

Identifikation und Katalogisierung retinal exprimierter Gene

der Fakultät für Biologie
der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines Doktors
der Naturwissenschaften

von
Jens Meyer
aus Emden

vorgelegte
D i s s e r t a t i o n

2006

Tag der mündlichen Prüfung: 15.11.2006

Dekan: Prof. Dr. F. Schöfl
1. Berichterstatter: Prof. Dr. E. Zrenner
2. Berichterstatter: Prof. Dr. F. Götz

„Wir stehen selbst enttäuscht und sehen betroffen
den Vorhang zu und alle Fragen offen.“

Berthold Brecht

Der gute Mensch von Sezuan

„Wissenschaft ist Landgewinnung im Meer der Unwissenheit.“

Sir Isaac Newton

Philosophiae Naturalis Principia Mathematica

Danksagung

Herrn **Prof. Dr. Eberhart Zrenner** danke ich für die Möglichkeit, den Zivildienst und die Doktorarbeit in der Abteilung II der Universitäts-Augenklinik Tübingen gemacht zu haben. Herrn **Prof. Dr. Friedrich Götz** gilt mein Dank für die freundliche Übernahme des Zweitgutachtens.

Ganz herzlich bedanke ich mich bei Herrn **Dr. Bernd Wissinger** für die überlassenen Fragestellungen, die sorgfältige Korrektur der vorliegenden Arbeit sowie für die eingeräumten Freiheiten während des Projektes. Weiterhin für die Bereitschaft zur konstruktiven Diskussion unerwarteter Resultate.

Bei Herrn **Dr. Reinhold Brückner** möchte ich mich bedanken für die Hilfsbereitschaft bei der Durchsicht des Manuskripts sowie die aufschlußreichen (forschungs-)politischen Gespräche.

Im weiteren gilt mein Dank:

Valeria Roni for spiritual, cultural and intellectual support. And for enduring my antithetic point of view.

Christiane „dark lady of DNA“ Wolf für ein geteiltes Tübingen-Bild und gemeinsame kulinarische Exzesse.

Ronald Carpio for enjoying/enduring sarcastic comments and the German weather.

Jochen Scholl und erneut **Christiane Wolf** für die auf höchstem Niveau ausgefochtenen Billard-Schlachten und die Gespräche über Sinn und Unsinn von Vitaminen.

Meinen Leidensgenossen, den Doktoranten **Marcel Alavi**, **Susann Dangel**, **Nico Fuhrmann**, **Véronique Kitiratschky**, **Katja Koeppen** und **Simone Schimpf** sowie den PostDocs **Susanne Kohl** und **Nicole Weißschuh** danke ich für die angenehme Arbeitsatmosphäre und die frauenspezifischen Gesprächsthemen beim Mittagstisch. Nicht zu vergessen der Dank an die heimlichen Stars des Labors, die Damen und Herren von der technischen Front: **Britta Baumann** für viele hilfreiche Tipps, **Simone Schaich** und **Jochen Scholl** wegen kompetenter Hilfe bei der Optimierung der Bakterienpräparation. **Sabine Tippmann** für cinematische Inspiration und nicht zuletzt **Eva Weber** für freundliche Gespräche an der Bench.

Dr. Michael Becker für technische Reparaturen in Not und die Einsicht, dass fünf Sicherheitskopien besser sind als vier.

Dr. Christoph Linnemann für die geduldige Einführung in die Welt der RealTime-PCR

PD Dr. Kay Nieselt-Struwe und **PD Dr. Jochen Scheel** für initiale Beratung bei der Auswahl des Datenbanksystems.

Prof. Dr. Hans Probst, **Dr. Maria van Betteraey** und **Dr. Hans-Jörg Riedinger** für die freundliche Unterstützung bei den Hypoxie-Experimenten.

Herzlichsten Dank an **Dr. Carsten Pusch** für die initiale Durchsicht des Manuskripts und den geteilten Filmgeschmack.

Ralf Schäfer (nicht nur, aber auch) für unschätzbare Hilfe bei der Installation und Bedienung von MySQL und Linux.

Trotz und wegen allem ergeht auch ein Dank an **Michael Schlenker** für die Programmierung beider Versionen des Querymasters.

Mein ganz besonderer Dank gilt **meinen Eltern** für einfach alles.

Inhaltsverzeichnis

	Vorbemerkung	1
1	Einleitung	2
1.1	Das Auge	2
1.2	Die Retina	3
1.3	Erblindungen als Folge einer Hypoxie	5
1.4	cDNA-Datenbanken und Data Mining	10
1.4.1	Erstellung einer cDNA-Bibliothek	11
1.4.2	Bioinformatische Auswertung einer cDNA-Bank	13
1.4.3	Metaanalyse	15
2	Material	18
2.1	Verbrauchsmaterialien und Reaktionskits	18
2.2	Enzyme	19
2.3	Antibiotika	20
2.4	Längenstandards	20
2.5	Elektrophorese-Puffer	20
2.6	Ladepuffer	20
2.7	Kompetente Zellen	21
2.8	Geräte	21
2.9	Medien	22
2.10	PCR-Puffer	23
2.11	Primerliste	24
2.12	Verwendete Computerprogramme	24
3	Methoden	25
3.1	Molekularbiologische Methoden zur Erstellung einer cDNA-Bank	25
3.1.1	Zellkultur WERI	25
3.1.2	Kryokonservierung von WERI-Zellen	25
3.1.3	Auftauen von WERI-Zellen	25
3.1.4	Aufschluss mit Trizol	26
3.1.5	CsCl-Dichtegradientenzentrifugation	26

3.1.6	Aufreinigung von Poly-A mRNA durch Oligotex-Säulen	27
3.1.7	Erstellung einer retinalen cDNA-Bibliothek durch das Superscript Plasmid System	27
3.1.8	Transformation durch Elektroporation	28
3.1.9	Bakterienanzucht	29
3.1.10	Glycerinkultur	29
3.1.11	Präparation bakterieller Plasmid-DNA durch R.E.A.L. Prep 96	29
3.1.12	Bestimmung von Insertgrößen	30
3.1.13	Elektrophoretische DNA-Auftrennung in Agarosegelen	31
3.1.14	Elektrophoretische RNA-Auftrennung in Agarosegelen	31
3.1.15	Polymerase Kettenreaktion (PCR)	32
3.1.16	Cycle-Sequenzierung	32
3.1.17	Elektrophoretische Auftrennung durch den ABI Prism 377 DNA-Sequencer	33
3.2	Molekularbiologische Methoden zur Bestimmung von Genregulationen durch Hypoxie	34
3.2.1	Zellkulturen unter hypoxischen Bedingungen	34
3.2.2	Bestimmen der WERI-Zellzahl in einer Neubauer-Zählkammer	34
3.2.3	Isolierung totaler RNA durch das RNeasy Mini Kit-System	34
3.2.4	Aufbewahrung von RNA	34
3.2.5	Photometrische Konzentrationsbestimmung von DNA und RNA	35
3.2.6	Quantitative Real Time PCR (Q-PCR)	36
3.3	Bioinformatik	38
3.3.1	Verwendete Internetseiten	38
3.3.2	Verwendete Programmiersprachen	38
3.3.2.1	Tcl/Tk 8.4.2	38
3.3.2.2	Python 2.2.2	39
3.3.3	Verwendete Tcl/Tk-Module	39
3.3.3.1	SDX zum Wrappen der Querymaster-Module	39
3.3.3.2	XOTcl 1.0	40
3.3.3.3	tDOM 0.8	40
3.3.3.4	tkcon 2.3	41
3.3.3.5	Tcllib	41
3.3.3.6	MySQLTcl 2.5	41
3.3.4	Verwendete Clustering-Module	42

3.3.4.1	StackPACK 2.2	42
3.3.4.2	Stack_ImportFasta / Stack_ImportGenebank	42
3.3.4.3	Cross_match 0.990319	43
3.3.4.4	D2_cluster	44
3.3.4.5	Phrap 0.990319	44
3.3.4.6	CRAW	45
3.3.4.7	Link_out	45
3.3.5	Automatisierte Blast-Abfrage und Annotation	46
3.3.5.1	Querymaster I	46
3.3.5.2	Querymaster II	46
3.3.6	Verwaltung der MySQL-Ergebnisse durch PhpMyAdmin 2.6.1	49
3.3.7	Datenbankmanagement	49
3.3.7.1	MySQL 3.23.27	49
3.3.7.2	Erstellung der „nr-eye“-Datenbank	50
4	Ergebnisse	51
4.1	Erstellung und Analyse einer cDNA-Bank aus WERI <i>Rb1</i> -Zellen	51
4.1.1	Generierung der cDNA-Bank	51
4.1.2	Optimierung der Plasmidextraktion aus der cDNA-Bank	51
4.1.3	Auswertung der Sequenzen aus der cDNA-Bank	53
4.2	Einfluss hypoxischer Bedingungen auf WERI <i>Rb1</i> -Zellen	57
4.2.1	Auswahl der Real Time-Primer	57
4.2.2	Anzucht unter hypoxischen und normoxischen Bedingungen	57
4.2.3	Auswertung der Real Time-Ergebnisse	58
4.3	Erstellung einer vollständig annotierten humanen retinalen in vitro-Datenbank	61
4.3.1	Verlaufdiagramm der Datenbankerstellung	61
4.3.2	Auswahl der eingesetzten Sequenzen	61
4.3.3	Clustern der ESTs/cDNAs mit dem StackPACK-Programm	62
4.3.4	Entwicklung des Querymaster II	64
4.3.4.1	Auswahl der Programmiersprache für den Querymaster II	64
4.3.4.2	Blast-Analyse der geclusterten Sequenzen durch den Querymaster II	65
4.3.4.3	Generieren eines „best hits“ aus einer Blast-Ergebnisliste	67

4.3.4.4	Ansteuerung externer Datenbanken zur Annotierung der Blastresultate	69
4.3.4.5	AbSpeicherung der kombinierten Resultate in eine relationale MySQL-Datenbank	70
4.4	Auswertungen der MySQL-Datenbanken	72
4.4.1	Umfang der Ausgangsdaten	72
4.4.2	Erstellung und Auswertung der „nr-eye“-Datenbank	73
4.4.3	Auswertung der TRJ8-Datenbank	79
5.	Diskussion	81
5.1	Diskussion der cDNA-Bibliothek	81
5.1.1	Auswertung der WERI Rb1-Klone durch den QM I	81
5.1.2	Aussicht	82
5.2	Hypoxie-induzierte Expressionsanalyse der Zelllinie WERI Rb1	83
5.3	Erstellung und Analyse der „nr-eye“-Datenbank	87
5.3.1	Zielsetzung	87
5.3.2.	Clusterung durch StackPACK	87
5.3.3	Ansteuerung der NCBI-Datenbanken durch den QM II	88
5.3.4	Ansteuerung der MySQL-Tabellen durch PHPMyAdmin	90
5.3.5	Auswertung der MySQL-Tabellen	90
5.3.6	Ausblick	93
6.	Zusammenfassung	96
7.	Literaturverzeichnis	99
8.	Appendix	111
8.1	Liste der informativen Klone aus der lokal erstellten humanen retinalen cDNA-Bank	111
8.2	Abkürzungsverzeichnis	133

Vorbemerkung

Bis zum sechsten vorchristlichen Jahrhundert war unsere westliche, damals griechische, Kultur in Mythen und Legenden gefangen. Praktische Erfahrungswerte waren zwar vorhanden, es fehlte aber die Kenntnis der Zusammenhänge, eine Systematik und vor allem eine anerkannte Methode zur Erlangung von gültigem Wissen. Götter und Halbgötter nahmen die Rolle ein, die heute die Naturgesetze spielen. Im sechsten Jahrhundert vor Christi geriet die bis dahin gültige Überlieferung wegen offensichtlicher Widersprüche und Unzulänglichkeiten in die Kritik. In den folgenden zwei Jahrhunderten entstand aus diesem kritischen, vernunftbetonten Ansatz die klassische griechische Wissenschaft (Thales, Pythagoras, Heraklit). Diese war bereits recht hoch entwickelt, jedoch hauptsächlich schöngeistig orientiert. In der Renaissance und besonders der Aufklärung wurden schließlich die Methoden entwickelt, auf denen die moderne Wissenschaft fußt. Besonders zu erwähnen sind hier die Empirik, nach der alle Erkenntnis nur aus der Erfahrung abgeleitet werden darf (Hobbes, Locke, Hume) und der Reduktionismus, d.h. die Vorgehensweise, alle nicht holistisch erfassbaren Probleme in kleinere, lösbare Teilprobleme aufzugliedern (Descartes, Leibnitz, aber zum Teil bereits Demokrit). In der Moderne geriet die letzte Methode trotz unzweifelhafter Vorteile immer mehr in Verruf, weil komplexe Probleme wie etwa der Treibhauseffekt sich nicht gänzlich durch Betrachtungen einzelner Teilaspekte lösen lassen. In der Biologie können monogene Erbkrankheiten mit den klassischen Sichtweisen sehr gut erklärt werden, diese versagen jedoch weitgehend bei der Ergründung komplexer Erkrankungen, welche z.B. polygen und/oder durch Umwelteinflüsse begründet sind. Durch die Entwicklung von hochparallelisierten Untersuchungs- und Auswertungsmethoden wie der Chiptechnologie oder der Bioinformatik erschließen sich zunehmend auch diese Krankheitsbilder. Somit wird ein Paradigmenwechsel eingeleitet, der uns erneut eine holistische Betrachtungsweise ermöglichen könnte.

1. Einleitung

1.1 Das Auge

Das menschliche Auge hat einen Durchmesser von etwa 24 mm. Der kugelförmige Augapfel (*Bulbus oculi*) ist in die Augenhöhle (*Orbita*) eingebettet, die von Stirnbein, Jochbein und Oberkieferknochen gebildet wird. Er umschließt die mit Kammerwasser gefüllte vordere und hintere Augenkammer sowie den Glaskörper (*Corpus vitreum*). Der Augapfel wird von der Lederhaut, Aderhaut und Netzhaut ausgekleidet. Die aus derbem Bindegewebe bestehende Lederhaut (*Sklera*) bildet die äußere Schicht. Sie geht im vorderen Teil in die durchsichtige Hornhaut über. Die Hornhaut richtet das auffallende Licht, welches auf sie trifft, als Sammellinse nach innen und hilft sie zu ordnen, so dass auf der Netzhaut ein scharfes Bild entstehen kann. Auf die Lederhaut folgt nach innen zu die gut durchblutete Aderhaut (*Chorioidea*). Pigmente in, bzw. vor der Aderhaut absorbieren das Licht, das die Retina durchdringt. An die Aderhaut schließt sich nach innen zu die Netzhaut (Retina) an, von der die einfallenden Lichtreize aufgenommen und die entsprechenden Erregungen über den Sehnerv zum Gehirn weitergeleitet werden. Die vordere Augenkammer wird hinten durch die ringförmige Regenbogenhaut (Iris) begrenzt, die von Ausläufern der Aderhaut als auch der Netzhaut gebildet wird. Sie gibt dem Auge die charakteristische Färbung und absorbiert außerhalb der Sehöffnung einstrahlendes Licht. Die Regenbogenhaut liegt der Augenlinse auf und umgrenzt die Pupille, welche die Sehöffnung darstellt. Hinter Pupille und Regenbogenhaut, in eine Ausbuchtung des Glaskörpers eingebettet, liegt die Linse. Sie ist aus Schichten unterschiedlicher Brechkraft aufgebaut und wird von einer durchsichtigen, elastischen Membran umschlossen. Die Aufhängevorrichtung, durch welche die Linse in ihrer Lage festgehalten wird, besteht aus Zonulafasern, die dem Zilliarkörper des Auges entspringen. Der Zilliarkörper hat einen ringförmigen Muskelstreifen (Zilliarmuskel), bei dessen Kontraktion die Zonularfasern erschlaffen, so dass die Linsenwölbung zunimmt. Erschlafft der Muskel, so wird die Linse durch die Zugwirkung der Zonulafasern flachgezogen. Durch diese Veränderung ihrer Brechkraft ermöglicht die Linse das Nah- und Fernsehen. Hornhaut, Linse, vorderer Augenkammer und Glaskörper bilden den bildgebenden (dioptrischen) Apparat des Auges. Dabei ist die Flüssigkeit in der vorderen Augenkammer, das Kammerwasser, optisch so der Hornhaut angepasst, dass beide das Licht annähernd gleich stark brechen. An der gesamten Brechkraft des dioptrischen Apparates ist die Linse mit 2-19

Einleitung

Dioptrien und die Hornhaut mit 43 Dioptrien beteiligt. Das vom dioptrischen Apparat abgebildete Bild wird von der Netzhaut rezipiert und in Nervenimpulse umgewandelt.

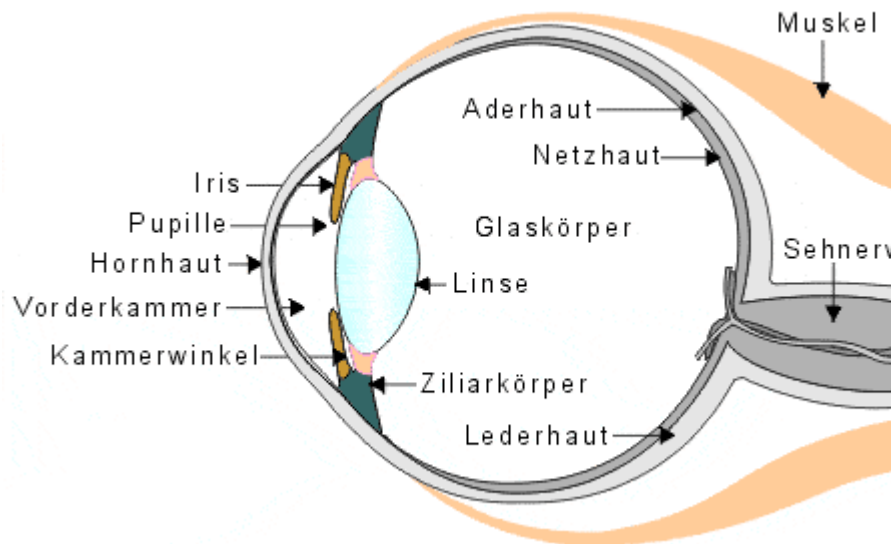


Abbildung 1: Darstellung des menschlichen Auges.
(aus: http://www.optik-rauschmann.de/Das_Auge.htm)

1.2 Die Retina

In der lichtabgewandten Netzhaut liegen die Zapfen und die Stäbchen. Die Stäbchen sind etwa 10.000mal lichtempfindlicher als die Zapfen und haben ihre höchste Dichte in der Peripherie der Netzhaut, sie sind für das skotopische Sehen bei Dämmerlicht zuständig. Ihr Photopigment, das Rhodopsin hat eine maximale Absorption bei 498 nm. Im Zentrum der Netzhaut überwiegen die Zapfen. Beim Menschen gibt es drei Typen, die mit Photopigmenten mit unterschiedlichen spektralen Absorptionsmaxima (Rotopsin: $\lambda=563$ nm, Grünopsin: $\lambda=533$ nm und Blauopsin: $\lambda=437$ nm) ausgestattet sind. Sie sind für das photopische Sehen bei hohen Lichtintensitäten und das Farbsehen verantwortlich. Die höchste Dichte erreichen die Zapfen in der Sehgrube (*Fovea centralis*), die inmitten des gelben Flecks (*Macula lutea*) liegt. Der gelbe Fleck ist daher der Ort der besten Auflösung und Farbunterscheidung. In der menschlichen Netzhaut gibt es etwa 125 Millionen Sehzellen, dabei etwa zwanzigmal mehr Stäbchen als Zapfen.

Einleitung

Die linsenseitig gelegenen Fortsätze der Netzhautganglienzellen vereinigen sich zum Sehnerv (*Nervus opticus*), der nahe dem Netzhautzentrum die Netzhaut durchdringt und nach hinten aus dem Auge austritt. An dieser Stelle, dem blinden Fleck, enthält die Netzhaut keine Sehzellen, so daß eine Lichtempfindung fehlt.

Untersucht man die Feinstruktur der Retina, fällt deren schichtweiser Aufbau auf:

Das retinale Pigmentepithel bildet die äußerste Schicht der Retina. Ihre Zellen verhindern eine störende Reflexion des einfallenden Lichtes. Die Kerne der Photorezeptorzellen bilden die äußere Körnerschicht (*outer nuclear layer*). Die anschließende äußere plexiforme Schicht (*outer plexiform layer*) wird aus den synaptischen Endigungen der Photorezeptoren und Dendriten der Bipolar- und Horizontalzellen gebildet. Die bipolaren Nervenzellen stehen in synaptischer Verbindung zu den Photorezeptoren. Während in der übrigen Retina mehrere Rezeptoren konvergent auf eine bipolare Zelle und mehrere bipolare Zellen wiederum auf eine Ganglienzelle geschaltet sind, wird in der Fovea centralis, dem Ort des schärfsten Sehens je eine Rezeptorzelle über eine einzelne bipolare Zelle mit einer Ganglienzelle verbunden.

Die horizontale Verschaltung wird durch Horizontal- und Amakrinzellen gewährleistet, wobei erstere Verbindungen zwischen den Photorezeptorzellen und letztere zu den Bipolarzellen, den Ganglienzellen und/oder untereinander ausbilden. Die Zellkörper von Bipolar-, Amakrin- und Horizontalzellen bilden die innere Körnerschicht (*inner nuclear layer*).

Die innere plexiforme Schicht wird aus den Axonen der Bipolarzellen und Amakrinzellen, sowie den Dendriten der Ganglienzellen gebildet.

In der Ganglienzellschicht (*ganglion cell layer*) führen die Neuriten der Ganglienzellen zunächst marklos bis zum Blinden Fleck und danach markhaltig als Nervus Opticus zum Gehirn.

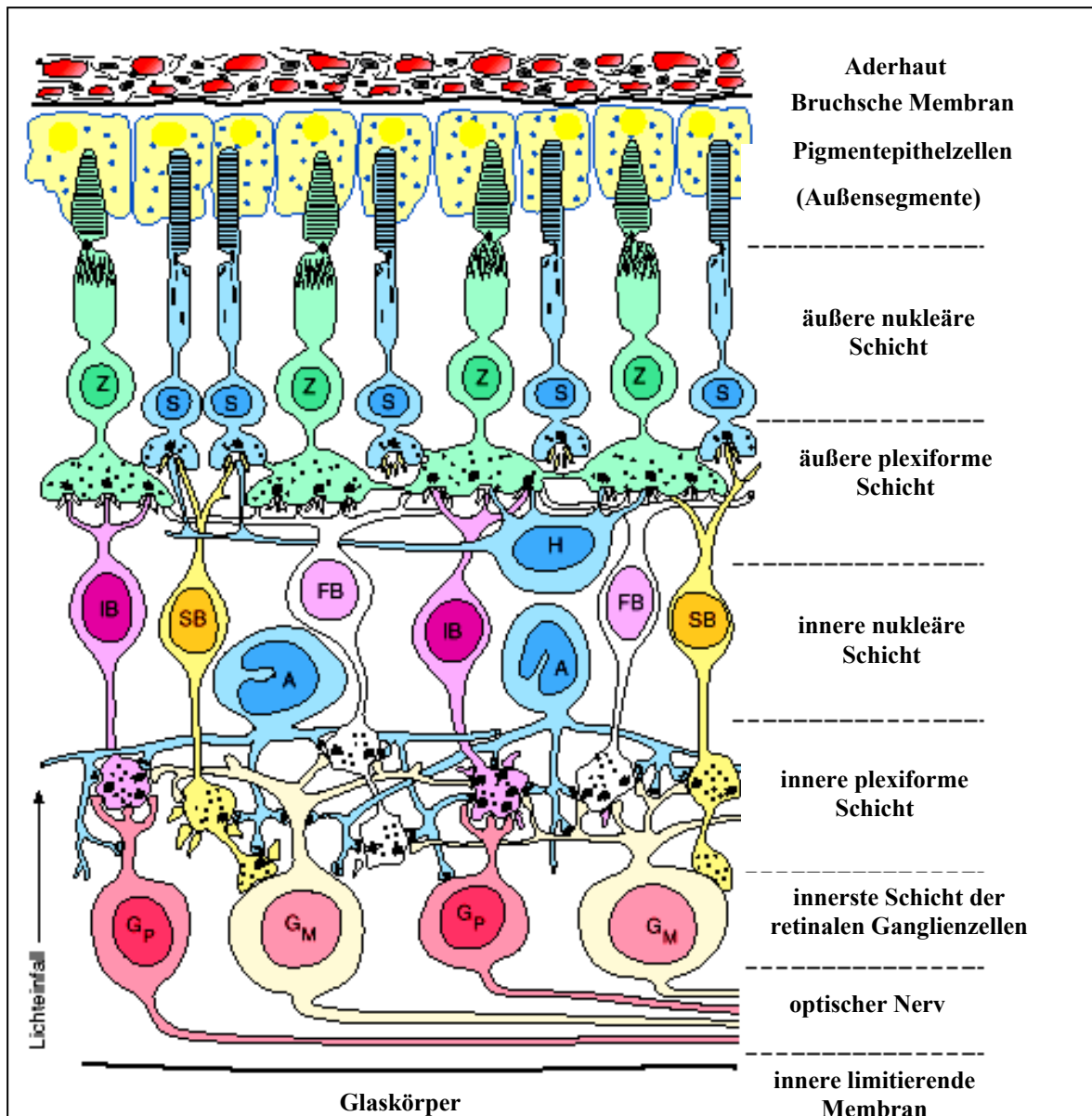


Abbildung 2: Darstellung der retinalen Zellschichten (aus Deetjen und Speckmann, 1999). Die Retina wird unterteilt in drei zellulären Schichten: die Ganglienzellschicht, die innere Körnerschicht und die äußere Körnerschicht. Die synaptische Verschaltung der zellulären Schichten geschieht in der inneren und der äußeren plexiformen Schicht.

1.3 Erblindungen als Folge einer Hypoxie

Die Weltgesundheitsorganisation bezifferte die Zahl der Blinden im Jahre 2004 weltweit mit 37 Millionen Menschen. Unter Blindheit versteht man den Verlust der Sehkraft, die Einschränkung der Sehfähigkeit auf unter 2% oder die Einschränkung des Gesichtsfeldes auf 5 Grad und weniger. Häufige Erblindungsursachen sind Infektionskrankheiten, altersabhängige Makulardegeneration, diabetische Retinopathie, etc. Während in der Vergangenheit die technischen Möglichkeiten überwiegend nur zur Aufklärung

Einleitung

monogenetischer Erkrankungen ausreichen, führt der technische Fortschritt vermehrt zur Aufklärung polygener und/oder umweltbedingter Erkrankungen. Wichtige Ursachen umweltinduzierter visueller Schädigungen sind die Auswirkungen eines verminderten Sauerstoffpartialdrucks.

Sauerstoff fungiert bei der oxidativen Phosphoryllierung als Elektronenakzeptor und ist somit essentiell für die Energiegewinnung in der Atmungskette. Zellen in höheren Organismen haben deswegen vielfache Adaptionsmechanismen entwickelt, um auf wechselnde Sauerstoffpartialdrücke angemessen reagieren zu können. Der Organismus reagiert als Ausgleich für ein gesunkenes Sauerstoffniveau mit Hyperventilation, verstärkter Produktion von roten Blutkörperchen und vermehrter Bildung von Kapillaren (Czyzk-Krzeska, 1997).

Auf der zellulären Ebene gewinnt mit fortschreitender Hypoxie die anaerobe Glycolyse gegenüber der oxidativen Phosphoryllierung an Bedeutung, zum Beispiel in stark beanspruchten Muskelzellen. Weiterhin wird u.a. mehr Glucose aufgenommen und das Niveau der entsprechenden Stressproteine wie z.B. *VEGF* angepasst (Bunn und Poyton, 1996).

Die Regulation durch Hypoxie-Stress induzierter Proteine geschieht überwiegend durch transkriptionelle Induktion über die Bindung des Transkriptionsfaktors *HIF1* (*Hypoxia Inducible Factor 1*) an die *HRE*-Kontrollregion (*Hypoxia Response Element*) der betreffenden Gene (Semenza *et al.*, 1991). Die zentrale *HRE*-Kontrollregion besteht aus einer RCGTG-Sequenz, die zuerst in der 3'-Enhancer-Region des Erythropoietins gefunden wurde (Semenza und Wang, 1992).

HIF1 ist ein Heterodimer und besteht aus je einem von drei möglichen Alpha-Untereinheiten, sowie einem HIF-1 β -Protein. Das *HIF1 α* -Protein ist 826 Aminosäuren lang und vermittelt die zentrale Antwort auf Sauerstoffmangel. *HIF2 α* übernimmt ähnliche Funktionen wie 1 α , wird allerdings nur in Macrophagen und Endothelzellen exprimiert (Ema *et al.*, 1997). Hingegen wird vermutet, dass *HIF3 α* ein negativer Regulator der HIF-regulierten Genexpression ist (Hara *et al.*, 1992). *HIF1 β* wurde als Aryl-Hydrocarbon-Nuclear-Rezeptor-Translokator (*ARNT*) identifiziert, welcher permanent und unabhängig vom Sauerstoff-Partialdruck gebildet wird und der Aktivierung des Dimers dient (Chapman-Smith *et al.*, 2004). Die Alpha-Untereinheiten werden ebenfalls unabhängig vom O₂-Level exprimiert, jedoch unter normoxischen Bedingungen durch den Ubiquitin-Proteasom-Pfad abgebaut: *HIF1* ist ein sehr

Einleitung

instabiles Protein, welches unter Normbedingungen eine Halbwertszeit von 10 Minuten hat (Salceda and Caro, 1997), was eine schnelle Anpassung an die Umweltbedingungen ermöglicht.

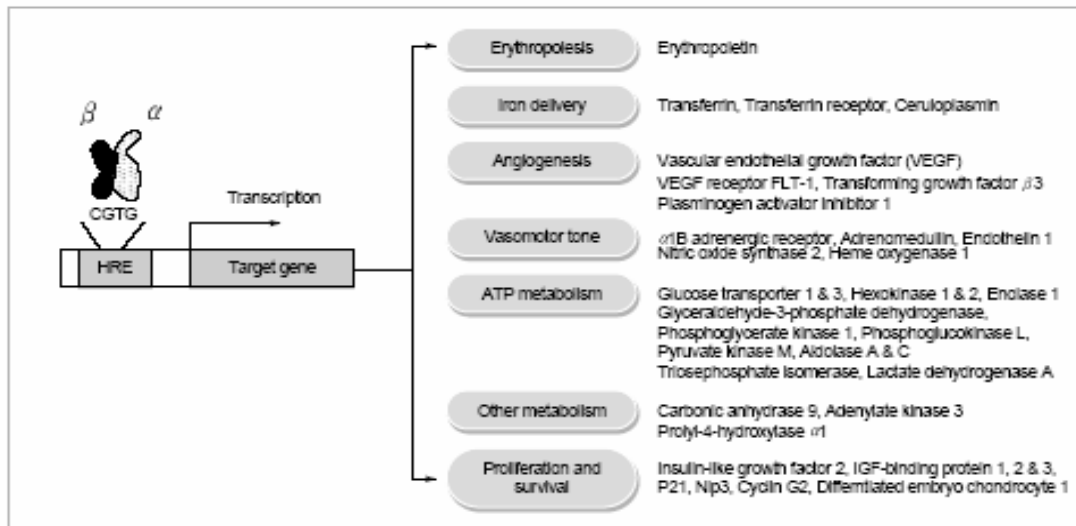


Abbildung 3: Übersicht über ausgewählte *HIF1*-regulierte Gene durch Bindung des *HIF*-Dimers an die *HRE*-Kontrollregion.
(aus: Chun *et al.*, 2002)

Eine Vielzahl unterschiedlicher Erblindungsursachen wird direkt oder indirekt mit einer Hypoxie in Verbindung gebracht:

Die diabetische Retinopathie ist eine der häufigsten Gründe für eine Blindheit in der westlichen Welt (Brinchmann-Hansen, 1992). Nach 15 Jahren sind 90% der Typ-I-Diabetiker und 25% der Typ-II-Diabetiker von visuellen Funktionsausfällen betroffen. Eine schlechte Diätführung, bzw. Blutzuckereinstellung begünstigt die Entstehung des Krankheitsbildes. 30% aller Erblindungen in Europa sind auf die diabetische Retinopathie zurückzuführen.

Die initialen Ursachen für die diabetische Retinopathie sind vielfältig: Durch die Hyperglykämie verschlechtern sich die rheologischen Eigenschaften der roten Blutkörperchen und somit auch die Versorgungsleistungen über die Kapillaren in der Netzhaut (Ditzel und Standl, 1975). Hinzu kommt die Verdickung der Basalmembranen der Kapillaren (Kern und Engermann, 1996), die veränderte Dissektionskurve für glycosyliertes Hämoglobin, sowie der höhere Sauerstoffverbrauch der Zellen bei höherem Glucoseangebot (van den Enden *et al.*, 1994). All diese Faktoren führen bei theoretisch ausreichendem Sauerstoffangebot zu einer funktionellen Hypoxie.

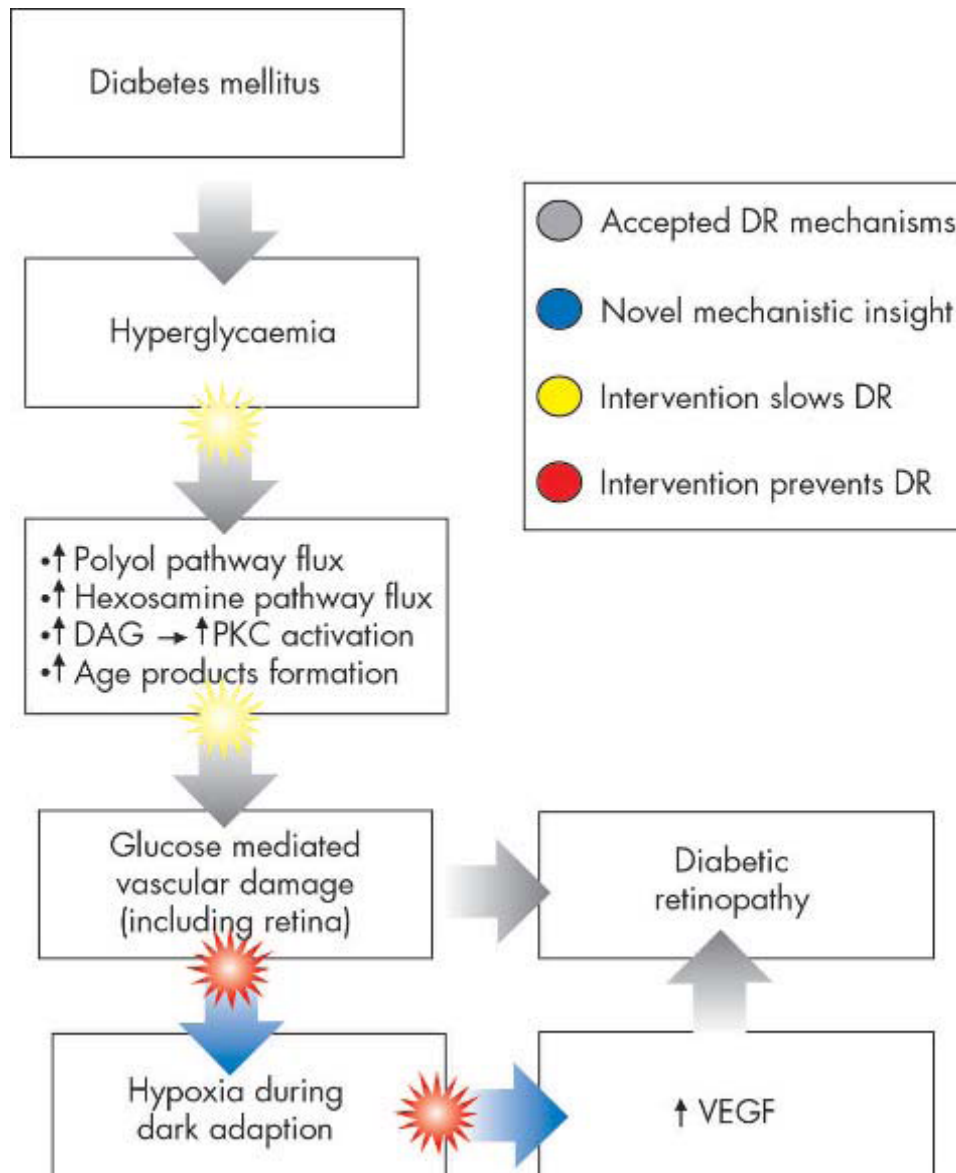


Abbildung 4: Pathomechanismus der diabetischen Retinopathie (DR). Weitere Abkürzungen: DAG: Diacylglycerol, PKC: Proteinkinase C, Age: Endprodukte fortgeschrittener Glycosylierung, VEGF: Vaskulärer endothelialer Wachstumsfaktor. (aus: Arden *et al.*, 2005)

Das Auge reagiert durch Neubildung von Versorgungskapillaren, um die Sauerstoffversorgung zu erhöhen, Das für die Neovaskularisierung wichtigste Protein, das *VEGF* (*Vascular Endothelial Growth Factor*) wird besonders durch *HIF1*, aber auch direkt durch Hyperglykämie hochreguliert. *VEGF* wird in den Müller-Zellen gebildet, die Rezeptoren für VEGF befinden sich gehäuft in Zellen der inneren Retina. *VEGF* schädigt aber auch die Endothelzellen und erzeugt somit die für die Verminderung des Visus hauptsächlich verantwortlichen Mikroläsionen (Boulton *et al.*, 1996). Eine Übersicht dieser Mechanismen findet sich in Abbildung 4.

Einleitung

Eine weitere, möglicherweise durch Hypoxie verursachte oder begünstigte Krankheit ist das Glaucom. Es handelt sich um eine optische Neuropathie, bei der es zum selektiven Untergang der retinalen Ganglionzellen und zu strukturellen Änderungen des Sehnervenkopfes kommt. In einem biphasischen Prozess kommt es zunächst zu einer Schädigung der retinalen Ganglienzellen durch ein mechanisches Trauma, eine degenerative Erkrankung oder auch eine Hypoxie, etwa durch eine Blutgefäßverengung (Kaushik *et al.*, 2003). Sekundär werden umliegende Zellen durch die untergehenden Zellen geschädigt, bis größere Läsionen entstehen (Lynch *et al.*, 1994).

Als einer der molekularen Gründe für den Untergang der retinalen Ganglienzellen wird die Glutamat-induzierte Excitotoxizität diskutiert. Sehr hohe extrazelluläre Glutamatkonzentrationen aktivieren eine Vielzahl von Rezeptoren, z.B. NMDA-Rezeptoren, die für einen massiven Calziumeinstrom in die Zelle verantwortlich sind. Dies führt wiederum zu einer für die Zelle schädlichen Aktivierungskaskade von Nucleasen, Proteasen und Lipasen. Diese schädigen die Zelle soweit, dass der Anteil an freien Radikalen steigt, was über die Aktivierung eines Stickoxid-vermittelten Pfades schließlich zur Apoptose der Zelle führt (Naskar *et al.*, 2001).

Als direkte Folge einer *HIF1*-Induktion werden ebenfalls Stickoxidsynthetasen hochreguliert, welche zur Initiierung von Apoptose führen können (Palmer *et al.*, 1998). Weiterhin führt eine Hypoxie dazu, dass in der Atmungskette Sauerstoff als Elektronenakzeptor fehlt. Deshalb akkumulieren Elektronen in den Mitochondrien und bilden reaktive Sauerstoffradikale, was zu Zellschäden und letztendlich auch zur Apoptose führen kann (Park *et al.*, 1997). Somit kann ein Glaukom direkt und indirekt durch Hypoxie ausgelöst werden.

Auch bei der altersabhängigen Maculadegeneration (AMD) wird eine Hypoxie als ursächlicher Faktor diskutiert, obwohl die genauen Mechanismen der Pathogenese noch unklar sind. Allein in Deutschland leben knapp zwei Millionen Betroffene, bei denen ab dem 50. Lebensjahr eine progressive Verschlechterung des Visus eintritt. Man unterscheidet zwischen einer langsam voranschreitenden „trockenen“ Form, die durch eine O₂-Unterversorgung der Retina im Bereich der *Macula densa* entsteht und der sich rasch verschlimmernden „feuchten“ Form, bei der sich an der Netzhautgrenze neue, oft leckende Gefäßknäuel bilden.

Man beobachtet im alternden Auge den Verlust von Stäbchen, die Verdickung der Bruchschens Membran und eine Verminderung des chorioidalen Blutflusses und somit der

Sauerstoffversorgung, wobei die kausalen Zusammenhänge unklar sind. Durch die so verminderte Bildung des *VEGF* kommt es zu einer Kapillaratrophie, was die Hypoxie verstärkt (Witmer *et al.*, 2003). Zwei Folgen werden diskutiert: Die Hypoxie sorgt für eine verstärkte Bildung von VEGF in der Retina, das aufgrund der schlechteren Permeabilität der Bruch'schen Membran am falschen Ort zu einer Neovaskularisierung führt (Spilsbury *et al.*, 2000). Daneben wird diskutiert, dass der Sauerstoffmangel den Untergang des retinalen Pigmentepithels durch eine Atrophie bedingt (Arden *et al.*, 2005).

Eine häufige Ursache der Hypoxie in der Hornhaut ist das Tragen von Kontaktlinsen. Die Sauerstoffversorgung der Cornea wird bei Kontaktlinsenträgern stark beeinträchtigt, große Anstrengungen werden daher unternommen, die O₂-Permeabilität der Kontaktlinsen zu erhöhen, ohne andere Funktionsparameter zu beeinträchtigen (Fonn *et al.*, 2005).

Die Retina ist besonders sensitiv gegenüber hypoxischen Bedingungen, da die ca. 150 Millionen Stäbchen zu den stoffwechselaktivsten Zellen überhaupt gehören. Hinzu kommt, dass die Kapillaren aus dem *Corpus Vitreus* die innere Retina durchziehen, wodurch die äußeren Segmente nicht optimal versorgt werden. Wenn also die Stäbchen besonders viel Sauerstoff verbrauchen, z.B. während der Dunkeladaption, kann es in der inneren Retina zu stark hypoxischen Bedingungen kommen, besonders wenn weitere einschränkende Begleitumstände dies unterstützen (Arden *et al.*, 2005).

1.4 cDNA-Datenbanken und Data Mining

Bei der Aufklärung der Ursachen erblicher oder erworbener Krankheiten des menschlichen Auges ist die Kenntnis des Transkriptoms der Retina, des Retinoms von zentraler Bedeutung. Die Darstellung des Retinoms gliedert sich in drei große Analyseschritte: die physikalische Erstellung einer entsprechenden cDNA-Bibliothek (1.4.1), die bioinformatische Auswertung der Bibliothek (1.4.2) und abschließend die Metaanalyse, in der eigene und externe Daten verknüpft und abgeglichen werden (1.4.3). Auch hier gewinnt der Zugriff auf bereits bekanntes Wissen und deren Aufarbeitung („*Data Mining*“) mehr und mehr an Bedeutung.

1.4.1 Erstellung einer cDNA-Bibliothek

Die Erstellung und Auswertung einer cDNA-Datenbank ist ein mehrstufiger Prozess. Zunächst wird das zu untersuchende Gewebe isoliert, die mRNA extrahiert, in cDNA umgeschrieben und schließlich kloniert und sequenziert. Zur Vermeidung von Sequenzierfehlern wird die cDNA häufiger und/oder überlappend sequenziert, da besonders Beginn und Ende einer Sequenz fehlerträchtig sind. Dies gilt auch für die Erstellung genomischer Contigs: so wurde z.B. ein typischer BAC-Klon für das Humanen Genomprojekt 6-10fach sequenziert (International humane genome sequencing consortium, 2004).

Einen Meilenstein bei der Identifizierung und Katalogisierung des humanen Transkriptoms stellte die Einführung der EST-Klonierung dar. Ein EST (*Expressed Sequence Tag*) ist ein Sequenzabschnitt eines cDNA-Klons (Adams *et al.*, 1991), also eine in DNA umgeschriebene mRNA, die in ein Plasmid kloniert und anschließend ein- oder beidseitig ansequenziert wird. ESTs bieten einen schnellen Weg zur Identifizierung von Genen, deren Expression und Regulation (Vasmatis *et al.*, 1998), sowie zur Aufklärung zur Diversität transkribierter Sequenzen wie Polymorphismen oder Splicevarianten (Wolfberg und Landsman, 1997).

Das Expressionsniveau eines Genes kann je nach Gewebetyp, physiologischen Zuständen und/oder Lebensalter divergieren. Reassortationskinetik-Analysen haben ergeben, dass die mRNA-Expressioslevel in einer typischen somatischen Zelle sich grob in drei Klassen gliedern lassen: 10-15 mRNAs sind superprävalent und repräsentieren 10-20% der Gesamt-mRNA, 1.000-2.000 mRNAs bilden 40-45% und die restlichen 40-45% werden durch 15.000-20.000 mRNAs ausgemacht (Bishop *et al.*, 1974; Davidson und Britten, 1979), wobei diese Zahlen je nach Spezialisierungsgrad des untersuchten Zelltyps variieren können. Diese Überrepräsentierung bestimmter Gene führt dazu, dass die komplette Sequenzierung einer herkömmlichen cDNA-Bank zu unwirtschaftlich redundanten Analysen führen würde. Um dies zu vermeiden, können prinzipiell zwei Methoden zur Anwendung kommen: Bei der Subtraktion wird die zu untersuchende cDNA gegen einen bekannten cDNA-Kontrollansatz hybridisiert. Identische cDNA-Sequenzen werden gebunden, nicht zur Deckung zu bringende Sequenzen entkommen der Hybridisierung und können so angereichert und analysiert werden. Bei der Normalisierung hingegen hybridisieren in Einzelstränge umgewandelte cDNA-Moleküle mit sich selbst, Redundante Moleküle finden leichter ihren Gegenstrang als seltene Transkripte und werden separiert. Dabei kommt es allerdings nicht zu einer vollständigen

Eliminierung überrepräsentierter Gene, sondern nur zu einem Abflachen der Verteilungskurven (Bonaldo *et al.*, 1996).

Hierbei korreliert die Menge der gebildeten mRNA nur grob mit der tatsächlichen Relevanz eines Gens, da ebenfalls die Halbwertszeit der mRNA, bzw. des prozessierten Polypeptids und dessen Reaktivität (Bindungskoeffizient, Stoffumsatz, etc.) berücksichtigt werden muss. Dennoch bleibt die mRNA-Differenzmenge eines Gens zumindest ein guter Indikator für dessen zelluläre Bedeutung. Die Transkriptionslevel einzelner Gene können über einen Northern-Blot (Sambrook *et al.*, 1989) oder Real-Time-PCR (Lee *et al.*, 1993; Livak *et al.*, 1995) quantifiziert werden. Für die hochparallelisierte Analysen multipler Transkriptionen oder des gesamten Transkriptomkomplements einer Zelle, bzw. eines Gewebes wird auf die SAGE- (Velculescu *et al.*, 1995) oder zunehmend Mikroarray-Technologie (Skena *et al.*, 1995) zurückgegriffen.

Ein weiteres Problem bei der Erstellung einer cDNA-Bank ist, dass ein großer Anteil der generierten Klone nicht die vollständige Transkript-Sequenz umfasst. Typischerweise erhält man einen gewissen Anteil leerer Vektoren, oder Plasmide, die selbstligierte Linkerpaare, bzw. lediglich den Poly-A-Anteil enthalten, verkürzte cDNAs oder Chimären zweier verschiedenen cDNA-Abschnitte, die bei der Ligation fälschlicherweise verknüpft wurden. Um längere oder vollständige cDNA-Sequenzen zu erhalten, werden bei Generierung von cDNA-Banken neben einer initialen Größenfraktionierung der cDNA üblicherweise Variationen des Cap-Trapping durchgeführt (Edery *et al.*, 1995; Carninci *et al.*, 1996), wodurch nicht-degradierte mRNAs als Templates angereichert werden. Einzelne ESTs können durch eine RACE-PCR komplettiert werden (Frohman, 1993). Zu beachten ist weiterhin, dass die Einbringung der cDNA in die Bakterienzellen ein kritischer Schritt ist, da manche Klonierungsstrategien wie die Transfektion kürzere Plasmide bevorzugen (Ohara und Temple, 2001).

Der Durchbruch der DNA-Sequenzierung begann mit der Entwicklung der Didesoxy-Kettenabbruchs-Methode (Sanger *et al.*, 1977). Im selben Jahr wurde mit dieser Methode das erste Gen sequenziert (Seeburg *et al.*, 1977). Eine wichtige Verbesserung stellte die Verwendung von fluoreszenzmarkierten Primern und die Datenauswertung über einen Computer dar, was als Geburtsstunde der modernen Bioinformatik gelten dürfte (Strauss *et*

al., 1986) und zur Entwicklung des ersten automatisierten DNA-Sequenzierers durch Applied Biosystems (Gocayne *et al.*, 1987) führte.

Zur Aufklärung des menschlichen Genoms wurde 1990 das Humane Genomprojekt gestartet. Bereits 1996 waren mehr als die Hälfte der humanen Gene durch systematische EST-Sequenzierprojekte identifiziert (Hillier *et al.*, 1996). Im Februar des Jahres 2001 legten das internationale Humangenom-Sequenzierkonsortium (International humane genome sequencing consortium, 2001) und Celera Genomics (Venter *et al.*, 2001) eine vorläufige, nahezu vollständige Sequenz des humanen Genoms vor, welche 2004 verbessert wurde (International humane genome sequencing consortium, 2004). Hierbei wurde die Anzahl der Lücken von 147.821 auf 341 gesenkt und die durchschnittliche Contig-Länge um den Faktor 475 auf knapp 40 Megabasen gesteigert. Von diesen Lücken bestanden 33 fast vollständig aus Heterochromatin und umfassten knapp 200 Megabasen. Während die geschätzte Zahl der von humanen Genom codierten Gene zu Beginn des Genomprojekts auf etwa 100.000 geschätzt wurde, sank diese Zahl zunächst auf 30-40.000 (Venter *et al.*, 2001) und wurde 2004 mit 22.287 Genloci angegeben, worunter 19.438 bekannte und 2.188 prädizierte Gene fielen. Es wurden verschiedene 34.214 Transkripte ermittelt, also durchschnittlich 1,54 Transkripte pro Locus. Die Anzahl der Exone wurde mit 231.667 angegeben, was 10,4 Exone per Locus und 9,1 pro Transkript bedeutete. Die Gesamtlänge aller codierenden Exone summiert sich danach auf 34 Megabasen, was 1,2% des Euchromatins entspricht. Die nicht translatierten Bereiche der Transkripte wurden auf 21 Megabasen, bzw. auf 0,7% des Euchromatins geschätzt (International humane genome sequencing consortium, 2004).

1.4.2 Bioinformatische Auswertung einer cDNA-Bank

Nach Erstellen und Sequenzieren der cDNA-Bank ist im Auswertungsschritt ein *in silico*-Clustering der cDNA-Sequenzen vorteilhaft. Hier werden die gewonnenen Sequenzen miteinander verglichen und bei hinreichender Ähnlichkeit virtuell verbunden. Dies ermöglicht zum einen die Eliminierung von Redundanzen zur effizienteren Weiterverarbeitung der Daten, zum anderen können überlappende Sequenzen zu einem längeren und somit informativeren Abschnitt verschmolzen werden. Zu beachten ist, dass das Clustern immer ein Balanceakt zwischen dem Vereinen ähnlicher, aber nicht zusammengehöriger (falsch positiver) Sequenzen und dem Nichtvereinen zusammengehöriger (falsch negativer)

Sequenzen ist. Bekannte Clustering-Programme sind u.a. die TGI Clustering tools (The Institute for Genomic Research, 2003), ClustalW (Chenna *et al.*, 2003) und STACK (Christoffels *et al.*, 2001).

Ein weiteres, wichtiges Anwendungsgebiet der Bioinformatik ist die Identifizierung möglicher Gene auf der Grundlage eines sequenzierten Genomabschnitts durch Genprädiktionsprogramme. Dabei werden zwei grundsätzliche Methoden unterschieden: ein Ansatz ist die Suche nach Sequenzähnlichkeiten zu bereits bekannten Genen (z.B. in anderen Species), der andere (*ab initio*) sucht nach kurzen Sequenzmotiven wie Startcodons oder Splice Stellen, bzw. gleicht Unterschiede im Codon-Usage zwischen codierenden und nicht-codierenden Abschnitten ab (Wang *et al.*, 2004). Die Programme, welche die erste Methode nutzen sind neben den BLAST-Anwendungen z.B. PROCRUSTES (Gelfand *et al.*, 1996) und GeneWise (Birney und Durbin, 2000), für die zweite Methode stehen die Programme GENSCAN (Burge und Karlin, 1997) und AUGUSTUS (Stanke und Waack, 2003) als modernste Verfahren zur Verfügung. Geschätzte 3.000 Transkriptionsfaktoren regulieren die Expression im menschlichen Genom. Zur Aufklärung der jeweiligen Bindungsstellen kommen spezialisierte Varianten der *ab initio*-Suchprogramme wie MOST (motif searching web tool) (Pizzi *et al.*, 2005) oder YMF (*yeast motif finder*) (Sinha und Tompa, 2003) zum Einsatz.

Die Abspeicherung der gewonnenen Daten geschieht meist in speziell angelegten Datenbanken. Die Vorteile der Verwendung von Datenbanken liegen sowohl in ihrer Selektierbarkeit, als auch in ihrer Flexibilität. Während einerseits nur der für die relevante Fragestellung interessante Genomausschnitt abgespeichert und bearbeitet wird, was teure Systemressourcen und Rechenzeiten spart, können andererseits Datensätze leicht gelistet, fusioniert und/oder selektiert werden. Besonders interessant sind so genannte verknüpfte Abfragen, bei denen zwei oder mehrere Abfragen durch Boolesche Operatoren verbunden werden. Datenbanken, bei denen Verknüpfungen möglich sind, nennt man relationale Datenbanken (z.B. SQL-Varianten oder ORACLE), während klassische Varianten hierarchisch strukturiert sind.

Datenbanken spielen bei genomischen Forschungsprojekten eine essentielle Rolle, man unterscheidet mehrere Funktionsebenen:

Organspezifische Datenbanken sind u.a. beschrieben für die Leber (Yano *et al.*, 2001) und das Herz (Megy *et al.*, 2002). Organismenspezifische Datenbanken existieren z.B. für den Menschen (Humanes Genomprojekt), für die Maus (Okazaki *et al.*, 2002) oder den Ackerschachtelhalm (Seki *et al.*, 2004). Funktionelle oder Fragestellungs-spezifische Datenbanken sind z.B.: HAGR (Human Ageing Genomic Resources) (de Magalhaes *et al.*, 2005) für genetische Seneszenzstudien, JCSGCS (Joint Center for Structural Genomics Core Screen) (Page und Stevens, 2004) zur Findung optimaler Proteinkristallisierungsstrategien oder PlantMarkers (Rudd *et al.*, 2005) zur Analyse von QTLs (*Quantitative Trait Loci*) bei Pflanzen.

1.4.3 Metaanalyse

So genannte Metatools extrahieren Informationen nach definierten Abfrageparametern aus unterschiedlichen *in silico*-Datenbanken und verknüpfen diese zu einem neuen Ergebnissatz. Von der Methodik erinnert dieses Vorgehen an verknüpfte Abfragen mit Booleschen Operatoren aus relationalen Datenbanken, wobei jedoch die Datenextraktion an die Syntax der jeweiligen Quelldatenbanken angepasst werden muss (Köhler, 2004). Speziell die Kompatibilität und eindeutige Zuordnung der jeweiligen Datensätze aus historisch gewachsenen Datenbanken ist ein großes Problem (Mons, 2005).

Es existieren eine Vielzahl dieser Metatools: Das Programm CLOE dient der vergleichenden Analyse von Ergebnissen aus Mikroarray-Experimenten zwischen verschiedenen Species (Pellegrino *et al.*, 2004), H-ANGEL stellt integrative Expressionsprofile aus verschiedenen Quellen dar (Tanino, 2005), ONDEX extrahiert und vergleicht graphen-basierte Datensätze (z.B. Enzymaktivitäten oder Proteininteraktionen) (Köhler *et al.*, 2005), wohingegen das SRS-Integrationsmodul diverse Datenbanken des EMBL zusammenfasst (Zdobnov *et al.*, 2002).

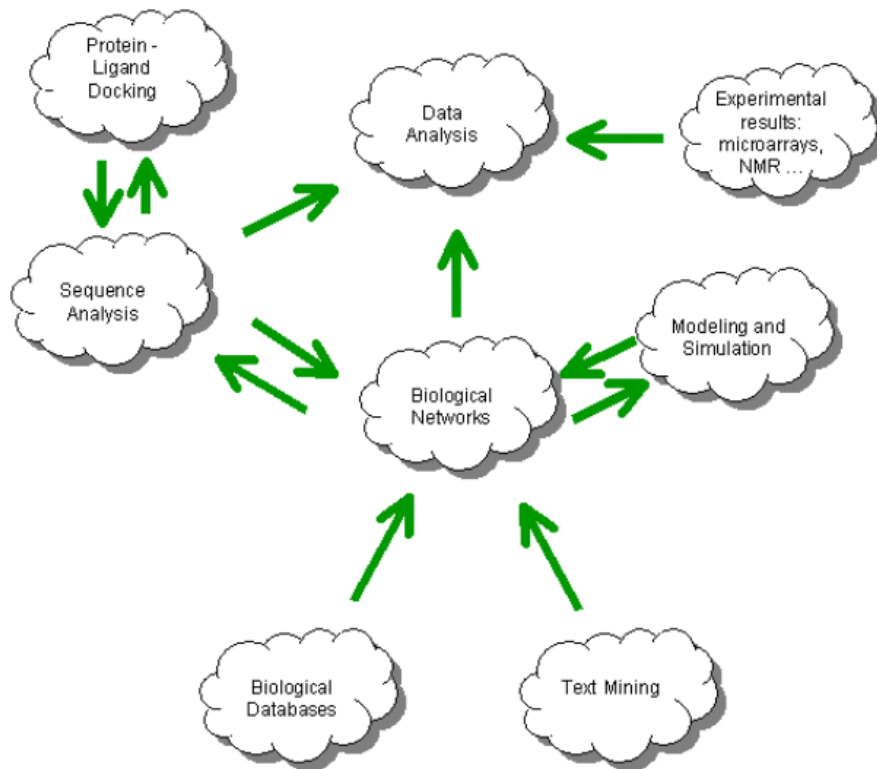


Abbildung 5: Übersicht über mögliche Datenflüsse innerhalb eines Data Mining-Programmes.
(aus: Köhler *et al.*, 2005)

Die transkriptionelle Komplexität der Retina mit ca. 50 unterschiedlichen neuronalen und glialen Zelltypen machte eine hochparallelisierte, von bioinformatischen Methoden unterstützte Analyse unumgänglich (Swaroop und Zack, 2002). Für Mikroarray-basierte Untersuchungen wurde ein Set von 10.034 cDNA-Fragmenten definiert, die Gene repräsentieren, welche an der Entwicklung, Funktionserhaltung, bzw. Pathogenese der Netzhaut beteiligt sind und die in der Retina exprimiert werden (Chowers *et al.*, 2003a). Analog dazu wurde ein Expressionsprofil der humanen Retina mittels SAGE-Technologie erstellt, in dem 228.192 Tags und 26.335 unterschiedliche Transkripte gelistet wurden (Sharon *et al.*, 2002). In einem rein bioinformatischen Ansatz verglichen Katsanis und Mitarbeiter 40.000 humane retinalen ESTs mit dem Gesamtset von 1.400.000 humanen ESTs und konnten 925 ESTs als spezifisch oder bevorzugt in der Retina exprimiert beschreiben. Für die bis dato 51 unklonierten Retinopathien konnten 42 putative Kandidatengene beschrieben werden, (Katsanis *et al.*, 2002). Eine Übersicht über lokalisierte, bzw. klonierte retinale Krankheitsgene findet sich in Abbildung 6. In einem anderen Ansatz wurden 481.137 in der Retina oder dem retinalen Pigmentepithel exprimierte Primärdatensätze aus 27 Studien zusammengefasst und bioinformatisch bearbeitet. Nach dem Clustern wurden noch 15.645 Transkript-Datensätze erhalten, die über einen LocusLink-Eintrag verfügten. Von diesen wurden 2.608 Transkripte nur je einmal in den berücksichtigten Studien identifiziert

(Singletons) und wurden aus Sicherheitsgründen nicht weiter untersucht. Die übrigen, durch mehrere Sequenzen repräsentierten 13.037 cDNA-Cluster lieferten 1.411 ausschließlich in der Retina und 246 lediglich im retinalen Pigmentepithel exprimierte Transkripte (Schulz *et al.*, 2004). Für eine vergleichende Untersuchung der Expressionsniveaus in verschiedenen Gewebetypen des Auges wurde das Set an cDNA-Fragmenten für eine Mikroarray-Analyse auf ca. 30.000 ausgedehnt. Neben vertieften Einsichten über Genregulationsprozesse im menschlichen Auge erbrachte diese Studie eine Anzahl von 3.727 spezifisch in der Netzhaut exprimierten Genen (Diehn *et al.*, 2005).

Es existiert zwischenzeitlich eine Reihe von Datenbanken, die dem Transkriptom der menschlichen Retina gewidmet sind, z.B. NEIBank (Wistow, 2002; Wistow, 2005), Retinome (Schulz *et al.*, 2004) oder Retbase (Katsanis *et al.*, 2002). Die RetNet-Datenbank (Daiger *et al.*, 1998) liefert eine Übersicht über die bislang kartierten Gene/Genorte für erbliche Netzhauterkrankungen.

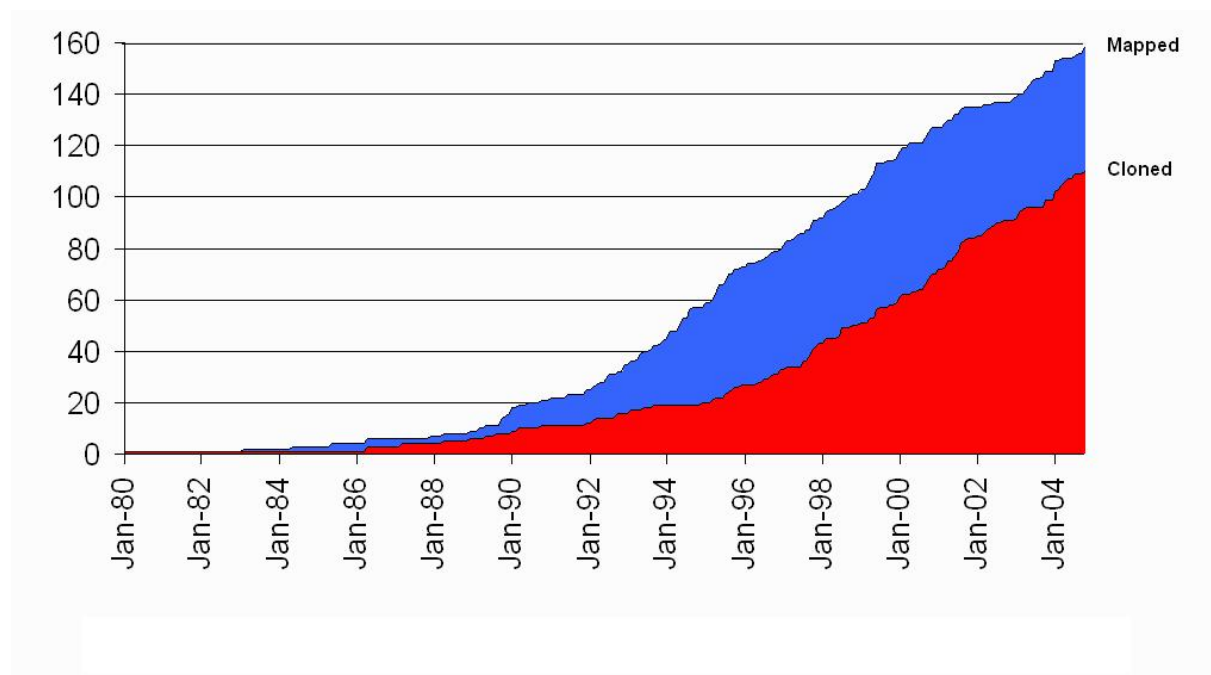


Abbildung 6: Anzahl lokalisierter und klonierter retinaler Krankheitsgene des Menschen. (aus: <http://www.sph.uth.tmc.edu/Retnet/sum-dis.htm#A-genes>)

2. Material

2.1 Verbrauchsmaterialien und Reaktionskits

Applied Biosystems, Foster City, USA	ABI PRISM BigDye Terminator v. 3.1 Cycle Sequencing Ready Reaction Kit
	ABI PRISM BigDye Terminator v. 3.2 Cycle Sequencing Ready Reaction Kit
	SYBR Green RT-PCR MasterMix
Biorad Laboratories, Richmond, USA	Bromphenolblau
	Ficoll 400
	Xylencyanol
Difco, Detroit, USA	Bacto-Agar
	Bacto-Trypton
Eppendorf, Hamburg, D	Perfect Prep Plasmid 96-Kit
Fluka Chemika / Fluka Chemie AG, Buchs, H	Borsäure
	Chloroform
	Isoamylalkohol
	β -Mercaptoethanol
	Natriumchlorid
	Natriumhydroxid
FMC Bioproducts, Rockland, USA	Low Melting NuSieve GTG Agarose
	SeaKam GTG Agarose
ICN Biomedical Inc., Aurora, USA	Hefeextrakt
	5-bromo-4-chloro-3-indolyl-beta-D-Galactopyranosid (X-Gal)
Invitrogene, Carlsbad, USA	Agarose Electrophoresis Grade, Ultra Pure
	Concert 96 Plasmid Purification-Kit
	Ethylendiamintetraacetat (EDTA)
	Fötales Kälberserum
	Glycerol
	L-Glutamin
	Phosphat-gepufferte Saline (PBS)
	RPMI 1640 Zellkulturmedium mit L-Glutamin
	SuperScript Plasmid System for cDNA

Material

	Synthesis and Plasmid Cloning
	Trizol
Macherey & Nagel, Düren, D	NucleoSpin Robot-96 Plasmid Kit
Merck, Darmstadt, D	Essigsäure
Nunclon, Roskilde, DK	30 ml Zellkulturflaschen
	200 ml Zellkulturflaschen
Roth & Co., Karlsruhe, D	Ampicillin
	Dimethylsulphoxid (DMSO)
	Roti-Phenol
Qiagen, Hilden, GER	Oligotex mRNA-Purification Kit
	Qiaprep 96 Turbo BioRobot Kit
	R.E.A.L Prep96 Kit
	RNeasy Mini Kit
Serva Feinbiochemica GmbH & Co KG, Heidelberg, D	Ethidiumbromid
Sigma-Aldrich Vertriebsgesellschaft GmbH, Deisenhofen, D	DEPC
	Tris-Base
	Tris pH 8,0
USB Amersham, Cleveland, Ohio, USA	dATP
	dCTP
	dGTP
	dTTP
	Gelatine

2.2 Enzyme

Eppendorf, Hamburg, GER	Taq DNA-Polymerase
Qiagen, Hilden, GER	Taq DNA-Polymerase
New England Biolabs, Beverly, USA	Restriktionsenzyme <i>Not I</i> , <i>Sal I</i>

Material

2.3 Antibiotika

Invitrogene, Carlsbad, USA
Roth & Co., Karlsruhe, D

Streptomycin
Ampicillin

2.4 Längenstandards

New England Biolabs, Beverly, USA

Eigenherstellung

1kb-Leiter
 λ DNA/Hind III
Jens Eigenbau 2 („JEB2“-Marker
(pBR322/*AluI*))

2.5 Elektrophorese-Puffer

10 x Laufpuffer für Sequenzierungen

60 g Tris-Base
30,5 g Borsäure
40 ml 0,25 M EDTA pH 8,0
ad 1000 ml H₂O_{dd}

50 x TAE-Puffer

242 g Tris-Base
57,1 ml Essigsäure
200 ml 0,25 M EDTA pH 8,0
ad 1000 ml H₂O_{dd}

10 x TBE-Puffer

108 g Tris-Base
55 g Borsäure
8 ml 0,25 M EDTA pH 8,0
ad 1000 ml H₂O_{dd}

2.6 Ladepuffer

10 x DNA-Ladepuffer für Agarosegele

20% Ficoll 400
100 mM EDTA pH 8,0
0,2% Bromphenolblau
0,2% Xylencyanol

RNA-Ladepuffer für Agarosegele

5 μ l 1 M Tris pH 8
5 μ l 0,2 M EDTA pH 9
5 μ l 5 M NaCl
125 μ l Formamid
200 μ l H₂O DPEC-behandelt

Material

50 µl 10 x Agarose Ladepuffer
50 µl Ethidiumbromid (10 µg/ml)

2.7 Kompetente Zellen

Invitrogene, Carlsbad, USA

ElektroMAX-Zellen
(*Escherichia coli*: DH10B):
F⁻ *mcrA* Δ(*mrr-hsdRMS-mcrBC*)
φ80*dlacZ*ΔM15 Δ*lacX74 deoR recA1 endA1*
araD139 Δ(*ara, leu*)7697 *galU galK λ⁻ rpsL*
nupG

2.8 Geräte

AEG, Nürnberg, D

Mikrowelle

Applied Biosystems, Foster City, USA

ABI Prism 377 DNA-Sequenzier

ABI Taqman 7000

GeneAmp PCR System 2400

GeneAmp PCR System 9600

Bachhofer, Reutlingen, D

Heizblock

Speed Vac

Transilluminator

Beckmann, München, D

L80 Ultrazentrifuge

Biometra, Göttingen, D

Absaugpumpe

Biorad, Richmond, USA

Agarose-Gelelektrophoresekammern

Eppendorf, Hamburg, D

Splid-lid Quad Vacuum Manifold

PERFECT prep-96 VAC-Kit

Dr. Koch Computertechnik AG, Tübingen, D

Pentium 4-Computer

2 x 2,4 GHz Pentium 4-Prozessoren

4 x 1 Gbyte RAM

180 Gbyte Western Digital Caviar Festplatte

Heraeus, Osterode, D

Biofuge 13,

Biofuge 15,

Brutschrank

Herolab, Wiesloch, D

Geldokumentationssystem (E.A.S.Y)

IKAMAG, Conroe USA

Magnetrührer

Material

Mettler, Giessen, D	Feinwaage PJ 400
Milipore, Georgia, USA	Mili-Q UF Plus
Scientific Industries, Florenz, IT	Vortex Genie 2
Sigma, Osterode, D	4K10 Zentrifuge

2.9 Medien

Alle Nährmedien wurden für 20 min bei 121°C und 2 bar Druck autoklaviert. Antibiotika, X-Gal und IPTG wurden ggf. nach Abkühlen der Nährmedien aus sterilen Stammlösungen zugesetzt.

LB-Medium	10 g Bacto-Trypton 5 g Hefeextrakt 5 g NaCl ad 1 l H ₂ O _{dd}
LB-Agar	10 g Bacto-Trypton 5 g Hefeextrakt 5 g NaCl 15 g Agar ad 1 l H ₂ O _{dd}
2 x YT-Medium	16 g Bacto-Trypton 10 g Hefeextrakt 5 g NaCl ad 1 l H ₂ O _{dd}
LB-Superbroth	32 g Bacto-Trypton 20 g Hefeextrakt 5 g NaCl 5 ml 1N NaOH ad 1000 ml H ₂ O _{dd}
SOC-Medium	2% Bacto-Trypton 0,5% Hefeextrakt 10 mM NaCl 2,5 mM KCl

Material

2.10 PCR-Puffer

dNTP-Lösung	5 mM dATP 5 mM dCTP 5 mM dGTP 5 mM dTTP 5 mM Tris pH 7,0
10 x Ampli I-Puffer	100 mM Tris pH 8,3 500 mM KCl 15 mM MgCl ₂ 0,01% Gelatine
10 x Ampli II-Puffer	100 mM Tris pH 8,6 500 mM KCl 15 mM MgCl ₂ 0,01% Gelatine
10 x Ampli III-Puffer	100 mM Tris pH 8,9 500 mM KCl 15 mM MgCl ₂ 0,01% Gelatine
10 x BOM9	100 mM Tris, pH 8,9 250 mM KCl 30 mM MgCl ₂
10 x BOM12	100 mM Tris, pH 8,3 500 mM KCl 30 mM MgCl ₂
10 x BOM15	100 mM Tris, pH 8,6 500 mM KCl 30 mM MgCl ₂
10 x BOM16	100 mM Tris, pH 8,9 500 mM KCl 10 mM MgCl ₂
10 x BOM17	100 mM Tris, pH 8,9 500 mM KCl 20 mM MgCl ₂
10 x BOM18	100 mM Tris, pH 8,9 500 mM KCl 30 mM MgCl ₂
10 x BOM27	100 mM Tris, pH 8,9 750 mM KCl 30 mM MgCl ₂

2.11 Primerliste

Alle Primer für die RT-PCR (3.2.6) wurden bei Eurogentec synthetisiert, die Sequenzierprimer stammen von GIBCO BRL und wurden für Arbeitslösungen auf eine Konzentration von 5 pmol/μl in H₂O_{dd} eingestellt.

Verwendungszweck	Name	Sequenz (5` - 3`)
RT-PCR – GAPDH forward	RTGAPDHF	TTC GTC ATG GGT GTG AAC C
RT-PCR – GAPDH reverse	RTGAPDHR	GCA TGG ACT GTG GTC ATG AG
RR-PCR – 18S forward	RT18SF	CGG CTA CCA CAT CCA AGG AA
RT-PCR – 18S reverse	RT18SR	GCT GGA ATT ACC GCG GCT
RT-PCR – VEGF forward	RTVEGFF	TCG CTT ACT CTC ACC TGC TTC TG
RT-PCR – VEGF reverse	RTVEGFR	CCG GGA CAT CTG CCA GTG
RT-PCR – EPO forward	RTEPOF2	CTG TCG CAT GGA CAC TCC AGT
RT-PCR – EPO reverse	RTEPOR2	AGT TCC TCT GGC CCC TGA GA
cDNA-Sequenzierung forward	M13 forward	GTA AAA CGA CGG CCA GT
cDNA-Sequenzierung reverse	T7 extended	TAA TAC GAC TCA CTA TAG GG

2.12 Verwendete Computerprogramme

Systat Software GmbH, Erkrath, D	Sigmaplot
DNASTAR Inc., Madison, USA	Lasergene 5
Applied Biosystems, Foster City, USA	Primer Express v 2.0
	Sequence Detection System ABI Prism 7700

3 Methoden

3.1 Molekularbiologische Methoden zur Erstellung einer cDNA-Bank

3.1.1 Zellkultur WERI

1 ml in flüssigem Stickstoff gelagerte WERI-Zellen wurden auf Eis langsam aufgetaut. Die Zellen wurden zweimal mit 10 ml vorgewärmtem Kulturmedium (RPMI 1640 + 10% FCS) gewaschen und 10 Minuten lang bei 250 g abzentrifugiert. Das Pellet wurde resuspendiert und in eine 30 ml Nunclon-Zellkulturflasche mit 10 ml Kulturmedium passagiert, wo die Zellen bei 37°C und 5% CO₂-Partialdruck inkubiert wurden. Verfärbte sich der Farbindikator von rosa ins gelbliche, wurde das Medium wie oben beschrieben gewechselt. Wuchsen die Zellen zu dicht, wurde ein entsprechender Anteil der WERI-Zellen bei der nächsten Passage verworfen. Da die Zellen bei jeder Zellteilung mutieren können, sollte die Kultur nicht länger als 50 Passagen benutzt werden.

3.1.2 Kryokonservierung von WERI-Zellen

Aliquots von ca. 1×10^7 Zellen wurden zweimal mit PBS gewaschen und in 1 ml 90% FCS/10% DMSO resuspendiert. Die Aliquots wurden in Einfrierröhrchen langsam auf -70°C abgekühlt und nach 1-2 Tagen in flüssigen Stickstoff überführt.

3.1.3 Auftauen von WERI-Zellen

Eingefrorene Zellen wurden im Wasserbad bei 37°C aufgetaut, in 10 ml Kulturmedium aufgenommen und zweimal mit vorgewärmtem PBS gewaschen. Die Abzentrifugation erfolgte bei 250 g für 10 Minuten.

3.1.4 Aufschluss mit Trizol

Nach Abzentrifugation bei 500 g und Waschen mit PBS wurde das WERI-Rb1-Zellpellet (ca. 10^7 Zellen) mit 1 ml Trizol-Reagenz versetzt und mit einer Pipette gut durchgemischt. Bei Trizol handelt es sich um eine monophasische Lösung aus saurem Phenol und Guanidiniumisothiocyanat. Bei der Trizol-Behandlung lysieren die Zellen und die Proteine einschließlich endogener RNAsen werden denaturiert.

Die Suspension wurde 5 min bei Raumtemperatur stehen gelassen, in ein Eppendorf-Cup überführt und mit 200 μ l Chloroform versetzt. Das Gemisch wurde 15 Sekunden kräftig geschüttelt und wieder zwei Minuten bei Raumtemperatur stehen gelassen. Danach wurde der Ansatz bei 12.000 rpm in einer Tischzentrifuge für 15 Minuten bei 4°C zentrifugiert. Die obere wässrige Phase wurde abgenommen, mit 200 μ l Chloroform gemischt sowie erneut geschüttelt und abzentrifugiert. Die obere Phase wurde in ein neues Reaktionsgefäß überführt, mit 500 μ l Isopropanol gut vermischt und 10 Minuten bei Raumtemperatur stehen gelassen. Nach einer weiteren Zentrifugation bei 12.000 rpm bei 4°C für 30 min wurde der Überstand vorsichtig abgezogen, das Pellet wurde mit 1 ml kaltem 80%igem Ethanol gewaschen und bei 4°C für 5 Minuten bei 7.500 rpm zentrifugiert. Der Überstand wurde dekantiert, das Pellet kurz getrocknet und in 50 μ l DEPC-Wasser aufgenommen.

3.1.5 CsCl-Dichtegradientenzentrifugation

Für die Generierung von cDNA-Bibliotheken ist es sehr wichtig, Reste genomischer DNA aus RNA-Präparationen zu entfernen, da sonst leicht ein Priming von genomischen DNA-Templates und folglich ein Einklonieren von DNA-Fragmenten geschehen könnte. Eine besonders effiziente Trennung von RNA und DNA kann durch eine Cäsiumchlorid-Dichtegradientenzentrifugation erzielt werden. Hierbei wird eine CsCl-Lösung mit hoher Dichte starken Zentrifugalkräften ausgesetzt, so dass sich ein Dichtegradient ausbildet: In Richtung Rotornabe herrscht eine geringere Beschleunigung, die Lösung ist hier weniger dicht als weiter außen. Dispergierte Partikel reichern sich im Gradienten an derjenigen Position an, die ihrer eigenen Dichte im Lösemittel entspricht. Die Genauigkeit der Trennung einzelner Komponenten erreicht die vierte Nachkommastelle der Dichte.

Zunächst wurde das RNA-Präzipitat in ein Opti-Seal-Röhrchen pipettiert, welches mit einer sterilen CsCl-Lösung (1,55 g/ml) aufgefüllt und auf 0,01 g genau austariert wurde. Die CsCl-

Lösung enthielt 0,1 % DEPC als RNase-Inhibitor. Das Röhrchen wurde mit einem Stopfen verschlossen, wobei darauf zu achten war, dass keine Flüssigkeit oberhalb des O-Ringes verblieb. Die Zentrifugation erfolgte in einem NVT-90-Rotor bei 20°C für 16 h bei 65.000 rpm, wodurch die Gesamt-RNA am Boden des Röhrchens pelletierte (Glisin *et al.*, 1974). Nach Beendigung des Laufes wurde der Überstand vorsichtig mit einer Pasterupipette entfernt. Das Pellet wurde in 200 µl DEPC-behandeltem H₂O_{dd} mit 0,1 mM EDTA (pH 7,5) aufgelöst.

3.1.6 Aufreinigung von Poly-A mRNA durch Oligotex-Säulen

Bei diesem Schritt wird aus der zuvor gewonnenen Gesamt-RNA die polyadenylierte mRNA mittels Affinitätschromatographie isoliert.

Das Oligotex-Verfahren nutzt die Tatsache, dass die meiste eukaryontische mRNA am 3`-Ende ein Homopolymer aus 20-250 Adenosin-Nukleotiden besitzt (Poly A-tail), während strukturelle RNAs nicht polyadenyliert sind.

Die Trennung der mRNA von anderen Nucleotiden erfolgt durch Hybridisierung des Poly A-tails an Oligo-dT-Primer, welche an eine Festphasen-Matrix gebunden sind. Da hohe Salzkonzentrationen zur Hybridisierung notwendig sind, kann die gebundene mRNA nach Auswaschung der nicht gebundenen Moleküle durch einen Puffer mit geringerer Ionenstärke eluiert werden (Kuribayashi-Otha *et al.*, 1993).

Die Durchführung erfolgte nach Angaben des Herstellers (Qiagen).

3.1.7 Erstellung einer retinalen cDNA-Bibliothek durch das Superscript Plasmid System

Die cDNA-Synthese und Klonierung erfolgte mit dem SuperScript Plasmid System for cDNA Synthesis and Plasmid Cloning-Kit der Firma Qiagen.

In einem ersten Schritt wurde die aus den WERI-Zellen isolierte und aufgereinigte mRNA (3.1.5) mit einem *NotI*-primer-Adapter inkubiert. Dieser Primer-Adapter hybridisiert über seinen Poly-T-Schwanz an den 3`-Poly A-Schwanz der mRNA und liefert gleichzeitig eine *NotI*-Schnittstelle für die nachfolgende Klonierung in den pSport2-Vektor. Die Erststrangsynthese erfolgte mit dem Enzym SuperScript II RT, einer modifizierten MMLV- (*Moloney Murine Leukaemia Virus*) Reverse Transkriptase. Die Zweitstrangsynthese erfolgte

nach dem Prinzip des „*nick translational replacement*“, wobei Strangbrüche in der mRNA erzeugt wurden, die Ansatzpunkte für die Synthese des Zweitstranges durch *E.coli* Polymerase I liefern. Die Bruchstellen zwischen den neu synthetisierten Einzelsträngen wurden durch eine Ligase geschlossen. Durch Zugabe von T4 DNA-Polymerase wurden eventuell überstehende Nucleotide für die anschließende blunt-end-Ligation entfernt. An die Enden der cDNA wurden doppelsträngige Adapter ligiert, die *Sall*-Schnittstellen enthalten. Durch Doppelverdau mit *Sall* und *NotI* erhält man cDNA-Moleküle mit den entsprechenden Überhängen am 5`- und 3`-Ende. Um möglichst lange Fragmente zu erhalten, wurde die cDNA-Schar einer chromatographischen Größenfraktionierung mittels einer Cellulose-Säule unterzogen. Bei der Elution wurden getrennte Größenfraktionen aufgefangen und die Fraktion der größten cDNA mit dem pSport2-Vektor ligiert. 1 µl Aliquots des Ligationsansatzes wurden entnommen und in elektrokompente ELECTROMax DH10B-Zellen transformiert (3.1.8). 1:100-Verdünnungen des Transformationsansatzes wurden auf IPTG/XGal/Ampicillin-Platten mit dem Drigalski-Spatel ausgestrichen und über Nacht bei 37°C angezogen (Sambrook *et al.*, 1989). Die Durchführung geschah nach Angaben des Herstellers.

Abweichungen vom Protokoll des SuperScript Plasmid System-Kits:

- Alle Waschschrte der präzipitierten cDNA wurden mit 80%igem statt mit 70%igem Ethanol durchgeführt, um die mögliche Ausbeute zu erhöhen.
- Bei der Transformation der *E.coli*-Zellen wurde die Zeit der phänotypischen Expression vor dem Ausstreichen auf die Agarplatte von 60 auf 30 Minuten gesenkt, damit durch eine eventuelle Teilung der transformierten Zellen während dieser Zeit nicht unnötig identische Klone analysiert würden.

3.1.8 Transformation durch Elektroporation

Elektrokompente *E.coli* DH10B-Zellen (20 µl) wurden auf Eis aufgetaut, mit 1 µl Ligationsansatz vermischt und in vorgekühlte sterile Elektroporationsküvetten mit einem Elektrodenabstand von 2 mm pipettiert. Die Elektro-Transformation mit dem Bio-Rad Gene Pulser erfolgt bei einer Spannung von 1,8 kV, 200 Ω Widerstand und 25 µF Kondensatorkapazität. Anschließend wurden die transformierten Bakterien mit 1 ml vorgewärmtem SOC-Medium (37 °C) ausgespült und für 45 min im Bakterienschüttler zur

phänotypischen Expression inkubiert (37 °C, 225 rpm). Der Ansatz wurde 1:100 mit vorgewärmtem LB-Medium verdünnt, von dieser Verdünnung wurden je 200 µl auf Ampicillin- und IPTG/X-Gal-haltigen LB-Agarplatten ausplattiert und über Nacht im Brutschrank bei 37 °C inkubiert. Die gewachsenen Bakterienkolonien zeigten eine blaue, bzw. weiße Färbung.

Aufgrund der Expression des *lacZ*-Genes im Vektor sind Kolonien ohne Insert blau gefärbt, die weißen Zellen hingegen sollten ein Insert tragen.

3.1.9 Bakterienanzucht

Eine gepickte Kolonie (aus 3.1.8) wurde mit einem sterilen Zahnstocher in 1,3 ml LB-Medium mit 150 µg/ml Ampicillin überführt und wurde über Nacht bei 37°C auf dem Schüttelinkubator (225 rpm) in einem 96er Wachstumsblock kultiviert. Diese Kultur diente als Vorkultur, um ein gleichmäßiges Wachstum aller Klone zu gewährleisten, da sich die Zellen zum Zeitpunkt der Ernte in der Plateauphase befinden sollten. Danach wurden 75 µl für eine Glycerinkultur abgenommen (3.1.10). Mit der Multipipette wurden je 10 µl der Kultur aufgezogen und in 1,3 ml 2xYT-Medium mit 150 µg/ml Ampicillin als Starterkultur überführt. Diese Kulturen wuchsen 20 h bei 37 °C auf dem Schüttelinkubator in QIAGEN 96er Wachstumsblöcken, welche mit einem luftdurchlässigen AirPore Tape Sheet überzogen worden waren.

3.1.10 Glycerinkultur

Zur dauerhaften Lagerung der rekombinanten Bakterien wurden 75µl einer Kultur abgezogen und mit 75 µl einer sterilen 50% Glycerin-Lösung versetzt. Diese Mischung wurde in 96er Multititerplatten bei -80°C eingefroren.

3.1.11 Präparation bakterieller Plasmid-DNA durch R.E.A.L. Prep 96

Die Plasmid-DNA der transformierten Klone (3.1.8) wurde nach der Anzucht der Bakterienkulturen (3.1.9) mittels der Methode der alkalischen Lyse (Birnboim und Doly,

1979; Birnboim 1983) isoliert. Dazu wurden die Bakterien in den 96er Wachstumsblöcken abzentrifugiert (Sigma 4K10 Zentrifuge, 1.500 g, 5 min, 4°C), das Zellpellet resuspendiert und die Zellwand danach unter Zugabe von Natronlauge und SDS lysiert. Das Lysat wurde danach durch Zugabe hochmolarer Salzlösungen neutralisiert. Dabei präzipitieren Proteine und andere Zellbestandteile einschließlich der bakteriellen DNA, während die Plasmid-DNA in Lösung bleibt. Das Lysat wurde durch einen Filter gesogen, der die unlöslichen Bestandteile abtrennte. Anschließend wurde das Lysat mit Isopropanol versetzt und zentrifugiert (Sigma 4K10 Zentrifuge, 2.500 g, 15 min, 20°C) und mit 0,5 ml 70% Ethanol gewaschen. Die pelletierte Plasmid-DNA wurde in 50 µl EB-Puffer (10 mM Tris, pH 8,0) aufgenommen. Die Präparation wurde entsprechend den Angaben des Herstellers mit dem R.E.A.L. Prep 96-Kit im 96er Maßstab durchgeführt.

Abweichungen:

- Die optionalen Schritte 6 und 7 (Erhitzen und Abkühlen des Lysates) des R.E.A.L.-Präparationsprotokolls wurden standardmäßig durchgeführt.

3.1.12 Bestimmung von Insertgrößen

Zur Bestimmung der Insertgrößen der klonierten cDNA wurden die Plasmide mit den Restriktionsenzymen verdaut, die auch zur Klonierung der cDNA in den Vektor verwendet wurden (*drop out*). Die verdaute Plasmid-DNA wurde elektrophoretisch aufgetrennt (3.1.13) und die Insertgröße im Vergleich zu bekannten Längenstandarts abgeschätzt.

Zur Größenbestimmung der in pSport2 klonierten cDNA wurde die Plasmid-DNA mit *NotI* und *Sall* verdaut.

Restriktionsansatz:

0,2 µg DNA

5 U *NotI*

5 U *Sall*

2 µl React 3-Puffer

X µl H₂O_{dd}

20 µl Gesamtvolumen

3.1.13 Elektrophoretische DNA-Auftrennung in Agarosegelen

Zur Bestimmung der Reinheit, von Fragmentgrößen und zur Abschätzung der Menge an DNA, bzw. RNA erfolgte eine Auftrennung mittels nativer Agarose-Gelelektrophorese. Die verwendete Agarose-Konzentration richtete sich nach dem gewünschten Auftrennungsbereich. Reine Mengenabschätzungen wurden in 1%igen Agarosegelen vorgenommen, während die Bestimmungen von Insertgrößen in 1,5%igen Gelen erfolgte (Sambrook *et al.*, 1989). Die Agarose wurde unter Aufkochen in 1 x TBE-Puffer geschmolzen. Nach kurzer Abkühlungsphase wurde der Gellösung unmittelbar vor dem Gießen Ethidiumbromid (0,1 µg / 100 ml) zugesetzt. Das ausgehärtete Gel wurde mit Laufpuffer (1 x TBE) übergossen und die mit Ladepuffer versetzten Proben wurden in die ausgesparten Geltaschen pipettiert. Zur Größenbestimmung wurden pro Kammreihe Längenstandards für den gewünschten Größenbereich aufgetragen. Bei 50-120 Volt Spannung erfolgte die Auftrennung der Proben. Die Detektion des in die DNA interkalierenden Ethidiumbromidfarbstoffes erfolgte mittels UV-Bestrahlung mit einem Transilluminator (Bachhofer, Reutlingen, $\lambda=254$ nm). Die Gele wurden mit einem Geldokumentationssystem (Enhanced Analysis System (E.A.S.Y), Herolab, Wiesloch) fotografiert und analysiert.

3.1.14 Elektrophoretische RNA-Auftrennung in Agarosegelen

Die analytische Auftrennung von RNA-Proben geschah analog zu 3.1.13, wurde aber mit folgenden Modifikationen durchgeführt:

- Verwendung des 1xTAE-Laufpuffers statt 1xTBE.
- Verwendung des RNA-Probenpuffers.
- Die Probe wurde zupipettiert, während bereits eine Spannung von 30 V anlag. Nach 5 Minuten Einlaufzeit wurde die Spannung auf die übliche Endspannung von 100 V erhöht.
- Benutzung der Low Melting NuSieve GTG Agarose.

3.1.15 Polymerase Kettenreaktion (PCR)

Die PCR (Mullis *et al.*, 1987) wurde zur Amplifikation des cDNA-Inserts bei Klonen mit geringen Plasmid-DNA-Ausbeuten angewendet. Entscheidend dafür war die Bewertung der Plasmidmenge nach elektrophoretischer Auftrennung (3.1.13).

Ein typischer 50 µl PCR-Ansatz enthielt 10-20 ng Plasmid-DNA, 10 mM Tris, pH 8,9, 50 mM KCl, 1,5-3 mM MgCl₂, 10 pmol jedes Primers (pUC/M13 forward, bzw. pUC/M13 reverse), 200 µM je dNTP und 1 U AmpliTaq- Polymerase. Alle PCR-Reaktionen wurden in dünnwandigen 200 µl-Reaktionsgefäßen in Thermocyclern (Gene Amp PCR Systems 2400 oder 9600) der Firma Perkin Elmer mit Deckelheizung durchgeführt.

3.1.16 Cycle-Sequenzierung

Die Basenabfolge der cDNA-Inserts der isolierten Plasmid-DNA (3.1.11) vorliegende Plasmides, bzw. PCR-Fragmentes (3.1.15) wurde mittels Dye-Terminator-Cycle-Sequencing-Verfahrens bestimmt, welches eine Weiterentwicklung der Kettenabbruchsreaktion nach Sanger (Sanger *et al.*, 1977) ist. Hierbei erfolgt durch eine thermostabile Taq-DNA-Polymerase die Neusynthese eines komplementären Stranges entlang einer Einzelstrangmatrize.

Dabei wird zunächst die doppelsträngige DNA durch Erhitzung in Einzelstränge aufgeschmolzen (Denaturierung). Im anschließenden Annealingschritt wird die Temperatur gesenkt, so dass sich der Sequenzierprimer an den Sequenzabschnitt anlagern kann. Am 3'-Ende dieses Primers wird die Replikationssynthese durch die DNA-Polymerase unter sukzessivem Einbau von dNTPs initiiert und fortgeführt (Syntheseschritt). Diese Reaktionsabfolge wird mehrfach wiederholt, so dass genügend Sequenzierprodukte für die Detektion mittels Fluoreszenz synthetisiert werden.

In einem bestimmten Mischungsverhältnis sind dem Reaktionsansatz ddNTPs zugesetzt, welchen die für die Ausbildung der Phosphodiesterbindung erforderliche Hydroxylgruppe am C3-Atom des Riboserestes fehlt. Aufgrund der rein statistischen Verteilung der ddNTP-Einbaus entsteht eine Schar unterschiedlich langer DNA-Fragmente mit invariantem 5'-Beginn (dem Primer) und variablem 3'-Ende (durch das jeweilige ddNTP determiniert). Die

hier verwendete DNA-Polymerase, die AmpliTaq DNA Polymerase, besitzt keine 3' - 5' Exonucleaseaktivität und baut die ddNTP's effizient ein.

Da die vier ddNTPs mit Fluoreszenzfarbstoffen mit unterschiedlichen Emissionswellenlängen verknüpft sind, kann die elektrophoretische Auftrennung der Sequenzierprodukte in einer Spur erfolgen.

Die Sequenzierreaktion wurde mit dem ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction Kit durchgeführt.

Die Sequenzierreaktion setzt sich bei einem Gesamtvolumen von 10 µl aus 2 µl BigDye TerminatorMix, 1 µl (5 pmol/µl) Primer und 250 ng PCR-Produkt, bzw. 500 ng Plasmid-DNA als Template zusammen. Die Cycle-Sequencing-Reaktionen wurden in dünnwandige 200 µl-Reaktionsgefäße pipettiert und auf Gene Amp PCR Systems 2400 oder 9600 Thermocyclern (Perkin Elmer) mit Deckelheizung durchgeführt. Als Standardbedingungen für die Cycle-Sequencing-Reaktionen wurden folgende Parameter verwendet: 2 min 96 °C, 25 Cyclen mit 30 s 96 °C, 10 s 53 °C und 4 min 60 °C.

Zur Abtrennung der nicht eingebauten ddNTPs wurde der Reaktionsansatz mit Isopropanol gefällt (Sambrook *et al.*, 1989). Der Überstand wurde vollständig abgesogen, das Pellet mit 200 µl 80%igem Ethanol gewaschen und anschließend im offenem Reaktionsgefäß für 1 min bei 60 °C vollständig getrocknet.

3.1.17 Elektrophoretische Auftrennung durch den ABI Prism 377 DNA-Sequencer

Die Sequenzanalyse am ABI 377-DNA Sequencer erfolgte mittels 4,8%igen PAA (Polyacrylamid)-Harnstoff-Gelen. Zur Herstellung des Gels wurde 18 g Harnstoff bei ca. 50°C in 6 ml PAGE-Plus (40%), 10 ml 5 x TBE und 18 ml HPLC-H₂O gelöst, anschließend durch einen Papierfilter (Sartorius, Göttingen) filtriert und 5 min entgast. Vor dem Gießen in die Gelvorrichtung wurden 350 µl 10%iges APS und 15 µl TEMED als Polymerisationsstarter zugegeben. Nach dem Stecken des Kamms polymerisierte das Gel in horizontaler Lage für ca. 2 h aus. Bei einer konstanten Leistung von 48 Watt über 12 h wurden die Proben auf einer Trennstrecke von 34 cm aufgetrennt. Als Laufpuffer wurde 1 x TBE verwendet.

Die Elektropherogramme wurden anschließend durch die Lasergene-Software in FASTA-Sequenzdateien überführt und manuell, bzw. durch das Querymaster-Programm ausgewertet.

3.2 Molekularbiologische Methoden zur Bestimmung von Genregulationen durch Hypoxie

3.2.1 Zellkulturen unter hypoxischen Bedingungen

Frisch gesplittete WERI-Rb1-Zellen (McFall *et al.*, 1977) wurden in einer Konzentration von ca. 10^4 Zellen/ml auf vier 14 cm-Petrischalen mit je 25 ml Medium verteilt. Zwei der Schalen wurden als Kontrollen in einem Brutschrank inkubiert, während die beiden anderen Schalen ohne Deckel in einer Hypoxiekammer bebrütet wurden. Hierbei handelte es sich um eine luftdicht abgeschlossene Kammer, in die ein Gemisch aus Stickstoff, 5% Kohlendioxid, 1% Sauerstoff und ein hoher Argonanteil geleitet wurde (Riedinger *et al.*, 1999). Argon ist schwerer als Luft und diente für den Fall eines Lecks als Schutzschicht über der offenen Zellkultur. Ein Befeuchtungssystem wirkte der Austrocknung entgegen. Die Zellen verblieben 8 h, bzw. 24 h bei 37°C in der Kammer. 200 µl Kultur wurden für eine Zellzählung mit der Neubauer-Zählkammer abgenommen. Danach wurden die Zellen bei 250 g abzentrifugiert und mit 4°C kaltem PBS gewaschen. Unmittelbar danach wurde mit der RNA-Extraktion begonnen, da nach spätestens 30 min die Zellen wieder auf den Sauerstoffnormaldruck einzustellen beginnen. Analog wurde mit den unter normoxischen Bedingungen gewachsenen Zellen aus dem Brutschrank verfahren.

3.2.2 Bestimmen der WERI-Zellzahl in einer Neubauer-Zählkammer

Ein Tropfen einer WERI-Kultur wurde auf eine Neubauer-Zählkammer getropft und lichtmikroskopisch ausgezählt. Die Zählkammer enthält 16 Kleinquadraten mit einer Seitenlänge von 0,025 mm und 0,1 mm Höhe, also 10^{-6} ml Volumen. 100 in einem Großquadrat ausgezählte Zellen entsprechen also einem Titer von 10^8 Zellen pro ml, wobei eventuelle Verdünnungen zu berücksichtigen sind.

WERI-Zellen neigen während des Wachstums zur Verklumpung. Um die korrekte Zellzahl zu bestimmen, wurde das auszuzählende Aliquot stark gevortext. Da die Konzentration der WERI-Zellen direkt ausgezählt werden konnte, wurde auf eine Verdünnung verzichtet. Im vorliegenden Experiment wurden vier Großquadrate ausgezählt und der Mittelwert ermittelt.

3.2.3 Isolierung totaler RNA durch das RNeasy Mini Kit-System

Das RNeasy-System wird bevorzugt verwendet zur parallelisierten Aufreinigung größenfraktionierter RNA, z.B. für quantitative RT-PCR (3.2.6) (Signoretti *et al.*, 2000). Die Vorteile dieses Systems liegen in der einfachen Handhabbarkeit, der Gewinnung hochreiner RNA, sowie guter quantitativer Ausbeuten (Mahadevappa *et al.*, 1999).

Die Zell-Proben wurden zunächst mit einem Puffer mit β -Mercaptoethanol und Guanidiniumisothiocyanat lysiert, wodurch gleichzeitig auch die RNAsen inaktiviert wurden. Nach Zugabe von Ethanol zur Steigerung der Bindungseffizienz wurde das Lysat durch eine Minisäule zentrifugiert, wobei RNA-Moleküle mit mehr als 200 bp Länge an eine Silicamembran binden. Zelldebris wurde durch einen vorgeschalteten Filter abgefangen. Nach zwei Waschrufen wurde die RNA aus der Minisäule mit DEPC-H₂O_{dd} eluiert.

Abweichungen von den Angaben des Herstellers:

- Der optionale DNase-Schritt wurde durchgeführt.
- Zur Erhöhung der Ausbeute wurde die Eluierung zweimal durchgeführt und die Eluate vereinigt. Das Volumen wurde anschließend durch Vakuumtrocknung in der Speedvac eingeengt.

3.2.4 Aufbewahrung von RNA

Aufgereinigte RNA wurde in DEPC-behandeltem H₂O_{dd} mit 0,1 mM EDTA (pH 7,5) bei -80°C aufbewahrt. Das Diethylpyrocarbonat (DEPC) ist ein reversibler Inhibitor von RNAsen und dient somit zur Stabilisierung der RNA. Da Nucleinsäuren im alkalischen Milieu ihr Löslichkeitsoptimum haben, wurde das EDTA zugesetzt (Sambrook *et al.*, 1989).

3.2.5 Photometrische Konzentrationsbestimmung von DNA und RNA

Eine sehr schnelle Methode zur Konzentrationsbestimmung von Nucleinsäurelösungen besteht in einer photometrischen Absorptionsmessung. Dazu wurden die Nucleinsäure-Proben so mit einem definierten Volumen an TE-Puffer oder H₂O_{dd} verdünnt, damit der angestrebte

Messwert von 0,1-1 erreicht werden konnte und die Absorption bei $\lambda=260$ nm und $\lambda=280$ nm in einer Quarzküvette (Vol.: 100 μ l) gemessen. Dabei entspricht eine Absorption bei $\lambda=260$ nm von 1,0 einer Konzentration von 50 μ g/ml doppelsträngiger DNA, bzw. 43 μ g/ml RNA. Aus dem Quotienten der OD bei $\lambda=260$ und $\lambda=280$ nm, der bei doppelsträngiger DNA zwischen 1,8 und 2,0 und bei reiner RNA bei ca. 1,5 liegen sollte, kann zudem auf die Reinheit der Nukleinsäurelösung geschlossen werden. Die Konzentrationsbestimmungen erfolgten nach folgender Gleichung:

$$\text{DNA-Konzentration } (\mu\text{g/ml}) = A_{260} \times \text{Verdünnungsfaktor} \times 50 \mu\text{g/ml}$$

$$\text{RNA-Konzentration } (\mu\text{g/ml}) = A_{260} \times \text{Verdünnungsfaktor} \times 43 \mu\text{g/ml}$$

$$A_x = \text{Absorption bei } \lambda = x \text{ nm}$$

3.2.6 Quantitative Real Time PCR (Q-PCR)

Die Q-PCR ist eine Methode zur quantitätsabbildenden Amplifikation definierter DNA, bzw. cDNA-Abschnitte. Von einer regulären PCR unterscheidet sich die Real Time-PCR durch die Zugabe eines interkalierenden Fluoreszenzfarbstoffes vor der Reaktion und Messung der Menge doppelsträngiger PCR-Produkte proportional zur Fluoreszenz nach jedem Reaktionszyklus. Ausgangssubstanz der RT-PCR ist nicht DNA, sondern mRNA, die zunächst in cDNA umgeschrieben werden muss.

Einer der wichtigsten Faktoren für eine effiziente und damit auch quantitative Amplifizierung der Proben ist die Auswahl der Primer. Deren Generierung erfolgte anhand der von Applied Biosystems festgelegten Designanweisungen (<http://docs.appliedbiosystems.com/pebi/docs/04304449.pdf>) unter Anwendung des Computerprogramms Primer Express v. 2.0. Die so erstellte Vorschlagsliste wurde manuell ausgewertet und je ein Primerpaar pro Zielgen ausgewählt (2.12).

Die Umschreibung der aus den normoxischen und hypoxischen WERI-Zellen gewonnenen RNA (3.2.3) in cDNA erfolgte in einem separaten Schritt mit Oligo-dT-Primern nach Empfehlungen von Applied Biosystems.

Die cDNA wurde 1:100 verdünnt und in die eigentliche Reaktion eingesetzt:

- 5 μ l verdünnte cDNA (Duplikate)
- 7,5 μ l 2x MasterMix (SYBR Green)
- 0,9 μ l genspezifischer Primer Forward (5 μ M)
- 0,9 μ l genspezifischer Primer Reverse (5 μ M)
- 0,7 μ l H₂O_{dd}

Für jedes Zielgen werden zwei *No Template*-Kontrollen ohne cDNA und je drei PCR-Reaktionsansätze mit cDNA angesetzt. Die Replikas der eigentlichen Reaktion wurden nach dem Standardprotokoll von Applied Biosystems im ABI Taqman 7000 durchgeführt.

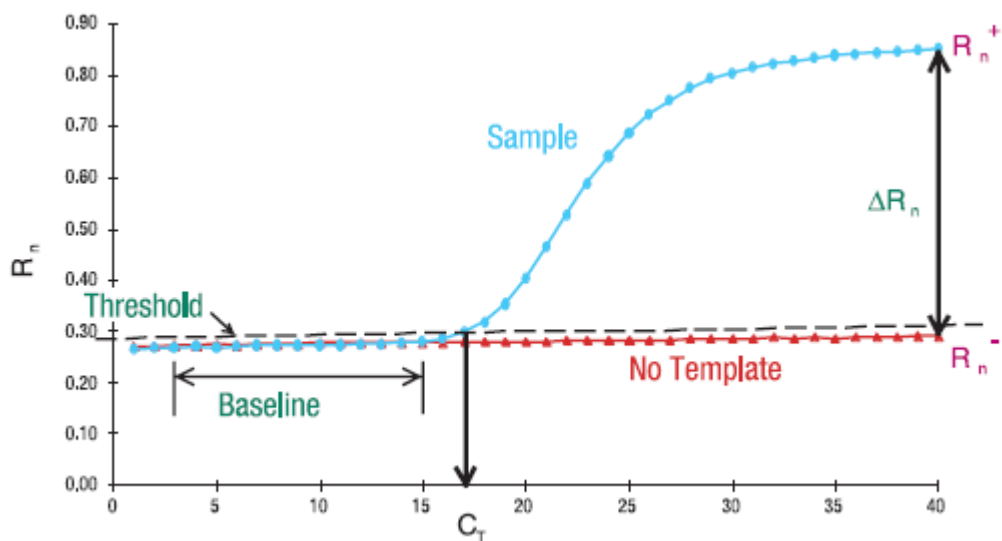


Abbildung 7: Verlaufsdigramm einer Real Time-PCR mit und ohne Template.
 C_t beschreibt die Zykluszahl, bei der die detektierte Fluoreszenz die Nachweisgrenze überschreitet.
 (Aus: <http://dna-9.int-med.uiowa.edu/RealtimePCRdocs/realtimeoverview.pdf>)

Die Ermittlung des Schwellenwertes C_t erfolgte durch Auswertung der Rohdaten durch das Sigmaplot-Programm. C_t beschreibt die Zykluszahl, bei der zum ersten Mal die Fluoreszenz einen Schwellenwert an der Basis der sigmoiden Verlaufskurve erreicht. Ein kleinerer C_t -Wert bei einer untersuchten Probe im Vergleich zu einer Vergleichsprobe bedeutet, dass die Ursprungsmenge der cDNA, bzw. der mRNA größer sein muss, da man weniger Verdopplungszyklen braucht, um die Nachweisgrenze zu überschreiten.

Erschwert wird die Berechnung von C_t dadurch, dass zwar theoretisch pro Zyklus die Produktmenge verdoppelt wird, diese maximale Reaktionseffizienz aufgrund einer Vielzahl limitierender Faktoren jedoch nie erreicht werden kann. Es musste daher vorab durch eine geometrische Verdünnungsreihe gezeigt werden, dass in dem zu messenden Konzentrationsbereich die reale Amplifikationseffizienz nicht schwankt. (Souazé *et al.*, 1996).

3.3 Bioinformatik

3.3.1 Verwendete Internetseiten

http://www.biotiquesystems.com	CRAW D2_cluster
http://www.ensembl.org	Ensembl-Homepage
http://www.equi4.com/starkit.html	sdx.kit
http://media.wu-wien.ac.at	XOTcl 1.0
http://www.ncbi.nlm.nih.gov	NCBI-Homepage
ftp://ftp.ncbi.nih.gov/blast/executables/release	Download Anwendungsprogramme z.B. formatdb und Blast
ftp://ftp.ncbi.nlm.nih.gov/blast/db/	Download allgemeiner Blast- Datenbanken, z.B. non redundant
ftp://ftp.ncbi.nih.gov/refseq/H_sapiens/mRNA_Prot/	Download humane refseq-Datenbank
http://www.phpmyadmin.net	PhpMyAdmin 2.6.1
http://www.phrap.org	Cross_match 0.990319
http://www.python.org	Python 2.2.2
http://www.sanbi.ac.za/codes	STACK_PACK 2.2
http://www.tcl.tk	Tcl/Tk 8.4.2
http://tcllib.sourceforge.net	tcllib
http://tdom.org	tDOM 0.8
http://tkcon.sourceforge.net	tkcon 2.3
http://www.xdobry.de/mysqltcl	MySQLTcl 2.5

3.3.2 Verwendete Programmiersprachen

3.3.2.1 Tcl/Tk 8.4.2

Tcl (*Tool Command Language*) und Tk (*Tool Kit*) werden selten getrennt erwähnt, obwohl es sich um zwei unterschiedliche Systeme handelt. Tcl ist die eigentliche Programmiersprache, wohingegen es sich bei Tk um eine in Tcl geschriebene Bibliothek von graphischen Unterstützungsprogrammen oder –programmteilen handelt.

Die Vorteile von Tcl/Tk sind die leichte Erlernbarkeit, die durch die Verwendung von Bibliotheken geringe Programmierzeit, sowie die Verwendbarkeit in unterschiedlichen

Betriebssystemen („*cross platform portability*“). Es handelt sich um ein Open Source-Produkt; es ist somit kostenlos und unterliegt einer ständigen Weiterentwicklung, welche zu erweiterten Verwendungsmöglichkeiten, verbesserter Stabilität und größerer Prozessierungsgeschwindigkeit führt.

Beide Querymaster-Versionen wurden ausschließlich in Tcl/Tk geschrieben. Als graphische Unterstützungsprogramme aus Tk wurden die GUI-Pakete BWidget, Tablelist und Icons importiert, wobei GUI für Graphical User Interface steht.

3.3.2.2 Python 2.2.2

Python ist wie Tcl/Tk eine bibliotheksorientierte Programmiersprache und dieser insgesamt funktionell sehr ähnlich. Unterschiede liegen darin, dass Python etwas schneller als Tcl/Tk ist und sich auf dem Markt besser durchgesetzt hat. Tcl/Tk hat hingegen Vorteile beim Zusammenspiel mit anderen Programmiersprachen („*glueing language*“) und gilt als portabler, d.h. läuft auf mehr Betriebssystemen.

Das Cluster-Programm StackPACK 2.2 ist in Python geschrieben und benötigt daher eine lauffähige Version dieser Programmiersprache.

3.3.3 Verwendete Tcl/Tk-Module

3.3.3.1 SDX zum Wrappen der Querymaster-Module

In der Entwicklungsphase müssen Programmmodule durch Eingaben in der Kommandozeile (shell-Oberfläche) bedient werden und liegen mehr oder weniger ungeordnet als Bündel von kleinen Dateien in einem Ordner. Um diese zusammengehörigen Anwendungspakete portabler, übersichtlicher und benutzerfreundlicher, z.B. durch bloße Anwahl mit der Maus aktivierbar zu machen, müssen sie gewrappt (etwa: verpackt, verhüllt) werden, so dass für den Benutzer die „Illusion einer Einzeldatei“, z.B. einer *.exe-Datei entsteht, während sich für die Programmroutinen wenig ändert. Kommerzielle Anwendungen liegen grundsätzlich gewrappt vor, da sonst ihr Quellcode einsehbar wäre.

Das Wrappen der Querymaster-Module (3.2.5) wurde mittels des SDX-Kits durchgeführt. SDX steht für *starkit developer extension*, eine Sammlung von Anleitungen und Bibliotheken,

die speziell für das Wrappen von Tcl/Tk-Modulen, bzw. das Verpacken durch Tcl/Tk-Skripte entwickelt wurden. Starkit ist wiederum ein Akronym für *StandAlone Runtime for Tclkit* und bezeichnet das eigentliche Programm zum Verpacken.

3.3.3.2 XOTcl 1.0

XOTcl ist eine Erweiterung für Tcl, die es erlaubt, objektorientiert zu programmieren, was beim ursprünglichen Tcl nicht unterstützt wird. Unter objektorientiertem Programmieren (OOP) versteht man die Unterteilung eines Programms in separate Untereinheiten (Module), wobei Daten und die Anweisungen, mit denen die Daten bearbeitet werden, an einer Stelle gespeichert werden. Durch diese so genannte Kapselung wird eine größere Zuverlässigkeit und Wiederverwendbarkeit des Programms erzielt, da fehlerhafte, bzw. veraltete Programmteile ohne Störung anderer Module ersetzt werden können. Ein zweiter Vorteil ist, dass durch diese Segmentierung eine Adressierung von Datenströmen leichter wird und somit fehlerhaftes Zuordnen seltener vorkommt.

XOTcl unterstützt dabei einige sehr fortschrittliche OOP-Mechanismen für die Wiederverwendung von bereits vorhandenem Code (Vererbung) und für aspektorientierte Programmierung. XOTcl ist somit anderen Programmiersprachen in Bereichen der OOP überlegen. Bei aspektorientierter Programmierung ist es möglich, Methodenaufrufe "abzufangen" und Aktionen zu definieren, die vor oder nach diesen Methodenaufrufen ausgeführt werden sollen. Übliche Aktionen für aspektorientiertes Programmieren sind hier z.B. Logging oder Filterung von Daten.

3.3.3.3 tDOM 0.8

TDOM ist ein Tcl-Skript und wird zur Datenverarbeitung von XML-Dateien benötigt. XML (*extensible markup language* - zu deutsch: Erweiterbare Auszeichnungssprache) ist ein Speicherformat, das zum systemübergreifenden Zugriff auf Daten aller Art, insbesondere aber auf Bildformate benutzt wird. Als zweiten Vorteil bietet XML die Möglichkeit, selbstdefinierte Datenstrukturen zu entwerfen, welche für die eigenen Systemanforderungen optimiert sind. Bei der Entwicklung des Querymasters II (3.2.5.2) wurde auf XML-Dateien und somit auf tDOM bei der Auswertung der Blast-Resultate, der

zugehörigen Unigene- und EntrezGene-Resultate zurückgegriffen. TDOM wurde weiterhin bei der Validierung und Reparatur nicht korrekt abgespeicherter EntrezGene-Dateien benötigt. Einer der größten Vorteile dieses Skripts ist seine geringe Speicherplatzanforderung.

3.3.3.4 tkcon 2.3

Tkcon ist eine sehr umfangreiche Erweiterung der standardmäßig mitgelieferten Arbeitsplatzkonsole von Tk. Wie bei vielen anderen Tcl/Tk-Skripten ist auch diese auf diversen Betriebssystemen lauffähig. Obwohl man damit auch die normale Betriebsoberfläche von Unix/Windows-Maschinen ersetzen könnte, wird tkcon primär als graphische Oberfläche zur detaillierten Anzeige laufender Prozesse in Tcl/Tk-Skripten benutzt. Beim Querymaster II wurde die von tkcon ausgegebene Konsole hauptsächlich zum Aufdecken fehlerhafter Prozeduren (debugging) verwendet.

3.3.3.5 Tcllib

Tcllib ist eine der populärsten Bibliotheken zur Programmiersprache Tcl. Sie enthält eine Menge Funktionen vor allem für internetbezogene und programmiertechnische Alltagsaufgaben und ist selbst vollständig in Tcl geschrieben. Somit stellt sie eine Erweiterung der Tk-Bibliothek dar, in der lediglich graphische Unterstützungsskripte enthalten sind.

3.3.3.6 MySQLTcl 2.5

Bei MySQLTcl handelt es sich um eine Kollektion von Tcl-Kommandos und globalen Arrays (d.h. indizierte Listen von Variablen, die innerhalb eines Programms oder Programmteils gleich sind), die Tcl-Programmen wie dem Querymaster II (3.2.5.2) Zugang zu einer MySQL-Datenbank oder einem MySQL-Server über die MySQL C-API Schnittstelle ermöglichen. C-API steht für *Application Programming Interface*, welches in der Sprache C geschrieben wurde.

3.3.4 Verwendete Clustering-Module

3.3.4.1 StackPACK 2.2

Eine cDNA-Datenbank wird idealerweise aus einer möglichst umfassenden und fehlerfreien Sammlung komplett annotierter, nicht redundanter *full-length* cDNAs eines Gewebetyps oder eines Organismus` bestehen. Selbst erstellte oder aus dem Internet importierte EST-Datenbanken genügen meist nicht allen dieser Kriterien. StackPACK ist ein in Python geschriebenes Programmpaket, das zur Generierung von qualitativ hochwertigen *in silico* EST-Datenbanken dient (Christoffels *et al.*, 2001). Die Module `cross_match`, `d2_Cluster`, `Phrap`, sowie `CRAW` mussten separat von StackPACK erworben und kompiliert werden.

Die Daten werden in der folgenden Reihenfolge prozessiert:

- 1) Einlesen durch `stack_Import`
- 2) Maskierung durch `Cross_match`
- 3) Grobes Vorclustern durch `d2_Cluster`
- 4) Stringentes Clustern durch `Phrap`
- 5) Subclustern durch `CRAW`
- 6) Bildung von *linked-clusters* durch `link_out` (optional)
- 7) Export der Daten zur weiteren Verarbeitung durch den Querymaster II

3.3.4.2 `Stack_ImportFasta` / `Stack_ImportGenebank`

Sequenzdaten wurden auf zwei Arten in das StackPACK-Programm importiert: a) im FASTA- oder b) im Genbank-Format. Genbank-Dateien gleichen Inhalts sind etwa viermal so groß wie FASTA-Dateien und somit etwas langsamer in der Prozessierung, können aber neben Singletons und Consensi zusätzlich *linked-clusters* (3.3.4.7) als dritte Resultatebene erzeugen.

3.3.4.3 Cross_match 0.990319

Cross_match ist ein Programm zur Maskierung unerwünschter Sequenzen, welche die Qualität des nachgeschalteten Clusters stören würden (Green, 1996). Es maskiert Repeats, Sequenzen niedriger Komplexität (z.B. Poly-A-Bereiche) und Vektoren. Sequenzvergleiche werden durch den Smith-Waterman-Gotoh-Algorithmus auf der Basis von Homologiesuchen durchgeführt. Die zu maskierenden Bereiche werden durch die Repbase-Update-Datenbank des Genetic Information Research Institute (G.I.R.I.) spezifiziert.

Abweichungen von den Grundeinstellungen:

- a) Da cross_match in der Grundeinstellung nur maximal 64.000 Datensätze, bzw. nur Sequenzen bis 64.000 bp Länge verarbeiten kann, wurde bei der Installation die „make manyreads“-Option gewählt.

- b) Standardmäßig werden maskierte Bereiche mit „N“ überschrieben. Dies sollte durch Änderung der StackPACK-Konfigurationsdatei in „X“ variiert werden, da es sonst zu einer Vermischung von Maskierung und Sequenzierfehlern kommen könnte. Verwendet man hingegen cross_match nicht für ein nachfolgendes Clustern der Sequenzen, sondern als reines Maskierungsprogramm und möchte die Sequenzen danach blasten, sollte auf die Grundeinstellung zurückgegriffen werden, da sonst die zu untersuchende Sequenz möglicherweise nicht als DNA-Strang erkannt wird.

- c) Die minmatch-Variable wurde von 10 auf 12 gesetzt. Minmatch definiert die minimale Anzahl von Matches als Schwellenwert, bei der eine Homologie durch cross_match als solche erkannt wird. Eine Erhöhung dieses Wertes führt zu einer drastisch beschleunigten Prozessierung, erkennt jedoch auch weniger Homologien zwischen eigenen Sequenzen und der Repbase-Update-Datenbank. Der Wert von 12 hat sich bei der Maskierung von Datenbanken als guter Kompromisswert herausgestellt (Phil Green, University of Washington, persönliche Mitteilung)

3.3.4.4 D2_cluster

D2_cluster (Burke *et al.* 1999) ist das erste Assemblierungsscript des StackPACK-Programms. Seine Aufgabe ist ein grobes und schnelles Vorclustern der Sequenzen in Indexklassen, die dann durch stringenteres Nachclustern durch Phrap (3.3.4.5) und CRAW (3.3.4.6) verfeinert wird.

Die angewendete Methode ist die einer agglomerativen Clusterbildung: jede Sequenz beginnt in einem eigenen Cluster („*Single cluster*“) und wird durch eine Serie von Abgleichsschritten mit anderen Clustern vereint (Johnson und Witchern 1994). Diesen Prozeß des Clusters bezeichnet man als „*minimal linkage clustering*“ (von anderen Autoren transitive Schließung genannt). Hierunter versteht man, dass selbst Sequenzen in einem Cluster zusammengefasst werden, die untereinander keine Ähnlichkeiten aufweisen, aber hinreichend homolog zu einer dritten Sequenz sind, die beide Ausgangssequenzen enthält (z.B. werden zwei unterschiedliche Exons in ein Verzeichnis geschrieben, sofern eine *full-length* cDNA als Bindeglied vorliegt). Der Schwellenwert für das Zusammenfassen von Sequenzen in einen Cluster wird durch zwei Kriterien festgelegt: die Fenstergröße bezeichnet die minimale Anzahl der zu vergleichenden Basenpaare per Berechnungsschritt und die Stringenz den prozentualen Anteil der Identität zweier Sequenzen (Hide *et al.* 1994).

D2_cluster sucht mit einer Fenstergröße von 150 Basenpaaren und einer Stringenz von 96% Identität ab, wobei Sequenzen mit weniger als 50 validen Basen nicht geclustert, sondern als Singletons abgespeichert werden. Nur A, C, G und T werden als valide Basenpaare gewertet, maskierte Bereiche werden nicht zu den validen Bereichen gezählt.

3.3.4.5 Phrap 0.990319

Phrap („*Phil's revised assembly program*“) dient zur Weiterverarbeitung der von d2_cluster erzeugten Daten, wobei Phrap einen stringenteren Algorithmus zur Abtrennung ungleicher Sequenzen benutzt und somit qualitativ bessere Alignments generiert (Gordon *et al.*, 1998). Die so erzeugten Singletons bleiben jedoch mit dem von d2_cluster erzeugten Cluster logisch verbunden, so dass sich eine aufgrund hoher Stringenz erfolgte Trennung homologer Sequenzabschnitte in der Auswertung nicht negativ auswirken sollte. Das Clustern ist grundsätzlich eine Gratwanderung zwischen der Erzeugung falsch negativer Datensätze (homologe Sequenzen werden nicht verbunden) und falsch positiver Datensätze (nicht

homologe Sequenzen werden verbunden). Der Vorteil einer Verknüpfung von D2-Cluster und Phrap liegt darin, dass bei dieser Vorgehensweise die Vorzüge beider Programme addiert werden (Adams *et al.*, 1995)

Phrap kann so eingestellt werden, dass es die von `cross_match` erzeugten Maskierungen ignoriert, falls z.B. kurze Repeats wertvolle Sequenzinformation darstellen würden. Da aber lange Repeats den Assemblierungsprozess empfindlich stören würden, wurde diese Option abgewählt. Weiterhin kann Phrap so konfiguriert werden, dass nur ein Sequenzstrang assembliert wird. Falls z.B. die cDNA in einem Projekt direktional kloniert und ausschließlich in einer Richtung sequenziert wird, kann der Rechenaufwand so ohne Informationsverlust erheblich verringert werden. In der vorliegenden Arbeit wurden beide Stränge zur Erstellung des Clusters herangezogen.

3.3.4.6 CRAW

CRAW ist das finale Assemblierungsprogramm in der StackPACK-Pipeline und baut auf den Ergebnissen von Phrap auf. CRAW bewertet die Sequenzcluster im Vergleich zur ermittelten Consensussequenz und bildet im Falle einer signifikanten Abweichung Subcluster (Chou und Burke, 1999). Diese Subcluster decken Inkonsistenzen der vorhergehenden Assemblierung auf und sind z.B. interessant für die Analyse von Splicevarianten eines Genes. CRAW kann aber auch zur Identifizierung von chimärischen Klonen, genomischer Kontaminierung und anderen Artefakten benutzt werden.

CRAW generiert Subcluster, falls die Sequenzidentität mit der Consensussequenz in einem 100 bp-Fenster geringer als 49% ist. Dabei werden die ersten 100 bp einer Sequenz ignoriert, um eine Auftrennung in Subcluster wegen nicht erkannter Vektorsequenzen oder mangelhafter Sequenziergenauigkeiten z.B. durch Primerwolken zu Beginn eines Datensatzes zu vermeiden. CRAW hängt damit sehr stark von der Qualität der Sequenzdaten ab, da zu kurze Sequenzen nicht ausreichend prozessiert werden können.

3.3.4.7 Link_out

Wenn eine klonierte cDNA jeweils vom 5`- und vom 3`-Ende her sequenziert wird, muss dies bei hinreichend großen Inserts nicht zu einer Überlappung beider Sequenzen führen.

Herkömmliche Verfahren würden diese beiden DNA-Sequenzen nicht in einem Cluster vereint werden. Sofern keine Chimären vorliegen, gehören beide Seiten des Inserts jedoch zur selben cDNA und wertvolle Information würde verloren gehen. Falls der zu analysierende Datensatz im Genbank-Format vorliegt, liest Link_out den Namen des annotierten Klons aus und speichert die 5'-Sequenz in dasselbe Verzeichnis wie die entsprechende 3'-Sequenz des Klons. Dieser Prozeß betrifft jedoch nur ESTs, die als Singletons vorliegen.

3.3.5 Automatisierte Blast-Abfrage und Annotation

3.3.5.1 Querymaster I

Der Querymaster I (QM I) ist ein Tcl/Tk-Programmpaket zur automatisierten Generierung von Blast-Ergebnisdateien auf einer Windows-Oberfläche als Alternative zum speicherintensiven lokalen blastall-Programm. Hierzu wurden die vom Lasergene-Programm erzeugten FASTA-Sequenzen in ein Zielverzeichnis kopiert, durch den QM I in Abfragedateien *.inq umgewandelt und über die Blast-API-Schnittstelle an das Internet-Blast-Portal des NCBI zum Homologieabgleich gesandt. Nach Beendigung des Blast-Prozesses wurden die entsprechenden Ergebnisdateien als *.out-Textdateien in das Ausgangsverzeichnis kopiert. Sollte ein Ergebnis nicht innerhalb eines vorgegebenen Zeitfensters vorliegen, wurde nach 10 Minuten jede Minute eine Ergebnisabfrage an das NCBI gesandt. Dieses Verfahren stellte sich als zu aggressiv heraus, da bei jeder Anfrage des QM I die Server des NCBI mit Rechenoperationen belastet wurden. Daher wurde die Anfragelast stufenweise auf ein Zehntel des ursprünglichen Wertes herabgesetzt. Die Auswertung der Blastergebnisse erfolgte nicht automatisiert.

3.3.5.2 Querymaster II

Der Querymaster II (QM II) stellt eine Weiterentwicklung des QM I dar. Der QM II ist ein linux-optimiertes Tcl/Tk-Programmpaket zur vollautomatischen Blast-Abfrage und Annotation von cDNA-Sequenzen.

Es besteht aus mehreren einzeln ansteuerbaren Modulen:

a) Das erste ist ein FASTA-Browser zur Validierung der eingelesenen DNA-Sequenzen. Da StackPACK ein ähnliches Validierungsskript zum Ausschluss von Proteinsequenzen benutzt, wurde der FASTA-Browser bei der Verarbeitung von StackPACK-generierten Sequenzclustern nicht verwendet.

b) Der zweite Programmteil ist eine graphische Oberfläche zur Ansteuerung einer lokalen blastall-Anwendung (Altschul *et al.*, 1990). Hier können Stapel-Sequenzen (in einer Datei vereinigte Sequenzen, die durch ein Trennungszeichen wie „>“ voneinander separiert werden) eingelesen werden, sowie die blast-Parameter und die abzugleichenden Datenbanksegmente ausgewählt werden. Sollte ein batch-Abfrage abgebrochen worden sein, erkennt das Programm durch Setzung von Markierungen, bis zu welcher Sequenz ein Ergebnis vorliegt und kann den Prozeß an dieser Stelle wieder aufnehmen. Änderungen der Standard-Abfrageparameter können in der .querymaster2rc-Datei vorgenommen werden. Die Ausgabe erfolgt als XML-Datei.

c) Das dritte Modul ist der Blast-Favoritengenerator. Üblicherweise besteht ein Blast-Resultat aus den 50 besten Ergebnisdatensätzen und den zugehörigen Alignments. Würde jedes Ergebnis für die nachfolgenden Analysen gleichwertig behandelt werden, würde sich die Datenmenge verfünffzigfachen, was ohne größeren Informationsgewinn sehr speicher- und rechenintensiv wäre. Der Favoritengenerator wertet die Liste der Blast-Ergebnisdatensätze nach bestimmten Kriterien aus und vergibt für jeden Datensatz einen entsprechenden Gewichtungsfaktor, die Bonus-Punktzahl, wobei die meisten Operatoren durch Wortsuche aus der Ergebnisüberschrift („Header“) gewonnen werden, so definiert z.B. das Wort „Homo“ oder „human“ die Sequenz als menschlichen Ursprungs. Das Alignment mit der höchsten Punktzahl wird als Favorit definiert und wird in nachgeschalteten Analyse-Modulen verwendet. Bei Datensätzen mit gleicher Bonus-Punktzahl wird der in der ursprünglichen Ergebnisfolge voranstehende Datensatz verwendet. Die Ergebnisse dieser Punkteverteilung werden in *.fav-Dateien zur Nachkontrolle gespeichert. Durch Änderung der Parameter in der KlassifikatorExample.tcl-Datei kann das Skript den jeweils gewünschten Bedingungen angepasst werden.

Die Punkteverteilung in der Grundeinstellung lautet übersetzt wie folgt:

Und-Bedingungen:	Punkte
human, mRNA-Derivat (z.B. cDNA, Protein, ...), kein Vektor, Ref-Sequenz, Blast-Hitscore > 200	100
non-human menschlich, mRNA-Derivat, kein Vektor, Ref-Sequenz, Blast-Hitscore > 200	50
kein Vektor, Blast-Hitscore > 100	10
keine Übereinstimmung mit oberen Kriterien	0

d) Die zum in c) ausgewählten Favoriten passende zugehörige Sequenzbeschreibung wurde ausgelesen und die enthaltenen Codes wie z.B. die Accession-Nummer wurden zur Datenabfrage der Unigene- und Entrezgene-Datenbanken über einen Browser benutzt. Die enthaltenen Unigene-, bzw. Entrezgene-Datensätze wurden als *.fav.ug oder *.fav.eg-Datei lokal abgespeichert. Vor der Umstellung Entrezgene beim NCBI wurde die LocusLink-Datenbank lokal angesteuert.

Obwohl nur geringe Serverlasten durch das web-basierte Ansteuern der NCBI-Datenbanken auftraten, wurde eine Zeitverzögerung zwischen den Abfragen vorgeschaltet, um etwaigen Überlastungen vorzubeugen.

Die Nachteile der Abfrage-Ansteuerung über einen Browser liegen in der geringeren Geschwindigkeit, die Vorteile in der kürzeren Programmierzeit und der fehlenden Notwendigkeit, die anzusteuernden Datenbanken lokal zu speichern und zu erneuern.

e) Die Ansteuerung der MySQL-Tabelle zur finalen Abspeicherung und Auswertung der annotierten Daten erfolgt über drei Module. Das erste Modul tablecreate.tcl erzeugt eine MySQL-Tabelle mit den benötigten Spalten und Formaten. Das zweite Modul batchtable.tcl füllt die Tabelle mit den aus den Entrezgene-, bzw. Unigene-Datensätzen ausgelesenen Daten. Das dritte Modul tabledrop.tcl löscht eine Tabelle, sollte sie nicht mehr benötigt werden. Die Datenbank-Konnektivität wird durch die MySQL-API-Schnittstelle einerseits und das MySQLTcl-Modul (3.2.3.6) andererseits gewährleistet.

3.3.6 Verwaltung der MySQL-Ergebnisse durch PhpMyAdmin 2.6.1

PhpMyAdmin ist ein in PHP (Hypertext Preprocessor) geschriebenes Skript, das die umfassende Administration von MySQL-Datenbanken über einen Webbrowser ermöglicht. In dieser Arbeit wurden damit die von StackPACK und Querymaster II generierten Datensätze visualisiert.

Die Vorzüge von PhpMyAdmin liegen in

- der intuitiveren Bedienbarkeit im Vergleich zur MySQL-Shell (graphische Shortcuts)
- der größeren Übersichtlichkeit der Tabellen durch eine graphische Oberfläche
- dem Zugang externer User auf die Datenbank durch Webansteuerung
- der kostenlosen Distribution (Open Source)

Um den Missbrauch durch externe Benutzer einzuschränken, wurde das Programm durch eine Passwortabfrage geschützt.

3.3.7 Datenbankmanagement

3.3.7.1 MySQL 3.23.27

Relationale Datenbanken speichern Daten in separaten Tabellen, anstatt sie alle in einem einzelnen großen Speicherraum abzulegen, wodurch sowohl große Geschwindigkeiten als auch Flexibilität bei Abfragen erzielt werden können. SQL („*Structured Query Language*“) ist eine standardisierte Sprache für relationale Datenbankoperationen. Hierunter versteht man die logische Verbindung der einzelnen Tabellen durch definierte Parameter, was verknüpfte Abfragen aus verschiedenen Tabellen ermöglicht.

MySQL ist die derzeit populärste Open Source SQL-Datenbank, die sich durch einfache Handhabbarkeit, große Stabilität, geringe Kosten und vor allem hohe Geschwindigkeiten auszeichnet (Widenius *et al.*, 2002).

Es kam die Version 3.23.27 zur Anwendung.

3.3.7.2 Erstellung der „nr-eye“-Datenbank

Zur Erstellung einer nicht-redundanten, annotierten Datenbank des menschlichen Auges wurden die folgenden MySQL-Befehle („Statements“) verwendet. Zum besseren Verständnis der Syntax sind die benötigten Statements in der rechten Spalte erklärt.

MySQL-Statement	Bedeutung
SELECT * ,	Primärbefehl: Anzeige aller Tabelleninhalte
COUNT (*) AS Anzahl	Sekundärbefehl I: allen Ergebniszeilen eine neue Indexspalte „Anzahl“ zuweisen, und die Menge der identischen Ergebnisse darin zählen
FROM clones AS c, blast AS b, unigene AS u, entrezgene AS e	Festlegung der abzufragenden, bzw. zu vereinenden Tabellen und Definition einer verkürzten Schreibweise
WHERE c.id = b.clone_id AND b.fav_entrez_gene_id = e.id AND b.fav_unigene_id = u.id	Primäre Bedingung: unter welchen Voraussetzungen findet eine Verknüpfung der Tabellen statt (in diesem Falle bei gleicher Accession-, bzw. CID oder GI-Nummer)
AND (u.cDNA_sources LIKE "%eye%")	Sekundäre Bedingung: nur laut Unigene im Auge transkribierte ESTs benutzen
GROUP BY b.fav_hit_def ORDER BY Anzahl DESC	Sekundärbefehl II (bezieht sich auf COUNT): alle Ergebniszeilen mit gleichem Treffer („definition line“ des Blastresultats) zusammenfassen, Multiplizität anzeigen und in absteigender Reihenfolge ausgeben

4 Ergebnisse

4.1 Erstellung und Analyse einer cDNA-Bank aus WERI Rb1-Zellen

4.1.1 Generierung der cDNA-Bank

Humane Retinae sind nur sehr limitiert verfügbar; zudem ist die Qualität humaner mRNA aus Spenderorganen sehr unterschiedlich. Daher wurde auf das Modellsystem der humanen Retinoblastom-Zelllinie WERI *Rb1*-Zellkultur (McFall *et al.*, 1977) als mRNA-Quelle für die Erstellung der cDNA-Bank zurückgegriffen.

Es wurden sukzessive 14 200 ml Nunclon-Zellkulturflaschen mit WERI-Zellen angezogen und abgeerntet (3.1.1). Die Zellen wurden mit Trizol aufgeschlossen (3.1.4). Da die Qualität der gewonnenen RNA ein wesentlicher Faktor für die Generierung einer qualitativ hochwertigen cDNA-Bank ist, wurde die Gesamt-RNA aus der Trizol-Extraktion einer Cäsiumchlorid-Dichtegradientenzentrifugation unterzogen (3.1.5). Aus der so gewonnenen Gesamt-RNA wurde die mRNA mittels Oligotex-Säulen isoliert (3.1.6). Die photometrische Mengenbestimmung (3.2.5) ergab 2,3 µg mRNA; diese wurden komplett zur Generierung der cDNA-Bank in das SuperScript Plasmid System for cDNA Synthesis and Plasmid Cloning eingesetzt (3.1.7). Nach der säulenchromatographischen der cDNA wurde die größte Fraktion ligiert und kloniert. Die beim Ausstrich auf IPTG/XGal-Ampicillin-Platten enthaltenen weißen Klone wurden gepickt und in 96er Wachstumsblöcken kultiviert (3.1.9). Insgesamt konnten ca. 3.000 Klone mit einer durchschnittlichen Insertlänge von ca. 1 Kb gewonnen werden.

4.1.2 Optimierung der Plasmidextraktion aus der cDNA-Bank

Das Finden optimaler Wachstumsparameter für die Anzucht der cDNA-Klone ist ein multiphasischer Prozess. Initial wurden die Plasmid-Extraktionskits diverser Anbieter auf ihre Funktionalität getestet: das Perfect Prep Plasmid 96-Kit von Eppendorf, das NucleoSpin Robot-96 Plasmid Kit von Macherey & Nagel, das Concert 96 Plasmid Purification-Kit der Firma GibCo, das R.E.A.L.-Kit der Firma Qiagen, sowie das Qiaprep 96 Turbo BioRobot Kit von Qiagen. Die anfänglichen Standardbedingungen waren 20 h Wachstum bei 37°C unter

Ergebnisse

225 rpm auf dem Schüttelinkubator bei Zugabe von 100 µg Ampicillin/ml in 1,2 ml LB-Medium. Hierbei zeigte es sich, dass das neu auf den Markt kommende Concert 96 Plasmid Purification-Kit der Firma GibCo eine deutlich höhere hohe Anzahl von Ausfällen bei der Präparation von 96 nicht identischen Klonen erbrachte: Die Ausfallquote lag bei ca. einem Drittel gegenüber 20-25% bei anderen Anbietern. Folglich wurde dieses Kit von den weiteren Optimierungsschritten ausgeschlossen.

Zunächst wurden drei Wachstumsmedien LB, 2xYT und Terrific Broth getestet: da Terrific Broth aufgrund eines sehr starken Wachstums die Filter der Präparations-Kits verstopfte, erbrachte 2xYT die beste DNA-Ausbeute und wurde für die weiteren Schritte verwendet. Weiterhin wurden Wachstumszeiten von 16 h, 20 h und 24 h getestet und standardgemäß bei 20 h belassen. Die Menge des eingesetzten Mediums wurde leicht variiert und auf 1,3 ml angehoben. Um etwaige Kontaminationen aufgrund der Verwendung des luftdurchlässigeren AirPore Tape Sheets entgegenzuwirken, wurde die Menge des Ampicillins verändert und auf 150 µg/ml eingestellt. Zur Nivellierung heterogener Wachstumskurven wurde eine Vorkultur angesetzt; von dieser wurden 10 µl als Starterkultur eingesetzt.

Eine erhebliche Verringerung der Sequenzierausfälle konnte beim R.E.A.L.-Kit durch das kurzfristige Erhitzen des neutralisierten Lysats nach dem Passieren des Filters zur Inaktivierung etwaiger Nucleasen erzielt werden.

Nach Auswertung der Optimierungsschritte zeigte sich, dass die Präparation mittels des Qiaprep 96 Turbo BioRobot Kits die geringste Anzahl an Sequenzierausfällen nach sich zog. Aufgrund des besseren Preis-Leistungs-Verhältnisses wurde jedoch das nur marginal schlechtere R.E.A.L.-Kit der Firma Qiagen bevorzugt.

Nach Auswertung der ersten Sequenzierungen wurde festgestellt, dass eine große Anzahl von Klonen mit identischen Inserts vorlag. Nach der Elektroporation (3.1.7) wurde die Bakteriensuspension nach Angaben der Hersteller 60 Minuten zur phänotypischen Expression mit SOC-Medium bei 37°C geschüttelt. Die Generationszeit von *E.coli* unter optimalen Bedingungen liegt bei knapp 25 Minuten, daher könnte es bereits zu Wachstumsprozessen und so zum Ausstreichen identischer Kolonien gekommen sein. Um dies zu verhindern, wurde bei der Klonierung eines zweiten Aliquots der längsten Insertfraktion die phänotypische Expression auf 30 Minuten verkürzt. In der Folge sank die Anzahl der gewachsenen Klone um etwa die Hälfte, die Anzahl der identischen Inserts etwa um ein Viertel.

Nach erfolgter Analyse aller Sequenziererergebnisse waren noch knapp 5% aller Klone nicht auswertbar, die anfänglichen Ausfälle also um ca. 80% reduziert. Um herauszufinden, ob diese restlichen Sequenzierausfälle auf grundsätzliche Schwankungen in der Präparation zurückzuführen waren, wurde ein erfolgreich sequenzierter Klon mittlerer Insertlänge 66fach angezogen und mit dem R.E.A.L.-Kit aufgereinigt. Obwohl alle Wachstumskammern ein gleich großes Zellpellet nach der Abzentrifugation erbrachten, ergaben drei Klone bei einer Analyse im Gel keine, bzw. nur eine sehr dünne Bande. Dieser systematische Fehlerquotient korrelierte gut mit den Sequenzierausfällen nach erfolgter Optimierung.

Um Mengenschwankungen auszugleichen, wurden bei nicht sequenzierbaren Proben die aufgereinigten Eluate als Ausgangsmaterial für eine PCR mit den randständigen Sequenzierprimern *forward* und *reverse* genommen, um eventuell vorhandene, aber in der Konzentration zu geringe Mengen vorliegende Plasmide zu vervielfältigen und erneut zu sequenzieren. (3.1.15). Nach der Optimierung der PCR-Bedingungen wurde unter Verwendung von 2 µl Plasmideluat und AmpliTaq I-Puffer folgendes Touchdownprotokoll verwendet:

94°C	4 min	1 Zyklus
94°C	30 sec	
56°C	30 sec	10 Zyklen
72°C	1,5 min	
94°C	30 sec	
53°C	30 sec	25 Zyklen
72°C	1,5 min	
72°C	5 min	1 Zyklus

4.1.3 Auswertung der Sequenzen aus der cDNA-Bank

Im Anschluß an die Sequenzierungen wurden die Resultate im FASTA-Format gespeichert und für eine Blast-Analyse gegen das *non redundant*-Segment von GenBank die non redundant-Datenbank, bzw. im Falle uninformativer Ergebnisse gegen die dbEST-Datenbank des NCBI unter Anwendung der Standardparameter durchgeführt. Die Blast-Abfragen erfolgten anfangs manuell über das Web-Interface von NCBI und später unter Anwendung

des Abfrage-Tools Querymaster I. Die Ergebnisse wurden manuell in eine Access-Tabelle übertragen. Ausgehend vom informativsten Blast-Resultat (Parameter analog zu 3.3.5.2) wurden zugehörige Unigene- und LocusLink-Daten der entsprechenden Sequenz über das Web-Interface des NCBI abgefragt und, soweit vorhanden, der Access-Tabelle zur vollständigen Annotation beigelegt.

Von 943 der 960 cDNA-Klone (98%) konnten Sequenzdaten gewonnen werden. 42 der Klone enthielten kein Insert, 14 lediglich den poly-A-tail ohne verwertbare spezifische Sequenzabschnitte und 83 Klone stellten sich als Kontamination mit einem RBP4-Konstrukt heraus. Unter den verbliebenen 804 Klonen wurden 23 cDNA-Sequenzen mitochondrialen Ursprungs detektiert. Da mitochondriale Transkripte zumindest kurze poly-A-tails besitzen, ist die Präsenz entsprechender Klone in der cDNA-Bank nicht überraschend.

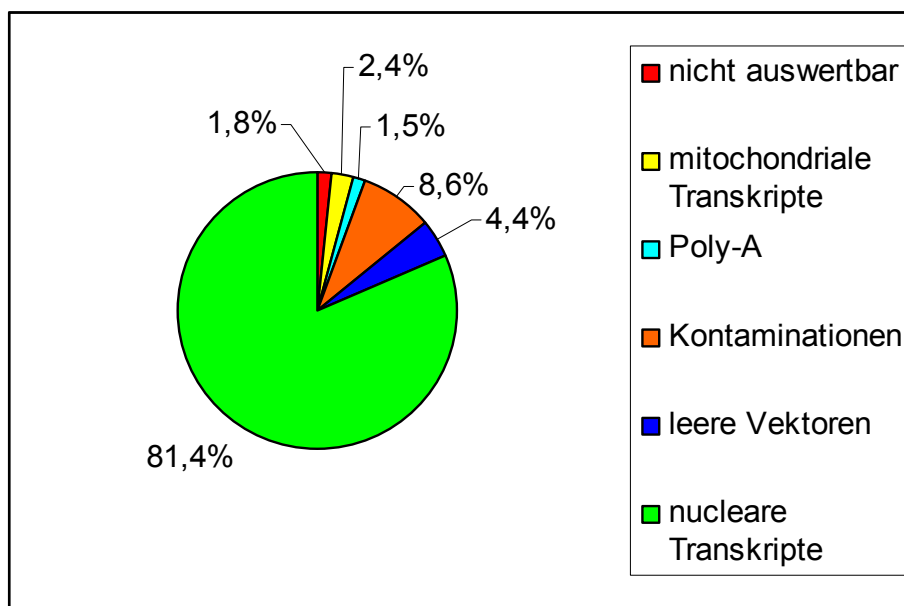


Abbildung 8: Verteilungsdiagramm der 960 analysierten cDNA-Klone.

Die Auswertung der übrigen 781 cDNA-Klone erbrachte eine Zahl von 466 verschiedener, nicht redundanter Transkriptspezies. Redundante Sequenzen (identische, bzw. überlappende Sequenzen) wurden in der weiteren Auswertung nicht berücksichtigt, da das Projekt auf die Erfassung der Transkript-Komplexität, bzw. der Identifikation neuer, unbekannter Transkripte abzielte.

Die 466 verschiedenen, nicht-redundanten cDNAs wurden bzgl. der abzuleitenden Funktion der Genprodukte eingeteilt (Gen-Ontologie), entweder durch Analyse der zugeordneten

LocusLink-Daten oder mittels einer Ableitung aus dem Namen des Gens (z.B. Glycerinphosphatdehydrogenase).

Hierbei wurde aus Gründen der Übersichtlichkeit pro EST nur je eine Definitions-Ontologie zugeordnet. Diese singuläre Klassifizierung war dann problematisch, sobald es mehr als ein Attribut gab oder verschiedene Ontologieebenen (z.B. EC-Klassifikation neben tatsächlichen Funktionsdefinitionen) vorlagen. Deshalb wurde eine Informationsdichtehierarchie bestimmt (so ist die Bezeichnung „Transkriptionsfaktor“ informativer als „DNA-bindendes Protein“) und der informativste Begriff als Primärfunktion gewählt.

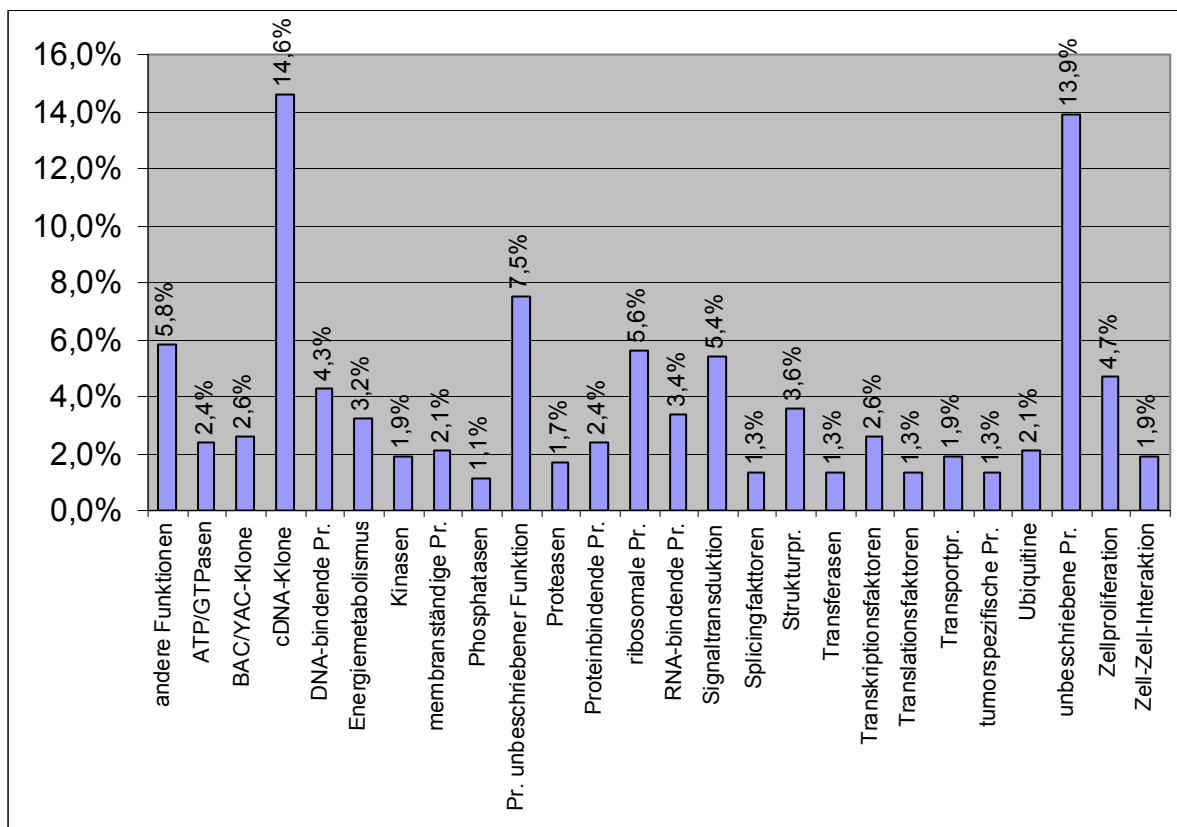


Abbildung 9: Verteilungsdiagramm der 466 informativen Klone nach Ontologien. Pr. = Protein.

Unter „andere Funktionen“ wurden Ontologien subsumiert, die grundsätzlich bestimmbar waren, die aber weniger als fünfmal auftraten, wie etwa Lyasen oder Peroxidasen. Als ESTs für „unbeschriebene Proteine“ wurden prädizierte oder herkömmlich kaum annotierte Eiweiße wie z.B. KIAA- oder LOC-Proteine in Abgrenzung zu „Proteinen unbeschriebener Funktion“ klassifiziert, die namentlich ausgezeichnet waren, bei denen aber eine Einteilung in die die ontologischen Gruppen nicht möglich war.

Insgesamt 119 der 466 verschiedenen, nicht-redundanten Klone (25,5%) wiesen eine starke Ähnlichkeit zu BAC/YAC- und cDNA-Klonen, sowie zu unbeschriebenen Proteinen auf,

waren also nicht vollständig beschrieben und stellen somit prinzipiell Kandidatengene für unbekannte vererbliche Krankheiten des Auges dar. Da Homologien zu bereits beschriebenen Genen auf der Proteinebene möglicherweise stärker konserviert worden wären, wurden diese Klone einer blastx-Homologiesuche gegen bekannte Proteine unterzogen. Diese war jedoch relativ unergiebig und wurde durch eine tblastx-Analyse ersetzt. Bei 72 Klonen war keine signifikante Proteinhomologie nachweisbar. 25 Klone wiesen Ähnlichkeiten zu humanen Proteinen oder translatierten cDNA-Klonen auf und die restlichen 22 zu Proteinen oder ESTs anderer Spezies.

Um vollständigere Sequenzen zu erhalten, wurden einige der 119 Klone mit *reverse*-Primern sequenziert. Weiterhin diente diese Analyse zu einer der Chimärenkontrolle, zum anderen hätte es bei längeren Inserts dazu kommen können, dass keine Homologie hätte gefunden werden können, weil zum Zeitpunkt der Analyse lediglich das 3`-Ende des entsprechenden Gens bekannt war. Dieser Untersuchungsschritt war nicht sonderlich erfolgreich, da einerseits über eine Reihe von Adenosinresten sequenziert werden musste und die Sequenzierqualität aus verschiedenen Gründen (z.B. *Slippage*, Verbrauch des ATP im Sequenziermix) stark abnahm, zum anderen konnte nur zwei abweichende Homologieergebnisse bestimmt werden, die aber zu keinem bereits bekannten Protein, sondern nur zu einem abweichenden cDNA-Klon führten.

Abschließend wurden die cDNA- und BAC/YAC-Klone mittels des VecScreen-Programms des NCBI auf nicht von der Lasergene-Software erkannte Vektorreste und repetitive Sequenzen getestet. Letztere deuten u.U. auf ein Mispriming des Oligo-dT-Primers bei der Erststrangsynthese der cDNA und somit auf eine nichtcodierende Sequenz hin. Von diesen zeigten lediglich 11 Klone signifikante Spuren von repetitiven Sequenzen.

4.2 Einfluss hypoxischer Bedingungen auf WERI Rb1-Zellen

4.2.1 Auswahl der Real Time-Primer

Der Effekt lang anhaltender hypoxischer Bedingungen auf retinales Gewebe, wie er z.B. durch einen Schlaganfall hervorgerufen wird, könnte durch Mikroarray-Analysen am Modell der Retinoblastom-Zelllinie WERI Rb1 bestimmt werden. In einem Vorversuch wurden mögliche Expressionsunterschiede ausgewählter Gene durch quantitative Real Time-PCR-Experimente (Q-PCR) untersucht.

Die Auswahl sollte einen Standard umfassen, dessen Expression nicht vom Sauerstoffniveau beeinflusst wird und solche Gene, die stark auf sinkende Sauerstoffpartialdrücke reagieren. Als Standard wurde die 18S-rRNA ausgewählt (Razeghi *et al.*, 2003), als variabel exprimierte Gene die von *HIF-1* (Chun *et al.*, 2002) regulierten *VEGF* (*Vascular Endothelial Growth Factor*) (Vaupel, 2004), *EPO* (Erythropoietin) (Jelkmann, 2004), bzw. *GAPDH* (Glycerinaldehyd-3-Phosphat-Dehydrogenase) (Semenza, 2000). Die vom Computerprogramm Primer Express v. 2.0 (3.2.6) vorgeschlagenen Primerpaare für die entsprechenden Gene wurden manuell überprüft und gegebenenfalls übernommen.

4.2.2 Anzucht unter hypoxischen und normoxischen Bedingungen

Um die Auswirkungen einer Ischämie, z.B. durch einen Schlaganfall zu simulieren, wurde der Sauerstoffpartialdruck im hypoxischen Versuchsansatz auf 1% gesenkt und 8, bzw. 24 Stunden inkubiert. Parallel dazu wurde je eine Anzuchtschale unter normoxischen Bedingungen als Vergleichswert abgeerntet. Die Durchführung erfolgte in Hypoxiekammern der Abteilung Probst des Physikalisch-chemischen Instituts Tübingen (3.2.1). Die Aufreinigung der Proben erfolgte zunächst nach dem Trizol-Protokoll (3.1.4) (Daten nicht gezeigt). Da hier die Real Time-Ergebnisse qualitativ schlechter ausfielen (keine glatten Signalkurven, stärkere Streuung der C_T-Werte um den Mittelwert), wurde die RNA-Extraktion später mit dem RNeasy-System (3.2.3) durchgeführt.

4.2.3 Auswertung der Real Time-Ergebnisse

Zum Test etwaiger Auswirkungen der Hypoxie auf die Regulation der betreffenden Gene wurden die RNA-Proben nach einer RNeasy-Extraktion (3.2.3) durch eine Q-PCR (3.2.6) quantifiziert. Die gewonnenen Daten wurden mit Hilfe der ABI Prism 7000 SDS Software ausgewertet (Abb. 10). Für die vier getesteten Konditionen (8 h, bzw. 24 h unter hypoxischen, bzw. normoxischen Bedingungen) wurden dreifache Messungen durchgeführt, aus denen ein Mittelwert gebildet wurde.

Die Quantifizierung der Matrizenmenge erfolgte hierbei nicht über die Messung des Endprodukts, sondern mittels der Bestimmung des C_T -Wertes, d.h. der Zyklenzahl, bei der das Fluoreszenzsignal des gebildeten PCR-Produkts den Detektionsschwellenwert überschreitet. Der C_T -Wert verhält sich hierbei umgekehrt proportional zum Logarithmus der Ausgangsmenge des Transkripts (Higuchi *et al.*, 1993).

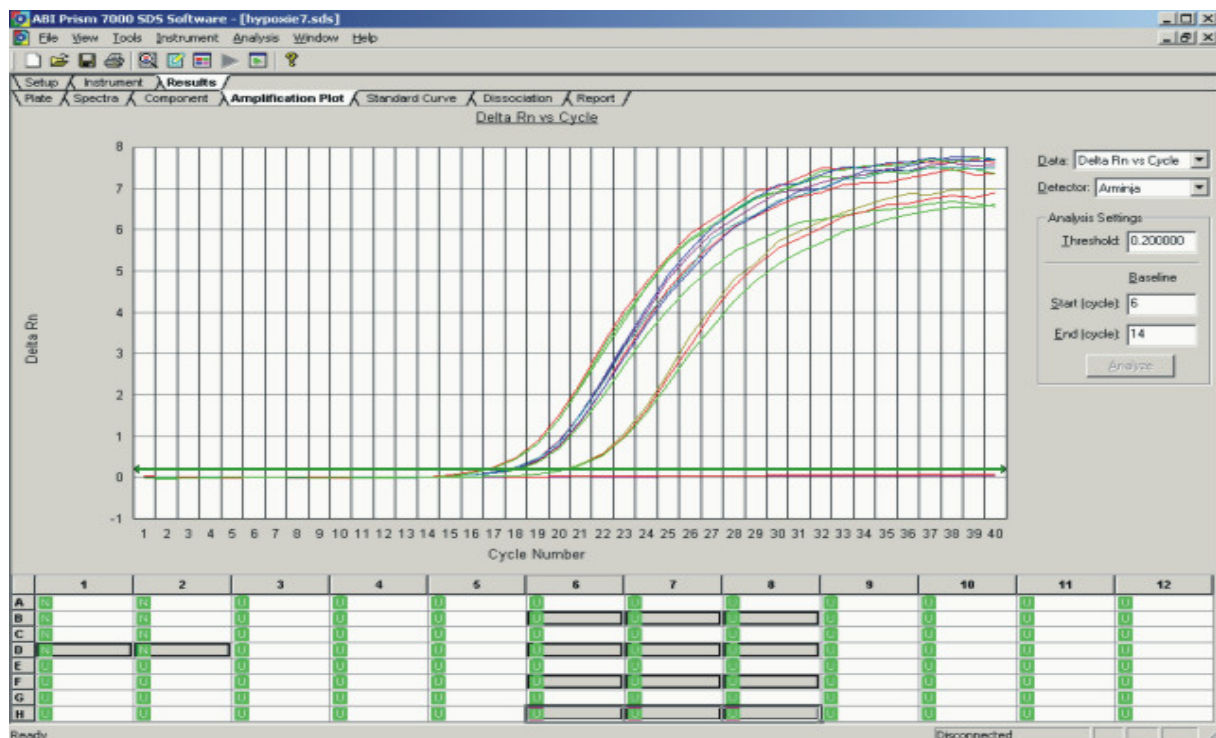


Abbildung 10: Auswertung der Realtime-PCR durch die AbiPrism 7000 SDS Software.

Da die Vermehrungsrate einer PCR-Reaktion jedoch von vielen Faktoren abhängt, kann nicht direkt auf die initiale Matrizenmenge geschlossen werden (Bustin, 2000). Ein indirektes Quantifizierungsverfahren stellt die $\Delta\Delta C_T$ -Methode dar, welche die unbekannte Menge eines Transkripts A mit der bekannten Menge eines Transkripts B vergleicht (Livak und

Ergebnisse

Schnittgen, 2001). In unserem Fall wurden die Hypoxie-induzierbaren Gene *GAPDH*, *EPO* und *VEGF* mit dem inerten Standard 18S rRNA aus derselben Probe verglichen.

Messung	C _t -Wert Probe 1	C _t -Wert Probe 2	C _t -Wert Probe 3	Mittelwert C _t
18S rRNA – 8h Hypoxie	8,55	8,75	8,66	8,653
18S rRNA – 24h Hypoxie	8,93	9,00	9,15	9,027
18S rRNA – 8h Normoxie	Ausfall	9,09	9,21	9,150
18S rRNA – 24h Normoxie	8,09	8,16	Ausfall	8,125
EPO – 8h Hypoxie	27,10	27,09	26,97	27,053
EPO – 24h Hypoxie	26,59	26,65	26,50	26,580
EPO – 8h Normoxie	26,74	26,50	26,55	26,597
EPO – 24h Normoxie	26,52	26,42	26,45	26,463
GAPDH – 8h Hypoxie	11,56	11,29	11,30	11,383
GAPDH – 24h Hypoxie	10,41	10,15	10,54	10,367
GAPDH – 8h Normoxie	10,88	10,70	10,15	10,577
GAPDH – 24h Normoxie	12,38	12,34	12,23	12,317
VEGF – 8h Hypoxie	17,92	18,07	18,00	17,997
VEGF – 24h Hypoxie	16,68	16,68	16,82	16,727
VEGF – 8h Normoxie	17,65	17,72	17,71	17,693
VEGF – 24h Normoxie	20,27	20,35	20,38	20,333

	ΔC_t EPO-18S	ΔC_t GAPDH-18S	ΔC_t VEGF-18S
8 h Hypoxie	18,400	2,730	9,344
24 h Hypoxie	17,553	1,340	7,700
8 h Normoxie	17,443	1,427	8,543
24 h Normoxie	18,338	4,192	12,208

Ergebnisse

Zwei Proben der 18S rRNA-Kontrolle wichen aus unbekanntem Gründen um mehr als einen Zyklus vom C_T -Mittelwert ab und wurden als Ausfall gewertet; die anderen Werte divergierten nur wenig. Bei der Auswertung waren folgende Ergebnisse auffällig: Nach 24 Stunden Hypoxieexposition wie auch nach 8 Stunden unter Normbedingungen wurde etwa die halbe Menge an 18S rRNA-Transkript detektiert wie nach 24 Stunden unter normalem Sauerstoffpartialdruck. Während der 18S rRNA-Ansatz als gleichbleibend stark exprimierter Standard vorgesehen war, waren diese Unterschiede zu klein, um signifikant zu sein (Dr. Christoph Linnemann, Neurologisches Institut Tübingen, pers. Mitteilung). Die C_T -Werte von Erythropoietin blieben innerhalb einer Schwankungsbreite von einem Zyklus, die Transkription von *EPO* unterlag somit keiner signifikanten Regulation. Die *GAPDH*-Expression war unter 24 h Hypoxie und 8 h Normoxie-Einwirkung vergleichbar hoch, sank jedoch im Vergleich dazu nach acht Stunden bei vermindertem Sauerstoffpartialdruck um mehr als Faktor zwei und verringerte sich nach 24 h Normbedingungen um fast Faktor acht. Ausgehend von 8 h Normoxie sank die Expression von VEGF nach weiteren 16 h Normbedingungen um etwa Faktor 12,5, wohingegen die Transkriptmengen unter Hypoxieeinfluß ebenfalls unerwartet nahezu unverändert blieben.

Aufgrund dieser den Erwartungen widersprechenden Ergebnisse unterblieb die Überführung dieses Versuchsaufbaus in ein Expressionschip-Experiment.

4.3 Erstellung einer vollständig annotierten humanen retinalen in vitro-Datenbank

4.3.1 Verlaufsdiagramm der Datenbankerstellung

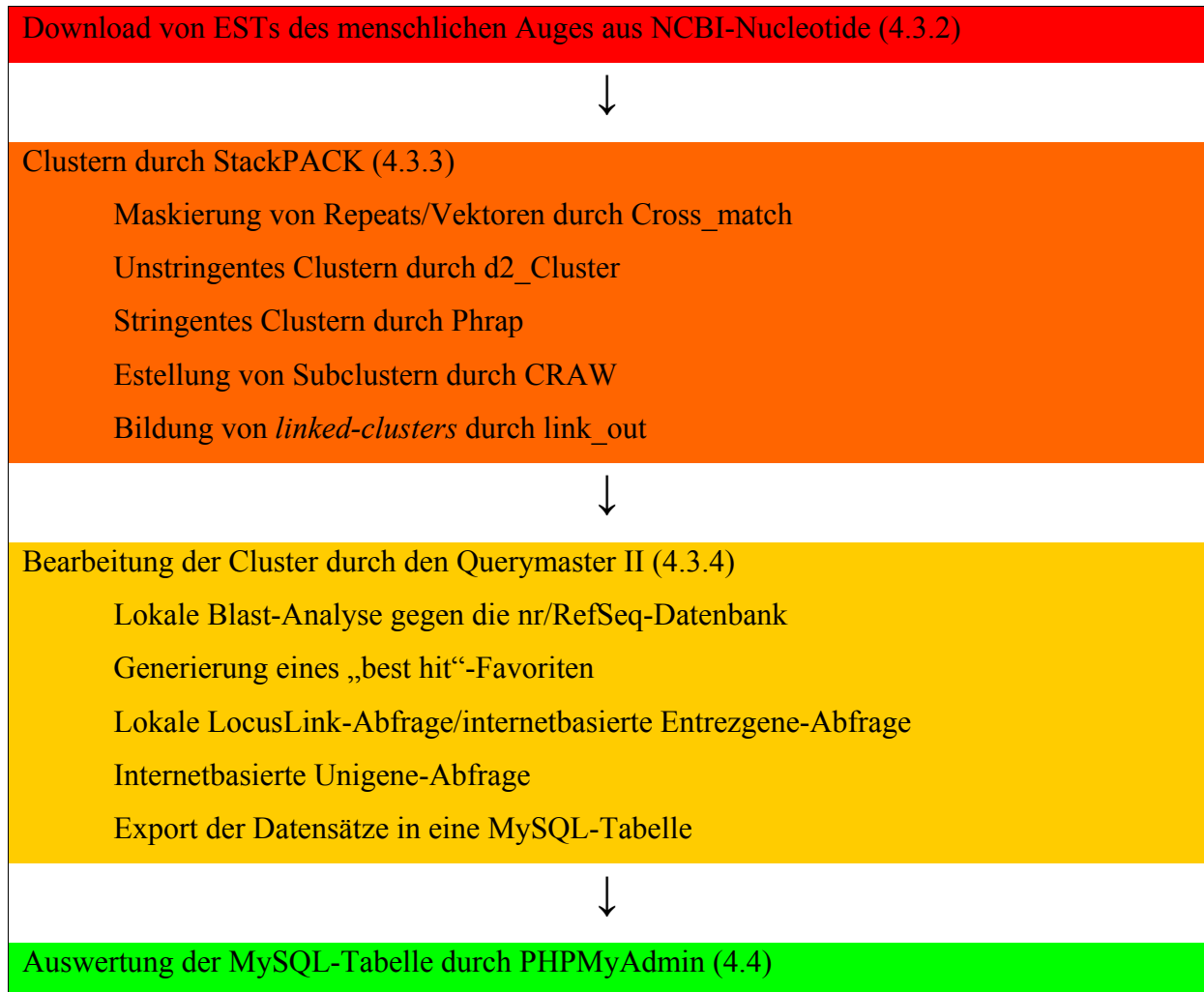


Abbildung 11: Flussdiagramm der einzelnen Arbeitsschritte zur Erstellung und Auswertung der Datenbanken TRJ8 und TRV1.

4.3.2 Auswahl der eingesetzten Sequenzen

Die selbst generierte und durch Sequenzanalyse dargestellte cDNA-Bibliothek (4.1) umfasste nur einen Bruchteil des beschriebenen humanen Retinoms. Zur Identifizierung und Katalogisierung möglichst aller ESTs der Netzhaut sollten ergänzend entsprechende externe Datenbanken ausgewertet werden. Hierzu wurden als erster Schritt die bekannten humanen retinalen ESTs vom Entrez Nucleotide-Server (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=search&DB=nucleotide>) als Stapeldatei heruntergeladen. Um ein

Clustern nicht überlappender, zu einem Klon gehöriger Sequenzen durch Link_out (3.3.4.7) zu ermöglichen, wurde das Genebank-Speicherformat gewählt. Es wurden folgende Filter verwendet:

```
homo [organism] AND (retina OR eye OR "pineal gland") NOT (contig OR bac OR yac OR pac OR "complete sequence" OR "contain*") AND (mrna OR est OR cdna)
```

„Eye“ und „pineal gland“ wurden als Alternativparameter zu „retina“ verwendet, da die Netzhaut ein Bestandteil des Auges ist und in vielen Studien keine Subdifferenzierung vorgenommen wurde, bzw. möglich war. Die Zirbeldrüse wiederum ist genetisch eng verwandt mit der Netzhaut und exprimiert bekanntermaßen viele Gene, die sonst lediglich in Photorezeptoren exprimiert werden (Vigh *et al.*, 2002).

Die ausschließenden Bedingungen wurden hinzugenommen, um das Abspeichern größerer genomischer Sequenzbereiche (Chromosome, BACs, etc.) zu verhindern, bei denen es sich nicht selbst um ESTs handelte, die aber Gene enthielten. Ein Clustern mit diesen Kontaminationen hätte nicht nur die Rechenzeit extrem erhöht – abhängig von der Qualität der verwendeten Skalierungsalgorithmen geht eine Zunahme der abzugleichenden Basenpaarmenge mit einer mehr oder minder gedämpften Quadrierung der benötigten Zeit einher – sondern auch die Qualität der Cluster hätte sich aufgrund der Bildung uninformativer Contigs dramatisch verschlechtert.

Für den ersten Versuchslauf testretjens8 (TRJ8) wurden 180.785 humane retinale ESTs als Batch heruntergeladen. Der zweite Ansatz testretvalerial (TRV1) umfasste knapp ein Jahr später bei gleichem Filter bereits 194.935 Sequenzen. Die Batchdateien waren unkomprimiert ca. 640, bzw. 680 Megabyte groß.

4.3.3 Clustern der ESTs/cDNAs mit dem StackPACK-Programm

Bei der Importierung in die StackPACK-Programmpipeline erfolgte ein Herausfiltern namentlicher Dubletten und Proteinsequenzen durch Auslesen der Kopfzeile, bzw. Abgleichen der Sequenz mit dem IUPAC-Code für Nucleotide, was insgesamt zu einer Eliminierung von ca. einem Promill der Datensätze führte (3.3.4.2). Nach Maskierung

störender Sequenzelemente wie Vektoren, Repeats, etc. (3.3.4.3) wurden die Datensätze anschließend durch das Modul D2_cluster (3.3.4.4) grob in Cluster (Cl) vorsortiert, welche dann durch das Phrap-Programm (3.3.4.5) bei signifikanten Sequenzunterschieden in Contigs (Ct) aufgeschlüsselt wurden. Jede singuläre oder zu einem Contig zusammengefasste Sequenz lieferte nach der finalen Assemblierung durch CRAW (3.3.4.6) eine Consensus-Sequenz (Cn). Durch das Link_out-Unterprogramm (3.3.4.7) wurden ggf. nichtüberlappende, aber zum selben Klon gehörige Cluster zu Superclustern (Ln) zusammengefasst.

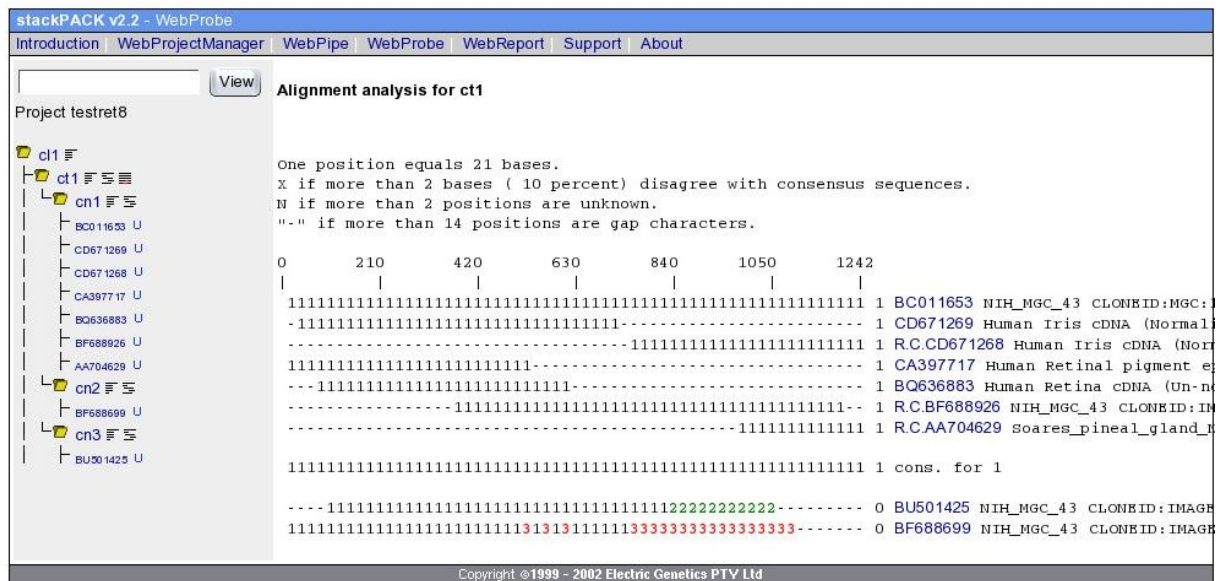


Abbildung 12: Graphische Darstellung eines stackPACK-Sequenzclusters.

Von den 180.785 Ausgangs-Sequenzen des TRJ8-Ansatzes entfielen 32.526 auf singuläre Sequenzen und 12.969 auf Multisequenz-Cluster (MSCs), die sich wiederum aufspalteten in 11.783 unabhängige MSCs und 1.186 MSCs, die durch Link_out in 556 Supercluster zusammengefasst wurden. Somit belief sich der Clusterungsfaktor (nicht-singuläre Ausgangssequenzen zu finalen MSCs) auf ca. 11,4 zu eins.

Der aus 194.935 Sequenzen bestehende TRV1-Ansatz spaltete sich auf in 33.879 Singletons, 13.669 ungelinkte MCSs, mit 12.528 unverbundene Consensi und 1.141 MSCs in 523 Superclustern. Der Clusterungsfaktor betrug hier durchschnittlich 11,8.

Der gesamte Clusterungsprozess dauerte insgesamt vier, bzw. fünf Tage, wobei drei komplette Tage Rechenzeit auf den initialen Vereinigungsschritt durch D2_cluster (3.3.4.4)

entfielen. D2_cluster war so eingestellt worden, dass beide vorhandenen Prozessoren angesteuert wurden, d.h. dass sich die Rechenzeit halbierte.

Die durch StackPACK bearbeiteten Sequenzen wurden in je drei Stapeldateien exportiert: a) Singletons, b) die Consensi der unabhängigen MSCs und c) die Consensi der MSC-Supercluster. Diese Batchdateien wurden vom Querymaster II zur weiteren Analyse importiert.

4.3.4 Entwicklung des Querymaster II

Der Querymaster I (3.3.5.1) war lediglich in der Lage, gestapelte Blastanfragen in mittlerer Geschwindigkeit über das Internet zu vermitteln. Um die automatisierte Generierung einer vollständig annotierten humanen retinalen Datenbank zu ermöglichen, musste das Leistungsspektrum erweitert werden:

Der Querymaster II (3.3.5.2) wurde entworfen, um die von StackPACK generierten Cluster durch eine Pipeline der vom NCBI bereitgestellten Einzelmodule blastall, unigene sowie locuslink, bzw. entrezgene zu schleusen. Die gesammelten Einzelergebnisse sollten in einer relationalen Datenbank vereinigt und ausgewertet werden.

4.3.4.1 Auswahl der Programmiersprache für den Querymaster II

Als erster Schritt der Softwareentwicklung musste die zu verwendende Programmiersprache festgelegt werden. Maschinensprachen wie C⁺⁺ werden durch Computer erheblich schneller interpretiert als Bibliothekssprachen wie Python (3.3.2.2) oder Tcl/Tk (3.3.2.2) und finden für rechenintensive Arbeitsschritte Anwendung (Wang, 2000). Bibliothekssprachen benötigen hingegen u. U. nur ein Zehntel des Programmcodes, was zu entsprechend verkürzten Entwicklungszeiten und –kosten führt (Webster und Francis, 2000). Da der Querymaster II (QM II) selbst keine sonderlich rechenintensiven Schritte bearbeiten sollte, wurde prinzipiell eine der Bibliothekssprachen erwogen. Tcl/Tk gilt als „glueing language“, seine besonderen Vorteile liegen in der effizienten Verknüpfung anderssprachiger Module (Ousterhout, 1995; Michael Schlenker, Institute for Science Networking Oldenburg, persönliche Mitteilung). Dies entsprach exakt dem Anforderungsprofil, weswegen Tcl/Tk verwandt wurde.

4.3.4.2 Blast-Analyse der geclusterten Sequenzen durch den Querymaster II

Die Verarbeitung geclusterter EST/cDNA-Sequenzen erfolgte über Homologieabgleiche mit bekannten Sequenzen mittels der Blast-Analyse (Altschul *et al.*, 1990). Der QM II steuerte hierbei ein lokal installiertes blastall-Modul über eine graphische Schnittstelle an, die Ergebnisausgabe der aus der Stapeldatei ausgelesenen Abfragesequenzen erfolgte in Form durchnummerierter einzelner XML-Dateien (z.B. 1000001.xml). Der XML-Code ist ein standardisiertes Exportformat, das leicht portabel und auszulesen ist (Behme und Mintert, 2000). Die vorgegebenen Blastparameter (Filtersätze, Prozessoranzahl, etc.) wurden in der .querymaster2rc-Datei abgespeichert und könnten dort mit einem Texteditor gegebenenfalls angepaßt werden. Einzelne Variationen eines Experiments können über die graphische Oberfläche vorgenommen werden (Abb. 13) Zwecks Debugging und Qualitätskontrolle wurden der Initialisierungsstatus und die Ergebnisausgabe nicht nur in einer Log-Datei gespeichert, sondern wurden auch simultan in der Konsole (Abb. 13) ausgegeben. Der Umfang des Loggings wurde nach Fertigstellung des Programmteils zur Steigerung der Leistungsfähigkeit minimiert. Abgesehen von der Verwendung beider Computerprozessoren wurde das blastn-Programm unter Standardparametern betrieben. Das formatdb-Programm diente zur Formatierung der angesteuerten NCBI-Datenbanken.

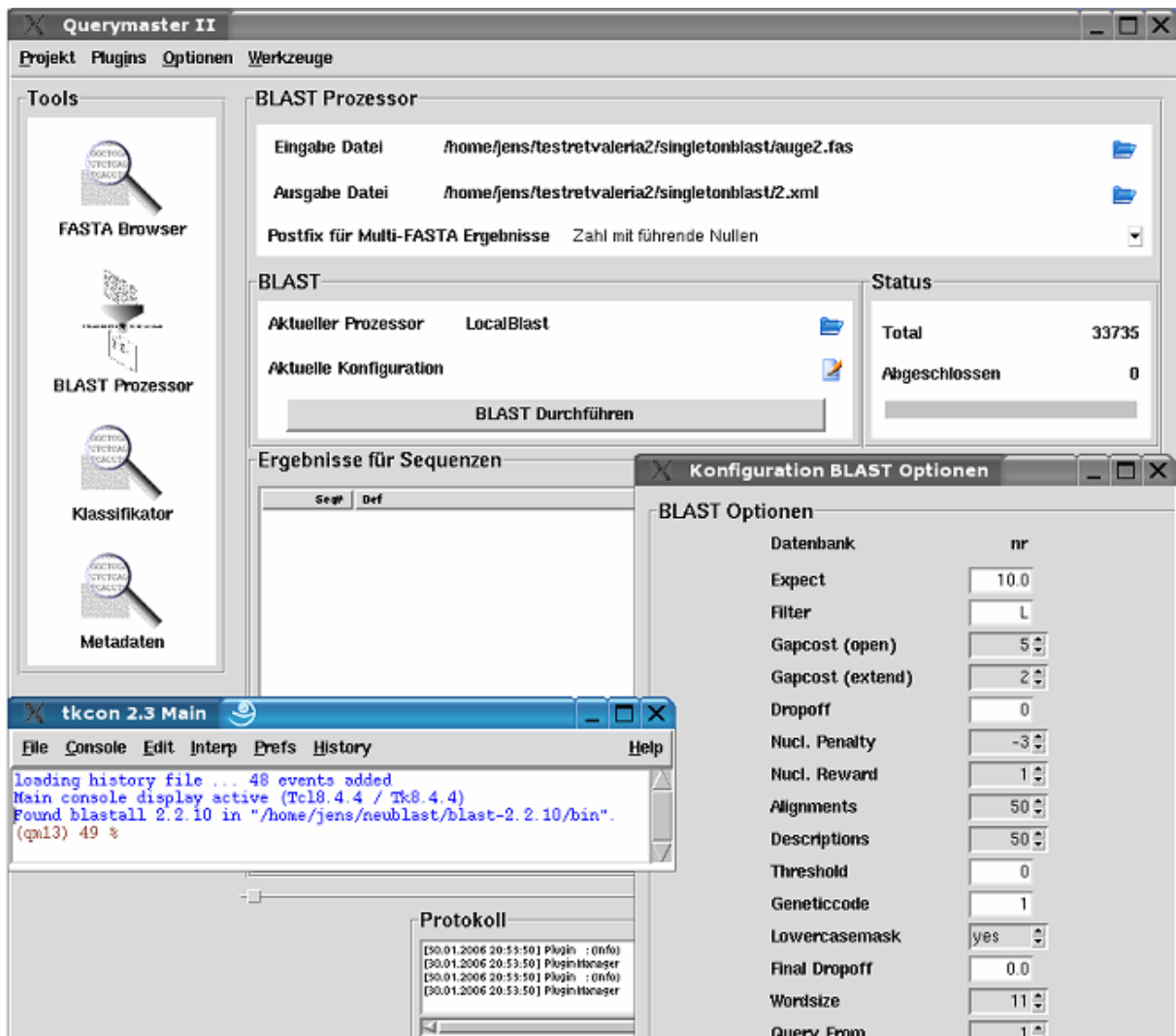


Abbildung 13: Graphisches Frontend des QM II während der Initialisierung einer Blast-Analyse.

Um initial einen möglichst umfassenden Überblick über das Transkriptom zu bekommen, wurde der TRJ8-Ansatz gegen die nr-Datenbank geblastet unter Verwendung der blastall-Version 6. Die nr-Bibliothek umfasst alle nicht redundanten GenBank-, EMBL-, DDBJ- und PDB-Sequenzen. Durch einen Programmierfehler in der blastall-Software wurden die erstellten Ergebnisdateien nach erfolgtem Kopieren in den Zielordner nicht wieder aus dem virtuellen Speicher gelöscht. Es folgten abhängig von der Größe der jeweiligen Dateien nach durchschnittlich 2.000 Sequenzen Systemabstürze durch Speicherüberlauf. Hierbei mußten die abgearbeiteten Sequenzen manuell aus der Eingangsstapeldatei gelöscht und der Prozeß neu gestartet werden. Die bei gegebener Rechnerleistung theoretisch zu erreichende Geschwindigkeit lag bei etwa 1.000 Sequenzanfragen pro Tag und Prozessor. Insgesamt benötigte die Blast-Analyse der TRJ8-Dateien 29 Tage. Ohne den Programmfehler hätte dieser Prozeßabschnitt etwa 20 Tagen in Anspruch genommen.

Es wurde daher darauf verzichtet, die Cluster mit der est-Datenbank des NCBI abzugleichen solange nur die fehlerhafte blastall-Version verfügbar war. Da die est-Datenbank mit 3,5 Gigabyte etwa doppelt so groß wie die nr-Bank war, hätte die zu erwartende Rechenzeit ohne Berücksichtigung unterschiedlicher Sequenzlängen bei ca. zwei Monaten gelegen. Weiterhin standen dem Rechner nach Abzug der Systemprozesse noch ca. 3,5 Gigabyte freies RAM zur Verfügung, daher hätte es eventuell zu einem Swappen (Datenauslesen von der Festplatte durch fehlenden virtuellen Speicher) kommen können, was bei den betroffenen Sequenzen zu einer tausendfachen Prozessverlangsamung hätte führen können.

Die TRV1-Dateien wurden mit der blastall-Version 10 bearbeitet, bei welcher der erwähnte Fehler beseitigt worden war. Weiterhin wurde nicht die nr-Datenbank angesteuert, sondern lediglich die humanen Referenzsequenzen des NCBI (RefSeq-Datenbank). Die Blast-Analyse Blasten wurde erfolgreich über Nacht abgeschlossen. Die Geschwindigkeit des Blast-Prozesses hängt neben unterschiedlich effizient programmieren Softwareversionen hauptsächlich von der Länge und Menge der eingegebenen sowie der abzugleichenden Sequenzen ab. Da die RefSeq-Datenbank bei vergleichbaren Abfragesequenzen um zwei Zehnerpotenzen kleiner als ihr nr-Pendant war, war die Analyse entsprechend schneller beendet.

4.3.4.3 Generieren eines „best hits“ aus einer Blast-Ergebnisliste

Das blastall-Programm bewertet die Homologie zwischen einer Eingangssequenz und aller in einer jeweiligen Zieldatenbank gelisteten Vergleichssequenzen (Altschul *et al.*, 1990). Unter Verwendung der Standardparameter wird eine Liste von 50 Referenzsequenzen („hits“) ausgegeben. Die Reihenfolge in der Liste des Blastergebnisses wird durch den Homologiewert (Hit-Score) bestimmt, wobei die der Eingangssequenz ähnlichste Vergleichssequenz an Platz eins („best hit“) liegt.

Bei der manuellen Auswertung der Blastresultate der lokal erstellten cDNA-Bibliothek (4.1) fiel auf auf, dass der von Blast generierte „best hit“ nicht zwangsläufig dem tatsächlich informativsten Resultat entsprach, z.B. wenn ein BAC-Klon einen längeren Match gab als eine Sequenz für ein prädiktiertes Protein.

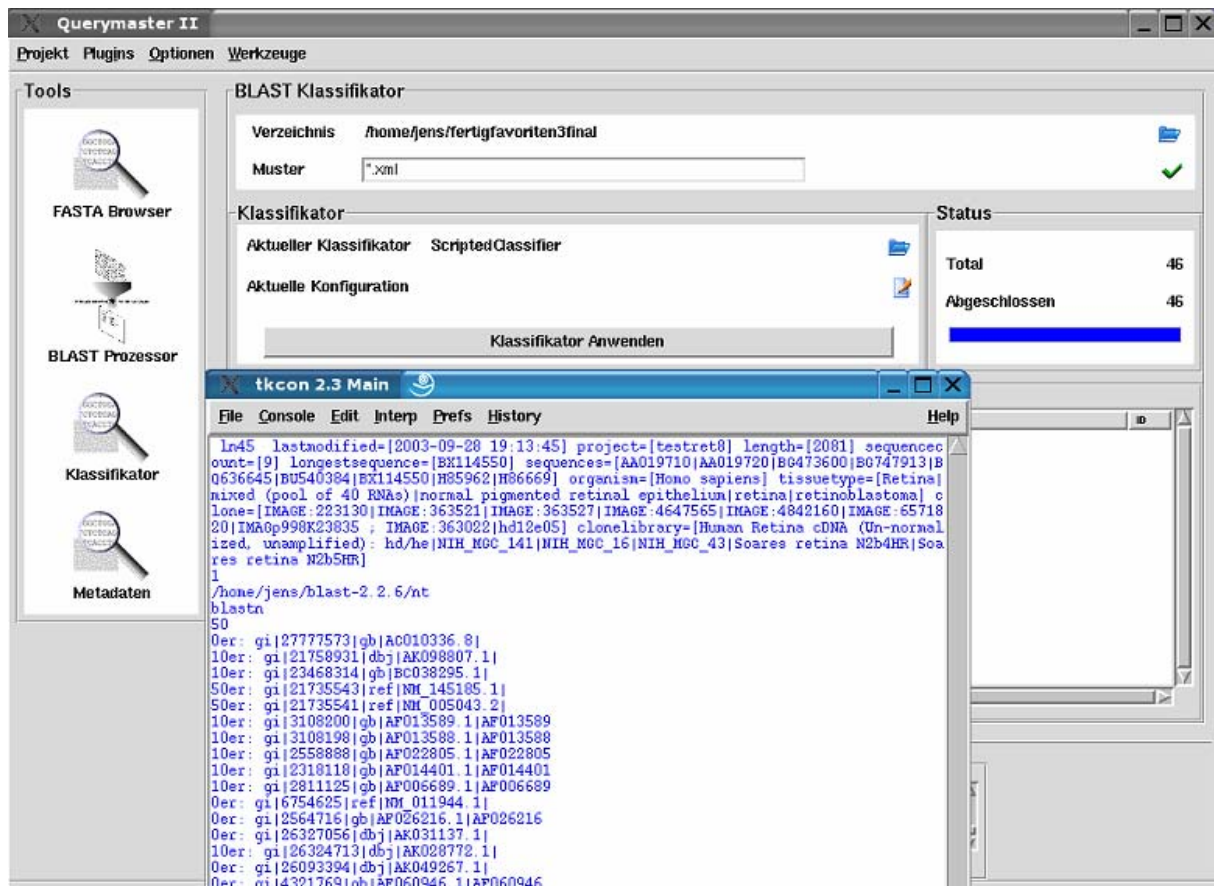


Abbildung 14: Graphisches Frontend des QM II nach der Bestimmung des Favoriten durch den BLAST-Klassifikator. Die graphische Ausgabe der Bonuspunkte jeder Query-Sequenz erfolgte temporär in der tkcon-Konsole.

Daher wurde ein Favoriten-Generierungsscript entwickelt, um nach individuellen Parametern die Suche nach dem informativsten Blast-Hit zu verfeinern (3.3.5.2c). Das Klassifikatorscript wurde wie die .querymaster2rc-Datei zur Blast-Ansteuerung aus dem gewrappten (3.3.3.1) und damit unzugänglichen QM II-Programm ausgelagert, um durch Anpassung der Booleschen Operatoren mittels eines Texteditors möglichst flexibel agieren zu können. Dies entspricht bei umfangreichen Programmen (z.B. StackPACK) einer üblichen Vorgehensweise. Nach Aktivierung dieses Moduls wurde jedem „hit“ der Blastresultat-Liste ein Bonuswert zugeordnet (Abb. 14), der humanen Referenzsequenzen den Vorzug gibt. Die Ergebnisse wurden in einer *.xml.fav-Datei gespeichert. Lediglich der durch den Klassifikator ausgewählte Favorit wurde zur Ansteuerung sekundärer Datenbanken wie Entrezgene (4.3.4.4) benutzt, da die nachgeschaltete Analyse vieler Homologe zu einer multiplizierten Rechenzeit, Speicherplatzverknappung und hoher Komplexität und Unübersichtlichkeit der Daten geführt hätte.

Der TRJ8-Versuchslauf wurde mit den in 3.3.5.2c aufgeführten Parametern durchgeführt, wobei die zu bewertende Ergebnisdatei aus den 50 ähnlichsten Sequenzen bestand. Da die ref-Datenbank nur aus humanen Referenzsequenzen bestand, verblieb als einziges Differenzierungsmerkmal nur der Hit-Score. Um Systemressourcen zu sparen wurde somit beim TRV1-Versuch die Menge der möglichen Resultate im Blast-Filter auf 1 gesenkt.

4.3.4.4 Ansteuerung externer Datenbanken zur Annotierung der Blastresultate

Nicht alle für eine vollständige Funktionsbeschreibung eines Gens notwendigen Informationen sind im Genbank-Eintrag enthalten. Zu diesem Zweck ist es sinnvoll, Informationen aus nachgeschalteten Datenbanken hinzuzuziehen. Die Unigene-Datenbank bietet sich dafür an, welche Daten zur Expression, Homologien und chromosomalen Lokalisierung transkribierter Sequenzen liefert (Boguski und Schuler, 1995; Pontius *et al.*, 2003). Die zweite in dieser Arbeit herangezogene Datenbank des NCBI war LocusLink (Wheeler *et al.*, 2000), welches zwischenzeitlich von EntrezGene (Maglott *et al.*, 2005) abgelöst wurde. In LocusLink, bzw. EntrezGene finden sich unter anderem Daten zu genetischen Ontologien (Camon *et al.*, 2003), textbasierte Zusammenfassungen wichtiger Informationen zum Gen, sowie der Name und Synonyme des Gens. Die Ansteuerung beider Datenbanken erfolgte durch Auslesen der *accession-number* des Favoriten aus der zugehörigen XML-Blastdatei (3.3.5.2d). Auf das Einbinden weiterer frei zugänglicher Datenbanken wie z.B. Ensembl wurde aus terminlichen Gründen und wegen geringer zusätzlicher Informationsdichte verzichtet. Weiterhin wurden nicht alle Informationen aus den angesteuerten Datenbanken eingelesen, zum Teil aufgrund redundanter Daten (chromosomale Lokalisierung), wegen geringer Relevanz für das vorliegende Projekt (Homologien) oder wegen höherer Speicherplatzanforderungen in der MySQL-Tabelle (graphische Abbildung des betreffenden Genes). Da im Bedarfsfall individuelle Gene leicht direkt durch Eingabe des vorliegenden *Accession*-Codes in die entsprechenden Datenbank-Oberfläche weiteranalysiert werden könnten, erschien das Auslesen aller möglichen Daten in die MySQL-Tabelle auch aus diesem Grund verzichtbar.

Zu Beginn der Arbeiten wurde das LocusLink-Modul lokal angesteuert, was sich programmiertechnisch u.a. wegen mangelhafter Dokumentation des NCBI und unausgereifter Pipeline-Konnektivität als äußerst anspruchsvoll erwies. Für das Unigene-Modul wurde

hingegen das Web-Interface des NCBI benutzt (3.3.5.2d). Besaß der „best hit“ eine Unigene-Codierung, wurde die Ergebnisseite als „HTML-Screenshot“ heruntergeladen und in eine XML-Datei überführt. Während der TRJ8-Ansatz mit dem lokalen LocusLink-Modul analysiert wurde, wurde für den TRV1-Datensatz auf ein ebenfalls internetbasiertes EntrezGene-Abfragemodul zurückgegriffen. Pro Datenbankanfrage wurde ein Zeitbedarf von durchschnittlich 8 Sekunden ermittelt. Der duale Prozessor des Linuxrechners ermöglichte ein paralleles Ansteuern beider NCBI-Datenbanken, so dass dieser Bearbeitungsschritt in etwa einer halben Woche abgeschlossen werden konnte. Theoretisch wäre es durch Hyperthreading möglich gewesen, mehr als zwanzig simultane QM II-Webanfragen zu initiieren, da kaum limitierende Prozessorlasten erzeugt wurden. Obwohl die verwendete „Screenshot“-Technologie keine große Serverlast beim NCBI erzeugen sollte, wurde ein Interface zur Vorgabe eines Verzögerungswertes zwischen zwei Abfragen eingerichtet. Standardmäßig wurde hier ein Wert von 500 Millisekunden eingegeben.

4.3.4.5 Abspeicherung der kombinierten Resultate in eine relationale MySQL-Datenbank

Parallel zur Entwicklung des QM II musste die Frage beantwortet werden, welcher Datenbanktyp Verwendung finden sollte, was u.a. auch das Speicherformat der Zwischenergebnisse (z.B. XML) determinierte. Prinzipiell sollte eine relationale Datenbank (3.3.7) zur Anwendung kommen, um verknüpfte Abfragen möglich zu machen. Das für die lokale cDNA-Bibliothek verwendete ACCESS schied aus mehreren Gründen aus. Verknüpfte Abfragen sind hier nur sehr eingeschränkt möglich und der Speicherplatz ist systembedingt für einzelne Datensätze auf 64 Kilobyte beschränkt, was das Abspeichern langer Sequenzen unmöglich macht. Weiterhin ist ACCESS nicht in der Lage, die zu erwartenden umfangreichen Datenmengen dieses Projektteils effizient verwalten zu können, was zu langen Zugriffszeiten, bzw. auch zu Systemabstürzen hätte führen können. ORACLE-Datenbanken hingegen waren für unseren Ansatz und die vorliegenden Ressourcen zu überdimensioniert (Priv. Doz. Jochen Scheel, MPI für Entwicklungsgenetik, Tübingen; pers. Mitteilung). Für andere Datenbanksysteme wie INFORMIX, DB2 oder dBASE konnten aufgrund mangelnder Verbreitung keine Erfahrungswerte erfragt werden. Somit schieden sie ebenfalls aus.

Es verblieben noch zwei potentielle Datenbanksysteme: MySQL und PostgreSQL. Beide basieren auf dem relationalen SQL (3.3.7). MySQL und PostgreSQL unterscheiden sich

hauptsächlich darin, dass MySQL zwar schneller ist, aber in Versionen unter 5.0 keine Transaktionen unterstützt (Dr. Kay Nieselt, Zentrum für Bioinformatik, Tübingen, pers. Mitteilung). Transaktionen sind Datentransfers, bei denen zwingend gewährleistet ist, dass zwei Prozesse verbunden bleiben, z.B. eine Sollbuchung bei der Geldentnahme am Bankautomaten. Diese zusätzliche Sicherheit war bei gegebener Systemstabilität weniger wichtig als der Geschwindigkeitsvorteil, zumal zu Beginn des Projektes die mögliche Berechnungsdauer einer Abfrage an die relationale Datenbank unklar war. MySQL war wegen der Verwendung des StackPACK-Programmes ohnehin installiert worden, weswegen sich ebenfalls ein Zurückgreifen auf diesen Datenbanktyp anbot.

Da StackPACK v.2.2 für MySQL-Versionen bis 3.23.27 optimiert war, wurde für die Datenbankanalyse ebenfalls die Version 3.23.27 verwendet.

Aus den vier Datenquellen Cluster-Stapeldatei, Blast-Resultat, Unigene- und LocusLink, bzw. Entrezgene wurden die definierten Inhalte ausgelesen und in vier MySQL-Untertabellen importiert.

Das NCBI verwendete für die XML-Codierung der EntrezGene-Ergebnisse ein vom Standard abweichendes, selbstdefiniertes Format (siehe 3.3.3.3), welches dafür sorgte, dass beim Import der Daten des TRV1-Ansatzes ca. 30% der Datensätze nicht in die EntrezGene-Untertabelle eingelesen werden konnten. Bei diesen Sequenzen führten u.a. erratisch gesetzte Determinanten für ein Zeilenende zu Problemen bei der Definition des auszulesenden Abschnitts. Aufgrund mangelhafter Dokumentation des NCBI konnte dieser Fehler nur durch das Zwischenschalten eines Reparaturmoduls behoben werden. Dieses definiert basierend auf einer statistischen Analyse der Datensätze einen Syntax-Kanon und korrigiert abweichende Einträge durch Überschreibung.

Der benötigte Speicherplatz lag für den TRJ8- und TRV1-Ansatz bei jeweils etwa 60 Gigabyte, wobei allein für die Blastresultate in XML-Codierung pro Sequenz durchschnittlich 100 KByte benötigt wurden. Vom Download der Sequenzen bis zum Import der Daten in die MySQL-Tabelle wurden bei der Annotierung des TRV1-Ansatz umgerechnet knapp zwei benötigt.

Die Analyse der MySQL-Tabellen geschah durch das graphische *Frontend* PhpMyAdmin (3.3.6).

4.4 Auswertungen der MySQL-Datenbanken

4.4.1 Umfang der Ausgangsdaten

Insgesamt wurden bei der Analyse der TRJ8-Datenbank 45.692 Favoriten erhalten, von denen 22.282 einen LocusLink-Eintrag und 24.966 einen Unigene-Eintrag hatten. Durch Optimierung der Abfrageparameter beim QM II sowie qualitativen Verbesserungen der abgefragten Datenbanken in der Zeit nach dem TRJ8-Versuchslauf konnten für den TRV1-Ansatz höhere Ergebnisraten erzielt werden:

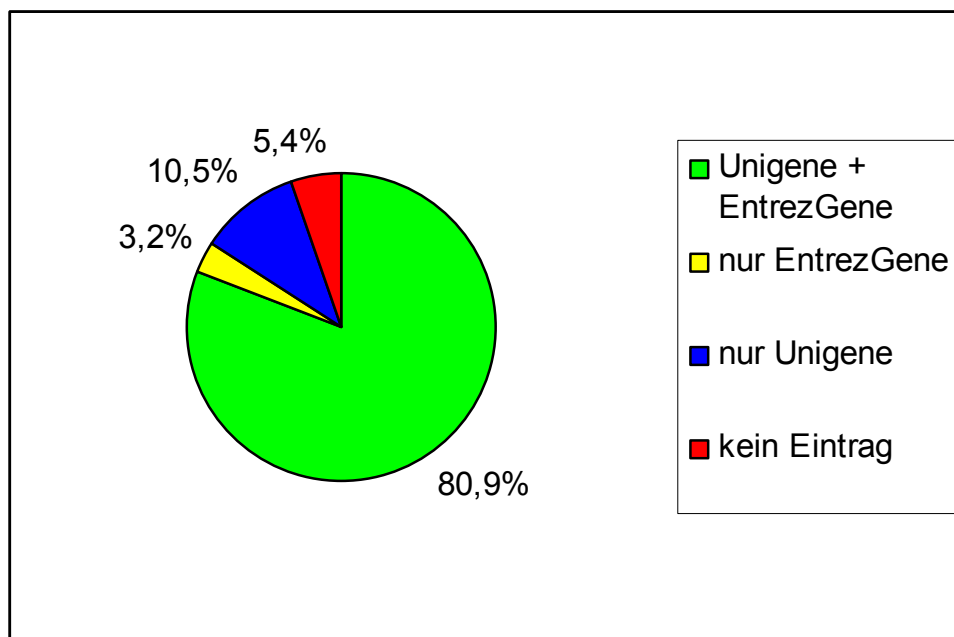


Abbildung 15: Verteilungsdiagramm der Blast-Favoriten von TRV1 mit Unigene-, bzw. EntrezGene-Dateneintrag.

Von den 48.969 Favoriten hatten 39.630 sowohl Einträge in Unigene als auch in EntrezGene, 1.589 nur ein EntrezGene-Ergebnis, für 5.130 Datensätzen lag lediglich der Unigene-Verweis vor, und für 2.620 Favoriten konnte weder ein Unigene noch ein Entrez-Eintrag gefunden werden.

Ergebnisse

Datenbank *testretvaleria1* - Tabelle *unigene* auf *localhost*

Struktur Anzeigen SQL Teilw. anzeigen Einfügen Exportieren Operationen Optionen Leeren Löschen

Zeige Datensätze 60 - 89 (44760 insgesamt)

SQL-Befehl : [Bearbeiten] [SQL erklären] [PHP-Code erzeugen]
`SELECT *
FROM `unigene` LIMIT 60 , 30`

Zeige : 30 Datensätze, beginnend ab 90
untereinander angeordnet und wiederhole die Kopfzeilen nach 100 Datensätzen.

	id	unigene_id	file	genetic_map	cDNA_source
Bearbeiten Löschen	61	Hs.70903	/home/jens/testretvaleria2/consensus/1000061.xml.fav.entrezunigene	20pter-q11.23	Bladder, Bone, Brain, Cervix, Colon, Eye, Kidney, ...
Bearbeiten Löschen	62	Hs.528439	/home/jens/testretvaleria2/consensus/1000062.xml.fav.entrezunigene	11q14.1	Bladder, Blood Bone, Brain, Colon, Eye, Larynx, L...
Bearbeiten Löschen	63	Hs.523510	/home/jens/testretvaleria2/consensus/1000063.xml.fav.entrezunigene	11p15.1-p14	Bladder, Blood Bone, Bone Marrow, Brain, Cervix, ...
	64	Hs.177576	/home/jens/testretvaleria2/consensus/1000064.xml.fav.entrezunigene		Bladder, Blood

Abbildung 16: Graphische Ausgabe der *unigene*-Tabelle von TRV1 durch PHPMYAdmin.

4.4.2 Erstellung und Auswertung der „nr-eye“-Datenbank

Zur Erstellung einer vollständig annotierten humanen Datenbank des menschlichen Auges aus den vorliegenden MySQL-Rohdaten mussten primär drei Anforderungen erfüllt sein: a) sollten die Einzeltabellen fusioniert werden können, b) nur im Auge exprimierte ESTs abgebildet sein und c) verbliebene Redundanzen entfernt werden. Die so erstellte virtuelle Ergebnisdatenbank verwendete den jüngeren TRV1-Ansatz und wird im Folgenden „nr-eye“ genannt (3.3.7.2).

Für eine Zusammenfassung der vier erzeugten Einzeltabellen *clones* (aus StackPACK exportierte Sequenz plus Kopfzeile), *blast*, *entrezgene* und *unigene* wäre die Verwendung eines Primärschlüssels sinnvoll gewesen, d.h. eines in allen Tabellen identischen Wertes, um eine korrekte und schnelle Zuordnung aller Daten zu gewährleisten. Da das lokale Blast-Programm keinen solchen Primärschlüssel verwendet, musste ein aufwendigerer MySQL-Befehl zur Fusion der Einzeltabellen konzipiert werden: So wurden *clones* und *blast* über die gleiche *accession-number* verbunden, ein Abgleich von *blast* und *entrezgene* erfolgte über die GI-Nummer, wohingegen *unigene* über den CID-Code mit *blast* zusammengeführt wurde.

Die Selektion nur im Auge exprimierter Gene geschah unter Anwahl der Spalte *cDNA_sources* aus der *unigene*-Tabelle und Verbindung mit einem Booleschen AND-Operator.

Da das CRAW-Modul von StackPACK die bei der Clusterung gebildeten Contigs z.T. in verwandte oder ähnliche MSCs unterteilte und exportierte (z.B. Splice-Isoformen, siehe Abb. 12), enthielten die Blastresultate bei unterschiedlicher Abfragesequenz z.T. gleiche Favoriten. Um diese Redundanzen für die nachfolgende Analyse auszublenden, wurden entsprechende Treffer durch COUNT-, GROUP- und ORDER-Statements in einer Gruppe zusammengefasst. Die Multiplizität wurde durch einen COUNT-Befehl ermittelt und in einer neuen Indexspalte „Anzahl“ ausgewiesen. So würden vier unterschiedliche Sequenzen mit Favorit-Gen *OPAI* in der *nr-eye*-Datenbank nur einmal, aber mit dem „Anzahl“-Wert „4“ ausgegeben werden, wobei das angezeigte Ergebnis dem mit dem höchsten *Hit-Score* entspräche. Die Gruppierung erfolgte dabei anhand der „*definition line*“, die aus dem jeweiligen favorisierten Blast-Resultats ausgelesen wurde. Ein Beispiel für eine „*definition line*“ wäre z.B.: „Homo sapiens optic atrophy 1 (autosomal dominant) (*OPAI*), nuclear gene encoding mitochondrial protein, transcript variant 5, mRNA“.

In Abgrenzung zum sequenzbasierten Clustern durch StackPACK wird dieser rein datenbankgestützte Vorgang im Folgenden als „MySQL-Clustern“ bezeichnet.

Abbildung 17 zeigt die Notwendigkeit dieses Schrittes zur Redundanzbeseitigung, etwa 70% der ermittelten Cluster zeigten eine solche Redundanz. Das multipelste Gen, der Interleukin 28-Rezeptor alpha, trat dabei sechszigmal auf. Dieser Effekt kann auf der Erstellung von Subclustern aufgrund von a) unterschiedlich schlechter Qualität der Ursprungssequenzen oder b) Splicevarianten oder c) Polymorphismen oberhalb der definierten Filterwerte von CRAW beruhen.

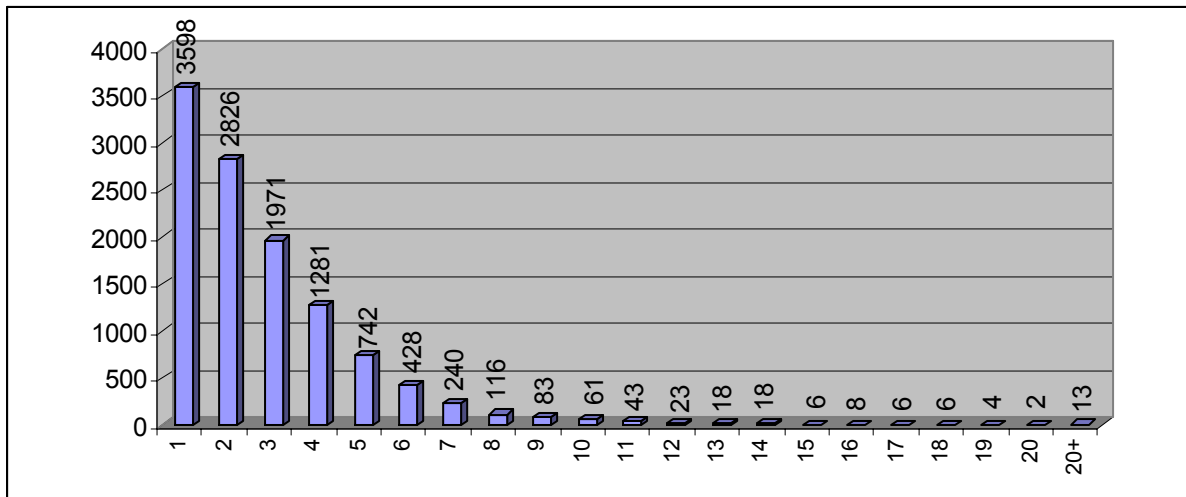


Abbildung 17: Verteilung der Redundanz in der „nr-eye“-Datenbank.

Die 48.969 Favoriten wurden durch das MySQL-Clustern auf 14.533 nichtredundante Datensätze, also auf etwa ein Drittel der Ursprungsmenge reduziert. Von den 14.533 nichtredundanten Datensätzen zeigten 11.493 laut Unigene eine Expression im Auge und 12.635 im Gehirn auf. 10.982 sind in beiden Geweben exprimiert und 1.387 in keinem der beiden. Um zu testen, wie viele der in *nr-eye* gelisteten Ergebnisse vollständig annotiert waren, wurde die Datenbank hinsichtlich des Anteils von Datensätzen mit bestimmten Einträgen analysiert. Die Ergebnisse umfassten den Wert NULL (leerer Eintrag) oder einen beliebigen anderen Inhalt. Zur Vermeidung mehrdeutiger Angaben wurde der englische NCBI-Begriff aus dem XML-Quelltext als Zeilenname beibehalten.

	NULL	Nicht NULL
Description	9	11.484
gene-name	2.479	9.014
Geneontology	2.451	9.042
genetic map	0	11.493
MIM-link	5.033	6.460
Summary	5.083	6.410

Es wurde weiter untersucht, inwieweit sich dieses Expressionsprofil zwischen den Gruppen der MSCs und Singletons unterschied. Hierzu wurden die beiden Datensätze separat MySQL-geclustert, weswegen die Einzelsummen der Cluster größer als die oben genannte Gesamtzahl waren.

Ergebnisse

Diese Qualitätsanalyse hatte den Sinn, dass im Falle einer signifikant schlechteren Informationsdichte der Singletons, bzw. einer ausreichenden Datenfülle durch die MSCs bei späteren Versuchsläufen auf eine Bearbeitung der Singletons möglicherweise verzichtet werden könnte.

	Auge & Gehirn	Auge insg.	Gehirn insg.	Weder noch	MySQL-Cluster gesamt
Consensus	7.372	7.599	7.861	380	8.468
Singleton	8.526	8.932	10.000	1.256	11.672

Als weiterer Qualitätsparameter wurde die Score-Werte-Verteilung der Blast-Ergebnisse zwischen Consensus- und Singleton-Sequenzen verglichen. Um dies zu untersuchen, wurde die „nr-eye“-Datenbank um den Parameter des Hit-Scores von Blast erweitert.

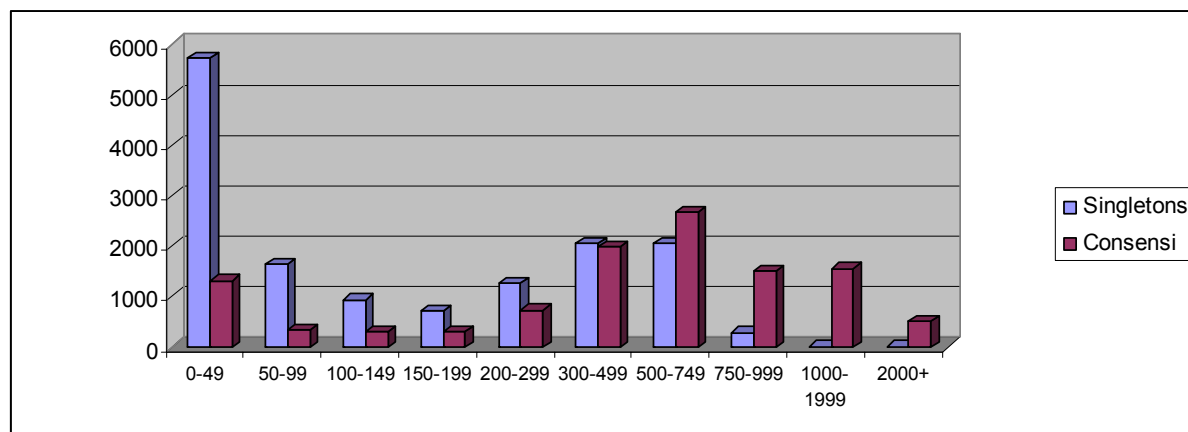


Abbildung 18: Aufschlüsselung der MSCs und Singletons aus „nr-eye“ nach Hit-Score.

Das Ergebnis höherer Score-Werte bei den MSCs ist nicht unerwartet, da für die MSCs als Contig mehrerer Sequenzen im Mittel sowohl eine bessere Sequenzgenauigkeit als auch eine größere matchende Gesamtlänge zu erwarten ist

Nach Abschluß der rein qualitativen Analysen wurde die nr-eye-Datenbank auch inhaltlich ausgewertet. Während bei der lokal erstellten WERI-Rb1-Datenbank manuell eine Primärfunktion des korrespondierenden Proteins aus den LocusLink-Daten ausgewählt werden konnte (Abbildung 9), war dies aufgrund der Zahlenmenge hier nicht mehr möglich. Als Beispiel für die Auswertungsmöglichkeiten der Datenbank wurde lediglich nach auftretenden Begriffen in der gene-ontology-Spalte der *entrezgene*-Tabelle gefiltert. So waren

hier Mehrfachnennungen möglich, z.B. konnte ein ATP-bindendes Protein erneut in der Spalte für Kinasen auftreten.

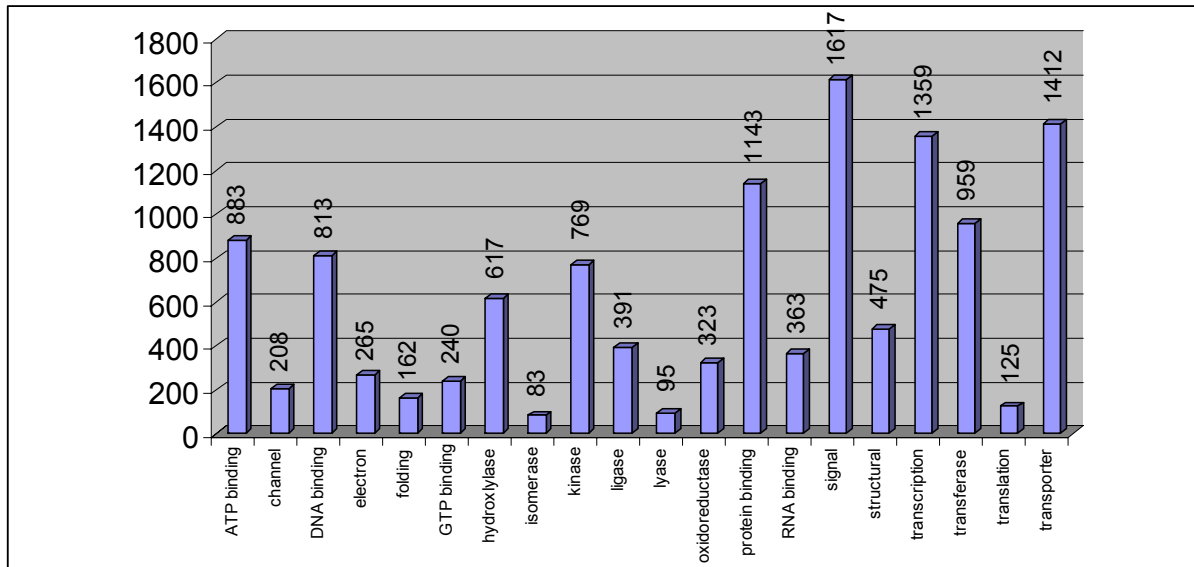


Abbildung 19: Beispiel für eine Aufschlüsselung der *nr-eye*-Ergebnisse nach Ontologie. Auf der X-Achse abgebildet sind die verwendeten englischen Filterbegriffe.

Für die Identifizierung lokaler Kandidatengene für erbliche Erkrankungen des Auges oder die Indetifikation lokaler/chromosomaler Cluster ist eine Auswertung der „*nr-eye*“-Gene bezüglich ihrer chromosomalen Lokalisation unerlässlich. Diesen wurden die bekannten, bzw. prädiktierten Gene aus der Ensembl-Datenbank (www.ensembl.org) gegenübergestellt:

Chromosom	„ <i>nr-eye</i> “-Gene	Bekannte Gene (Ensembl)	Prädikierte Gene (Ensembl)
1	1.129	1.988	2.281
2	754	1.246	1.482
3	642	1.033	1.168
4	438	743	866
5	535	834	970
6	576	1.050	1.152
7	542	916	1.116
8	415	692	794
9	456	778	919
10	438	730	862
11	682	1.264	1.462

Ergebnisse

12	613	1.009	1.104
13	193	318	399
14	363	646	733
15	376	589	766
16	533	839	957
17	726	1.104	1.257
18	172	267	322
19	696	1.337	1.468
20	376	592	631
21	80	243	271
22	242	471	552
X	381	766	931
Y	8	13	37

Es zeigt sich, dass der Anteil der im Auge exprimierten Gene bei allen Chromosomen relativ homogen verteilt ist, wobei die Chromosome 21, X und Y vergleichsweise wenige dieser Gene besitzen.

Neben der chromosomalen Verteilungen der „nr-eye“-Gene ist es möglich, auch eine genauere subchromosomale Lokalisation vorzunehmen. Das akrozentrische Chromosom 21 enthielt die wenigsten gelisteten Gene und wurde daher exemplarisch analysiert:

Lage	Gensymbol
21q11.1	<i>CNN2</i>
21q11.2	<i>LOC3888B, similar to ancyrin domain 20A</i>
21q21	<i>C21orf66</i>
21q21- 21q22.1	<i>GABPA</i>
21q21.2	<i>APP, CYYRI</i>
21q21.3	<i>ADAMTS5, c21orf66, CRYZL1, MRLPL39</i>
21q21.3- 21q22.1	<i>MRPS</i>
21q22	<i>c21orf66, WDR</i>

Ergebnisse

21q22.1	<i>c21orf59, c21orf91, CXADR, GART, HLCS, HUNK, IFNAR, IFNAR2, MRAP, NCAM2, SFRS15, TIAMI, ZNFZ94</i>
21q22.1- 21q22.2	<i>DSCR1, SON, ATP50, IL10RB, ITSN1</i>
21q22.11	<i>BACH1, c21orf6, c21orf45, c21orf63, OLIG1</i>
21q22.12	<i>C21orf51, CLIC6</i>
21q22.13	<i>CHAF1B, DRYK1A, ZCWCC3</i>
21q22.2	<i>DSCAM, DSCR3, SYNJI, WDR9</i>
21q22.3	<i>ABCG1, ADARP1, AGPAT3, BACE2, c21orf2, c21orf25, c21orf33, c21orf56, c21orf58, c21orf106, c21orf124, CBS, COLGA1, COLGA2, CSTB, DNA segment 2056, DESC2, FTCD, HRMT1L1, ITGB2, LSS, MCM3AP, MX1, PCBP3, PCNT2, PKNOX1, PRDM15, PTTG1IP, RIPK4, S100B, SLC19A1, Tmprss2, UBE2G2, WDR4, WRB</i>

Eine Lokalisation bezüglich physikalischer Sequenzpositionen ist mit dem QM II prinzipiell ebenfalls möglich, sofern die Positionsdaten in den abgefragten Datenbanken vorliegen und die Ausleseskripte entsprechend angepasst werden.

Abschließend wurden versuchsweise anhand dreier Parameter (Lokalisation auf Chromosom 16, Koexpression in Auge und Gehirn und mitochondriale Ontologie) Kandidatengene für einen möglichen zweiten Locus des humanen Gens Optikusatrophie gesucht. Hierbei konnten 29 Gene für das fragliche Intervall in der „*nr-eye*“-Datenbank identifiziert werden (Daten nicht gezeigt).

4.4.3 Auswertung der TRJ8-Datenbank

Im Unterschied zum TRV1-Ansatz wurde diese Datenbank unter Ansteuerung der *non redundant*-Datenbank statt der humanen Referenzsequenzen und unter Verwendung eines aktiven Favoritenscriptgenerators (3.3.5.2c) erstellt. Aufgrund ihrer geringeren Aktualität (dieser Versuchslauf wurde etwa 12 Monate vor dem TRV1-Ansatz durchgeführt), wird auf die explizite Darstellung der prinzipiell ähnlichen Ergebnisse verzichtet. Da sich die Rechenzeiten zwischen beiden Verfahren um mehrere Wochen unterschieden (4.3.4.2),

Ergebnisse

sollte getestet werden, ob eine Konzentrierung auf die menschlichen Sequenzen wie im TRV1-Ansatz zu einem nennenswerten Verlust an Information hätte führen können:

Von den 22.282 LocusLink-Einträgen der ausgewerteten Favoriten von TRJ8 waren 21.902 dem Menschen zuzuordnen, 299 *Mus musculus*, 29 *Drosophila melanogaster*, 23 *Rattus norvegicus*, 19 *Danio rerio*, 3 *Caenorhabditis elegans*, 2 *Bos taurus*, 2 *Xenopus laevis*, 1 *Gallus gallus*, 1 *Canis familiaris* und 1 Sequenz erbrachte eine Homologie zu *Sus scrofa*. Insgesamt waren also lediglich knappe 1,5 % der Datensätze nicht humanen Ursprungs, was angesichts des enormen Rechenzeitunterschiedes ein praktischer Grund für die Favorisierung des TRV1-Ansatzes war.

Eine genaue Analyse der Effizienz des Favoritenscriptes für den TRJ8-Ansatz unterblieb, da zum Zeitpunkt des TRJ8-Versuchslaufes noch kein Import dieser spezifischen Information in die MySQL-Tabelle realisiert war. Eine manuelle Auswertung der Blast-Ergebnisse von jeweils 50 MSCs und Singletons ergab, dass in mehr als 5% Favoriten gewählt wurden, die nicht an erster Stelle des ursprünglichen Blast-Rankings standen und somit ursprünglich möglicherweise weniger vollständig annotierte Kandidaten ausgewertet worden wären.

5. Diskussion

5.1 Diskussion der cDNA-Bibliothek

5.1.1 Auswertung der WERI Rb1-Klone durch den QM I

Mit der generierten WERI Rb1-cDNA-Bank sollte zum einen die Frage beantwortet werden inwieweit diese Zelllinie als effiziente Quelle zur Identifizierung neuer retinal exprimierter Gene des Menschen dienen kann. Zum anderen sollten die isolierten Sequenzen ggf. als Ausgangsmaterial für einen retinaspezifischen Expressionschip bilden.

Auf den ersten Blick überraschte die geringe Anzahl redundanter Sequenzen in der cDNA-Bank nach Abzug der Klonierungsartefakte (Abb. 8), da die erwartete Mengen-Verteilung von Transkripten (Bishop *et al.*, 1974; Davidson und Britten, 1979) in einem wie der Retina hochspezialisierten Gewebes deutlich von den beobachteten Zahlen abwich. Dies ist erstens möglicherweise auf die relativ geringe Anzahl analysierter Klone zurückzuführen. Eine genauere Statistik hätte jedoch eines Mehrfachen an sequenzierten Klonen bedurft. Zudem war zu Beginn des Projekts eine weit höhere Anzahl humaner Gene prädiziert worden, welche zwischenzeitlich jedoch nach unten korrigiert wurde so dass sich eine Fortführung des Projekts aus Mangel an neuen cDNAs nicht mehr anbot.

Weiterhin stellt eine Zellkultur nur ein Modellsystem dar, dessen Expressionsprofil nicht notwendigerweise die tatsächlichen Verhältnisse in der Retina widerspiegeln muss. So wiesen z.B. Proteine mit Signaltransduktions-Funktionen in der lokalen WERI Rb1-cDNA-Bank mit jeweils 5,4% (Abb. 9) eine gleiche Häufigkeit wie bei Untersuchungen der ARPE-19 Zelllinie auf (Sharma *et al.*, 2005). Bei der durch den QM II erstellten Gesamt-Datenbank wurde mit 14,0% fast der dreifache Wert gemessen (Abb. 19), was durch eine schwächere Ausdifferenzierung der Retinoblastomzellen erklärt werden könnte.

Grundsätzlich erwies sich die WERI Rb1-Zelllinie als Quelle retinal exprimierter Gene jedoch als gut geeignet, wie die Vielzahl der gemäß den Unigene-Ergebnissen im Auge exprimierten ESTs belegte. Dies deckte sich mit den Befunden anderer Autoren (Reinisalo *et al.*, 2003), wobei Retinoblastome generell als gutes Modellsystem zur Identifizierung zur Analyse des Transkriptoms des Auges gelten (Bernstein *et al.*, 1994; Sharma *et al.*, 2005).

Auf eine tiefergehende Sequenzierung der cDNA-Bank, sowie eine Optimierung der cDNA-Bibliothek durch eine *full-length*-Anreicherung (Carninci *et al.*, 1996) oder Normalisierung (Bonaldo *et al.*, 1996) wurde verzichtet, da bei 960 analysierten Klonen keine neuen cDNAs/Gene identifiziert werden konnten und somit auch bei weiteren Versuchen mit einer extrem geringen Neuidentifizierungsrate zu rechnen war.

5.1.2 Aussicht

Die Zielsetzung, die selbst generierten Sequenzen (4.1) für ein Update automatisiert durch den QM II (3.3.5.2) annotieren zu lassen konnte nicht realisiert werden. Hierzu wäre es zunächst nötig gewesen, die Sequenzen von einzelnen Datensätzen in eine Batch-Datei zu überführen. Hiernach wäre sie analog zu den NCBI-Datensätzen durch das StackPACK-Programm (3.3.4.1) geclustert, durch den QM II geblastet und annotiert worden. Das erste Problem war, dass nicht im Genbank-Format vorliegende Batch-Sequenzen zwar geclustert werden konnten, die Consensi aber wegen eines Programmfehlers nicht exportiert und somit nicht annotiert werden konnten. Dieses Problem erwies sich als nicht lösbar. Wegen der damit verbundenen hohen Ausfallraten, der geringen zu erwartenden Informationsfülle und der fortgeschrittenen Zeit wurde auf eine Installation einer alternativen Clusterungssoftware verzichtet. Dieses könnte für zukünftige Arbeiten nachgeholt werden.

Es war zunächst geplant gewesen, die lokal erstellte cDNA-Datenbank als Ausgangsmaterial für die Entwicklung eines retinalen Expressionschip zu benutzen. Der Vorteil eines gewebespezifischen Arrays liegt in den geringeren Kosten pro Chip sowie im besseren Signal/Rausch-Verhältnis, da nicht alle Gene als Target aufgetragen werden müssen. Zwischenzeitlich haben sich die kommerziell erwerblichen Oligonucleotid-Chips jedoch durchgesetzt und werden zu einem sehr günstigen Preis angeboten, weswegen dieses Teilprojekt nicht verwirklicht wurde.

5.2 Hypoxie-induzierte Expressionsanalyse der Zelllinie WERI Rb1

Sowohl *VEGF* als auch *EPO* und *GAPDH* werden durch den Transkriptionsfaktor *HIF-1* reguliert (Chun *et al.*, 2002). *HIF-1* wird unter normoxischen und hypoxischen Bedingungen gebildet, jedoch bei normaler Sauerstoffversorgung proteasomal abgebaut. *HIF-1* verstärkt als Transkriptionsfaktor die Expression von *VEGF*, um durch Neovaskularisierung die Sauerstoffversorgung des betreffenden Gewebes zu verbessern (Shweiki *et al.*, 1992). Die Expression von *EPO* kann bis zu 100fach erhöht werden (Semenza *et al.*, 1991), was die Erythrozytenneubildung stark beschleunigt (Hochachka *et al.*, 1996). *GAPDH* wird wie andere Glycolyseenzyme verstärkt gebildet, um die durch die verringerte oxidative Phosphorylierung stark verminderte ATP-Produktion wieder aufzufangen (Hochachka und Lutz, 2001). Die 18S-rRNA hingegen gilt als effizientes Beispiel einer inerten Kontrolle (Suzuki *et al.*, 2000).

Überraschenderweise zeigte die Q-PCR-Analyse keine generelle Aktivierung der *HIF-1*-regulierten Gene, aber eine starke Verminderung von *VEGF* nach 24 h bei Normbedingungen, sowie eine etwas schwächere Absenkung der *GAPDH*-Transkriptmenge unter denselben Bedingungen.

Es stellt sich die Frage, ob es in der Zeit zwischen Beendigung der Hypoxie und der Extraktion der mRNA zu einer Reversion des Expressionslevels der betreffenden Gene kam. Dies ist jedoch unwahrscheinlich, da erstens die Proben auf Eis lagen und deutlich kürzer als die halbe Stunde normoxischen Bedingungen ausgesetzt waren, unter denen eine Reversion theoretisch möglich gewesen wäre (Dr. Hans-Jörg Riedinger, Physikalisch-chemisches Institut Tübingen, pers. Mitteilung). Zweitens erweisen sich die Transkripte der ausgewählten Gene in postmortalem menschliche Gewebe auch nach 48 Stunden mit der Ausnahme einer mit der Zeit steigenden *EPO/GAPDH*-Relation als bemerkenswert stabil (Zhao *et al.*, 2006), was gegen einen sehr schnellen Adaption in unserem Ansatz spricht. Zudem hätten zunächst überschüssige *HIF-1*-Proteine abgebaut werden müssen, bevor sich die nachgeschalteten Zielgene in ihrer Expression wieder an Normbedingungen hätten anpassen können.

Erschwert wird die Bewertung von Q-PCR-Ergebnissen dadurch, dass Zellen aufgrund der vorliegenden Stressbedingungen eine erhöhte Apoptoserate haben und die absterbenden

Zellen mit ihrem geänderten Expressionsverhalten die Gesamtmessung beeinträchtigen können (Dr. Perikles Simon, Institut für Sportmedizin Tübingen, pers. Mitteilung).

Möglicherweise war der Sauerstoffgehalt mit 1% auch zu hoch angesetzt, um eine signifikante Regulation zu messen. In Säugergewebe kann unter normalen Bedingungen der Sauerstoffpartialdruck unter 10% liegen, wobei im Falle eines Schlaganfalls, einer Ischämie oder im Inneren eines etablierten Tumors der O₂-Partialdruck bis auf 0,02% absinkt (Semenza, 2000). Auf der anderen Seite entsprach der gewählte Wert in etwa dem Ansatz anderer Autoren (Maslim *et al.*, 1997; Ernens *et al.*, 2006).

Weiterhin ist der Einfluss externer Faktoren wie die Zusammensetzung des Wachstumsmediums auf die Vergleichbarkeit von Expressionsexperimenten nicht zu unterschätzen (Sharma *et al.*, 2005). Da die Transkriptionslevel nicht vergleichend unter Verwendung anderer Wachstumsmedien gemessen wurden, kann ein diesbezüglicher Effekt nicht grundsätzlich ausgeschlossen werden.

HIF1-regulierte Gene zeigen in manchen Fällen eine sehr zelltypspezifische, von externen Stimuli (z.B. transkriptionelle Cofaktoren oder Zell-Zell-Kontakte) abhängige Aktivierung, was die schwache Reaktion in WERI *Rb1*-Zellen erklären könnte (Bracken *et al.*, 2003). Es wäre in diesem Fall jedoch recht ungewöhnlich, wenn dies auf alle gewählten Gene zuträfe, da diese normalerweise elementar für eine Hypoxieantwort sind (Liu und Simon, 2004).

Auch unterliegt die Expression von *HIF-1* neben der Regulation durch Sauerstoff noch anderen Faktoren: So beeinflussen reaktive Sauerstoffderivate wie freie Radikale die Stabilität von *HIF-1* (Chandel *et al.*, 2000). Zudem erhöht ein Sauerstoffmangel den Partialdruck von Stickoxiden, da *HIF-1* die Expression von Stickoxidsynthetasen erhöht (Palmer *et al.*, 1998). Stickoxide können die Expression mancher *HIF-1*-regulierten Genen wie *EPO* oder *VEGF* reprimieren (Chun *et al.*, 2002), was die Ergebnisse der Q-PCR-Versuche erklären könnte. Dieser Steuerungseffekt scheint jedoch dosisabhängig zu sein: Unterhalb einer Konzentration von 0,5 mM NO_x gibt es Hinweise auf eine Reprimierung von *HIF-1* (und somit auch eine Reprimierung positiv durch *HIF-1* regulierter Gene wie *VEGF*), oberhalb dieser Konzentration wird im Gegenteil eine Aktivierung gemessen (Kimura *et al.*, 2000). Die Relevanz dieser externen Faktoren für das vorliegende Experiment ist jedoch ungeklärt.

Das signifikante Absinken der *GAPDH*-Transkriptmenge nach 24 Stunden unter Normbedingungen könnte dadurch bedingt sein, dass die Zellen bei optimaler Sauerstoffversorgung und 24 h Inkubationszeit bereits einen Teil des Mediums verbraucht

haben, bzw. hemmende Metabolite in hoher Konzentration vorliegen. Als Folge des sinkenden Zuckerangebotes könnten die Zellen u.U. die Transkription ihrer Glycoyse-Gene reprimiert haben. Der Farbindikator des Mediums zeigte zwar noch keinen Umschlag ins Gelbliche, doch war bereits ein leichtes Aufhellen der Färbung zu beobachten.

Der niedrige Level an Erythropoietin-Transkripten, bzw. die schwache Regulation in der WERI *Rb1*-Zelllinie hängen möglicherweise damit zusammen, dass *EPO* zu 90% in der Niere gebildet und damit generell nur schwach in anderen Geweben exprimiert wird (Jelkmann, 2004). In einer intakten Mausretina wurde hingegen eine Anhebung von *EPO* um ca. Faktor acht unter hypoxischen Bedingungen beobachtet (Grimm *et al.*, 2002).

Eine ebenfalls sehr überraschende Beobachtung war die signifikante Absenkung der *VEGF*-Transkriptmenge nach einem Tag Normoxie und die fehlende Induktion bei Sauerstoffmangel. *VEGF* zeigt unter hypoxischen Bedingungen ein abweichendes Spliceing, was möglicherweise die Menge an prozessiertem Template reduzierte. Sechs Isoformen für alternatives Spliceing aufgrund von variablen physiologischen Milieus sind bei *VEGF* bekannt. Dem steht entgegen, dass das Exon 8, welches die Primerbindungsstellen enthält, normalerweise nicht entfernt wird, allerdings ist das Spliceingmuster bei *VEGF* noch weitgehend unverstanden (Gora-Kupilas und Josko, 2005). Die genetische Regulation von *VEGF* ist sehr komplex: *VEGF* kann u.a. inhibiert werden durch TNF- α , Selen oder Wildtyp-p53-Protein, wobei unklar ist, warum diese Faktoren im 24 h Normoxie-Versuch eine mögliche Bedeutung haben sollten und in den anderen Versuchsläufen nicht. (Lin und Chao, 2005).

Für das Auge könnte diese Messung bedeuten, dass hypoxische Bedingungen auf Ebene der Genregulation weit weniger dramatische Folgen nach sich ziehen als bislang angenommen. Das Auge und insbesondere die Netzhaut als stoffwechselaktives Organ, bzw. Gewebe ist in der Tat oft einem hypoxischen Milieu ausgesetzt. Durch den starken Sauerstoffverbrauch der Stäbchen liegen Teile der äußeren Retina unter sehr sauerstoffarmen Bedingungen und sind daher möglicherweise entsprechend unsensitiv (Arden *et al.*, 2005). Zudem sind Retinoblastomzellen als Krebszellen möglicherweise auf hypoxische Umgebungen adaptiert. So führen Mutationen im *p53*-Gen zu einer erhöhten Produktion von *VEGF* (Royds *et al.*, 1998), so dass eine weitere Hochregulation u.U. nicht mehr ohne weiteres möglich ist. Es ist gut denkbar, dass die Netzhaut als komplexes Gewebe durch die Auswirkungen einer

Diskussion

Hypoxie leichter zu schädigen ist als isolierte Zellen in einer Kultur. Generell ist bei der Interpretation der Ergebnisse zu beachten, dass die WERI *Rb1*-Zelllinie nur ein Modellsystem ist und als solches nicht unbedingt repräsentativ für die Verhältnisse in einer hypoxischen Retina ist.

5.3 Erstellung und Analyse der „nr-eye“-Datenbank

5.3.1 Zielsetzung

Die geringe Zahl von Sequenzen die bei der Pilotanalyse der WERI *Rb1*-Datenbank gewonnen wurde ist für die Findung von Kandidatengenomen für genetische Erkrankungen des Auges zu gering, weswegen die vom NCBI bereitgestellten Sequenzen für eine Analyse des menschlichen Retinoms einbezogen werden sollten. Um die Datenmengen zu bewältigen, sollte nach einer Datenclustering mit dem QM II und der anschließenden Aufarbeitung der Ergebnisse durch eine relationale Datenbank eine annotierte, redundanzbereinigte Transkript-Bibliothek des menschlichen Auges auf der Basis bereits vorhandener cDNA- und EST-Sequenzen erstellt werden.

Erfolg und Praktikabilität eines solchen Ansatzes messen sich neben der Qualität der Ergebnisse an den dafür benötigten Systemressourcen wie Speicherplatz und Rechenzeiten.

5.3.2. Clusterung durch StackPACK

Das erste Modul war das Clusterungsprogramm StackPACK. Die Reduktion der Datenmenge durch StackPACK betrug beim TRV1-Ansatz etwa ein Viertel (4.3.3), wobei der Clusterungsfaktor, also der Quotient aus zusammengefassten Sequenzen zu den gebildeten MSCs, bei ca. 11:1 lag. Diese Clusterung reduzierte gleichermaßen die Systemanforderungen, wie es die Qualität der Sequenzen durch Clusterung und Consensusbildung für nachfolgende Schritte verbesserte.

Da die vorhandenen Blast-Resultate die Anzahl der vermuteten im Auge exprimierten Gene (Schulz *et al.*, 2004) und sogar die Gesamtzahl der prädiktierten Gene (International Human Genome Sequencing Consortium, 2004) des Menschen bei weitem überschritten, mussten noch immer Redundanzen vorhanden sein. Zum einen ist das Clustern trotz sinnvoller Filtersets grundsätzlich eine statistische Abwägung zwischen falsch-negativen und falsch-positiven Ergebnissen und kann daher nie restlos alle Dubletten beseitigen, zum zweiten werden ähnliche Sequenzen wie z.B. Isoformen eines Contigs oder Hybridklone von StackPACK in unterschiedliche Consensi abgespeichert (Abb. 14). Bei diesen Sequenzen

handelte sich somit um „informative Redundanzen“, die aufgrund möglicher nachgeschaltete Analysen (z.B. für Detailbetrachtungen eines Kandidatengens) nicht gelöscht wurden. Zudem bestand die Möglichkeit des Nachfilterns durch MySQL. Aus diesen Gründen waren die Clusterungs-Parameter des StackPACK nicht stringenter eingestellt worden.

Grundsätzlich erfüllte StackPACK die gestellten Anforderungen. Einziges Manko war die fehlende Exportierungsmöglichkeit der Link-out-Sequenzen (nicht überlappende MSCs). Deren Anzahl lag jedoch nur bei etwa einem Promill der exportierten Daten und umfasste lediglich bereits aus den Consensi bekannten Gene. Damit war der Verlust nicht relevant. Möglicherweise basierte der Exportfehler darauf, dass StackPACK für die Linuxversion Red Hat optimiert war, lokal jedoch SuSe verwendet wurde.

Die benötigte Rechenzeit hielt sich mit etwa einer halben Woche im Rahmen dessen, womit z.B. unter Verwendung von ClustalX bei gegebener Rechenleistung zu rechnen gewesen wäre (Jeanmougin *et al.*, 1998).

5.3.3 Ansteuerung der NCBI-Datenbanken durch den QM II

Mit dem Querymaster II wurde eine neue Plattform programmiert und implementiert, die folgende Anforderungen erfüllte:

- die halbautomatisierte Blast-Analyse der geclusterten Importsequenzen
- die Auswahl eines definierteren Favoriten aus der Liste der jeweiligen Blast-*Alignments*
- die Ansteuerung ausgewählt werden sollte. Darauf folgend sollten Unigene und LocusLink, bzw. EntrezGene angesteuert und die zusammengehörenden Datensätze in eine MySQL-Tabelle geschrieben werden.

Der Import der mit StackPACK geclusterten Daten in den QM II sowie deren Einspeisung in das lokale Blastmodul geschahen ohne größere Probleme. Ein massiver Programmfehler in der lokalen blastall-Version 6 führte zu Systemabstürzen und einer Verzögerung der Bearbeitung durch manuelle Nachbearbeitungen beim TRJ8-Ansatz. Durch die Installation der blastall-Version 10 konnte dieser Fehler für den Versuchslauf TRV1 beseitigt werden. Jener konnte durch die Installation einer neueren Blastversion für TRV1 beseitigt werden.

Diskussion

Die Entwicklung des Favoritenscripts war für die Auswahl eines Favoriten in TRJ8 wichtig, da der informativste Treffer („*best hit*“) in mehr als 5% der Fälle nicht dem obersten Treffer in der Reihe der Blastresultate entsprachen. Die Bedeutung des Favoritenscripts verschwand jedoch mit der Verwendung von RefSeq (Pruitt *et al*, 2005) als Referenzdatenbank für die Blast-Analyse, da es außer dem Hit-Score keine unterscheidbaren Faktoren mehr gab und das Favoritenscript deshalb de facto abgeschaltet werden konnte. Insgesamt zeigte sich, dass im Vergleich zur hohen erforderlichen Rechenleistung und ca. 98,5% menschlicher „Best Hits“ beim TRJ8-Versuchslauf der Blast-Abgleich gegen die humanen RefSeq-Sequenzen ein qualitativ gutes Ergebnis in angemessener Zeit erbrachte.

Nach den Erfahrungen mit dem zwischenzeitlich eingestellten lokal installierten LocusLink wurde ein etwas langsamerer, aber stabilerer und leichter zu programmierender internetbasierter Ansatz zur Ansteuerung von Unigene und EntrezGene genutzt. Mit dieser Lösung konnte innerhalb zweier Tage der komplette Datensatz analysiert werden, was ein akzeptabler Zeitrahmen war, der durch Aufsplitten der Datensätze mittels Hyperthreading weiter hätte verkürzt werden können.

Der Import der in XML-Formaten abgespeicherten Einzelresultate verlief nach Lösung der durch die XML-Variante hervorgerufenen Hindernisse bei der Darstellung der EntrezGene-Daten (4.3.4.5) ohne Probleme und in ausreichender Geschwindigkeit. Ein Nachteil des XML-basierten Ansatzes war Datengröße der Blast-Resultate mit durchschnittlich je ca. 100 KByte. Dieser Speicherumfang konnte bei TRV1 durch die Beschränkung auf lediglich eine Iteration reduziert werden. Nach Beendigung beider Hauptversuche war die Festplatte des benutzten Rechners dennoch beinahe komplett belegt. Die Rohdaten werden nach dem Export in die MySQL-Tabelle, bzw. Sicherung durch das MySQLdump-Programm jedoch nicht mehr zwingend benötigt. Da diese Daten wieder gelöscht werden sollten, wurde auf eine Verlinkung mit dem MySQL-Frontend verzichtet – geplant war zunächst, dass durch Anwahl eines Buttons die jeweilige XML-Datei in einem Browser geöffnet werden sollte.

Insgesamt erfüllte der QM II abgesehen von der sehr langen Entwicklungszeit alle an ihn gestellten Anforderungen vollständig. Selbstverständlich bleibt der QM II nur solange funktionell, wie sich die angesteuerten Datenbanken weder nennenswert in Format und Inhalt verändern. Die vom NCBI gelieferten Datenbanken haben sich im Laufe der Zeit jedoch sehr

verändert und erweitert (Wheeler *et al.* 2006), so dass ein kontinuierlicher Anpassungsprozeß für eine weiterführende Nutzung des QM II erforderlich wäre.

5.3.4 Ansteuerung der MySQL-Tabellen durch PHPMyAdmin

Die vier Einzeltabellen wurden durch einen MySQL-Befehl virtuell in eine einzige Ergebnistabelle überführt, die dann ausgewertet wurde. Hierbei waren die Bearbeitungszeiten einer MySQL-Abfrage länger als die erwarteten Zehntelsekunden. Hauptursache der Ladezeiten dürfte der Mangel eines einheitlichen Primärschlüssels in den Ursprungsdateien sein. So musste zum einen die umständliche Fusionierung der Tabellen und zum anderen die Umrechnung auf eine graphische Ausgabe berechnet werden. Trotz dieser Umstände dauerte die Berechnung komplexer Abfragen zwischen wenigen Sekunden und einer halben Minute. Daher wurden die vorab angestellten Überlegungen, das Transaktionstaugliche PostgreSQL wegen seiner langsameren Abfragen nicht zu benutzen gegenstandslos, wengleich die Verwendung von MySQL keine Nachteile nach sich zog. Einfache Abfragen wurden mit dem graphischen Frontend PHPMyAdmin durchgeführt. Ergebnisse aus verknüpften Tabellen mit mehr als 300 Zeilen verzögerten sich bei PHPMyAdmin erheblich, weswegen z.T. LIMIT-Befehle die Ausgabe reduzierten, diese konnte dann sequentiell ausgewertet werden. Alternativ konnte die schnellere Kommandozeilenausgabe von MySQL genutzt werden.

Wegen der aufgeführten hohen Ausgabezeiten bei PHPMyAdmin waren Alternativen hin zu einer anderen graphischen Oberfläche wie z.B. MySQLFront 3.2 (www.mysqlfront.de) erwogen worden. Die damaligen verfügbaren Alternativen zeigten jedoch ähnliche Ladezeiten und waren z.T. erheblich unintuitiver in der Bedienung. Möglicherweise könnten die Funktionsparameter durch ein Update verbessert werden.

5.3.5 Auswertung der MySQL-Tabellen

Nur etwa 50% aller Blastresultate des TRJ8-Ansatzes lieferten einen LocusLink, bzw. Unigene-Eintrag. Beim jüngeren TRV1-Ansatz verfügten bereits ca. 80% sowohl über einen EntrezGene-, bzw. Unigeneverweis und nur ca. 5% waren diesbezüglich gar nicht annotiert. Bei weiter fortschreitendem Annotierungsprozeß des humanen Genoms kann man davon

ausgehen, dass die noch vorhandenen Informationslücken in der Zukunft geschlossen werden. Hierbei ist zu berücksichtigen, dass die Annotierungstiefe sehr unterschiedlich ist: In der *nr-eye*-Datenbank gibt es chromosomale Lokalisierung für alle gelisteten Gene, während nur etwa 80% einen Ontologie-Kommentar und nur ca. 60% einen MIM-Link aufwiesen.

Im Zusammenhang mit der Identifizierung von Kandidatengen verdienen hier MIM-gelinkte Gene besonderes Augenmerk. Die MIM-Datenbank des NCBI enthält ca. 16.000 Einträge zu menschlichen Genen und hereditären Erkrankungen, wovon mehr als 10.000 diese Einträge Angaben zu Genloci und phänotypische Beschreibungen enthalten (McKusick, 1998). Hat ein potentiell Kandidatengen einen krankheitsbezogenen MIM-Eintrag ohne Relevanz für das Auge, kann es möglicherweise ausgeschlossen werden. Es ist aber auch denkbar, dass eine nur partiell charakterisierte Mutation ebenfalls die Sehfähigkeit einschränken könnte (z.B. bei polygen begründeten Syndromen).

Von besonderem Interesse bei der *nr-eye*-Datenbank ist der mit 30% recht hohe Anteil an einzeln vorliegenden Consensi- oder Singleton-Sequenzen (Abb. 18). Es wäre im Einzelfall zu prüfen, ob diese Sequenzen tatsächlich im Auge, bzw. der Retina exprimiert werden oder sie durch mangelhafte Gewebepräparation oder unvollständige Reprimierung („*leaky expression*“) in entsprechenden cDNA-Banken gefunden wurden. Es ist umstritten, ob singuläre Sequenzen als nicht vertrauenswürdig eingestuft werden, oder ob sie z.B. als seltene Transkripte und somit als interessante Kandidatengene gelten sollten (Schulz *et al.*, 2004). Aus Gründen der Vollständigkeit wurden sie in die Auswertung einbezogen.

Zur Qualitäts-Analyse wurde der Hit-Score nach MSCs und Singletons aufgeschlüsselt (Abb. 18). Mit fast 6.000 Sequenzen wies ein überraschend hoher Anteil von Singletons einen Blastscore von 50 oder weniger aus. Eine stichprobenhafte Analyse ordnete diese Ergebnisse sehr kurzen, qualitativ schlechten DNA-Sequenzen, gehäuft aus Soares-Bibliotheken, (Gerhard *et al.*, 2004) zu. Da die zugeordneten Blast-Favoriten jedoch überwiegend eine Expression im Auge sowie eine Homologie zu einem bekannten Gen aufwiesen, wurden sie nicht aussortiert. Nicht überraschend hingegen war, dass MSCs durchschnittlich einen besseren Hit-Score aufwiesen, da sie erstens, aus überlappenden Sequenzen bestehend, eine größere Gesamtlänge hatten und zweitens beim Alignment Sequenzfehler entfernt werden.

Eine Analyse zur Koexpression in Auge und Gehirn wurde exemplarisch durchgeführt, um die Koexpression in diesen ontogenetisch eng verwandten Geweben zu untersuchen. Der Grad der gemeinsamen Expression zwischen Auge und Gehirn entspricht bekannten Schätzungen, wonach zwei beliebige Gewebe zwischen 73 und 84% ihres Transkriptoms teilen (Brentani *et al.*, 2003). Nur etwa 500 der 14.533 nichtredundanten Transkripte wurden im Auge und nicht im Gehirn exprimiert; ein Wert, der deutlich unter den vergleichbaren Studien (Schulz *et al.*, 2004) lag. Dabei ist zu berücksichtigen, dass dort z.T. andere Datenquellen und Ausschlussgewebe bearbeitet, zum anderen waren die vorliegenden Studien ein Jahr jünger. Durch die Vergrößerung der Datensätze und empfindlichere Messverfahren bei nie vollständiger Transkriptions-Reprimierung wird der Anteil selektiv in einem einzigen Gewebe exprimierter Gene zukünftig weiter schrumpfen, bzw. nur noch von der Wahl der quantitativen Ausschlussfilter abhängen. Der Prozentsatz der ausschließlich im Auge exprimierten Gene dürfte bei zusätzlichen berücksichtigten Geweben weiter sinken.

Fraglich ist, ob die gefundene Verteilung der Gene in der *nr-eye*-Datenbank tatsächlich qualitativ die Genexpression im Auge wiedergibt: Die Retina besteht aus mindestens 55 Zelltypen, die abhängig von ihrem quantitativen Anteil mehr oder minder gut repräsentiert sind (McKay *et al.*, 2004). So ist z.B. die Wahrscheinlichkeit, dass ein ausschließlich im Blauzapfen exprimiertes Gen in der Datenbank repräsentiert ist, deutlich geringer als für ein Gen, welches in Stäbchen exprimiert wird, da ca. 200 mal mehr Stäbchen in der menschlichen Netzhaut vorhanden sind. Weiterhin sind Spenderaugen alter Menschen gegenüber ihren jüngeren Pendanten überrepräsentiert. Alterndes Gewebe verhält sich das globale Expressionsprofil betreffend jedoch z.T. deutlich anders als der entsprechende normale Gewebetyp (Zou *et al.*, 2000), ein Effekt, der auch in der Retina beschrieben wurde (Chowers *et al.*, 2003a). Zudem könnten schwierig zu amplifizierende mRNAs in einer sequenzierten cDNA-Bank unterrepräsentiert sein.

Aufgeschlüsselt nach Chromosomen lag der Anteil der in *nr-eye* gelisteten Gene im Vergleich zu den von Ensembl prädiktierten Gesamtgenzahl bei den Chromosomen 1-19 zwischen 46 und 58%, den höchsten Anteil zeigte das Chromosom 20 mit 59,6%. Aus unbekanntem Gründen fiel der retinale Anteil bei den verbliebenen Chromosomen deutlich geringer aus: Chr. 21: 29,5%, Chr. 22: 43,8% und das X-Chromosom mit 40,9%. Interessant in diesem Zusammenhang sind Modelle, die von einer Clusterung gewebsspezifisch exprimierter Gene in bestimmten chromosomalen Abschnitten ausgehen. So konnten retinale Gencluster auf

bestimmten Bereichen der Chromosomen 6, 11, 16 und 19 lokalisiert werden (Schulz *et al.*, 2004). Ohne Korrelierung der gelisteten Gene aus „*nr-eye*“ mit physikalischen Positionsdaten konnten in der vorliegenden Arbeit keine retinalen Gencluster ermittelt werden. Diese Verteilungsanalyse könnte nach erfolgter Implementierung der Positionsdaten mit dem QM II durchgeführt werden.

Als primäres Sinnesorgan des Menschen zeichnet sich das Auge durch hohe Stoffwechselaktivität und eine Vielzahl an Signaltransduktions-Prozessen aus, weswegen der hohe Anteil an ATP/GTP-bindenden (9,8%), Transport- (12,3%) und Signaltransduktions-Genen (14,1%) sowie Kinasen (6,7%) nicht überrascht. Diese Ergebnisse scheinen auf den ersten Blick höher als vergleichbare Beobachtungen: In einer Microarray-Analyse an einer ARPE-19-Zellkultur wurde ein Anteil von 7,4% Signaltransduktions-Proteinen, 2,8% Transportproteinen, sowie 3,5% Kinasen (Sharma *et al.*, 2005) gemessen, was jedoch auf eine geringere Ausdifferenzierung in der Zellkultur zurückzuführen sein könnte. Generell erwies es sich aufgrund der Verwendung abweichender Suchbegriffe, Clusterungsparameter, Datensätze und unterschiedlicher Schwerpunkte bei der Datenauswertung als schwierig, Expressionsdaten aus verschiedenen Experimenten zu vergleichen. So lieferte eine SAGE-Analyse vierer humanen Retinae einen Anteil von 9,6% an in die Signaltransduktion involvierten Genen sowie 4,2% von Phototransduktions-relevanten Genen (Sharon *et al.*, 2002), wohingegen ein Mikroarray-Experiment an ebenfalls vier Retinae nur einen Anteil von 2% Signatransduktions-Genen erbrachte, aber 12% der Phototransduktion zuordnete (Chowers *et al.*, 2003b). Die Unterscheidung zwischen Signal- und Phototransduktion wurde in der vorliegenden Arbeit nicht getroffen, der Signaltransduktions-Anteil lag hier bei 14,1%, was der Summe beider Teilmengen entspricht.

5.3.6 Ausblick

Aufgrund der Tatsache, dass StackPACK v2.2 aller Voraussicht nach nicht mehr weiterentwickelt werden wird und die vorhandene Version zudem aus Kompatibilitätsgründen die Installation neuerer MySQL-Serien blockiert, wäre es für zukünftige Projekte möglicherweise komfortabler, auf andere Clusterprogramme auszuweichen. Eine sehr verbreitete und laufend aktualisierte Software ist ClustalX (Jeanmougin *et al.*, 1998), wohingegen MUSCLE (Edgar, 2004) oder Kalign (Lassmann und Sonnhammer, 2005) bei

ähnlicher oder besserer Qualität Geschwindigkeitsvorteile gegenüber ClustalX von bis zu Faktor 10 aufweisen. Diese Effizienzunterschiede variieren jedoch stark mit Qualität, Länge und Menge der eingespeisten Sequenzen, weswegen diese Angaben vorsichtig zu rezipieren sind. Aufgrund standardisierter Exportformate steht es nicht zu befürchten, dass diese Programme inkompatibel zu QM II wären, obwohl dies nachzuprüfen bliebe.

Alternativ oder ergänzend kann die Zeit zur Erstellung einer *in silico* Transkript-Datenbank des Auges nicht nur durch Zukauf eines leistungsfähigeren Computers oder durch Installation alternativer Programme gelöst werden, sondern auch durch Verteilen der Rechenleistung auf einen vernetzten Verbund von Rechnern. Hierbei ermittelt ein kleines Programm die freie Rechenkapazität eines Computers im Cluster und übergibt ein entsprechendes Datenpaket an diesen. Der QM II ist grundsätzlich plattformunabhängig und clustertauglich programmiert worden, daher müsste nur ein Clusterprogramm wie z.B. wackamole 2.1.1 (<http://www.backhand.org/wackamole/>) installiert werden, um die Rechenleistung zu vervielfachen.

Ein großer Schritt zum Verständnis des Retinoms wäre zweifelsfrei die Einbindung der annotierten Resultate in Expressionsprofil-Datenbanken, um eine Interaktion mit anderen Genen zu verstehen. Zwei interessante Datenbanken für einen solchen Ansatz wären NCBI GEO (Barrett *et al.*, 2005) oder Affymetrix (Sharma *et al.*, 2005). Hierfür müssten lediglich Browser- und Parsingscripte der Unigene-, bzw. EntrezGene-Module des QM II erweitert werden und die abgefragten Resultate in einem XML-Code vorliegen, also die Konnektivität gewährleistet sein, damit eine Einbindung in die bestehende MySQL-Datenbank problemlos vonstatten gehen könnte.

Exemplarisch wurde ein möglicher zweiter Locus für das humane *OPAI*-Gen gesucht. Nach einem Update der „*nr-eye*“-Datenbank sollen zukünftig auch andere Kandidatengene für Erkrankungen des menschlichen Auges durch Verknüpfung bekannter Determinaten wie z.B. genetische Ontologien und chromosomale Positionen untersucht werden, wie sie in der RETNET-Datenbank zu finden sind.

Der QM II ist nicht beschränkt auf die Verarbeitung humaner retinaler Sequenzen, sondern könnte z.B. auch für andere Organismen oder Gewebe analoge Datenbanken erstellen. Weiterhin sind trotz mittlerweile guter Filtersysteme eukaryontische Datenbanken noch

Diskussion

immer mit falsch annotierten bakteriellen Kontaminationen versehen (PD Reinhold Brückner, Mikrobiologie Kaiserslautern, pers. Mitteilung). Durch entsprechende Auswahl eines Filters für das Favoritenscript des QM II könnte eine solche Verunreinigung im Sinne eines Qualitätsmanagements erkannt werden. Gleiches gilt z.B. für nicht ausreichend entfernte Vektorenreste. Weiterhin könnte dasselbe genetische Ausgangsmaterial für die Erstellung einer Homologiedatenbank genutzt werden, indem in einem Paralleldurchlauf nach der Blast-Analyse durch ein anderes Favoritenscript präferentiell nach beispielsweise Maus-Homologien selektiert wird. Durch Verbinden beider Tabellen und entsprechende Filtersets könnte nach Funktionsunterschieden oder -ähnlichkeiten gesucht werden. Dies ist deshalb interessant, da viele Krankheiten erst über ein Homolog aus anderen Organismen identifiziert werden (Delettre *et al.*, 2000).

Zusammenfassung

Zielsetzung der vorliegenden Arbeit war die Identifizierung und Katalogisierung retinal exprimierter Gene. Dies sollte anhand dreier Teilprojekte verwirklicht werden.

Zunächst wurde ausgehend von isolierter mRNA aus WERI *Rb1*-Zellkulturen eine physikalische cDNA-Bank erstellt, deren Sequenzen als Grundlage für einen retinaspezifischen Expressionschip dienen sollten.

Hierfür wurden initial 960 Klone präpariert und sequenziert. Diese Sequenzen wurden zu Beginn des Projektes manuell, später mit Hilfe des Querymaster I-Moduls einer Blast-Analyse unterzogen und die Ergebnisse tabellarisch ausgewertet. Die auftretenden Sequenzierausfälle wurden durch Optimierung der Wachstums- und Extraktionsprotokolle auf knapp 2% gesenkt. Nach Abzug weiterer 16% informativer Ausfälle (leere Vektoren, mitochondriale Sequenzen, zu kurze Inserts, etc.) blieben 781 prinzipiell informative Klone übrig, von denen 466 Sequenzen nicht redundant vorlagen. Von diesen wurden die zugehörigen Blast-, LocusLink- und Unigene-Einträge tabellarisch erfasst und ausgewertet. 25,5% wiesen zum Zeitpunkt der Analyse eine Homologie zu nicht vollständig annotierten Sequenzen wie cDNA-Klonen, hypothetischen Proteinen, etc. auf. Diese stellen grundsätzlich interessante Kandidatengene für bislang nicht klonierte erbliche Augenkrankheiten dar. Da jedoch keine gänzlich unbeschriebenen Sequenzen ermittelt werden konnten, wurde dieses Teilprojekt zugunsten einer *in silico* cDNA-Bibliothek des Auges aufgegeben. Auf eine Vervollständigung der Sequenzen durch RACE-Experimente oder initiale *full-length*-Klonierung sowie eine Senkung der Redundanzen durch Normalisierung wurde daher ebenfalls verzichtet.

Ein zweiter Ansatz sollte den Einfluss hypoxischer Bedingungen auf die Genexpression in WERI *Rb1*-Zellen untersuchen, da Sauerstoffmangel als Ursache oder begünstigender Faktor vieler Erkrankungen des menschlichen Auges wie AMD, des Glaukoms, oder diabetischer Retinopathie diskutiert wird. Als Vorversuch zur geplanten Globalanalyse durch Mikroarrays wurden die Expressionslevel von 18S-rRNA, *GAPDH*, *EPO* und *VEGF* unter normoxischen sowie hypoxischen Bedingungen nach 8, bzw. 24 Stunden Exposition verglichen. Während 18S-rRNA als Standard diente, sollte die Expression der anderen Gene unter Sauerstoffmangel signifikant durch den Hypoxie-induzierten Transkriptionsfaktor *HIF-1* verstärkt werden. Entgegen den Erwartungen zeigte sich in den meisten Fällen nur eine

geringe Regulation. Signifikant waren lediglich eine Verringerung des Expressionslevels von *GAPDH* sowie von *VEGF* nach 24 Stunden unter Normbedingungen. Die Gründe hierfür sind unbekannt, könnten aber u.a. darin begründet liegen, dass das Modellsystem Retinoblastom im Vergleich zur Netzhaut möglicherweise insgesamt weniger sensitiv auf Sauerstoffmangel reagiert. Weiterhin könnten veränderte NO_x-Konzentrationen u.U. zu einer Reprimierung von *HIF-1*-regulierten Genprodukten geführt haben. Es ist denkbar, dass die Absenkung der *GAPDH*-Transkriptmenge auf einen Verbrauch der Glucose im Anzuchtmedium, bzw. die Anhäufung hemmender Metabolite zurückzuführen ist. Aufgrund der Daten aus diesem Vorversuch unterblieb eine Analyse des Einflusses der Hypoxie auf die Genregulation in der menschlichen Netzhaut auf der Basis eines Expressionschips.

Das dritte Teilprojekt bildete den Hauptteil der vorliegenden Arbeit: zur Analyse des Retinoms und Identifizierung möglicher Kandidatengene für erbliche Erkrankung der Netzhaut sollte eine vollständig annotierte *in silico*-Datenbank des menschlichen Auges erstellt werden. Hierfür wurden in zwei Ansätzen 180.785 (TRJ8), bzw. 194.935 (TRV1) entsprechende ESTs aus der NCBI-Datenbank importiert und mittels des StackPACK-Programms geclustert. Zur Auswertung beider Ansätze wurde lokal eine Annotationssoftware entwickelt, welche jede Sequenz aus der geclusterten Stapeldatei einer Blast-Abfrage unterzog, aus der jeweiligen Blast-Resultatsliste den informativsten „best hit“ auswählte, ausgehend von diesem die Unigene- sowie die LocusLink, bzw. die EntrezGene-Datenbanken des NCBI angesteuerte und schließlich ausgewählte Ergebnisinhalte in eine MySQL-Datenbank überführte. Hierbei wurde der erste Ansatz gegen die *non redundant*-Bibliothek geblastet, was Rechenzeiten von etwa einem Monat nach sich zog. Unter Verwendung eines verbesserten Blast-Algorithmus` sowie unter Ansteuerung der kleineren humanen RefSeq-Datenbank konnte die Berechnungszeit auf ca. ein Zehntel gesenkt werden.

Insgesamt wurden bei der Analyse der TRJ8-Datenbank 45.692 Blastresultate gezählt, von denen 22.282 einen LocusLink-Eintrag und 24.966 einen Unigene-Eintrag aufwiesen. Durch zwischenzeitliche qualitative Verbesserungen der NCBI-Datenbanken konnten für den TRV1-Ansatz folgende Zuordnungen getroffen werden: Von den 48.969 Ergebnissen aus der Blast-Analyse hatten 39.630 sowohl Einträge für Unigene als auch für EntrezGene, 1.589 nur ein EntrezGene-Ergebnis, bei 5.130 Datensätzen lag lediglich der Unigene-Verweis vor, und bei 2.620 Blast-Resultaten konnte weder ein Unigene noch ein Entrez-Eintrag gefunden werden.

Bei der Analyse der Blast-Ergebnisse zeigte sich auch nach der Clusterung noch ein hohes Maß an Redundanz, da StackPACK in der Grundeinstellung Cluster in Subcluster aus sehr ähnlichen, aber nicht identischen Sequenzen (z.B. Isoformen) unterteilt und diese exportiert. Hierbei lagen etwa 70% der im Auge exprimierten Transkripte mindestens doppelt vor. Die Eliminierung dieser Redundanzen erfolgte durch eine Filteroperation innerhalb der Datenbank („MySQL-Clustern“). Die ausgeblendeten Subcluster bleiben somit für eventuell nachfolgende Analysen erhalten, weswegen auf eine Änderung der Exportparameter verzichtet wurde. Insgesamt lagen 11.493 nicht-redundante Sequenzen vor, die im menschlichen Auge exprimiert werden. Diese wiesen unterschiedliche Annotationsgrade auf, z.B. lag bei allen gelisteten Genen eine chromosomale Lokalisierung vor, jedoch nur zu 80% ein Ontologie-Kommentar und nur zu 60% einen MIM-Link. Diese Sequenzen wurden für weitere Analysen zur „nr-eye“-Datenbank vereint.

„Nr-eye“ besteht aus MSCs und singulären Sequenzen, wobei letztere in der Blast-Analyse erwartungsgemäß kürzere *Alignments* und eine geringere Informationsdichte erbrachten. Dennoch waren die Qualitätsunterschiede nicht groß genug, um diese grundsätzlich vor einer Datenbankerstellung auszuschließen. Exemplarisch für die Möglichkeiten der Datenbankanalyse wurde die chromosomale Verteilung der im menschlichen Auge exprimierten Gene untersucht, welche ohne die Einbindung physikalischer Positionsdaten keine Expressionscluster aufschlüsseln konnte. Für die Chromosomen 1-19 ergab sich eine relativ homogene Verteilung retinaler Gene zwischen 46-58%. Hierbei ist unklar, ob die „nr-eye“-Datenbank die tatsächlichen Verhältnisse im menschlichen Auge widerspiegelt: Eine Verteilungsanalyse der gelisteten Gene hinsichtlich ihrer Funktion deckte sich nur zum Teil mit Vergleichswerten aus der Literatur, wobei auch diese untereinander divergierten. Dies ist vor allem auf unterschiedliche Meßmethoden und Datenquellen sowie uneinheitliche Ontologie-Aufschlüsselungen zurückzuführen. Abschließend erbrachte eine verknüpfte Abfrage anhand dreier bekannter Determinanten 29 Kandidatengene für einen möglichen weiteren Locus der Optikusatrophie Typ Kjer.

Zusammenfassend zeigte sich die Programmpipeline von StackPACK über den Querymaster II bis hin zur MySQL-Datenbank als ein intuitiv bedienbares und effizientes Instrument, welches das annotierte Transkriptom des menschlichen Auges halbautomatisiert und in angemessener Geschwindigkeit und Qualität darstellen konnte.

7. Literaturverzeichnis

- Adams, M.D., Kelley, J.M., Gocayne, J.D., Dubnick, M., Polymeropoulos M.H., Xiao, H., Merril, C.R., Wu, A., Olde, B., Moreno, R.F., Kervelage, A.R., McConbie, W.R., Venter, J.C.** (1991) Complementary DNA Sequencing: Expressed Sequence Tags and Human Genome Project. *Science*; **252**: 1651-1656.
- Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A., Bult, C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D., White, O., Sutton, G., Blake, J.A., Brandon, R.C., Chiu, M.W., Clayton, R.A., Cline, R.T., Cotton, M.D., Earle-Hughes, J., Fine, L.D., Fitz-Gerald, L.M., FitzHugh, W.M., Fritchman, J.L., Geoghagen, N.S.M., Glodek, A., Gnehm, C.L., Venter, C., et al.** (1995) Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. *Nature*; **377 (suppl.)**: 3-17.
- Altschul, S.F., Gish, W., Miller, Myers, E.W., Lipman, D.J.** (1990) Basic local alignment search tool. *J. Mol. Biol.*; **215**: 403-410.
- Arden, G. B., Sidman, R. L., Arap, W., Schlingemann, R. O.** (2005) Spare the rod and spoil the eye. *Br. J. Ophthalmol.*; **89**: 764-769.
- Barrett, T., Suzek, T.O., Troup, D.B., Wilhite, S.E., Ngau, W.C., Ledoux, P., Rudnev, D., Lash, A.E., Fujibuchi, W., Edgar, R.** (2005) NCBI GEO: mining millions of expression profiles-database and tools. *Nucleic Acids Research*; **33**: D562-566.
- Beck, I., Ramirez, S., Weinmann, R., Caro, J.** (1991) Enhancer element at the 3'-flanking region controls transcriptional response to hypoxia in the human erythropoietin gene. *J. Biol. Chem.*; **266**: 15563-15566.
- Bernstein, S.L., Kutty, G., Wiggert, B., Albert, D.M., Nickerson, J.M.** (1994) Expression of retina-specific genes by mouse retinoblastoma cells.
- Behme, H., Mintert, S.** (2000) XML in der Praxis – Extensible Markup Language für Profis. (<http://www.linkwerk.com/pub/xmlidp/2000/>)
- Birney, E., Durbin, R.** (2000) Using GeneWise in the Drosophila annotation experiment. *Genome Research*; **10**: 547-548.
- Bishop, J.O., Morton, J.G., Rosbash, M., Richardson, M.** Three abundance classes in HeLa cell messenger (1974) *Nature*; **250**: 199-204.
- Birnboim, H.C., Doly J.** (1979) A rapid alkaline lysis procedure for screening recombinant plasmid DNA. *Nucl. Acids Res*; **7**: 1513 – 1522.
- Birnboim, H.C.** (1983) A rapid alkaline lysis extraction method for the isolation of plasmid DNA. *Methods Enzymol*; **100**: 243-255.
- Boguski, M.S., Schuler, G.S.** (1995) ESTablishing a human transcript map. *Nature Genetics*; **10**: 369-371.

- Bonaldo, M.F., Lennon, G., Soares, M.B.** (1996) Normalization and Subtraction: Two approaches to facilitate gene discovery. *Genome Research*; **6**: 791-806.
- Boulton, M., Foreman, D., Williams, G., McLeod, D.** (1998) VEGF localisation in diabetic retinopathy. *Br. J. Ophthalmol.*; **82**: 561-568.
- Bracken, C.P., Whitelaw, M.L., Peer, D.J.** (2003) The hypoxia-inducible factors: key transcriptional regulators of hypoxic responses. *Cell Mol. Life Sci.*; **60**: 1376-1393.
- Brentani, H., Caballero, O.L., Camargo, A.A., da Silva, A.M., da Silva, W.A. Jr, Dias Neto E, Grivet, M., Gruber, A., Guimaraes, P.E., Hide, W. et al.** (2003) The generation and utilization of a cancer-orientated representation of the human transcriptome by using expressed sequence tags. *Proc. Natl. Acad. Sci.*; **100**: 13418-13423.
- Brinchmann-Hansen, O., Dahl-Jorgensen, K., Sandvik, L., Hanssen, K.F.** (1992) Blood glucose concentration and progression of diabetic retinopathy; the seven years result of the Oslo studies. *Brit. Med. J.*; **304**: 19-22.
- Bunn, H.F., Poyton, R.O.** (1996) Oxygen sensing and molecular adaptation to hypoxia. *Physiol. Rev.*; **76**: 839-885.
- Burge, C., Karlin, S.** (1997) Prediction of complete gene structure in human genomic DNA. *J. Mol. Biol.*; **268**: 78-94.
- Burke, J.P., Wang, W. Hide, W., Dawson D.** (1998) Alternative gene form discovery and candidate gene selection from gene indexing projects. *Genome Research*; **8**: 276-290.
- Burke, J.P., Davison, D., Hide, W.** (1999) d2_cluster: a validated method for clustering EST and full-length cDNA sequences. *Genome Research*; **9**: 1135-1142.
- Bustin, S.** (2000) Absolute quantification of mRNA using Real-time reverse transcription polymerase chain reaction assays. *J. Mol. Endocrinol.*; **25**: 169-193.
- Camon, E., Magrane, M., Barrell, D., Binns, D., Fleischmann, W., Kersey, P., Mulder, N., Oinn, T., Maslen, J., Cox, A., Apweiler, R.** (2003) The Gene Ontology Annotation (GOA) project: implementation of GO in Swiss-Prot, TrEMBL and InterPro. *Genome Research*; **4**: 662-672.
- Carninci, P., Kvam, C., Kitamura, A., Ohsumi, T., Okazaki, Y., Itoh, M., Kamiya, M., Shibata, K. Sasaki, N., Izawa, M., Muramatsu, M., Hayashizaki, Y., Schneider, C.** (1996) High-efficiency full-length cDNA cloning by biotinylated CAP trapper. *Genomics*; **1**: 327-336.
- Chandel, N.S., McClintock, D.S., Feliciano, C.E., Wood, T.M., Melendes, J.A., Rodriguez, A.M., Schumacker, P.T.** (2000) Reactive oxygen species generated at mitochondrial complex III stabilize hypoxia-inducible factor-1alpha during hypoxia: a mechanism for O₂ sensing. *J. Biol. Chem.*; **275**: 25130-25138.

- Chapman-Smith A., Lutwyche J.K., Whitelaw M.L.** (2004) Contribution of the Per/Arnt/Sim (PAS) domains to DNA binding by the basic helix-loop-helix PAS transcriptional regulators. *J. Biol. Chem.*; **7**: 5353-62.
- Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T.J., Higgins, D.G., Thompson, D.G.** (2003) Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Research*; **31**: 3497-3500.
- Chou A., Burke J.** (1999) CRAWview: for viewing splicing variation, gene families, and polymorphism in clusters of ESTs and full-length sequences. *Bioinformatics*; **15**: 376-381.
- Chowers, I., Liu, D., Farkas, R.H., Gunatilaka, T.L., Hackam, A.S., Bernstein, S.L., Campochiaro, P.A., Parmigiani, G., Zack, D.J.** (2003) Gene expression variation in the adult human retina. *Hum. Mol. Genet.*; **22**: 2881-2893.
- Chowers, I., Gunatilaka, T.L., Farkas, R.H., Qian, J., Hackam, A.S., Duh, E., Kageyama, M., Wang, C., Vora, A., Campochiaro, C.A., Zack, D.J.** (2003) Identification of novel genes preferentially expressed in the retina using a custom human retina cDNA microarray. *Invest. Ophthalm. Vis. Sci.*; **44**: 3732-3741.
- Christoffels, A., van Gelder, A., Greyling, G., Miller, R., Hide, T., Hide, W.** (2001) STACK: Sequence Tag Alignment and Consensus Knowledgebase. *Nucleic Acids Research*; **29**: 234-238.
- Chun, Y.S., Kim, M.S., Park, J.W.** (2002) Oxygen-dependent and -independent regulation of HIF-1 α . *J Korean Med Sci*; **17**: 581-8.
- Curwen, V., Eyraas, E., Andrews, T.D., Clarke, L., Mongin, E., Searle, S.M., Clamp, M.** (2004) The Ensembl automatic gene annotation system. *Genome Research*; **5**: 942-950.
- Czyzyk-Krzeska, M.F.** (1997) Molecular aspects of oxygen sensing in physiological adaptation to hypoxia. *Respir. Physiol.*; **110**: 99-111.
- Daiger S.P., Rossiter B.F., Greenberg J., Christoffels A., Hide W.** (1998) Data services and software for identifying genes and mutations causing retinal degeneration. *Invest. Ophthalmol. Vis. Sci.*; **39**: S295.
- Davidson, E.H., Britten, R.J.** (1979) Regulation of gene expression: Possible role of repetitive sequences. *Science*; **204**: 1052-1059.
- Delettre, C., Lenaers, G., Griffoin, J.M., Gigarel, N., Lorenzo, C., Belenguer, P., Pelloquin, L., Grosgeorge, J., Turc-Carel, C., Perret, E., Astarie-Dequeker, C., Lasquellec, L., Arnaud, B., Ducommun, B., Kaplan, J., Hamel, C.P.** (2000) Nuclear gene OPA1, encoding a mitochondrial dynamin-related protein, is mutated in dominant optic atrophy. *Nat. Genet.*; **2**: 207-210.
- De Magalhaes, J.P., Costa, J., Toussaint, O.** (2005) HAGR: The human ageing genomic resources. *Nucleic Acids Research*; **33**: 537-543.

- Diehn, J.J., Diehn, M., Marmor, M.F., Brown, P.O.** (2005) Differential gene expression in anatomical compartments of the human eye. *Genome Biology*; **6**: R74.
- Ditzel, J., Standl, E.** (1975) The oxygen transport system of red blood cells during diabetic ketosis and recovery. *Diabetologia*; **4**: 255-260.
- Edery, I., Chu, L.L., Sonenberg, N., Pelletier, J.** (1995) An efficient strategy to isolate full-length cDNA based on the mRNA-cap retention procedure (CAPture). *Mol. Cell. Biol.* **15**: 3363-3371.
- Edgar, R.C.** (2004) MUSCLE a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*; **5**: 113.
- Ema, M., Taya, S., Yokotani, N., Sogawa, K., Matsuda, Y., Fujii-Kuriyama, Y.** (1997) A novel bHLH-PAS factor with close sequence similarity to hypoxia-inducible factor I alpha regulates the VEGF expression and is potentially involved in lung and vascular development. *Proc. Natl. Acad. Sci.*; **94**: 4273-4278.
- Ernens, I., Goodfellow, S.J., Innes, F., Kenneth, N.S., Derblay, L.E., White, R.J., Scott, P.H.** (2006) Hypoxic stress suppresses RNA polymerase III recruitment and tRNA gene transcription in cardiomyocytes. *Nucleic Acids Research*; **1**: 286-294.
- Fonn, D., Sweney, D., Holden, B.A., Cavanagh, D.** (2005) Corneal oxygen deficiency. *Eye contact lens.*; 23-27.
- Frohman, M.A.** Rapid amplification of complementary DNA-ends for generation of full-length complementary DNAs: thermal RACE. *Methods Enzymol.*; **218**: 340-356.
- Gelfand, M.S., Mironov, A.A., Pevzner, P.A.** (1996) Gene recognition via spliced sequence alignment. *Proc. Natl. Acad. Sci.*; **93**: 9061-9066.
- Gerhard, D.S., Wagner, L., Feingold, E.A., Shenmen, C.M., Grouse, L.H., Schuler, G., Klein, S.L., Old, S., Rasooly, R., Good, P., Guyer, M., Peck, A.M., Derge, J.G., Lipman, D., Collins, F.S., Jang, W., Sherry, S., Feolo, M., Misquitta, L., Lee, E., Rotmistrovsky, K., Greenhut, S.F., Schaefer, C.F., Buetow, K., Bonner, T.I., Haussler, D., Kent, J., Kiekhaus, M., Furey, T., Brent, M., Prange, C., Schreiber, K., Shapiro, N., Bhat, N.K., Hopkins, R.F., Hsie, F., Driscoll, T., Soares, M.B., Casavant, T.L., Scheetz, T.E., Brown-stein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Piao, Y., Dudekula, D.B., Ko, M.S., Kawakami, K., Suzuki, Y., Sugano, S., Gruber, C.E., Smith, M.R., Simmons, B., Moore, T., Waterman, R., Johnson, S.L., Ruan, Y., Wei, C.L., Mathavan, S., Gunaratne, P.H., Wu, J., Garcia, A.M., Hulyk, S.W., Fuh, E., Yuan, Y., Sneed, A., Kowis, C., Hodgson, A., Muzny, D.M., McPherson, J., Gibbs, R.A., Fahey, J., Helton, E., Ketteman, M., Madan, A., Rodrigues, S., Sanchez, A., Whiting, M., Madari, A., Young, A.C., Wetherby, K.D., Granite, S.J., Kwong, P.N., Brinkley, C.P., Pearson, R.L., Bouffard, G.G., Blakesly, R.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Griffith, M., Griffith, O.L., Krzywinski, M.I., Liao, N., Morrin, R., Palmquist, D., Petrescu, A.S., Skalska, U., Smailus, D.E., Stott, J.M., Schnerch, A., Schein, J.E., Jones, S.J., Holt, R.A., Baross, A., Marra, M.A., Clifton, S., Makowski, K.A., Bosak, S., Malek, J.; MGC Project Team.** (2004) The status, quality, and expansion of the NIH

- full-length cDNA project: the Mammalian Gene Collection (MGC). *Genome Res.*; **10B**: 2121-2127.
- Gocayne J., Robinson D. A., FitzGerald M. G., Chung F. Z., Kerlavage A. R., Lentes, K. U., Lai J., Wang C. D., Fraser C. M., Venter J. C.** (1987) Primary structure of rat cardiac beta-adrenergic and muscarinic cholinergic receptors obtained by automated DNA sequence analysis: further evidence for a multigene family. *Proc. Natl. Acad. Sci.*; **84**: 8296-8300.
- Gora-Kupilas, K., Josko, J.** (2005) The neuroprotective function of vascular endothelial growth factor (VEGF). *Folia Neuropathol.*; **1**: 31-39.
- Gordon, D., Abajian, C., Green, P.** (1998) Consed: a graphical tool for sequence finishing. *Genome Research*; **8**: 195-202.
- Green, P.** (1996) PHRAP
(<http://www.genome.washington.edu/uwgc/analysistools/phrap.htm>.
phg@u.washington.edu.).
- Grimm, C., Wenzel, A., Groszner, M., Mayser, H., Seeliger, M., Samardzija, M., Bauer, C., Gassmann, M., Reme, C.E.** (2002) HIF-1-induced erythropoietin in the hypoxic retina protects against light-induced retinal degeneration. *Nat. Med.*; **7**: 718-724.
- Hara, S., Hamada, J., Kobayashi, C., Kondo, Y., Imura, N.** (2001) Expression and characterization of hypoxia-inducible factor (HIF)-3alpha in human kidney: suppression of HIF-mediated gene expression by HIF-3alpha. *Biochem. Biophys. Res. Commun.*; **287**: 808-813.
- Hide, W., Burke J., Davison D.** (1994) Biological evaluation of d2, an algorithm for high performance sequence comparison. *J. Comp. Biol*; **1**: 199-215.
- Higuchi, R., Fockler, C., Dolinger, G., Watson, R.** (1993) Kinetic PCR: Real-time monitoring of DNA amplification reactions. *Biotechnology*; **11**: 1026-1030.
- Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston, R., Williamson, A., Wohldmann, P., Wilson, R.** (1996) Generation and analysis of 280,000 human expressed sequence tags. *Genome Research*; **6**: 807-828.
- Hochachka, P.W., Buck, L.T., Doll, C.J., Land, S.C.** (1996) Unifying theory of hypoxia tolerance: Molecular/metabolic defense and rescue mechanisms for surviving oxygen lack. *Proc. Natl. Acad. Sci.*; **93**: 9493-9498.
- Hochachka, P.W., Lutz, P.L.** (2001) Mechanism, origin and evolution of anoxia tolerance in animals. *Comp. Biochem. Physiol. B. Biochem. Mol. Biol.*; **130**: 435-459.
- International humane genome sequencing consortium.** (2001) Initial sequencing and analysis of the human genome. *Nature*; **409**: 860-921.

- International humane genome sequencing consortium.** (2004) Finishing the euchromatic sequence of the human genome. *Nature.*; **431**: 931-945.
- Jeanmougin, F., Thompson, J.D., Gouy, M., Higgins, D.G., Gibson, T.J.** (1998) Multiple sequence alignment with ClustalX. *Trends Biochem. Sci.*; **10**: 403-405.
- Jelkmann, W.** (2004) Molecular biology of erythropoietin. *Intern Med.*; **8**: 649-659.
- Johnson, R.A. and Wichern, D.W.** (1994) Applied multivariate statistical analysis 3rd ed. Prentice-Hall Englewood Cliffs, NY.
- Kaushik, S., Pandav, S. S., Ram, J.** (2003) Neuroprotection in glaucoma. *J. of Postgrad. Med.*; **49**: 90-95.
- Katsanis, N., Worley, K.C., Gonzalez, G., Ansley, S.J., Lupski, J.R.** (2002) A computational/functional genomics approach for the enrichment of the retinal transcriptome and the identification of positional candidate retinopathy genes. *Proc. Natl. Acad. Sci.*; **22**: 14326-14331.
- Kern, T. S., Engerman, R. L.** (1996) Capillary lesions develop in retina rather than in cerebral cortex in diabetes and experimental galactosaemia. *Arch. Ophthalmol.*; **114**: 306-310.
- Kimura, H., Weisz, A., Kurashima, Y., Hashimoto, K., Ogura, T., D'Acquisto, F., Addeo, R., Makuuchi, M., Esumi, H.** (2000) Hypoxia response element of the human vascular endothelial growth factor gene mediates transcriptional regulation by nitric oxide: control of hypoxia-inducible factor-1 activity by nitric oxide. *Blood*; **1**: 189-197.
- Köhler, J.** (2004) Integration of live science databases. *Drug Discov. Today: BioSilico*; **2**: 61-69.
- Köhler, J., Rawlings, C., Verrier, P., Mitchell, R., Skusa, A., Ruegg, A., Philippi, S.** (2005) Linking experimental results, biological networks and sequence analysis methods using ontologies and generalised data structures. *In Silico Biol.*; **1**: 33-44.
- Kuribayashi-Otha, K., Tamatsukuri, S., Hikata, M., Myamoto, C., Furuichi, Y.** (1993) Application of the oligo(dT)30-latex for rapid purification of poly (A)+ mRNA and for hybrid subtraction with in situ reverse transcribed cDNA. *Biochimica et Biophysica Acta*; **1156**: 204-212.
- Lassmann T., Sonnhammer, E.L.L.** (2005) Kalign – an accurate and fast multiple sequence alignment algorithm. *BMC Bioinformatics*; **6**: 298.
- Lee, L.G., Connell, C.R., Bloch, W.** (1993) Allelic discrimination by nick translation PCR with flourogenic probes. *Nucleic Acids Research*; **11**: 3761-3766.
- Lin, L.D., Chao, T.D.** (2005) Vascular endothelial growth factor in thyroid cancers. *Cancer Biother. Radiopharm.*; **6**: 648-661.

- Liu, L., Celeste Simon, M.** (2004) Regulation of transcription and translation by hypoxia. *Cancer Biology & Therapy*; **6**: 492-497.
- Livak, K.J., Flood, S.J., Marmaro, J., Giusti, W., Deetz, K.** (1995) Oligonucleotides with fluorescent primers dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. *PCR Methods Appl.*; **4**: 357-362.
- Livak, K.J., Schmittgen, T.D.** (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2^{-Delta Delta C(T)} Method. *Methods*; **25**: 402-408.
- Lynch D.R., Dawson T.M.** (1994) Secondary mechanisms in neuronal trauma. *Curr. Opin. Neurol.*; **7**: 510-6.
- Magott, D., Ostell, J., Pruitt, K.D., Tatusova, T.** (2005) Entrez Gene: gene centered information at NCBI. *Nucleic Acids Res.*; **33 (Database Issue)**: D54-D58.
- Mahadevappa, M. and Warrington, J. A.** (1999) A high density probe array method using 10- to 100-fold fewer cells. *Nat. Biotechnol.* **17**: 1134-1136.
- Maslim J., Valter, K., Egensperger, R., Holländer, H., Stone, J.** (1997) Tissue oxygen during a critical developmental period controls the death and survival of photoreceptors. *Invest. Ophthalmol. Vis. Sci.*; **9**: 1667-1677.
- Mc Fall R.C., Sery T.W., Makadon M.** (1977) Charakterisation of a new continuous cell line derived from a human retinoblastoma. *Cancer Res*; **37**: 1003-10.
- Mc Kay, G.J., Campbell, L., Oliver, M., Brockbank, S., Simpson, D.A.C., Curry W.J.** (2004) Preparation of planar retinal specimens: verification by histology, mRNA profiling and proteome analysis. *Mol. Vision*; **10**: 240-247.
- Megy, K., Audic, S., Claverie, J.M.** (2002) Herat specific genes revealed by expressed sequence tag (EST) sampling. *Genome Biol.*; **12**: 0074.1-0074.11.
- Mons, B.** (2005) Which gene did you mean? *BMC Bioinformatics*; **6**: 142.
- Mullis KB, Faloona FA** (1987) Specific synthesis of DNA in vitro via a polymerase-catalysed chain reaction. *Methods Enzymol*; **155**: 335-350.
- Naskar R, Dreyer E.B.** (2001) New horizons in neuroprotection. *Surv. Ophthalmol.*; **45 (Suppl 3)**: 250-256.
- Ohara, O., Temple, G.** (2001) Directional cDNA library construction assisted by the in vitro recombination reaction. *Nucleic Acids Research*; **29**: e22.
- Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C., Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusica, V., Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani T.A., Fletcher, C.F., Forrest, A.,**

- Frazer, K.S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A., Kawaji, H., Kawasaki Y., Kedzierski, R.M., King, B.L., Konagaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W.J., Perte, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E., Hayashizaki, Y.; FANTOM Consortium; RIKEN Genome Exploration Research Group Phase I & II Team. (2002) Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature*. **420**: 563-73.
- Ousterhout, J.K.** (1995) Tcl und Tk. Addison Wesley Publishing Company, München.
- Page, R., Stevens, R.C.** (2004) Crystallization data mining in structural genomics: using positive and negative results to optimize protein crystallization screens. *Methods*; **3**: 373-389.
- Palmer, L. A., Semenza, G. L., Stoler, M. H., Johns, R. A.** (1998) Hypoxia induces type II NOS gene expression in pulmonary artery endothelial cells via HIF-1. *Am. J. Physiol.*; **274**: 1212-1219.
- Pearson, W.R.** (1990) Rapid and sensitive sequence comparison with FASTP and FASTA. In Doolittle, R.F. (ed.), *Molecular Evolution: computer analysis of protein and nucleic acid sequences*, *Methods In Enzymol. Academic Press, San Diego*; 63-98.
- Pellegrino, M., Provero, P., Silengo, L., Di Cunto, F.** (2004) CLOE: identification of putative functional relationships among genes by comparison of expression profiles between two species. *BMC Bioinformatics*; **5**: 179.
- Pizzi, C., Bortoluzzi, S., Bisognin, A., Coppe, A., Danieli, G.A.** (2005) Detecting seeding motifs in DNA sequences. *Nucleic Acids Research*; **33**: e135.
- Pontius J.U., Wagner L., Schuler, G.D.** (2003) UniGene: a unified view of the Transcriptome. In: *The NCBI Handbook*. Bethesda (MD): National Center for Biotechnology Information.
- Pruitt, K., Tatusova, T., Maglott, D.** (2005) NCBI Reference Sequence Project. *Nucleic Acids Research*; **33**: D501-D504.

- Pugh, C.W., Tan, C.C., Jones, R.W., Ratcliffe, P.J.** (1991) Functional analysis of an oxygen-regulated transcriptional enhancer lying 3' to the mouse erythropoietin gene. *Proc. Natl. Acad. Sci.*; **88**: 10553-10557.
- Razeghi, P., Essop, M.F., Huss, J.M., Abbasi, S., Manga, N., Taegtmeier, H.** (2003) Hypoxia-induced switches of myosin heavy chain iso-gene expression in rat heart. *Biochem. Biophys. Res. Commun.*; **4**: 1024-1027.
- Reinisalo, M., Urtti, A., Honkakoski, P.** (2003) Retina-specific gene expression and improved DNA transfection in WERI-Rb1 retinoblastoma cells. *Biochim. et Biophys. Acta*; **3**: 169-176.
- Riedinger H.J., van Betteraey M, Probst H.** (1999) Hypoxia blocks in vivo initiation of simian virus 40 replication at a stage preceding origin unwinding. *J. Virol.*; **73**: 2243-2252.
- Royds, J.A., Dower, S.K., Qwarnstrom, E.E., Lewis, C.E.** (1998) Response of tumour cells to hypoxia : role of p53 and NFK β . *Mol. Pathol.*; **51**: 55-61.
- Rudd, S., Schoof, H., Mayer, K.** (2005) PlantMarkers – a database of predicted molecular markers from plants. *Nucleic Acids Research*; **33**: 628-632.
- Salceda, S., Caro, J.** (1997) Hypoxia-inducible factor 1 alpha (HIF-1alpha) protein is rapidly degraded by the ubiquitin-proteasome system under normoxic conditions. *J. Biol. Chem.*; **272**: 22642-22647.
- Sambrook, J. et al., eds.** (1989) Molecular cloning: a laboratory manual. 2nd ed., Cold Spring Harbor, N.Y.; Laboratory Press.
- Sanger F., Nicklen S., Coulson A.R.** (1977) DNA Sequencing with Chain-Terminating Inhibitors. *Proc. Natl. Acad. Sci.* **74**: 5463-5467.
- Schena, M., Shalon, D., Davis, R.W., Brown, P.O.** (1996) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*; **270**: 467-470.
- Schulz, H.L., Götz, T., Kaschkötö, J., Weber, B.H.F.** (2004) The retinome – defining a reference transcriptome of the adult mammalian retina/retinal pigment epithelium. *BMC Genomics*; **5**: 50.
- Seeburg, P.H., Shine, J., Martial, J.A., Ullrich, A., Gooman, H., M., Baxter, J.D.** (1977) Nucleotide sequence of the human gene coding for a polypeptide hormone. *Trans. Assoc. Am. Physicans.*; **90**: 109-116.
- Seki, M., Satou, M., Sakurai, T., Akiyama, K., Iida, K., Ishida, J., Nakajima, M., Enju, A., Narusaka, M., Fujita, M., Oono, Y., Kamei, A., Yamaguchi-Shinozaki, K., Shinozaki, K.** (2004) RIKEN Arabidopsis full-length (RAFL) cDNA and its applications for expression profiling under abiotic stress conditions. *J. Exp.Botany*; **395**: 213-223.

- Semenza, G.L., Nejfelt, M.K., Chi, S.M., Antoarakis, S.E.** (1991) Hypoxia-induced nuclear factors bind to an enhancer element located 3' to the human erythropoietin gene. *Proc. Natl. Acad. Sci.*; **88**: 5680-5684.
- Semenza, G.L., Wang, G. L.** (1992) A nuclear factor induced by hypoxia via de novo protein synthesis binds to the human erythropoietin gene enhancer at a site required for transcriptional activation. *Mol. Cell. Biol.*; **12**: 5447-5454.
- Semenza, G.L.** (2000) HIF1: Mediator of physiological and pathophysiological responses to hypoxia. *J. Appl., Physiol.*; **88**: 1474-80.
- Sharma R.K., Orr, W.E., Schmitt, A.D., Johnson D.A.** (2005) A funktional profile of gene expression in ARPE-19 cells. *BMC Ophthalm.*; **5**: 25.
- Sharon, D., Blackshaw, S., Cepko, C.L., Dryja, T.P.** (2002) Profile of the genes expressed in the human peripheral retina, macula, and retinal pigment epithelium determined through serial analysis of gene expression (SAGE). *Proc. Acad. Natl. Sci.*; **99**: 315-320.
- Shweiki, D., Itin, A., Soffer, D., Keshet, E.** (1992) Vascular endothelial growth factor induced by hypoxia may mediate hypoxia-initiated angiogenesis. *Nature*; **359**: 843-845.
- Signoretti, S., Waltregny, D., Isaac, B., Lin, D., Garraway, L., Yang, A., Montironi, R., McKeon, F., Loda, M.** (2000) p63 is a prostate basal cell marker and is required for prostate development. *Am J. Pathol.*; **157**: 1769- 1775.
- Sinha, S., Tompa, M.** (2003) YMF: a program for discovery of novel transcription factor binding sites by statistical overrepresentation. *Nucleic Acids Research*; **31**: 3586-3588.
- Souazè, F., Ntodou-Tomé, A., Tran, C.Y., Rostene, W., Forgez, P.** (1996) Quantitative RT-PCR : limits and accuracy. *Biotechniques*; **21**: 280-285.
- Spilsbury K., Garrett K.L., Shen W. Y., Constable I.J., Rakoczy P.E.** (2000) Overexpression of vascular endothelial growth factor (VEGF) in the retinal pigment epithelium leads to the development of choroidal neovascularization. *Am. J. Pathol.*; **1**: 135-144.
- Stanke, M., Waack, S.** (2003) Gene prediction with a hidden Markow model and a new intron submodel. *Bioinformatics*; **19 (Suppl. 2)**: II215-225.
- Strauss, E. C., Kobori, J. A., Siu, G., Hood, L. E.** (1986) Specific-primer-directed DNA sequencing. *Anal. Biochem.*; **154**: 353-360.
- Sutton, G., White, O. Adams, M.D., Kerlavage, A.R.** (1995) TIGR assembler: a new tool for assembling large shotgun sequencing projects. *Genome Sci. Technol.*; **1**: 9-18.
- Suzuki, T., Higgins, P.J., Crawford, D.R.** (2000) Control selection for RNA quantitation. *Biotechniques*; **2**: 332-337.

- Swaroop, A., Zack, D.** (2002) Transcriptome analysis of the retina. *Genome Biol.*; **8**: 1022.1-1022.4.
- Tanino, M., Debily, M.A., Tamura, T., Hishiki, T., Ogasawara, O., Murakawa, K., Kawamoto, S., Itoh, K., Watanabe, S., de Souza, S.J., Imbeaud, S., Graudens, E., Eveno, E., Hilton, P., Sudo, Y., Kelso, J., Ikeo, K., Imanishi, T., Gojobori, T., Auffray, C., Hide, W., Okubo, K.** (2005) The Human Anatomic Gene Expression Library (H-ANGEL), the H-Inv integrative display of human gene expression across disparate technologies and platforms. *Nucleic Acids Research*; **33**: 567-572.
- The Institute for Genomic Reserch.** (2003) TGI clustering tools (<http://www.tigr.org/software>)
- Van den Enden, M.K., Nyengaard, J.R., Ostrow, E., Burgan, J.H., Williamson, J.H.** (1996) Elevated glucose levels increase retinal glycolysis and sorbitol pathway metabolism. *Invest. Ophthalmol. Vis. Sci.*; **53**: 2968-2975.
- Vasmatis, G., Essand, M., Brinkmann, U., Lee, B., Pastan, I.** (1998) Discovery of three genes specifically expressed in human prostate by expressed sequence tag database analysis. *Proc. Natl. Acad. Sci.*; **95**: 300-304.
- Velculescu, V.E., Zhang, L., Vogelstein, B., Kinzler, K.W.** (1995) Serial analysis of gene expression. *Science*; **270**: 484-487.
- Venter, J.C., et al.** (2001) The sequence of the human genome. *Science*; **291**: 1304-1351.
- Vigh, B., Manzano, M.J., Zadori, A., Frank, C.L., Lukats, A., Rohlich, P., Szel, A., David, C.** (2002) Nonvisual photoreceptors of the deep brain, pineal organs and retina. *Histol. Histopathol.*; **17**: 555-590.
- Wang, W.** (2000) Programmieren für Dummies. MITP-Verlag, Bonn.
- Wang, Z., Chen, Y., Li, Y.** A brief review of computational gene prediction methods. (2004) *Geno. Prot. Bioinfo.*; **2**: 216-221.
- Webster, T., Francis, A.** (2000) Tcl/Tk für Dummies. MITP-Verlag, Bonn.
- Wistow, G.** (2002) A project for ocular bioinformatics: NEIBank. *Mol. Vision*; **8**: 161-163.
- Wistow G.** (2005) The NEIBank project for ocular genomics: Data-mining gene expression in human and rodent eye tissues. *Prog. Retin. Res.*; [Elektronische Publikation vor Erscheinungsdatum].
- Wheeler, D.L., Chappey, C., Lash, A.E., Leipe, D.D., Madden, T.L., Schuler, G.D., Tatusova, T.A., Rapp, B.A.** (2000) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*; **1**: 10-14.
- Wheeler, D.L., Barrett, T., Benson, D.A., Bryant, S.H., Canese, K., Chetvernin, V., Church, D.M., DiCuccio, M., Edgar, R., Federhen, S., Geer, L.Y., Helmberg, W., Kapustin, Y., Kenton, D.L., Khovayko, O., Lipman, D.J., Madden, T.L., Maglott, D.R., Ostell, J., Pruitt, K.D., Schuler, G.D., Schriml, L.M., Sequeira, E., Sherry,**

- S.T., Sirotkin, K., Souvorov, A., Starchenko, G., Suzek, T.O., Tatusov, R., Tatusova, T.A., Wagner, L., Yaschenko, E.** (2006) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*; **34 (Database issue)**: D173-180.
- Widenius, M., Axmark, D., MySQL AB Documentation Team.** (2002) MySQL – Das offizielle Handbuch. MITP-Verlag Bonn.
- Witmer, A. N., Vrensen, G. F., Van Noorden, C. J., Schlingemann, R.O.** (2003) Vascular endothelial growth factors and angiogenesis in eye disease. *Prog. Retin. Eye Res.*; **1**: 1-29.
- Wolfberg T. G., Landsman, D.** (1997) A comparison of expressed sequence tags (ESTs) to human genome sequences. *Nucleic Acids Research*; **25**: 1626-1632.
- Worley K.C., Wiese, B.A., Smith, R.** (1995) BEAUTY: an enhanced BLAST-based search tool that integrates multiple biological information resources into sequence similarities results. *Genome Res.*; **5**: 173-184.
- Yano, N. Habib, N.A., Fadden, K.J., Yamashita, H., Mitry, R., Jauregui, H., Kane, A., Endoh, M., Rifai, A.** (2001) Profiling the adult human liver transcriptome: analysis by cDNA array hybridization. *J. Hepatol.*; **35**: 178-186.
- Zdobnov, E.M., Lopez, R., Apweiler, R., Eitzold, T.** (2002) The EBI SRS server – recent developments. *Bioinformatics*; **18**: 368-373.
- Zhao, D., Zhu, B.L., Ishikawa, T., Quan, L., Li, D.R., Maeda, H.** (2006) Real-time RT-PCR quantitative assays and postmortem degradation profiles of erythropoietin, vascular endothelial growth factor and hypoxia-inducible factor 1 alpha mRNA transcripts in forensic autopsy materials. *Leg. Med. (Tokyo)*; **2**: 132-136.

8. Appendix

8.1 Liste der informativen Klone aus der lokal erstellten humanen retinalen cDNA-Bank

Name des Klons	Blast-Homologie ("best hit")	Ontologie
2b2	12q BAC RP11-473M14 (Roswell Park Cancer Institute Human BAC Library)	?
6c7	3'-phosphoadenosine 5'-phosphosulfate synthetase (PAPSS), exon 12	?
4b6	602377364F1 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:4508391 5'.	?
5f9	602507167F1 NIH_MGC_79 Homo sapiens cDNA clone IMAGE:4604176 5'	?
8h2	602617526F1 NIH_MGC_79 Homo sapiens cDNA clone IMAGE:4731199 5'.	?
9f8	603022793F1 NIH_MGC_114 Homo sapiens cDNA clone IMAGE:5193335 5'.	?
10d9	7b56d05.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3232233 3'.	?
1c11	actin related protein 2/3 complex, subunit 4 (20 kD) (ARPC4)	cell motility, actin cytoskeleton, Arp2/3 protein complex, cytoskeletal structural protein, Cell structure, Cell elongation, Cytoskeletal, Cell migration/motility, Complex assembly protein, Actin-cytoskeleton associated
4c8	adaptor-related protein complex 3, beta 2 subunit (AP3B2)	coatamer, vesicle transport, post Golgi transport, secretory vesicle membrane, Vesicle coat protein, Secretory vesicles
10a5	AFG3 ATPase family gene 3-like 2 (yeast) (AFG3L2), nuclear gene encoding mitochondrial protein	chaperone, mitochondrion, adenosinetriphosphatase, Hydrolase, Chaperones, Protein degradation, mitochondrial, Protease (other than proteasomal)
9g11	AGENCOURT_6401179 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:5498497	?
1f7	AGENCOURT_6418922 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5535027 5'.	?
2d1	AGENCOURT_6707852 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5753118 5'.	?
6h6	AGENCOURT_7552390 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:6061295 5'.	?
1b5	AGENCOURT_7809481 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:6053434 5'.	?
6d4	AGENCOURT_7846167 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:6064210 5'.	?
5d11	AGENCOURT_8727552 NIH_MGC_47 Homo sapiens cDNA clone IMAGE:6340648	?
6b5	AI680875 tx42c08.x1	?

Appendix

2f8	NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:2272238 3'. aj98d02.s1	?
4b9	Soares_parathyroid_tumor_NbHP A Homo sapiens cDNA clone IMAGE:1404483 3' similar to contains Alu repetitive element; AL534778 LTI_FL013_FBrn1	?
2e9	Homo sapiens cDNA clone CS0DF006YH16 5 prime. AL566029 LTI_FL013_FBrn1	?
9a7	Homo sapiens cDNA clone CS0DF016YF01 3 prime. aldo-keto reductase family 1, member A1 (aldehyde reductase) (AKR1A1)	glucose metabolism, aldehyde reductase, aldehyde metabolism, electron transporter, Oxidoreductase
3g1	aldolase A, fructose-bisphosphate	glycolysis, fructose metabolism, striated muscle contraction, fructose-bisphosphate aldolase, Lyase, General cellular role
5g2	anti-oxidant protein 2 (non-selenium glutathione peroxidase, acidic calcium-independent phospholipase A2) (KIAA0106)	non-selenium glutathione peroxidase
9d3	asparaginyl-tRNA synthetase (NARS)	cytoplasm, soluble fraction, protein biosynthesis, asparagine-tRNA ligase, Ligase, Protein synthesis, tRNA synthetase, Cytoplasmic, RNA-binding protein, RNA processing/modification
2d2	at65d06.x1 Barstead colon HPLRB7 Homo sapiens cDNA clone IMAGE:2376875 3'.	?
8d8	ATP synthase, H ⁺ transporting, mitochondrial F0 complex, subunit d (ATP5H)	this is one of the chains of the nonenzymatic component (cf(0) subunit) of the mitochondrial atpase complex. it has no apparent bacterial homolog and its exact function is unknown.
1e1	ATP synthase, H ⁺ transporting, mitochondrial F0 complex, subunit F6 (ATP5J), nuclear gene encoding mitochondrial protein	transporter, mitochondrion, energy pathways, membrane fraction, adenosinetriphosphatase, mitochondrial inner membrane, hydrogen-transporting two-sector ATPase, Hydrolase, Energy generation, Mitochondrial, Unspecified membrane, Active transporter, primary
9h8	ATPase, Na ⁺ /K ⁺ transporting, beta 3 polypeptide (ATP1B3)	small molecule transport, sodium/potassium-exchanging ATPase, sodium/potassium-transporting ATPase, Integral membrane, Unspecified membrane
8a3	attractin precursor (ATRN) gene, exon 3	plasma membrane, extracellular space, integral plasma membrane protein, Feeding, Hydrolase, Control of body size, Anti-pathogen response, Integral membrane, Receptor (signalling), Protease (other than proteasomal), Extracellular (excluding cell wall)
2b10	AU119848 HEMBA1 Homo sapiens cDNA clone HEMBA1006776 5'.	?
10h11	au88f12.y1 Schneider fetal brain 00004 Homo sapiens cDNA clone IMAGE:2783375 5'.	?
9a12	autism-related protein 1 (KIAA0442)	autism-related protein
1f2	Autosomal Highly Conserved Protein (LOC221659)	?
1d3	AV700422 GKC Homo sapiens cDNA clone GKCDSh12 3'.	?
10b11	AV722143 HTB Homo sapiens	?

Appendix

8a2	cDNA clone HTBATH12 5'. AV749684 NPC Homo sapiens cDNA clone NPCAYD07 5'. (fragment)	?
1d6	BAC clone GS1-39E22 from 5q31, complete sequence.143875 bp	?
4c5	BAC clone RP11-298H3 from 2, complete sequence	LINE/L2 von 165 bis 263, kein poly a, keine Fragmentierung, schwache homologie zu h.s. ventral anterior homeobox 2 (VAX2)
10c2	BAC clone RP11-429J10 from 2	?
7h10	BAC clone RP11-429O22 from 4, complete sequence	LINE/L1 von 2 bis 92, kein poly a, keine fragmentierung
10e12	BAC clone RP11-460N20 from 7, complete sequence	großes insert, keine fragmentierung, kein poly a, sine/alu repeat von 361 bis 476
9d8	basigin (OK blood group) (BSG)	signal transduction, peripheral plasma membrane protein, cell surface receptor linked signal transduction, Extracellular matrix maintenance
6b12	BCL2-associated athanogene (BAG1)	cytoplasm, anti-apoptosis, apoptosis inhibitor, receptor signalling protein, cell surface receptor linked signal transduction, Cell death/Apoptosis, Inhibitor or repressor
6f7	BCL2-like 12 (proline rich) (BCL2L12)	?
4f2	binder of Arl Two (BART1)	signal transduction, small GTPase regulatory/interacting protein,
7h2	bone morphogenetic protein receptor, type IA (BMPR1A) / ALK-3	integral membrane protein, TGFbeta receptor signalling pathway, Integral membrane, Receptor (signalling)
9g10	breast carcinoma amplified sequence 2 (BCAS2)	spliceosome, RNA processing, pre-mRNA splicing factor, RNA splicing, Nuclear, Spliceosomal subunit, RNA processing/modification
9b12	C18248 Human placenta cDNA (TFujiwara) Homo sapiens cDNA clone GEN-559E04 5'.	?
5a1	C6orf48: mRNA for G8 protein (G8 gene, located in the class III region of the major histocompatibility complex)	major histocompatibility complex antigen
2g4	calcium channel, voltage- dependent, alpha 1F subunit (CACNA1F)	vision, synaptic transmission, small molecule transport, voltage-gated calcium channel, voltage- sensitive calcium channel, Photoreception, Neuronal transmission, Integral membrane, Channel [passive transporter]
3h10	calmodulin 2 (phosphorylase kinase, delta) (CALM2)	calcium binding, phosphorylase kinase
4g3	calpain, small subunit 1 (CAPNS1)	calpain, positive control of cell proliferation, Protein degradation, Regulatory subunit
9b11	catenin (cadherin-associated protein), delta 2 (neural plakophilin-related arm-repeat protein) (CTNND2)	cytoplasm, cell adhesion, signal transduction, Cell structure, Neuronal development, Cytoplasmic
9c12	CD2-associated protein (CD2AP)	cytoplasm, cell shape control, actin cytoskeleton, signal transduction, protein complex assembly, cytoskeletal structural protein, substrate-bound cell migration, cell extension, Cell structure, Cytoskeletal, Cell migration/motility, Cell assembly protein
5a8	Cdc14B2 phosphatase	?
6g11	CDC28 protein kinase 2 (CKS2)	cell proliferation, regulation of CDK activity, Regulatory subunit
8f1	cDNA DKFZp434B049	?
2c9	cDNA DKFZp434B2411 / cDNA	?

Appendix

	DKFZp434O0827	
8a9	cDNA DKFZp434E082	?
4h12	cDNA DKFZp434K052	?
10c7	cDNA DKFZp564C053	?
1e4	cDNA DKFZp686C117	?
1e5	cDNA FLJ10533 fis	?
3e10	cDNA FLJ10991 fis	?
8g12	cDNA FLJ12443 fis	?
8d7	cDNA FLJ12670 fis	?
1h2	cDNA FLJ12754 fis, clone NT2RP2001268, highly similar to Homo sapiens mRNA for KIAA0810 protein.3804 bp	?
7g9	cDNA FLJ13618 fis	?
1c2	cDNA FLJ13783 fis, clone PLACE4000558, weakly similar to PROBABLE UBIQUITIN CARBOXYL-TERMINAL HYDROLASE FAF (EC	?
2f3	cDNA FLJ13888 fis	?
8f3	cDNA FLJ14725 fis, highly similar to Homo sapiens chromodomain- helicase-DNA-binding protein	?
7f12	cDNA FLJ14942 fis, highly similar to INSERTION ELEMENT IS1 PROTEIN INSB	?
9c9	cDNA FLJ20585 fis	?
7b12	cDNA FLJ21548 fis	?
9h12	cDNA FLJ21770 fis / cDNA FLJ31303 fis	?
3h8	cDNA FLJ22290 fis	?
10c6	cDNA FLJ23460 fis	?
8a1	cDNA FLJ23482 fis	?
8b10	cDNA FLJ30020 fis	?
3e1	cDNA FLJ30088 / cDNA FLJ39355 , hum zu Homo sapiens likely ortholog of mouse MAPK-interacting and spindle- stabilizing protein (MISS), mRNA	?
1a3	cDNA FLJ30653 fis	?
2c11	cDNA FLJ31373 fis / cDNA FLJ14028 fis	?
2a7	cDNA FLJ32031 fis	?
1c1	cDNA FLJ37630 fis / cDNA FLJ34422 fis, clone HHDPC2005185, moderately similar to STEROID RECEPTOR PROTEIN DG6	?
1d10	cDNA FLJ39478 fis / cDNA FLJ32554 fis	?
10h9	cDNA FLJ90320 fis / cDNA FLJ10905 fis	?
9a1	CDw92 antigen (CDW92)	?
10d3	cell division cycle 25C (CDC25C), transcript variant 1,	cell proliferation, control of mitosis, mitotic start control point, Control of Cell Proliferation
1f3	cell division cycle associated 4 (CDCA4)	?
10a4	cellular retinoic acid binding protein 2 (CRABP2),	retinoid binding, signal transduction, transcription regulation, epidermal differentiation, Small molecule- binding protein, Epidermal Development and Maintenance
3d4	ceroid-lipofuscinosis, neuronal 2,	lysosome, peptidase, neurogenesis, lipid metabolism,

Appendix

	late infantile (Jansky-Bielschowsky disease) (CLN2)	proteolysis and peptidolysis, Hydrolase, Protein degradation, CNS-specific functions, Lysosome/vacuole, Protease (other than proteasomal)
9h10	CGI-120 protein (LOC51644)	?
6g1	CGI-147 protein (LOC51651)	?
10f5	CGI-152 protein (CGI-152),	?
2b4	CGI-78 protein (LOC51107)	?
2h2	chaperonin containing TCP1, subunit 3 (gamma) (CCT3),	cytosol, chaperone, cytoskeleton, protein folding, Chaperones, Cytoplasmic, Cytoskeletal, Tubulin-cytoskeleton associated
5h6	chromobox homolog 3 (HP1 gamma homolog, Drosophila) (CBX3)	nucleus, chromatin binding, Nuclear, DNA-binding protein, Chromatin/chromosome structure,
9c2	chromodomain helicase DNA binding protein 1-like (CHD1L)	DEAD/DEAH box helicase, DEAD-like helicases superfamily, SNF2 and others N-terminal domain, helicase superfamily c-terminal domain, Helicases conserved C-terminal domain
9a4	chromosome 11, clone RP11-2C23, complete sequence	kein poly a, nicht fragmentiert, kein repeat
2d5	chromosome 17, clone hRPK.271_K_11, complete sequence	?
2d11	chromosome 5 clone CTB-43E15	?
6c9	chromosome 8, clone RP11-730G20, complete sequence	kein poly a, kein repeat, keine fragmentierung
6a3	cisplatin resistance-associated overexpressed protein (LUC7A)	nucleus, RNA splicing, RNA splicing, Nuclear, Serine rich protein
1b8	clathrin, light polypeptide (Lca)	Vesicle coat protein, Secretory vesicles
4d11	CLLL7 protein (FLJ10716)	?
1h3	clone IMAGE:5225645, mRNA	?
6b8	clone MGC:20579 IMAGE:4300679, mRNA, complete cds	?
2h5	clone RP11-730G20	?
5d6	coactivator for steroid receptors (COASTER)	coactivator for steroid receptors
5d12	coagulation factor VIII-associated (intronic transcript) (F8A)	possible housekeeping role
8g5	cold inducible RNA binding protein (CIRBP)	RNA binding, cold response, Cell stress, RNA-associated, RNA-binding protein
1d4	cone photoreceptor cGMP-gated channel alpha subunit (CNGA3)	vision, membrane fraction, signal transduction, small molecule transport, ligand-gated ion channel, integral plasma membrane protein, Photoreception, Integral membrane, Unspecified membrane, Channel [passive transporter]
2b9	cone photoreceptor cyclic nucleotide-gated channel beta subunit (CNGB3) oder: proteasom, macropain!!!	vision, signal transduction, small molecule transport, integral plasma membrane protein, intracellular cGMP activated cation channel, Photoreception, Integral membrane, Channel [passive transporter]
4h2	core1 UDP-galactose:N-acetylgalactosamine-alpha-R beta 1,3-galactosyltransferase (C1GALT1)	?
3d12	cyclin-dependent kinase inhibitor 2A (melanoma, p16, inhibits CDK4) (CDKN2A)	nucleus, oncogenesis, tumor suppressor, cell cycle arrest, cell cycle checkpoint, regulation of CDK activity, negative control of cell proliferation, cyclin-dependent protein kinase inhibitor, Nuclear, Regulatory subunit, Inhibitor or repressor
7b3	cytochrome b5 reductase 1 (B5R.1) (LOC51706),	Oxidoreductase, FAD-binding, NAD-binding,
1d1	cytochrome c oxidase subunit VIII	Mitochondrion, energy pathways, cytochrome-c

Appendix

	(COX8)	oxidase, Energy generation, General cellular role, Mitochondrial
9a2	cytosolic phospholipase A2 beta (cPLA2 beta) precursor RNA (PLA2G4B)	phospholipid metabolism, calcium-dependent cytosolic phospholipase A2
6b2	D component of complement (adipsin) (DF)	complement factor D, complement activation, serine-type peptidase, proteolysis and peptidolysis, Hydrolase, Anti-pathogen response, Protease (other than proteasomal)
5f5	D11lgp1e-like (LGP1)	?
8g4	DDM36 (DDM36)	?
4a8	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 15 (DDX15)	nucleus, RNA helicase, mRNA processing, ATPase, RNA splicing, Hydrolase, Nuclear, RNA-associated, RNA-binding protein
2a4	death associated protein 3 (DAP3), transcript variant, 2, nuclear gene encoding mitochondrial protein 1, nuclear gene encoding mitochondrial protein	induction of apoptosis by extracellular signals, Cell death/Apoptosis
10e11	dendritic cell protein (GA17)	?
4c9	desmuslin (DMN) / synemin (SYN gene), isoform B	alpha-dystrobrevin, intermediate filament protein,
7d9	differentially expressed in FDCP (mouse homolog) 6,	?
10a8	dihydrolipoamide dehydrogenase (E3 component of pyruvate dehydrogenase complex, 2-oxo-glutarate complex, branched chain keto acid dehydrogenase complex) (DLD)	mitochondrion, energy pathways, dihydrolipoamide dehydrogenase, catabolic carbohydrate metabolism, Energy generation, Oxidoreductase, Mitochondrial
1e7	DKFZp564N083_s1 564 (synonym: hfbr2) Homo sapiens cDNA clone DKFZp564N083 3'.	?
6d2	DNA sequence from clone RP11-168O16 on chromosome 1, complete sequence [Homo sapiens]	poly a am anfang, SINE/MIR von 73 bis 277
3d7	DNA sequence from clone RP11-214M7 on chromosome 1, complete sequence [Homo sapiens]	keine repeats, kein poly a, keine fragmentierung
3f4	DNA sequence from clone RP1-69B10 on chromosome 11p13	?
10g6	DNA-dependent protein kinase catalytic subunit (DNA-PKcs) mRNA, complete cds	protein binding, possibly be a Ca ²⁺ -binding
5b2	downstream neighbor of SON (DONSON)	?
4d10	EAP30 subunit of ELL complex (EAP30)	RNA polymerase II transcription factor, transcription regulation from Pol II promoter, Pol II transcription, DNA-binding protein
10a10	EBNA1 binding protein 2 (EBNA1BP2)	membrane fraction, Integral membrane
3g12	electron-transfer-flavoprotein, alpha polypeptide (glutaric aciduria II) (ETFA), nuclear gene encoding mitochondrial protein	mitochondrion, mitochondrial matrix, electron transfer flavoprotein, Energy generation, Mitochondrial
3b4	enolase 1, (alpha),	DNA binding, transcription factor, transcription co-repressor, phosphopyruvate hydratase, repression of transcription from Pol II promoter, Lyase, Pol II transcription, DNA-binding protein, Inhibitor or

Appendix

9g7	epithelial protein lost in neoplasm beta (EPLIN)	repressor, DNA-associated focal adhesion, actin cytoskeleton, Cytoskeletal, Cell junction, Actin-cytoskeleton associated
9a9	eukaryotic translation elongation factor 1 alpha 1,	cytoplasm, GTP binding, protein synthesis elongation, eukaryotic translation elongation factor 1, Hydrolase, Protein synthesis, Cytoplasmic, Ribosome-associated, GTP-binding protein/GTPase, Protein synthesis factor
1h7	eukaryotic translation elongation factor 1 gamma (EEF1G)	forms a nucleotide exchange complex with elongation factor 1 beta (EEFB2)
6e6	eukaryotic translation initiation factor 2, subunit 3 (gamma, 52kD) (EIF2S3)	GTPase, translation factor, cytosolic small ribosomal (40S)-subunit, eukaryotic translation initiation factor 2 complex, Hydrolase, Protein synthesis, Cytoplasmic, GTP-binding protein/GTPase, Protein synthesis factor
2h4	eukaryotic translation initiation factor 3, subunit 2 (beta, 36kD)	translation factor, translational regulation, initiation, eukaryotic translation initiation factor 3 complex, Protein synthesis, Cytoplasmic, Protein synthesis factor
1a7	fatty acid desaturase 1 (FADS1)	C-5 sterol desaturase, fatty acid desaturation, integral membrane protein, Oxidoreductase, Unspecified membrane
5h12	fatty acid desaturase 3 (FADS3)	fatty acid desaturase
1g10	F-box protein FBX30 (FBX30)	?
8f2	fibrillarin (FBL)	RNA binding, rRNA processing, small nucleolar RNA, RNA-associated, RNA-binding protein, RNA processing/modification, Nuclear nucleolus
1c9	formin binding protein 3 (FNBP3)	mRNA splicing, protein binding, signal transduction, proteolysis and peptidolysis
8h12	fusion, derived from t(12;16) malignant liposarcoma (FUS)	nucleus, RNA binding, Nuclear, RNA-associated, RNA-binding protein, RNA processing/modification
5e6	G8 protein (G8 gene, located in the class III region of the major histocompatibility complex)	major histocompatibility complex antigen
2c8	GCN5 general control of amino-acid synthesis 5-like 1 (yeast) (GCN5L1)	general control of amino-acid synthesis
7f7	general transcription factor II, i (GTF2I), transcript variant 3	protein binding, signal transduction, transcription factor, ranscription initiation from Pol II promoter, general RNA polymerase II transcription factor, Activator, Nuclear, Cytoplasmic, Neuronal development, Anti-pathogen response, DNA-binding protein,
2d9	genomic DNA, chromosome 11 clone:RP11-142N13,	?
3c12	genomic DNA, chromosome 11q clone:RP11-147I3, complete	?
8h3	germline T-cell receptor beta chain Dopamine-beta-hydroxylase-like, TRY1, TRY2, TRY3, TCRBV27S1P, TCRBV22S1A2N1T, TCRBV9S1A1T, TCRBV7S1A1N2T, TCRBV5S1A1T, TCRBV13S3, TCRBV6S7P, TCRBV7S3A2T, TCRBV13S2A1T, TCRBV9S2A2PT, TCRBV7S2A1N4T, TCRBV13S9/13S	?
5a3	glucosidase, beta; acid (includes glucosylceramidase) (GBA)	lysosomal, membrane bound, glycosyl hydrolase, gaucher's disease,

Appendix

3c6	glutamate-cysteine ligase, catalytic subunit (GCLC),	circulation, cysteine metabolism, glutamate metabolism, glutathione biosynthesis, glutamate--cysteine ligase, Ligase, Amino-acid metabolism
6f5	glyceraldehyde-3-phosphate dehydrogenase (GAPD)	Oxidoreductase
1f9	glyoxalase I (GLO1)	carbohydrate metabolism, lactoylglutathione lyase, Lyase
3b7	growth arrest and DNA-damage-inducible, alpha (GADD45A),	nucleus, apoptosis, DNA repair, stress response, cell cycle arrest, cell cycle control, regulation of CDK activity, Cell stress, Nuclear, Cell death/Apoptosis
8d3	growth arrest and DNA-damage-inducible, gamma (GADD45G),	DNA repair, stress response, activation of MAPKKK, Cell stress
2f4	growth suppressor 1 (GROS1)	growth suppressor
8d9	guanine nucleotide binding protein (G protein), beta polypeptide 2-like 1 (GNB2-RS1)	signal transduction, heterotrimeric G protein, protein kinase C activation, protein kinase C binding protein, peripheral plasma membrane protein
6f8	guanine nucleotide binding protein alpha s long (GNASL)	olfaction, plasma membrane, Golgi trans cisterna, adenylate cyclase activation, peripheral plasma membrane protein, Golgi to secretory vesicle transport, heterotrimeric G-protein GTPase, alpha subunit, Hydrolase,
1h6	guanine nucleotide binding protein-like 1 (GNL1)	GTP binding, signal transduction, Hydrolase, GTP-binding protein/GTPase
1b9	H1 histone family, member 0	nucleus, DNA binding, Nuclear, DNA-binding protein, Chromatin/chromosome structure,
2b3	H1 histone family, member 2 (H1F2)	nuclear, condensation of nucleosome chains
2a2	H3 histone, family 3A (H3F3A)	histone, Nuclear, General cellular role
3f10	H4027H10-3 NIA Mouse 7.4K cDNA Clone Set Mus musculus cDNA clone H4027H10 3'.	maus-homologie!
10d8	hepatitis delta antigen-interacting protein A (DIPA)	virulence
4e9	HepG2 3' region cDNA	? Ras homolog in Maus
9f10	heterogeneous nuclear ribonucleoprotein A1	cytoplasm, nucleoplasm, RNA binding, mRNA processing, nuclear RNA-nucleus export, heterogeneous nuclear ribonucleoprotein, Nuclear, RNA-associated, RNA-binding protein, RNA processing/modification, Nuclear-cytoplasmic transport
4e1	heterogeneous nuclear ribonucleoprotein A2/B1 (HNRPA2B1), transcript variant A2	nucleus, RNA binding, RNA processing, heterogeneous nuclear ribonucleoprotein, Nuclear, RNA-associated, RNA-binding protein, RNA processing/modification
10g8	heterogeneous nuclear ribonucleoprotein D (AU-rich element RNA binding protein 1, 37kDa) (HNRPD), transcript variant 1	nucleus, RNA binding, RNA catabolism, RNA processing, RNA turnover, Nuclear, RNA-associated, RNA-binding protein, RNA processing/modification
8a5	heterogeneous nuclear ribonucleoprotein D-like (HNRPDL)	nucleus, DNA binding, RNA binding, RNA processing, poly(A) binding, heterogeneous nuclear ribonucleoprotein,
9b6	heterogeneous nuclear ribonucleoprotein K (HNRPK), transcript variant 1	nucleus, nucleoplasm, RNA binding, RNA processing, heterogeneous nuclear ribonucleoprotein, RNA-associated, RNA processing/modification
5e4	hexokinase 2 (HK2)	hexokinase, cell cycle control, glucose catabolism, glucose catabolism, glucose metabolism, mitochondrial outer membrane, Transferase, Other kinase, Mitochondrial
4a1	high-mobility group (nonhistone chromosomal) protein 2,	DNA binding, nuclear chromosome, transcription factor, DNA-binding protein, DNA-associated (direct

Appendix

		or indirect)
1f8	histone acetyltransferase (HBOA)	nucleus, DNA replication, protein binding, transcription regulation, H3/H4 histone acetyltransferase, DNA synthesis, Transferase, Nuclear, Chromatin/chromosome structure,
8h9	histone deacetylase 6 (HDAC6)	deacetylation of lysine residues on the n-terminal part of the core histones (h2a, h2b, h3 and h4), role in transcriptional regulation, cell cycle progression and developmental events (by similarity), nuclear, hd subfamily 2.
10a12	homolog of Yeast RRP4 (ribosomal RNA processing 4), 3'-5'-exoribonuclease (RRP4)	nucleus, cytoplasm, 7S RNA binding, rRNA processing, 3'-5' exoribonuclease, Hydrolase, RNA-associated, RNA-binding protein, RNA processing/modification, Nuclease [endo, exo, ribo, deoxyribo]
6a4	host cell factor 2 (HCF-2)	nucleus, viral life cycle, transcription co-activator, RNA polymerase II transcription factor, transcription regulation from Pol II promoter, Activator, Pol II transcription, DNA-associated
10f12	HS1 binding protein (HAX1)	mitochondrion, nuclear membrane, soluble fraction, endoplasmic reticulum, Soluble, Nuclear, Mitochondrial
9e5	hypothetical protein BC009518 (LOC90799)	?
6f3	hypothetical protein BC012010 (LOC113174)	?
7g11	hypothetical protein DJ328E19.C1.1 (DJ328E19.C1.1),	?
8a7	hypothetical protein DKFZp564B1023 (DKFZP564B1023)	?
4b4	hypothetical protein FLJ10006 (FLJ10006)	?
4f9	hypothetical protein FLJ10006 (FLJ10006)	?
4c6	hypothetical protein FLJ10983 (FLJ10983)	?
8d11	hypothetical protein FLJ12443 (FLJ12443)	?
10c1	hypothetical protein FLJ12788 (FLJ12788)	?
9e1	hypothetical protein FLJ13111 (FLJ13111)	?
1g8	hypothetical protein FLJ14494 (FLJ14494)	?
8c2	hypothetical protein FLJ22060 (FLJ22060)	?
9d2	hypothetical protein FLJ22530 (FLJ22530)	?
7f4	hypothetical protein FLJ23018 (FLJ23018)	?
5e9	hypothetical protein FLJ23209 (FLJ23209)	?
6b11	hypothetical protein FLJ23251 (FLJ23251)	?
9e4	hypothetical protein FLJ31795 (FLJ31795)	?
10b10	hypothetical protein HSPC155 (HSPC155)	?
7b11	hypothetical protein HSPC219 (HSPC219)	?

Appendix

9g2	hypothetical protein LOC221302 (LOC221302)	?
8e10	hypothetical protein LOC51256 (LOC219624)	?
10g4	hypothetical protein LOC51320, hom zu Mus musculus similar to PEM-3 [Ciona savignyi]	?
10h12	hypothetical protein MGC13090 (MGC13090)	?
7b10	hypothetical protein MGC20496 (MGC20496)	?
8f7	hypothetical protein MGC20781 (MGC20781)	?
10g9	hypothetical protein MGC2734 (MGC2734)	?
9a8	hypothetical protein MGC8721 (MGC8721)	?
1c3	ig79c02.x1 HR85 islet Homo sapiens cDNA 3'.	?
10c12	IL5-CI0150-121100-244-b04 CI0150 Homo sapiens cDNA.	?
6d6	imageqc_9_2001/snq350bdf41.x 1 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:5449711 5'.	?
1a6	ING1 gene, exons 1a, 1b, 1c.	tumor suppressor, negative control of cell proliferation, Inhibitor or repressor, Control of Cell Proliferation
1g4	Inositol 1,4,5-trisphosphate receptor type 3 (Type 3 inositol 1,4,5-trisphosphate receptor) (Type 3 InsP3 receptor) (IP3 receptor isoform 3) (InsP3R3) (LOC221498),	?
8g8	insulinoma-associated protein IA-6 (INSM2)	?
4d7	integral inner nuclear membrane protein (MAN1)	membrane fraction, nuclear inner membrane, integral protein, Nuclear, Integral membrane, Unspecified membrane
2c10	integrin, beta 1 (fibronectin receptor, beta polypeptide, antigen CD29 includes MDF2, MSK12) (ITGB1), transcript variant 1A,	cell adhesion receptor, homophilic cell adhesion,
4g6	interferon gamma receptor 1 (IFNGR1)	immune response, plasma membrane, response to viruses, signal transduction, interferon-gamma receptor, antiviral response protein, resistance to pathogenic bacteria, Integral membrane, Receptor (signalling)
7b8	karyopherin (importin) beta 3 (KPNB3)	nucleus, cytoplasm, nuclear pore, GTPase inhibitor, RAN protein binding, importin, beta-subunit, NLS-bearing substrat-nucleusi
8b12	karyopherin (importin) beta 3 (KPNB3)	nucleus, cytoplasm, nuclear pore, GTPase inhibitor, RAN protein binding, importin, beta-subunit, NLS-bearing substrate-nucleus import, nuclear localization sequence binding, Nuclear, Protein translocation, Nuclear-cytoplasmic transport, Receptor (protein)
1f1	K-EST0057018 S3SNU16s1 Homo sapiens cDNA clone S3SNU16s1-22-G08 5'.	?
5e5	KIAA0036 protein (KIAA0036)	?
5g6	KIAA0102 protein (KIAA0102)	?
10b6	KIAA0117 protein (KIAA0117),	?

Appendix

	hom zu Mus musculus similar to Hypothetical RNA-binding protein	
7g3	KIAA0191 protein (KIAA0191)	?
10g12	KIAA0232 protein (KIAA0232)	?
2e5	KIAA0523 protein (KIAA0523)	?
6e5	KIAA0556 protein (KIAA0556)	?
10h8	KIAA0766 protein (KIAA0766)	?
3e7	KIAA0910 protein (KIAA0910)	?
9e11	KIAA0947 protein (KIAA0947)	?
6g9	KIAA1042 protein (KIAA1042)	?
7c7	KIAA1058 protein (KIAA1058)	?
1d9	KIAA1170 protein (KIAA1170)	?
8g10	KIAA1243 protein (KIAA1243),	?
9c11	KIAA1323 protein (KIAA1323)	?
8f4	KIAA1333 protein (KIAA1333)	?
7f6	KIAA1387 protein (KIAA1387)	?
2b5	KIAA1409 protein (KIAA1409)	?
3c8	KIAA1423 protein (KIAA1423)	?
8h7	KIAA1663 protein (KIAA1663)	?
6g8	lamin B1 (LMNB1)	lamin, nuclear lamina, structural protein, Cell structure, Nuclear, Cytoskeletal, Intermediate-filament associated
3a6	laminin receptor 1 (67kD, ribosomal protein SA) (LAMR1)	Oncogenesis, cell adhesion, translational regulation, laminin receptor protein, integral plasma membrane protein, cytosolic small ribosomal (40S)-subunit, cell surface receptor linked signal transduction, Protein synthesis, Receptor (signalling),
5h7	latent transforming growth factor beta binding protein 4 (LTBP4)	calcium binding, protein binding, extracellular matrix, Cell-to-cell signalling, Extracellular matrix (cuticle and basement membrane)
6d9	likely ortholog of mouse thyroid hormone receptor-associated protein 100 kDa (TRAP100)	ATP/GTP-binding site, thyroid hormone, presumptive zinc finger, suggesting nuclear hormone receptor interaction.
10e3	LOC147525 (LOC147525)	?
2b12	LOC152359 (LOC152359)	?
3h2	LOC152359 (LOC152359)	?
2g7	LOC161735 (LOC161735), mRNA / cDNA: FLJ22290 fis	?
5b11	LOC163078 (LOC163078)	?
4e5	LOC200339 (LOC200339)	?
10b3	LOC220662 (LOC220662)	?
8d5	LOC221640 (LOC221640)	?
3c2	LOC93487 (LOC93487)	?
1a5	lung cancer candidate FUS1 (FUS1)	cell proliferation, cell-cell signalling, Cell-to-cell signalling, Control of Cell Proliferation
2h6	macrophage myristoylated alanine-rich C kinase substrate (MACMARCKS),	?
4a5	malate dehydrogenase 2, NAD (mitochondrial) (MDH2), nuclear gene encoding mitochondrial protein	oxidative decarboxylation of malate, mitochondrial, pyruvate, may participate in the malate-aspartate shuttle,
2h9	MCM4 minichromosome maintenance deficient 4 (S. cerevisiae) (MCM4)	nucleus, chromatin, DNA binding, DNA replication, ATPase, DNA synthesis, Hydrolase, Nuclear, DNA-binding protein
9d11	MDN1, midasin homolog (yeast) (MDN1)	?
8d4	megakaryocyte-associated tyrosine kinase (MATK)	protein binding, soluble fraction, cell cycle control, cell proliferation, mesoderm development, protein phosphorylation, protein tyrosine kinase, positive control of cell proliferation, Transferase
10f10	melanoma antigen, family A, 12	may play a role in tumor transformation or aspects of

Appendix

10d4	(MAGEA12) membrane-bound transcription factor protease, site 1 (MBTPS1)	tumor progression. serine-type endopeptidase, endoplasmic reticulum lumen, proteolysis and peptidolysis, Hydrolase, Protease (other than proteasomal)
2g3	methylenetetrahydrofolate reductase (MTHFR) gene, 3' UTR	circulation, amino acid metabolism, methylenetetrahydrofolate reductase (NADPH), Oxidoreductase, Amino-acid metabolism
1g5	mitochondrial carrier homolog 1 (MTCH1)	?
7a7	mitochondrial ribosomal protein L14 (MRPL14), nuclear gene encoding mitochondrial protein	mitochondrion, ribosomal, nuclear
10c10	mitochondrial ribosomal protein L32 (MRPL32)	?
9e7	mitochondrial ribosomal protein L4 (MRPL4)	?
9h4	mitochondrial ribosomal protein L41 (MRPL41), nuclear gene encoding mitochondrial protein	mitochondrial
8a4	mitochondrial ribosomal protein L9 (MRPL9)	mitochondrial, belongs to the I9p family of ribosomal proteins
1b3	mitochondrial ribosomal protein S18C (MRPS18C)	?
10b7	mitogen-activated protein kinase kinase 2 (MAP2K2)	ser/thr family of protein kinases, map kinase kinase subfamily, activates the erk1 and erk2 map kinases
3b3	mitogen-activated protein kinase-activated protein kinase 3 (MAPKAPK3)	nucleus, stress response, MAP kinase kinase, signal transduction, Cell stress, Transferase, Nuclear
4a11	MRS2-like, magnesium homeostasis factor (<i>S. cerevisiae</i>) (MRS2L)	?
4g8	myeloid cell leukemia sequence 1 (BCL2-related) (MCL1)	apoptotic program, heat shock response, developmental processes, Differentiation, Cell death/Apoptosis, Inhibitor or repressor
10c8	myotubularin related protein 7 (MTMR7)	protein dephosphorylation, protein tyrosine phosphatase, Hydrolase
2c5	myristoylated alanine-rich protein kinase C substrate (MARCKS)	cell motility, plasma membrane, actin cytoskeleton, calmodulin binding, actin cross-linking, Cytoskeletal, Cell migration/motility, Actin-cytoskeleton associated
4h5	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 5 (16kD, SGDH) (NDUFB5)	complex I (NADH to ubiquinone), NADH dehydrogenase (ubiquinone), NADH dehydrogenase (ubiquinone), Energy generation, Oxidoreductase, Mitochondrial
2g1	nardilysin (N-arginine dibasic convertase) (NRD1)	metalloendopeptidase, proteolysis and peptidolysis, neuromuscular junction development, neuromuscular junction development, Hydrolase, Protease (other than proteasomal)
8c8	neogenin homolog 1 (chicken) (NEO1)	cell adhesion, cell motility, cell adhesion, plasma membrane, tumor suppressor, cell-cell signalling, developmental processes, integral membrane protein, integral plasma membrane protein
1h4	N-ethylmaleimide-sensitive factor attachment protein, gamma (NAPG)	membrane fusion, intra Golgi transport, inter-Golgi transport vesicle, Secretory vesicles
5c5	neural precursor cell expressed, developmentally down-regulated 5 (NEDD5)	?
8e9	neural proliferation, differentiation and control, 1 (NPDC1)	suppresses oncogenic transformation in neural and non- neural cells, down-regulates neural cell proliferation, might be involved in transcriptional regulation (by similarity).

Appendix

5a4	neurogenic differentiation 1 (NEUROD1)	RNA polymerase II transcription factor, Differentiation, Neuronal development
10g1	neurogenic differentiation 4 (Neurod4) gene, exon 2 and complete cds	RNA polymerase II transcription factor, Differentiation, Neuronal development
2b7	N-methyltransferase (GAMT), transcript variant 1	methyltransferase, muscle contraction, creatine biosynthesis, Muscle action, Other metabolism, Transferase, CNS-specific functions
8f9	NPDC1-like protein (NPDC1)	neural proliferation, differentiation and control, 1
9e12	nuclear distribution gene C homolog (A. nidulans) (NUDC)	cell cycle control, cell proliferation, developmental processes, Other development, Control of Cell Proliferation
6e3	nuclear pore complex protein (NUP107)	?
8c7	nuclease sensitive element binding protein 1 (NSEP1),	nucleus, DNA binding, transcription factor, double-stranded DNA binding, single-stranded DNA binding, response to pest/pathogen/parasite, transcription from Pol II promoter,
1d2	nucleolin (NCL)	nucleolus, RNA binding, Nuclear, RNA-associated, RNA-binding protein, RNA processing/modification, Nuclear nucleolus
7e4	nucleophosmin (nucleolar phosphoprotein B23, numatrin) (NPM1)	nucleolus, Nuclear
5e10	nucleosome assembly protein 1-like 1 (NAP1L1)	DNA replication, nucleosome assembly, chromatin assembly complex, positive control of cell proliferation, DNA synthesis, Nuclear, Chromatin/chromosome structure, Complex assembly protein,
10f6	nyctalopin (NYX)	congenital stationary night blindness
5d8	partial LIMD1 gene for LIM domains containing 1, exons 1-2	zinc binding, protein binding, tumor suppressor, signal transduction, developmental processes, transcription regulation, cell growth and maintenance
2h7	partial mRNA for TBP-associated factor 170 (TAFII170)	DNA binding, TFIID complex, ATPase, transcription co-repressor, transcription regulation from Pol III promoter, general RNA polymerase II transcription factor, repression of transcription from Pol II promoter, Hydrolase, Pol II + Pol III transcription
9a5	partial RANBP7 gene for RanBP7/importin7 and partial ZNF143 gene	?
4d9	partial XPGC gene, exon 1 / excision repair cross-complementing rodent repair deficiency, complementation group 5 (xeroderma pigmentosum, complementation group G (Cockayne syndrome))	endodeoxyribonuclease, nucleotide-excision repair, transcription-coupled repair, DNA repair, Hydrolase, Nuclease [endo, exo, ribo, deoxyribo]
3e8	PC4 and SFRS1 interacting protein 1	transcriptional coactivator, p52, interacts with splicing factor ASF/SF2.
1e6	PC4 and SFRS1 interacting protein 2 (PSIP2)	?
4a7	pelota homolog (Drosophila) (PELO)	?
4c2	peptide:N-glycanase similar to yeast PNG1 (FLJ11005),	peptide:N-glycanase
6e12	peroxiredoxin 1 (PRDX1)	skeletal development, cell proliferation
8b8	peroxiredoxin 3 (PRDX3)	alkyl hydroperoxide reductase, Cell stress, Oxidoreductase
6e4	phosphate cytidyltransferase 2, ethanolamine (PCYT2)	phospholipid biosynthesis; cdp-ethanolamine synthesis; second step, phosphorylethanolamine transferase

Appendix

9f7	phosphofructokinase, muscle (PFKM)	glucose metabolism, phosphofructokinase, glycolysis regulation, Energy Storage
9f1	phospholipase A2, group IVB (cytosolic) (PLA2G4B)	phospholipid metabolism, calcium-dependent cytosolic phospholipase A2
4b2	pinin, desmosome associated protein (PNN)	nucleus, cell adhesion, cell adhesion, plasma membrane, structural protein, intermediate filament, intercellular junction, cell shape and cell size control, Cell structure, Cell junction, Intermediate-filament associated
7e3	pituitary tumor-transforming 1 interacting protein,	may facilitate pttg nuclear translocation, pituitary tumor-transforming gene protein binding factor
5b7	pleiomorphic adenoma gene-like 2 (PLAGL2)	shows weak transcriptional activatory activity, nuclear (probable), zink finger
1h9	poly (ADP-ribose) glycohydrolase (PARG)	cytoplasm, poly(ADP-ribose) glycohydrolase, Hydrolase, Cytoplasmic
6b3	poly(rC) binding protein 2 (PCBP2), transcript variant 1	RNA-associated, RNA-binding protein, RNA processing/modification
1g9	polyamine N-acetyltransferase (SSAT2)	polyamine N-acetyltransferase
4h8	polycystic kidney disease 1 (autosomal dominant) (PKD1),	homophilic cell adhesion, cell-cell matrix adhesion, integral membrane protein, embryogenesis and morphogenesis, integral plasma membrane protein, calcium-independent cell-cell matrix adhesion, Integral membrane, Osmoregulation and Excretion,
2e4	polymerase (DNA directed), alpha (POLA)	nuclear, substrate recognition, polymerase
5e2	polyubiquitin UbC	ubiquitin, polyubiquitin
3b2	potential membrane protein C14orf1	?
7g1	PPAR binding protein (PPARBP)	nucleus, transcription co-activator, Activator, Nuclear, Other development, Pol II transcription
5d1	pre-B-cell leukemia transcription factor 4 (PBX4)	homeobox, transkription factor
8b11	PRO1073 mRNA	?
3c4	profilin 2 (PFN2), transcript variant 2	Regulatory subunit, binds to actin, affects structure of the cytoskeleton, profilin family.
9f3	prohibitin (PHB)	tumor suppressor
9b4	proliferating cell nuclear antigen (PCNA)	nucleus, DNA repair, DNA binding, DNA replication, cell proliferation, cell cycle control, delta-DNA polymerase cofactor, proliferating cell nuclear antigen, DNA synthesis, Nuclear, Control of Cell Proliferation, DNA polymerase or subunit
7b2	prominin-like 1 (mouse) (PROML1)	integral plasma membrane protein, Integral membrane
7c10	prosaposin (variant Gaucher disease and variant metachromatic leukodystrophy) (PSAP),	lipid binding, lipid transport, enzyme activator, extracellular space, integral membrane protein, glycosphingolipid metabolism, Soluble, Integral membrane, Intercellular transport, Unspecified membrane, Extracellular (excluding cell wall)
4d1	prostaglandin I2 (prostacyclin) receptor (IP) (PTGIR)	cell-cell signalling, prostoglandin I receptor, integral plasma membrane protein, G-protein signalling, linked to cyclic nucleotide second messenger, Cell-to-cell signalling, Receptor (signalling)
10d6	proteasome (prosome, macropain) 26S subunit, non-ATPase, 11 (PSMD11)	26S proteasome, Protein degradation, Proteasome subunit
2g12	proteasome (prosome, macropain) 26S subunit, non-ATPase, 3 (PSMD3)	regulatory subunit, atp-dependent degradation of ubiquitinated proteins.
9c4	proteasome (prosome, macropain) 26S subunit, non-ATPase, 7 (Mov34 homolog)	26S proteasome, Protein degradation, Proteasome subunit

Appendix

9d5	(PSMD7) proteasome (prosome, macropain) subunit, alpha type, 3 (PSMA3)	26S proteasome, Protein degradation, Proteasome subunit
3c3	protein kinase C substrate 80K-H (PRKCSH) / glucosidase II beta subunit	protein kinase / glucosidase II beta subunit
5e8	protein phosphatase 1, regulatory (inhibitor) subunit 11 (PPP1R11)	soluble fraction, protein phosphatase inhibitor, Inhibitor or repressor
5a6	protein phosphatase 1, regulatory (inhibitor) subunit 15B (PPP1R15B)	protein phosphatase 1
9f11	15B (PPP1R15B) protein phosphatase 1A (formerly 2C), magnesium-dependent, alpha isoform (PPM1A)	protein dephosphorylation, protein phosphatase type 2C, Cell stress, Hydrolase
3e6	protein phosphatase 1B (formerly 2C), magnesium-dependent, beta isoform	protein dephosphorylation, protein phosphatase type 2C, Hydrolase
6g3	protein phosphatase 4 (formerly X), catalytic subunit (PPP4C)	microtubule organization, serine/threonine protein phosphatase, cytoplasmic and nuclear; centrosomes.
8h6	protoporphyrinogen oxidase (PPOX)	mitochondrion, heme biosynthesis, electron transporter, protoporphyrinogen oxidase, Other metabolism, Oxidoreductase, Mitochondrial
10d10	PRP4 pre-mRNA processing factor 4 homolog (yeast) (PRPF4)	nucleus, mRNA splicing, RNA processing, pre-mRNA splicing factor, RNA splicing, Nuclear, RNA- associated, RNA-binding protein, Spliceosomal subunit, RNA processing/modification
5b6	PTK2 protein tyrosine kinase 2 (PTK2)	cytoskeleton, protein kinase, focal adhesion kinase, protein phosphorylation, integrin receptor signal signalling pathway, Cell structure, Transferase, Cytoplasmic, Cytoskeletal, Unspecified membrane, Actin-cytoskeleton associated
6d7	putative L-type neutral amino acid transporter (KIAA0436)	?
1a2	pyrophosphatase (inorganic) (PP)	phosphate metabolism, inorganic diphosphatase, Hydrolase, Other metabolism, Other phosphatase
5f8	RAB1A, member RAS oncogene family (RAB1A)	vesicle transport, RAB small monomeric GTPase, Hydrolase, GTP-binding protein/GTPase
1g6	RAB2, member RAS oncogene family (RAB2)	ER to Golgi transport, RAB small monomeric GTPase, Hydrolase, GTP-binding protein/GTPase
9c5	RAD21 homolog (S. pombe) (RAD21)	DNA recombination, meiotic recombination, double- strand break repair, DNA repair
9e9	RAN, member RAS oncogene family (RAN)	mitosis, nucleus, chromatin, nuclear pore, DNA metabolism, cell cycle control, signal transduction, mitotic spindle assembly, RAN small monomeric GTPase, nucleocytoplasmic transport, DNA synthesis, Hydrolase, GTP-binding protein/GTPase
4f4	ras homolog gene family, member A (ARHA)	oncogenesis, RHO small monomeric GTPase, RHO protein signal transduction, actin cytoskeleton reorganization, Cell structure, Hydrolase, GTP- binding protein/GTPase
9e8	RAS p21 protein activator (GTPase activating protein) 3 (Ins(1,3,4,5)P4-binding protein) (GAP1IP4BP)	plasma membrane, GTPase activator, signal transduction, Activator, GTPase activating protein
10e9	RB1-inducible coiled-coil 1 (RB1CC1)	coiled-coil
3d11	RC0-LT0001-131299-031-g09 LT0001 Homo sapiens cDNA.	?
4c3	regulator of chromosome condensation 1 (CHC1) / RCC1	nucleus, DNA packaging, chromosome condensation, nucleocytoplasmic transport, RAN

Appendix

		guanyl-nucleotide exchange factor, Nuclear-cytoplasmic transport, Chromatin/chromosome structure, Nuclear import/export protein, Guanine nucleotide exchange
8f8	restin (Reed-Steinberg cell-expressed intermediate filament-associated protein) (RSN)	endosome, microtubule binding, intermediate filament, microtubule cytoskeleton, non-selective vesicle transport, Cell structure, Cytoskeletal, Endosome/Endosomal vesicles, Tubulin-cytoskeleton associated, Intermediate-filament associated
2d12	reticulon 3 (RTN3)	extracellular space, integral membrane protein. endoplasmic reticulum (potential).
2b11	retinol binding protein (RBP)	retinoid binding, extracellular space, Other metabolism, Small molecule-binding protein
2g5	rhodopsin (opsin 2, rod pigment) (retinitis pigmentosa 4, autosomal dominant)	vision, G-protein linked receptor, rhodopsin mediated signalling, integral plasma membrane protein, phototransduction, visible light, G-protein linked receptor, protein signalling pathway
4e7	ribosomal protein L10a (RPL10A)	embryogenesis and morphogenesis, cytosolic large ribosomal (60S)-subunit,
1c4	ribosomal protein L11 (RPL11)	RNA binding, protein biosynthesis, structural protein of ribosome, cytosolic large ribosomal (60S)-subunit, Protein synthesis, Cytoplasmic, General cellular role, Ribosomal subunit, Ribosome-associated
6c3	ribosomal protein L15 (RPL15)	RNA binding, protein biosynthesis, structural protein of ribosome, Protein synthesis, Cytoplasmic, General cellular role, Ribosomal subunit, RNA-binding protein, Ribosome-associated
1f5	ribosomal protein L19 (RPL19)	RNA binding, protein biosynthesis,
10g11	ribosomal protein L3 (RPL3)	RNA binding, protein biosynthesis, structural protein of ribosome, cytosolic large ribosomal (60S)-subunit, Protein synthesis, Cytoplasmic, General cellular role, Ribosomal subunit, Ribosome-associated
2a1	ribosomal protein L32 (RPL32)	RNA binding, structural protein of ribosome, cytosolic large ribosomal (60S)-unit, Protein synthesis, General cellular role, Cytoplasmic, Ribosomal subunit, RNA-binding protein, Ribosome-associated
1e3	ribosomal protein L36a (RPL36A)	?
8b5	ribosomal protein L41 (RPL41)	RNA binding, protein biosynthesis, structural protein of ribosome, cytosolic large ribosomal (60S)-subunit, Protein synthesis, Cytoplasmic, General cellular role, Ribosomal subunit, Ribosome-associated
3b8	ribosomal protein L7a (RPL7A)	RNA binding, protein biosynthesis, structural protein of ribosome, cytosolic large ribosomal (60S)-subunit, Protein synthesis, Cytoplasmic, General cellular role, Ribosomal subunit, Ribosome-associated
10a11	ribosomal protein L9 (RPL9)	ribosome, RNA binding, protein biosynthesis, structural protein of ribosome, Protein synthesis, Cytoplasmic, General cellular role, Ribosomal subunit, Ribosome-associated
7h3	ribosomal protein S10 (RPS10)	RNA binding, protein biosynthesis, structural protein of ribosome, cytosolic small ribosomal (40S)-subunit, Protein synthesis, Cytoplasmic, General cellular role, Ribosomal subunit, RNA-binding protein, Ribosome-associated
2e8	ribosomal protein S16 (RPS16)	RNA binding, protein biosynthesis, structural protein of ribosome, cytosolic small ribosomal (40S)-subunit, Protein synthesis, Cytoplasmic, General cellular role, Ribosomal subunit, RNA-binding protein, Ribosome-associated
9g5	ribosomal protein S18 (RPS18)	RNA binding, protein biosynthesis, small ribosomal subunit, structural protein of ribosome, Cytoplasmic, General cellular role, Ribosomal subunit, Ribosome-

Appendix

10f7	ribosomal protein S20 (RPS20)	associated RNA binding, protein biosynthesis, structural protein of ribosome, cytosolic small ribosomal (40S)-subunit, Protein synthesis, Cytoplasmic, General cellular role, Ribosomal subunit, RNA-binding protein, Ribosome-associated
9g12	ribosomal protein S24 (RPS24)	RNA binding, protein biosynthesis, structural protein of ribosome, cytosolic small ribosomal (40S)-subunit, Protein synthesis, Cytoplasmic, General cellular role, Ribosomal subunit, RNA-binding protein, Ribosome-associated
1b7	ribosomal protein S25 (RPS25)	RNA binding, protein biosynthesis, structural protein of ribosome, cytosolic small ribosomal (40S)-subunit, Cytoplasmic, General cellular role, Ribosomal subunit, Ribosome-associated
1h12	ribosomal protein S27 (metallopanstimulin 1) (RPS27)	RNA binding, protein biosynthesis, structural protein of ribosome, cytosolic small ribosomal, (40S)-subunit, Protein synthesis, Cytoplasmic, General cellular role, Ribosomal subunit, RNA-binding protein, Ribosome-associated
5g1	ribosomal protein S3A (RPS3A)	RNA binding, protein biosynthesis, structural protein of ribosome, cytosolic small ribosomal, (40S)-subunit, Protein synthesis, Cytoplasmic, General cellular role, Ribosomal subunit, RNA-binding protein, Ribosome-associated
8f6	ribosomal protein, large, P0 (RPLP0)	RNA binding, protein biosynthesis, structural protein of ribosome, cytosolic large ribosomal (60S)-subunit, Protein synthesis, Cytoplasmic, General cellular role, Ribosomal subunit, Ribosome-associated
1a9	ribosomal protein, large, P1 (RPLP1)	RNA binding, protein biosynthesis, structural protein of ribosome, cytosolic large ribosomal (60S)-subunit, Protein synthesis, Cytoplasmic, General cellular role, Ribosomal subunit, Ribosome-associated
8b7	ring finger protein (C3HC4 type) 8 (RNF8)	?
5b12	RNA-binding region (RNP1, RRM) containing 2 (RNPC2)	nucleus, transcription, protein binding, transcription co-activator
8c10	sarcosine dehydrogenase (SARDH)	electron transporter, mitochondrial matrix, sarcosine dehydrogenase, Other metabolism, Oxidoreductase, Mitochondrial
9g8	SET translocation (myeloid leukemia-associated) (SET)	nucleus, oncogenesis, DNA replication, nucleosome assembly, DNA replication factor, protein phosphatase inhibitor, DNA synthesis, Nuclear, Inhibitor or repressor, DNA polymerase or subunit
10e1	SFRS protein kinase 2 (SRPK2)	nucleus, RNA splicing, protein kinase, spliceosome assembly, RNA splicing, Transferase, Nuclear
7e12	SH3-domain binding protein 5 (BTK-associated) (SH3BP5),	cytoplasm, signal transduction, SH3/SH2 adaptor protein
4g4	SIGNAL RECOGNITION PARTICLE RECEPTOR ALPHA SUBUNIT (SR-ALPHA) (DOCKING PROTEIN ALPHA) (DP-ALPHA) (LOC219818)	?
3a9	signal sequence receptor, beta (translocon-associated protein beta) (SSR2)	endoplasmic reticulum, signal sequence receptor, integral membrane protein, co-translational membrane targeting, Protein translocation, Integral membrane, Receptor (protein translocation)
1a10	signal sequence receptor, gamma (translocon-associated protein gamma) (SSR3)	microsome, signal sequence receptor
3h3	Similar to actin, gamma 1	cytoskeletal structural protein, Cell structure, Cytoskeletal

Appendix

9d1	similar to Antigen KI-67 (LOC157313)	?
9g9	similar to Basement membrane-specific heparan sulfate proteoglycan core protein precursor (HSPG) (Perlecan)	?
6h12	similar to KIAA1245 protein (LOC115144)	?
2d3	Similar to neural cell adhesion molecule 1 (NCAM1)	cell adhesion, plasma membrane, cell-cell signalling, synaptic transmission, integral membrane protein, Neuronal development, Neuronal transmission, CNS-specific functions, PNS-specific functions, Adhesin/agglutinin, Cell-to-cell signalling,
3a7	similar to Neural cell adhesion molecule 1, 180 kDa isoform precursor (N-CAM 180) (NCAM-180) (LOC143913),	?
2d4	Similar to prominin (mouse)-like 1 (PROML1)	Integral plasma membrane protein,
2a9	similar to protease (prosome, macropain) 26S subunit	ATPase
3c11	similar to putative transmembrane protein; homolog of yeast Golgi membrane protein Yif1p (Yip1p-interacting factor) (LOC90522)	? Protein phosphatase?
1a4	similar to RIKEN cDNA 1700102P08 gene (LOC151907),	?
6f1	Similar to RNA binding motif protein, X chromosome,	?
4h6	similar to sialyltransferase 7 ((alpha-N-acetylneuraminyl 2,3-betagalactosyl-1,3)-N-acetyl galactosaminide alpha-2,6-sialyltransferase) E (MGC3184)	?
6g2	Similar to ubiquitin carboxy-terminal hydrolase L1,	ubiquitin, hydrolase
4b11	sin3-associated polypeptide, 18kD (SAP18)	transcription co-repressor, histone deacetylase complex, transcription regulation from Pol II promoter, Nuclear, Pol II transcription, Inhibitor or repressor, Chromatin/chromosome structure, DNA-associated (direct or indirect)
2d10	single stranded DNA binding protein-3 (SSDP3) gene, exon 17 / myo-inositol 1-phosphate synthase A1 (LOC245744)	?
5h8	small nuclear ribonucleoprotein polypeptide C (SNRPC)	snRNP U1e, mRNA splicing, small nuclear ribonucleoprotein, RNA splicing, Nuclear, RNA-associated, RNA-binding protein, Spliceosomal subunit, RNA processing/modification
10f1	small nuclear ribonucleoprotein polypeptide N	spliceosome, mRNA splicing, small nuclear ribonucleoprotein, RNA splicing, Nuclear, RNA-associated, CNS-specific functions, RNA-binding protein, Spliceosomal subunit, RNA processing/modification
5e11	small nuclear ribonucleoprotein polypeptides B and B1 (SNRPB) SNRPB: small nuclear ribonucleoprotein polypeptides B and B1	spliceosome, mRNA splicing, small nuclear ribonucleoprotein, RNA splicing, Nuclear, RNA-associated, RNA-binding protein, Spliceosomal subunit, RNA processing/modification
10d2	SMC1 structural maintenance of chromosomes 1-like 1 (yeast) (SMC1L1)	motor, mitosis, nucleus, mitotic chromosome, adenosinetriphosphatase, mitotic chromosome segregation, Nuclear, Motor protein, Chromatin/chromosome structure, DNA-associated

Appendix

7a12	smoothelin (SMTN), transcript variant 2	(direct or indirect) cytoskeleton, actin binding, muscle development, actin cytoskeleton, smooth muscle contraction, structural protein of muscle, cell shape and cell size control, Cell structure, Muscle action, Cytoskeletal,
3f2	solute carrier family 17 (sodium-dependent inorganic phosphate cotransporter), member 7 (SLC17A7) / BNPI mRNA for brain-specific Na-dependent inorganic phosphate cotransporter	phosphate transport, integral membrane protein, sodium-dependant phosphate transporter, Integral membrane, Unspecified membrane
2f6	somatic cytochrome c (HCS)	cytoplasm, mitochondrion, induction of apoptosis, Energy generation, Cell death/Apoptosis, Mitochondrial
9d9	spastic paraplegia 7, paraplegin (pure and complicated autosomal recessive) (SPG7)	chaperone, peptidase, neurogenesis, mitochondrion, adenosinetriphosphatase, proteolysis and peptidolysis, Hydrolase, Energy generation, Protein degradation, Neuronal development, Mitochondrial, Protease (other than proteasomal)
4g7	spliced UHG / U22 small nucleolar	small nucleolar
10e8	splicing factor 3a, subunit 3, 60kD (SF3A3)	nucleus, spliceosome, mRNA splicing, mRNA processing, pre-mRNA splicing factor, RNA splicing, Nuclear, RNA-associated, RNA-binding protein, Spliceosomal subunit, RNA processing/modification
1f4	splicing factor, arginine/serine-rich 11 (SFRS11)	nucleus, mRNA splicing, mRNA processing, pre-mRNA splicing factor, RNA splicing, Soluble, Nuclear, Spliceosomal subunit, RNA processing/modification
6g4	splicing factor, arginine/serine-rich 4 (SFRS4),	nucleus, mRNA splicing, mRNA processing, pre-mRNA splicing factor, RNA splicing, Nuclear, RNA-associated, RNA-binding protein, Spliceosomal subunit, RNA processing/modification
8b6	splicing factor, arginine/serine-rich 7 (SFRS7)	nucleus, mRNA splicing, mRNA processing, pre-mRNA splicing factor, RNA splicing, Nuclear, RNA-associated, RNA-binding protein, Spliceosomal subunit, RNA processing/modification
4e4	sporulation-induced transcript 4-associated protein SAPLa (C11orf23)	?
3h9	stathmin 1/oncoprotein 18 (STMN1)	cytosol, cytoplasm, signal transduction, cell growth and maintenance, intracellular signalling cascade, Soluble, Cytoplasmic
3f6	stearoyl-CoA desaturase (delta-9-desaturase) (SCD),	Oxidoreductase, fatty acid desaturase family, integral membrane protein. endoplasmic reticulum (probable), acyl-coa desaturase
10h3	stromal cell derived factor receptor 1 (SDFR1), transcript variant alpha	belonging to the Ig superfamily, believed to be involved in cell-cell interactions or cell-substrate interactions
10g7	succinate dehydrogenase complex, subunit C, integral membrane protein, 15kDa (SDHC), nuclear gene encoding mitochondrial protein	cytochrome b, mitochondrion, mitochondrial membrane, tricarboxylic acid cycle, Energy generation, Oxidoreductase, Mitochondrial, Unspecified membrane
6c2	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 4 (SMARCA4),	nucleus, helicase, nucleoplasm, transcription factor, adenosinetriphosphatase, transcription co-activator, transcription regulation from Pol II promoter, Activator, Hydrolase, Nuclear, Pol II transcription,
1b6	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily	nucleus, DNA helicase, DNA metabolism, DNA dependent adenosinetriphosphatase, transcription regulation from Pol II promoter, Hydrolase, Nuclear,

Appendix

	a-like 1 (SMARCAL1)	Pol II transcriptionm, DNA-binding protein, Chromatin/chromosome structure,
9a6	synaptogyrin 1 (SYNGR1)	integral plasma membrane protein, Integral membrane
6c12	synaptojanin 1 (SYNJ1),	secretory vesicle, phosphate metabolism, synaptic vesicle endocytosis, non-selective vesicle transport, inositol-1,4,5-triphosphate 5-phosphatase, Hydrolase, Other phosphatase, Other vesicles of the secretory/endocytic pathways
6e8	synovial sarcoma translocation gene on chromosome 18-like 1 (SS18L1)	belongs to the ss18 family
6f11	TATA element modulatory factor 1 (TMF1)	transcription co-factor, transcription regulation, transcription from Pol II promoter, RNA polymerase II transcription factor, Pol II transcription, DNA-binding protein,
1g3	TBP-interacting protein (TIP120A)	?
6f6	TCF19: transcription factor 19 (SC1)	cell cycle control, cell proliferation, transcription factor, transcription regulation from Pol II promoter, Control of Cell Proliferation
9f2	tetraspan 3 (TSPAN-3)	cell motility, cell adhesion, cell proliferation, integral membrane protein, mystery cell fate differentiation, Integral membrane, Unspecified membrane
10c9	tetratricopeptide repeat domain 3 (TTC3)	?
10b12	Thymidine kinase 1, soluble (TK1)	cytoplasm, thymidine kinase, nucleobase, nucleoside, nucleotide and nucleic acid metabolism, Cytoplasmic
5a5	thyroid autoantigen 70kD (Ku antigen) (G22P1)	nucleus, DNA ligation, DNA binding, membrane fraction, DNA recombination, ATP dependent DNA helicase, double-strand break repair, double-stranded DNA binding,DNA-nonhomologous end-joining, DNA repair, Nuclear, Anti-pathogen response, DNA-associated,
4f6	transaldolase 1 (TALDO1)	transaldolase, carbohydrate metabolism
8h8	transducin-like enhancer of split 1 (E(sp1) homolog, Drosophila) (TLE1)	nucleus, signal transduction, developmental processes, histogenesis and organogenesis, Nuclear, Cell fate specification
2d7	transketolase (Wernicke-Korsakoff syndrome) (TKT)	transketolase, Transferase
9d12	translocase of outer mitochondrial membrane 20 (yeast) homolog (KIAA0016)	?
6e7	transmembrane 4 superfamily member tetraspan NET-6 (NET-6)	membrane fraction, integral plasma membrane protein, Integral membrane, Unspecified membrane
1g1	tubulin, alpha, ubiquitous	microtubule, structural protein, Cell structure,
5d2	tubulin, beta 5	tubulin
10b8	tubulin-specific chaperone a (TBCA)	cochaperone, beta-tubulin folding, post-chaperonine tubulin folding pathway, Chaperones, Cytoskeletal, Tubulin-cytoskeleton associated
10a3	tumor protein, translationally-controlled 1 (TPT1)	extracellular space, Anti-pathogen response, Extracellular (excluding cell wall)
3d10	tx42c08.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:2272238 3'.	?
8g2	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta polypeptide (YWHAZ)	protein binding, Regulatory subunit
9h6	ubiquinol-cytochrome c reductase	aerobic respiration, oxidative phosphorylation,

Appendix

	core protein II (UQCRC2)	mitochondrial inner membrane, Energy generation, General cellular role, Mitochondrial
3b11	ubiquitin B (UBB)	ubiquitin, protein modification, Protein conjugation factor
6h2	ubiquitin carboxyl-terminal esterase L1 (ubiquitin thiolesterase) (UCHL1)	ubiquitin-dependent protein degradation, PNS-specific functions, Protease (other than proteasomal)
6d3	ubiquitin hydrolyzing enzyme I (UBH1)	ubiquitin carboxyl-terminal hydrolase, peptidase family c19
3h1	ubiquitin specific protease 22 (USP22)	ubiquitin specific protease
2f11	ubiquitin-conjugating enzyme E2 variant 1 (UBE2V1),	cell cycle control, polyubiquitylation, DNA repair regulation, ubiquitin conjugating enzyme
6b10	ubiquitin-conjugating enzyme E2G 2(homologous to yeast UBC7),	protein modification, DNA repair regulation, ubiquitin-protein ligase, proteolysis and peptidolysis, ubiquitin conjugating enzyme, Ligase, Protein degradation, Protein conjugation factor
5c4	ubiquitin-conjugating enzyme E2N (UBC13 homolog, yeast) (UBE2N)	protein modification, DNA repair regulation, ubiquitin-protein ligase, proteolysis and peptidolysis, ubiquitin conjugating enzyme, Ligase, Protein degradation, Protein conjugation factor
1d8	UDP-glucose dehydrogenase (UGDH)	electron transporter, UDP-glucose metabolism, UDP-glucose 6-dehydrogenase, UDP-glucuronate biosynthesis, glycosaminoglycan biosynthesis, Other metabolism, Oxidoreductase
1e10	UI-E-CL1-afe-n-12-0-UI.r2 UI-E-CL1 Homo sapiens cDNA clone UI-E-CL1-afe-n-12-0-UI 5'.	?
6e2	UI-E-EJ1-ajz-a-23-0-UI.r1 UI-E-EJ1 Homo sapiens cDNA clone UI-E-EJ1-ajz-a-23-0-UI 5'.	?
6c6	UI-E-EO0-aia-k-23-0-UI.r1 UI-E-EO0 Homo sapiens cDNA clone UI-E-EO0-aia-k-23-0-UI 5'.	?
9f9	UI-E-EO1-ajc-n-20-0-UI.s1 UI-E-EO1 Homo sapiens cDNA clone UI-E-EO1-ajc-n-20-0-UI 3'.	?
4d3	UI-H-BI3-alc-b-10-0-UI.s1 NCI_CGAP_Sub5 Homo sapiens cDNA clone IMAGE:2736330 3'.	?
1e2	UPF3A (UPF3A)	binds to spliced mRNAs upstream of exon-exon junctions, part of mRNP complexes, nuclear export, participate in nonsense-mediated mRNA decay, vesicle targeting, ER to Golgi transport, endoplasmic reticulum, integral membrane protein, non-selective vesicle transport, Docking protein, Integral membrane
7c3	vesicle trafficking protein (SEC22C), transcript variant 2	cytoplasm, protein binding, Cytoplasmic
1f10	von Hippel-Lindau binding protein 1 (VBP1)	
9b7	WAS protein family, member 1 (WASF1)	cytoskeleton, cell motility, actin cytoskeleton, actin polymerising, protein complex assembly, cell shape and cell size control
3b9	WD repeat domain 1 (WDR1) transcript variant 2,	hearing, protein binding, disassembly of actin filaments
9f12	ws39b04.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2499535 3'.	?
9a10	WW domain binding protein 11 (WBP11)	nucleus, single-stranded DNA binding
10f3	xenotropic and polytropic retrovirus receptor (XPR1)	receptor, pathogenesis, plasma membrane, transmembrane receptor, G-protein linked receptor, integral plasma membrane protein, G-protein linked receptor protein signalling pathway, Pathogenic

Appendix

7b9	yw59e09.r1 Soares_placenta_8to9weeks_2N bHP8to9W Homo sapiens cDNA clone IMAGE:256552 5'	Invasion, Integral membrane, Receptor (signalling) ?
10d11	ze72c10.r1 Soares_fetal_heart_NbHH19W Homo sapiens cDNA clone IMAGE:364530 5'	?
5h4	zo74h04.r1 Stratagene pancreas (#937208) Homo sapiens cDNA clone IMAGE:592663 5' similar to contains Alu repetitive element;contains element TAR1 repetitive element	?
8e8	zt75f04.s1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:728191 3' similar to gb:M99435 TRANSDUCIN-LIKE ENHANCER PROTEIN 1 (HUMAN)	?

8.2 Abkürzungsverzeichnis

AMD	<i>age relatated macular dystrophy</i>
AS	Aminosäure
API	<i>application programming interface</i>
APS	Ammoniumpersulfat
ATP	Adenosintriphosphat
BAC	<i>bacterial artificial chromosome</i>
BLAST	<i>basic local search alignment tool</i>
bp	Basenpaar
cDNA	komplementäre Desoxyribonukleinsäure
CsCl	Cäsiumchlorid
C _y	Schwellenwert für Q-PCR
dATP	Desoxy-adenosin-5'-triphosphat
dCTP	Desoxy-cytidin-5'-triphosphat
DDBJ	<i>DNA Data Bank of Japan</i>
ddNTP	Didesoxyribonukleotidtriphosphat
dGTP	Desoxy-guanosin-5'-triphosphat
DEPC	Diethylpyrocarbonat
DMSO	Dimethylsulfoxid
DNA	Desoxyribonucleinsäure
dNTP	Desoxy-nukleosid-5'-triphosphat
dTTP	Desoxy-thymidin-5'-triphosphat
dUTP	Desoxy-uridin-5'-triphosphat
<i>E. coli</i>	<i>Escherichia coli</i>
EMBL	Europäisches Molekulargenetisches Labor
EDTA	Ethylendiamintetraessigsäure
<i>EPO</i>	Erythropoietin
EST	<i>expressed sequence tag</i>
FCS	Fetales Kälberserum
<i>GAPDH</i>	Glycerinaldehydphosphat-Dehydrogenase
G.I.R.I	<i>Genetic Information Research Institute</i>
GUI	<i>graphic user interface</i>
HIF1	<i>hypoxia induced factor 1</i>
HPLC	<i>high pressure liquid chromatography</i>
IPTG	Isopropyl-1-thio-β-D-galactosid
kb	Kilobase = 1000 bp
KCl	Kaliumchlorid
LB	Luria Bertani
MA	Milliampere
Mb	Mega-Base
MgCl ₂	Magnesiumchlorid
MIM	<i>Mendeleian Inheritance in Man</i>
MMLV	<i>Moloney Murine Leukaemia Virus</i>
MSC	Multisequenz-Cluster
mM	Millimolar
mRNA	<i>messenger RNA</i>
NaCl	Natriumchlorid
NaOH	Natriumhydroxid
NCBI	National Center for Biotechnology Information

Appendix

NMDA	N-methyl-D-Aspartat
OD	optische Dichte
OOP	objektorientiertes Programmieren
<i>OPAI</i>	Optikus Atrophie Locus 1
PAA	Polyacrylamid
PAGE	Polyacrylamid-Gelelektrophorese
PAC	<i>P1-Phage derived Artificial Chromosome</i>
PBS	<i>phosphate-buffered saline</i>
PDB	<i>protein database</i>
PCR	Polymerasekettenreaktion
pH	<i>potentia hydrogenii</i>
PHP	<i>Hypertext Preprocessor</i>
Phrap	<i>Phil's revised assembly program</i>
pM	picomolar
QM	Querymaster
Q-PCR	<i>quantitative Real Time-PCR</i>
RACE	<i>rapid amplification of cDNA ends</i>
RAM	<i>random access memory</i>
R.E.A.L.	<i>rapid extraction alkaline lysis</i>
RNA	Ribonukleinsäure
RNase	Ribonuklease
rpm	Umdrehungen pro Minute
RT	Raumtemperatur
RT-PCR	<i>Real Time-PCR</i>
SAGE	<i>serial analysis of gene expression</i>
SDS	Natriumdodecylsulfat
SDX	<i>starkit developer extension</i>
STS	<i>sequence tagged site</i>
Starkit	<i>stand alone runtime for tool kit</i>
SQL	<i>structured query language</i>
TAE	Tris-Acetat-EDTA
Taq	<i>Thermus aquaticus</i>
TBE	Tris-Borat-EDTA-Puffer
TBS	Tris-buffered-saline
TCL/Tk	<i>tool command languaget/tool kit</i>
TE	Tris-EDTA-Puffer
TEMED	N, N, N', N'-Tetramethylethylendiamin
TGI	<i>The Institute for Genomic Research</i>
Tris	Tris(hydroxymethyl-)aminomethan
U	Units
UV-Licht	Ultraviolettes Licht
<i>VEGF</i>	<i>vascular endothelial growth factor</i>
X-Gal	5-Brom-4-Chlor-3-indolyl- β -D-Galactosid
XML	<i>extensible markup language</i>
YAC	<i>yeast artificial chromosome</i>

Lebenslauf

Persönliche Daten:

Name: Jens Meyer
Geburtsdatum: 11.09.1972
Geburtsort: Emden
Familienstand: ledig

Schulbildung:

1979 – 1983 Grundschule Wolthusen, Emden
1983 – 1985 Wallschule (Orientierungsstufe), Emden
1985 – 1992 Gymnasium am Treckfahrtstief, Emden

Universitätsausbildung:

1992 – 1995 Studium der Biologie an der er Westfälischen Wilhelms-Universität, Münster
1995 – 1997 Studium der Biologie an der Eberhard Karls-Universität, Tübingen mit dem Hauptfach Humangenetik und den Nebenfächern Mikrobiologie und Biochemie
1998 Diplomarbeit am Lehrstuhl für Mikrobielle Genetik der Eberhard Karls-Universität, Tübingen unter der Leitung von Prof. F. Götz mit dem Titel: „Molekularbiologische Charakterisierung des Phosphotransferase-Systems in *Staphylococcus xyosus*“
1998 – 1999 Zivildienst an der Universitätsaugenkinik Tübingen
2000 Ausbildung zum Juniorteamer des Centrums für Angewandte Politikforschung der Ludwig Maximilian-Universität, München
2000 Beginn eines BWL-Studiums an der Fernuniversität Hagen
2000 – 2006 Dissertation an der Universitätsaugenkinik Tübingen unter Leitung von Prof. E. Zrenner mit dem Titel „Identifikation und Katalogisierung retinal exprimierter Gene“

Eigene Veröffentlichungen

Publikationen:

Jacobi, F.K., Leo-Kottler, B., Mittelviehhaus, K., Zrenner, E., Meyer, J., Pusch, C.M., Wissinger, B. (2001) Segregation patterns and heteroplasmy prevalence in Leber's hereditary optic neuropathy. *Invest. Ophthalmol. Vis. Sci.*; **6**: 1208-14.

Jacobi, F.K., Meyer, J., Pusch, C.M., Wissinger, B. (2001) Quantitation of heteroplasmy in mitochondrial DNA mutations by primer extension using Vent_R(exo-) DNA polymerase and RFLP analysis. *Mutat. Res.*; **1**: 141-51.

Buchbeiträge:

Jancovic, I., Meyer, J., Brückner, R. (2003) Catabolite control protein CcpA-dependent glucose repression in *Staphylococcus xylosum*: efficient activation of CcpA by glucose transported independently from the phosphotransferase system. *Regulatory Networks in Prokaryotes*; Horizon Scientific Press, Norwich.

Poster:

Jacobi, F.K., Baumann, B., Meyer, J., Leo-Kottler, B., Mittelviehhaus, K., Pusch, C.M., Wissinger, B.

Häufigkeit und Stabilität heteroplasmatischer Mutationen bei der Leberschen Optikusneuropathie.

Fortüne-Kolloquium, Tübingen, Oktober 2000

Meyer, J., Schlenker, M., Feldkämper, M., Schaich, S., Wissinger, B.

Identifizierung und Katalogisierung retinal exprimierter Gene.

Fortüne-Kolloquium, Tübingen, Oktober 2002

Feldkämper, M., Meyer, J., Wissinger, B., Schäffel, F.

Identification of genes expressed in chicken retinal cDNA library.

9th International conference on Myopia, Hongkong, November 2002