteban C, Raya M, Kawakami H, Marti M, et al. (2006) Wnt/beta-catenin signaling regulates vertebrate limb regeneration. *Genes Dev* 20: 3232-3237.
153. Haynes T, Gutierrez C, Aycinena JC, Tsonis PA, Del Rio-Tsonis K (2007) BMP signaling mediates stem/progenitor cell-induced retina regeneration. *Proc Natl Acad Sci U S A* 104: 20380-20385.
154. Grogg MW, Call MK, Okamoto M, Vergara MN, Del Rio-Tsonis K, et al. (2005) BMP inhibition-driven regulation of six-3 underlies induction of newt lens regeneration. *Nature* 438: 858-862.
155. Bon E, Steegers R, Steegers EA, Ursem N, Charif H, et al. (2009) Proteomic Analyses of the Developing Chicken Cardiovascular System. *J Proteome Res*: 268-274.
156. Lam L, Tsoutsman T, Arthur J, Semsarian C (2009) Differential protein expression profiling of myocardial tissue in a mouse model of hypertrophic cardiomyopathy. *J Mol Cell Cardiol*: Epub ahead of print.
157. Bousette N, Kislinger T, Fong V, Isserlin R, Hewel JA, et al. (2009) Large-scale characterization and analysis of the murine cardiac proteome. *J Proteome Res* 8: 1887-1901.
158. Grant JE, Bradshaw AD, Schwacke JH, Baicu CF, Zile MR, et al. (2009) Quantification of protein expression changes in the aging left ventricle of *Rattus norvegicus*. *J Proteome Res* 8: 4252-4263.
159. Zatyka M, Ricketts C, da Silva Xavier G, Minton J, Fenton S, et al. (2008) Sodium-potassium ATPase 1 subunit is a molecular partner of Wolframin, an endoplasmic reticulum protein involved in ER stress. *Hum Mol Genet* 17: 190-200.
160. Berger BJ, Muller TS, Buschmann IR, Peters K, Kirsch M, et al. (2003) High levels of the molecular chaperone Mdg1/ERdj4 reflect the activation state of endothelial cells. *Exp Cell Res* 290: 82-92.
161. Mardones GA, Snyder CM, Howell KE (2006) Cis-Golgi matrix proteins move directly to endoplasmic reticulum exit sites by association with tubules. *Mol Biol Cell* 17: 525-538.
162. Liang J, Yin C, Doong H, Fang S, Peterhoff C, et al. (2006) Characterization of erasin (UBXD2): a new ER protein that promotes ER-associated protein degradation. *J Cell Sci* 119: 4011-4024.
163. Helenius A, Marquardt T, Braakman I (1992) The endoplasmic reticulum as a protein-folding compartment. *Trends Cell Biol* 2: 227-231.
164. Harding HP, Zhang Y, Ron D (1999) Protein translation and folding are coupled by an endoplasmic-reticulum-resident kinase. *Nature* 397: 271-274.
165. Calton M, Zeng H, Urano F, Till JH, Hubbard SR, et al. (2002) IRE1 couples endoplasmic reticulum load to secretory capacity by processing the XBP-1 mRNA. *Nature* 415: 92-96.
166. Zhu C, Johansen FE, Prywes R (1997) Interaction of ATF6 and serum response factor. *Mol Cell Biol* 17: 4957-4966.

Literaturverzeichnis

167. Oh-hashii K, Koga H, Ikeda S, Shimada K, Hirata Y, et al. (2009) CRELD2 is a novel endoplasmic reticulum stress-inducible gene. *Biochem Biophys Res Commun* 387: 504-510.
168. Palmer S, Groves N, Schindeler A, Yeoh T, Biben C, et al. (2001) The small muscle-specific protein Csl modifies cell shape and promotes myocyte fusion in an insulin-like growth factor 1-dependent manner. *J Cell Biol* 153: 985-998.
169. Harvey RP (1996) NK-2 homeobox genes and heart development. *Dev Biol* 178: 203-216.
170. Tanaka M, Chen Z, Bartunkova S, Yamasaki N, Izumo S (1999) The cardiac homeobox gene *Csx/Nkx2.5* lies genetically upstream of multiple genes essential for heart development. *Development* 126: 1269-1280.
171. Chen J, Shpall RL, Meyerdierks A, Hagemeyer M, Bottger EC, et al. (2000) Interferon-inducible Myc/STAT-interacting protein Nmi associates with IFP 35 into a high molecular mass complex and inhibits proteasome-mediated degradation of IFP 35. *J Biol Chem* 275: 36278-36284.
172. Bannasch D, Weis I, Schwab M (1999) Nmi protein interacts with regions that differ between MycN and Myc and is localized in the cytoplasm of neuroblastoma cells in contrast to nuclear MycN. *Oncogene* 18: 6810-6817.
173. Schlierf B, Lang S, Kosian T, Werner T, Wegner M (2005) The high-mobility group transcription factor Sox10 interacts with the N-myc-interacting protein Nmi. *J Mol Biol* 353: 1033-1042.
174. Lee ND, Chen J, Shpall RL, Naumovski L (1999) Subcellular localization of interferon-inducible Myc/stat-interacting protein Nmi is regulated by a novel IFP 35 homologous domain. *J Interferon Cytokine Res* 19: 1245-1252.
175. Zhu M, John S, Berg M, Leonard WJ (1999) Functional association of Nmi with Stat5 and Stat1 in IL-2- and IFN γ -mediated signaling. *Cell* 96: 121-130.
176. Fillmore RA, Mitra A, Xi Y, Ju J, Scammell J, et al. (2009) Nmi (N-Myc interactor) inhibits Wnt/beta-catenin signaling and retards tumor growth. *Int J Cancer* 125: 556-564.
177. Herrmann J, Borkham-Kamphorst E, Haas U, Van de Leur E, Fraga MF, et al. (2006) The expression of CSRP2 encoding the LIM domain protein CRP2 is mediated by TGF-beta in smooth muscle and hepatic stellate cells. *Biochem Biophys Res Commun* 345: 1526-1535.
178. Yet SF, Folta SC, Jain MK, Hsieh CM, Maemura K, et al. (1998) Molecular cloning, characterization, and promoter analysis of the mouse *Crp2/SmLim* gene. Preferential expression of its promoter in the vascular smooth muscle cells of transgenic mice. *J Biol Chem* 273: 10530-10537.
179. Jain MK, Fujita KP, Hsieh CM, Endege WO, Sibinga NE, et al. (1996) Molecular cloning and characterization of SmLIM, a developmentally regulated LIM protein preferentially expressed in aortic smooth muscle cells. *J Biol Chem* 271: 10194-10199.
180. Clowes AW, Reidy MA, Clowes MM (1983) Kinetics of cellular proliferation after arterial injury. I. Smooth muscle growth in the absence of endothelium. *Lab Invest* 49: 327-333.

181. Eppenberger-Eberhardt M, Flamme I, Kurer V, Eppenberger HM (1990) Reexpression of alpha-smooth muscle actin isoform in cultured adult rat cardiomyocytes. *Dev Biol* 139: 269-278.
182. Bao S, Zhu J, Garvey WT (1998) Cloning of Rab GTPases expressed in human skeletal muscle: studies in insulin-resistant subjects. *Horm Metab Res* 30: 656-662.
183. Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, et al. (2009) InterPro: the integrative protein signature database. *Nucleic Acids Res* 37: D211-215.