Molekularbiologische Untersuchungen der frühen Herz-und Linsenregeneration am Molch *Notopthalmus viridescens* mit Hochdurchsatzanalysen

Inauguraldissertation
zur Erlangung des Grades eines Doktors der Medizin
des Fachbereichs Medizin
der Justus-Liebig-Universität Gießen

vorgelegt von Michel, Christian Sebastian
aus Gelnhausen

Gießen, 2013

Aus dem Max-Planck-Institut für Herz-und Lungenforschung in Bad Nauheim unter
Leitung von Prof. Dr. Dr. Thomas Braun.

Gutachter: Prof. Dr. Dr. Thomas Braun
Gutachter: Prof. Dr. Heinrich Sauer

Tag der Disputation
03.12.2013

Meinen Eltern

Inhaltsverzeichnis

1. Einleitung

Im Mittelpunkt dieser Arbeit steht die Untersuchung regenerativer Prozesse im rot gepunkteten grünen Wassermolch (*Notophthalmus viridescens*), insbesondere an Herz und Linse, mit Hilfe von Hochdurchsatzanalysen. Mit der Erforschung regenerativer Vorgänge im Molch erhofft man Erkenntnisse zu erlangen, um diese für medizinische Therapien nutzen zu können. Die Motivation, die mit der Erforschung regenerativer Prozesse einhergeht, ist der Wunsch Ansätze zu finden, um eventuell Querschnittsgelähmten oder auch Patienten nach myokardialem Infarkt zu helfen, ihre Erkrankung zu heilen.

Für Säugetiere im Allgemeinen besitzt die Leber ein hohes Regenerationspotential. Pionierarbeit auf diesem Gebiet leisteten Higgins und Anderson im Jahre 1931, als sie zeigen konnten, dass nach partieller Resektion der Leber eine Regeneration des ektomierten Areals auftrat[1]. Neben der Leber besitzen Haut, Darm, Niere, Muskel und in Teilen das Nervensystem die Fähigkeit zur Regeneration[2]. Im Gegensatz zu Säugetieren weist der Molch, zur Klasse der Amphibien gehörend, ein außergewöhnlich hohes Regenerationspotential auf (Abbildung Nr.1). Erste wissenschaftliche Untersuchungen zu den Fähigkeiten von Molchen und Salamandern wurden bereits vor mehr als 200 Jahren von Lorrenzo Spallanzani durchgeführt. Vor allem das Potential Beine und Schwänze regenerieren zu können wurde beschrieben[3]. Im Laufe der letzten Jahrzehnte konnte von Forschern weiterhin gezeigt werden, dass sich neben Beinen und Schwänzen auch Augenlinse[4], Herz[5], Gehirn und Rückenmark[6] regenerieren können. Neben Salamandern und Molchen besitzen Planarien[7] und einige Fischarten, wie zum Beispiel der Zebrafisch[8], ein ebenfalls ausgeprägtes Regenerationspotential.

Der Molch, als Modellorganismus für die Regeneration, ist verglichen mit dem Zebrafisch auf molekularbiologischer Ebene bisher in geringerem Maße erforscht worden. Das Genom des Molches ist bis dato weitestgehend unerforscht, da es etwa die 10 fache Größe des menschlichen Genoms besitzt und eine Sequenzierung zur Zeit noch sehr teuer ist[9]. Dennoch kann durch den Einsatz adaptierter Ansätze, wie beispielsweise EST-Sanger Sequenzierungen, cDNA Microarrays, RNAseq und massenspektrometrische Analysen, ein tieferer Einblick in den Ablauf der Regenerationsprozesse des Molches erreicht werden. Neue Daten über den Ablauf dieser Prozesse dürften den molekularbiologischen Zugang zum Molch erleichtern und

gezielte Untersuchungen einzelner Kandidaten während der Regeneration befördern. Mit dieser Dissertation sollen vor allem die Abläufe früher regenerativer Prozesse unter dem besonderen Aspekt der Herz-und Linsenregeneration beleuchtet werden. Im nachfolgenden Kapitel wird eine Einführung in den aktuellen Stand der Regenerationsforschung gegeben, unter eingehender Betrachtung der Regeneration im Molch und dessen Bedeutung als Modellorganismus.

Abbildung Nr. 1: Photographie des *Notophthalmus viridescens*

Der Molch *Notophthalmus viridescens*, zur Ordnung der Schwanzlurche gehörend, besitzt im adulten Zustand eine Körperlänge zwischen 7 und 15 cm und hält sich überwiegend im Wasser auf. Der Südosten Kanadas, die Ostküste Nordamerikas bis in den Golf von Mexiko ist sein natürliches Vorkommensgebiet.

1.1. Allgemeine Regeneration und deren molekularbiologische Prozesse

Die Fähigkeit zur Regeneration ist im Tierreich weit verbreitet, allerdings variiert diese teilweise sehr von einem zum anderen Organismus. Planarien sind in der Lage fehlende Köpfe und komplett fehlende Körper, von kleineren Fragmenten ausgehend, zu regenerieren, während Salamander nicht in der Lage sind nach Dekapitation weiterzuleben, aber andere Körperteile und Organe regenerieren können. Aus diesen Beobachtungen heraus ergeben sich viele Fragen, darunter: Wie startet ein Regenerationsprozess? Welche Zellen sind daran beteiligt? Wie werden sie aktiviert und wie weiß das regenerierende Organ, wann der Regenerationsprozess abgeschlossen ist?
Um etwa 1700 führte der Schweizer Zoologe Abraham Trembley erste Versuche zur Regenerationsfähigkeit der Hydra durch und konnte zeigen, dass es nach Teilung der Hydra in beiden Teilen zu einer vollständigen Wiederherstellung kommt und dass es

möglich ist, einen Teil einer Hydra einer anderen Hydra aufzupfropfen und diese miteinander zusammenwachsen[10]. Etwa 70 Jahre später konnte Spallanzani zeigen, dass auch höher entwickelte Organismen, wie Urodele (Schwanzlurche) komplette Gliedmaßen regenerieren können[3].

Hydra sind Nesseltiere, die zwei Keimblätter und eine Körperachse mit zwei Polen besitzen. Der obere Pol, aus Kopf mit Tentakeln und Mund bestehend, und der untere Pol, der auch als Fuß bezeichnet wird[11]. Die beiden äußeren Keimblätter besitzen jeweils nur eine einzige Zellschicht. Die innere Zellschicht besteht aus endodermalen Zellen und die äußere Zellschicht aus ektodermalen Zellen. Innerhalb der ektodermalen Zellschicht befinden sich interstitielle Stammzellen, die der Selbsterneuerung dienen und Neuronen sowie sekretorische Zellen und Gameten bilden können (Abbildung Nr. 2)[12].

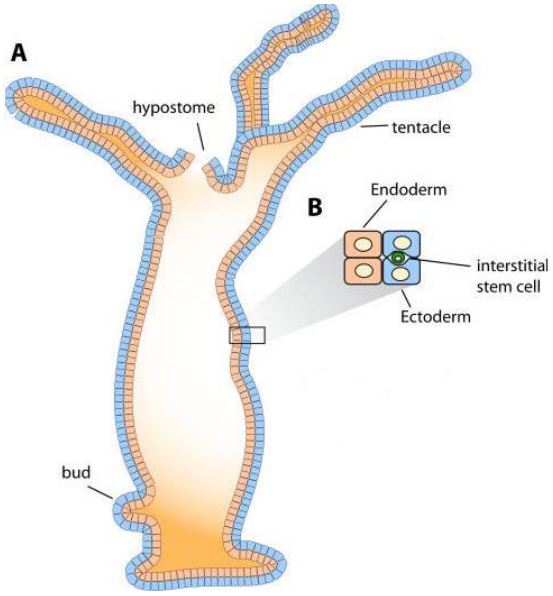


Abbildung Nr. 2: Schematischer Aufbau einer Hydra

(A) Übersicht über den Aufbau einer Hydra mit Tentakeln (tentacle), Mundöffnung (hypostome) und einzelner Knospe (bud), die der asexuellen Reproduktion dient. (B) Doppelte Zellschicht, bestehend aus Endoderm (endoderm) und Ektoderm (ectoderm), sowie interstitiellen Stammzellen (interstitial stem cells). Entnommen aus[8].

Alle drei Stammzelltypen besitzen eine unbeschränkte Möglichkeit der Selbsterneuerung. Um diese Stammzellfunktion auch im adulten Zustand beibehalten zu können, muss eine ständige Balance zwischen Differenzierung und Selbsterneuerung eingehalten werden. Zu Signalwegen und Faktoren von denen bekannt ist, dass sie die

Differenzierung von Stammzellen in der Hydra regulieren, zählen der Wnt, Notch, Glycogen synthase kinase-3 beta und Map-kinase-CREB Signalweg[12-14]. Im Wnt3 Signalweg können bereits 15 Minuten nach Dekapitation erste Veränderungen registriert werden[15]. Interessanter Weise ist dabei die Höhe der Dekapitation ein entscheidendes Kriterium. Bei apikaler Kopf Regeneration findet das Remodeling unter Abwesenheit von apoptotischen Prozessen und unter verstärkter Zellproliferation statt, wobei epitheliale Zellen die Wnt3 Expression rapide heraufregulieren. Bei basaler Kopf Regeneration kommt es zur Apoptose von Zellen, die aus der Linie der interstitiellen Linie stammen, wie zum Beispiel von Neuronen und sekretorischen Zellen. Diese bilden eine der Hauptquellen für die Wnt3 Produktion, die in den umgebenden Zellen den Wnt-β-catenin Signalweg aktivieren und eine rasche Zellteilung, auch als epimorph-ähnlich beschrieben, einleiten[16]. Aus diesem Beispiel heraus wird ersichtlich, dass Regeneration ein hoch dynamischer Prozess ist, indem die Wundheilung und die Wiederentwicklung von fehlenden Strukturen in Zusammenhang gesehen werden müssen und nicht voneinander getrennte Abläufe sind. Allerdings muss man bezüglich der außergewöhnlich hohen Regenerationsfähigkeit der Hydra einwerfen, dass sie im Vergleich zu Vertebraten einen sehr hohen Anteil an adulten pluripotenten Stammzellen besitzt und ihre Gewebs- und Organkomplexität, ebenfalls mit Vertebraten verglichen, wesentlich niedriger entwickelt ist.
Schwanzlurche, wie der Molch, oder Fische können keine inneren Organe de novo regenerieren. Partielle Gewebe-oder Zellentfernung von Herz, Gehirn, Schwanz oder auch Bein führt zur Einleitung eines Regenerationsprozesses[17]; eine Totalektomie von Herz oder auch Gehirn führt zum Tod des Tieres.

Historisch gesehen unterscheidet man die Regeneration nach partieller Schädigung in eine mit und eine ohne Blastembildung. Diese Sichtweise ist mittlerweile allerdings sehr umstritten[18, 19]. Als Blastem bezeichnet man nach neueren Forschungsergebnissen eine Mixtur von Zellen mit unterschiedlicher, teilweise beschränkter Differenzierungsfähigkeit und Gewebsherkunft, die im Zusammenspiel einen Regenerationsprozess koordiniert[8]. Jahrzehntelang war weitestgehend akzeptiert, dass ein Blastem aus multi- oder pluripotenten Zellen besteht. Histologisch beschrieb man das Blastem als ein homogenes Gebilde[20].
Die Fähigkeit zur Blastembildung von Salamandern ist Gegenstand jahrelanger Forschungen, die jedoch noch immer viele Fragen unbeantwortet lassen oder deren

Ergebnisse teilweise kontrovers zueinander stehen. Wie neuere Forschungsergebnisse nahelegen, soll es sich bei einem Blastem eher um ein heterogenes Gebilde aus eingeschränkten Vorläuferzellen handeln. Erste Beweise hierfür wurden 2004 von Gargioli und Slack erbracht, als sie für die *Xenopus* Schwanz Regeneration zeigen konnten, dass unterschiedliche Gewebe von verschiedenartigen Vorläufer Zellpools stammen[21]. Weitergehende Untersuchungen mit transgenen Axolotls, die konstitutiv ein GFP-Transgen exprimierten, ermöglichten die Beobachtung, dass markierte Muskelfasern und Satellitenzellen bevorzugt Muskeln regenerieren können und nicht zur Knorpel-oder Epidermisbildung beitragen. Epidermale Zellen hingegen waren in der Lage Knorpel und Bindegewebe aufzubauen, und Schwannzellen konnten nur Schwannzellen bilden[22]. Dabei konnte auch festgestellt werden, dass Knorpelzellen im Gegensatz zu Schwannzellen eine proximo- distale Positionsidentität besitzen, da sie sowohl den für die distale Positionsbestimmung assoziierten Marker HOXA13, als auch das Zellkernprotein MEIS 1+2, einen weiteren Positionsbestimmer, exprimieren. Auf Basis dieser Erkenntnisse änderte sich der konzeptionelle Aufbau eines Blastemas und unterstützt die These, dass es sich bei der Gewebsneubildung nicht um eine Dedifferenzierung zu pluripotenten Zellen handelt, von denen die Regeneration ausgeht. Die Frage, inwieweit anwesende gewebsspezifische Stamm-oder Progenitorzellen eine Rolle spielen und die Regeneration beeinflussen, ist Gegenstand intensiver Forschungsbemühungen[23].

Ein weiterer Aspekt, der bei der Erforschung der Blastem-gestützten Beinregeneration des Salamanders gefunden wurde, sind die Erkenntnisse, dass es wahrscheinlich im Laufe der Evolution zur Entwicklung regenerationsspezifischer Salamanderproteine gekommen ist. Ein Beispiel hierfür wäre PROD1, dass entscheidend die proximo-distale Identität beeinflusst[24, 25]. Dies könnte eine Erklärung dafür sein, warum einige Tiere in der Lage sind Körperteile zu regenerieren und andere nicht. Die lokale Evolution der Beinregeneration von Salamandern, so wird spekuliert, könnte ihren Ausgangspunkt in der Perm Zeit haben, als sich die Vorfahren der heutigen Salamander von der Vertebratenlinie abgezweigt haben[26]. Die Entdecker von PROD1 beschreiben es als ein Protein, dass hochkonservierte zelluläre Mechanismen, die sowohl in regenerierenden, als auch in nicht regenerierenden Tieren vorhanden sind, beeinflusst. Es ist ebenfalls bekannt, dass zwei Bindungsstellen im PROD1 Promoter des Axolotls für den bereits erwähnten proximalen Positionsbestimmer MEIS 1+2 bestehen und Mutationen zu Veränderungen im proximalen Blastem führen. Diese Hypothese steht

im Gegensatz zur weit verbreiteten und populäreren Annahme, dass es sich bei der Beinregeneration der Salamander um eine Fähigkeit handelt, die bei der Entwicklung höher entwickelter Vertebraten verloren gegangen ist[27]. Als Beispiel wird hierfür von Alexandra Bely die Verteilung der Fähigkeit zur Flossenregeneration von Fischen angeführt[28]. Im Kontext phylogenetischer Vergleiche der Regenerationsfähigkeit führt sie an, dass diese Fähigkeit für die frühen Vertreter bestehe und spätere Vertreter diese Fähigkeit verloren hätten. Gründe für den Verlust der Regenerationsfähigkeit könnten darin bestehen, dass ein Totalverlust der betroffenen Struktur die Funktion des Organismus in geringerem Maße schädigt als eine partielle Regeneration, die eventuell ein hohes Maß an Ressourcen verbraucht. Eine weitere Erklärung könnte sein, dass die Fähigkeit zur Regeneration keinen signifikanten Selektionsvorteil erbrachte und als neutrales Merkmal verloren gegangen sein könnte. Welche der beiden Hypothesen zur Regeneration zutreffend ist, könnte durch ein tiefergehendes Verständnis, wie Regenerationsprozesse ablaufen, eventuell beantwortet werden.

1.2. Linsenregeneration im Molch

Neben der Bein und Schwanzregeneration besitzen einige Amphibien, wie prämetamorphe Frösche und adulte Molche, die Fähigkeit zur Augenlinsenregeneration[29]. Erstmals wurde diese Fähigkeit für den adulten Molch von Colucci im Jahre 1891[30] und von Freeman im Jahre 1963 für prämetamorphe Frösche entdeckt[31].

Im Molch konnte die Regenerationsfähigkeit der Linse nach wiederholter Entfernung und anschließender Regeneration bis ins hohe Alter (30 Jahre) gezeigt werden. Im Zeitraum von 18 Jahren wurde einem Molch 16mal hintereinander die Augenlinse entfernt, und dabei gezeigt, dass sich die Linse in Form, Größe, Transparenz und Genexpressionsmuster gegenüber intakten Linsen von 14 Jahre alten Molchen kaum unterscheidet. Verzögerungen im Regenerationsprozess wurden über die gesamte Zeit hin nicht beobachtet[4]. Der Prozess der Linsenregeneration im Molch unterscheidet sich von der im Frosch insbesondere im initialen Stadium in einigen Punkten[32]. Ausgangspunkt für die Regeneration der Linse im Molch ist nicht die Kornea, sondern sind die dorsalen Iris-Pigmentepithelzellen. Innerhalb der ersten 4 Tage nach Lentektomie kommt es zur Dedifferenzierung, Verlust der Pigmentierung und beginnender Proliferation der Iris-Pigmentepithelzellen (Abbildung Nr. 3)[33, 34].

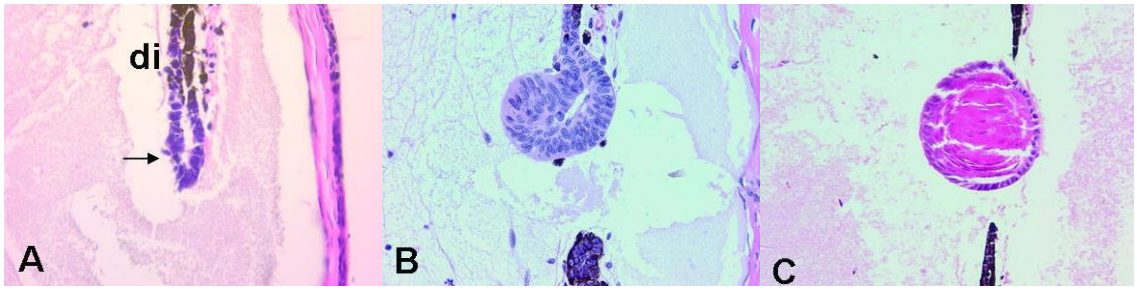


Abbildung Nr. 3: Linsenregeneration im Molch

(A) 10 Tage post Lentektomie, Enstehung eines ersten Linsen-Vesikels. (B) Bildung von ersten Linsenfasern nach 15 Tagen der Regeneration. (C) 20 Tage nach Lentektomie, Bildung einer differenzierten Augenlinse. Entnommen aus[35].

In den ersten Tagen leiten Moleküle wie PAX6, SOX2, mafB sowie FGF2 einen Wiedereintritt in den Zellzyklus verbunden mit einer Dedifferenzierung, sowohl in dorsalem, als auch in ventralem Iris-Pigmentepithel, ein. Der Wiedereintritt in den Zellzyklus und die Genexpression im ventralen Irispigmentepithel wird später allerdings unterdrückt[36]. In der dorsalen Iris kommt es nachfolgend zu einer gegenüber der Ventralen deutlich erhöhten FGF2 Expression und zur Aktivierung von Wnt-Signalweg assoziierten Molekülen, die wiederum in der dorsalen Iris eine Transdifferenzierung und eine Expression von Genen wie PROX1 und SOX1 bewirken[37, 38]. Zusätzlich haben Genexpressionsmuster, basierend auf Hochdurchsatzanalysen gezeigt, dass während der Dedifferenzierung auch Gene aus den Bereichen der RAS family, der P53 tumour suppressor family, der TNF family, der Retinoblastoma und Jun transcription factor family eine Rolle spielen[39].

Der Einfluss der extrazellulären Matrix auf die Regeneration der Molchlinse ist bisher nur in einigen Ansätzen erforscht. Mitte der 80er Jahre wurden erstmals von Kulyk et al. Beobachtungen zu erhöhten Mengen an Glykosaminoglykanen und Hyaluronat im dorsalen Teil gegenüber dem ventralen Teil der Iris festgestellt[40]. Das Glykoprotein Fibronectin wurde wenig später als eine weitere Komponente der extrazellulären Matrix identifiziert, das sich am unbeschädigten Auge an der basalen Oberfläche der Iris und bei beginnender Depigmentierung des Iris- Pigmentepithels vermehrt an der basolateralen und apikalen Zelloberfläche befindet und die dedifferenzierenden Zellen umgibt[41]. Zu Veränderungen der extrazellulären Matrix bei der Regeneration der

Augenlinse und der Dedifferenzierung der Iris-Pigmentepithelzellen wurden im Jahr 2010 neue Ergebnisse von Godwin et al. veröffentlicht, die die Formation eines Fibrin Propf im dorsalen Teil der Iris und dessen Rolle beschreiben. Fibrin Pröpfe sind bekannt für ihre adhäsiven Eigenschaften für Leukozyten und Wachstumsfaktoren[42]. Der Fibrin Propf soll eine lokalisierte Expression von Wachstumsfaktoren, wie FGF2, initiieren und den bereits erwähnten Wnt-Signalweg aktivieren[43]. Bei Blockierung der Thrombin Aktivierung und einer nachfolgenden Verhinderung der Entstehung eines Fibrin Propf konnte gezeigt werden, dass die Linse des Molches sich nicht mehr regenerieren kann[44]. Diese Erkenntnisse sind auch von vergleichbaren Gewebeverletzungen im Säugerbereich bekannt, die im Zusammenhang einer Fibrin und Thrombin Aktivierung eine verstärkte FGF2 Expression zeigen[45].

Ein weiterer wichtiger Aspekt, der bei Hochdurchsatzanalysen in der Regeneration der Linse des Molches gefunden wurde, ist die Expression von Genen, die in pluripotenten Zellen exprimiert werden (stem cell pluripotency factors). Während der Dedifferenzierung der Zellen des Iris-Pigmentepithels wurde eine Akkumulation des Proteins Nucleostemin in den Nukleoli gefunden. Es ist aus der Säugerwelt bekannt, dass Nucleostemin im Nukleolus von Stammzellen ein erhöhtes Expressionslevel besitzt, in den Zell-Zyklus eingreift und den Transkriptionsfaktor P53 moduliert[46]. Ebenso konnte die Expression von drei weiteren „stem cell factors“ gefunden werden, darunter SOX2, KLF4 und C-MYC. Andere spezifische Gene für pluripotente Zellen wie NANOQ und OCT4 konnten über den gesamten Verlauf der Linsenregeneration nicht gefunden werden. Eine mögliche Erklärung für dieses Phänomen könnte darin liegen, dass die dedifferenzierte Pigmentepithelzelle ausschließlich in der Lage ist eine neue Linse zu regenerieren und nicht in vollem Maße pluripotenten Stammzellcharakter besitzt[47].

Weitere Gene, die eine Rolle bei der Dedifferenzierung, der anschließenden Transdifferenzierung und innerhalb der extrazellulären Matrix eine Rolle spielen, konnten bisher nur teilweise gefunden werden. Insbesondere Untersuchungen sehr früher Zeitpunkte in Kombination mit Hochdurchsatzanalysen, wie cDNA-Microarrays, könnten dabei neue Antworten liefern.

1.3. Regeneration des Herzens

Die Herzregeneration in Säugetieren, Amphibien und Fischen gehört seit einigen Jahrzehnten zu den mit am meist erforschten Feldern der modernen Medizin. Einer der dominierenden Gründe liegt darin, dass kardiovaskuläre Erkrankungen in hoch entwickelten Industrienationen in Statistiken zur Todesursache an erster Stelle stehen und nach Schätzungen der World Health Organization (WHO) auch weiterhin die häufigste Todesursache bleiben werden[48].

Das menschliche Herz zeigt verglichen mit anderen menschlichen Organen, wie Leber, Haut, Skelettmuskel, Lunge und Knochen, nur eine sehr begrenzte Kapazität zur Regeneration. Hauptgrund, nach Meinung vieler Forscher, ist eine nur gering ausgeprägte Fähigkeit von adulten Kardiomyozyten wieder in den Zellzyklus einzutreten und sich zu teilen. Menschliche Kardiomyozyten im adulten Zustand bestehen überwiegend aus einem Zellkern mit tetraploidem oder auch polyploidem Genom. Postnatal besitzen diese Kardiomyozyten noch ein diploides Genom und sind noch einige Monate in der Lage zu proliferieren[49]. Mit Hilfe der ^{14}C Radiokarbon Methode konnten, von zeitnah verstorben Menschen verschiedenen Alters, Raten bezüglich der jährlichen Zellerneuerung (Turnover) von Kardiomyozyten berechnet werden. Im Alter von 25 Jahren mit einer Rate der Zellerneuerung von etwa 1% im Jahr, reduziert sich diese auf 0,45% im Alter von 75 Jahren. Dies würde bedeuten, dass sich ca. 45% aller Kardiomyozyten im Laufe eines Lebens erneuern, aber auch etwa 55% der Kardiomyozyten von Geburt bis Lebensende sich nicht erneuern[50]. In anderen Veröffentlichungen mit Messungen von Zellzyklus Markern wie Ki67 und der Inkorporation von Nukleotid-Analoga, wie Iododeoxyuridin, in neu synthetisierte DNA, wurde Raten von etwa 30% Zellerneuerung pro Jahr beschrieben[51]. Gründe für diese deutlichen Unterschiede in den Ergebnissen, liegen zum einen am geringeren Turnover von Kardiomyozyten verglichen mit umgebenden Stromazellen, und zum anderen an den jeweils einzelnen Messmethoden und deren Fehlern, wie z.B. die Autofluoreszenz des Herzens bei der Ki67 Methode oder kleinste Verunreinigungen in der Durchflusszytometrie bei der ^{14}C Radiokarbon Methode, die eine große Streuung hervorrufen können[52]. Diese Messungen zeigen jedoch eine partiell vorhandene Fähigkeit des menschlichen Herzens zur Zellerneuerung bzw. Regeneration und wiederlegen das Paradigma, dass adulte Kardiomyozyten nicht ersetzt werden können.

Prinzipiell gibt es dabei drei unterschiedliche Wege der Zellerneuerung bzw. Regeneration von Kardiomyozyten die diskutiert werden. Die erste geht von sich teilenden reifen Kardiomyozyten, die zweite von dedifferenzierenden Kardiomyozyten und die dritte von sich differenzierenden Vorläuferzellen aus[53-57]. Alle drei beschriebenen Wege sind allerdings in unterschiedlichen Organismen und unter teilweise sehr verschiedenen pathologischen Bedingungen erforscht worden und erlauben daher nur bedingt Vergleiche untereinander.

1.3.1. Herzregeneration in der Maus

Innerhalb der Säugetiere gilt die Maus aufgrund einer Vielzahl von genetischen Manipulations-und Untersuchungsmöglichkeiten und ihrer genetischen Nähe zum Menschen als Modellorganismus für die Bemessung des Potentials der Herzregeneration. Ebenso, wie das menschliche Herz, ist das Herz der Maus ein 4 Kammersystem mit zwei getrennten Kreisläufen, und die Mehrheit der Kardiomyozyten verliert postnatal die Fähigkeit sich zu teilen.

Interessanterweise besteht, wie kürzlich entdeckt, die Fähigkeit der Herzregeneration noch in der Neonatalperiode[58]. Nach Resektion des Apex (ca. 15% des gesamten Ventrikels) am ersten Tag nach Geburt wird der initial entstandene Wundverschluss (blood clot) von regenerierenden Kardiomyozyten innerhalb der nächsten 21 Tage ersetzt. Form, Anatomie und Funktion des Herzens sind danach identisch mit einem ungeschädigten Herzen. Ausgangspunkt der sich teilenden und proliferierenden Kardiomyozyten waren bereits existierende Kardiomyozyten, die durch eine Myh6-Cre Mauslinie zurückverfolgt werden konnten. Die Proliferation der Kardiomyozyten nach apikaler Resektion ist nicht beschränkt auf den Amputationsbereich, sondern vollzieht sich eher über das gesamte Myokard[56]. Eine Resektion des Apex am siebten Tage nach Geburt endete in Bildung einer fibrotischen Narbe.

In verschiedenen Studien wurde versucht einen Wiedereintritt in den Zellzyklus einzuleiten, so zum Beispiel mit dem Zellzyklus Aktivator cyclin D2 und einer transgenen Expression in Kardiomyozyten. Nach Induktion eines Myokardinfarktes wiesen die transgen veränderten Herzen eine verbesserte Funktion und eine geringere Infarktnarbengröße verglichen mit ungeschädigten Mausherzen auf[59]. Als weiterer Kandidat neben cyclin D2 wurde der Wachstumsfaktor Neuregulin1 (NRG1) zusammen mit dem Thyrosinkinase-Rezeptor EB4 gefunden. Nach Myokardinfarkt und

Substitution von NRG1 per intraperitonealer Injektion konnte ein Wiedereintritt von Kardiomyozyten in den Zellzyklus beobachtet werden, sowie Karyokinese und anschließende Zytokinese. Auch diese Herzen zeigten eine deutliche verbesserte Funktion und eine um ca. 44% verringerte Narbengröße[60]. Der Effekt von NRG1 bleibt jedoch auf mononukleare Kardiomyozyten beschränkt, die einen Anteil von 10% der Kardiomyozyten im adulten Maus Herzen haben. Die Anzahl von Zellen, die durch NRG1 stimuliert werden und in vivo eine Zytokinese einleiten, liegt daher nur bei ca. 0,3% der mononuklearen Zellen[61].

Neben der Erforschung der Stimulation bereits vorhandener adulter Kardiomyozyten liegt ein weiterer Schwerpunkt in der Rolle von Stammzellen, epikardialen Zellen und Fibroblasten als potentielle Quelle neuer Kardiomyozyten. In einer Studie wurden Stammzellen, die den Thyrosinkinaserezeptor c-Kit exprimieren, in infarzierte Herzen injiziert. Die Ergebnisse zeigten, dass die infarzierten Herzen mit c-Kit positiven Zellen, jedoch nicht mit mesenchymalen Stammzellen, eine deutlich verbesserte Ejektionsfraktion hatten und eine Differenzierung von endogenen kardialen Progenitorzellen zu Kardiomyozyten auslösen. Eine Transdifferenzierung der c-KIT positiven Zellen wurde nicht beobachtet. Die transplantierten Zellen scheinen parakrine Effekte zu besitzen und über bisher unbekannte Mediatoren neue Kardiomyozyten entstehen zu lassen[62].

Ein weiterer Aspekt in der Herzregenerationsforschung ist der Einfluss der extrazellulären Matrix (ECM) auf Umbau- und Regenerationsprozesse. Etwa 1/3 aller Zellen im adulten Herz sind Kardiomyozyten, während die restlichen Zellen überwiegend aus Fibroblasten bestehen, die einen Großteil der ECM Proteine bilden. Die primären Komponenten der ventrikulären ECM bestehen aus Kollagen Typ I und zu einem geringen Anteil aus Kollagen Typ III, sowie eine Reihe weiterer Strukturproteine, die abhängig von Schädigung-und Umbauprozessen erheblich variieren. Diese Strukturproteine spielen eine wichtige Rolle in der Regulation inflammatorischer, regenerativer, fibrotischer und angiogenetischer Signalwege. Periostin, Osteopontin, sowie Thrombospondin 1,2 und 4 (TSP-1, TSP-2, TSP-4), SPARC, Tenascin-C und Mitglieder der CCN-Familie sind einige dieser und beeinflussen diverse Signalwege[63].

Die Rolle der 6 Mitglieder der CCN-Familie (CYR61, CTGF, NOV, WISP1, WISP2, WISP3) sind im Herzen allerdings bisher nur zum Teil erforscht. Generell besitzen die

Mitglieder dieser Familie eine wichtige Rolle bei Reparaturvorgängen in Geweben und bei fibrotischen Veränderungen (Abbildung Nr. 4).

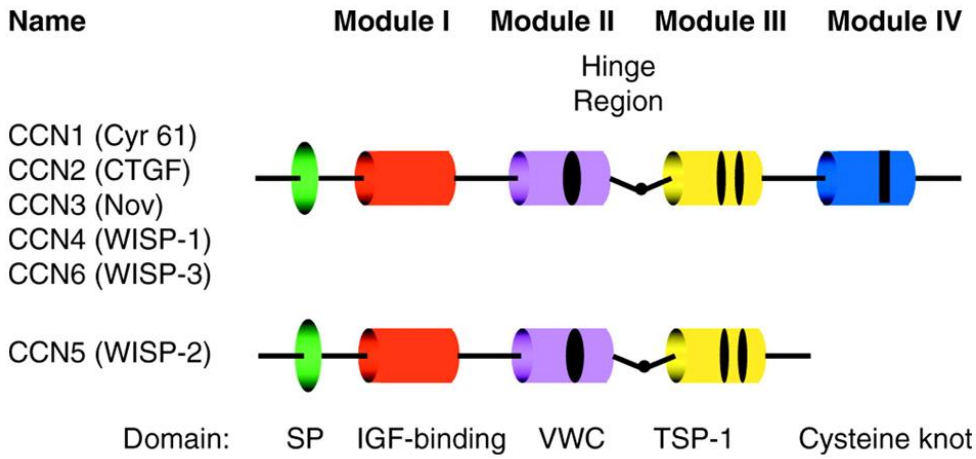


Abbildung Nr. 4: Struktur der CCN-Familienmitglieder

Die CCN-Familienmitglieder besitzen eine gemeinsame Struktur, die aus einem Signalpeptid (SP), einer IGF-Bindungs-Domäne (Module I), einer von Willebrand Typ C Domäne (Module II), einer Thrombospondin-1 Domäne (Module III) und einem Cystein-Knoten (Module IV) bestehen. Entnommen aus [64].

Im Bereich des Herzens vermutet man für CYR61 (CCN1) Effekte auf inflammatorische Leukozyten, Fibroblasten und Kardiomyozyten in Infarkt-und Herzinsuffizienzmodellen[65]. In adulten Herzen konnte in vivo bei mechanischem Stress und unter erhöhten ventrikulären Drücken eine vermehrte Produktion von CCN1 in Kardiomyozyten beobachtet werden[66]. CTGF (CCN2), das zweite Mitglied der CCN- Familie, besitzt eine Rolle im Bereich infarzierter Herzareale, dessen erhöhte Expression durch Angiotensin II und den TGF-β/ Smad3 Signalweg gesteuert wird[67]. Bei in vitro CTGF Überexpressionsmodellen zeigten sich Überlebensvorteile für Kardiomyozyten und profibrotische Ereignisse[68]. CTGF gilt als einer der Schlüsselmediatoren in der Wundheilung[69].

Für NOV (CCN3), WISP2 (CCN5) und WISP3 (CCN6) gibt es nur wenige Erkenntnisse zur Rolle in Homöostase und bei Herzerkrankungen[63]. Die Expression von WISP1 (CCN4) ist bei myokardialen Infarkten erhöht und lokalisiert sich auf die Infarkt-Randbereiche und später auf das remodelierende Myokardgewebe. TNF-α gilt dabei als einer der Mediatoren für die Expression von CCN4 und konnte in vitro zeigen,

dass es Kardiomyozyten zur Hypertrophie stimulieren und kardiale Fibroblasten zur Proliferation bewegen kann[70].

1.3.2. Herzregeneration im Molch

Der Molch, *Notophthalmus viridescens*, ist wie bereits in der Einleitung erwähnt, ein Meister der Regeneration und in der Lage, nach partieller Schädigung sein Herz zu regenerieren[71]. Sein Herz besteht im Gegensatz zur Maus oder dem Menschen nur aus einem Ventrikel und zwei Atrien. Auch die Nährstoffversorgung läuft nicht über Koronararterien, sondern wird über im Ventrikel und Atrien befindliches Blut bewerkstelligt. Trotz einiger Unterschiede auf anatomischer, physiologischer und zellulärer Ebene kommt das Molchherz, verglichen mit dem Herz des Zebrafisches, einem anderen Modellorganismus der Herzregeneration, dem menschlichen Herzen am Nächsten[72].
Erste Untersuchungen zur Regenerationsfähigkeit des Molchherzens wurden bereits 1974 von Oberpriller et al. veröffentlicht. Die von ihm verwendete Schädigungsmethode bestand in einer Entfernung von etwa 1/8 des Ventrikels im Bereich des Apex und anschließender Untersuchung verschiedener Regenerationszeitpunkte bis 30 Tage nach Schädigung. In diesem Zeitraum beobachtete er einen initialen Prozess der Blut-Propf Bildung, gefolgt von einer Koagulations-Nekrose und einer Makrophagen Aktivität, sowie regenerative Aktivitäten des Herzmuskels, die mit einer Bindegewebsformation (Narbe) endeten. Oberpriller et al. konnten mit ihren Studien erstmalig demonstrieren, dass eine Proliferation von Kardiomyozyten im Bereich des Wundareals entsteht[73]. In einer späteren Veröffentlichung konnte gezeigt werden, dass adulte Kardiomyozyten in vitro zur Mitose und anschließenden Zytokinese im Stande sind[74]. Molch-Kardiomyozyten im Ventrikel sind zu etwa 98% mononukleär und diploid im ungeschädigten Gewebe und etwa die Hälfte unterläuft eine Runde der DNA-Synthese während der Regeneration, wobei ca. 90 % mononukleär und diploid nach Abschluss der Regeneration verbleiben.
Auf Basis einer anderen Schädigungsmethodevon Laube et al., bei der mit einer Pinzette eine mechanische Herzquetschung vorgenommen wurde, erkannte man, dass es nach Schädigung zu einer ausgeprägten Reduktion sarkomerer Proteine im Myokard, verbunden mit einer partiellen Dedifferenzierung adulter Kardiomyozyten, kommt. Etwa 14 Tage nach Schädigung fand man keine Reduktion sarkomerer Proteine im

Myokard mehr vor[5]. Mit der Schädigungsmethode einer mehrfachen Quetschung des Ventrikels versucht man durch Zerstörung und Desorganisation des Ventrikels einem Modell eines myokardialen Infarktes beim Menschen am nächsten zu kommen.

Eine weitere Schädigungsmethode, die von einer schwedischen Arbeitsgruppe 2011 publiziert wurde, besteht in der Entfernung eines lateralen Segmentes des Ventrikels. In ihren Untersuchungen konnte diese Arbeitsgruppe aufgrund deutlich verlängerter postoperativer Beobachtungszeiten (70 Tage) feststellen, dass das operative entfernte Ventrikel-Gewebe vollständig mit Kardiomyozyten ersetzt wurde[71]. Eine Narbenbildung, wie von Oberpriller et al. nach 30 Tagen beschrieben wurde, konnte nach 60 Tagen nicht mehr beobachtet werden[73]. Die Erkenntnis einer narbenfreien Regeneration des Molchherzens deckt sich mit kürzlich veröffentlichten Ergebnissen aus unserer Arbeitsgruppe innerhalb eines Untersuchungszeitraumes bis 200 Tage nach Schädigung (Abbildung Nr. 5)[75].

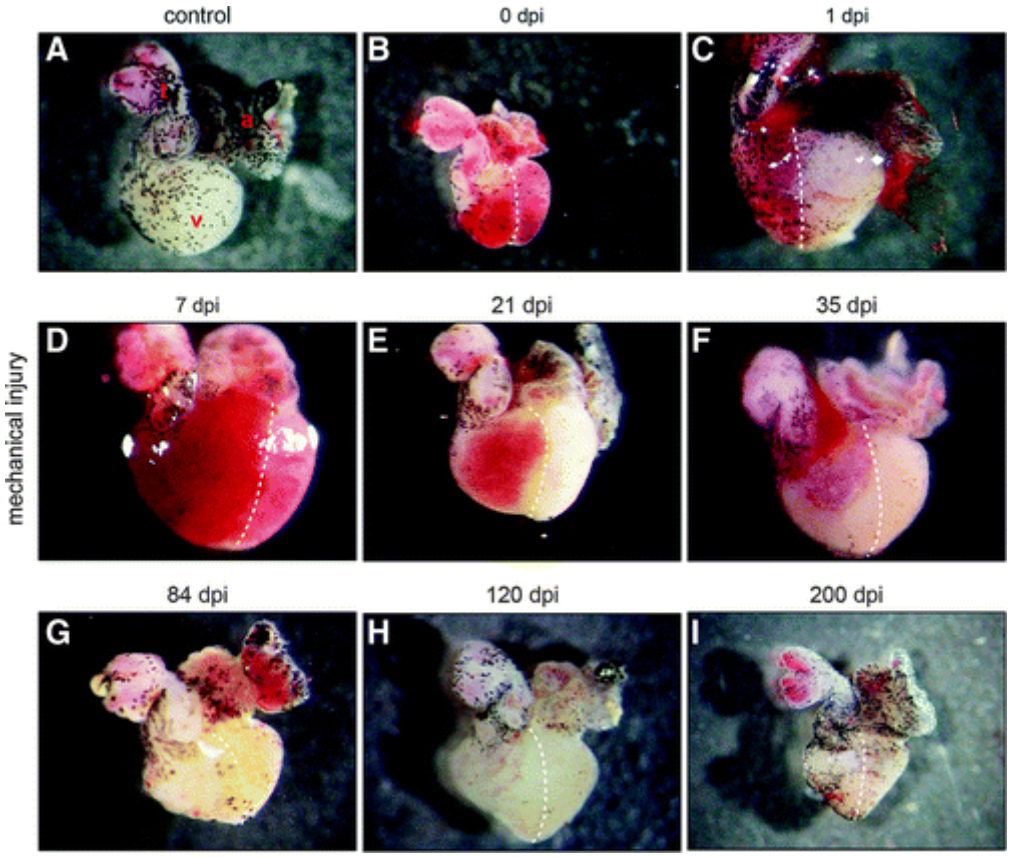


Abbildung Nr. 5: Makroskopischer Überblick Molch-Herzregeneration

(A-I) Vom ungeschädigten zum regenerierenden Herzen bis 200 Tage nach Schädigung; das Molchherz besteht aus einem Ventrikel (v), zwei Atrien (a) und einem Truncus

aorticus (t). 84 Tage nach Schädgung ist kein makroskopischer Unterschied zwischen geschädigtem und ungeschädigtem Herzen mehr erkennbar. Entnommen aus [75].

Neben der Entdeckung einer narbenlosen Regeneration des Molch Herzens wurde bei der Analyse einiger kardialer Transkriptionsfaktoren (NKX2.5, ISLET1, HAND2) durch die schwedische Arbeitsgruppe ein deutlich niedrigeres Expressionslevel eine Woche nach Schädigung festgestellt. Um diesen Zeitraum lassen sich erste Zellproliferationsvorgänge im entfernten Ventrikel Areal erkennen. 23 Tage nach Schädigung jedoch befanden sich die genannten Transkriptionsfaktoren auf ihrer Ausgangsexpression oder darüber hinaus erhöht vor. Insbesondere der kardiale Transkriptionsfaktor GATA4 zeigte nach 23 Tagen eine deutlich erhöhte Expression, was die Autoren dieser Veröffentlichung vermuten ließ, dass es sich bei diesem Transkriptionsfaktor um einen wesentlichen Bestandteil zur Aktivierung von Zellen, die für die Regeneration verantwortlich sind, handelt. Ansätze, die von Forschergruppen momentan intensiv verfolgt werden, ist die Markierung einzelner Zelllinien, um deren einzelnen Beitrag zur Herzregeneration besser verstehen zu können. Die Markierung einzelner Zelllinien ermöglicht es eine genauere Aussage zu treffen, um welche Zellen es sich handelt, die z.B. den Transkriptionsfaktor GATA4 dreiundzwanzig Tage nach Schädigung exprimieren[71].
Aus den Beobachtungen von Oberpriller et al. ist bekannt, dass bereits nach sieben Tagen erste proliferierende Kardiomyozyten zu beobachten sind[74]. Die Prozesse, die zu dieser Proliferation führen, sind bis dato so gut wie unerforscht. Eine erste Studie zu diesen Prozessen innerhalb der ersten Tage nach Schädigung lassen sich erstmals in einer cDNA Microarray Studie von Borchardt et al. im Jahr 2010 finden[76]. In dieser Untersuchung konnte festgestellt werden, dass in den ersten Tagen auf Transkriptomebene eine Reihe von Prozessen, angefangen bei der Wundantwort über die Regulation der Extrazellulär-Matrix, Zellzyklus-Kontrolle, Proliferations-und Differenzierungprozesse stattfinden. Untersuchungen auf Proteomebene sowie Erkenntnisse über potentiell Molch spezifische Moleküle in der frühen Herz Regeneration wurden bislang nicht durchgeführt und sind Gegenstand dieser Arbeit. Aus Veröffentlichungen zur Bein Regeneration des Molches ist die Bedeutung der frühen Wundantwort und Molch-spezifischer Moleküle bekannt, so dass die Vermutung naheliegt, dass dies in der Herz Regeneration ebenfalls der Fall sein könnte[77-79].

2. Veröffentlichungen

2.1. Veröffentlichung Nr. 1 Erstautor:

„Spiked-in pulsed in vivo labeling identifies a new member of the CCN family in regenerating newt hearts"

Mario Looso[1,2], Christian S. Michel[1,2], Anne Konzer[1], Marc Bruckskotten[1], Thilo Borchardt[1], Marcus Krüger[1], Thomas Braun[1,3]

[1.] Max-Planck-Institute for Heart and Lung Research, Ludwigstrasse 43, 61231 Bad Nauheim, Germany

These authors contributed equally to this work

Veröffentlicht in:
J Proteome Res. 2012 Sep 7;11(9):4693-704. doi: 10.1021/pr300521p. Epub 2012 Aug 27, PMID: 22891955.

2.1.1. Inhaltsangabe

In der vorliegenden Veröffentlichung konnte gezeigt werden, dass Molch-spezifische extrazelluläre Matrixproteine und inflammatorische Prozesse zu Beginn der Herzregeneration eine Rolle spielen. Diese Beobachtung wurde durch den Einsatz moderner Hochdurchsatzanalysen ermöglicht.

2.1.2. Eigenanteil

Aufgrund der mangelnden Kenntnisse, welche Moleküle und Signalwege in der frühen Herzregeneration eine Rolle spielen, entwickelte Herr Dr. Looso auf Basis seiner im Jahre 2010 veröffentlichten Methode zur Identifizierung von unbekannten Proteomen (Advanced identification of proteins in uncharacterized proteomes by pulsed in vivo stable isotope labeling-based mass spectrometry.[80]), diese um eine Quantifizierung weiter. Auf Basis dieser Proteom Quantifizierung wurden ungeschädigte Herzen mit Herzen 6 Stunden nach Schädigung verglichen. Bei diesem Herzschädigungsmodell wurden die Molchherzen einer Herzquetschung unterzogen. Zu diesem Zwecke wurden von mir von thorakaler Seite die Brusträume diverser Molche eröffnet, und mit einer stumpfen Pinzette der rechte Teil des Ventrikels 15mal einer Quetschung unter Vermeidung einer Ventrikelperforation unterzogen. Anschließend wurde der äußere

Brustraum mit einem Hautkleber verschlossen. Nach 0 bzw. 6 Stunden wurden die ungeschädigten und geschädigten Herzen entnommen und für die massenspektrometrische Analyse von Frau Konzer vorbereitet. Die anschließende Auswertung wurde von Herrn Dr. Looso durchgeführt und es konnten 1353 Proteine akkurat quantifiziert werden. Dabei waren 72 Proteine raufreguliert und 32 zeigten eine niedrige Regulation. Zur Überprüfung inwieweit die neu entwickelte Methode mit dem Expressionslevel auf Transkriptomebene übereinstimmt, führte ich für 12 deregulierte Kandidaten qRT-PCRs durch und verglich diese mit den Proteom Expressionswerten.
Bereits zuvor und nach Sichtung der Ergebnisse der qRT-PCR Ergebnisse fiel ein Kandidat auf, der ein hohes Expressionslevel und eine leichte Ähnlichkeit zu mehreren Mitgliedern der CCN Familie aufwies. Da die Nukleotidsequenz nicht vollständig war, musste diese mit 5`und 3`Race PCRs verlängert werden. Dadurch konnten alle 4 typischen Domänen der CCN Familie identifiziert werden, was letztendlich auch die Zuordnung zu dieser Proteinfamilie zuließ. In öffentlichen Datenbanken waren bis dato für den Molch nur CTGF, das zweite Mitglied der CCN-Familie, bekannt gewesen. Um weitere Mitglieder der CCN-Familie im Molch zu identifizieren, suchte ich mit Hilfe von degenerierten Primern auf Basis multipler Alignements in hochkonservierten Bereichen zwischen diversen Spezies (*H. sapiens, M. musculus, R. norvegicus, X. tropicalis, D. rerio, B. taurus, G. gallus*) nach weiteren CCN Mitgliedern im Molch. Gefundene PCR-Fragmente wurden anschließend in Vektoren kloniert und per Sanger-Sequenzierung ausgewertet. Für *N. viridescens* konnte ich CYR61, NOV und WISP1 identifizieren. Im Anschluss führte ich einen phylogenetischen Vergleich auf Proteinebene zwischen den Mitgliedern der CCN Familie des *N. viridescens* und 4 weiteren Spezies (*H sapiens, M. musculus, X. tropicalis, D. rerio*) durch. Dieser phylogenetische Vergleich unterstützte in hohem Maße unsere These, dass es sich bei diesem Kandidaten, um ein Molch bzw. Amphibien spezifisches CCN-Mitglied handelt.
Zur genaueren Charakterisierung der Mitglieder der CCN Familie im Molchherzen führte ich qRT-PCRs sowohl im ungeschädigten als auch im geschädigten Herzen durch. Hierbei zeigte sich für den neuen Kandidaten, genannt nsCCN (newt specific CCN), ein deutlich angestiegenes Expressionslevel in den ersten Stunden nach Schädigung. Um zu erfahren von welchen Zellen die Expression von nsCCN ausgeht, fertigte ich in situ-Hybridisierungen an. Dadurch konnte eine fast exklusive Expression von nsCCN in geschädigtem Gewebe festgestellt werden. Weiterhin wurde eine endokardiale Zellpopulation als Quelle der Expression identifiziert. Ein anderes

Mitglied der CCN Familie, CTGF, zeigte erst nach einigen Tagen der Regeneration ein signifikant erhöhtes Expressionslevel.
Statistische Analysen, falls nicht genauer im Manuskript erläutert, wurden mit Graphpad Prism Software 4 durchgeführt. Abbildungen und Figuren im Manuskript wurden von Dr. Looso und mir erstellt. Das Manuskript wurde von Prof. Dr. Dr. Thomas Braun, Dr. Looso und mir geschrieben und zur Publikation eingereicht.

Kurz-Zusammenfassung der durchgeführten Experimente:

- Durchführung von Operationen am Molch und Gewinnung von Gewebematerial
- qRT-PCR Analysen von regulierten Proteom-Kanidaten in ungeschädigtem und geschädigtem Herzen
- 5`und 3´Race PCRs von nsCNN sowie der neu gefundenen CCN-Familienmitglieder
- Multiple Alignements für unbekannte CCN-Familienmitglieder; Auswahl degenerierter Primer mit anschließender PCR in multiplen Geweben (Gehirn, Lunge, Leber, Milz, Herz, Bein, Schwanz, Auge); Klonierung gefundener Fragmente in Vektoren sowie Sanger-Sequenzierung
- Phylogenetischer Vergleich der CCN-Familienmitglieder
- qRT-PCR Analysen und in situ Hybridisierungen der CCN-Familienmitglieder in ungeschädigtem und geschädigten Herzen

Unterschrift des Betreuers bezüglich der Angaben zum Eigenanteil:

Ort, Datum	Unterschrift Prof. Dr. Dr. T. Braun

2.1.3. Originalpublikation

"Spiked-in Pulsed in Vivo Labeling Identifies a New Member of the CCN Family in Regenerating Newt Hearts"

Mario Looso[1,2], Christian S. Michel[1,2], Anne Konzer[1], Marc Bruckskotten[1], Thilo Borchardt[1], Marcus Krüger[1], Thomas Braun[1,3]

[1.] Max-Planck-Institute for Heart and Lung Research, Ludwigstrasse 43, 61231 Bad Nauheim, Germany

These authors contributed equally to this work

Max-Planck-Institute for Heart and Lung Research, Ludwigstr. 43, 61231 Bad Nauheim, Germany

Proteome Res., 2012, *11* (9), pp 4693–4704; DOI: 10.1021/pr300521p
Publication Date (Web): August 14, 2012
*E-mail: thomas.braun@mpi-bn.mpg.de. Phone: 49-6032-705-1101.

2.1.3.1. Abstract

The newt *Notophthalmus viridescens*, which belongs to the family of salamanders (Urodela), owns remarkable regenerative capacities allowing efficient scar-free repair of various organs including the heart. Salamanders can regrow large parts of the myocardium unlike mammals, which cannot replace lost cardiomyocytes efficiently. Unfortunately, very little is known about the molecules and the regulatory circuits facilitating efficient heart regeneration in newts or salamanders. To identify proteins that are involved in heart regeneration, we have developed a pulsed SILAC-based mass spectrometry method based on the detection of paired peptide peaks after $^{13}C_6$-lysine incorporation into proteins in vivo. Proteins were identified by matching mass spectrometry derived peptide sequences to a recently established normalized newt EST library. Our approach enabled us to identify more than 2200 nonredundant proteins in the regenerating newt heart. Because of the pulsed in vivo labeling approach, accurate quantification was achieved for 1353 proteins, of which 72 were up- and 31 down-regulated with a (|log 2 ratio| > 1) during heart regeneration. One deregulated member was identified as a new member of the CCN protein family, showing a wound specific

activation. We reason that the detection of such deregulated newt-specific proteins in regenerating hearts supports the idea of a local evolution of tissue regeneration in salamanders. Our results significantly improve understanding of dynamic changes in the complex protein network that underlies heart regeneration and provides a basis for further mechanistic studies.

Keywords:
heart regeneration; **proteomics**; **mass spectrometry**; **newt**

2.1.3.2. Introduction

Adult tetrapod vertebrates differ significantly in their regenerative capacities. Complete regeneration in adult mammals is limited to few tissues such as the liver,(1) although newborn mammals are able to repair an extended number of organs including the heart.(2) In contrast, various species of salamanders own a remarkable regenerative potential unique among tetrapod vertebrates. Salamanders regenerate even complex structures such as appendages,(3, 4) jaws, ocular tissues,(5, 6) parts of the central nervous system(7) and the heart,(8-10) among others. Remarkably, the regenerative potential of newts does not seem to decline with age or repetition, since newts are able to regenerate extracted lens tissue for at least 18 times over a 16-year time period without noticeable molecular or morphological alterations.(11) The regenerative capacity of salamanders seems to be associated with the ability to generate a blastema composed of undifferentiated progenitors from differentiated cells, although the developmental potential of dedifferentiated cells seems to be lower than initially anticipated.(12) In contrast, the contribution of activated stem cells to regeneration and blastema formation is still under debate.

In the newt heart, regeneration is associated with accumulation of dedifferentiated cardiomyocytes, which proliferate after ventricular amputation or damage.(13-15) Dedifferentiated newt cardiomyocytes show a certain plasticity when transplanted heterotopically into amputated newt limb stumps but not undamaged limbs by contributing to newly forming skeletal muscle tissue.(16) Similarly, the zebrafish has also been demonstrated to regenerate cardiac muscle after partial resection through activation and expansion of cardiomyocyte populations,(17, 18) although it is not known whether zebrafish cardiomyocytes are able to adapt different fates. The ability of damaged salamander hearts for reconstitution ad integrum sharply contrasts to the

response of infarcted mammalian hearts, which primarily react by fibrosis and scarring, resulting in impaired ventricular pump function.

It has been generally assumed that regeneration is an ancestral or primordial property of metazoa that was selectively maintained in salamanders but was lost in most other vertebrates. However, it has also been postulated that regenerative capacity is a unique feature of salamanders, which locally evolved in the salamander branch after separation from the main vertebrate tree.(19) Arguments for the latter hypothesis are mainly derived from the identification of molecules only found in regenerative species emphasizing the idea that a relatively small number of taxon-specific components utilizes a largely conserved cellular machinery, present in both regenerative and nonregenerative taxa, to drive regeneration. Prod1, a salamander-specific protein that mediates nerve dependent signals to the regenerating blastema and contributes to the proximodistal identity in the limb bud is a typical example for such a molecule.(20) Further support for the local evolution hypothesis depends on the identification of additional species-specific proteins in regenerating salamander tissues. Unfortunately, only limited genomic and proteomic information is available for salamanders and newts in particular, which renders a difficulty in such an analysis. At present only very few sequences (less than 200 nonredundant protein sequences) for newt proteins are deposited in public databases supplemented by 9696 ESTs,(21) which has greatly impeded high throughput analysis on the proteome level in the past.

Recently, a newt database was established, which harbors available sequence, annotation and expression data of newts.(22) Incorporation of additional sequence information by next generation sequencing techniques into this database accompanied by improved assembly algorithms will allow identification of hitherto unknown transcripts in salamanders in the near future. Yet, even the availability of nucleotide sequences does not always allow unequivocal identification of corresponding proteins, in particular when no obvious homologues to other species are found (i.e., salamander-specific sequences). The distinction of cloning artifacts and noncoding RNAs from protein-coding mRNAs is difficult, especially if a comparison to candidate mRNAs using a substitution-matrix (i.e., homologous sequence) is not possible. Moreover, several further problems exist that impede the identification of peptides in partially characterized genomes. A potential solution for this problem might be de novo protein sequencing, which continues to represent a challenge (reviewed(23, 24)). In addition, other approaches have been proposed to deal with this issue (reviewed(25, 26)).

Nowadays, the generation and sequencing of EST libraries or next-generation-sequencing projects together with shotgun proteomics seem to be the most economical approach to characterize the proteome of organisms with limited genome information.(27)

Originally, the SILAC technology was developed to quantify proteins in cell culture.(28) Later on, it was used for protein quantification in mice and flies in vivo.(29, 30) In vivo SILAC is one of the most versatile tools to quantify changes within the proteome in living organisms,(29, 31) although other labeling approaches like metabolic N15 labeling(32) or chemical labeling with ICAT(33) or ICPL(34) exist. The main disadvantages of metabolic in vivo approaches are the relatively high costs and that it is time-consuming to achieve full labeling of tissues. Pulsed protein labeling with stable isotopes of amino acids (pSILAC(35)) reduces both the cost and the time but still allows efficient quantification of proteins,(36) characterization of proteomes in vitro,(37) and improved peptide identification in reverse translated EST libraries.(27)

Here, we used a new pulsed SILAC like spike-in approach for high throughput quantitative proteome analysis of the regenerating heart of red spotted newts in vivo. Peptide sequences obtained by mass spectrometry analysis were matched to a normalized cDNA library (generated from regenerating newt heart tissue), since no protein database for newts was available. We identified more than 2200 nonredundant proteins including salamander specific proteins. Detailed sequence analysis of proteins with no apparent homologues in other species led to the identification of nsCCN, a new member of the CCN family of growth factors, which was specifically up-regulated in endocardial cells of injured areas of the heart. The CCN family is a secreted group of proteins, playing a role in intercellular signaling like mitosis, adhesion, apoptosis, extracellular matrix production, growth arrest and migration. In situ hybridizations allowed us to verify a strong continuous expression located in the injured area of the early regenerating heart.

Our quantitative approach enabled us for the first time to monitor changes in protein levels during the early newt heart regeneration including components of the early wound response(38) and the immune system.(38, 39)

2.1.3.3. Results

Identification of Proteins in Regenerating Newt Hearts Using Mass Spectrometry and Mascot Searches in Normalized Newt EST Databases

To obtain information about putative protein coding RNAs in regenerating newt hearts we generated a normalized cDNA library and sequenced 50 000 ESTs. Details about the library can be found and accessed via the newt repository http://newt-omics.mpi-bn.mpg.de. Sequences were annotated by homology searches against multiple databases using the BLAST tools. Mass spectrometry data were generated using protein lysates from undamaged and damaged newts hearts 6 h after injury and utilized for mascot searches in the annotated cDNA library based on principles described elsewhere.(27) We grouped identified proteins into three separated entities starting from the library annotation files. The first group consisted of proteins with high quality annotations (high homology hits to NR nucleotides and NR proteins), the second group consisted of proteins with EST annotations only, and the third group consisted of proteins without any annotation. Proteins within the first group were further assigned to Gene Ontology annotations. Gene Ontology assignments for mouse, human, and zebrafish were used since no Gene Ontology annotation file was available for amphibians. In total, we assigned 2072 identified proteins to functional annotations (Table S1, Supporting Information).

Quantification of Newt Proteins in Regenerating Newt Hearts by Spiked-in Pulsed in Vivo Labeling

To assess turnover rates of identified proteins, we measured incorporation of heavy $^{13}C_6$-lysine after feeding the newts with a $^{13}C_6$-lysine-labeled tissue mouse diet over time. Regeneration of newt tissues by heart injury, limb amputation, or tail amputation was initiated after an average tissue labeling of 50% was reached. Regenerating tissues were collected from different time points and tissues and pooled to generate a labeled protein mix for the spike-in approach (R_P, see Figure 1) (metabolic labeled regeneration mix(40)). The complex labeled protein mix was measured by LC–MS to determine the specific labeling rate for each single protein n, defined as ratio pool of protein n (R_{Pn}). Measured labeling ratios were calculated as % $^{13}C_6$-lysine incorporation rate (% label = (labeling-ratio × 100)/(labeling-ratio + 1)). In total, we identified 2074 proteins within the heavy pool based on our normalized EST library.

Next, we used the heavy pool as a spike-in label taking advantage of the defined H/L ratios for each single protein n (R_{Pn}). To identify proteins that were deregulated during

early heart regeneration, we combined the metabolic labeled regeneration mix with lysates from nonregenerating hearts (R_0) and with lysates derived from regenerating hearts 6 h after injury (R_{6h}, Figure 1).

Combined lysates were measured, and heavy to light ratios were determined. Resulting H/L peptide pairs (and ratios) were used for both improved peptide identification(27) and for protein quantification based on the pulsed in vivo labeling spike-in approach. For the spike-in approach, we used a refined and modified normalization method first applied to N15 labeled mouse neurons.(37) Peptide peak intensities of mixed measured pools were normalized to ratios of individual pools of lysates derived from regenerating labeled tissues R_P (Figure 2).

Since the labeling rate was dependent on the turnover of each single protein we normalized each protein by its own "Spike-in" ratio.

The protein specific ratios of the heavy pool R_P was defined as $\text{Ratio } R_P = \frac{\text{Heavy}_P}{\text{Light}_P}$

The protein specific ratio of undamaged heart (R_0) was defined as $\text{Ratio } R_0 = \frac{\text{Heavy}_P}{\text{Light}_P + \text{Light}_0}$

The protein specific ratio of regenerating heart 6 h after injury (R_{6h}) was defined as

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

Ratios were inverted (indicated with *) and normalized (indicated with N)

$$\text{Ratio } R_{N0}{}^* = \frac{\text{Light}_0}{\text{Light}_P + \text{Heavy}_P}$$

By solving the equation to light 0 and inserting it

$$R_0{}^* - R_P{}^* = \frac{\text{Light}_0 + \text{Light}_P}{\text{Heavy}_P} - \frac{\text{Light}_P}{\text{Heavy}_P} = \frac{\text{Light}_0}{\text{Heavy}_P} \text{Light}_0 = (R_0{}^* - R_P{}^*) \times \text{Heavy}_P$$

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

We received

The same calculation was done for R_{N6h}*. The resulting two ratios R_{N0}* and R_{N6h}* were used for relative quantifications:

$$R_{exp}{}^* = \log_2\left(\frac{\frac{(R_{0h}{}^* - R_P{}^*)}{R_P{}^* + 1}}{\frac{(R_0{}^* - R_P{}^*)}{R_P{}^* + 1}}\right) = \log_2\left(\frac{R_{6h}{}^* - R_P{}^*}{R_0{}^* - R_P{}^*}\right)$$

$$f(R_{exp}{}^*) = \begin{cases} \text{upregulated}, & R_{exp}{}^* > 0 \\ \text{downregulated}, & R_{exp}{}^* < 0 \end{cases}$$

For localization and spread, see Figure S1 (Supporting Information). Normalization was performed for all proteins, which were identified within the heavy pool R_P, and at both

time points if the H/L ratio variability was smaller than 50%. Finally, we normalized all ratios by the mean of all quantified proteins (factor 0.29). Our quantification approach yielded a protein list of 1353 proteins, out of which 103 were regulated during regeneration by a (|log 2 ratio| > 1). 72 proteins were up-regulated and 31 proteins were down-regulated (Table S2, Supporting Information). Inspection of exemplary peptides of deregulated proteins corroborated regulatory changes of corresponding proteins (Figure S2, Supporting Information). To analyze whether changes of protein concentrations went along with changes of the corresponding mRNAs, we performed qRT-PCR analysis for 6 candidates selected from the up-regulated and down-regulated protein group, respectively. All tested candidates showed changed mRNA levels that corresponded to changes of respective protein levels (Figure 3).

Pulsed labeled based quantification of protein concentrations in complex samples is more precise than label-free quantification if completely labeled proteins are used as internal standards (as demonstrated for example in Lys8 ($^{13}C_6$,$^{15}N_2$) labeled flies(30)). To evaluate whether our spike-in approach provides similar advantages, we compared data obtained by the spike-in approach to results from an additional label free protein quantification experiment. The pulsed labeled based approach yielded an almost 50% decrease of the deviation of the ratio of two measurements compared to intensity based, label free quantifications of uninjured and regenerating newt heart tissues (6 h after damage, Figure 4).

Although the increase in precision was even better when completely labeled flies were used(30) we concluded that the spike-in approach, which is far less cumbersome and cost-intensive in amphibians compared to full labeling, offers distinctive advantages compared to label-free quantification strategies.

Regenerating Newt Hearts Contain Several New Urodele-Specific Proteins Including a New Member of CCN Protein Family

To identify new proteins, which might be specific for newts and/or processes unique for newts such as regeneration of complex tissues, we analyzed potential proteins, which were either annotated to uncharacterized EST clones or possessed no annotations at all. We detected 86 protein sequences, which seem to exist exclusively in newts or in closely related amphibians. A search for conserved sequence motifs or domains in the Pfam database(41) revealed 17 highly conserved structures (Table S3, Supporting Information) including 9 protein domains, 5 Pfam B domains, and 3 protein families.

Next, we selected a DAN domain containing protein for a more detailed phylogenetic evaluation and expression analysis. We found that the DAN domain was located in a cDNA clone of 989 bp, which also contained a second conserved domain, TSP-1. 5′ RACE PCR experiments using RNA from regenerating newt hearts helped to extend the cDNA sequence to 1854 bp containing a potential open reading frame flanked by a 5′ and a 3′ untranslated region (UTR) including a poly A tail. In addition to the DAN and TSP-1 domains, the open reading frame now contained a signal peptide and an IGF (insulin like growth factor) binding domain as well as a VWC (von Willebrand factor type C) domain (Figure S3, Supporting Information). The combination of these domains is a hallmark of growth factors of the CCN family, which so far comprises 6 members (CCN1 to CCN6) in different species (reviewed(42)). Although the domain structure of the new sequence matched all definitions of a CCN family member, we only found comparatively weak homologies to mammalian members of the CCN family outside of the conserved domains (Figure 5). Obviously, this complicated a comparison of the new sequence to other CCN family members. We therefore attempted to identify all existing members of the CCN family in newts. Database searches of sequences derived from our normalized cDNA libraries (Plate151_O3, Contig22962, ref|NP_001026734.1| cysteine-rich, angiogenic inducer, 61 [*Gallus gallus*], *E*-value 7×10^{-080}) and (Plate095_E11, Contig 8784, Q98TQ8 Connective tissue growth [*Gallus gallus*], *E*-value 4×10^{-075}) readily identified direct homologues of CCN1 (= Cyr61) and CCN2 (= CTGF) in newts. Next, we designed degenerated primers based on conserved sequence motifs identified by multiple sequence alignments of CCN family members from various taxa. We were able to amplify sequences from newts that showed clear homology to mammalian CCN3 (= Nov) and CCN4 (= Wisp1) (Figure 5). We did not amplify sequences in newts with significant homologies to CCN5 and CCN6 despite considerable efforts using RNA from different tissues, physiological conditions, and various development stages. These results suggest that CCN5 and CCN6 might not exist in newts, although a definitive answer awaits full sequencing and annotation of the newt genome. Instead, a new CCN family member with strong similarity within the conserved domains seems to have evolved in newts, which we named nsCCN (newt-specific CCN). Taken together, we identified 5 members of the CCN family of growth factors in newts, of which four are direct homologues to CCN1–CCN4, while the fifth represents a new, newt-specific family member (Figure 5).

nsCCN Is Specifically Induced in Endocardial Cells Located in Regenerating Areas of Damaged Hearts

The identification of a newt-specific member of the CCN family of growth factors in regenerating hearts on proteome level suggested that nsCCN might be involved in the regulation of heart regeneration. Therefore, we determined the expression pattern of nsCCN and other members of the CCN family in intact and regenerating hearts. We found a robust expression of CCN1 (= Cyr61) and CCN2 (= CTGF) in smooth muscle cells and cardiomyocytes of undamaged newt hearts by in situ hybridization (Figure 6A), which was further confirmed by qRT-PCR analysis (Figure 6B). No expression was found for CCN3 (= Nov) and CCN4 (= Wisp1) in intact (Figure 6B) and early regenerating hearts. qRT-PCR analysis uncovered a low-level expression of the new family member nsCCN in undamaged hearts, which was not detectable by in situ hybridization (Figure 6A,B). This situation changed dramatically after mechanical injury in regenerating hearts. We observed a strong up-regulation of nsCCN as early as 2 h after injury in endocardial cells located in damaged but not in intact areas of the heart, while cardiomyocytes and other cell types did not seem to express nsCCN at high levels (Figure 6C). Further analysis of the time course of nsCCN expression during regeneration by qRT-PCR revealed a persisting expression of nsCCN throughout the first 24 h after injury with a peak between six and 12 hours (Figure 6D). We also detected a moderate up-regulation of CCN1 (= Cyr61) during regeneration with an early expression peak at 2 h that returned to basal levels within 12 h. CCN2 (= CTGF) expression was not elevated during the first 24 h of heart regeneration (Figure 6D) but was strongly up-regulated 7 d after injury, while CCN4 (= Wisp1) was only slightly up-regulated 14 d after injury (see Figure S4, Supporting Information).

Functional Annotation of the Proteome of Regenerating Newt Hearts

To obtain a detailed view about the biological processes that are active and dynamically regulated during newt heart regeneration, we assigned all detected proteins to Gene Ontology annotations. Since no Gene Ontology annotation file is available for amphibians, we used a stepwise Gene Ontology assignment approach that included a number of different organisms (annotation via blast tools). First, we selected all EST clones in our newt EST library for which we had identified a peptide by mass spectrometry. Next, we searched for homologies of these protein coding candidates in the Uniprot database and sorted all significant hits (E-value < 10^{-20}) by source

organisms. For each organism, we selected only the most significant hit. Hits for mouse, human and zebrafish proteins were checked for available GO annotations (Uniprot GO annotation file). This approach enabled us to annotate 1826 protein groups to human GO terms, 1832 protein groups to zebrafish GO terms, and 2014 protein groups to mouse GO terms (Table S1, Supporting Information).

To investigate which biological process is most prominent in regenerating heart tissue, we searched for over-represented Gene Ontology (GO) terms. Restriction of the analysis to up-regulated proteins yielded a GO enrichment tree for the biological processes ontology (Figure S5, Supporting Information). Among others, we identified enriched terms such as protein activation cascade, complement activation, defense response, immune effector process and activation of immune response in this tree. Assigned proteins within these terms included A2M, F10, CFB, C4A,C2, PRDX2, FCN3 and a specific isoform of ANXA1. A2M is a nonspecific protease inhibitor involved in host defense mechanisms.(43) F10 is a trypsin-like serine protease that plays a key role in the blood coagulation cascade. CFB, C2 and C4A are members of the complement cascade (involved in the alternative pathway), PRDX2 has an antiapoptosis function(44) and is enriched within heart tissues (BioGPS(45)). FCN3 is an activator of the complement system and ANXA1 is a well-known anti-inflammatory protein. Further analysis of the ontology cellular component revealed an enrichment of the term extracellular region (1.96×10^{-4}), further supporting the idea of multiple changes in the extracellular matrix (where all CCN family members are located).

Finally, GO enrichment analysis of down-regulated proteins yielded only a single significantly enriched term, named lipid binding. Five protein assignments were found in this term including the annexin family members ANXA1 and ANXA4, HADHB, FABP3, and PLA2G4C. Initially, we were surprised to identify ANXA1 in the group of down-regulated proteins since it was already detected in the group of up-regulated proteins. Closer inspection of this conflicting set of data revealed that we identified two very similar but not identical proteins both with high significant hits to annexin A1 [*Xenopus laevis*, *E*-value up-regulated protein: 7×10^{-56}; *E*-value down-regulated protein: 4×10^{-107}] that were inversely regulated (Figure S6, Supporting Information). Clearly, further studies are needed to investigate whether these two isoforms are indeed inversely regulated and serve different functions. The HASHB protein is located in the mitochondria and was linked to mitochondrial trifunctional protein deficiency disease.(46) The FABP3 protein is a muscle specific protein, highly expressed in the

heart muscle (BioGPS(45)) and known as marker for acute myocardial infarction.(47, 48) Recently, it was also demonstrated that FABP3 mRNA is down-regulated after birth during growth of the pectoralis muscle,(49) which might represent a parallel to the growth of cardiac muscle after injury. Finally the PLA2G4C protein is a phospholipase not described to be involved in regeneration processes so far.

2.1.3.4. Discussion

Identification of proteins involved in heart regeneration is instrumental to gain insight into the regulatory circuits that facilitate efficient heart regeneration in newts and salamanders. Moreover, it is vital to identify putative species-specific proteins in salamanders to support (or falsify) the hypothesis of a local evolution in salamanders, which provides enhanced regenerative capacity. Conversely, the absence of salamander-specific proteins might argue for a general loss of regenerative potential in most tetrapod vertebrates during evolution with the exception of salamanders. The lack of comprehensive genomic and proteomic data has delayed our understanding of regenerative processes in newts, although generation of complex EST libraries,(21)next-generation sequencing approaches, establishment of dedicated newt databases,(22) and several screening approaches(50) have been helpful to identify molecules involved in regeneration. Yet, it is relatively difficult to identify known and new species-specific proteins solely based on nucleotide-sequences due to the presence of noncoding RNAs, cloning artifacts and sequencing errors. Recently, we developed a new method to analyze the proteome organisms for which virtually no proteome sequence information is available.(27) Here, we demonstrate that a variation of that method based on pulsed in vivo labeling allows unprecedented insights into early regenerative processes in the heart of newts including quantitative changes of proteins during the early course of regeneration.

Advantages and Limitations of Pulsed in Vivo Labeling for Quantitative Proteomics in the Regenerating Newt Heart

The pulsed in vivo labeling method combined with bioinformatical analysis of complex EST libraries resulted in identification of more than 2200 and quantification of 1353 newt proteins, respectively. Our method is not as exact as quantification methods based on complete tissue labeling(29, 30) but far less cumbersome and expensive though applicable for a large range of experimental conditions. Compared to label free

quantification, the pulsed in vivo labeling approach demonstrated superior accuracy, which improved resolution of analysis. Initially, a significant amount of proteins resisted our quantification attempts, although we quantified expression changes of more than 1300 proteins during regeneration. This limitation was due to the origin of the labeled spike-in pool, which initially consisted primarily of nonregenerating tissues. Of course, proteins, which are completely absent in nonregenerating tissues, cannot be labeled and hence cannot serve as a standard during quantification. Therefore, we detected only proteins in regenerating tissues that were already present in intact tissues to a certain degree. To overcome this limitation, we included labeled tissue from several different regenerative conditions and stages. Although this maneuver increased the number of quantifiable proteins, a significant number still resisted our quantification attempts. We reason that this is due to the high complexity of the spike-in mix and the relatively low concentration of some of the proteins from regenerative stages, which will be even further diluted when mixed with extracts from nonregenerating tissues. In principle, this restriction might be overcome by preparing separate spike-in mixes from different regenerative stages and by combining each of these mixes separately with extracts from regenerating tissues. Yet, such an approach would increase measuring/instrument time significantly and would also make data analysis more complicated.

Detection and Quantification of a New Urodele-Specific Protein Up-Regulated during Heart Regeneration by Pulsed Labeling

One of the primary rationales for our proteomics analysis was the identification and quantification of proteins that may play a crucial role in heart regeneration of salamanders. Comparison of newt EST clones with a positive peptide identification to all published nucleotide sequences yielded multiple up- or down-regulated newt proteins with homologues in other vertebrate species. In addition, a group of proteins was detected, which showed no significant homology to other organisms. In an attempt to verify the relevance of known and newly detected proteins for regeneration, we determined their spatial and temporal expression pattern after mechanical heart injury.

One of the detected molecules with a dynamic expression pattern and a lack of an obvious homologue contained IGFBP, VWC, TSP1 and CTCK domains, which led us to the conclusion that we discovered a new, species-specific member of the CCN family, which consists of 6 members in mammals and is highly conserved. Two

mammalian CCN members [CCN1 = secreted cysteine-rich protein cyr61 (Score = 126 bits (316), Expect = 6×10^{-27}) and CCN4 = WNT1 inducible signaling pathway protein 1 (Score = 112 bits (279), Expect = 1×10^{-23})] showed a limited similarity to our new protein. CCN molecules are secreted matricellular proteins that interact with integrins and proteoglycans thereby influencing different processes such as cell differentiation, proliferation, adhesion, migration, apoptosis, ECM production, angiogenesis, and fibrogenesis.(51) CCN1, i.e., binds to IGF, several integrines (αvβ3, αvβ5, α6β1, αIIbβ3, αMβ2), and Heparan Sulfate ProteoGlycans (HSPG)(52) affecting Wnt signaling, NF-kB signaling, tyrosin kinase signaling and Akt signaling.(53) During ischemia CCN1 interacts with the beta 1 integrin-Akt and extracellular signal-regulated kinase (ERK) pathways preventing apoptosis of cardiomyocytes.(54) High throughput studies in mammals indicate that CCN1 and CCN2 are highly expressed in smooth muscle and cardiomyocytes.(55)

CCN4 (WISP1), another member of the family, has also been demonstrated to play a role in mammal heart regeneration.(56) In total we detected five members of the CCN family in newts including the new molecule nsCCN and direct homologues of CCN1 to CCN4, while CCN5 and CCN6 do not seem to exist in newts. Three CCN family members (nsCCN, CCN1, and CCN2) were expressed in the intact or regenerating heart. In particular, the newly discovered urodele-specific nsCCN showed an intriguing expression pattern: it was specifically up-regulated in endocardial cells located in damaged by not intact areas of the regenerating heart. Since the endocardium plays a crucial role in the regulation of regenerative cardiogenesis,(57) it is tempting to speculate that nsCCN is an important component of the regulatory machinery that drives this process. Further functional studies will determine the role of nsCCN in species-specific regenerative responses.

Pulsed Labeling Based Proteomic Analysis of Hearts Regeneration Reveals Differences of Wound-Healing Responses between Salamanders and Mammals

In mammals, wound healing of major injuries often leads to scarring, while salamanders are able to completely restore tissue functions. Hence, it seems reasonable to assume significant differences in wound healing between both taxa. In fact, analysis of the distribution of GO terms of regulated proteins in regenerating hearts of newts revealed a specific mode of wound response that differed substantially from responses of damaged mammalian hearts, although some processes such as activation of the complement

system were similar. For example, we observed an increase of components and activators of the complement system in newt hearts including CFB, C2 and C4A, which are members of the complement cascade, and FCN3, an activator of the complement cascade. Most likely, myocardial cell necrosis results in release of molecules, which activate the complement system.(58) Activation of the complement system will contribute to recruitment of mononuclear cell, which in turn might damage surviving myocardial cells and promote adverse remodeling processes upon extended or exaggerated accumulation in damaged areas. In this context, it was interesting to note that several anti-inflammatory and antiapoptotic proteins, which were up-regulated in regenerating newt hearts, and other proteins regulatingapoptosis and inflammation showed an inverse expression pattern compared to damaged hearts of mammals. We detected that PRDX2, a protein linked to inhibition of apoptosis and ANXA1, a protein connected to suppression of inflammatory processes via regulation of intracellular Ca2+ signaling were strongly up-regulated. A different isoform of ANXA1 was down-regulated in the early regenerating heart, which might indicate opposing functions of both proteins. Interestingly, another member of the annexin family (ANXA4) was also down-regulated during early newt heart regeneration. We also found down-regulation of pro-apoptotic proteins such as BCL2L13(59) suggesting an enhanced propensity of newt hearts to inhibit apoptosis and to restrict inflammatory responses. A further hint for the restriction of inflammatory processes was provided by the down-regulation of fatty acid binding protein 3 (FABP3), which in mammals positively correlates with increased accumulation of TNFα and sFas during myocardial damage(60) and serves as a marker for acute myocardial infarction(47, 48, 61) and cell membrane damage.(60) Further insights into differences of early wound responses between newts and mammals came from the expression pattern of the arginine-rich, ER stress response protein ARMET. This protein, which is secreted after myocardial infarction in mice, has been postulated to protect cardiomyocytes from ischemia/reperfusion injury.(62) Surprisingly, ARMET was slightly down-regulated (−0.85) during early newt heart regeneration, which again might indicate a mitigated stress response.

Taken together, our quantitative in vivo analysis of regenerating newt heart yielded new insights into the molecular machinery that allows complete functional reconstitution of the myocardium in newts. We obtained evidence for species-specific molecules (such as nsCCN) and regulatory processes (such as anti-inflammatory and antiapoptotic signaling), which support the hypothesis of a local evolution of regenerative responses

in newts. A more detailed comparison to processes that are active in damaged mammalian hearts might uncover additional pathways and molecules, which make tissue regeneration in both taxa fundamentally different. Further functional studies primarily focusing on putative species–species proteins will help to pin down the key players that control regeneration, although such studies are still a challenge in newts.

2.1.3.5. Materials and methods

Normalized cDNA Library

The cDNA library was generated from newt hearts using 13 different conditions (undamaged, sham operated, 30 min, 2 h, 6 h, 12 h, 1 d, 2 d, 4 d, 7 d, 14 d, 21 d, and 35 d after injury). For each time-point total RNA of eight hearts was isolated and reverse transcripted to cDNA using the SMART method. Normalization was done applying the DSN method at Evrogen, Moskau (http://www.evrogen.com/). Details about the library were published recently;(22) sequences, annotations and corresponding peptides are publicly available at the newt repository, http://newt-omics.mpi-bn.mpg.de. Sequences for CYR61, NOV, WISP1 and nsCCN were transmitted to NCBI and can be found by Genbank accession numbers: CYR61- JX094346, NOV- JX094347, WISP1- JX094348, nsCCN- JX094349.

Animal Treatment

3–4 years old adult newts were obtained from Charles Sullivan, Inc., Newt Farm, Nashville, Tennessee. Animals were kept at 20 °C in aerated single aquaria and habituated to mouse tissue diet by manual feeding in 3 day intervals for 4 weeks. Animals, which were fed with the mouse diet (chopped liver tissue), did not show any discernible clinical effects compared to animals on a regular tubifex diet during this period. No signs of wasting or weight loss were seen during the feeding period. To incorporate $^{13}C_6$-lysine into newt tissues, tissue from mice were used that had been labeled with a customized mouse diet containing $^{13}C_6$-lysine.(29) Newts were anesthetized with 0.1% Ethyl 3-aminobenzoate, methanesulfonic acid solution (Sigma) when the labeling rate achieved a minimum of 50% and subjected to regeneration by tail tip amputation (1 cm), limb amputation, or heart damage.(16) After surgical intervention newts were incubated for 2 h in 0.5% Sulfamerazine solution (Sigma) to prevent infections. Regenerating tissues were isolated at different time points (heart

undamaged, heart 1 day regeneration, heart 28 days regeneration, tail undamaged, tail blastema) after deep anesthetization and decapitation and immediately processed for further analysis. Undamaged hearts (0 h) and immediate early regenerating hearts (6 h) were collected from adult newts maintained on regular tubifex diet (5 hearts per group). Concentrations of protein lysates from different tissues were determined using the Bradford assay (Biorad, Cat. Nr. 5000113, 5000114) before mixing lysates from labeled and unlabeled tissues a 1:1 ratio.

Sample Preparation, HPLC, and Mass Spectrometry

Heart tissue was homogenized in SDS lysis buffer (4% SDS in 100 mM Tris/HCl pH 7.6) using an Ultra-Turrax (Ika). For complete cell lysis, samples were heated shortly at 95 °C prior to DNA shearing by sonication. After lysates were clearified by centrifugation at 16000*g* for 5 min, protein concentration was estimated using the DC protein assay (Biorad). Next, samples were loaded on a filter unit (MW cut off 30 kDa) according to the filter aided sample preparation (FASP) protocol.(63) In brief, the proteins were reduced with 0.1 M dithiothreitol (DTT), alkylated with 0.05 M iodoacetamide and digested with the endopeptidase Lys-C (Wako). To reduce sample complexity, peptides were separated by isoelectric focusing using an OFF-gel fractionator (Agilent). Finally, peptides were desalted and purified by stop and go extraction (STAGE) tips.(64)

Reverse-phased chromatography was performed using an Agilent 1200 nanoflow LC system (Agilent Technologies). The LC system was coupled to a LTQ Orbitrap XL instrument (Thermo Fisher Scientific) equipped with a nanoelectrospray source (Proxeon). Chromatographic separation of peptides was achieved with custom-made columns, filled with reverse-phase ReproSil-Pur C18-AQ 3 μm resin (Dr. Maisch GmbH). LysC-digested peptide mixtures were autosampled at a flow rate of 0.5 μL/min and then eluted with a linear gradient at a flow rate of 0.2 μL/min. The mass spectrometer was operated in the data-dependent mode to automatically measure MS and MS^2. LTQ-FT full scan MS spectra (from m/z 350 to m/z 1750) were acquired with a resolution of R = 60 000 at m/z 400. The five most intense ions were sequentially isolated and fragmented in the linear ion trap both by using collision-induced dissociation (CID). In addition peptides were also fragmented with the higher energy collision dissociation (HCD) as described.(65)

Analysis of LC–MS/MS Data

Raw data files were converted to Mascot generic format files by MaxQuant.(66) The Mascot search engine (version 2.2.02) was used for database searches and protein identification using the following search parameters: LysC digestion, 2 missed cleavages. Carbamidomethylation of cysteine was set as fixed modification and oxidation of methionines was selected as variable modification. The maximum allowed mass deviation for MS and MS^2 scans was 20 ppm and 0.5 Da, respectively. For peptide identification, we searched for peptide assignments on our in-house generated EST database from regenerating newt hearts. The database included 48537 N.v. ESTs, reverse translated in 3 reading frames. All databases were generated as DECOY target databases.(67) A minimum peptide length of 6 amino acids and two peptides per protein group including one unique peptide (CID method, see above) was necessary to define a positive hit (Table S4, Supporting Information). False discovery rates were based on reverse sequence matches in the combined DECOY target databases. Our maximum false discovery rate was set below 1% for peptide and protein identifications. RAW files are available from the proteome commons repository or on request from the authors. The files can be downloaded from https://proteomecommons.org/dataset.jsp?id=nSFtHPGpn%2BqPYPRcc3%2FNGKbHFpPoBh5m8fLzbkRXChOdEyGUoLguR0CTQA6F7wF%2FCZ0Z7jdO89t2H2hDjsxz%2FNzINpoAAAAAAAAB1g%3D%3D, and the hash code is nSFtHPGpn+qPYPRcc3/NGKbHFpPoBh5m8fLzbkRXChOdEyGUoLguR0CTQA6F7wF/CZ0Z7jdO89t2H2hDjsxz/NzINpoAAAAAAAAB1g==.

Analysis of Protein Ratios

Incorporation of $^{13}C_6$-lysine into proteins at different time points of regeneration was calculated by the ratio of H/L (heavy/light) peptide peaks using the MaxQuant software tool.(66) The average labeling rate was calculated using all detected peptides with a H/L ratio. To generate the complete identification list, all detected masses from three measurements (heavy pool, undamaged and 6 h regenerating) were used for protein/peptide identification in the cross database search approach. Only proteins that were identified in all three samples were used for quantification purposes.

Protein Classification with Gene Ontology

Newt EST sequences with peptide identifications were assembled on the nucleotide level. We performed BLAST searches against Uniprot databases(68) (Swissprot and Trembl) and collected Gene Ontology identifiers from high quality homologue hits (*E*-value threshold < 10^{-20}) since virtually no amphibian Gene Ontology annotation files were available (only a small *Xenopus* Gene Ontology annotation file exists). To avoid an overlap of different protein homology assignments, we only took the best-rated homology hit with GO annotations from either mouse, human, zebrafish, chicken or *Xenopus*. For further analysis GO annotations from each group of organisms were processed individually.

To evaluate molecular processes that are active during heart regeneration, we searched for overrepresented GO terms by calculating *p*-values for each organism in overrepresentation statistics (GOrilla(69)). Enriched GO terms were further analyzed in the group of proteins that showed dynamic regulation during regeneration. The sum of all detected proteins served as a baseline control.

qRT-PCR Transcript Verification

Total RNA of uninjured, 2, 6, 12, and 24 h regenerating newt heart tissues was isolated using TRIzol reagent (Invitrogen) according to instruction of the manufacturer. One microgram of total RNA was used for reverse transcription with SuperscriptII (Invitrogen) following standard procedures. Real-time PCR was performed using the iCycler (Bio-Rad) and ABsolute QPCR SYBR Green Fluorescein Mix (ABgene, U.K.). A list of the primers is supplied (Table S5, Supporting Information).

In Situ Hybridization

Hearts were damaged by squeezing the left halve of the ventricle with small forceps. Uninjured and regenerating hearts (2 and 6 h after injury) were processed using standard in situ hybridization protocols. Heart tissues were fixed in 4% paraformaldehyde, embedded in paraffin wax, and sectioned at 10 μm. In situ hybridization was done by hybridization of a digoxigenin-labeled RNA probe. Bound probes were detected by anti-DIG antibodies (Roche Diagnostics, Germany) followed by BM-purple staining. Images were taken using a Zeiss Axioplan2 microscope equipped with a ×5/0.15 and a ×40/0.75 Plan-Fluar objective, an Axiocam and the Axiovision 4.6 software (all from Carl Zeiss MicroImaging, Inc.).

Statistical Analysis

Not otherwise specified statistical analyses were performed using GraphPad Prism software version 4. Statistical significance between two groups was determined by a two-paired, untailed Student's *t*-test (*, $p < 0,05$; **, $p < 0,01$).

Supporting Information

Supporting Figures S1–S6 and Tables S1–S5. **Table S1**. Gene Ontology assignments. The file includes all identifier mappings based on strict sequence similarity searches that were used for Gene Ontology enrichment analysis. It assigns Uniprot protein accession numbers for three different taxa to newt sequence IDs. Column leading identifier for each protein group represents IDs for all identified proteins. Yellow columns represent a subgroup for which a corresponding zebrafish orthologue was found. Blue columns represent a subgroup for which a human orthologue was found. Red columns represent a subgroup for which a mouse orthologue was found. Magenta columns sum up all newt IDs with an assigned orthologue independent of taxa. **Table S2**. List of regulated proteins, calculated by pulsed labeling. The file describes all quantified proteins with protein annotations and calculated expression values. Column ID gives the protein ID, column peptides gives the number of identified peptides, column unique peptides gives the subgroup of unique mapping peptides, column sequence coverage describes sequences covered by peptides, column molecular weight gives the total mass in kDa, ratio pool describes heavy pool ratios that were used for the spike in approach, ratio Mix 0 h shows measured ratios for the undamaged heart mixed with the heavy pool, ratio mix 6 h shows measured ratios for the damaged heart combined with the heavy pool, annotation gives the best similarity search hit, R norm mix 0 is the calculated ratio for the undamaged light heart (as explained in the manuscript), R norm mix 6 h is the corresponding ratio for the regenerating light heart, log2 Rnorm Mix6/r Norm Mix0 is the direct ratio that describes the fold change, and column mean normalized is the final normalized ratio with mean expression 0. **Table S3**. Selection of conserved motifs within coding newt ESTs. The file lists identified and quantified protein sequences that were detected to harbor known sequence domains/family pattern/Pfam B motifs. It includes sequence identifiers, name and type of conserved sequence motifs and the corresponding significance. **Table S4**. List of all proteins detected in database searches. The file lists all proteins/protein groups that were identified (only a subgroup was quantified). Column names are in accordance to Table S2. **Table S5**. List of primers

used for qRT-PCR analysis. The file lists all PCR primers that were used for qRT-PCR analysis. This material is available free of charge via the Internet at http://pubs.acs.org.
The authors declare no competing financial interest.

Acknowledgment

This work was supported by the Max-Planck-Society, the DFG (SFB/TR81 and Br1416), the Excellence Initiative "Cardiopulmonary System", the University of Giessen-Marburg Lung Center (UGMLC), the Cell and Gene Therapy Center (CGT) supported by the HMWK, and the Deutsches Zentrum für Herz- und Kreislaufforschung.

2.1.3.6. References

This article references 69 other publications.

1. 1
   Fausto, N.; Campbell, J. S.The role of hepatocytes and oval cells in liver regeneration and repopulation *Mech*. *Dev*. 2003, 120 (1) 117– 30
2. 2.
   Porrello, E. R.; Mahmoud, A. I.; Simpson, E.; Hill, J. A.; Richardson, J. A.; Olson, E. N.; Sadek, H. A.Transient regenerative potential of the neonatal mouse heart *Science* 2011, 331 (6020) 1078– 80
3. 3.
   Tassava, R. A.; Huang, Y.Tail regeneration and ependymal outgrowth in the adult newt, *Notophthalmus viridescens*, are adversely affected by experimentally produced ischemia *J*. *Exp*. *Zool*., *Part A* 2005, 303 (12) 1031– 9
4. 4.
   Wong, C. J.; Liversage, R. A.Limb developmental stages of the newt *Notophthalmus viridescens Int*. *J*. *Dev*. *Biol*. 2005, 49 (4) 375– 89
5. 5.
   Kimura, Y.; Madhavan, M.; Call, M. K.; Santiago, W.; Tsonis, P. A.; Lambris, J. D.; Del Rio-Tsonis, K.Expression of complement 3 and complement 5 in newt limb and lens regeneration *J*. *Immunol*. 2003, 170 (5) 2331– 9
6. 6.
   Henry, J. J.; Tsonis, P. A.Molecular and cellular aspects of amphibian lens regeneration *Prog*. *Retinal Eye Res*. 2010, 10.1016/j.preteyeres.2010.07.002
7. 7.
   Parish, C. L.; Beljajeva, A.; Arenas, E.; Simon, A.Midbrain dopaminergic neurogenesis and behavioural recovery in a salamander lesion-induced regeneration model *Development* 2007, 134 (15) 2881– 7
8. 8.
   Bader, D.; Oberpriller, J.Autoradiographic and electron microscopic studies of minced cardiac muscle regeneration in the adult newt, notophthalmus viridescens *J*. *Exp*. *Zool*. 1979, 208 (2) 177– 93
9. 9.
   Bader, D.; Oberpriller, J. O.Repair and reorganization of minced cardiac muscle in the adult newt (*Notophthalmus viridescens*) *J*. *Morphol*. 1978, 155 (3) 349– 57
10. 10.
    Borchardt, T.; Braun, T.Cardiovascular regeneration in non-mammalian model systems: what are the differences between newts and man? *Thromb*. *Haemostasis* 2007, 98 (2) 311– 8
11. 11.
    Eguchi, G.; Eguchi, Y.; Nakamura, K.; Yadav, M. C.; Millan, J. L.; Tsonis, P. A.Regenerative capacity in newts is not altered by repeated regeneration and ageing *Nat*. *Commun*. 2011, 2, 384
12. 12.
    Kragl, M.; Knapp, D.; Nacu, E.; Khattak, S.; Maden, M.; Epperlein, H. H.; Tanaka, E. M.Cells keep a memory of their tissue origin during axolotl limb regeneration *Nature* 2009, 460 (7251) 60– 5
13. 13.
    Oberpriller, J. O.; Oberpriller, J. C.; Matz, D. G.; Soonpaa, M. H.Stimulation of proliferative events in the adult amphibian cardiac myocyte *Ann*. *N*. *Y*. *Acad*. *Sci*. 1995, 752, 30– 46

14. 14.
Oberpriller, J. O.; Oberpriller, J. C.Response of the adult newt ventricle to injury *J*. *Exp*. *Zool*. 1974, 187 (2) 249– 53
15. 15.
Oberpriller, J.; Oberpriller, J. C.Mitosis in adult newt ventricle *J*. *Cell Biol*. 1971, 49 (2) 560– 3
16. 16.
Laube, F.; Heister, M.; Scholz, C.; Borchardt, T.; Braun, T.Re-programming of newt cardiomyocytes is induced by tissue regeneration *J*. *Cell Sci*. 2006, 119 (Pt 22) 4719– 29
17. 17.
Kikuchi, K.; Holdway, J. E.; Werdich, A. A.; Anderson, R. M.; Fang, Y.; Egnaczyk, G. F.; Evans, T.; Macrae, C. A.; Stainier, D. Y.; Poss, K. D.Primary contribution to zebrafish heart regeneration by gata4(+) cardiomyocytes *Nature* 2010, 464 (7288) 601– 5
18. 18.
Jopling, C.; Sleep, E.; Raya, M.; Marti, M.; Raya, A.; Izpisua Belmonte, J. C.Zebrafish heart regeneration occurs by cardiomyocyte dedifferentiation and proliferation *Nature* 2010, 464 (7288) 606– 9
19. 19.
Garza-Garcia, A. A.; Driscoll, P. C.; Brockes, J. P.Evidence for the local evolution of mechanisms underlying limb regeneration in salamanders *Integr*. *Comp*. *Biol*. 2010, 50 (4) 528– 35
20. 20.
Kumar, A.; Godwin, J. W.; Gates, P. B.; Garza-Garcia, A. A.; Brockes, J. P.Molecular basis for the nerve dependence of limb regeneration in an adult vertebrate *Science* 2007, 318 (5851) 772– 7
21. 21.
Borchardt, T.; Looso, M.; Bruckskotten, M.; Weis, P.; Kruse, J.; Braun, T.Analysis of newly established EST databases reveals similarities between heart regeneration in newt and fish *BMC Genomics* 2010, 11, 4
22. 22.
Bruckskotten, M.; Looso, M.; Reinhardt, R.; Braun, T.; Borchardt, T.Newt-omics: a comprehensive repository for omics data from the newt *Notophthalmus viridescens Nucleic Acids Res*. 2011, 40 (Database issue) D895– 900
23. 23.
Standing, K. G.Peptide and protein de novo sequencing by mass spectrometry *Curr*. *Opin*. *Struct*. *Biol*. 2003, 13 (5) 595– 601
24. 24.
Liska, A. J.; Popov, A. V.; Sunyaev, S.; Coughlin, P.; Habermann, B.; Shevchenko, A.; Bork, P.; Karsenti, E.Homology-based functional proteomics by mass spectrometry: application to the Xenopus microtubule-associated proteome *Proteomics* 2004, 4 (9) 2707– 21
25. 25.
Yates, J. R., 3rdDatabase searching using mass spectrometry data *Electrophoresis* 1998, 19 (6) 893– 900
26. 26.
Choudhary, J. S.; Blackstock, W. P.; Creasy, D. M.; Cottrell, J. S.Matching peptide mass spectra to EST and genomic DNA databases *Trends Biotechnol*. 2001, 19 (10 Suppl.) S17– 22

27. 27.
Looso, M.; Borchardt, T.; Krueger, M.; Braun, T.Advanced identification of proteins in uncharacterized proteomes by pulsed in vivo SILAC *Mol*. *Cell*. *Proteomics* 2010, 10.1074/mcp.M900426-MCP200
28. 28.
Ong, S. E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M.Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics *Mol*. *Cell*. *Proteomics* 2002, 1 (5) 376– 86
29. 29.
Kruger, M.; Moser, M.; Ussar, S.; Thievessen, I.; Luber, C. A.; Forner, F.; Schmidt, S.; Zanivan, S.; Fassler, R.; Mann, M.SILAC mouse for quantitative proteomics uncovers kindlin-3 as an essential factor for red blood cell function *Cell* 2008, 134 (2) 353– 64
30. 30.
Sury, M. D.; Chen, J. X.; Selbach, M.The SILAC fly allows for accurate protein quantification in vivo *Mol*. *Cell*. *Proteomics* 2010, 10.1074/mcp.M110.000323
31. 31.
Boettger, T.; Beetz, N.; Kostin, S.; Schneider, J.; Kruger, M.; Hein, L.; Braun, T.Acquisition of the contractile phenotype by murine arterial smooth muscle cells depends on the Mir143/145 gene cluster *J*. *Clin*. *Invest*. 2009, 119 (9) 2634– 47
32. 32.
Fritz, I.Studies on the incorporation and release of N15 by tissue proteins of rats fed N15-glycine. I. The protein metabolic activity of salivary glands *J*. *Dent*. *Res* 1955, 34 (3) 435– 45
33. 33.
Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R.Quantitative analysis of complex protein mixtures using isotope-coded affinity tags *Nat*. *Biotechnol*. 1999, 17 (10) 994– 9
34. 34.
Schmidt, A.; Kellermann, J.; Lottspeich, F.A novel strategy for quantitative proteomics using isotope-coded protein labels *Proteomics* 2005, 5 (1) 4– 15
35. 35.
Doherty, M. K.; Hammond, D. E.; Clague, M. J.; Gaskell, S. J.; Beynon, R. J.Turnover of the human proteome: determination of protein intracellular stability by dynamic SILAC *J*. *Proteome Res*. 2009, 8 (1) 104– 12
36. 36.
Schwanhausser, B.; Gossen, M.; Dittmar, G.; Selbach, M.Global analysis of cellular protein translation by pulsed SILAC *Proteomics* 2009, 9 (1) 205– 9
37. 37.
Liao, L.; Park, S. K.; Xu, T.; Vanderklish, P.; Yates, J. R., 3rdQuantitative proteomic analysis of primary neurons reveals diverse changes in synaptic protein content in fmr1 knockout mice *Proc*. *Natl*. *Acad*. *Sci*. *U*. *S*. *A*. 2008, 105 (40) 15281– 6
38. 38.
Frangogiannis, N. G.The immune system and cardiac repair *Pharmacol*. *Res*. 2008, 58 (2) 88– 111
39. 39.
Frangogiannis, N. G.Targeting the inflammatory response in healing myocardial infarcts *Curr*. *Med*. *Chem*. 2006, 13 (16) 1877– 93

40. 40.
Geiger, T.; Cox, J.; Ostasiewicz, P.; Wisniewski, J. R.; Mann, M.Super-SILAC mix for quantitative proteomics of human tumor tissue *Nat*. *Methods* 2010, 7 (5) 383– 5
41. 41.
Finn, R. D.; Tate, J.; Mistry, J.; Coggill, P. C.; Sammut, S. J.; Hotz, H. R.; Ceric, G.; Forslund, K.; Eddy, S. R.; Sonnhammer, E. L.; Bateman, A.The Pfam protein families database *Nucleic Acids Res*. 2008, 36 (Database issue) D281– 8
42. 42.
Katsube, K.; Sakamoto, K.; Tamamura, Y.; Yamaguchi, A.Role of CCN, a vertebrate specific gene family, in development *Dev*., *Growth Differ*. 2009, 51 (1) 55– 67
43. 43.
Onara, D. F.; Forlenza, M.; Gonzalez, S. F.; Rakus, K. L.; Pilarczyk, A.; Irnazarow, I.; Wiegertjes, G. F.Differential transcription of multiple forms of alpha-2-macroglobulin in carp (*Cyprinus carpio*) infected with parasites *Dev*. *Comp*. *Immunol*. 2008, 32 (4) 339– 47
44. 44.
Lim, S. K.; Kim, J. C.; Moon, C. J.; Kim, G. Y.; Han, H. J.; Park, S. H.Formaldehyde induces apoptosis through decreased Prx 2 via p38 MAPK in lung epithelial cells *Toxicology* 2010, 271 (3) 100– 6
45. 45.
Wu, C.; Orozco, C.; Boyer, J.; Leglise, M.; Goodale, J.; Batalov, S.; Hodge, C. L.; Haase, J.; Janes, J.; Huss, J. W., 3rd; Su, A. I.BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources *Genome Biol*. 2009, 10 (11) R130
46. 46.
Purevsuren, J.; Fukao, T.; Hasegawa, Y.; Kobayashi, H.; Li, H.; Mushimoto, Y.; Fukuda, S.; Yamaguchi, S.Clinical and molecular aspects of Japanese patients with mitochondrial trifunctional protein deficiency *Mol*. *Genet*. *Metab*. 2009, 98 (4) 372– 7
47. 47.
Tanaka, T.; Hirota, Y.; Sohmiya, K.; Nishimura, S.; Kawamura, K.Serum and urinary human heart fatty acid-binding protein in acute myocardial infarction *Clin*. *Biochem*. 1991, 24 (2) 195– 201
48. 48.
Xie, P. Y.; Li, Y. P.; Chan, C. P.; Cheung, K. Y.; Cautherley, G. W.; Renneberg, R.A one-step immunotest for rapid detection of heart-type fatty acid-binding protein in patients with acute coronary syndromes *J*. *Immunoassay Immunochem*. 2010, 31 (1) 24– 32
49. 49.
Teltathum, T.; Mekchay, S.Proteome changes in Thai indigenous chicken muscle during growth period *Int*. *J*. *Biol*. *Sci*. 2009, 5 (7) 679– 85
50. 50.
Kumar, A.; Gates, P. B.; Brockes, J. P.Positional identity of adult stem cells in salamander limb regeneration *C*. *R*. *Biol*. 2007, 330 (6–7) 485– 90
51. 51.
Lau, L. F.; Lam, S. C.The CCN family of angiogenic regulators: the integrin connection *Exp*. *Cell Res*. 1999, 248 (1) 44– 57
52. 52.

Pi, L.; Ding, X.; Jorgensen, M.; Pan, J. J.; Oh, S. H.; Pintilie, D.; Brown, A.; Song, W. Y.; Petersen, B. E.Connective tissue growth factor with a novel fibronectin binding site promotes cell adhesion and migration during rat oval cell activation *Hepatology* 2008, 47 (3) 996– 1004

53. 53.
Latinkic, B. V.; Mercurio, S.; Bennett, B.; Hirst, E. M.; Xu, Q.; Lau, L. F.; Mohun, T. J.; Smith, J. C.Xenopus Cyr61 regulates gastrulation movements and modulates Wnt signalling *Development* 2003, 130 (11) 2429– 41

54. 54.
Yoshida, Y.; Togi, K.; Matsumae, H.; Nakashima, Y.; Kojima, Y.; Yamamoto, H.; Ono, K.; Nakamura, T.; Kita, T.; Tanaka, M.CCN1 protects cardiac myocytes from oxidative stress via beta1 integrin-Akt pathway *Biochem*. *Biophys*. *Res*. *Commun*. 2007, 355 (3) 611– 8

55. 55.
Lin, M. T.; Chang, C. C.; Chen, S. T.; Chang, H. L.; Su, J. L.; Chau, Y. P.; Kuo, M. L.Cyr61 expression confers resistance to apoptosis in breast cancer MCF-7 cells by a mechanism of NF-kappaB-dependent XIAP up-regulation *J*. *Biol*. *Chem*. 2004, 279 (23) 24015– 23

56. 56.
Venkatachalam, K.; Venkatesan, B.; Valente, A. J.; Melby, P. C.; Nandish, S.; Reusch, J. E.; Clark, R. A.; Chandrasekar, B.WISP1, a pro-mitogenic, pro-survival factor, mediates tumor necrosis factor-alpha (TNF-alpha)-stimulated cardiac fibroblast proliferation but inhibits TNF-alpha-induced cardiomyocyte death *J*. *Biol*. *Chem*. 2009, 284 (21) 14414– 27

57. 57.
Kikuchi, K.; Holdway, J. E.; Major, R. J.; Blum, N.; Dahn, R. D.; Begemann, G.; Poss, K. D.Retinoic acid production by endocardium and epicardium is an injury response essential for zebrafish heart regeneration *Dev*. *Cell* 2011, 20 (3) 397– 404

58. 58.
Pinckard, R. N.; Olson, M. S.; Giclas, P. C.; Terry, R.; Boyer, J. T.; O'Rourke, R. A.Consumption of classical complement components by heart subcellular membranes in vitro and in patients after acute myocardial infarction *J*. *Clin*. *Invest*. 1975, 56 (3) 740– 50

59. 59.
Kataoka, T.; Holler, N.; Micheau, O.; Martinon, F.; Tinel, A.; Hofmann, K.; Tschopp, J.Bcl-rambo, a novel Bcl-2 homologue that induces apoptosis via its unique C-terminal extension *J*. *Biol*. *Chem*. 2001, 276 (22) 19548– 54

60. 60.
Setsuta, K.; Seino, Y.; Ogawa, T.; Ohtsuka, T.; Seimiya, K.; Takano, T.Ongoing myocardial damage in chronic heart failure is related to activated tumor necrosis factor and Fas/Fas ligand system *Circ*. *J*. 2004, 68 (8) 747– 50

61. 61.
Muehlschlegel, J. D.; Perry, T. E.; Liu, K. Y.; Fox, A. A.; Collard, C. D.; Shernan, S. K.; Body, S. C.Heart-type fatty acid binding protein is an independent predictor of death and ventricular dysfunction after coronary artery bypass graft surgery *Anesth*. *Analg*. 2010, 111 (5) 1101– 9

62. 62.
Tadimalla, A.; Belmont, P. J.; Thuerauf, D. J.; Glassy, M. S.; Martindale, J. J.; Gude, N.; Sussman, M. A.; Glembotski, C. C.Mesencephalic astrocyte-derived

neurotrophic factor is an ischemia-inducible secreted endoplasmic reticulum stress response protein in the heart *Circ*. *Res*. 2008, 103 (11) 1249– 58
63. 63.
Wisniewski, J. R.; Zougman, A.; Mann, M.Combination of FASP and StageTip-based fractionation allows in-depth analysis of the hippocampal membrane proteome *J*. *Proteome Res*. 2009, 8 (12) 5674– 8
64. 64.
Rappsilber, J.; Ishihama, Y.; Mann, M.Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics *Anal*. *Chem*. 2003, 75 (3) 663– 70
65. 65.
Olsen, J. V.; Macek, B.; Lange, O.; Makarov, A.; Horning, S.; Mann, M.Higher-energy C-trap dissociation for peptide modification analysis *Nat*. *Methods* 2007, 4 (9) 709– 12
66. 66.
Cox, J.; Mann, M.MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification *Nat*. *Biotechnol*. 2008, 26 (12) 1367– 72
67. 67.
Elias, J. E.; Gygi, S. P.Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry *Nat*. *Methods* 2007, 4 (3) 207– 14
68. 68.
Barrell, D.; Dimmer, E.; Huntley, R. P.; Binns, D.; O'Donovan, C.; Apweiler, R.The GOA database in 2009—an integrated Gene Ontology Annotation resource *Nucleic Acids Res*. 2009, 37 (Database issue) D396– 403
69. 69.
Eden, E.; Navon, R.; Steinfeld, I.; Lipson, D.; Yakhini, Z.GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists *BMC Bioinf*. 2009, 10, 48

2.1.3.7. Figures:

Figure 1:

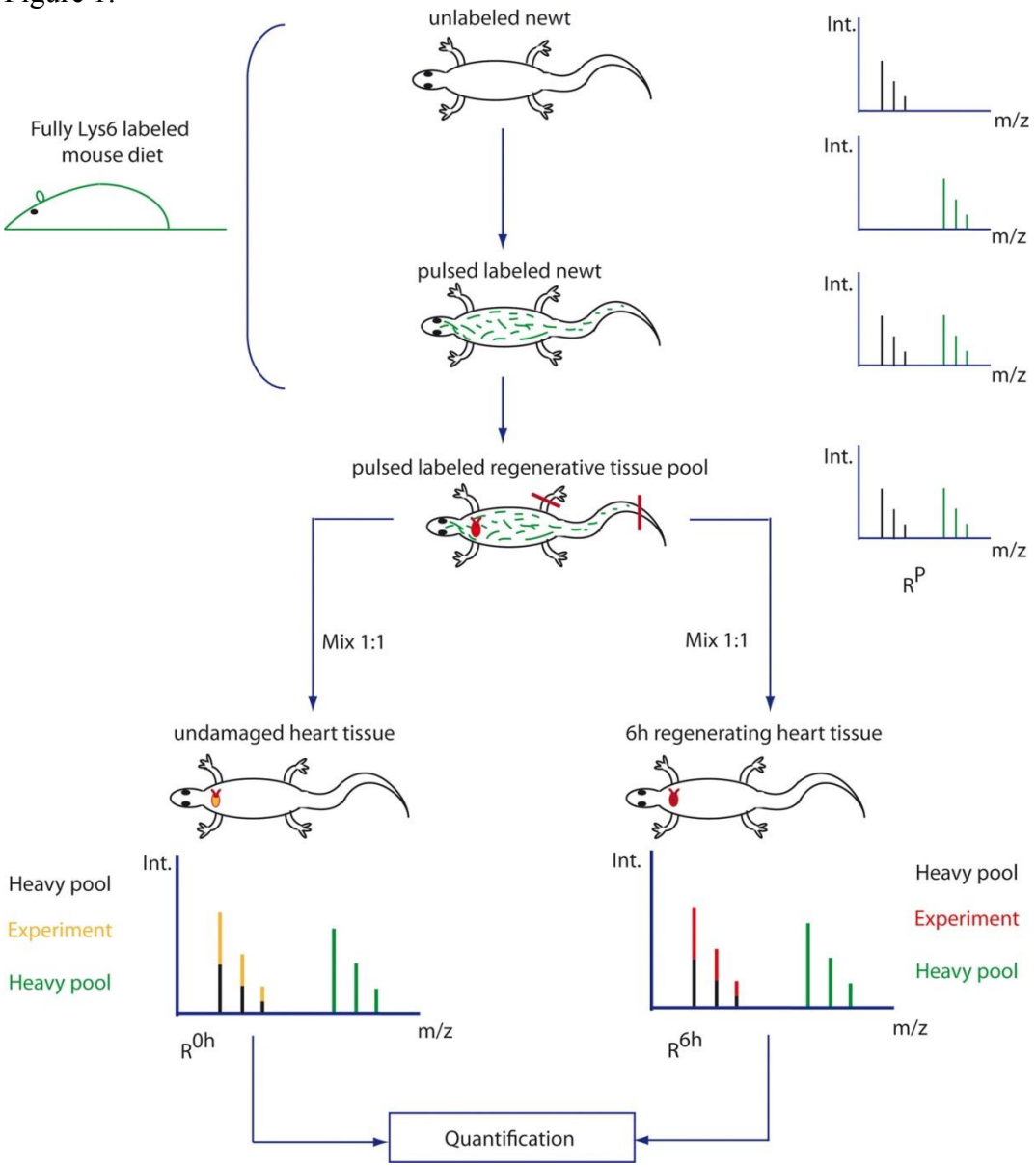


Figure 1. Schematic representation of the in vivo labeling strategy. Newts are labeled with stable isotopes of amino acids (Lys [C^{13}]) by feeding them a fully labeled mouse tissue diet. Partially labeled newts are subjected to regeneration in different tissues such as tail, limb, or heart by amputation or squeezing. Partially labeled proteins are pooled and used as a spike-in reference. Pools are measured by mass spectrometry, and heavy/light ratios are calculated for each individual protein. The heavy pool is mixed with light (not labeled) undamaged tissue or regenerating tissues at time points of interest. Spectra from heavy and light peptides represent signals from the light and the (incompletely) labeled heavy pool (as indicated by colors) and are used for exact peptide quantification.

Figure 2:

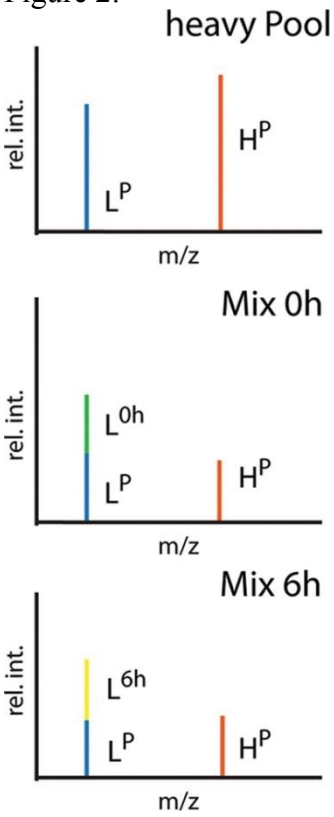


Figure 2. Schematic representation of pulsed labeled peptide peaks. The heavy pool contains a schematic H/L peptide peak pair, in which the heavy peptide is partially labeled. LP (blue) indicates the light peak and HP (red) indicates the heavy peak. The mass difference between the peaks is 6 Da. Mix 0 h contains a schematic H/L peptide peak pair obtained after combining a heavy reference pool and a light pool derived from undamaged tissue (0 h). LP and HP indicate peaks related to the mixed-in heavy pool. L0h (green) indicates the light peak from the experiment. Mix 6 h contains a schematic H/L peptide peak pair obtained after combining a heavy reference pool and a light experiment of regenerating tissue (6 h after damage). LP and HP indicate peaks related to the mixed-in heavy pool, L6h (yellow) indicates the light peak from the experiment.

Figure 3:

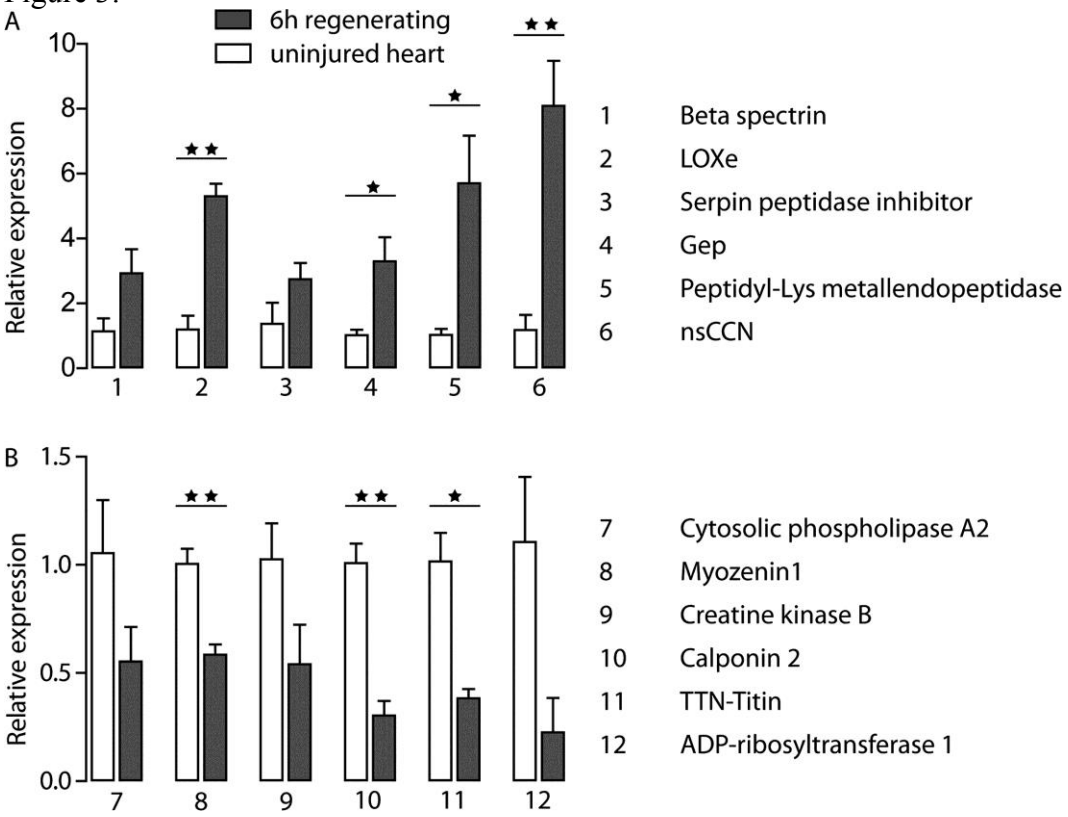


Figure 3. Validation of deregulated proteins on the transcriptome level via qRT-PCR. (A) mRNA expression profiles of proteins that were up-regulated in pulsed labeled experiments. (B) mRNAs of proteins that were down-regulated. Light bars indicate expression levels in uninjured heart tissue, and dark bars indicate expression levels in regenerating hearts 6 h after injury. Expression levels were normalized against rp21. Values are shown as an average of triplicates. Error bars represent standard deviations: * $p < 0.05$, ** $p < 0.01$.

Figure 4:

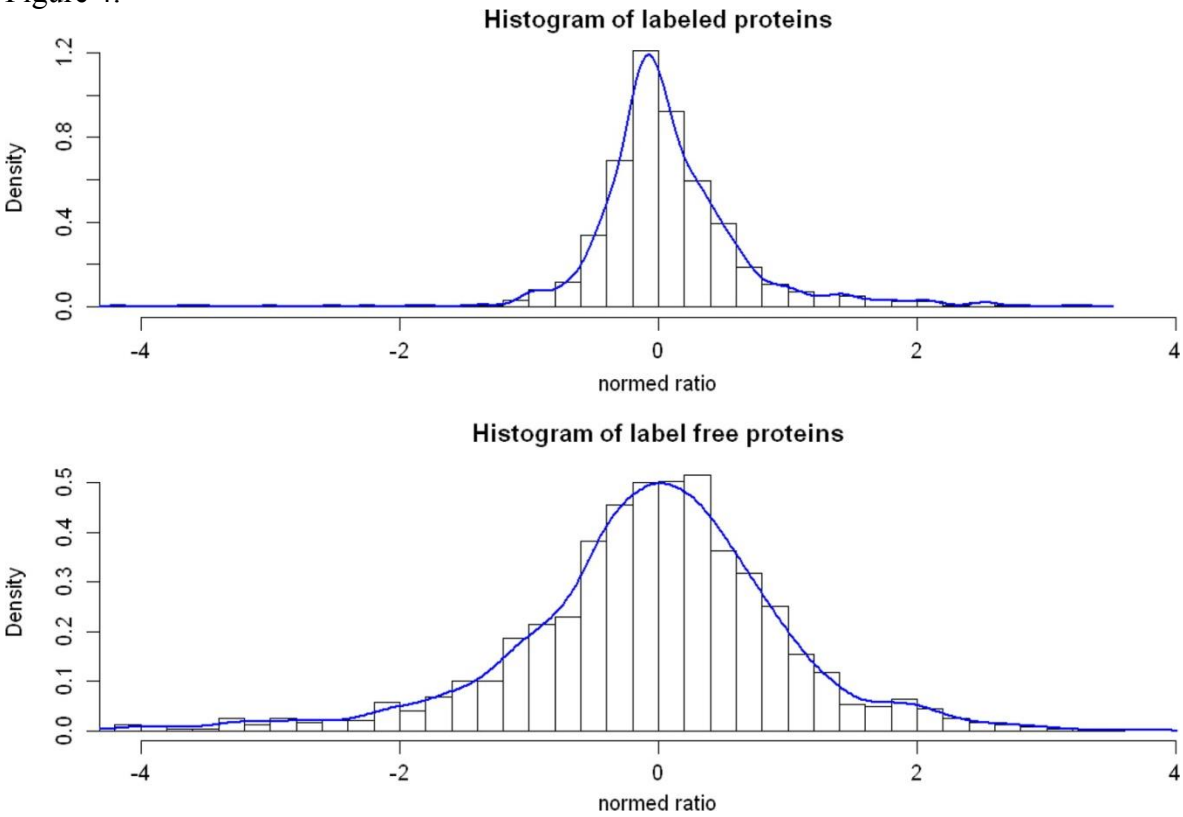


Figure 4. Pulsed labeled based quantification is more accurate than label-free quantification. A comparison of pulsed labeled based quantification (A) to label-free quantification (B) is shown. *X*-axis shows protein ratios in bins (log 2), and the *Y*-axis represents the total number of proteins per bin. Standard deviations of nonlabeled protein quantifications are almost twice as large as in pulsed labeled labeled protein quantifications.

Figure,5:

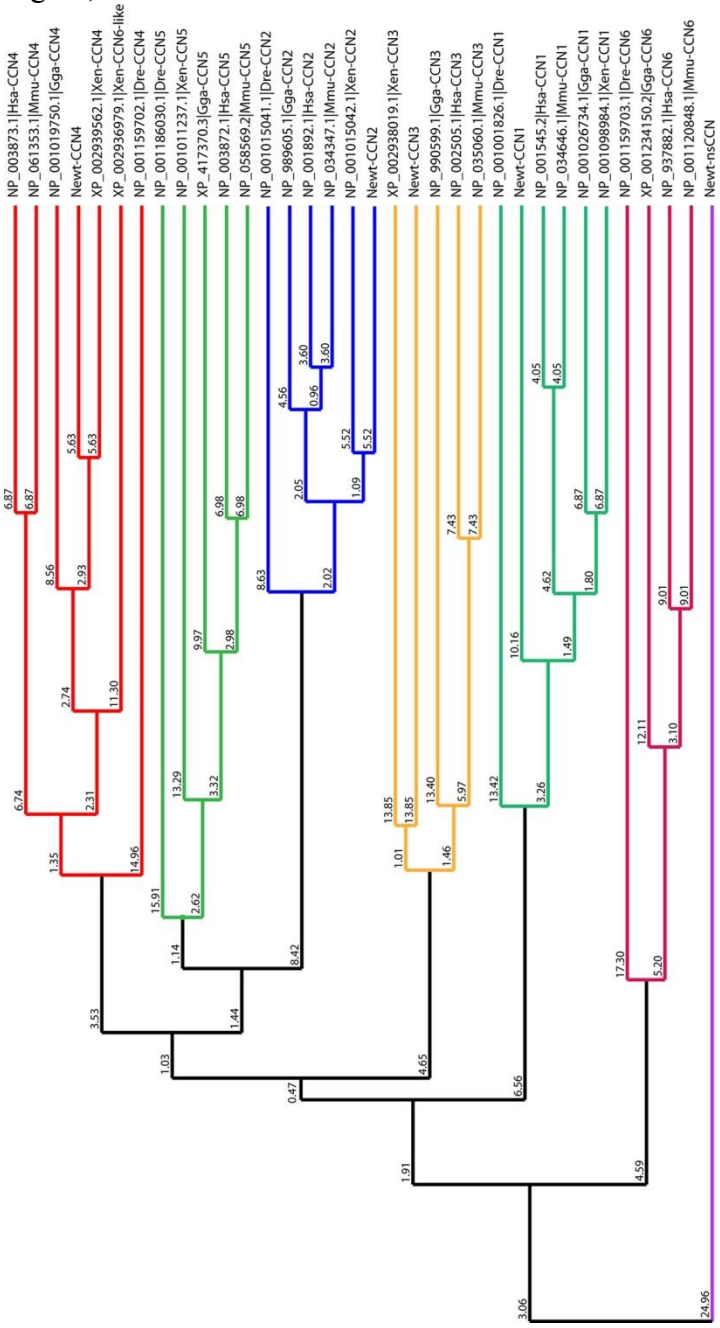


Figure 5.,Phylogenetic tree of CCN family members and position of the newt specific gene nsCCN. On the basis of multiple protein sequence alignments of all 5 newt CCN family members and all 6 human, zebrafish, Xenopus, chicken, and mouse CCN family members, an average distance phylogenetic tree was calculated on the basis of a measure of similarity between each pair of sequences. Numbers in the tree indicate the branch length, which correspond to the distance measure used to construct the tree. Sequences were taken from the NCBI RefSeq database. nsCCN does not locate close to any known CCN family member in mouse, human, zebrafish, Xenopus, chicken, and newt, indicating that it represents a new gene that has separated relatively early during evolution from other members of the CCN family.

Figure 6:

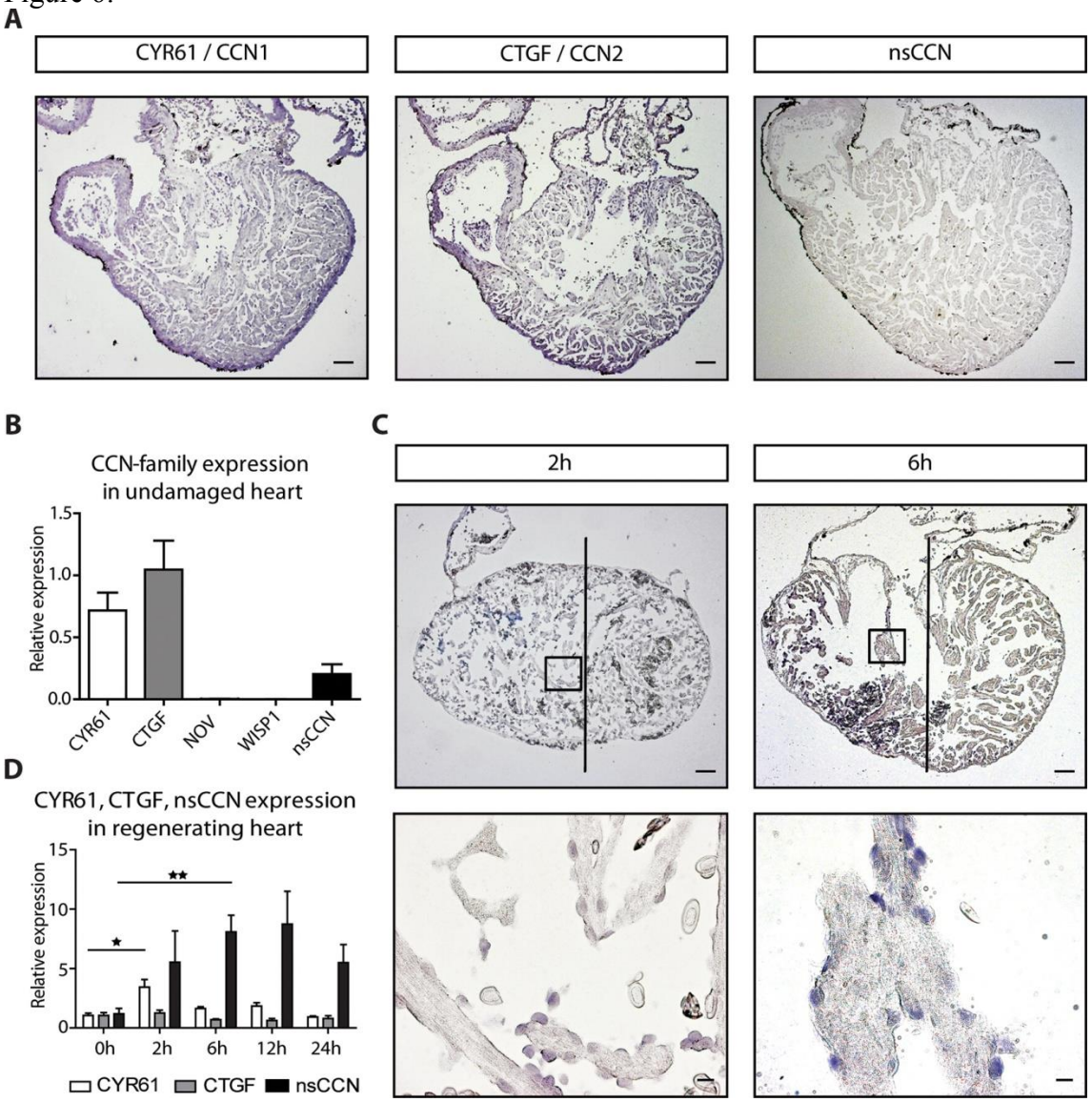


Figure 6. Expression analysis of different CCN family members in uninjured and regenerating hearts after mechanical injury. (A) In situ hybridization analysis of CYR61, CTGF, and nsCCN expression in uninjured hearts reveals high basal expression of CYR61 and CTGF but low expression of nsCCN in cardiomyocytes and smooth muscle cells. (B) qRT-PCR expression analysis of different CCN family members in intact hearts. (C) In situ hybridization analysis of nsCCN expression in regenerating heart 2 and 6 h after injury (5× magnification, scale bar 0.1 mm). The horizontal line demarcates regenerating (left) and nonregenerating (right) areas of the ventricle. nsCCN is expressed in endocardial cells located in damaged areas of the heart (40× magnification, scale bar 10 μm). (D) Expression analysis of CYR61, CTGF, and nsCCN in regenerating heart. All qRT-PCR expression levels were normalized to rp21. Values are shown as average of triplicates. Error bars represent standard deviations: * $p < 0.05$, ** $p < 0.01$.

2.2. Veröffentlichung Nr. 2 Erstautor:

„A microarray analysis of gene expression patterns during early phases of newt lens regeneration"

Konstantinos Sousounis,[1] Christian S. Michel,[2] Marc Bruckskotten,[2] Nobuyasu Maki,[1] Thilo Borchardt,[2] Thomas Braun,[2] Mario Looso,[2] and Panagiotis A. Tsonis[1]

[1]Department of Biology and Center for Tissue Regeneration and Engineering at Dayton, University of Dayton, OH

[2]Max-Planck-Institute for Heart and Lung Research, Ludwigstrasse 43, 61231 Bad Nauheim, Germany

Corresponding author.

The first two authors contributed equally to this work.

Correspondence to: Mario Looso, Max-Planck-Institute for Heart and Lung Research, Ludwigstrasse 43, 61231 Bad Nauheim, Germany; Phone: +49 (0)6032 705-1763, FAX: +49 (0)6032 705-419; email: Mario.Looso@mpi-bn.mpg.de

Veröffentlicht in:

Mol Vis. 2013;19:135-45. Epub 2013 Jan 28. PMID: 23378727 ; PMCID: PMC3559099

2.2.1. Inhaltsangabe

In dieser Veröffentlichung konnte gezeigt werden, dass bereits am ersten Tag nach erfolgter Linsenschädigung erste DNA Reparaturmechanismen, Zellzyklus Regulationen und eine Vielzahl identischer Prozesse auf dorsaler wie auf ventraler Irisseite ablaufen, aber auch einige Unterschiede im Expressionsmuster festzustellen sind. Diese Erkenntnisse wurden durch den Einsatz moderner Hochdurchsatzanalysen (cDNA Microarrays) erzielt.

2.2.2. Eigenanteil

Der Molch ist, wie bereits in der Einleitung erwähnt, in der Lage seine Augenlinse komplett zu regenerieren. Um besser verstehen zu können, welche Prozesse und Signalwege nach Schädigung und früher Regeneration eine Rolle spielen, wurde eine

Microarray Analyse an den Tagen 1, 3 und 5 nach Schädigung durchgeführt. Dazu wurde an den entsprechenden Tagen das komplette Auge entnommen und die RNA der dorsalen und ventralen Iris extrahiert. Diese wurde anschließend in cDNA umgeschrieben und mit zwei unterschiedlichen Fluoreszenzfarbstoffen (Alexa 555, Alexa 647) markiert.

Auf den verwendeten Microarray Chips sind jeweils etwa 100.000 individuelle Klone des *N. viridescens* gespottet worden (Beschreibung der verwendeten cDNA Bibliothek in „Newt-omics: a comprehensive repository for omics data from the newt Notophthalmus viridescens“[81]). Im Anschluss wurde die umgeschriebene fluoreszenzmarkierte cDNA auf diese Microarrays hybridisiert. Hierbei wurde ungeschädigte ventrale oder dorsale Iris mit geschädigter dorsaler und ventraler an den Tagen 1,3, und 5 verglichen. Für jeden Zeitpunkt wurde ein biologisches Replikat erstellt, sowie für jedes biologische Replikat ein technisches Replikat mit Fluoreszenzfarbstoffwechsel (dye-swap).

Diese Microarrays wurden von mir anschließend mit dem Gene Pix 4000b Microarray Scanner eingescannt und mit Hilfe der Acuity 4.0 Software analysiert. Für die Analyse der Microarray Daten verwendete ich zu Beginn drei unterschiedliche Einstellungen um hochqualitative Spots (high quality spots=HQS) von normalen (normal quality spots=NQS) oder niedrig-qualitativen Spots (low quality spots=LQS) unterscheiden zu können. Für die publizierten Ergebnisse wurden ausschließlich die hochqualifizierten Spots verwendet. Anschließend wurden die Daten normalisiert basierend auf der Lowess Normalisierung, die für zweifarbige Arrays geeignet ist. Die Lowess Normalisierung ist besonders geeignet nicht lineare Zusammenhänge in Messdaten herauszurechnen. Eine ausführliche Beschreibung der Lowess Normalisierung findet sich in Dudoit et al. (2002)[82].

Nach Normalisierung wurde ein Mittelwert über die biologischen und technischen Replikate für jeweils jeden hochqualitativen Spot gebildet. Für Werte mit einer $\log_2$ Ratio >0,9, <-0,9 fanden sich über alle drei Zeitpunkte, sowie dorsale und ventrale Iris 804 regulierte Spots. Für etwa 10% der regulierten Spots lagen noch keine Sequenzinformationen vor, sodass ich diese nachsequenzierte. Der Grund liegt darin, dass im Rahmen der Erstellung des Arrays von etwa 100.00 Spots überwiegend Spots sequenziert worden waren, die eine Regulation in der Herzregeneration zeigten und dadurch einige Spots unsequenziert blieben. Mit Erhalt aller Sequenzen konnte ich diese mit dem Programm SeqMan assemblieren und erhielt für die gewählten Parameter

(>80% identische Basenpaare, mindest Sequenzlänge 100 Basenpaare) 467 verschiedene Contigs. Die Contigs hatten eine durchschnittliche Sequenzlänge von 657 Basenpaaren. Diese 467 Contigs wurden von mir anschließend mittels BLAST (Informationen hierzu unter: www. ncbi.nlm.nih.gov/blast; Altschul et al. [83]), annotiert und von Herrn Bruckskotten in die am MPI gepflegte Datenbank für das Molchtranskriptom und Proteom eingepflegt (Zugang zur Datenbank www.newt-omics.mpi-bn.mpg.de). Als Cut-off für eine Annotation wurde ein e-value von e-15 gewählt.

Die regulierten Contigs wurden ebenfalls von mir geclustert und ein Expressionsmuster (heatmap) erstellt. Hierbei zeigte sich, dass eine Vielzahl regulierter Contigs in dorsaler und ventraler Iris identisch sind. Im Anschluss wurden für die deregulierten Contigs soweit möglich GO-Terme zugeteilt (siehe Eden et al., 2009). Diese zeigten eine Anreicherung von Kandidaten im Bereich Zell-Zyklus, Zell-Reparatur, Redox-Homeostase, Extrazellulär-Region und aus dem mitochondrialen Bereich.

Das Manuskript wurde von Konstantinos Sousounis, Dr. Mario Looso, Prof. Panagiotis A. Tsonis und mir verfasst und zur Publikation eingereicht. Figuren und Abbildungen in der Publikation wurden von Konstantinos Sousounis und mir erstellt.

Kurz-Zusammenfassung der durchgeführten Experimente:

- Scan von 48 Microarrays und anschließende Analyse mit Acuity 4.0 Software
- Plasmidpräparation (Miniprep) von 80 deregulierten Spots und anschließende Sanger-Sequenzierung
- Assemblierung deregulierter Sequenzen mit anschließender Annotation mittels BLAST; Clustering und Erstellung eines Expressionsmusters
- GO-Term Zuteilung

Unterschrift des Betreuers bezüglich der Angaben zum Eigenanteil:

_____ _____

Ort, Datum Unterschrift Prof. Dr. Dr. T. Braun

2.2.3. Originalpublikation:

"A microarray analysis of gene expression patterns during early phases of newt lens regeneration"

Konstantinos Sousounis,[1] Christian S. Michel,[2] Marc Bruckskotten,[2] Nobuyasu Maki,[1] Thilo Borchardt,[2] Thomas Braun,[2] Mario Looso,[2] and Panagiotis A. Tsonis[1]

[1]Department of Biology and Center for Tissue Regeneration and Engineering at Dayton, University of Dayton, OH

[2]Max-Planck-Institute for Heart and Lung Research, Ludwigstrasse 43, 61231 Bad Nauheim, Germany

Corresponding author.

The first two authors contributed equally to this work.

Correspondence to: Mario Looso, Max-Planck-Institute for Heart and Lung Research, Ludwigstrasse 43, 61231 Bad Nauheim, Germany; Phone: +49 (0)6032 705-1763, FAX: +49 (0)6032 705-419; email: Mario.Looso@mpi-bn.mpg.de

Received June 27, 2012; Accepted January 28, 2013.

Copyright © 2013 Molecular Vision.

This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

2.2.3.1 Abstract

Purpose

Notophthalmus viridescens, the red-spotted newt, possesses tremendous regenerative capabilities. Among the tissues and organs newts can regenerate, the lens is regenerated via transdifferentiation of the pigment epithelial cells of the dorsal iris, following complete removal (lentectomy). Under normal conditions, the same cells from the ventral iris are not capable of regenerating. This study aims to further understand the initial signals of lens regeneration.

Methods

We performed microarray analysis using RNA from a dorsal or ventral iris isolated 1, 3, and 5 days after lentectomy and compared to RNA isolated from an intact iris. This analysis was supported with quantitative real-time polymerase chain reaction (qRT-PCR) of selected genes.

Results

Microarrays showed 804 spots were differentially regulated 1, 3, and 5 days post-lentectomy in the dorsal and ventral iris. Functional annotation using Gene Ontology revealed interesting terms. Among them, factors related to cell cycle and DNA repair were mostly upregulated, in the microarray, 3 and 5 days post-lentectomy. qRT-PCR for rad1 and vascular endothelial growth factor receptor 1 showed upregulation for the dorsal iris 3 and 5 days post- lentectomy and for the ventral iris 5 days post-lentectomy. Rad1 was also upregulated twofold more in the dorsal iris than in the ventral iris 5 days post-lentectomy ($p<0.001$). Factors related to redox homeostasis were mostly upregulated in the microarray in all time points and samples. qRT-PCR for glutathione peroxidase 1 also showed upregulation in all time points for the ventral and dorsal iris. For the most part, mitochondrial enzymes were downregulated with the notable exception of cytochrome c–related oxidases that were mostly upregulated at all time points. qRT-PCR for cytochrome c oxidase subunit 2 showed upregulation especially 3 days post-lentectomy for the dorsal and ventral iris ($p<0.001$). Factors related to extracellular matrix and tissue remodeling showed mostly upregulation (except collagen I) for all time points and samples. qRT-PCR for stromelysin 1/2 alpha and avidin showed upregulation in all the time points for the dorsal and ventral iris.

Conclusions

The results show that the dorsal iris and the ventral iris follow the same general pattern with some distinct differences especially 5 days after lentectomy. In addition, while the expression of genes involved in DNA repair, redox homeostasis, and tissue remodeling in preparation for proliferation and transdifferentiation is altered in the entire iris, the response is more prominent in the dorsal iris following lentectomy.

2.2.3.2 Introduction

Whole organ regeneration is promising for medical applications. Many organ systems and models are being used for their ability to regenerate certain tissues and organs [1]. The red-spotted newt, *Notophthalmus viridescens*, has the ability to regenerate, after partial removal, many organs and tissues, including the heart [2], brain [3], tail, and limbs [4]. Newts also have the unique ability to regenerate whole organs such as the retina and lens of the eye even as adults. The lens is regenerated by transdifferentiation of pigmented epithelial cells (PECs) of the iris. Dorsal PECs are the only ones involved in lens regeneration while ventral PECs do not participate in the process. Dorsal and ventral PECs reenter the cell cycle at day 4 post-lentectomy, and by days 8–10, a dedifferentiated vesicle is formed in the dorsal iris. Thus, analysis of global gene expression in the dorsal and ventral iris could provide versatile answers why cells from one side can regenerate a lens and why the same cells from another side cannot. In addition, when individual genes were studied, important regulatory genes were expressed on the dorsal and ventral iris [5]. A small-scale microarray analysis using 373 genes showed similar expression pattern when examined at day 8 after lentectomy, which demarcates the stage of dedifferentiation and vesicle formation [6]. Congruent results were achieved by achieving microRNA expression data [7,8] and proteome analysis [9]. Consequently, our hypothesis is that the ventral iris initiates the process but is then inhibited. This is also implied by studies in which the ventral iris was induced to transdifferentiate by exogenous factors [10] or by studies in which gene expression-suppressive histone modifications were found specifically in the ventral iris [11].

To obtain a better idea of and new insights into this regulation, we examined gene expression patterns during the early phases of lens regeneration (1 day post-lentectomy to cell cycle reentry). We compared gene expression at day 0 (intact iris) with expression at days 1, 3, and 5 post-lentectomy. In the current study, we use custom-made microarrays generated from expressed sequence tags during heart regeneration [12]. Our data clearly indicate that the expression patterns between the dorsal and ventral iris are quite similar; however, there are distinct differences in the later stages (dorsal 5 days compared to ventral 5 days after lentectomy). Further, functional annotation and quantitative real-time polymerase chain reaction (qRT-PCR) analysis revealed that PECs prepare the ground for the main events of cell cycle reentry and for transdifferentiation even at the early stages.

2.2.3.3. Methods

Animals

Red-spotted newts, *Notophthalmus viridescens*, were obtained from Charles Sullivan Inc. Newt Farm (Nashville, TN). Procedures involving animals were approved by the Institutional Animal Care and Use Committee (IACUC) of the University of Dayton. Newts were anesthetized in 0.1% ethyl 3-aminobenzoate, methanesulfonic acid salt (MS222; Signa-Aldrich, St. Louis, MO) in phosphate buffer saline (PBS; 37 mM NaH_2PO_4 monohydrate, 58 mM Na_2HPO_4 anhydrous, pH 7.0) and then euthanized by decapitation.

Lentectomy – sample preparation

Corneas were split between the dorsal and ventral area with a scalpel. Tweezers were used to pull the lens out. Both lenses were removed from each newt. Newts were then placed in appropriate humidified containers for 1, 3, or 5 days. At the desired time points, whole eyeballs were removed and placed in RNAlater Solution (Life technologies, Grand Island, NY). Subsequently the eyeballs were dissected to obtain the dorsal and the ventral irises, which were separated in different RNase-free microtubes (Fisher Scientific, Pittsburgh, PA) with RNAlater Solution.

Microarrays

RNA pools from the samples were reverse transcribed to double-stranded cDNA using the SMART™ cDNA Library Construction Kit (Clontech, Mountain View, CA). This method allows amplification of double stranded full length cDNA by anchoring the 5' end of the isolated mRNA with a specific oligonucleotide [13]. Normalization of cDNA was performed based on DSN technology (Evrogen, Moscow, Russia). Cloned cDNAs were amplified to generate a library of more than 100,000 independent clones.

After plating, 100,000 individual clones were picked, and the inserts were amplified with PCR. After purification, inserts were resuspended in an appropriate spotting buffer (200 ng/µl PCR product in 3X SSC/1.5 M betaine) and spotted onto two sets of glass microarrays (Nexterion slides E, Schott). For each sample, 25 ventral or dorsal irises were pooled, and RNA was extracted with TRIzol (Life technologies). RNA was amplified with MessageAmp II aRNA Kit (Life technologies). For two-color microarray hybridization, cDNA labeling was performed with the SuperScript Plus Direct Labeling

Kit (Life technologies), using Alexa 555 and Alexa 647 dye coupled nucleotides (Life technologies). Undamaged ventral or dorsal irises were compared with ventral or dorsal irises 1, 3, and 5 days after lentectomy. Two samples per time point were technically replicated by dye swap to reach n=4 per experiment. Scanning of microarrays was performed on a GenePix 4000B Microarray Scanner Molecular Devices LLC, Sunnyvale, CA) with GenePix Pro 6.0 Software, and data were statistically analyzed with Acuity 4.0 Software (Molecular Devices LLC). A complete list of all expressed expressed sequence tags (ESTs) during lens regeneration is available as a downloadable tab-delimited file. The links to these lists can be found at the bottom of each dorsal and ventral expression page (Expression search). Sequencing of the cDNA library and the sequence assembly are described in [12].

Assembly of deregulated lens sequences

Sequences corresponding to 804 significantly regulated spots were assembled with SeqMan, version 8.0.2, with a minimum match percentage of 80% and a minimum sequence length of 100 bp, receiving 467 contigs with an average sequence length of 657 bp. For hierarchical clustering and heat map visualization, we used the Perseus software package. The expression values of the replicates were combined by mean. For visualization, we filtered the rows to have at least two valid values.

Functional annotation

Assembled contig sequences were assigned to Gene Ontology (GO) terms by the BLAST2GO tool. For similarity searches, we used the NCBI NR database (blastx) with an e-value cutoff e^{-10} [14,15]. Contigs that were consistently up- or downregulated in all time points were obtained with perl scripts. A Venn diagram was created using VENNY.

RNA extraction for quantitative real-time polymerase chain reaction

RNA extraction was performed following the TRIzol Reagent protocol (Life technologies). The tubes were centrifuged briefly, and RNAlater Solution was removed. For 20 iris pieces (10 newts; 20 irises), 500 µl of TRIzol Reagent were added, and using a pestle, the samples were homogenized and incubated at room temperature for 5 min. To eliminate pigments released to the solution, the samples were centrifuged for 10 min at 12,000 ×*g* at 4 °C. Briefly, chloroform was added, as the manufacturer suggests.

From the aqueous phase, 200 µl were removed to limit the amount of DNA contamination. RNA precipitation was performed with 100% isopropanol, and RNA was washed with 75% ethanol. RNA resuspension was performed with RNase-free water. RNA quality and quantity were checked with a NanoDrop 2000 spectrophotometer (Thermo Scientific, Waltham, MA). A clear peak in the 260 nm and an A260/A280 ratio >2 indicated high-quality RNA for qRT-PCR. RNA used for this experiment was from different animals than the one used to probe the microarrays.

Reverse transcription reaction

Reverse transcription (RT) reactions were performed with the First-strand cDNA Synthesis kit (GE Healthcare, Piscataway, NJ) following the recommended protocol for 1 µg total RNA and oligo(dT) primers for positive RT reactions and half the volumes for the negative RT reactions. For the negative RT reaction, no primers were added, and the samples were incubated for 5 min at 98 °C for enzyme inactivation before RNA was added. The success of the RT reaction was checked with polymerase chain reaction using TaKaRa Ex Taq (TaKaRa, Otsu, Shiga, Japan) and primers against *rpl27* and *gapdh* genes.

Quantitative real-time polymerase chain reaction

qRT-PCR was performed in a Bio-Rad iCycler (Bio-Rad, Hercules, CA). qRT-PCR reactions were performed using iQ SYBR Green Supermix (Bio-Rad) and following the manufacturer's protocol for a total volume of 25 µl. Melt curve analysis was performed for primer specificity. For each qRT-PCR run, a concentration gradient of the target gene cloned cDNAs was used. The concentration of the target genes in the samples was calculated using the standard curve made from the known concentration gradient (R^2 >0.98) and the number of cycles (Ct). The concentration of the target gene in the samples was normalized against the housekeeping gene *rpl27* (6), and the relative expression level compared to the intact dorsal iris was calculated (day 0). All samples were run in triplicate. Statistical significance was calculated using the Student *t* test. Table 1 shows the primers that were used. All primers were tested for specificity in known newt sequences using the Basic Local Alignment Search Tool [16]. Annealing temperatures were checked by locating only the appropriate size band using polymerase chain reaction followed by agarose gel electrophoresis.

2.2.3.4. Results and discussion

Array expression data

Microarray analysis obtained 804 spots with differential expression 1, 3, or 5 days post-lentectomy in the dorsal or ventral iris. Combining the expression values of the replicates yielded 467 different sequences, which we refer to as contigs. Appendix 1 contains the list of all assembled contigs, their annotation, expression in the microarrays, and an identifier that can be used to retrieve more information from the newtomics database. Differentially regulated contigs were clustered and visualized with a heat map (Figure 1). A general pattern that emerges from a visual inspection of the heat map is that there is common up- or downregulation, when compared with the intact iris (0 day) in the dorsal and ventral iris (see clusters A and B and part of Cluster C, Figure 1). In other words, genes that show upregulation during 1, 3, and 5 days post-lentectomy in the dorsal iris show the same differential expression in the ventral iris (Cluster B, Figure 1). Reversely, this is the case for several downregulated contigs, too (Cluster A, Figure 1). In addition, Figure 2 presents a direct comparison of dorsal and ventral genes that are consistently up- or downregulated without regarding single time points. Forty-six contigs out of 72 (63.9%) and 46 out of 57 (80.7%) are commonly upregulated in the dorsal and ventral iris, respectively. Fifty-two contigs out of 126 (41.3%) and 52 out of 68 (76.5%) are commonly downregulated in the dorsal and ventral iris, respectively. These results strengthen the hypothesis that dorsal and ventral irises initiate the same first steps of lens regeneration. In a different comparison, we examined which genes are regulated at dorsal day 5 compared with day 1 (D day 1, V day 1 downregulation versus D day 5 upregulation and opposite). This analysis could potentially reveal genes related to transdifferentiation. Interestingly, among regulated genes we found cytoskeletal organization-related proteins (stathmin 1 [17], svil protein [18]) and cell pluripotency-maintenance factor (rtf1 [19]; Figure 2B).

In addition to the clusters, we identified another remarkable pattern. This pattern is defined by an inversely regulated time point (five days; Cluster D; Figure 1). Other time points are not regulated. This cluster consists of 94 non-redundant members. When we examined this cluster for potential enrichment of assigned function, we found only 14 candidates with a similarity to public available sequences. However, most of the candidates lack counterparts in higher vertebrates, even as the corresponding sequences have a sufficient sequence length (see Table 2). Within the annotated contigs, we found

an isoform of suppression of tumorigenicity 7, known to play a crucial role in tumor suppression through regulating genes involved in maintaining the cellular structure [20] or remodeling the extracellular matrix [21]. Altogether this indicates that Cluster D might represent newt specific genes that are inversely regulated in the dorsal and ventral iris 5 days after lentectomy.

Functional annotation

Of the 467 contigs, 265 have a significant sequence similarity to public available proteins, whereas 192 have assigned GO terms. We further investigated these contigs and identified that the most interesting functional annotation, in concordance with other regeneration model systems, is related to cell cycle, proliferation, extracellular region, redox homeostasis, mitochondrion, DNA repair, and parental GO terms. The list of GO annotated contigs is shown in Appendix 2.

Cell cycle reentry, proliferation, and DNA repair

Previous studies revealed that by day 4 post-lentectomy the dorsal and ventral irises reenter the cell cycle [22]. We found that bccip homolog and myeloid leukemia factor 1, two proteins that cause cell cycle arrest, were slightly downregulated 3 days after lentectomy and then were upregulated 5 days after lentectomy in the dorsal iris. In the ventral and dorsal iris in all time points, peroxiredoxin 1, a protein that has been shown to play a role in proliferation and to be expressed in melanosomes, was upregulated (Table 3) [23,24]. In addition, tmsb4, also known as thymosin beta 4, was upregulated in all time points in the ventral and dorsal iris. This protein is known for actin organization and the S phase entry of the cell cycle. Among the DNA repair and proliferation factors, we found expression of rad1 and vascular endothelial growth factor receptor 1 (VEGFR1). Rad1 is a protein that plays a role in DNA repair of double-strand breaks and is part of the 9.1.1 complex with Rad9 and Hus1 [25]. VEGFR1 is a receptor that is important in the proliferation of endothelial cells, angiogenesis, and cancer. qRT-PCR analysis demonstrated Rad1 was upregulated in the dorsal iris 3 days ($p<0.001$) and 5 days ($p<0.001$) compared to 0 day post-lentectomy. Further, rad1 was upregulated in the ventral iris only 5 days ($p<0.05$) compared to 0 day after lentectomy. In addition, the dorsal iris showed upregulation of rad1 compared to the ventral iris 3 days ($p<0.05$) and 5 days ($p<0.001$) post-lentectomy (Figure 3). The other candidate, VEGFR1, was upregulated in the ventral and dorsal iris 1 day

($p<0.001$) and 5 days ($p<0.05$) compared to 0 day post-lentectomy. Overall, during regeneration, the dorsal iris has a higher VEGFR1 expression level compared to the ventral iris. These results are consistent with the onset of proliferation during the early phases of lens regeneration and are in context with the robustness in DNA repair inferred during regenerative activities. For example, Eguchi et al. showed that lens regeneration is not affected after repeated lentectomies (18 times over 16 years) in the Japanese newt, *Cynops pyrrhogaster* [26].

We think that these patterns reflect the preparation of the iris day 1 post-lentectomy to initiate reentry in the cell cycle to proliferate. This was accompanied by DNA repair factors that will most likely be activated to maintain the integrity of DNA especially in the dorsal iris (twofold upregulation of rad1 in dorsal versus ventral, Figure 3). These findings support the thesis that the dorsal iris prepares more rigorously for regenerating the lens.

Reactive oxygen species and mitochondria-related proteins

As shown in Table 4, enzymes related to redox homeostasis are upregulated in all time points of the ventral and the dorsal iris. Glutathione peroxidase 1, peroxiredoxin mitochondrial-like (peroxiredoxin 3), sh3 domain-binding glutamic acid-rich-like protein 3 (SH3BGRL3), and thioredoxin are proteins responsible for reducing reactive oxygen species (ROS) to protect the cells from ROS-related stress and apoptosis. We confirmed these findings derived from the microarray data with qRT-PCR of glutathione peroxidase 1 (Gpx1). Gpx1 was upregulated 1, 3, and 5 days after lentectomy in the dorsal ($p<0.001$ for all days) and ventral iris ($p<0.001$ for 1 and 5 days, $p<0.05$ for 3 days) compared to the intact iris. Gpx1 was upregulated in the dorsal iris 1 day ($p<0.05$) and 3 days ($p<0.001$) compared to the ventral iris (Figure 3). Peroxiredoxin 6, which was downregulated, seems to be linked with the redox system the cells chose to reduce ROS.

Another obvious pattern was formed by mitochondrial enzymes related to ketone metabolism, ATP synthesis, and purine biosynthesis that are downregulated in the dorsal and ventral iris in all investigated time points. However, the cytochrome c oxidases were mostly upregulated. Cytochrome c oxidase subunit 2 is a mitochondrial component involved in oxidative phosphorylation. Dysregulation of this gene is mostly due to mitochondrial-related stress [27]. Since lentectomy is a surgical procedure, it naturally creates stress. Cytochrome c oxidase subunit 2 was upregulated in the ventral iris 1 day ($p<0.05$) and 3 days ($p<0.001$) compared to 0 day post-lentectomy (qRT-PCR

validation, Figure 3). Cytochrome c oxidase subunit 2 was also upregulated in the dorsal iris 3 days ($p<0.001$) and 5 days ($p<0.05$) compared to 0 day post-lentectomy. In addition to these common regulations, there was an inverse differential expression between the dorsal and ventral iris at day 1 ($p<0.05$) and 3 ($p<0.05$) with increased expression levels in the ventral iris.

Additionally, glutathione peroxidase 1 and cytochrome oxidase subunit 2 had opposite differential expression between the dorsal and ventral irises 1 and 3 days post-lentectomy. *Gpx1* was upregulated in the dorsal iris after 1 day ($p<0.05$) and 3 days ($p<0.001$) post-lentectomy compared to the ventral iris. Cytochrome c oxidase subunit 2 was upregulated in the ventral iris 1 day ($p<0.05$) and 3 days ($p<0.05$) post-lentectomy compared to the dorsal iris. Both genes returned to the same expression levels 5 days post-lentectomy ($p>0.05$; Figure 3). These two genes are connected via ROS. Stress caused by ROS can affect mitochondria in which Gpx1 can reduce ROS [28]. The high expression level of *gpx1* in the dorsal iris has a potential role in reducing ROS faster than the ventral iris, affecting the rate of cell activation and responsiveness in early stages of regeneration.

To sum up, these results indicate that stress caused by ROS is answered through upregulating of enzymes that reduce ROS in the iris, thus protecting the cells. ROS might cause mitochondrial stress, which is reflected by downregulation of certain mitochondrial enzymes. It is not yet clear whether ROS is a signal of regeneration or ROS is produced by the surgical operation through cutting the cornea. However, these patterns provide an impetus for further investigation.

Extracellular matrix and transdifferentiation

Another group of annotated proteins can be described as extracellular matrix–related proteins (Table 5). Most of these proteins are upregulated in the dorsal and ventral iris in all time points. For example, clusterin is known to play a role in extracellular matrix organization by acting as a stress-induced secreted chaperone protein and having the ability to inhibit stress-induced protein precipitation [29]. Mmp18 and stromelysin 1/2 alpha are matrix metalloproteinases that degrade certain types of matrix, including collagen III, IV, IX, and X. Their role is critical for transdifferentiation since extracellular matrix and remodeling proteins are needed for cell fate change. qRT-PCR validation of stromelysin 1/2 alpha confirmed upregulation in the dorsal and ventral irises 1 day ($p<0.001$ for both), 3 days ($p<0.05$ for both), and 5 days ($p<0.001$ for both)

compared to 0 day after lentectomy. There was a slight upregulation of stromelysin 1/2 alpha in the dorsal iris compared to the ventral iris at day 1 post-lentectomy ($p<0.05$). The high expression of stromelysin 1/2 alpha is consistent with other studies in which it was upregulated in the dorsal iris and the ventral iris 8 days post-lentectomy compared to the intact iris [6]. The results allow us to assume that the process of tissue remodeling starts immediately after lentectomy since the matrix metalloproteinases were upregulated 1,200-fold at 1 day post-lentectomy in the ventral and dorsal irises ($p<0.001$). In the context of highly increased expression levels of matrix metalloproteinases, collagen I was downregulated, a result consistent with the fact that the cells degraded the collagen around them for later proliferation. Another protein known to be located in the extracellular matrix is avidin. qRT-PCR confirmed the expression data of our microarrays and showed upregulation in all investigated days after lentectomy for the ventral and dorsal irises compared to the intact iris. Avidin is thought to have an impact as a defense protein against microbial infections that could be caused through the surgery process [30].

Concluding remarks

With this report, we further expand our knowledge of gene expression during early lens regeneration in *Notophthalmus viridescens*. Our results clearly indicate that during early stages there are not many qualitative differences between the dorsal iris and the ventral iris. Rather, quantitative differences might be critical for the dorsal and ventral iris's contribution in lens regeneration. The lack of gene annotation for a group of potential newt/amphibian specific genes points toward the hypothesis that some regeneration-related proteins might be evolutionarily conserved only in amphibians. However, certain signatures for regeneration have been verified. Cell proliferation was combined with a robust quality check mechanism to keep the DNA intact and to be passed to new cells without mutations. We also report that the iris employs enzymes to reduce ROS. Both homeostasis mechanisms were upregulated on the site where the regeneration process occurs, the dorsal iris. In addition, as soon as 1 day after lentectomy, cells started to prepare for the transdifferentiation by upregulation of certain matrix metalloproteinases and regulation of their extracellular matrix, a step crucial to the upcoming differentiation to lens cells. As more studies on global gene expression become available, certain differences might lead to elucidating the much-coveted mechanisms of regenerative processes.

Acknowledgments

This work was supported by NIH grant EY10540 to PAT. KS has been supported in part by the University of Dayton, Office of Graduate, Professional & Continuing Education through the Graduate Student Summer Fellowship Program. Author contributions: KS, CSM, ML and PAT co-wrote the paper. NM performed experiments for the preparation of tissues and probes. KS and CSM performed the microarray analysis. KS, MB and ML contributed to the bioinformatical analysis. MB administrated the database. TBo hybridized the microarrays. TB, ML and PAT gave technical advice and contributed to the study design. All authors read and approved the final manuscript. Dr. Mario Looso (Mario.Looso@mpi-bn.mpg.de) and Panagiotis A. Tsonis (ptsonis1@udayton.edu) contributed equally to this work thus they both should be considered as corresponding authors.

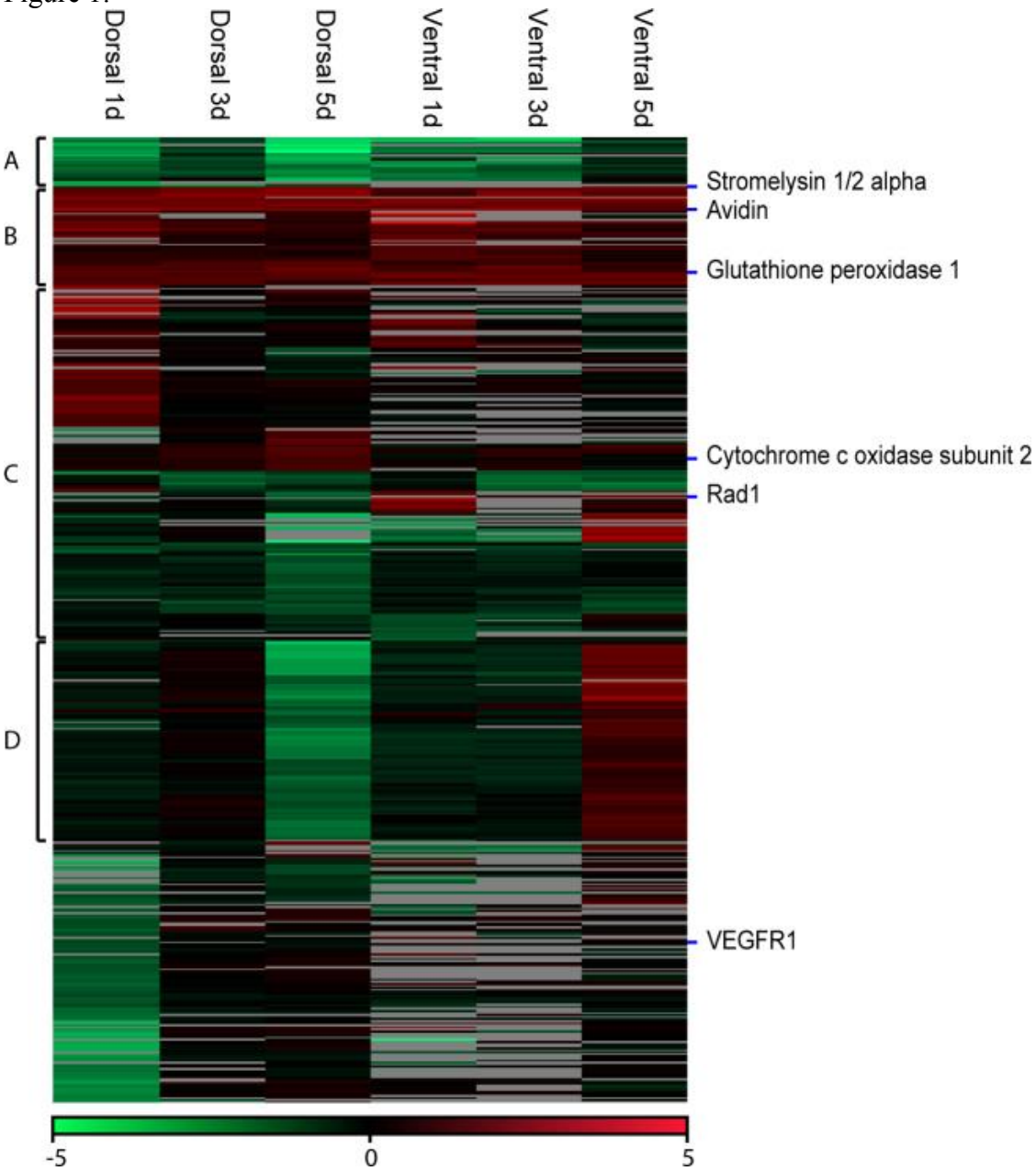
2.2.3.5. References

1. Baddour JA, Sousounis K, Tsonis PA. Organ repair and regeneration: an overview. Birth Defects Res C Embryo Today. 2012;96:1–29.
2. Oberpriller JO, Oberpriller JC. Response of the adult newt ventricle to injury. J Exp Zool. 1974;187:249–53.
3. Okamoto M, Ohsawa H, Hayashi T, Owaribe K, Tsonis PA. Regeneration of retinotectal projections after optic tectum removal in adult newts. Mol Vis. 2007;13:2112–8.
4. Tsonis PA. Limb Regeneration. Cambridge University Press1996.
5. Henry JJ, Tsonis PA. Molecular and cellular aspects of amphibian lens regeneration. Prog Retin Eye Res. 2010;29:543–55.
6. Makarev E, Call MK, Grogg MW, Atkinson DL, Milash B, Odelberg SJ, Tsonis PA. Gene expression signatures in the newt irises during lens regeneration. FEBS Lett. 2007;581:1865–70.
7. Tsonis PA, Call MK, Grogg MW, Sartor MA, Taylor RR, Forge A, Fyffe R, Goldenberg R, Cowper-Sal-lari R, Tomlinson CR. MicroRNAs and regeneration: Let-7 members as potential regulators of dedifferentiation in lens and inner ear hair cell regeneration of the adult newt. Biochem Biophys Res Commun. 2007;362:940–5.
8. Nakamura K, Maki N, Trinh A, Trask HW, Gui J, Tomlinson CR, Tsonis PA. miRNAs in newt lens regeneration: specific control of proliferation and evidence for miRNA networking. PLoS ONE. 2010;5:e12058.
9. Roddy M, Fox TP, McFadden JP, Nakamura K, Del Rio-Tsonis K, Tsonis PA. A comparative proteomic analysis during urodele lens regeneration. Biochem Biophys Res Commun. 2008;377:275–9.
10. Grogg MW, Call MK, Okamoto M, Vergara MN, Del Rio-Tsonis K, Tsonis PA. BMP inhibition-driven regulation of six-3 underlies induction of newt lens regeneration. Nature. 2005;438:858–62.
11. Maki N, Tsonis PA, Agata K. Changes in global histone modifications during dedifferentiation in newt lens regeneration. Mol Vis. 2010;16:1893–7.
12. Bruckskotten M, Looso M, Reinhardt R, Braun T, Borchardt T. Newt-omics: a comprehensive repository for omics data from the newt Notophthalmus viridescens. Nucleic Acids Res. 2012;40:D895–900.
13. Zhu YY, Machleder EM, Chenchik A, Li R, Siebert PD. Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. Biotechniques. 2001;30:892–7.
14. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics. 2005;21:3674–6.
15. Myhre S, Tveit H, Mollestad T, Laegreid A. Additional gene ontology structure for improved biological reasoning. Bioinformatics. 2006;22:2020–7.
16. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215:403–10.
17. Larsson N, Marklund U, Gradin HM, Brattsand G, Gullberg M. Control of microtubule dynamics by oncoprotein 18: dissection of the regulatory role of multisite phosphorylation during mitosis. Mol Cell Biol. 1997;17:5530–9.
18. Crowley JL, Smith TC, Fang Z, Takizawa N, Luna EJ. Supervillin reorganizes the actin cytoskeleton and increases invadopodial efficiency. Mol Biol Cell. 2009;20:948–62.

19. Ding L, Paszkowski-Rogacz M, Nitzsche A, Slabicki MM, Heninger AK, de Vries I, Kittler R, Junqueira M, Shevchenko A, Schulz H, Hubner N, Doss MX, Sachinidis A, Hescheler J, Iacone R, Anastassiadis K, Stewart AF, Pisabarro MT, Caldarelli A, Poser I, Theis M, Buchholz F. A genome-scale RNAi screen for Oct4 modulators defines a role of the Paf1 complex for embryonic stem cell identity. Cell Stem Cell. 2009;4:403–15.
20. Charong N, Patmasiriwat P, Zenklusen JC. Localization and characterization of ST7 in cancer. J Cancer Res Clin Oncol. 2011;137:89–97.
21. Hooi CF, Blancher C, Qiu W, Revet IM, Williams LH, Ciavarella ML, Anderson RL, Thompson EW, Connor A, Phillips WA, Campbell IG. ST7-mediated suppression of tumorigenicity of prostate cancer cells is characterized by remodeling of the extracellular matrix. Oncogene. 2006;25:3924–33.
22. Eguchi G, Shingai R. Cellular analysis on localization of lens forming potency in the newt iris epithelium. Dev Growth Differ. 1971;13:337–49.
23. Kawai S, Takeshita S, Okazaki M, Kikuno R, Kudo A, Amann E. Cloning and characterization of OSF-3, a new member of the MER5 family, expressed in mouse osteoblastic cells. J Biochem. 1994;115:641–3.
24. Chi A, Valencia JC, Hu ZZ, Watabe H, Yamaguchi H, Mangini NJ, Huang H, Canfield VA, Cheng KC, Yang F, Abe R, Yamagishi S, Shabanowitz J, Hearing VJ, Wu C, Appella E, Hunt DF. Proteomic and bioinformatic characterization of the biogenesis and function of melanosomes. J Proteome Res. 2006;5:3135–44.
25. Pichierri P, Nicolai S, Cignolo L, Bignami M, Franchitto A. The RAD9–RAD1-HUS1 (9.1.1) complex interacts with WRN and is crucial to regulate its response to replication fork stalling. Oncogene. 2012;31:2809–23.
26. Eguchi G, Eguchi Y, Nakamura K, Yadav MC, Millán JL, Tsonis PA. Regenerative capacity in newts is not altered by repeated regeneration and ageing. Nat Commun. 2011;2:384.
27. Jenko K, Karouna-Renier NK, Hoffman DJ. Gene expression, glutathione status, and indicators of hepatic oxidative stress in laughing gull (Larus atricilla) hatchlings exposed to methylmercury. Environ Toxicol Chem. 2012;31:2588–96.
28. Esposito LA, Kokoszka JE, Waymire KG, Cottrell B, MacGregor GR, Wallace DC. Mitochondrial oxidative stress in mice lacking the glutathione peroxidase-1 gene. Free Radic Biol Med. 2000;28:754–66.
29. Jones SE, Jomary C. Clusterin. Int J Biochem Cell Biol. 2002;34:427–31. Review.
30. Korpela J. Avidin, a high affinity biotin-binding protein, as a tool and subject of biological research. Med Biol. 1984;62:5–26.

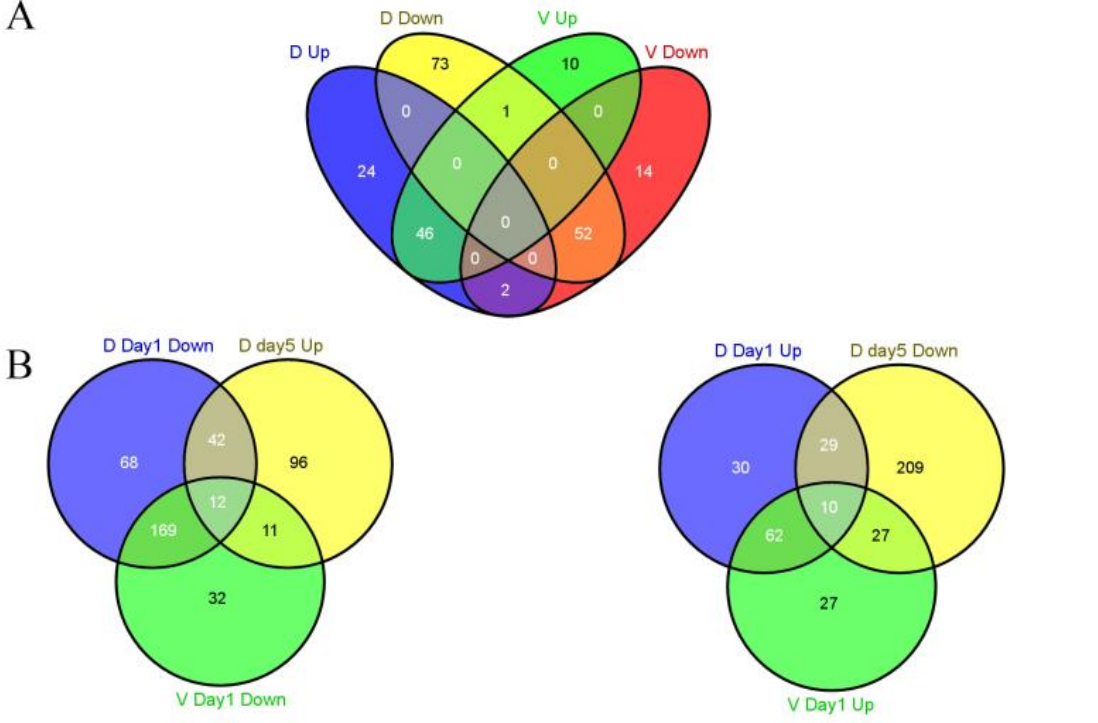
2.2.3.6. Figures and tables:

Figure 1:



Heat map of expression patterns derived from the microarrays. The heat map is subdivided into four clusters depending on the expression patterns. The location of the genes used for "quantitative real-time (qRT)-PCR analysis is shown on the heat map. Only contigs with a minimum number of two valid values have been selected for the heat map. For the heat map, we took 467 candidates into account, but this number was reduced to 465 candidates by filtering for valid values. Within the clusters, we had 23 transcripts within cluster A, 48 transcripts within Cluster B, 171 transcripts within Cluster C, and 94 transcripts within Cluster D.

Figure 2:

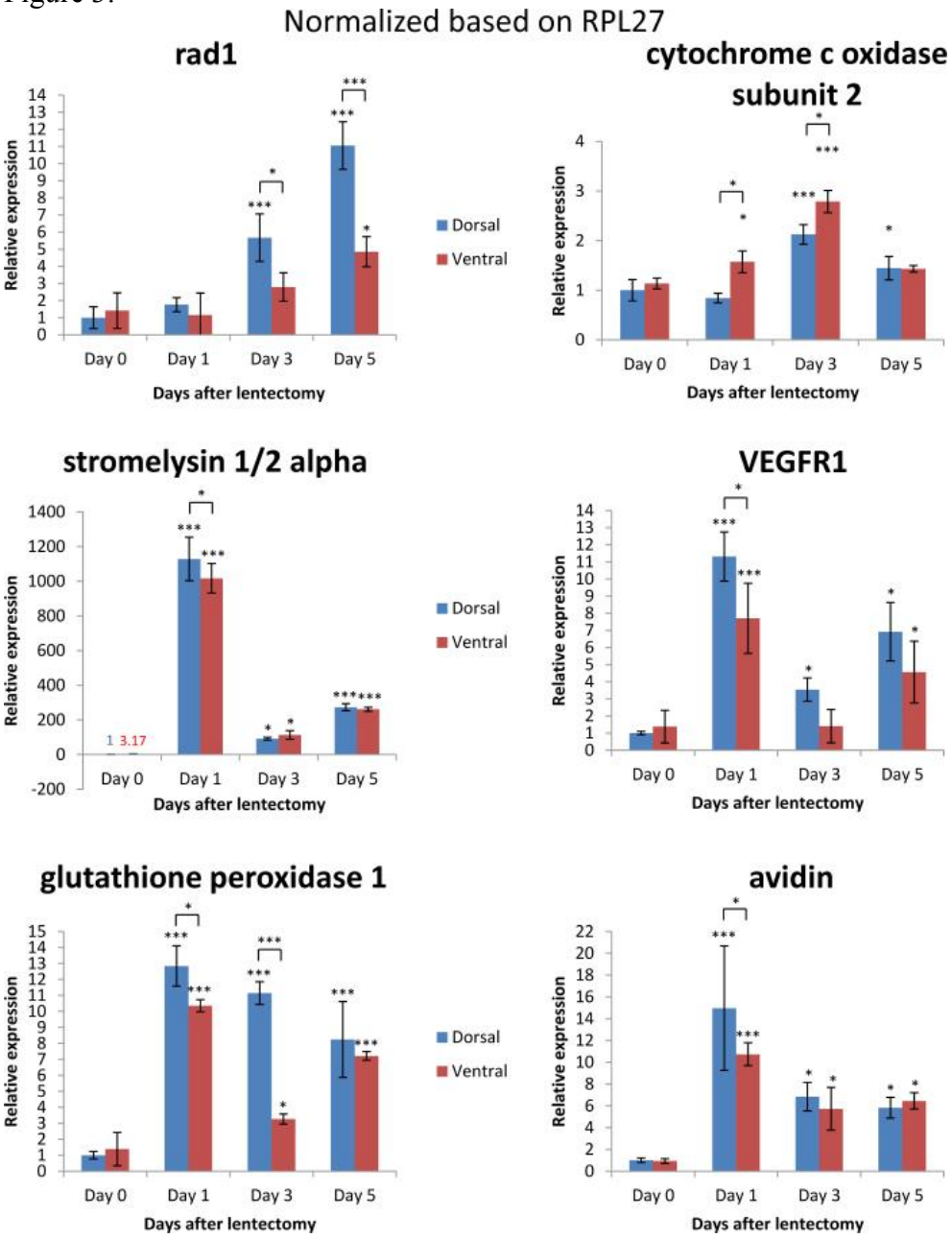


stathmin 1 oncoprotein
myeloid leukemia factor 1
myelin p2 protein
high-mobility group nucleosome binding domain
gamma-secretase subunit pen-2
cysteine-rich protein 1

svil protein
rtf1 protein
origin recognition complex subunit 2
neurofilament heavy polypeptide
lysozyme g
immunoglobulin m heavy chain constant region
high choriolytic enzyme 1 precursor
hemoglobin subunit beta-1

Expression comparison among dorsal/ventral iris in selected time points. **A**: Venn diagram for contigs consistently up- or downregulated in dorsal or ventral iris during all the time points. D up: Contigs upregulated in the dorsal iris during all the time points. D down: Contigs downregulated in the dorsal iris during all the time points. V up: Contigs upregulated in the ventral iris during all the time points. V down: Contigs downregulated in the ventral iris during all the time points. **B**: Venn diagrams for contigs that are downregulated in the dorsal iris and the ventral iris at day 1, and upregulated in the dorsal iris at day 5, and the opposite. Annotated genes are included below each Venn graph. D: dorsal iris; V: ventral iris.

Figure 3:



Relative expression of rad1, cytochrome oxidase subunit 2, stromelysin 1/2 alpha, vascular endothelial growth factor receptor 1 (VEGFR1), glutathione peroxidase 1, and avidin genes after quantitative real-time (qRT)-PCR 0, 1, 3, and 5 days after lentectomy for dorsal and ventral pigmented epithelial cells (PECs) normalized with housekeeping gene ribosomal protein large 27 (RPL27). One asterisk (*) indicates p value smaller than 0.05 ($p<0.05$). Three asterisks (***) indicate p value smaller than 0.001 ($p<0.001$). Asterisk located above the expression columns indicates significance between the sample at a particular time point and the intact iris. Asterisk located above the black line indicates significance between the samples of the same day. The statistical analysis used is the Student t test. Bars on graph indicate standard deviation.

Table 1

List of primers for genes tested by qRT-PCR and annealing temperatures used for their respective target genes.

Genes	Annealing temperature – band size
RPL27	55 °C - ~100 bp
Forward 5′ – 3′	ATTTATGAAACCCGGGAAGG
Reverse 5′ – 3′	CCAGGGCATGACTGTAAGGT
GAPDH	55 °C – 509 bp
Forward 5′ – 3′	GCCTCCTGTACTACCAACTG
Reverse 5′ – 3′	CCCCACTCGTTGTCATACCA
Cytochrome c oxidase subunit 2	61 °C – 355 bp
Forward 5′ – 3′	ACACTAACGCAATAGACGCACAAGA
Reverse 5′ – 3′	ACGCCCATTGAGGGGACTGC
Rad1	61 °C – 341 bp
Forward 5′ – 3′	TGCGTGCCTCGACAACGTCC
Reverse 5′ – 3′	TCACCACCCCACCTTCCTCCA
Stromelysin 1/2 a	61 °C – 168 bp
Forward 5′ – 3′	GGGGGACAAAGACTCTCCCCGA
Reverse 5′ – 3′	CTGGTGTTCTTCAGTGTCCGGGT
VEGFR1	61 °C – 227 bp
Forward 5′ – 3′	TCCTGCAGCAGCCTGACCTTG
Reverse 5′ – 3′	GTTTGGGGCTGTGACTCGGC
Avidin	60 °C – 103 bp
Forward 5′ – 3′	TCGTTTCTCCTCTGACGGGCT
Reverse 5′ – 3′	TGCCCTGCCCAGGTGGTGAT
Glutathione peroxide 1	60 °C – 104 bp
Forward 5′ – 3′	GCTGGTGGTGCTGGGCTTCC
Reverse 5′ – 3′	ACCCTTTTCCTGGACGAACGTACT

Table 2

List of annotated contigs found in cluster D with sufficient sequence length.

Contigname	Accession	Description	Organism	Score	E value	Length
Contig_122	ADJ80991.1	CR1–2	Lycodichthys dearborni	90.5	1e-17	638
Contig_180	NP_001006347.1	OTU domain-containing protein 6B	Gallus gallus	152	4e-42	493
Contig_185	EGV97740.1	Pol polyprotein	Cricetulus griseus	75.5	4e-15	381
Contig_187	XP_002737624.1	PREDICTED: polyprotein-like	Saccoglossus kowalevskii	128	5e-30	755
Contig_2	ACJ43736.1	NADH dehydrogenase subunit 5	Notophthalmus viridescens	390	6e-129	863
Contig_207	XP_003216825.1	PREDICTED: transmembrane protein 205-like	Anolis carolinensis	230	6e-73	672
Contig_215	XP_002942833.1	PREDICTED: androgen-induced gene 1 protein-like	Xenopus (Silurana) tropicalis	201	8e-61	761
Contig_258	ACO09746.1	Lysozyme g	Osmerus mordax	238	2e-76	653
Contig_277	ABI93642.1	suppression of tumorigenicity 7, isoform a, 5 prime	Neofelis nebulosa	330	3e-110	808
Contig_287	XP_003221516.1	PREDICTED: 60S ribosomal protein L9-like	Anolis carolinensis	248	4e-81	454
Contig_302	XP_418054.2	PREDICTED: prolargin	Gallus gallus	271	5e-86	791
Contig_34	XP_002193668.1	PREDICTED: eukaryotic translation initiation factor 2B, subunit 3 gamma	Taeniopygia guttata	325	6e-106	761
Contig_399	NP_990623.1	GTPase HRas precursor	Gallus gallus	338	4e-115	721

Contigname	Accession	Description	Organism	Score	E value	Length
Contig_46	BAE25396.1	unnamed protein product	Mus musculus	122	3e-28	1106

Table 3

Selected contigs related to cell cycle, proliferation and DNA repair. Values are log_2fc

Function	Contig Annotation	Dorsal Day 1	Dorsal Day 3	Dorsal Day 5	Ventral Day 1	Ventral Day 3	Ventral Day 5
Cell cycle, proliferation and DNA repair	bccip homolog	-	−0.17	1.60	-	-	-
	myeloid leukemia factor 1	−1.63	−0.12	0.13	−0.14	-	-
	peroxiredoxin-1	2.13	1.85	2.65	2.02	2.09	1.78
	vascular endothelial growth factor receptor 1	−1.31	-	−0.18	-	-	0.14
	26s proteasome complex subunit dss1	−0.83	0.55	−4.14	−2.42	−2.14	0.10
	tmsb4×protein	1.11	1.62	1.76	0.83	1.49	0.93
	tubulin beta 6	1.86	2.32	1.90	2.50	1.89	1.81
	cell cycle checkpoint protein rad1-like	−1.42	−0.12	−1.77	-	0.95	2.07
	cyclin-dependent kinase 7	-	-	−0.12	−1.82	-	-

Table 4:

Selected contigs related redox homeostasis and mitochondria. Values are log_2fc

Function	Contig Annotation	Dorsal Day 1	Dorsal Day 3	Dorsal Day 5	Ventral Day 1	Ventral Day 3	Ventral Day 5
redox homeosta	glutathione	2.21	1.96	1.40	1.82	1.74	1.29

Function	Contig Annotation	Dorsal Day 1	Dorsal Day 3	Dorsal Day 5	Ventral Day 1	Ventral Day 3	Ventral Day 5
sis	peroxidase 1						
	peroxiredoxin-mitochondrial-like	0.85	0.49	0.28	1.69	1.04	0.14
	peroxiredoxin-1	2.13	1.85	2.65	2.02	2.09	1.78
	sh3 domain-binding glutamic acid-rich-like protein 3	0.39	0.86	1.55	0.56	0.96	0.78
	thioredoxin	1.38	0.74	0.37	0.80	0.39	0.12
	redox-regulatory protein fam213a isoform 1 precursor	-	0.45	1.47	0.29	-	0.31
	peroxiredoxin 6	−0.27	−0.72	−1.39	−0.35	−0.71	−0.51
mitochondria	a chain crystal structures of transition state analog inhibitors of inosine monophosphate cyclohydrolase	−2.11	−1.06	−2.08	−2.08	−1.42	−0.52
	acetyl-mitochondrial precursor	−0.33	-	−0.65	−1.34	−1.33	−0.61

Function	Contig Annotation	Dorsal Day 1	Dorsal Day 3	Dorsal Day 5	Ventral Day 1	Ventral Day 3	Ventral Day 5
	adp atp translocase 1	−1.39	−0.33	−0.30	−0.67	−0.64	−0.09
	isocitrate dehydrogenase	−0.47	−1.56	−1.40	−0.41	−1.83	−1.71
	succinate dehydrogenase	−2.54	-	−0.11	−0.23	−0.30	-
	cytochrome b-c1 complex subunit mitochondrial-like	−2.59	0.24	0.42	0.30	0.29	0.11
	cytochrome c oxidase subunit 2	1.09	0.62	1.08	−0.51	0.38	−0.37
	cytochrome c oxidase subunit 3	-	0.42	1.81	-	0.80	-
	cytochrome c oxidase subunit iv isoform 1	0.33	0.64	1.34	0.18	0.56	−0.15
	cytochrome testis-specific	1.12	1.20	0.53	0.55	1.30	0.88

2.3. Veröffentlichung Nr. 3 Koautor:

„A *de novo* assembly of the newt transcriptome combined with proteomic validation identifies new protein families expressed during tissueregeneration"

Mario Looso1*, Jens Preussner1*, Konstantinos Sousounis3, Marc Bruckskotten1, Christian S Michel1, Ettore Lignelli1, Richard Reinhardt2, Sabrina Höffner1, Marcus Krüger1, Panagiotis A Tsonis3§, Thilo Borchardt1§ and Thomas Braun1§

1 Max-Planck-Institute for Heart and Lung Research, Ludwigstrasse 43, 61231 Bad Nauheim,
Germany

2 Max-Planck Genome Centre Cologne, Carl-von-Linné-Weg 10, 50829 Köln, Germany

3 Department of Biology and Center for Tissue Regeneration and Engineering at Dayton,
University of Dayton, OH 45469-2320, USA

*These authors contributed equally to this work

§Corresponding author

Email addresses: Thomas.Braun@mpi-bn.mpg.de / Thilo.Borchardt@mpi-bn.mpg.de / ptsonis1@udayton.edu

Veröffentlicht in:
Genome Biology 2013, 14:R16 doi:10.1186/gb-2013-14-2-r16; PMID:23425577

2.3.1. Inhaltsangabe

In der vorliegenden Veröffentlichung wurde erstmalig ein fast vollständiges Transkriptom für den Molch erstellt und dabei neue Proteinfamilien gefunden, die in der Regeneration reguliert werden. Dadurch wird der bislang schwierige Zugang zum Transkriptom des Molches deutlich erleichtert.

2.3.2. Eigenanteil

Um ein möglichst vollständiges Transkriptom für *N. viridescens* zu erhalten, wurden bereits vorhandene und neu sequenzierte Datensätze in der vorliegenden Veröffentlichung neu assembliert und mit Hilfe massenspektrometrsich erhaltener

Daten auf einen Protein kodierenden Charakter hin validiert. In Ermangelung eines sequenzierten Genoms, bestand daher die Möglichkeit mit Hilfe von identifizierten Peptiden zwischen kodierenden und nicht kodierenden RNA Sequenzen zu unterscheiden. Dadurch konnte die Entdeckung neuer, zuvor unbekannter Protein Familien erzielt werden, die in unterschiedlichen Geweben und während verschiedener Regenerationsprozesse exprimiert werden.
Zusammen mit Herrn Preussner evaluierte ich das neu assemblierte Transkriptom auf PCR-Ebene mit 32 zufällig ausgewählten Sequenzen, deren Sequenzlängen alle übereinstimmten. Dies ist ein Beleg für die akkurate Assemblierung der Sequenzen und der erhaltenen Contigs. Auf Basis der bioinformatisch gefundenen neuen Protein Familien untersuchte ich deren Expressionsmuster auf mRNA Ebene in unterschiedlichen ungeschädigten Geweben (Herz, Gehirn, Milz, Auge, Leber, Schwanz, Bein, Lunge) mit Hilfe von qRT-PCRs. Im Anschluss erfolgte in Zusammenarbeit mit Ettore Lignelli und Konstantinos Sousounis die Erstellung eines mRNA Expressionsmusters der neu gefundenen Proteinfamilien in regenerierenden Herzen, Linsen und Beinen.

Kurz-Zusammenfassung der durchgeführten Experimente:

- Durchführung von Operationen am Molch und Gewinnung von Gewebematerial
- PCR-Analysen von 32 zufälligen Sequenzen und anschließender Abgleich der Sequenzlängen
- qRT-PCR Analysen der 4 neu gefundenen Protein-Familien und deren mRNA-Expressionsmuster in ungeschädigten Gewebe (Herz, Gehirn, Milz, Auge, Leber, Schwanz, Bein, Lunge); sowie in Kooperation die Erstellung eines mRNA-Expresionsmusters in regenerierenden Geweben (Herz, Linse, Bein).

Unterschrift des Betreuers bezüglich der Angaben zum Eigenanteil:

_____ _____

Ort, Datum Unterschrift Prof. Dr. Dr. T. Braun

2.3.3. Originalpublikation

"A *de novo* assembly of the newt transcriptome combined with proteomic validation identifies new protein families expressed during tissueregeneration"

Mario Looso1*, Jens Preussner1*, Konstantinos Sousounis3, Marc Bruckskotten1, Christian S Michel1, Ettore Lignelli1, Richard Reinhardt2, Sabrina Höffner1, Marcus Krüger1, Panagiotis A Tsonis3§, Thilo Borchardt1§ and Thomas Braun1§

1 Max-Planck-Institute for Heart and Lung Research, Ludwigstrasse 43, 61231 Bad Nauheim,
Germany
2 Max-Planck Genome Centre Cologne, Carl-von-Linné-Weg 10, 50829 Köln, Germany
3 Department of Biology and Center for Tissue Regeneration and Engineering at Dayton,
University of Dayton, OH 45469-2320, USA
*These authors contributed equally to this work
§Corresponding author
Email addresses: Thomas.Braun@mpi-bn.mpg.de / Thilo.Borchardt@mpi-bn.mpg.de / ptsonis1@udayton.edu

2.3.3.1. Abstract

Background

Notophthalmus viridescens, an urodelian amphibian, represents an excellent model organismto study regenerative processes, but mechanistic insights into molecular processes drivingregeneration have been hindered by paucity and poor annotation of coding nucleotidesequences. The enormous genome size and the lack of a closely related reference genomehave so far prevented assembly of the urodelian genome.

Results

We describe the *de novo* assembly of the transcriptome of the newt *Notophthalmus viridescens* and its experimental validation. RNA pools covering embryonic and larval

development, different stages of heart, appendage and lens regeneration, as well as acollection of different undamaged tissues were used to generate sequencing datasets on Sanger, Illumina and 454 platforms. Through a sequential *de novo* assembly strategy, hybriddatasets were converged into one comprehensive transcriptome comprising 120,922 nonredundant transcripts with a N50 of 975. 38,384 putative transcripts were annotated and around 15,000 transcripts were experimentally validated as protein coding by massspectrometry based proteomics. Bioinformatical analysis of coding transcripts identified 826proteins specific for urodeles. Several newly identified proteins establish novel proteinfamilies based on the presence of new sequence motifs without counterparts in public databases, while others containing known protein domains extend already existing families and also constitute new ones.

Conclusions

We demonstrate that our multistep assembly approach allows *de novo* assembly of the newt transcriptome with an annotation grade comparable to well characterized organisms. Our data provide the groundwork for mechanistic experiments to answer the question whether urodeles utilize proprietary sets of genes for tissue regeneration.

Keywords: *de novo* assembly, next generation sequencing, transcriptome, newt, regeneration, high throughput proteomics

2.3.3.2. Background

The regenerative potential of urodele amphibians and especially newts as adult individuals is known for more than 200 years. The complete regeneration of entire appendages [1] is one of the landmark abilities of newts accompanied by their ability to regenerate parts of the central nervous system [2, 3], the lens [4] and the heart (reviewed in [5, 6]). Compared to other animal models [7, 8] the potential of the adult red spotted newt for regeneration is remarkable. Newts do not lose the capacity to regenerate the lens even after repetitive tissue damage that continues over several years. Lenses remain indistinguishable in their molecular signature and morphology even after repetitive rounds of regeneration [9]. In sharp contrast, the ability of mammalian species to regenerate declines rapidly during postnatal life suggesting that the regenerative capacity in mammalians is inversely proportional to the age of an individual. At present, it is still unclear whether regeneration in mammals is a mere extension of embryonic development or represents an independent process. It seems likely that a thorough analysis of the molecular mechanisms of newt tissue regeneration will aid our understanding of regenerative processes and help to develop new therapeutic strategies.

Although the regenerative capability of the newt is extraordinary, it attracted less attention than other model organisms in recent decades. This is partly due to the comparatively long reproductive cycle of newts and their enormous genome size estimated to reach c*1010 bases, which is about 10-times the size of the human genome. Therefore, no genome sequencing approach was initiated so far and only about 140 annotated transcript and protein sequences are available in public databases (NCBI, as from September 2011). To overcome these obstacles, several initiatives were launched to obtain more detailed “omics” data. A set 11000 EST sequences [10] was uploaded to public databases and a mass spectrometry driven proteomics approach was able to identify peptides for more than 1000 newt proteins [11]. Furthermore, we devised a comprehensive newt data depository providing the ability to store, retrieve, link and visualize sequences, proteins and expression data [12]. This repository allows implementation of comprehensive datasets derived from next generation sequencing experiments and high throughput proteomics. Sequencing technologies have seen rapid progress in recent years in respect to the amount of base calls and prices. Despite these advancements and dramatic price cuts the large size of the newt genome still plagues a de novo genome project and makes it hardly affordable. An obvious solution to this problem is the analysis of transcriptomics data, but a detailed analysis of such data is

difficult in the absence of a comprehensive reference dataset. The availability of a detailed reference transcriptome of the newt *N.viridescens* will also yield functional insights and allow identification of new and known proteins that might be instrumental for tissue regeneration of urodelian amphibians.

Here, we present the de novo assembly of the newt transcriptome, based on hybrid sequencing datasets derived from Sanger, 454 Roche and Illumina platforms. Our approach, which generated over 38000 unique transcripts with high quality annotations, covers embryonic and larval development, different stages of heart, appendage and lens regeneration and a comprehensive collection of tissue specific transcripts. To exclude sequencing artefacts and verify coding sequences transcriptome data were matched to a large mass spectrometry derived proteomics dataset resulting in the identification of 14471 newt transcripts with approved protein-coding capacity. Further bioinformatical analysis disclosed several new protein families exclusive to urodelian amphibians of which some contain known domains from public databases, but also entirely new clusters of proteins sharing sequence motifs not known in other species. We reason that some of the proprietary newt proteins might play important roles in regeneration processes unique for urodeles.

2.3.3.3. Results

Library construction and *de novo* assembly strategy

To achieve a broad coverage of the newt transcriptome we used 48537 EST clones of a normalized cDNA library derived from regenerating newt hearts (uninjured, sham, 2h, 6h, 12h, 24h, 48h, 4d, 7d, 14d, 21d, and 35d after mechanical cardiac damage) previously described in [12]. In addition, we generated 807,466 reads from a complex normalized cDNA library by pyrosequencing using a 454 platform with an average read length of 310bp. The normalized library represented all stages of embryonic and larval development, different stages of heart-, appendage- and lens regeneration and a comprehensive collection of transcripts from multiple adult tissues (see methods). Finally, we produced a set of 679,816,626 Illumina paired end reads, (2x60bp, insert size 150 bp), which were derived from a cDNA library of both dorsal and ventral iris during lens regeneration, 4 and 8 days postlentectomy. Next, we evaluated four different approaches to achieve an optimal assembly of different sequence reads, since there is no gold standard for the combination of sequences derived from different sequencing

platforms. Our purpose was to enlarge the N50 (length N for which 50% of all bases in the assembly are located in a transcript of length L < N) and the total number of input sequences, the number of assembled transcripts over 1000 bp and to decrease the number of sequences shorter than 400 bp. The N50 value was only used as a surrogate parameter since we are well aware that N50 values might be affected by the presence of few, very long transcripts. Hence, we also tested effects of the choice of k-mers and the use of a reference mapping strategy, which are known to play crucial roles for the efficiency of an assembly [13].

(i) The first approach was based on initial mapping of 454 and Illumina reads to preassembled Sanger reads (reference mapping) to reduce complexity and redundancy of the datasets. This strategy also enabled us to determine the extent of the new sequence information that was added by 454 and Illumina sequencing. 454 and Illumina reads that remained after mapping were used for individual *de novo* assemblies. Almost 90% of the remaining 454 reads were assembled using MIRA while the extent of assembled Illumina PE reads using Velvet and Oases ranged from 20% to 40% depending on the k-mer choice. All resulting contigs from individual de novo assemblies and preassembled Sanger reads were pooled in a final assembly with 129474 transcripts and a N50 of 776 bp.

(ii) For the second approach, we focused on the PE information of short reads. We tested the influence of unpaired read mapping and subsequent scaffolding on the assembly by mapping all Illumina reads to preassembled Sanger reads (reference mapping) as in the first approach and by subsequent mapping to preassembled 454 contigs (reference mapping). Remaining reads were assembled *de novo* (Velvet and Oases) without considering paired end information. The significantly enlarged number of short contigs (373288 in n8dd pool to 696587 in n8dv pool) was scaffolded by SOPRA. All contigs from Sanger, 454 and scaffolded Illumina reads were then assembled by TGICL/CAP3. This second approach yielded a N50 of 753 bp, including 118416 transcripts.

iii) In the third approach we evaluated the influence of reference datasets on the outcome of the assembly. Since there is no reference genome or transcriptome available for the newt, we used an EST dataset from *Cynops pyrrhogaster,* which is closely related to *N.viridescence*. The EST dataset included 25747 sequences with an average length of 830 bp. Again we mapped our short reads to preliminary assembled Sanger reads and the remaining reads to sequences from *Cynops pyrrhogaster*. Residual non-

mapping reads were assembled *de novo* (Velvet and Oases) and all contigs (assembled 454 reads as well as reads mapped to *Cynops pyrrhogaster*, and de novo assembled Illumina reads) were assembled via TGICL/CAP3 followed by addition of preassembled sanger reads. The resulting assembly had a N50 of 801 bp including 151118 transcripts.

iv) Our final approach, which was eventually used for the generation of the final reference transcriptome, was based on a two-step strategy to reduce redundancies and to minimize computational time required for further analysis. In the first step, each sequence pool was assembled independently without a mapping step (Figure 1a). For the Illumina paired end reads, assembled by Velvet and Oases, we tested several k-mer parameters and continued with the best performing set-up (Figure 1b). The 454 sequence pool was assembled by MIRA, the Sanger sequence pool by CAP3 and MIRA (Figure 1b). In the second step, all resulting contigs irrespective of their length were used for a hybrid assembly performed by TGICL and CAP3 employing mgblast to remove redundancies. This strategy yielded 120922 putative non-redundant transcripts with an N50 of 975 bp (Figure 1c). The last strategy yielded the highest N50, without a significant drop in the number of individual transcripts compared to our other approaches and hence was chosen for all further annotation and verification steps (Additional file 1).

High quality annotation of the transcriptome

The final assembled transcriptome dataset contained 120922 transcripts, which provided a ca. 20-fold higher number of non-redundant assembled transcripts compared to previous studies [10, 12]. These transcripts were annotated by homology searches using the BLAST algorithms. To identify transcripts with a substantial similarity to known sequences, we set the e-value cutoff to e-15, although the total number of annotated sequences with a reasonable similarity dropped significantly compared to higher e-value cutoffs. At least one hit classified as “homology verified” was detected for 38384 individual transcripts. To determine similarities to sequences with known protein coding potential we performed searches against protein and nucleotide databases (NCBI nr and nt). To disclose additional similarities to sequences from organisms that are not included in the above-mentioned databases, such as other urodele amphibians, we also performed searches against EST databases (NCBI EST human, EST mouse and EST others).

To enable a preliminary functional analysis we needed high quality identifiers for identified transcripts. Therefore, we performed searches in Uniprot databases for the species mouse, human and cow, which show good quality of annotation. Additionally, we used Uniprot databases for zebrafish, Xenopus and chicken. The zebrafish served as another model organism for tissue regeneration, whereas Xenopus and chicken are the closest relatives to newts in the evolutionary tree with a substantial number of GO annotated proteins. For these searches, the e-value threshold was set to e-10 since many entries in the Uniprot database are manually curated. We generated functional annotations for 30760 transcripts including all searched species. Taken together, we annotated around 40 % of our complete de novo assembled transcript pool (not filtered for sequence length).

Furthermore, we evaluated the effect of transcript length and e-value cutoff on the rate of sequence annotation (Additional file 2). Not surprisingly, we found that sequence length filtering improved annotation rates considerably. Length filtering also helped to distinguish roughly between short non-coding sequences and sequences with coding potential. We first sorted all transcripts by sequence length and grouped them by bars of 50 bp that were plotted relative to their frequency (Figure 2). In the same graph, we plotted the subset of annotated transcripts relative to sequence length. The rate of 50% annotated transcripts was reached at a sequence length of 290 bp using an e-value threshold of -10 (Figure 2a) (including annotations in NCBI and Ensemble databases). The sequence length corresponding to 50% annotation rate depended on the e-value threshold and was continuously increasing to 320bp for e-15, 360bp for e-20 and 680bp for e-100 (Figure 2 b-d). Based on these findings, we defined a sequence length filter of 400 bp to distinguish between primarily coding sequences and sequences of mostly unknown function. The annotation rate for sequence length >= 400 bp with a threshold e-15 was 56%, which corresponds to 56401 remaining transcripts. This shows that the number of annotated transcripts is enriched among longer transcripts.

Bioinformatical assessment of the coverage of the newly established newt transcriptome

The main purpose of the generation of a reference newt transcriptome was to establish a comprehensive resource for future next generation sequencing and high throughput proteomics approaches. Hence, our transcriptome assembly and further validation by highthroughput protein analysis favoured high quality annotations and not inclusion of

maximum numbers of transcripts and proteins. The coverage of our reference transcriptome was estimated by determining the coverage of general signal transduction pathways and the rate of identified members of known gene families. We analyzed more than 850 signal transduction pathways listed in Biocarta (www.biocarta.com), KEGG [14], Reactome [15] and other databases and determined the rate of coverage using our transcripts with high quality annotations from Uniprot databases. For example, we covered over 80 % of the human p53 signaling pathway (58 components) listed in KEGG (Additional file 3). Similarly, our approach identified more than 80% of all TGF beta signaling pathway members (Additionalfile 4). Similar numbers were reached for other pathways analyzed. The lack of complete coverage might be due to the stringent cut-off criteria used for high quality annotations and/or might be caused by interspecies differences in the number and composition of signaltransduction pathways. Further sequencing projects and refined bioinformatical analysismight solve this conundrum in the future.

To further investigate the fragmentation grade of the transcriptome in the absence of a comprehensive genome data, we assigned orthologues by a recursive best mapping step to all taxa represented in the Uniprot database. We were able to identify 3771 ortholog pairs by sequence similarity and investigated the ortholog pair alignment length. We found newt transcripts that were between ~5% and ~25% longer than the corresponding orthologue sequences. Based on the identified length variation (25%) we assumed that all newt transcripts with alignments >75% of orthologue sequences are full length yielding 2000 (53%) “full length” candidates. The complete list of candidates is presented in Additional file 5.

Experimental validation of the protein coding potential of de novo assembled transcripts by high throughput mass spectrometry

To validate the coding potential of the newly established newt transcriptome, we performed numerous mass spectrometry (reverse-phase nano-LC-MS/MS) experiments and also used mass spectrometry data from earlier studies. Proteins were isolated from various newt tissues including heart [16], lens, tail [11], liver and spleen at different time points during regeneration and from uninjured tissues. Additionally, we isolated proteins from the newt derived myoblast-like cell line A1 [17] during different stages of differentiation into myotubes. The use of SILAC-labeled newt tissues enabled us to filter for mass shifted spectra, which increased the quality of the dataset considerably

[11]. Peptides identified by LCMS/ MS measurements were compared to a protein database generated by reverse translation of all potential coding sequences of the newt transcriptome using all possible reading frames.

We identified 55605 different peptides that matched to 14471 different transcripts, which corresponds to 11.97% of the total number of 120922 non-redundant transcripts. In total 11113 transcripts had at least two peptides. These numbers correspond well to results from similar studies and reflect the lower sensitivity of mass spectrometry based protein detection methods compared to nucleotide sequencing approaches [18]. The transcript with the highest number of peptides (266 peptides, 17373 nucleotides) coded for plectin, a ~4500 AS protein. This protein included 5 frame shifts. In total, 3618 transcripts carried a frame shift as identified by peptide assignments. See Additional file 6 for an example for the identification of ORF shifts.

Interestingly, we were unable to find any similarity for a substantial number of assembled sequences > 400bp to other transcripts or proteins. To estimate the coding potential, all transcripts were translated in six open reading frames (ORF). The longest ORF per transcript was plotted as a function of frequency and compared (i) to a randomly generated dataset containing transcripts of the same number and length and (ii) with transcripts of proved coding potential based on matching peptides identified by mass spectrometry. A substantial number of transcripts from the dataset containing sequences with no similarity exceeded the maximum coding potential of transcripts from the randomly generated control set (Figure 3).

We therefore concluded that the newt genome contains a large number of proprietary protein coding genes with limited similarity to known genes from other organisms. In the future, additional proteomics experiments using isolated cells and subcellular fractions together with the continuous increase of sensitivity and dynamic range of mass spectrometry instruments might allow detection of lower abundant proteins from newt tissues thereby enabling identification of more unknown urodelian-specific proteins.

Identification of new urodelian-specific proteins and protein families

We next wanted to characterize transcripts, which most likely encode new proteins not present in non-urodelian species. Such proteins might be involved in biological processes characteristic for newts such as tissue regeneration or reflect other species-specific properties. We focused on transcripts that either lacked similarities to any entries in public databases or showed sequence similarity exclusively in other urodelian

amphibians. To avoid sequencing artefacts we filtered for sequences that encode for peptides measured by mass spectrometry and hence represent valid protein coding genes.

We identified 583 protein-coding transcripts that did not show any hit in public databases and 243 protein coding transcripts with similarity to urodeles only (Figure 1c). Next, we screened the resulting 826 sequences for conserved domains or motifs using the Pfam database [19] to facilitate assignment of putative protein functions. We identified 145 defined domains within 131 transcripts (Additional file 7) while the remaining 695 transcripts did not contain any known motif. Domains located in urodelianspecific transcripts covered a wide spectrum of known domains, including an activin receptor domain (Figure 4a), which is also present in human mutant transforming growth factorreceptor beta receptor I fragment (A6MIV6_HUMAN). Other examples included the fascin domain (Figure 4b), which characterizes a family of structurally unique actin cross-linking proteins. The diversity of identified motifs suggests that various biological decisions are influenced by urodelian-specific proteins although a precise delineation of potentially affected processes is not possible at present.

To further investigate the tissue-specific distribution of some of these urodelian-specific transcripts we performed RT-PCR using a set of tissues including the heart, brain, spleen, eye, liver, tail, limb and lung. We found an increased expression of the activin receptor domain containing protein in the tail and a moderate expression in the heart, limb and lung (Figure 5a). A more basal expression was detected in the remaining tissues, which suggests that this protein is expressed in muscle containing tissues. To study whether the activin receptor domain containing gene responds to regenerative processes, we analyzed changes in expression during regeneration of the lens and heart by cDNA microarrays. The heart arrays are accessible via the newt repository (http://newtomics.mpi-bn.mpg.de, [12]). A detailed analysis of lens arrays will be published elsewhere [20]. The combined overview of heart and lens arrays is presented in Additional file 8 and Additional file 9. We detected a uniform expression in the regenerating heart tissue, and a slight upregulation in the dorsal and ventral iris (Additional file 10). Further validation by quantitative RT-PCR revealed a significant up regulation of the activin receptor domain containing gene in the regenerating heart 21h after injury and in the dorsal iris 3d after lentectomy (Additional file 11a). Furthermore, we found that two members of the fascin domain containing protein

family were exclusively expressed in the liver whereas another family member was highly expressed in the heart and the lung but only barely detectable in the liver (Figure 5b). Interestingly, we saw a strong expression of the heart/lung fascin domain containing gene during regeneration of the heart and lens with a strong up-regulation 2 days after cardiac injury, which persisted until 35d after injury with an expression peak at ~14d after injury (average ratio 2.7). This pattern was also corroborated by qRT-PCR analysis (Additional file 11b). A similar trend was observed in dorsal iris microarrays where the expression was increased 5 days after lentectomy (average ratio 1.3). The ventral iris showed no up-regulation at any time point (Additional file 10). Since lens regeneration originates from the dorsal iris but not from ventral side, we would like to speculate that this fascin domain containing gene is involved in initiation of the regenerative process. [21].

Next we tried to resolve more protein clusters in the remaining 695 proteins that do not contain any known domains or motifs. Therefore we performed PSI-blast searches (5 iterations, Cutoff 0.05). After manual inspection, we selected several clusters for further investigation. Finally, we used the PRATT tool [22] to scan all sequences as well as manually selected clusters identified by PSI-blast for new patterns not present in public databases such as PROSITE [23]. Using this approach we identified several clusters, which are likely to represent new protein families. One of these clusters consisted of five members represented by complete ORFs and additional 5'/3' UTRs (Figure 4c). All members of this cluster shared a common signal peptide, which indicates that these proteins are secreted [24]. The family is defined by the common motif (L-x(1,3)-C-L-x(2)-[AL]-L-x(3)-[AL]-[AET]-x(2)-[LV]-x- [AS]-[ILV]-x-[DQ]-[LV]-[LV]-C-[AC]-[FIV]-x(3)-[DN]-[EP]-[AIV]-[EK]-x-K-[EN]-x-L). Each sequence was covered by at least two peptides measured by mass spectrometry. All five family members lacked any similarity to known sequences from other urodeles. However, due to the limited sequence information available it is difficult to exclude that related proteins families do also exist in other urodeles. RT-PCR based expression analysis disclosed that all genes of this newly defined group were highly expressed in the tail, four were highly expressed in the limb and strongly transcribed in the liver but not in the limb (Figure 5c).

Additionally, one gene (candidate 1) showed a moderate expression in the eye. No member of this gene family yielded significant signals on our heart and lens microarrays. Quantitative RT-PCR analysis during lens regeneration revealed

significant expression changes of newly identified candidate genes 2-5. Candidate gene 2 was significantly down-regulated in the dorsal but not the ventral iris after lentectomy. Candidate gene 3 was also significantly downregulated in the dorsal iris with a similar trend in the ventral iris. In contrast, candidate 4 was significantly up-regulated in the dorsal iris during regeneration with a peak 3d after lentectomy. Lastly, candidate gene 5 showed a highly interesting, inverse expression pattern in the dorsal and ventral iris (Additional file 11c).

Finally we performed a combined PSIblast and PRATT analysis including all 826 sequences to identify protein clusters consisting of members with and without known domains. From 380 sequences that clustered with at least one other sequence, we identified additional protein clusters of which one example is a set of three sequences (Figure 4d). Two members were characterized by the presence of S100 and EF-hand domains identified by Pfam search, but did not show any similarity to urodele EST sequences while the additional member, which was mapped to the initial sequence by PSI Blast, showed similarities to known ESTs from other urodelian species. All transcripts contained a complete open reading frame (ORF) flanked by 5' and 3' UTR. Sequence similarities within this family were visualized by multiple sequence alignment (MSA) and display of sequence logos [25]. Our mass spectrometric measurements covered all family members with several peptides thereby corroborating the existence of respective proteins. Individual members of this family showed distinct expression patterns (Figure 5d). The first family member (efh_1) was present in virtually all tissues tested with strongest expression in lung and heart. The second family member showed a strong expression in the tail and limb and a moderate expression in the eye (efh_2). The third family member (efh_3) was only present in the tail and limb tissue suggesting a skeletal muscle specific expression pattern. Based on our microarray expressions we found a slight up-regulation of the efh_1 member in the heart 21d after injury (average ratio 0.6) and a down-regulation in the ventral iris (average ratio -1.7). A more detailed expression analysis of efhand family members during heart, limb and lens regeneration revealed a slight up-regulation of Efh1 in regenerating limbs starting from 0d to 14d and a strong increase in the dorsal iris between 1d and 3d (Additional file 11d). Efh2 was downregulated in the ventral iris but strongly up-regulated at 3d in the dorsal iris before its expression decreased during the course of regeneration (Additional file 11d). Finally, Efh3 was significantly up-regulated at 14 d in the regenerating limb, which was mirrored by increased expression in the dorsal iris 1d after injury. In contrast, Efh3 was

dramatically down-regulated after injury in the ventral iris, which does not contribute to the regenerating lens, (Additional file 11d). Nucleotide and amino acid sequences of candidate molecules are given in Additional file 12.

2.3.3.4. Discussion

High throughput sequencing has become an indispensable tool for whole-genome analysis of complex organisms, a trend that has also been fuelled by decreasing sequencing costs. Nevertheless, *de novo* whole-genome analysis is still cost extensive and needs specific bioinformatical expertise and dedicated computational equipment, especially for organisms with very large genomes. Alternatively, *de novo* whole-transcription analysis represents an attractive option to gain detailed insights into the genetic constitution of organism at a fraction of the costs and efforts needed for whole-genome analysis. In fact, this approach has made several organisms amendable for molecular analysis that have not been elicited for genome projects despite the valuable information that might be gained by comprehensive genetic analysis. Examples include exotic insects [26], crustaceans [27], planarians [28] and several more. Transcriptomes of such organisms were decoded by mapping transcripts to a reference genome of phylogenetic close relatives or by *de novo* assembly of transcripts. Assembled transcriptomes are used as reference sets for RNA expression analysis, as matrices for mass spectrometric driven peptide/protein identifications or for phylogenetic analyses. The availability of annotated transcriptomes does also enable RNAseq analyses of niche organisms, which bypasses previous limitations due to the lack of high density microarrays.

The red spotted newt *Notophthalmus viridescens* is an organism with remarkable abilities for tissue regeneration. Although the regenerative potential of this animal has been known since 1712 [29-31] and was intensively investigated in the 1950s to 1980s [32-39], no genome or larger sequencing project has been initiated so far, which is mainly due to the very large genome size (~c x 1010 bp) of newts. Here we describe the first comprehensive de novo assembled transcriptome together with a large-scale experimental validation of coding sequences.

Since the content and quality of a de novo assembled transcriptome strongly depends on the input material, choice of sequencing platforms and bioinformatical processing, we devised an integrated strategy to achieve the best possible outcome. (i) We used

complex RNA pools covering embryonic and larval development, the entire process of heart, appendage and lens regeneration and a comprehensive collection of tissue specific transcripts to limit the problem of incomplete representation of mRNAs. Furthermore, cDNA libraries were normalized to increase the probability to detect mRNAs expressed at low levels or in a small number of cells. (ii) We combined the output from different sequencing platforms, which each have specific advantages: long reads with high accuracy were obtained by Sanger sequencing; medium read length from normalized cDNA libraries at medium depth were generated by the 454 platform and short paired end reads with high sequencing depth from the Illumina platform. (iii) We tested five different assembly strategies in combination with several assembler software packages since most de novo assembly tools are not suited to process data from different platforms and the choice of adjustable parameters of the bioinformatics tool has a significant impact on the success of a de novo assembly. Our assembly strategy, which is based on preassembly of single sequencing sets using custom-configured assemblers followed by subsequent hybrid assembly into one transcriptome outperformed a "reference mapping" like approach based on sequences available for a close relative of the newt (*Cynops pyrrhogaster*). We obtained larger N50 values than in reference mapping approaches without a significant reduction of the overall number of unique transcripts.

The length, distribution and number of de novo generated transcripts give only a rough estimate about the quality of a de novo transcriptome assembly. It is crucial to accomplish a comprehensive annotation of all transcripts and validate putative coding RNAs. To achieve this goal we used multiple BLAST algorithms and fine-tuned e-value thresholds allowing detection of protein coding transcripts with different degrees of evolutionary conservation. The annotation rate of 56% that we achieved is similar to other published transcriptome assemblies (*Schmidtea mediterranea*, 6729 annotated from 18619 transcripts [28]) although we could not rely on comprehensive annotation of closely related organisms. Furthermore, the good coverage (ca. 80%) of components of known signal transduction pathways indicates that we identified the majority of protein coding transcripts that exist in *N. viridescens*.

Nevertheless, we would like to point out that we might have missed transcripts or proteins that are only lowly expressed in tissues studied, which might compromise our goal to achieve a comprehensive list of newt protein-coding genes to a certain extent. In this context it is interesting to note that even well annotated model organisms such as

zebrafish and the mouse still carry a large number of non-annotated transcripts. For example, 40% of all transcripts present on the Affymetrix Mouse 430-2 array have either no gene symbol (5553 probe sets), or only automatically assigned symbols such Riken (3494 probe sets), predicted (506 probe sets), hypothetical (322 probe sets), expressed (505 probe sets), cDNA (175 probe sets) and families with similarity (230 probe sets). The Affymetrix Zebrafish Genome Array shows an even higher rate of almost 45% weakly annotated transcripts. From 15122 entries 2773 missed any gene symbol and 3972 carried only automatically assigned symbols (wu:fa, si:zfos, sb:cb and others). Expression analysis of a selected subset of newly identified genes by RT-PCR and microarray analysis revealed tissue specific expression patterns as a pronounced response to heart and lens regeneration. A more thorough analysis of newly annotated transcripts during tissue renewal will help to define the regulatory network of genes controlling regeneration.

Taken together, our assembly and annotation strategy yielded annotated transcripts that are

close to organisms with complete genome information although the focus of our approach was to obtain transcripts with high quality annotation rather than a maximum number of transcripts. We do not claim that we have assembled and annotated the complete transcriptome since no sufficient sequence similarity fulfilling our significance criteria (e-15) was found for 44% of all assembled sequences (>400bp) but our dataset is sufficient to serve as a matrix for high-resolution expression analyses.

It seems reasonable to assume that a significant part of the non-annotated transcripts represent artefacts generated by misassemblies. Alternatively, such sequences might also represent noncoding RNAs or mRNAs coding for newt- or urodele-specific proteins. However, we did only find a limited number of non-coding RNAs. Screening for non-coding RNAs using the noncode database [40] yielded 17 and 24 hits with high sequence similarity when transcripts <400bp and >400bp were used, respectively. This lead us to the conclusion that a large number of our non-annotated sequences represents weakly conserved 5'/3' UTRs or that the newt owns a large number of RNAs with yet unknown function and similarity to other organisms. Since the lack of information about newt proteins makes it difficult to distinguish between these possibilities we took advantage of a large set proteomics data from newts obtained by mass spectrometry. This approach did not only enable us to validate almost 15,000 transcripts as protein coding but let also to the identification of 826 urodelian-specific proteins, which

showed either no sequence similarity at all (583 transcripts) or similarities to urodele EST sequences only (243 transcripts). We would like to emphasize that these numbers most likely represent only the tip of the iceberg since the current sensitivity and dynamic range of mass spectrometry does only allow measurement of high abundant proteins.

Some of the validated new proteins belong into new protein families, while others contain defined protein motifs (15%) or no discernable primary sequence feature. The detection of new protein families was particularly intriguing and might indicate that founding members of the newly discovered families did further evolve during urodele evolution to cope with species-specific requirements. In fact, it has been postulated that urodeles acquired regenerative capacity at the time when ancestral salamanders separated from the vertebrate tree [41] although more authors prefer the hypothesis that regeneration is a primordial property of metazoa [42], which was lost in most tetrapod vertebrates during evolution but selectively maintained in salamanders. Interestingly, previous studies identified newt-specific genes such as Prod1, a critical determinant of proximodistal identity in the limb bud, that mediates nerve dependent signals to the regenerating blastema by utilizing a conserved signaling machinery [43]. Another example is nsCCN, a newt specific member of the CCN protein family that is exclusively expressed in regenerating but not in uninjured hearts [44].

Further functional studies will reveal whether the newly discovered families are drivers of regenerative processes or serve another yet unknown cause. Of course, our conclusions are based on currently available sequence data leaving the possibility that the newly discovered urodelian-specific proteins exist also in other species, which have not been analysed so far. The newly established *de novo* transcriptome of *N. viridescens* will be an indispensable resource for a better understanding of regenerative events in newts and facilitate the identifications of molecules, conserved or urodele-specific, that control this fascinating process. We would also like to suggest that the combined use of transcriptomics and proteomics approaches provides a powerful means to address new model organisms and detect new protein coding genes.

Conclusions

Despite several obstacles to manipulate and analyse the red spotted newt *Notophthalmus viridescens* at the molecular level, the remarkable abilities of *Notophthalmus viridescens* for tissue regeneration define this organism as an excellent model to study

regeneration. The rapidly evolving techniques for genome/proteome analysis and genetic manipulation will help to understand regenerative processes at a functional level. Here, we describe the first comprehensive de novo assembled transcriptome of *Notophthalmus viridescens* combined with large-scale experimental validation of coding sequences. The use of complex RNA pools and normalized cDNA libraries allowed us to cover different biological processes including embryonic development, as well as heart, appendage and lens regeneration. De novo assembly of the newt transcriptome using different computational strategies was facilitated by combination of the output from different sequencing platforms, which each have specific advantages. The resulting 56% annotation rate of the transcriptome is similar to other transcriptome assemblies (*Schmidtea mediterranea*, 6729 annotated from 18619 transcripts [28]), which are less challenging compared to the newt. Finally, integration of transcriptomics and proteomics data allowed us to confirm the protein coding potential of almost 15,000 transcripts resulting in the identification of 826 urodelian-specific proteins. Several of these newly identified proteins represent new members of defined protein families or completely new protein families and show distinct expression profiles during regeneration. The newly established transcriptome of *N. viridescens* provides a matrix for high-resolution expression analyses and will be an indispensable resource for a better understanding of regenerative processes in newts at a molecular level.

2.3.3.5. Materials and methods

454 Sequencing

Total RNA from regenerating heart (11 timepoints 2 sham timepoints + undamaged), regenerating limb and tail (6 timepoints + undamaged), brain, eye, liver, lung, spleen, kidney, testes, ovaries, 1 cell embryo to larval stage 46 was extracted with Trizol (Invitrogen) following the instructions of the manufacturer. Double stranded cDNA was synthetized with the MINT Kit (Evrogen) and cDNA was normalized with the Trimmer Kit (Evrogen). Library preparation for sequencing was done according to the GS FLX Titanium protocol provided by the manufacturer.

Illumina Paired-End-Sequencing

Total RNA was extracted from the dorsal and ventral part iris four and eight days after lens removal. Ribosomal RNA was depleted using the RiboMinus Eukaryote Kit

(Invitrogen, Carlsbad, CA). 1 µg of depleted RNA was processed according to the Illumina mRNA sample preparation guide. A-tailed DNA was ligated with paired end adaptors using T4-DNA ligase provided by the Illumina RNA-seq kit (Illumina, San Diego, CA). Size selection of adaptor ligated DNA was performed by cutting the target fragment (400bp-450bp) from the DNA gel. Amplification of the cDNA library was obtained by in-gel PCR. Cluster generation and sequencing were performed according to the cluster generation and sequencing manual from Illumina (Cluster Station User Guide and Genome Analyzer Operations Guide). Sequencing was performed by Cofactor Genomics, St Louis, MO 63103, USA.

Sanger sequencing

Total RNA from regenerating heart (11 timepoints 2 sham timepoints + undamaged) was extracted with Trizol (Invitrogen). RNA was reverse transcribed to double stranded cDNA with the SMART method and cDNAs were normalized by the DSN method (Evrogen, Moscow, Russia). After cloning of cDNAs into pDNR-Lib vector, two independent bacterial libraries with more than 100.000 individual clones each were generated. After plating, 100.000 individual bacterial clones were picked and amplified in 96 well plates overnight. cDNA inserts were amplified by colony PCR. Products of PCR reactions were visually inspected on ethidium bromide stained gels and repeated, if failed. Bacterial cultures were cryostocked in 2 replicates into 384 well plates. PCR products were precipitated, washed and resuspended at 200ng/µl in an appropriate spotting buffer (3xSSC, 1,5M Betain) and again checked by visual inspection on ethidium bromide gels. 100.000 cDNA amplicons, including controls were spotted onto 2 sets of glass microarrays (Nexterion slide E, Schott). After microarray hybridization, all spots showing a significant deregulation together with other robustly detected array spots were selected for Sanger sequencing. Around 52.000 individual colonies were selected for reamplification and sequencing, yielding around 48.500 Sanger sequences of high quality. All individual sanger sequences are available via the newt repository newt-omics (*newt-omics.mpi-bn.mpg.de*) [12].

Quality control of sequences & *De Novo* assembly

Base calling for Sanger reads (48537) was performed by Phred. Primary clustering was done by wcd [45], assembly by cap3 [46] . The de novo assembly yielded 26594 unique transcripts. Illumina raw read quality was determined using FastQC

(http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc) quality control tool. Illumina sequencing reads below read quality threshold 20 were trimmed base by base from the 3'-end until the average quality of the read was >20. Paired end sequences, having one read with a length less than 35 were discarded before assembly (Additional file 1). *De novo* assembly of Illumina sequences was performed on each library separately using Velvet [47] and Oases [48]. To choose optimal parameters we evaluated summary statistics like N50, number of contigs and percentage reads assembled for k-mers 23, 25, 29 and 35. The Oases tool was run with a minimum transcript length of 100 bp, insert length of 150 bp and minimum coverage of 5. 454 sequences were de novo assembled by MIRA [49] after quality check and adapter clipping, using the following parameters: job=denovo,est,accurate,454 - fastq -notraceinfo -noclipping - AS:sep=yes:urd=no 454_SETTINGS -AS:mrl=100 - OUT:sssip=yes.

Generation of unique transcripts

To lower the redundancy resulting from individual assemblies, all Illumina transcripts from Oases, sanger transcripts from CAP3 and 454 transcripts from MIRA were pooled in one file. Using the TGICL/CAP3 pipeline [50] pooled transcripts were compared to themselves, using the mgblast (modified version of megablast [51]) algorithm. Clusters were generated with at least 90% sequence identity and a maximum unmatched overhang of 30 bp. For each cluster, all subcluster assembly results, i.e. the biggest transcripts, which are not contained in other sequences, were pooled with singletons and assembled a second time with CAP3. Resulting contigs and singletons from each cluster were stored in a globalContig and globalSinglets file. Remaining clusters with less than 15 transcripts per cluster were assembled together with CAP3 and added to a globalContigs or a globalSinglets file. Concatenation of these two files yielded non-redundant and unknown transcripts of five libraries.

Sequence Annotation and functional assignments

To annotate sequences obtained by *de novo* assembly, we performed sequence similarity searches using the BLAST algorithm. We implemented an automated annotation and quality filter pipeline, using the NCBI BLASTcl3 tool and UNIX shell scripts. The scripts performed blastn, blastx and tblastx searches on NCBI's nucleotide (nt), EST (human, mouse, other), proteins (nr) and HTGS databases. For the tblastx search of the NT database we used a HPC cluster and pipeline hosted at the GWDG (Gesellschaft für

wissenschaftliche Datenverarbeitung mbH Göttingen). We set the e-value threshold to e-15. We performed a quality rating by checking for keywords, which are represented in weak description lines (like 'mRNA','cDNA', 'clone' or 'genomic'). Detected sequence similarities containing one or more of such keywords were marked as low quality hits. We collected at least three top hits per taxon, BLAST algorithm and database. We performed these quality checks for 90 taxa in total. Data are accessible via **http://newt-omics.mpi-bn.mpg.de**. To assign functional annotations to de novo assembled transcripts BLAST searches against GO annotated Uniprot databases (*e*-value threshold < *e*-20) from mouse, human, zebrafish, chicken and cow were performed to cope with the limited Gene Ontology (GO) assignments for amphibians [52]. To avoid redundant functional assignments, the best-rated similarity hit with at least one GO annotation per taxon was chosen.

Peptide identification

Protein samples were isolated from newt heart, lens, tail, liver and spleen at different time points during regeneration or from uninjured tissues. Additionally, proteins from a newt derived myoblast like cell line (A1) during different stages of differentiation into myotubes were processed for mass spectrometry (reverse-phase nano-LC-MS/MS) measurements. Partially, tissues were labeled by stable isotope of amino acids (SILAC) in vivo as described [53]. Heavy amino acid derivates for Lysin (Lys6 and Lys8) and Arginin (Arg10) were used for metabolic labeling. Analysis and of individual mass spectrometry measurements are described elsewhere [53]. To identify mass spectrometry spectra, de novo assembled transcripts were translated into six reading frames to generate an Andromeda search engine compatible database [54]. Only reading frames longer than 25 amino acids were used for further analysis. The maximum false discovery rate was set below 1% for peptide and protein identifications using the DECOY target database approach [55]. For search transmission to Andromeda and peptide clustering, the MaxQuant software package (Version 1.2.0.18) was used [56].

Analysis of unknown, peptide verified sequences

Pfam [19] batch sequence search was used with default parameters to identify known domains within the subset of 826 coding transcripts without similarities to entries in public databases. Results were inspected manually. Detection of protein clusters sharing motifs or domains was done using PSI-Blast. For each sequence, a Blast database was

generated from 825 protein sequences excluding the query sequence. Five iterations of PSI-Blast searches were used with a cutoff value of 0.05. Resulting clusters with at least two sequences were inspected manually to find candidates for new motifs or domains. PRATT [22] was used with parameters –c% 0.6 –FL 4 –FN 3 –PX 2 to search for common patterns within all protein sequences to generate clusters of the size of 6 and maximizing fitness values of refined pattern. Pattern searches within selected clusters generated by PSI-Blast were performed with PRATT using default parameters. Signal peptide sequences were scanned with Signal P [24].

qRT-PCR transcript verification

Total RNA was isolated using TRIzol® reagent (Invitrogen™) or using the GE Healthcare kit (In case of the regenerating lens tissue) according to instructions of the manufacturers. 1µg of total RNA was used for reverse transcription using SuperscriptII®(Invitrogen™) following standard procedures. Real-time PCR was performed using the iCycler (Bio-Rad) and ABsoluteTM QPCR SYBR Green Fluorescein Mix (ABgene®,UK) or the iQTM SYBR® Green Supermix (Bio-rad) (for regenerating lens tissue). Expression levels were normalized based on RP21 or RPL27 (for regenerating lens tissue) housekeeping genes. A list of primers is supplied in the supplementary materials (Additional file 13).

Data access

Supplemental material is available for this article. The Illumina sequence data from this study have been submitted to the NCBI Sequence read archive under the accession number [ERP001353]. All assembled reads are publically available through the newt-omics database (http://newtomics.mpi-bn.mpg.de) and can be downloaded at: http://newt-omics.mpi-bn.mpg.de/De_Novo_Transcriptome_Newt.fasta . Mass spectrometry raw data and 454 reads have been uploaded to the Proteome Commons Tranche repository.

The link and hash key for A1 cell tissue: https://proteomecommons.org/dataset.jsp?id=E9gn3jtHz9%2FQUGK5WlBiB%2BM9oP6WYjljagq0cPNCdAgUvsL3s6NAQ32Kh%2BRkOKtbT22c0aTyEJ4rFq%2BpkdY977I6VdsAAAAAAAABzA%3D%3DE9gn3jtHz9/QUGK5WlBiB+M9oP6WYjljagq0cPNCdAgUvsL3s6NAQ32Kh+RkOKtbT22c0aTyEJ4rFq+pkdY977I6VdsAAAAAAAABzA==

The link and hash key for heart tissue: https://proteomecommons.org/dataset.jsp?id=nSFtHPGpn%2BqPYPRcc3%2FNGKbHFpPoBh5m8fLzbkRXChOdEyGUoLguR0CTQA6F7wF%2FCZ0Z7jdO89t2H2hDjsxz%2FNzINpoAAAAAAAAB1g%3D%3DnSFtHPGpn+qPYPRcc3/NGKbHFpPoBh5m8fLzbkRXChOdEyGUoLguR0CTQA6F7wF/CZ0Z7jdO89t2H2hDjsxz/NzINpoAAAAAAAAB1g==

The link and hash key for tail tissue: https://proteomecommons.org/dataset.jsp?id=Sh9IkJHrsCQLYOeNE2sf7gNiUsxpqGWRi23%2FWiTHFX3dXdNGJDMcqHD9LP9JYSgRc%2BJmHNB3lOnTXX5B1h66cfMigEkAAAAAAAAB0g%3D%3DSh9IkJHrsCQLYOeNE2sf7gNiUsxpqGWRi23/WiTHFX3dXdNGJDMcqHD9LP9JYSgRc+JmHNB3lOnTXX5B1h66cfMigEkAAAAAAAAB0g==

The link and hash key for lens cell tissue: https://proteomecommons.org/dataset.jsp?id=JwZJpuT9w4TPmouUz06eIVvBL726Fid%2BRxA8FNnbSMncIKA05OwOQEIGX8a%2BclxmVr7sSpo29%2BpnGLAmFL%2BPjIhj3vwAAAAAAAAB0A%3D%3DJwZJpuT9w4TPmouUz06eIVvBL726Fid+RxA8FNnbSMncIKA05OwOQEIGX8a+clxmVr7sSpo29+pnGLAmFL+PjIhj3vwAAAAAAAAB0A==

The link and hash key for other mixed tissue: https://proteomecommons.org/dataset.jsp?id=hPSvp96O5idgvaZy9R253%2FJAyYP5Qu%2Be%2Fw%2Fvej0Oq79ZDdJcmsynGVXIt%2F50mwx3eQDAj1UF726EmbeDtPUPDuWETb8AAAAAAAAB0A%3D%3DhPSvp96O5idgvaZy9R253/JAyYP5Qu+e/w/vej0Oq79ZDdJcmsynGVXIt/50mwx3eQDAj1UF726EmbeDtPUPDuWETb8AAAAAAAAB0A==

The link and hash key for 454 sequencing reads: https://proteomecommons.org/tranche/datadownloader.jsp?h=kdQqcmlwPViqk6Ep200pzjJnt6C73EkE5HpvvfSoAeVoK7t5iRfsC9jND%2B0jTUBOo6SnYZELlsh9PwMDd9RD94UYgA8AAAAAAAAB4g%3D%3DkdQqcmlwPViqk6Ep200pzjJnt6C73EkE5HpvvfSoAeVoK7t5iRfsC9jND+0jTUBOo6SnYZELlsh9PwMDd9RD94UYgA8AAAAAAAAB4g==

Abbreviations

EST: expressed sequence tag; GO: gene ontology; HPC: high performance computing; HTGS: high throughput genome sequencing; LC-MS/MS: liquid chromatography

coupled to a tandem mass spectrometry; MSA: multiple sequence alignment; NCBI: national center for biotechnology information; ORF:open reading frame; PE: paired end; SILAC:stable isotope labeling by amino acids in cell culture; UTR: untranslated region

Authors' contributions

ML, PAT, TBo and TB co-wrote the paper. ML, JP and MB performed the de novo assembly and bioinformatical analysis. KS contributed to the illumina sequencing, ML, TBo and RR contributed to the 454 and Sanger sequencing. CSM, EL and KS performed tissue preparation and RT-PCR experiments. ML, SH and MK contributed to the masspec analysis and peptide identification. MK, PAT, TBo and TB gave technical advice and contributed to the study design. All authors read and approved the final manuscript.

Acknowledgements

This work was supported by the Max-Planck-Society, the German Research Council (DFG), the Excellence Cluster Cardiopulmonary System (ECCPS), the University of Giessen- Marburg Lung Center (UGMLC), the Cell and Gene Therapy Center (CGT) of the University of Frankfurt and the EY10540 grant to PAT.

2.3.3.6. References

1. Brockes JP, Kumar A: **Appendage regeneration in adult vertebrates andimplications for regenerative medicine.** *Science* 2005, **310:**1919-1923.
2. Berg DA, Kirkham M, Beljajeva A, Knapp D, Habermann B, Ryge J, Tanaka EM, Simon A: **Efficient regeneration by activation of neurogenesis in homeostatically quiescent regions of the adult vertebrate brain.** *Development* 2010, **137:**4127-4134.
3. Berg DA, Kirkham M, Wang H, Frisen J, Simon A: **Dopamine controls neurogenesis in the adult salamander midbrain in homeostasis and during regeneration of dopamine neurons.** *Cell Stem Cell* 2011, **8:**426-433.
4. Tsonis PA: **Regeneration of the lens in amphibians.** *Results Probl Cell Differ* 2000, **31:**179-196.
5. Borchardt T, Braun T: **Cardiovascular regeneration in non-mammalian model systems: what are the differences between newts and man?** *Thromb Haemost* 2007, **98:**311-318.
6. Ausoni S, Sartore S: **From fish to amphibians to mammals: in search of novel strategies to optimize cardiac regeneration.** *J Cell Biol* 2009, **184:**357-364.
7. Tanaka EM, Reddien PW: **The cellular basis for animal regeneration.** *Dev Cell* 2011, **21:**172-185.
8. Sanchez Alvarado A, Tsonis PA: **Bridging the regeneration gap: genetic insights from diverse animal models.** *Nat Rev Genet* 2006, **7:**873-884.
9. Eguchi G, Eguchi Y, Nakamura K, Yadav MC, Millan JL, Tsonis PA: **Regenerative capacity in newts is not altered by repeated regeneration and ageing.** *Nat Commun* 2011, **2:**384.
10. Borchardt T, Looso M, Bruckskotten M, Weis P, Kruse J, Braun T: **Analysis of newly established EST databases reveals similarities between heart regeneration in newt and fish.** *BMC Genomics* 2010, **11:**4.
11. Looso M, Borchardt T, Kruger M, Braun T: **Advanced identification of proteins in uncharacterized proteomes by pulsed in vivo stable isotope labeling-based mass spectrometry.** *Mol Cell Proteomics* 2010, **9:**1157-1166.
12. Bruckskotten M, Looso M, Reinhardt R, Braun T, Borchardt T: **Newt-omics: a comprehensive repository for omics data from the newt Notophthalmus viridescens.** *Nucleic Acids Res* 2011.
13. Surget-Groba Y, Montoya-Burgos JI: **Optimization of de novo transcriptome assembly from next-generation sequencing data.** *Genome Res* 2010, **20:**1432-1440.
14. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M: **KEGG for integration and interpretation of large-scale molecular data sets.** *Nucleic Acids Res* 2012, **40:**D109-114.
15. Croft D, O'Kelly G, Wu G, Haw R, Gillespie M, Matthews L, Caudy M, Garapati P, Gopinath G, Jassal B, Jupe S, Kalatskaya I, Mahajan S, May B, Ndegwa N, Schmidt E, Shamovsky V, Yung C, Birney E, Hermjakob H, D'Eustachio P, Stein L: **Reactome: a database of reactions, pathways and biological processes.** *Nucleic Acids Res* 2011, **39:**D691-697.
16. Looso M, Michel CS, Konzer A, Bruckskotten M, Borchardt T, Kruger M, Braun T: **Spiked-in Pulsed in Vivo Labeling Identifies a New Member of the CCN Family in Regenerating Newt Hearts.** *J Proteome Res* 2012.
17. Tanaka EM, Gann AA, Gates PB, Brockes JP: **Newt myotubes reenter the cell cycle by phosphorylation of the retinoblastoma protein.** *J Cell Biol* 1997, **136:**155-165.

18. Drexler HC, Ruhs A, Konzer A, Mendler L, Bruckskotten M, Looso M, Gunther S, Boettger T, Kruger M, Braun T: **On marathons and sprints: an integrated quantitative proteomics and transcriptomics analysis of differences between slow and fast muscle fibers.** *Mol Cell Proteomics* 2011.
19. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer EL, Eddy SR, Bateman A, Finn RD: **The Pfam protein families database.** *Nucleic Acids Res* 2012, **40:**D290- 301.
20. Sousounis K, Michel CS, Bruckskotten M, Maki N, Borchardt T, Braun T, Looso M, Tsonis PA: **A microarray analysis of gene expression patterns during early phases of newt lens regeneration.** *Mol Vis* 2013, **19:**135-145.
21. Tsonis PA, Makarev E: **On dorsal/ventral-specific genes in the iris during lens regeneration.** *Cell Mol Life Sci* 2008, **65:**41-44.
22. Jonassen I, Collins JF, Higgins DG: **Finding flexible patterns in unaligned protein sequences.** *Protein Sci* 1995, **4:**1587-1595.
23. Sigrist CJ, Cerutti L, de Castro E, Langendijk-Genevaux PS, Bulliard V, Bairoch A, Hulo N: **PROSITE, a protein domain database for functional characterization and annotation.** *Nucleic Acids Res* 2010, **38:**D161-166.
24. Petersen TN, Brunak S, von Heijne G, Nielsen H: **SignalP 4.0: discriminating signal peptides from transmembrane regions.** *Nat Methods* 2011, **8:**785-786.
25. Crooks GE, Hon G, Chandonia JM, Brenner SE: **WebLogo: a sequence logo generator.** *Genome Res* 2004, **14:**1188-1190.
26. Poelchau MF, Reynolds JA, Denlinger DL, Elsik CG, Armbruster PA: **A de novo transcriptome of the Asian tiger mosquito, Aedes albopictus, to identify candidate transcripts for diapause preparation.** *BMC Genomics* 2011, **12:**619.
27. Zeng V, Villanueva KE, Ewen-Campen BS, Alwes F, Browne WE, Extavour CG: **De novo assembly and characterization of a maternal and developmental transcriptome for the emerging model crustacean Parhyale hawaiensis.** *BMC Genomics* 2011, **12:**581.
28. Adamidi C, Wang Y, Gruen D, Mastrobuoni G, You X, Tolle D, Dodt M, Mackowiak SD, Gogol-Doering A, Oenal P, Rybak A, Ross E, Sanchez Alvarado A, Kempa S, Dieterich C, Rajewsky N, Chen W: **De novo assembly and validation of planaria transcriptome by massive parallel sequencing and shotgun proteomics.** *Genome Res* 2011, **21:**1193-1200.
29. Reaumur R-AFd: **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.** 1712.
30. Trembley A: **Memoires pour servir a l'histoire d'un genre de polypes d'eau douce.** 1744.
31. Spallanzani L: **Prodromo di un opera da imprimersi sopra la riproduzioni animali.** 1768.
32. Sidman RL, Singer M: **Stimulation of forelimb regeneration in the newt, Triturus viridescens, by a sensory nerve supply isolated from the central nervous system.** *Am J Physiol* 1951, **165:**257-260.
33. Hay ED: **Effects of thyroxine on limb regeneration in the newt, Triturus viridescens.** *Bull Johns Hopkins Hosp* 1956, **99:**262-286.
34. Carlson BM: **The histology of inhibition of limb regeneration in the newt, Triturus, by actinomycin D.** *J Morphol* 1967, **122:**249-263.
35. Woolfitt R: **Lens regeneration from iris implants into the blastema of a regenerating newt limb.** *W V Med J* 1968, **64:**387.

36. Dresden MH: **Denervation effects on newt limb regeneration: DNA, RNA, and protein synthesis.** *Dev Biol* 1969, **19:**311-320.
37. Oberpriller J, Oberpriller JC: **Mitosis in Adult Newt Ventricle.** *J Cell Biol* 1971, **49:**560-563.
38. Reyer RW, Woolfitt RA, Withersty LT: **Stimulation of lens regeneration from the newt dorsal iris when implanted into the blastema of the regenerating limb.** *Dev Biol* 1973, **32:**258-281.
39. Bader D, Oberpriller JO: **Repair and reorganization of minced cardiac muscle in the adult newt (Notophthalmus viridescens).** *J Morphol* 1978, **155:**349-357.
40. Liu C, Bai B, Skogerbo G, Cai L, Deng W, Zhang Y, Bu D, Zhao Y, Chen R: **NONCODE: an integrated knowledge database of non-coding RNAs.** *Nucleic Acids Res* 2005, **33:**D112-115.
41. Garza-Garcia AA, Driscoll PC, Brockes JP: **Evidence for the local evolution of mechanisms underlying limb regeneration in salamanders.** *Integr Comp Biol* 2010, **50:**528-535.
42. Bely AE, Nyberg KG: **Evolution of animal regeneration: re-emergence of a field.** *Trends Ecol Evol* 2010, **25:**161-170.
43. Kumar A, Godwin JW, Gates PB, Garza-Garcia AA, Brockes JP: **Molecular basis for the nerve dependence of limb regeneration in an adult vertebrate.** *Science* 2007, **318:**772-777.
44. Looso M, Michel CS, Konzer A, Bruckskotten M, Borchardt T, Kruger M, Braun T: **Spiked-in pulsed in vivo labeling identifies a new member of the CCN family in regenerating newt hearts.** *J Proteome Res* 2012, **11:**4693-4704.
45. Hazelhurst S, Hide W, Liptak Z, Nogueira R, Starfield R: **An overview of the wcd EST clustering tool.** *Bioinformatics* 2008, **24:**1542-1546.
46. Huang X, Madan A: **CAP3: A DNA sequence assembly program.** *Genome Res* 1999, **9:**868-877.
47. Zerbino DR: **Using the Velvet de novo assembler for short-read sequencing technologies.** *Curr Protoc Bioinformatics* 2010, **Chapter 11:**Unit 11 15.
48. Schulz MH, Zerbino DR, Vingron M, Birney E: **Oases: Robust de novo RNA-seq assembly across the dynamic range of expression levels.** *Bioinformatics* 2012.
49. Chevreux B, Pfisterer T, Drescher B, Driesel AJ, Muller WE, Wetter T, Suhai S: **Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs.** *Genome Res* 2004, **14:**1147-1159.
50. Pertea G, Huang X, Liang F, Antonescu V, Sultana R, Karamycheva S, Lee Y, White J, Cheung F, Parvizi B, Tsai J, Quackenbush J: **TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets.** *Bioinformatics* 2003, **19:**651-652.
51. Zhang Z, Schwartz S, Wagner L, Miller W: **A greedy algorithm for aligning DNA sequences.** *J Comput Biol* 2000, **7:**203-214.
52. Barrell D, Dimmer E, Huntley RP, Binns D, O'Donovan C, Apweiler R: **The GOA database in 2009--an integrated Gene Ontology Annotation resource.** *Nucleic Acids Res* 2009, **37:**D396-403.
53. Looso M, Borchardt T, Krueger M, Braun T: **Advanced identification of proteins in uncharacterized proteomes by pulsed in vivo SILAC.** *Mol Cell Proteomics* 2010.
54. Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, Mann M: **Andromeda: a peptide search engine integrated into the MaxQuant environment.** *J Proteome Res* 2011, **10:**1794-1805.

55. Elias JE, Gygi SP: **Target-decoy search strategy for increased confidence in largescale protein identifications by mass spectrometry.** *Nat Methods* 2007, **4:**207-214.
56. Cox J, Mann M: **MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification.** *Nat Biotechnol* 2008, **26:**1367-1372.

2.3.3.7. Figures

Figure 1:

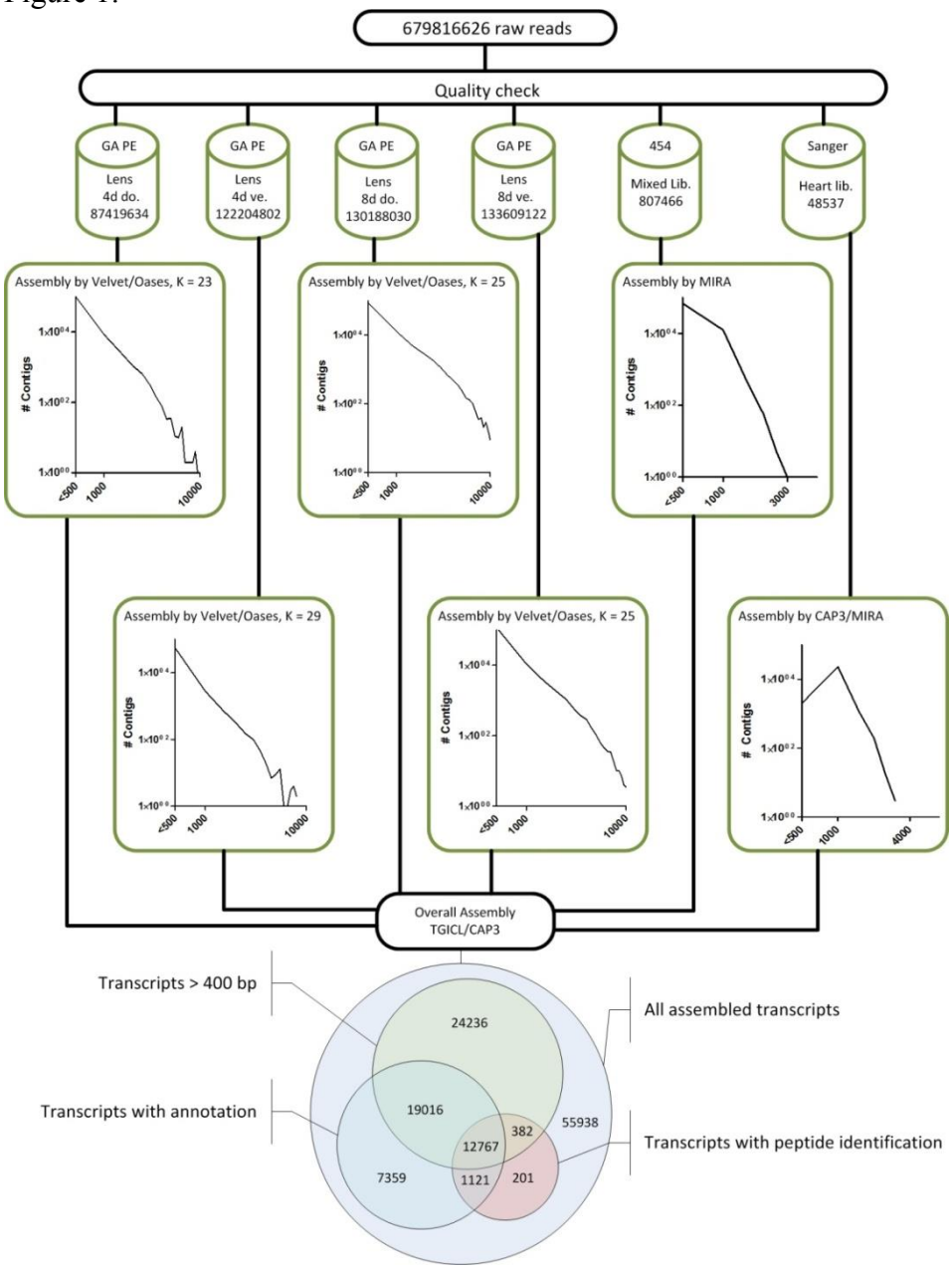


Figure 1: Workflow of the de novo assembly and resulting annotation steps. (A) Raw reads were filtered and assembled for each sequence pool. (B) Optimized assembly strategies for individual Illumina, 454 and Sanger sequence pools illustrated as number of Contig vs. length distribution. Velvet and Oases were used for the Illumina reads, MIRA for the 454 reads and CAP3 and MIRA for Sanger reads. Contigs from individual assemblies were merged by TGICL and CAP3, yielding 120922 unique transcripts (C) Venn diagram of all assembled transcripts, categorized into transcripts larger 400 bp, transcripts with annotation, and transcripts with peptide identification.

Figure 2:

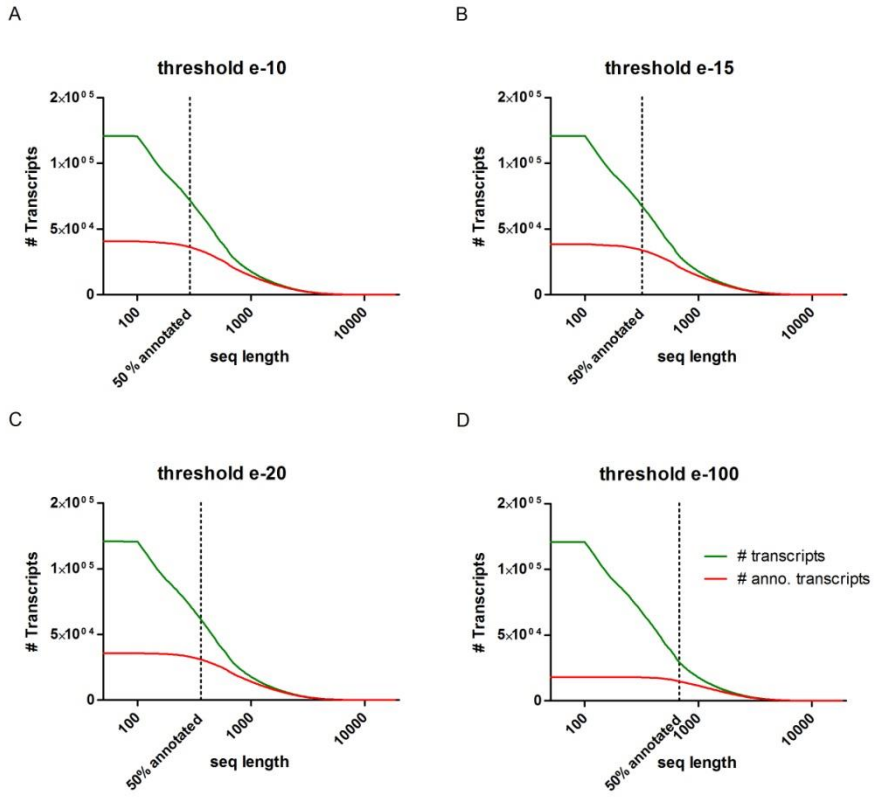


Figure 2: Influence of e-value cutoff on transcript annotation rate. (A-D) Number of annotated transcripts (y-axis, log scale) by cut-off e-value plotted against sequence length (xaxis log scale). The total number of transcripts is shown in green, the subset of annotated transcripts is shown in red. The number of transcripts with an annotation increases with the length of transcripts. The percentage of annotated transcripts depends on the cutoff e-value threshold. For example: an annotation of 50% of transcripts at 290 bp is reached at a cutoff of e-10 (a), this length increases to 680 bp when a cutoff of e-100 is used (d). We selected the minimum sequence length of 400 bp for our assembly to achieve a sufficient annotation rate of assembled sequences.

Figure 3

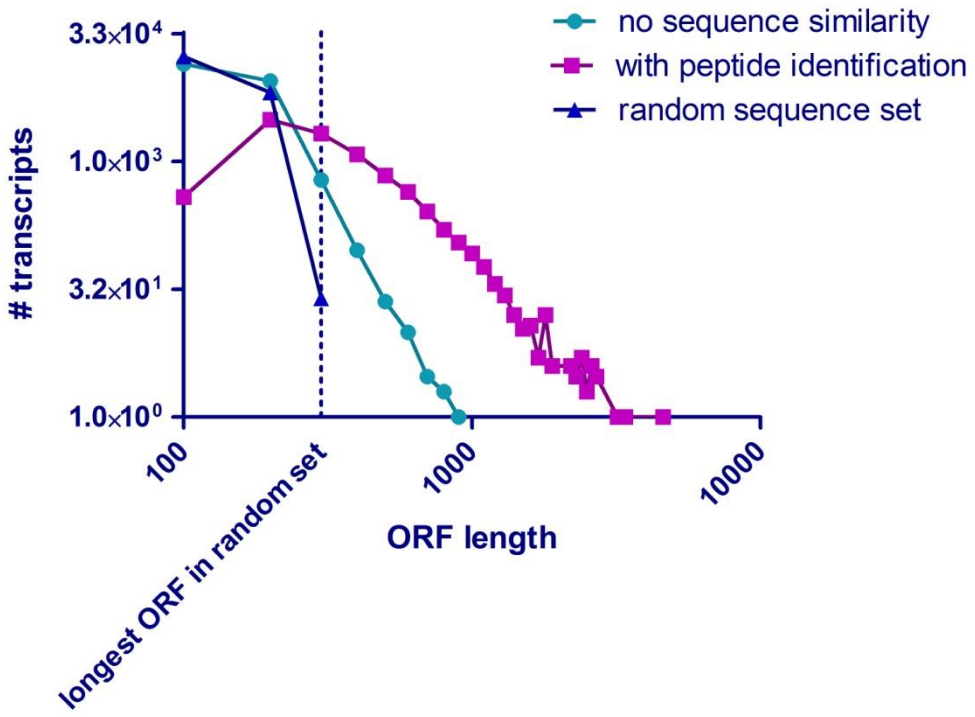


Figure 3: Estimation of coding potential in non-annotated transcripts. ORF coding potential in a randomly generated transcript set, verified transcripts without sequence annotation, and sequences with high quality annotations. All three transcript sets containing sequences longer than 400 bp were translated in 6 reading frames. The longest ORF per transcript was plotted against the number of transcripts per dataset. Note that a significant number of ORFs from the dataset without annotation (cyan line) exceeds the maximum ORF length of the randomly generated transcript. The non-annotated group contains an identical number of total transcripts with identical sequence length (blue line) indicating the existence of potential newt proteins that were not identified by proteomics.

Figure 4

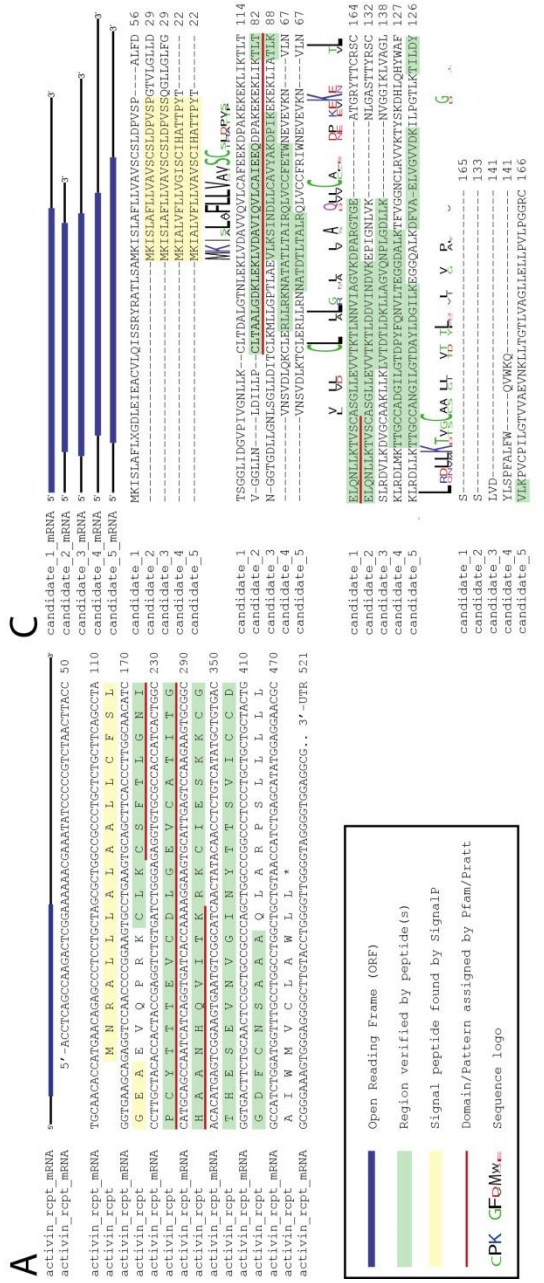


Figure 4: Identification of new urodele specific protein families. Schematic transcript sequence indicates flanking UTRs and included open reading frame with start and stop codon marked in blue. Amino acid regions with peptide coverage are marked in green, significant signal peptides are marked in yellow. Detected Domains or patterns are highlighted by a red underline. Within each MSA, a logo represents sequence conservation **A)** New protein including a Activin Receptor domain. **B)** Protein cluster of 4 members detected by common fascin domain. **C)** Sequence set representing a group of similar proteins detected by common pattern, missing any known domain. **D)** Combined sequence cluster, including one member with Pfam domains S100 and EF-hand like and three members without these domains.

Figure 5

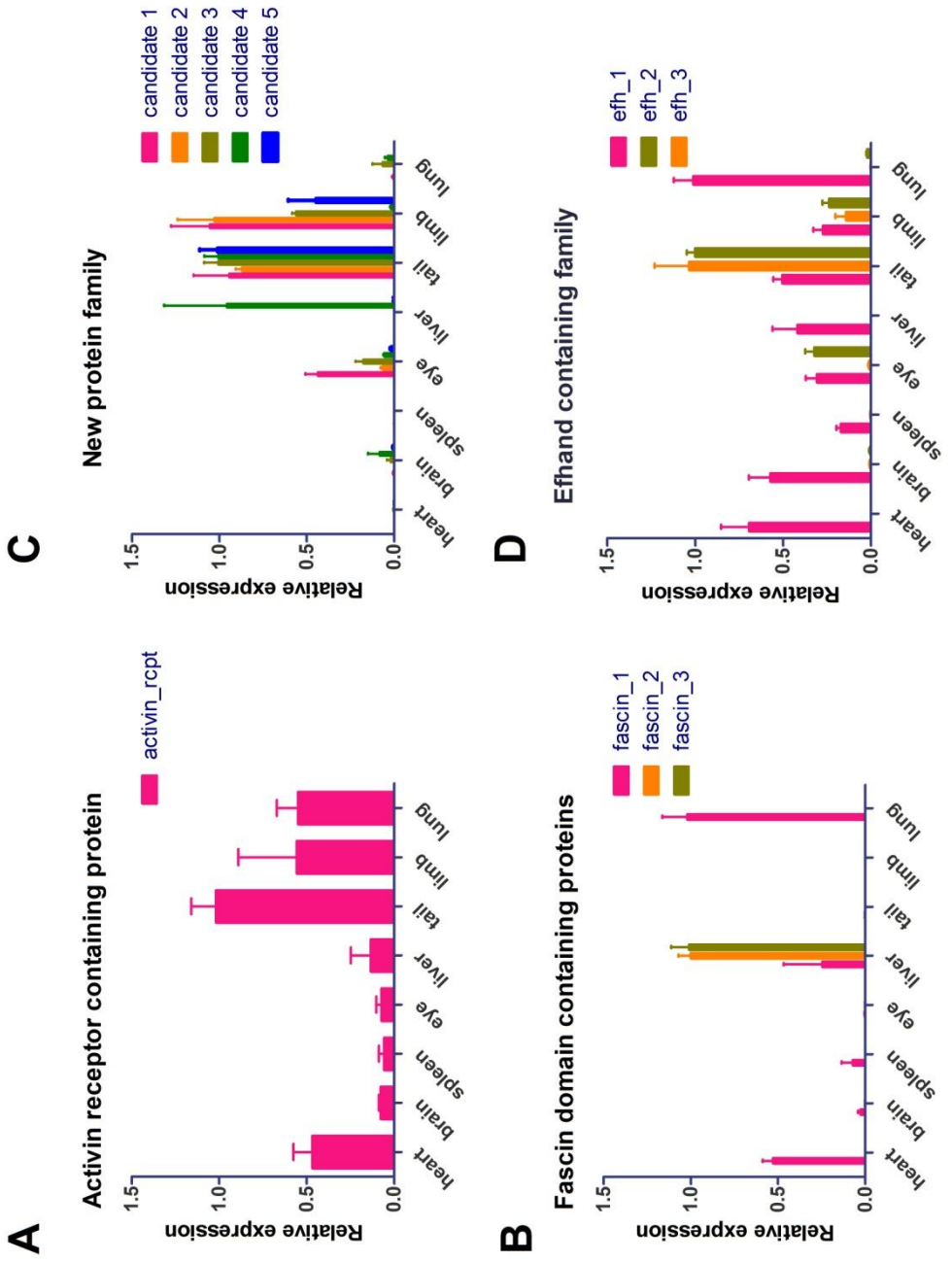


Figure 5: Expression of newly identified genes in adult tissues of the newt. Real time RT-PCR analysis ($n \geq 3$) of selected candidates and families in adult heart, brain, spleen, total eye, liver, tail, limb and lung. Tissues with highest expression level for each candidate were normalized to a relative value of 1.0.

3. Abschlussdiskussion

Am Molch stellt sich die Regenerationsforschung aufgrund des nicht sequenzierten Genoms und nur partiell vorhandenen Proteom und Transkriptom, verglichen mit anderen Modellorganismen wie dem Zebrafisch, schwieriger dar. Seine jedoch außergewöhnliche Fähigkeit zur Regeneration im adulten Zustand, sowie seine phylogenetische Nähe zum Menschen lassen es daher als hilfreich erscheinen, diese Regenerationsprozesse zu entschlüsseln und der Frage nachzugehen, ob das regenerative Potential des Molches durch eine lokale Evolution entstanden oder ein im Lauf der Evolution in höher entwickelten Organsimen verloren gegangene Fähigkeit ist.
Eine Möglichkeit, auch ohne sequenziertes Genom zu arbeiten und die dazu notwendigen Proteom und Transkriptom Daten zu erhalten, stellt der Einsatz moderner Hochdurchsatzanalysen, wie Massenspektrometrie, Microarrays und next generation sequencing (NGS) dar. In den vorliegenden drei Veröffentlichungen wurden neue Moleküle, die während der Herz-und Linsenregeneration eine potentielle Rolle spielen, sowohl auf Transkript- als auch auf Proteinebene gefunden.

3.1. Die Entdeckung von nsCCN und die frühe Wundantwort des Herzens

In der ersten Veröffentlichung liegt der Fokus auf den frühen Prozessen der Herzregeneration. Aus Forschungsergebnissen der Schwanz- und Beinregeneration ist bekannt, dass frühe inflamatorische und immunologische Prozesse entscheidend für den Regenerationsprozess sind[77-79]; ein Hauptgrund für die Wahl des Zeitpunktes der Protein Quantifizierung bereits sechs Stunden nach erfolgter Schädigung. Das Immunsystem und die an der Wundantwort beteiligten Moleküle in der frühen Herzregeneration im Molch sind bislang weitestgehend unbekannt gewesen. Bei nicht sequenziertem Genom ist die Entdeckung der dabei eventuell vorhandenen Spezies-spezifischen Moleküle auf reiner Nukleotidbasis aufgrund von nicht kodierender RNA, Klonierungsartefakten und Sequenzierungsfehlern eine komplizierte Aufgabe. Die Kombination von massenspektrometrischer Protein Identifizierung und einer bereits vorhandenen EST-Datenbank[76] ermöglichten dieses Vorhaben.
Durch die Methode der gewählten Protein Quantifizierung mit nicht vollständig gelabeltem Gewebe konnten 1353 Proteine quantifiziert werden. Eine verbesserte Quantifizierung ist prinzipiell durch eine vollständige Labelung möglich. Allerdings

dauert dieser Prozess aufgrund der langsamen Fortpflanzungszeit mehrere Jahre und verursacht darüber hinaus hohe Kosten in der Ernährung mit Aminosäuren, die eine Markierung mit stabilen schweren Isotopen besitzen. Die bei Beginn nur geringe Labelungsrate mit schweren Aminosäuren in ungeschädigtem Gewebe konnte durch Induktion einer Proliferation nach erfolgter Schädigung deutlich gesteigert werden, und dadurch auch eine bessere Quantifizierung ermöglichen. Diese könnte allerdings noch weiter verbessert werden. Anstatt dem durchgeführten Einsatz von gemischten Geweben, wären Messungen mit separierten Geweben von spezifischen Zeitpunkten eine Möglichkeit. Diese Form der Messungen benötigen lange Messzeiten und produzieren hohe Kosten.

Das nach Auswertung der Messergebnisse eingehender betrachtete Molekül, genannt „nsCCN“ (newt-specific CCN), zeigte nach vollständiger Entschlüsselung seiner Nukleotidsequenz alle 4 typischen Domänen (DAN, TSP-1, IGF, VWC), die Mitglieder der CCN-Familie aufweisen[63], was letztendlich auch die Zuordnung zu dieser Proteinfamilie ermöglichte. Die Frage, um welches Mitglied der CCN-Familie es sich bei diesem Kandidaten handelt, konnte nicht direkt beantwortet werden, da trotz aller vorhandener Domänen nur eine schwache Homologie zu anderen Familienmitgliedern bestand, die deutlich unterhalb der zu erwartenden Ähnlichkeit von CCN-Familienmitgliedern anderer Spezies untereinander lag. Bis auf CTGF (CCN2) waren keine weiteren Mitglieder der CCN Familie im Molch bekannt[84]. Ziel war es daher gewesen sämtliche Mitglieder der CCN-Familie im Molch zu identifizieren. Für CYR61, NOV und WISP1 konnten Sequenzen mit einer erwartet hohen Homologie, verglichen mit Säugetier-Sequenzen, gefunden werden. Versuche, Sequenzen mit einer Homologie zu WISP2 und WISP3 zu entdecken, schlugen fehl[85]. Auch in der de novo Assemblierung des Molch Transkriptoms in Veröffentlichung Nr. 3 konnten keine homologen Sequenzen gefunden werden. Das Nichtvorhandensein eines CCN Mitgliedes in der einen oder anderen Spezies ist bekannt, so wie z.B. im Zebrafisch NOV nicht vorhanden ist[86]. Eine abschließende Aussage über das Vorhandensein weiterer CCN-Familienmitglieder im Molch ist ohne ein sequenziertes Genom nicht möglich.

Im Anschluss erfolgte die Erstellung eines Expressionsmusters der 5 gefundenen CCN-Mitglieder in ungeschädigtem und regenerierendem Herzen. Hierbei zeigte sich in ungeschädigtem Herzgewebe ein kontinuierlich hohes Expressionsmuster von CYR61 (CCN1) und CTGF (CCN2), sowohl in qRT-PCRs, als auch in in-situ Hybridisierungen.

Signifikant geringer war die Expression des neuen Kandidaten nsCCN, sowie eine fast vollständige Abwesenheit von NOV (CCN3) und WISP1 (CCN4). Studien zum Expressionsmuster während der frühen Herzregeneration zeigten ein erhöhtes Expressionslevel von CYR61 und ein deutlich erhöhtes Expressionslevel von nsCCN. CTGF zeigte erst nach einigen Tagen ein ebenfalls signifikant erhöhtes Expressionslevel. Die Expression von NOV und WISP1 änderte sich verglichen mit den anderen CCN- Mitgliedern nur auf geringem Level. In situ Hybridisierungen von nsCCN in der frühen Herzregeneration zeigten dabei Anreicherungen der Expression in endokardialen Zellen. Aus Herzregenerationsstudien am Zebrafisch ist bekannt, dass endokardiale und epikardiale Zellen und deren exprimierte Proteine eine essentielle Rolle in der Kardiogenese besitzen[87]. Über die Rolle und mögliche Interaktionspartner von nsCCN in der frühen Herzregeneration kann aufgrund nicht weiter vorliegender Erkenntnisse nur spekuliert werden. Die Gewinnung weiterer Erkenntnisse verlangt die Möglichkeit einer genetischen Manipulation, die sich zu Zeit im Molch noch schwierig darstellen lässt. Ein erster gelungener Ansatz wurde Mitte des Jahres 2011 für die Linsenregeneration veröffentlicht[88]. Mit Hilfe von sogenannten Morpholinos, die zumeist aus antisense Oligonukleotiden verknüpft mit einem Octaguanadium Ring bestehen, konnte ein erfolgreicher Gen Knockdown in vivo erzielt werden. Die dabei verwendeten Injektionstechniken bestanden aus einer intraokularen für kurzfristige oder einer intraperitonealen für längerfristige Knockdowns. Intraperitoneale Injektionen mussten dabei täglich durchgeführt werden. Eine zwischen Linsen-und Herzschädigungsmodell auftretende Schwierigkeit könnte der noch teilweise intakte Gewebeverband im Molchherzen sein, während dieser bei der Augenlinsen Regeneration nach Entfernung entfällt. Nichtsdestotrotz könnte in dieser Methode eine Möglichkeit für eine genetische Manipulation liegen.

Ein weiterer Aspekt, der mit der Entdeckung des neu gefundenen Kandidaten einhergeht, ist die Hypothese einer lokalen Evolution der Regeneration im Molch. Erste Ansätze hierfür wurden bereits im Jahr 2002 [89] und 2009 für PROD1 [24] in der Bein Regeneration veröffentlicht. PROD1 gilt als ein wesentliches Gen für die proximo-distale Identität während der Beinregeneration. Gazia-Garcia et al. konnten zeigen, dass die Expression von PROD1 spezifisch für Molche und Salamander ist[24]. Dies könnte ein Beleg dafür sein, warum die Beinregeneration auf diese beiden Spezies beschränkt ist und durch eine lokale Evolution entstanden sein könnte. Durch die Entdeckung von nsCCN erfährt diese Hypothese neue Unterstützung. Es bleibt jedoch die Frage nach der

Rolle bzw. Bedeutung von nsCCN in der Herzregeneration offen. Mechanische Studien, die die Bedeutung ergründen, könnten Antworten darauf liefern.

Neben der Untersuchung des neu gefundenen Kandidaten stand, wie bereits erwähnt, die Betrachtung der Wundantwort und der ablaufenden immunologischen Prozesse im Blickpunkt. Ausgedehnte myokardiale Verletzungen innerhalb der Klasse der Säugetiere enden sehr häufig mit der Entstehung eines kollagenreichen Bindegewebes, das umgangssprachlich auch als Narbe bezeichnet wird[90]. Die Herzregeneration im Molch endet im Gegensatz zum Menschen nicht mit einer Narbenbildung, sondern mit neu gebildetem funktionell voll funktionsfähigem Herzgewebe[71]. Die Frage stellte sich dabei, wo Unterschiede bzw. Gemeinsamkeiten in der Wundantwort zwischen Molch und Mensch liegen könnten. Anfängliches Ziel war es eine allgemeine Übersicht über die Prozesse innerhalb der frühen Wundantwort zu erlangen. Basierend auf einer GO- Analyse konnte dabei gezeigt werden, dass signifikante Unterschiede und Gemeinsamkeiten in der Wundantwort bestehen. So zeigten insbesondere einige bekannte pro- und antiapoptotische Proteine im Molch eine inverse Regulation im Gegensatz zu geschädigten menschlichen Herzen. Antiapoptotische Proteine wie PRDX2, ANXA1 und proapoptotische Proteine wie BCL2L13 seien stellvertretend hierfür genannt. Aber auch inflammatorische Prozesse scheinen wesentlich milder in der frühen Wundantwort abzulaufen. FABP3, ein Fettsäure bindendes Protein, welches als serologischer Biomarker bei Myokardinfarkten im Menschen eingesetzt werden kann[91], zeigte eine runterregulierte Expression im Molch. Prozesse, wie die Aktivierung des Komplementsystems, finden sich jedoch in beiden Organismen wieder. Hierin könnte eventuell ein Ansatzpunkt für die Entwicklung von Therapien bei Herzinfarktpatienten liegen. Breit angelegte antiinflammatorische Therapien, wie Glukokortikoide, die das Immunsystem auf einer sehr breiten Basis hemmen, konnten in diversen Studien keinen Benefit für Patienten mit Myokardinfarkt erbringen[92, 93]. Ein besseres Verständnis der inflammatorischen Prozesse in einem Organismus, der in der Lage ist, sein geschädigtes Herz zu regenerieren, könnte dabei einen gezielten Ansatzpunkt liefern, welche inflammatorischen Prozesse unterdrückt werden müssten, um eine Narbenbildung zu verhindern. Kombinationen neuer antiinflammatorischer Therapien, lokal oder systemisch, mit einer perkutanen Koronarintervention (PCI) als Reperfusionstherapie wären eventuell eine neue Behandlungsmöglichkeit des akuten Myokardinfarktes[94].

3.2. Dorsales vs. ventrales Iris-Pigmentepithel in der frühen Linsenregeneration

Aufgrund der bereits bekannten und neu erworbenen Erkenntnisse über die Relevanz der frühen Wundantwort in der Bein-und Herzregeneration[79, 85] wurde in Zusammenarbeit mit dem Center for Tissue Regeneration and Engineeering der Universität Dayton (USA) unter Leitung von Prof. Tsonis ein Projekt zur Untersuchung der frühen Prozesse der Linsenregeneration durchgeführt. Im Rahmen einer Hochdurchsatzanalyse unter Verwendung von cDNA-Microarrays wurden dabei dorsale und ventrale Prozesse während der frühen Linsenregeneration miteinander verglichen. Die dabei verwendeten cDNA Microarrays und deren gespottete cDNA entstammen ausschließlich ungeschädigtem und regenerierendem Herzen von unterschiedlichen Zeitpunkten. Eventuell vorhandene Augen- bzw. Linsen spezifische Gene können demnach in der Expressionsanalyse durch das Design der Microarrays nicht berücksichtigt worden sein. Trotz der beschriebenen Einschränkungen konnten für die drei frühen Zeitpunkte (1. Tag, 3. Tag und 5. Tag nach Schädigung) insgesamt 467 verschiedene Sequenzen (Contigs) mit einer Deregulation identifiziert werden. Diese 467 Contigs stammen ausschließlich von hochqualitativen Spots, so dass bei geringerer stringenter Herangehensweise, noch deutlich höhere Zahlen an regulierten Contigs entstanden wären, bei damit einhergehender Zunahme an falsch positiven Kandidaten. Die 467 deregulierten Contigs wurden geclustert und mittels Heatmap visualisiert. Durch das Clustering konnte beobachtet werden, dass teilweise mehr als die Hälfte der Contigs auf dorsaler wie auf ventraler Irisseite in nahezu gleicher Weise reguliert waren. Dies bedeutet, dass eine Vielzahl von Prozessen auf dorsaler und ventraler Seite während der frühen Linsenregeneration in ähnlicher Weise ablaufen. Beispielhaft hierfür seien Gene aus dem Bereich der Redox Homöostase wie die Glutathione peroxidase 1, Peroxiredoxin-1, Thioredoxin und Peroxiredoxin 6 genannt. Ziel dürfte der Schutz vor oxidativen Schäden sein, die durch die Lentektomie entstanden sind[95].
Vor dem Hintergrund, dass in vivo die Linsenregeneration von der dorsalen Seite ausgeht, sind insbesondere Kandidaten, die auf dorsaler Seite eine differente Regulation erfahren von Bedeutung[96]. So wurden durch das Clustering 94 unterschiedliche regulierte Kandidaten zwischen dorsaler und ventraler Linse gefunden (fünf Tage post Lentektomie). Aus bereits veröffentlichten Untersuchungen ist bekannt, dass es im Bereich um Tag vier nach Lentektomie zu ersten Zellproliferationen und ab Tag acht zur sichtbar beginnenden Dedifferenzierung und Verlust der Pigmentierung der Iris-

Pigmentepithelzellen kommt[96]. Leider konnte nur für 14 der 94 Kandidaten eine Ähnlichkeit mit bekannten Genen aus anderen Organismen gefunden werden, trotz einer Sequenzlänge von im Durchschnitt mehr als 400 Basenpaaren. Die Frage stellt sich daher, um was für eine Art von Sequenzen es sich handelt. Dies könnten z.B. 5‘ bzw. 3‘ UTRs oder auch Molch spezifische long ncRNAs sein. Eventuell könnte man in manchen Fällen mit 5‘ bzw. 3‘ RACE-PCRs eine Antwort darauf erhalten. In Anbetracht der teilweise im Molch sehr langen 3`UTRs stößt eine RACE-PCR hierbei an ihre Grenzen. Eine Möglichkeit diese Frage jedoch zu beantworten, könnte der Einsatz einer weiteren tiefergehenden Transkriptom Sequenzierung oder eine Genom Sequenzierung sein.

Innerhalb der 14 Kandidaten mit einer Ähnlickkeit zu bekannten Genen fanden sich zwei äußerst Interessante. Zum einen wäre es ein Kandidat mit einer hohen Ähnlichkeit zum Tumorsuppressorgen suppression of tumorigenicity 7 (ST7) und zum anderen das Enzym GTPase HRas. Für das Tumorsuppressorgen ST7 liegen geringe Kenntnisse über dessen Einflüsse auf den Zellzyklus, die sich primär auf die Rolle in Tumoren beziehen, vor. Eine wesentliche Entdeckung in der Tumorforschung, die auch von Relevanz für die Linsenregeneration sein könnte, ist eine von ST7 gesteuerte Induktion von Genen die an Veränderungen der extrazellulären Matrix beteiligt sind wie SPARC, IGFBP und Matrix Metalloproteinasen[97]. Der zweite Kandidat, die GTPase HRas, auch bekannt unter dem Namen transforming protein p21, gilt als ein wichtiger Bestandteil bei der Regulation der Zellteilung[98], die durch Wachstumsfaktoren stimuliert werden[99]. Insbesondere die durch die Wundantwort entstehenden Fibrin Pröpfe bilden eine Vielzahl an Wachstumsfaktoren, die umgebende Zellen stimulieren[43]. Die differente Expression der GTPase HRas in dorsalem und ventralem Iris–Pigmentepithel könnte ein weiterer Baustein sein für die Enstehung der neuen Linse auf dorsaler Seite. Die Frage allerdings, welche Wachstumsfaktoren speziell eine erhöhte Expression der GTPase HRAs einleiten, kann nicht beantwortet werden und bedarf weiterer Untersuchungen.

In einem weiteren Schritt wurden den 467 gefundenen Kandidaten GO-Terme zugeteilt und eine funktionelle Annotation erstellt. Insbesondere Kandidaten aus den Bereichen Zellzyklus, Proliferation, Extrazellulär-Region, Redox-Homöostase, Mitochondrium und DNA-Reparatur zeigten eine Regulation in den ersten Tagen der Linsenregeneration. Da diese Veränderungen schon am ersten Tag nach Schädigung auftreten, liegt die Vermutung nahe, dass bereits in diesem frühen Zeitraum einige

Zellen wieder in den Zellzyklus eintreten und erste Proliferationen beginnen, die von Reparaturmechanismen intra-und extrazellulär begleitet werden. Auffällig war dabei eine signifikant höhere Expression von einzelnen Kandidaten aus dem Reparaturmechanismus Bereich in der dorsalen Iris. Im Kontext der Expression von ST7 und dessen beschriebenen Einfluss auf das extrazelluläre Milieu, ist die hohe Expression von Matrix Metalloproteinasen wie Mmp18 und Stromelysin ½ alpha in der dorsalen Iris ebenfalls als ein weiteres Element für die Regeneration von dorsaler Iris-Seite aus zu werten.

Die erstmalige Erkenntnis, dass bereits am ersten Tag nach Schädigung eine Vielzahl von Prozessen ablaufen, die die dorsalen Iris-Pigmentepithelzellen in die Lage versetzen zu proliferieren und die Basis zur späteren Transdifferenzierung legen, könnte dazu beitragen, den Fokus der Linsenregeneration vermehrt auf frühe Zeitpunkte zu richten.

Knockdown Experimente, z.B. mit Morpholinos, könnten eine Möglichkeit sein, die Rolle einzelner Moleküle besser zu verstehen, sowie Signalwege und Prozesse zu beschreiben[88]. Durch die die Verwendung von cDNA-Microarrays, einer Hochdurchsatzanalyse, konnte eine Vielzahl potentiell hochinteressanter Kandidaten der frühen Linsenregeneration entdeckt werden, deren Rolle, wie bereits erwähnt, zumeist im Unklaren verbleibt. Weitere Erkenntnisse, die die Entstehung der Linsenregeneration aus dem dorsalen Iris-Pigmentepithel unterstreichen sowie Moleküle, die die Relevanz der frühen Wundprozesse betonen, wurden in einer Vielzahl gefunden.

3.3. Neue Proteinfamilien in der Regeneration

Eine genauere Analyse des Transkriptoms der frühen Linsenregeneration dürfte mit einer Hochdurchsatzanalyse, wie einem next generation sequencing (NGS) Projekt möglich sein. Als Matrize für ein solches Sequenzierungsprojekt könnte das in Veröffentlichung Nr. 3 vorgestellte Transkriptom dienen. Im Gegensatz zum verwendeten Microarray bildet dieses Transkriptom eine Vielzahl unterschiedlicher Gewebe (Gehirn, Milz, Leber, Herz, Lunge, Auge, Ovar, Hoden, Bein, Schwanz), Entwicklungs-und Regenerationszeitpunkte ab.

Die zur Assemblierung verwendeten Sequenzierungsdatensätze entstammen drei unterschiedlichen Plattformen (EST-Sequenzen, Illumina-Sequenzen und 454-Sequenzen). In unterschiedlichen Assemblierungsschritten wurde aus zu Beginn vorhandenen 679.816.626 Sequenzen ein Transkriptom mit 120.922 nicht redundanten

Contigs erstellt. Die Methode, unterschiedliche Sequenzierungs Methoden miteinander zu kombinieren, ermöglicht es, die Vorteile die die einzelnen Technologien bieten zu nutzen und mögliche Nachteile der jeweils einzelnen auszugleichen. Die Methode der Sanger Sequenzierung bietet lange Leselängen, die mit einer hohen Fehlerfreiheit behaftet sind; Sequenzen, die auf Roches 454 Plattform sequenziert werden, mittlere Leselängen mit Schwierigkeiten in der Auflösung von Homopolymeren, sowie kurze Sequenzen mit einer sehr guten Sequenzierungstiefe die Illumina-Technologie bei Abstrichen in der Lesegenauigkeit[100, 101].

Von 120.922 nicht redundanten Contigs konnten durch die Verwendung unterschiedlicher Blast-Algorithmen 56% annotiert werden mit einem e-Value kleiner als e-15. Diese Annotationsrate ist vergleichbar mit anderen veröffentlichten Transkriptom Daten von Organismen, wie z.B. von Planarien[102].

Die Identifizierung neuer bzw. Molch spezifischer Protein kodierender Sequenzen stellt sich aufgrund des nicht sequenzierten Genoms, verglichen beispielsweise mit dem Zebrafisch, als Herausforderung dar. Eine Möglichkeit diese Hürde zumindest in Teilen zu umgehen, stellt die massenspektrometrische Peptid-Messung und eine anschließende Identifizierung selbiger durch das neue Transkriptom dar. Die Kombination dieser beiden Hochdurchsatzanalysen ermöglicht es vor allem für Sequenzen mit nicht vorhandener Annotation zu entscheiden, ob es sich um codierende oder um nicht codierende Sequenzen handelt. Im Rahmen dieser Analyse konnten 826 Urodel-spezifische Proteine identifiziert werden. Es bleibt zu vermuten, dass noch weitaus höhere Zahlen an Urodel spezifischen Proteinen vorhanden sind, die allerdings vor dem Hintergrund einer nicht sehr tiefgehenden massenspektrometrischen Analyse, die überwiegend hoch abundante Peptide identifiziert, durchgeführt wurde. Tiefergehende Analysen benötigen wiederum sehr lange Messzeiten und verursachen hohe Kosten. Aufbauend auf diesen Ergebnissen konnten innerhalb der neuen Kandidaten neue Proteinfamilien entdeckt werden, die anschließend mit qRT-PCRs auf ihre Expression hin in unterschiedlichen Geweben untersucht wurden. Neben gewebsspezifischen Expressionsmustern fanden sich auch deregulierte Kanidaten zu verschiedenen Zeitpunkten in der Linsen-, Herz- und Schwanzregeneration. Die Rolle, die diese Kandidaten innerhalb der Regeneration einnehmen, bleibt hingegen offen. Die nach aktuellem Kenntnisstand Urodel spezifischen Kandidaten sind am ehesten durch eine lokale Evolution entstanden. Funktionelle Studien werden dabei helfen zu erfahren, ob es sich bei diesen Kandidaten um Schlüsselakteure der Regeneration handelt oder sie

eine gänzlich andere Funktion besitzen. Grundsätzlich bleiben geringe Restzweifel, ob diese Kandidaten ausschließlich Urodel spezifisch sind, oder ob sie auch in anderen Organismen exprimiert werden, die bis dato unerforscht sind.

Regeneration gehört zu den interessantesten Prozessen im Tierreich und wird durch eine Reihe unterschiedlicher Faktoren gesteuert. Die Frage, warum einige Tiere in der Lage sind, Körperteile oder komplette Organe zu regenerieren und andere nicht, kann auch weiterhin nur in einem gewissen Rahmen beantwortet werden. Viele Hypothesen wurden bereits dazu entwickelt[28, 103-105]. Die Schwierigkeit der Beantwortung liegt vor allem in der hohen Varietät der Regeneration in den unterschiedlichen Spezies begründet. Als Beispiel dafür wäre allein die unterschiedliche Weise der Linsenregeneration in Frosch und Molch anzuführen[29].
Die rasch fortschreitende Entwicklung in der modernen Sequenzierungs-und Proteomics Technologie und in den Möglichkeiten einer genetischen Manipulation dürften jedoch in Zukunft eine weitaus genauere Beobachtung und Rückschlüsse über Gemeinsamkeiten und Unterschiede in Regenerationsprozessen zulassen. Eine wichtige Erkenntnis, die die Linsen-und Herzregenerationsstudien gezeigt haben, ist die hohe Relevanz der Wundantwort sowie der Zellzyklus- und Proliferationsprozesse in den ersten Stunden und Tagen nach Schädigung. Diese Erkenntnis dürfte auch von nicht unerheblichem Interesse für die Entwicklung medizinischer Therapien sein, insbesondere im Rahmen der frühen myokardialen Infarktbehandlung.
Mit der erstmaligen Erstellung eines *de novo* assemblierten Molch Transkriptoms wird der molekularbiologische Zugang zum Molch deutlich erleichtert und kann dabei als wichtige Grundlage bzw. Matrize für nachfolgende Sequenzierungsprojekte dienen. Inwieweit die Entdeckung Molch spezifischer Moleküle in der Regeneration eine Rolle spielt, werden erst zukünftige Analysen zeigen.
Die Untersuchung der Regenerationsprozesse im Molch bleibt nach wie vor ein wissenschaftliches Forschungsfeld in dem noch viele Fragen unbeantwortet bleiben. Eine Reihe hochinteressanter Kandidaten wurden in allen drei Veröffentlichungen gefunden, deren Charakterisierung und funktionelle Analyse die Möglichkeit bietet, dem Phänomen Regeneration näher zu kommen.

4. Zusammenfassung

Der rot gepunktete grüne Molch *Notophthalmus viridescens* besitzt die außerordentliche Fähigkeit verschiedene anatomische Strukturen wie Linse, Schwanz und Bein als auch kardiale Schädigungen ohne eine funktionelle Beeinträchtigung zu regenerieren. Diese Fähigkeit zu Regeneration im adulten Stadium verschafft dem Molch eine herausgehobene Position innerhalb der Vertebraten Familie. Trotz seiner außerordentlichen Fähigkeit ist nur wenig Wissen über frühe regenerative Prozesse vorhanden. Bei primärem Fokus auf der Herz-und Linsenregeneration konnten wir eine Reihe von Molekülen identifizieren, die eine potentiell wichtige Rolle in der frühen Regeneration spielen könnten.

Die Identifizierung und die erhöhte Expression eines Molch spezifischen Moleküls, genannt nsCCN und zur Protein Familie der CCNs gehörend, zeigte eine Wund-spezifische Aktivierung und eine auf endokardiale Zellen beschränkte Expression während der frühen Herzregeneration. Verschiedene vermutlich Molch spezifische Proteinfamilien wurden in einer Hochdurchsatzanalyse (NGS) gefunden und auf ihr Expressionsmuster hin in unterschiedlichen Geweben und regenerierenden Organen wie Herz, Linse und Bein untersucht. Die potenzielle Molch Spezifität dieser Moleküle lässt eine mögliche Rolle in der Regeneration vermuten und unterstützt die These einer lokalen Evolution der Regeneration im Molch. Zusätzlich untersuchten wir die frühe Wundantwort und Regeneration der Augenlinse. Basierend auf dem Wissen, dass die Regeneration der Linse von der dorsalen Iris ausgeht, stellte sich in einer Hochdurchsatzanalyse (cDNA Microarrays) heraus, dass eine Vielzahl von Prozessen während der frühen Wundantwort auf ventraler und auf dorsaler Iris ähnlich sind, und nur wenige Unterschiede im Expressionsmuster bestehen. Diese quantitativen Unterschiede könnten jedoch entscheidend für die Regeneration der Augenlinse von dorsaler Iris Seite aus sein.

Die in dieser Arbeit gewonnenen Erkenntnisse verbessern signifikant das Verständnis der dynamischen Veränderungen der frühen Regenerationsprozesse im Molch und bilden eine Basis für zukünftige funktionelle Analysen. Die Hoffnung, die mit dem Wissen über Prozesse der frühen Regeneration einhergeht, ist die Entwicklung gezielter therapeutischer Behandlungen für verschiedene Erkrankungen des Menschen.

Summary

The red spotted green newt *Notophthalmus viridescens* possesses the extraordinary ability to regenerate several anatomical structures like lens, tail and limb as well as cardial injuries without functional impairment. This regenerative capacity as adult animal provides the newt with an exceptional position among all vertebrates. Despite the important role there is only minor knowledge about early regeneration processes. By primarily focusing on the newt heart and lens we could identify new molecules that might play an important role in early regeneration processes.

The identification and upregulation of a newt specific molecule, named nsCCN and belonging to the CCN family, showed a wound specific activation and restricted expression in endocardial cells in heart regeneration. Several other probably newt specific protein families have been found by a next generation sequencing (NGS) run and further evaluated by their expression pattern in different tissues and regenerating structures like heart, lens and limb. The potential newt specifity of these molecules suggest a distinctive role in regeneration and support the idea of a local evolution of regeneration in newts. Additionally we analyzed the early wound response and regeneration in the lens. Well known that the regeneration process of the lens is driven by the dorsal iris, high throughput analyses by cDNA Microarrays demonstrated a lot of common pathways and a few changes of the expression pattern in dorsal and ventral iris in the early wound response. Quantitative differences might be critical for the dorsal and ventral iris contribution in lens regeneration.

These results significantly improve the understanding of dynamic changes in early regeneration processes and provide a basis for further mechanistic studies. The knowledge about these processes in the early newt regeneration hopefully leads to the specific development of different therapies of human diseases.

5. Abkürzungsverzeichnis

BLAST	Basic Alignment Search Tool
CCN	Familie von extrazellulären Matrix Proteinen (**C**YR61, **C**TGF, **N**OV)
cDNA	Komplementäre Desoxyribonukleinsäure mittels einer reversen Transkriptase
DNA	Desoxyribonukleinsäure
ECM	Extrazellulär-Matrix
EST	Expressed sequence tag
et. al	Et altera
evt.	eventuell
GO	Gene Ontology
MPI	Max-Planck-Institut
mRNA	Messenger Ribonukleinsäure
ncRNA	Nicht-codierende Ribonukleinsäure
NGS	Next Generation Sequencing
PCR	Polymerase Kettenreaktion
qRT-PCR	Quantitative Polymerase Kettenreaktion
RNA	Ribonukleinsäure
UTR	Untranslated region
vs.	versus
WHO	World Health Organization
z.B.	zum Beispiel

6. Abbildungsverzeichnis

Seite

7. Literaturverzeichnis

1. Higgins, G. and R. Anderson, *Experimental pathology of the liver. I. Restoration of the liver of the white rat following partial surgical removal*. Vol. 12. 1931: Arch Pathol. 186-202.
2. Bergmann, A. and H. Steller, *Apoptosis, stem cells, and tissue regeneration*. Sci Signal, 2010. 3(145): p. re8.
3. Spallanzani, L., *Prodromo di un' opera da imprimersi sopra le riproduzioni animali*. 1768, In Modena: Nella Stamperia di Giovanni Montanari. 102 p.
4. Eguchi, G., et al., *Regenerative capacity in newts is not altered by repeated regeneration and ageing*. Nat Commun, 2011. 2: p. 384.
5. Laube, F., et al., *Re-programming of newt cardiomyocytes is induced by tissue regeneration*. J Cell Sci, 2006. 119(Pt 22): p. 4719-29.
6. Berg, D.A., et al., *Efficient regeneration by activation of neurogenesis in homeostatically quiescent regions of the adult vertebrate brain*. Development, 2010. 137(24): p. 4127-34.
7. Sanchez Alvarado, A., *Planarian regeneration: its end is its beginning*. Cell, 2006. 124(2): p. 241-5.
8. Tanaka, E.M. and P.W. Reddien, *The cellular basis for animal regeneration*. Dev Cell, 2011. 21(1): p. 172-85.
9. Duellman, W.E. and L. Trueb, *Biology of amphibians*. 1994, Baltimore: Johns Hopkins University Press. xxi, 670 p.
10. Lenhoff, S.G., H.M. Lenhoff, and A. Trembley, *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*. 1986, Pacific Grove, CA: Boxwood Press. Getr. Zählung.
11. Bode, H., *Axis formation in hydra*. Annu Rev Genet, 2011. 45: p. 105-17.
12. Bosch, T.C., et al., *The Hydra polyp: nothing but an active stem cell community*. Dev Growth Differ, 2010. 52(1): p. 15-25.
13. Galliot, B., et al., *Hydra, a niche for cell and developmental plasticity*. Semin Cell Dev Biol, 2006. 17(4): p. 492-502.
14. Kasbauer, T., et al., *The Notch signaling pathway in the cnidarian Hydra*. Dev Biol, 2007. 303(1): p. 376-90.
15. Chera, S., et al., *Apoptotic cells provide an unexpected source of Wnt3 signaling to drive hydra head regeneration*. Dev Cell, 2009. 17(2): p. 279-89.
16. Galliot, B. and S. Chera, *The Hydra model: disclosing an apoptosis-driven generator of Wnt-based regeneration*. Trends Cell Biol, 2010. 20(9): p. 514-23.
17. Sanchez Alvarado, A. and P.A. Tsonis, *Bridging the regeneration gap: genetic insights from diverse animal models*. Nat Rev Genet, 2006. 7(11): p. 873-84.
18. Morgan, T.H., *REGENERATION AND LIABILITY TO INJURY*. Science (New York, N Y), 1901. 14(346): p. 235-48.
19. Agata, K., et al., *Intercalary regeneration in planarians*. Developmental dynamics : an official publication of the American Association of Anatomists, 2003. 226(2): p. 308-16.
20. Wallace, B.M. and H. Wallace, *Participation of grafted nerves in amphibian limb regeneration*. J Embryol Exp Morphol, 1973. 29(3): p. 559-70.
21. Gargioli, C. and J.M. Slack, *Cell lineage tracing during Xenopus tail regeneration*. Development, 2004. 131(11): p. 2669-79.
22. Kragl, M., et al., *Cells keep a memory of their tissue origin during axolotl limb regeneration*. Nature, 2009. 460(7251): p. 60-5.

23. Morrison, J.I., P. Borg, and A. Simon, *Plasticity and recovery of skeletal muscle satellite cells during limb regeneration*. FASEB J, 2010. 24(3): p. 750-6.
24. Garza-Garcia, A., et al., *Solution structure and phylogenetics of Prod1, a member of the three-finger protein superfamily implicated in salamander limb regeneration*. PLoS One, 2009. 4(9): p. e7123.
25. Kumar, A., P.B. Gates, and J.P. Brockes, *Positional identity of adult stem cells in salamander limb regeneration*. C R Biol, 2007. 330(6-7): p. 485-90.
26. Anderson, J.S., et al., *A stem batrachian from the Early Permian of Texas and the origin of frogs and salamanders*. Nature, 2008. 453(7194): p. 515-8.
27. Garza-Garcia, A.A., P.C. Driscoll, and J.P. Brockes, *Evidence for the local evolution of mechanisms underlying limb regeneration in salamanders*. Integr Comp Biol, 2010. 50(4): p. 528-35.
28. Bely, A.E., *Evolutionary loss of animal regeneration: pattern and process*. Integr Comp Biol, 2010. 50(4): p. 515-27.
29. Henry, J.J. and P.A. Tsonis, *Molecular and cellular aspects of amphibian lens regeneration*. Prog Retin Eye Res, 2010. 29(6): p. 543-55.
30. Colucci, V.S., *Sulla rigenerazione parziale dell' occhio nei Tritoni. Istogenesi e siluppo*. Vol. Ser V 1. 1891, Mem. della Roy. Acad. della Science dell' Inst. di Bologna. p. 593-629.
31. Freeman, G., *Lens Regeneration from the Cornea in Xenopus Laevis*. J Exp Zool, 1963. 154: p. 39-65.
32. Barbosa-Sabanero, K., et al., *Lens and retina regeneration: new perspectives from model organisms*. Biochem J, 2012. 447(3): p. 321-34.
33. Eguchi, G. and R. Shingai, *Cellular analysis on localization of lens forming potency in the newt iris epithelium*. Dev Growth Differ, 1971. 13(4): p. 337-49.
34. Yamada, T. and M. Roesel, *Control of mitotic activity in Wolffian lens regenerration*. J Exp Zool, 1971. 177(1): p. 119-28.
35. Call, M.K., M.W. Grogg, and P.A. Tsonis, *Eye on regeneration*. Anat Rec B New Anat, 2005. 287(1): p. 42-8.
36. Madhavan, M., et al., *The role of Pax-6 in lens regeneration*. Proc Natl Acad Sci U S A, 2006. 103(40): p. 14848-53.
37. Markitantova, Y.V., et al., *Location of the Prox1 gene expression during newt lens and retina regeneration*. Dokl Biol Sci, 2003. 391: p. 361-4.
38. Hayashi, T., et al., *Determinative role of Wnt signals in dorsal iris-derived lens regeneration in newt eye*. Mech Dev, 2006. 123(11): p. 793-800.
39. Maki, N., et al., *Expression profiles during dedifferentiation in newt lens regeneration revealed by expressed sequence tags*. Mol Vis, 2010. 16: p. 72-8.
40. Kulyk, W.M., S.E. Zalik, and E. Dimitrov, *Hyaluronic acid production and hyaluronidase activity in the newt iris during lens regeneration*. Exp Cell Res, 1987. 172(1): p. 180-91.
41. Elgert, K.L. and S.E. Zalik, *Fibronectin distribution during cell type conversion in newt lens regeneration*. Anat Embryol (Berl), 1989. 180(2): p. 131-42.
42. Sahni, A., et al., *Interleukin-1beta but not IL-1alpha binds to fibrinogen and fibrin and has enhanced activity in the bound form*. Blood, 2004. 104(2): p. 409-14.
43. Godwin, J.W., K.F. Liem, Jr., and J.P. Brockes, *Tissue factor expression in newt iris coincides with thrombin activation and lens regeneration*. Mech Dev, 2010. 127(7-8): p. 321-8.

44. Imokawa, Y., A. Simon, and J.P. Brockes, *A critical role for thrombin in vertebrate lens regeneration*. Philos Trans R Soc Lond B Biol Sci, 2004. 359(1445): p. 765-76.
45. Lee, J.G. and E.P. Kay, *FGF-2-induced wound healing in corneal endothelial cells requires Cdc42 activation and Rho inactivation through the phosphatidylinositol 3-kinase pathway*. Invest Ophthalmol Vis Sci, 2006. 47(4): p. 1376-86.
46. Maki, N., et al., *Rapid accumulation of nucleostemin in nucleolus during newt regeneration*. Dev Dyn, 2007. 236(4): p. 941-50.
47. Maki, N., et al., *Expression of stem cell pluripotency factors during regeneration in newts*. Dev Dyn, 2009. 238(6): p. 1613-6.
48. World-Health-Organization, *World Health Statistics 2008*. Leading causes of death, 2004 and 2030 compared 2008. p. 30.
49. Laflamme, M.A. and C.E. Murry, *Heart regeneration*. Nature, 2011. 473(7347): p. 326-35.
50. Bergmann, O., et al., *Evidence for cardiomyocyte renewal in humans*. Science, 2009. 324(5923): p. 98-102.
51. Kajstura, J., et al., *Cardiomyogenesis in the adult human heart*. Circ Res, 2010. 107(2): p. 305-15.
52. Steinhauser, M.L. and R.T. Lee, *Regeneration of the heart*. EMBO Mol Med, 2011. 3(12): p. 701-12.
53. Kikuchi, K., et al., *Primary contribution to zebrafish heart regeneration by gata4(+) cardiomyocytes*. Nature, 2010. 464(7288): p. 601-5.
54. Hsieh, P.C., et al., *Evidence from a genetic fate-mapping study that stem cells refresh adult mammalian cardiomyocytes after injury*. Nat Med, 2007. 13(8): p. 970-4.
55. Kajstura, J., et al., *Myocyte proliferation in end-stage cardiac failure in humans*. Proc Natl Acad Sci U S A, 1998. 95(15): p. 8801-5.
56. Porrello, E.R., et al., *Transient regenerative potential of the neonatal mouse heart*. Science, 2011. 331(6020): p. 1078-80.
57. Jopling, C., et al., *Zebrafish heart regeneration occurs by cardiomyocyte dedifferentiation and proliferation*. Nature, 2010. 464(7288): p. 606-9.
58. Porrello, E.R., et al., *MiR-15 family regulates postnatal mitotic arrest of cardiomyocytes*. Circ Res, 2011. 109(6): p. 670-9.
59. Hassink, R.J., et al., *Cardiomyocyte cell cycle activation improves cardiac function after myocardial infarction*. Cardiovasc Res, 2008. 78(1): p. 18-25.
60. Bersell, K., et al., *Neuregulin1/ErbB4 signaling induces cardiomyocyte proliferation and repair of heart injury*. Cell, 2009. 138(2): p. 257-70.
61. Braun, T. and S. Dimmeler, *Breaking the silence: stimulating proliferation of adult cardiomyocytes*. Dev Cell, 2009. 17(2): p. 151-3.
62. Loffredo, F.S., et al., *Bone marrow-derived cell therapy stimulates endogenous cardiomyocyte progenitors and promotes cardiac repair*. Cell Stem Cell, 2011. 8(4): p. 389-98.
63. Frangogiannis, N.G., *Matricellular proteins in cardiac adaptation and disease*. Physiol Rev, 2012. 92(2): p. 635-88.
64. Leask, A. and D.J. Abraham, *All in the CCN family: essential matricellular signaling modulators emerge from the bunker*. J Cell Sci, 2006. 119(Pt 23): p. 4803-10.

65. Yoshida, Y., et al., *CCN1 protects cardiac myocytes from oxidative stress via beta1 integrin-Akt pathway*. Biochem Biophys Res Commun, 2007. 355(3): p. 611-8.
66. Hilfiker-Kleiner, D., et al., *Regulation of proangiogenic factor CCN1 in cardiac muscle: impact of ischemia, pressure overload, and neurohumoral activation*. Circulation, 2004. 109(18): p. 2227-33.
67. Ahmed, M.S., et al., *Mechanisms of novel cardioprotective functions of CCN2/CTGF in myocardial ischemia-reperfusion injury*. Am J Physiol Heart Circ Physiol, 2011. 300(4): p. H1291-302.
68. Ohnishi, H., et al., *Increased expression of connective tissue growth factor in the infarct zone of experimentally induced myocardial infarction in rats*. J Mol Cell Cardiol, 1998. 30(11): p. 2411-22.
69. Alfaro, M.P., et al., *A physiological role for connective tissue growth factor in early wound healing*. Lab Invest, 2013. 93(1): p. 81-95.
70. Venkatachalam, K., et al., *WISP1, a pro-mitogenic, pro-survival factor, mediates tumor necrosis factor-alpha (TNF-alpha)-stimulated cardiac fibroblast proliferation but inhibits TNF-alpha-induced cardiomyocyte death*. J Biol Chem, 2009. 284(21): p. 14414-27.
71. Witman, N., et al., *Recapitulation of developmental cardiogenesis governs the morphological and functional regeneration of adult newt hearts following injury*. Dev Biol, 2011. 354(1): p. 67-76.
72. Borchardt, T. and T. Braun, *Cardiovascular regeneration in non-mammalian model systems: what are the differences between newts and man?* Thromb Haemost, 2007. 98(2): p. 311-8.
73. Oberpriller, J.O. and J.C. Oberpriller, *Response of the adult newt ventricle to injury*. J Exp Zool, 1974. 187(2): p. 249-53.
74. Matz, D.G., J.O. Oberpriller, and J.C. Oberpriller, *Comparison of mitosis in binucleated and mononucleated newt cardiac myocytes*. Anat Rec, 1998. 251(2): p. 245-55.
75. Piatkowski, T., et al., *Reconstitution of the myocardium in regenerating newt hearts is preceded by transient deposition of extracellular matrix components*. Stem Cells Dev, 2013.
76. Borchardt, T., et al., *Analysis of newly established EST databases reveals similarities between heart regeneration in newt and fish*. BMC Genomics, 2010. 11: p. 4.
77. Tassava, R.A. and R.M. Loyd, *Injury requirement for initiation of regeneration of newt limbs which have whole skin grafts*. Nature, 1977. 268(5615): p. 49-50.
78. Kato, T., et al., *Unique expression patterns of matrix metalloproteinases in regenerating newt limbs*. Dev Dyn, 2003. 226(2): p. 366-76.
79. Godwin, J.W. and J.P. Brockes, *Regeneration, tissue injury and the immune response*. J Anat, 2006. 209(4): p. 423-32.
80. Looso, M., et al., *Advanced identification of proteins in uncharacterized proteomes by pulsed in vivo stable isotope labeling-based mass spectrometry*. Mol Cell Proteomics, 2010. 9(6): p. 1157-66.
81. Bruckskotten, M., et al., *Newt-omics: a comprehensive repository for omics data from the newt Notophthalmus viridescens*. Nucleic Acids Res, 2012. 40(Database issue): p. D895-900.
82. Dudoit, S. and J. Fridlyand, *A prediction-based resampling method for estimating the number of clusters in a dataset*. Genome Biol, 2002. 3(7): p. RESEARCH0036.

83. Altschul, S.F., et al., *Basic local alignment search tool*. J Mol Biol, 1990. 215(3): p. 403-10.
84. Cash, D.E., et al., *Identification of newt connective tissue growth factor as a target of retinoid regulation in limb blastemal cells*. Gene, 1998. 222(1): p. 119-24.
85. Looso, M., et al., *Spiked-in pulsed in vivo labeling identifies a new member of the CCN family in regenerating newt hearts*. J Proteome Res, 2012. 11(9): p. 4693-704.
86. Fernando, C.A., et al., *Temporal and spatial expression of CCN genes in zebrafish*. Dev Dyn, 2010. 239(6): p. 1755-67.
87. Kikuchi, K., et al., *Retinoic acid production by endocardium and epicardium is an injury response essential for zebrafish heart regeneration*. Dev Cell, 2011. 20(3): p. 397-404.
88. Tsonis, P.A., et al., *Controlling gene loss of function in newts with emphasis on lens regeneration*. Nat Protoc, 2011. 6(5): p. 593-9.
89. da Silva, S.M., P.B. Gates, and J.P. Brockes, *The newt ortholog of CD59 is implicated in proximodistal identity during amphibian limb regeneration*. Dev Cell, 2002. 3(4): p. 547-55.
90. van den Borne, S.W., et al., *Myocardial remodeling after infarction: the role of myofibroblasts*. Nat Rev Cardiol, 2010. 7(1): p. 30-7.
91. Shand, J.A., D.J. McEneaney, and I.B. Menown, *Heart-type fatty acid-binding protein in the early diagnosis of acute myocardial infarction: a systematic review and meta-analysis*. Heart, 2011. 97(7): p. 605; author reply 605.
92. Madias, J.E. and W.B. Hood, Jr., *Effects of methylprednisolone on the ischemic damage in patients with acute myocardial infarction*. Circulation, 1982. 65(6): p. 1106-13.
93. Bush, C.A., W. Renner, and H. Boudoulas, *Corticosteroids in acute myocardial infarction*. Angiology, 1980. 31(10): p. 710-4.
94. Steffens, S., F. Montecucco, and F. Mach, *The inflammatory response as a target to reduce myocardial ischaemia and reperfusion injury*. Thromb Haemost, 2009. 102(2): p. 240-7.
95. Sarsour, E.H., et al., *Redox control of the cell cycle in health and disease*. Antioxid Redox Signal, 2009. 11(12): p. 2985-3011.
96. Tsonis, P.A., et al., *A newt's eye view of lens regeneration*. Int J Dev Biol, 2004. 48(8-9): p. 975-80.
97. Hooi, C.F., et al., *ST7-mediated suppression of tumorigenicity of prostate cancer cells is characterized by remodeling of the extracellular matrix*. Oncogene, 2006. 25(28): p. 3924-33.
98. Delprato, A., *Topological and functional properties of the small GTPases protein interaction network*. PLoS One, 2012. 7(9): p. e44882.
99. Ehrhardt, A., et al., *Distinct mechanisms determine the patterns of differential activation of H-Ras, N-Ras, K-Ras 4B, and M-Ras by receptors for growth factors or antigen*. Mol Cell Biol, 2004. 24(14): p. 6311-23.
100. Quail, M.A., et al., *A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers*. BMC Genomics, 2012. 13: p. 341.
101. Liu, L., et al., *Comparison of next-generation sequencing systems*. J Biomed Biotechnol, 2012. 2012: p. 251364.

102. Adamidi, C., et al., *De novo assembly and validation of planaria transcriptome by massive parallel sequencing and shotgun proteomics*. Genome research, 2011. 21(7): p. 1193-200.
103. Goss, R.J., *The evolution of regeneration: adaptive or inherent?* J Theor Biol, 1992. 159(2): p. 241-60.
104. Bely, A.E. and K.G. Nyberg, *Evolution of animal regeneration: re-emergence of a field*. Trends in ecology & evolution, 2010. 25(3): p. 161-70.
105. Reichman, O.J., *Evolution of Regeneration Capabilities*. American Naturalist, 1984. 123(6): p. 752-763.

8. Anhang

8.1. Zusatzmaterial Veröffentlichung Nr. 1

Supplementary figure 1. MA Plot of Protein Expression
Plot shows calculated ratios of all quantified proteins. Proteins without any homology to already known sequences are marked in red. The X-axis represents the calculated log2 fold change of proteins, the Y axis represents the total intensity of each protein.

Supplementary figure 2. Representative examples of deregulated peptides in three different measurements
Peptide spectra of three different proteins (column 1 to 3) and 3 measurements (rows 1-3) are shown. The upper row represents the heavy peptide pool, the middle row the combination of the labeled (heavy) pool with peptides from uninjured tissue, and lower row the combination of the heavy pool with peptides from regenerating hearts six hours after injury. Peptides within one column show different heavy/light ratios between different combination of tissue lysates indicating changes in expression levels. The protein Peptidyl-Lys Metalloendopeptidase (left column) is up-regulated. The heavy Pool ratio is almost 1, the ratio from the combination with uninjured tissue is 0.7 and the ratio from the combination with regenerating tissue is 0.3. The increased ratios of the un-labeled (light) spectrum to the labeled (heavy) spectrum between the middle and the lower rows indicate up-regulation. The expression level of protein Rab14 (middle column) did not change since ratios in the middle and lower rows are almost identical. The SPARC protein (right column) is down-regulated since the peptide ratio in the middle row is 0.3 and the peptide ratio in the lower row is 0.5 (not logged). SPACR is down-regulated since only the non-labeled (light) spectrum is decreased. 2

Supplementary figure 3. Predicted protein domains of nsCCN, a putative new member of the CCN family
SP = signal peptide, IGFP = insulin –like growth factor binding domain, VWFC_2 = Willebrand factor type C domain, TSP1 = Thrombospondin type 1 repeat domain, CTCK = cysteine-knotcontaining module.

Supplementary figure 4. Investigation of CCN3 and CCN4 expression in ongoing heart regeneration

Expression of CCN family members CCN2, CCN3 and CCN4 for up to 35 days after mechanical injury (qRT-PCR). Expression levels are relative to the CTGF expression level in uninjured hearts.

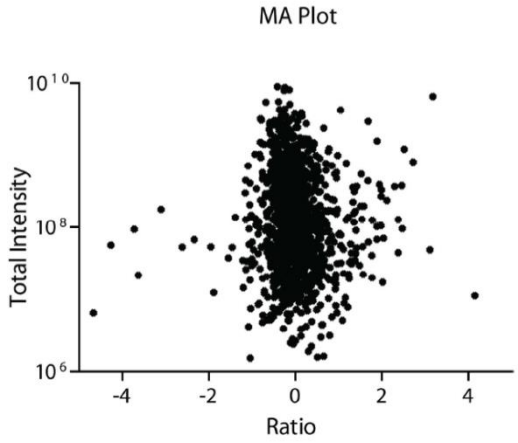
Supplementary figure 5. GO enrichment analysis reveals an overrepresentation of complement activation and inflammatory responses in regenerating newt hearts

Enrichment analysis of GO terms was calculated for all up-regulated proteins detected in the newt EST library (GO annotation file from homo sapiens). Enriched terms are marked in yellow (p value 10-3 to 10-5). The graph was generated using the GOrilla online tool 46.

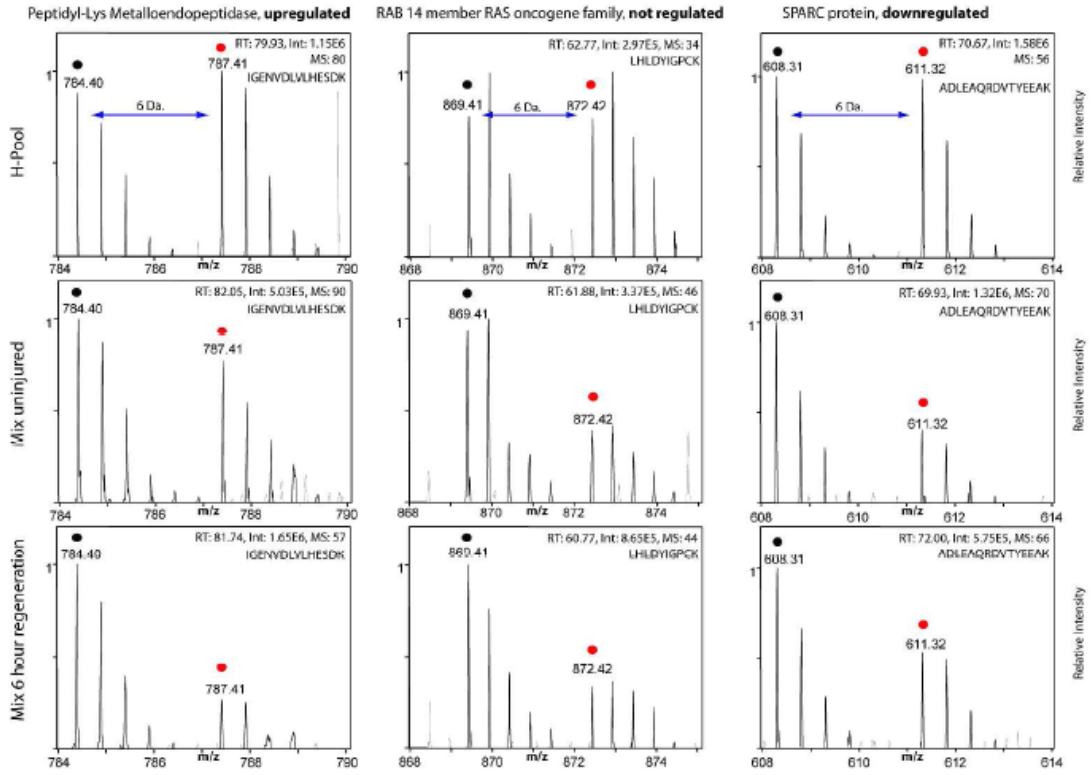
Supplementary figure 6. Newts possess two different Annexin A1 (ANXA1) isoforms that aredifferentially regulated during newt heart regeneration

Alignment of the two different ANXA1 isoforms that are either up-regulated (up) or downregulated (down) during heart regeneration. The core regions of the two identified ANXA1 isoforms show significant similarity to ref|NP_001080144.1| annexin A1 [Xenopus laevis]. The first line indicates the protein sequence of the up-regulated annexin Isoform (UP), the second line indicates the protein sequence of the down-regulated isoform (DOWN). Identical amino acids are highlighted in blue. Identified peptides in the up-regulated isoform are marked in yellow, peptides in the down-regulated form are marked in pink. On the bottom of the figure all identified peptides (Core region and flanking sequence) are listed.

Supplementary figure 1

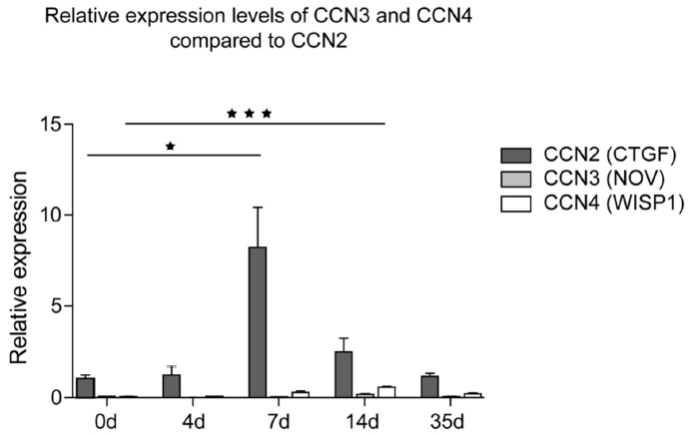


Supplementary figure 2

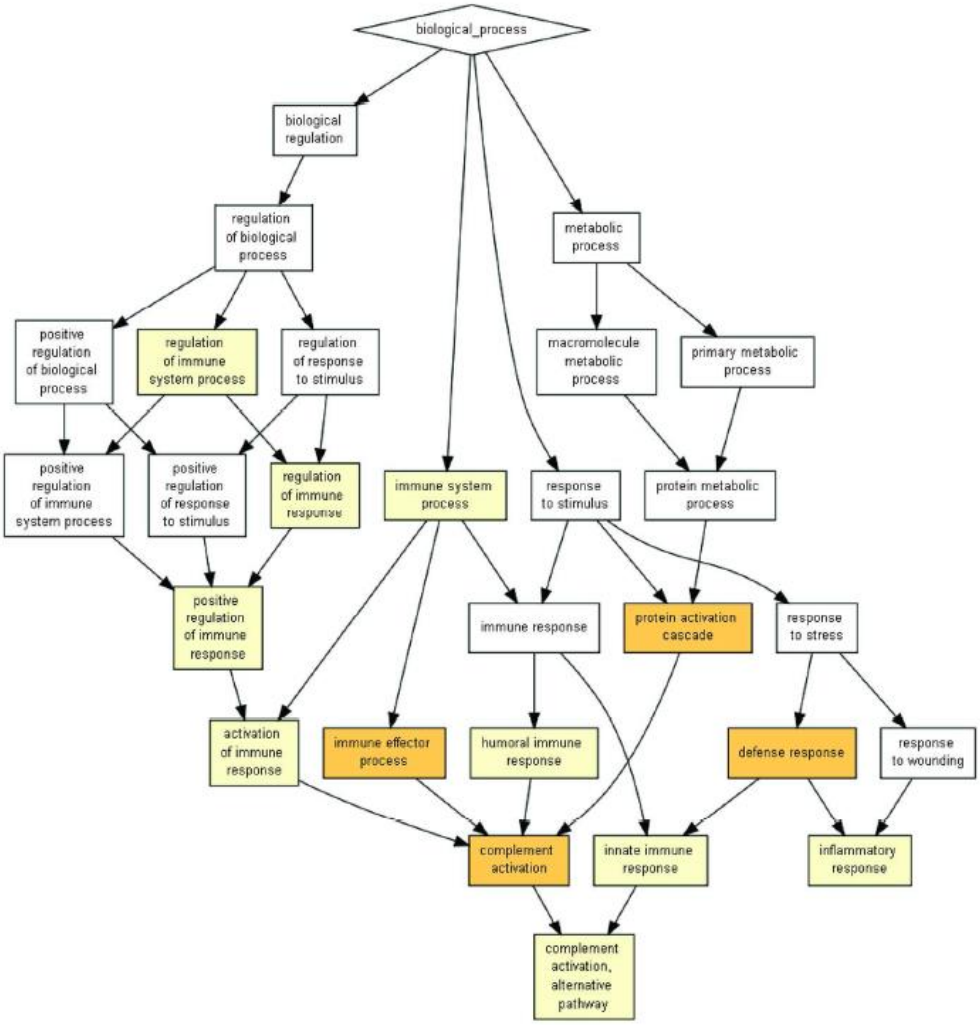


Supplementary figure 3

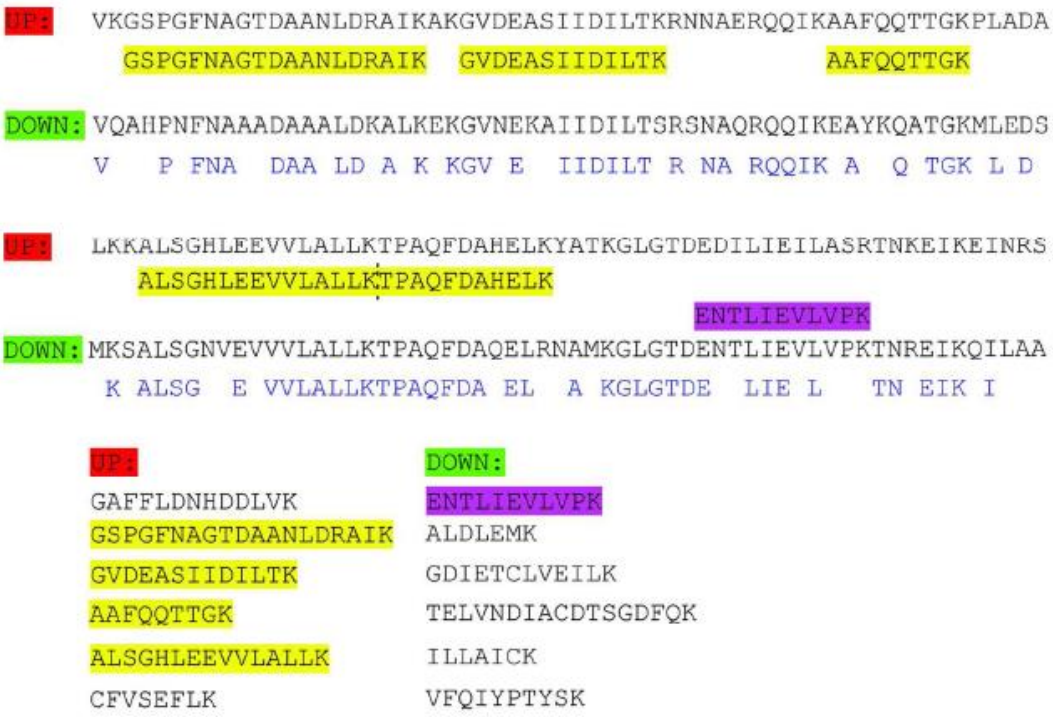
Supplementary figure 4



Supplementary figure 5



Supplementary figure 6



Supporting information:

MS excel:

http://pubs.acs.org/doi/suppl/10.1021/pr300521p/suppl_file/pr300521p_si_005.xlsx

http://pubs.acs.org/doi/suppl/10.1021/pr300521p/suppl_file/pr300521p_si_006.xlsx

http://pubs.acs.org/doi/suppl/10.1021/pr300521p/suppl_file/pr300521p_si_007.xlsx

http://pubs.acs.org/doi/suppl/10.1021/pr300521p/suppl_file/pr300521p_si_008.xlsx

http://pubs.acs.org/doi/suppl/10.1021/pr300521p/suppl_file/pr300521p_si_009.xlsx

8.2. Zusatzmaterial Veröffentlichung Nr. 2

Appendix 1. Array expression data.

To access the data, click or select the words "Appendix 1." This will initiate the download of a compressed (pdf) archive that contains the file.

http://www.molvis.org/molvis/v19/appendices/mv-v19-135-app1.pdf

Appendix 2. Functional annotation.

To access the data, click or select the words "Appendix 2." This will initiate the download of a compressed (pdf) archive that contains the file.

http://www.molvis.org/molvis/v19/appendices/mv-v19-135-app2.pdf

8.3. Zusatzmaterial Veröffentlichung Nr. 3

Additional files

Additional file 1: Performance of different assembly strategies. Performance comparison of four different assembly strategies comparing total number of transcripts, N50, transcripts larger 500 bp and transcripts larger 1000 bp.
http://genomebiology.com/imedia/7837603139140151/supp1.xlsx

Additional file 2: Overall distribution of transcript annotation rate as a function of sequence length. Transcript length (x axis, log scale) is plotted against the percentage of overall annotation (y axis). E-value cut-offs from e-10 to e-200 are marked in different colours. Dashed line demarks the sequence length above transcripts were chosen for further analysis.
http://genomebiology.com/imedia/1958137533914015/supp2.jpeg

Additional file 3: Coverage of de novo assembled newt transcript with high quality annotations on human signaling pathways. 58 members of the human p53 signaling pathway are matched by 47 proteins present in the assembled newt transcriptome. The use of high quality threshold criteria might have prevented detected of all family members.
http://genomebiology.com/imedia/9923170789140154/supp3.jpeg

Additional file 4: Coverage of de novo assembled newt transcript with high quality annotations on human signaling pathways. The TGF beta signaling pathway containing 51members is covered by 41 newt transcripts. Candidates identified by gene symbols are marked in pink, candidates that were not identified are marked in purple. Pathway nodes including multiple candidates that are only partially represented in the newt transcriptome are marked in dark red.
http://genomebiology.com/imedia/1810124991914016/supp4.jpeg

Additional file 5: List of all transcripts assigned to orthologues and corresponding length distribution in respect to the subject sequences.
http://genomebiology.com/imedia/7143004991401649/supp5.xlsx

Additional file 6: List of all potential protein-coding transcripts that were validated by corresponding peptides. The number of identified frameshifts and the total number of identified peptides is listed. The second sheet gives the example of a single candidate where peptides and alignments identified a frameshift in the nucleotide sequence.
http://genomebiology.com/imedia/1723716166914016/supp6.xlsx

Additional file 7: List of all identified PFAM domains in transcripts that lacked any sequence similarity to higher organisms. The second sheet includes the domains found in candidates presented in the manuscript.
http://genomebiology.com/imedia/1598623908914017/supp7.xlsx

Additional file 8: Comparative hierarchical clustering of heart and lens expression values. Hierarchical clustering of expressions levels in regenerating hearts (column 1-9), and lenses (dorsal iris: column 10-12; ventral iris:column 13-15) during regeneration. Only transcripts with valid array expressions for at least 13 columns are represented. The blue cluster represents a subset of transcripts that are down-regulated at least at two stages of heart and lens regeneration. The yellow cluster marks a set of transcripts that is up-regulated during late stages of heart regeneration but lacks an obvious pattern in the regenerating lens. The red cluster represents a set of transcripts that is inversely regulated at late stages of lens regeneration but lacks an obvious pattern in the in the regenerating heart with the exception of a smaller subfraction that was strongly up-regulated during early heart regeneration (6 hours after heart injury). The green cluster marks a set of transcripts that is uniformly upregulated during late stages of heart and lens regeneration. The purple cluster highlights a set of transcripts that is strongly up-regulated in regenerating lens and in the regenerating heart between 1 to 4d after damage. All heatmap members with cluster affiliation and expression values are provided in additional file 9.
http://genomebiology.com/imedia/3820627491401725/supp8.jpeg

Additional file 9: Expression data of all microarray spots presented in the heatmap (additional file 8).
http://genomebiology.com/imedia/8208115149197719/supp9.xlsx

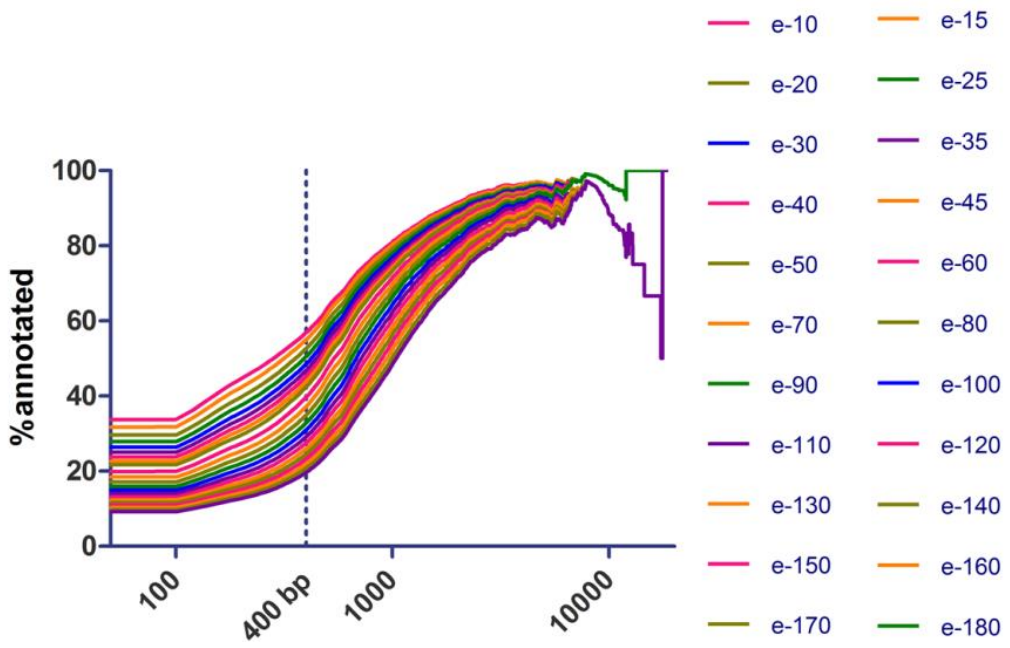
Additional file 10: Microarray expression data of regenerating heart and lens tissues are presented. All candidates are shown, which were described in the study and represented on microarrays.
http://genomebiology.com/imedia/7980503209197715/supp10.xlsx

Additional file 11: Expression of newly identified genes in regenerating newt tissues. Real time RT-PCR analysis ($n \geq 3$) of selected candidates in regenerating adult newt hearts, lenses and limbs. Values were normalized to the 0 time point and to tissues with highest expression levels.
http://genomebiology.com/imedia/9340917689197719/supp11.jpeg

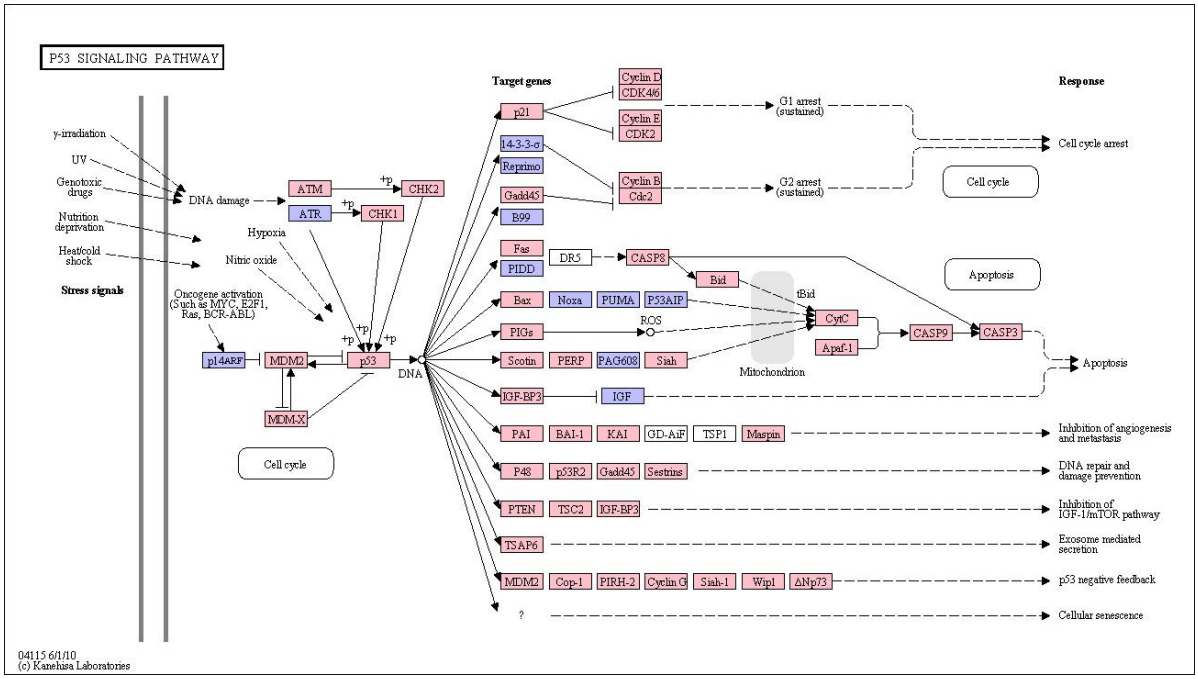
Additional file 12: Nucleotide and translated amino acid sequences of all candidate molecules described in the manuscript.
http://genomebiology.com/imedia/2012453715919771/supp12.txt

Additional file 13: List of all primers used.
http://genomebiology.com/imedia/1753700113919771/supp13.txt

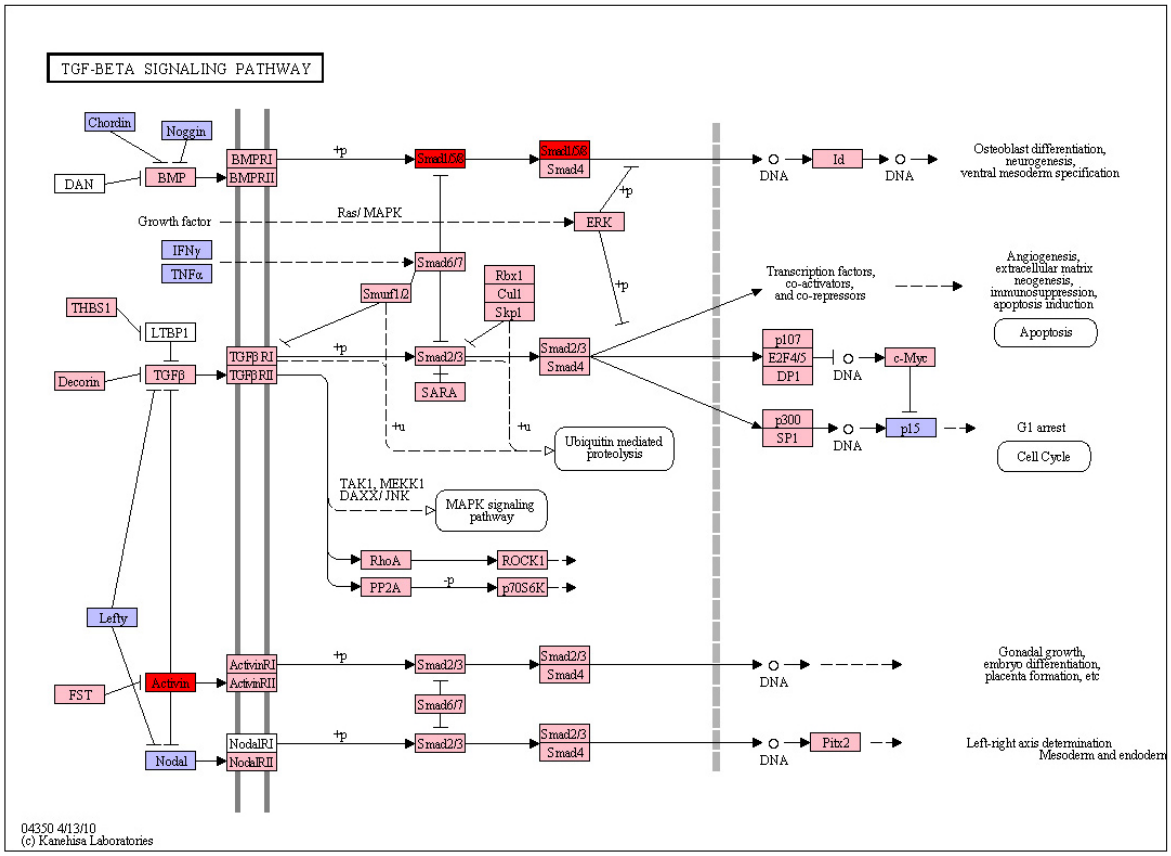
Additional file 2:



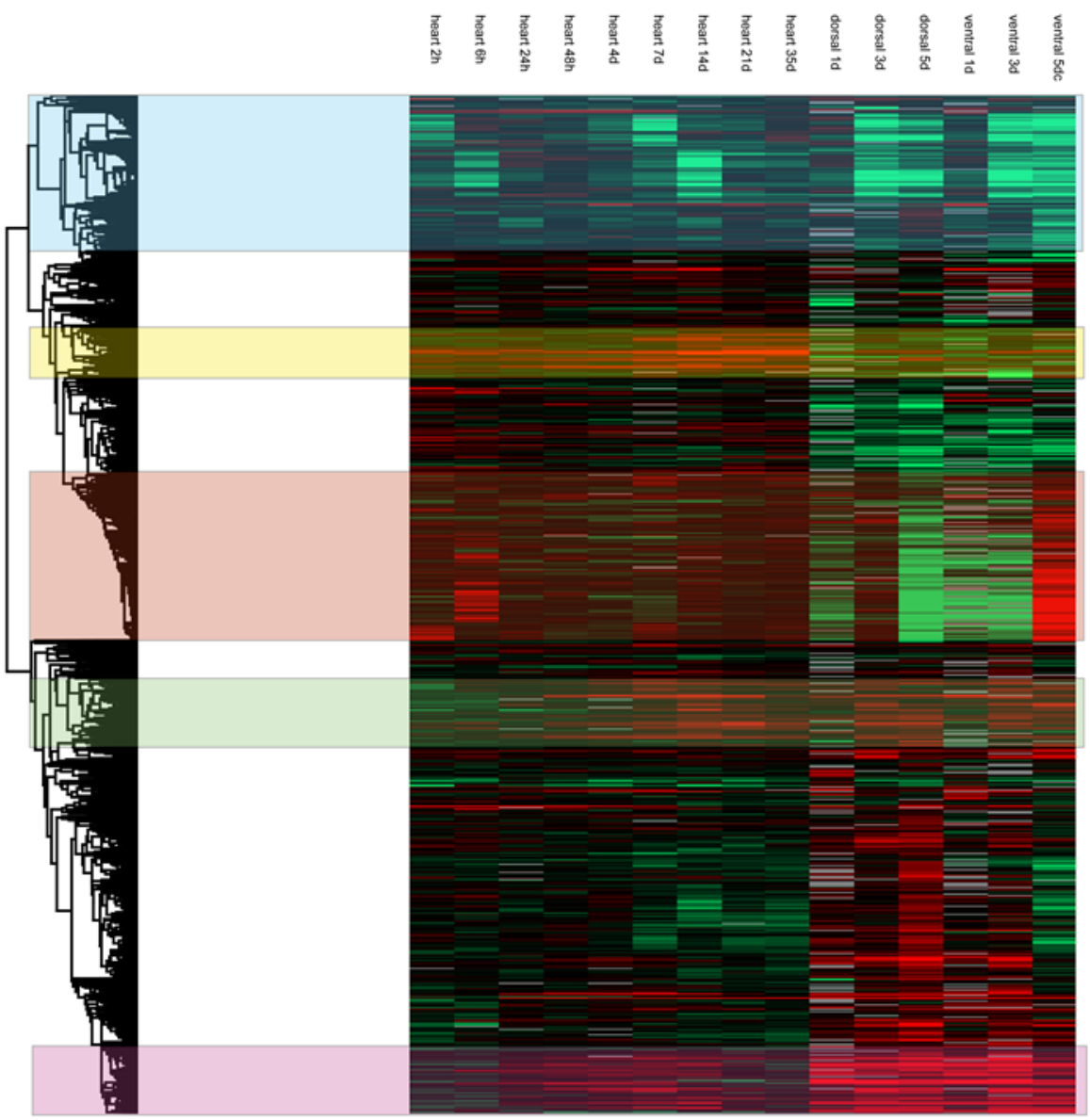
Additional file 3:



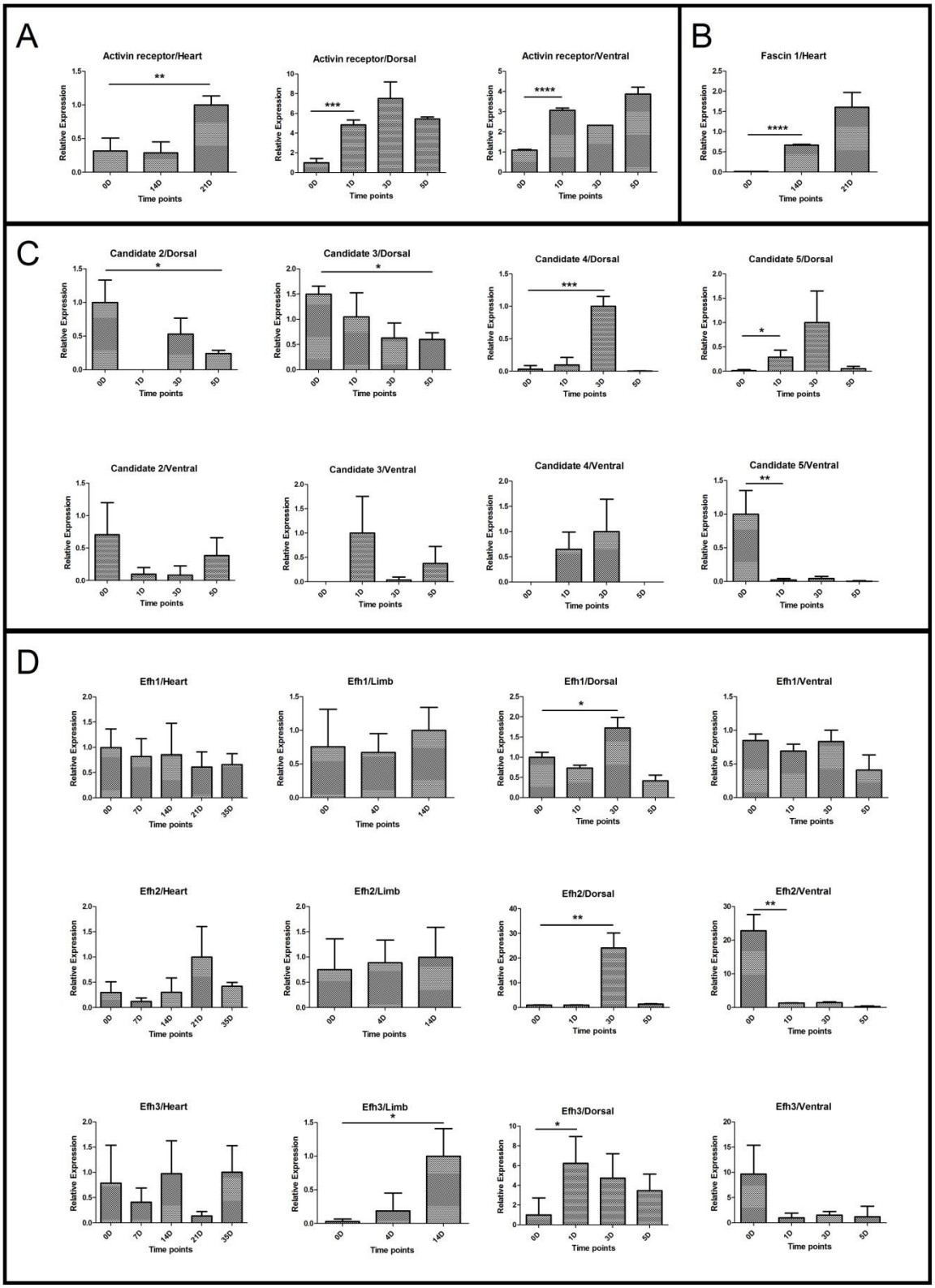
Additional file 4:



Additional file 8:



Additional file 11:



8.4. Erklärung zur Dissertation

„Hiermit erkläre ich, dass ich die vorliegende Arbeit selbständig und ohne unzulässige Hilfe oder Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Alle Textstellen, die wörtlich oder sinngemäß aus veröffentlichten oder nichtveröffentlichten 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 sowie ethische, datenschutzrechtliche und tierschutzrechtliche Grundsätze befolgt. Ich versichere, dass Dritte von mir weder unmittelbar noch mittelbar geldwerte Leistungen für Arbeiten erhalten haben, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen, oder habe diese nachstehend spezifiziert. Die vorgelegte Arbeit wurde weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde zum Zweck einer Promotion oder eines anderen Prüfungsverfahrens vorgelegt. Alles aus anderen Quellen und von anderen Personen übernommene Material, das in der Arbeit verwendet wurde oder auf das direkt Bezug genommen wird, wurde als solches kenntlich gemacht. Insbesondere wurden alle Personen genannt, die direkt und indirekt an der Entstehung der vorliegenden Arbeit beteiligt waren. Mit der Überprüfung meiner Arbeit durch eine Plagiatserkennungssoftware bzw. ein internetbasiertes Softwareprogramm erkläre ich mich einverstanden.“

_____ _____

Ort, Datum Unterschrift

8.5. Danksagung

Ich möchte mich bei allen bedanken, die mich bei der Entstehung dieser Arbeit unterstützt und begleitet haben. Mein Dank geht insbesondere an Prof. Dr. Dr. Thomas Braun für die Betreuung meiner Doktorarbeit in seiner Arbeitsgruppe und die vielen hilfreichen Ratschläge in diversen Meetings und Gesprächen. Auch möchte ich meinen lieben Eltern danken, ohne die das Studium und die nachfolgende Dissertation nicht möglich gewesen wären.

Während meiner Forschungstätigkeit im Labor und am Rechner standen mir immer zwei Personen zur Seite ohne die diese Arbeit nicht möglich gewesen wäre. Ich möchte mich bei Dr. Thilo Borchardt und Dr. Mario Looso für die jahrelange Unterstützung und Hilfe recht herzlich bedanken. Mit Ihrer konstruktiven Kritik und durch zahlreiche Diskussionen fand ich immer wieder eine Motivation das begonnene Projekt in dieser Form zu beenden.

Prof. Dr. Ralph Schermuly möchte für eine unkomplizierte und rasche Übernahme der Zweitkorrektur danken. Mein weiterer Dank gilt allen Mitarbeitern des Max-Planck-Instituts, die mich bei sowohl administrativen als auch labortechnischen Tätigkeiten unterstützt haben.