

## 3. Material und Methoden

### 3.1. Die Klon-Datenbank *CloneBase*

Aufgrund der gestellten Anforderungen an ein System zur Verwaltung der Klon-Bibliothek der Arbeitsgruppe und der zugehörigen Daten ist es sinnvoll, sämtliche Informationen in einer zentralen spezialisierten Datenbank zu verwalten, welche über das Netzwerk von verschiedenen Computern aus mittels definierter Mechanismen abfragbar ist.

Die Basistechnologie, die für die hierzu entwickelte Datenbank *CloneBase* genutzt wurde, ist ein *mySQL*-Datenbank-Managementsystem (*mySQL* AB, Uppsala/Schweden), das auf einem *Linux* Server-Computer (Suse 9.0; Suse Inc., San Francisco/USA) betrieben wird. Dieser Rechner ist mit 2 *Pentium 4*-Prozessoren (Intel Corp., Santa Clara, USA) mit 2,6 GHz Taktfrequenz und 2 GB 266 MHz RAM-Speicher, einem RAID-Speichersystem mit sechs 74 GB Festplattenkapazität und einem Gigabit-Netzwerkadapter ausgerüstet. Der Zugriff auf die *CloneBase*-Datenbank erfolgt über eine Internet-basierte Benutzeroberfläche, die von einem *Apache*-Webserver (Apache Software Foundation, Forest Hill/USA) des Server-Computers bereitgestellt wird.

Zur Einspeicherung und Aktualisierung von Annotationsdaten in die Datenbank wurden Programme in den Sprachen Java und Perl geschrieben. Java wurde 1995 als objektorientierte, architektur- und systemneutrale Sprache von Sun (Palo Alto, USA.) entwickelt und ist sowohl zur Entwicklung von speziellen Internet-Anwendungen (*Applets*) als auch für komplexe eigenständige Programme geeignet. Perl wurde ursprünglich 1987 (Version 1.0) von Larry Wall zur einfachen aber mächtigen Skript-Sprache für Text-Manipulation vorgestellt. Durch weltweit aktive Entwicklergruppen ist Perl zum idealen Werkzeug für ganz unterschiedliche Problemstellungen geworden und hat sich im Bereich der Bioinformatik als am weitesten verbreitete Sprache etabliert.

Die Aktualisierungs-Skripte nutzen Daten aus den folgenden öffentlich zugänglichen Datenbanken:

- euGene (<http://iubio.bio.indiana.edu:8089>, Gilbert 2002)

- HUGO (<http://www.gene.ucl.ac.uk/nomenclature>)
- GeneOntology System und Annotationen (<http://www.geneontology.org>, GeneOntology Consortium 2001)

und Daten aus der zwischen 2000 und 2003 abonnierten kommerziellen Datensammlung

- *ArrayScout*<sup>TM</sup> von LION Bioscience Heidelberg

Zur Abfragen der Klon-Daten wurden Internetseiten programmiert, welche HTML für statische Teile, Java-Script für Benutzer-Interaktionen und PHP für dynamische Inhalte nutzen.

Die *Hypertext Markup Language (HTML)* ist die Basissprache, welche die so genannten Internetbrowser-Programme „verstehen“, um vordefinierte Inhalte als Internet-Seiten darzustellen. Sie leitet sich aus der *Standardized General Markup Language (SGML)* ab und wird vom W3C-Consortium standardisiert (<http://www.w3.org/MarkUp>). Vordefinierte *Tags* (Formatierungs-Zeichen) umschließen in dieser Sprache den Inhalt und beschreiben so dessen Darstellung. Um Seiteninhalte dynamisch darstellen zu können, werden Skript-Sprachen genutzt, die ihrerseits HTML-Code generieren können und somit eine Benutzerinteraktion ermöglichen. *PHP* steht für die Bezeichnung *Hypertext Preprocessor*, die Sprache wird vom PHP-Konsortium standardisiert (<http://www.php.net/>). Verarbeitung der programmierten Funktionalität erfolgt dabei bereits auf dem Server-Computer. Dadurch können sämtliche Daten und andere Programme auf dem Server genutzt werden. Erst nach Abschluss von Berechnungen und Verarbeitungen werden ausschließlich die benötigten Daten über das Netzwerk geschickt. *JavaScript* ist im Gegensatz dazu eine Sprache, die in den *HTML*-Code integriert und auf dem Client-Computer ausgeführt wird. Hiermit werden kleinere Funktionen, die keine weiteren Daten vom Server benötigen, direkt auf dem Rechner des Benutzers ausgeführt.

### 3.2. Das Prozessdatensystem *QuickLIMS*

Das im Rahmen der vorliegenden Arbeit entwickelte Prozessdatensystem *QuickLIMS* wurde zur Kontrolle des Herstellungsvorganges von Microarrays sowie zur Archivierung der dabei involvierten Prozessdaten entwickelt. Das System basiert auf

einer Microsoft Access™ Datenbank (Microsoft Corp. Remond/USA) und wurde in der Sprache *Visual Basic for Applications* (VBA) programmiert. VBA leitet sich von der für das Betriebssystem Windows™ (Microsoft Corp. Remond/USA) entwickelten Sprache Visual Basic mit einem erweiterten Befehlssatz hauptsächlich für Office-Produkte (Microsoft Corp. Remond/USA) ab.

Die Anbindung des benutzten Pipettier-Roboters *Minitrak*, wie auch die Anbindung von *QuickLIMS* an die Klondatenbank *CloneBase*, erfolgt über *Active X Data-Objects* (ADO) und *Open Database Connectivity* (ODBC) Techniken, welche das Ausführen von SQL-Befehlen und den direkten Zugriff auf die Datenobjekte aus der entfernten Datenbank erlauben. Die *Structured Query Language* (SQL) ist die Abfragesprache, die von den gängigen Datenbank-Managementsystemen benutzt wird.

Um für die Produktion der Microarrays eine ausreichende Menge der gewünschten Transkripte zu erhalten, müssen diese mit Hilfe der Polymerase-Kettenreaktion (PCR) direkt aus den Übernachtskulturen von *E.coli* Bakterien vervielfältigt werden. Dies erfolgte in 96-Loch Mikrotiterplatten (AB Gene), die vom Roboter vollautomatisch bearbeitet werden können. Nach Zugabe der Komponenten für die PCR, dem Animpfen der Bakterien und dem Abschluss der Reaktion in PCR-Maschinen, übernimmt der Roboter die Fällung der Produkte durch Zugabe von Isopropanol und Natriumacetat.

Die Fragmente werden schließlich in eine *Spotting*-Lösung überführt. Jeweils vier 96er Platten werden zu einer 384er Klonplatte vereinigt, die vom *Spotting*-Roboter (*OmniGrid Microarrayer* von GeneMachines, USA) weiterverarbeitet werden kann.

### 3.3. Das Primergenerierungs-Werkzeug *AutoPrime*

Zur komfortablen Suche nach geeigneten Primern für die Methode der *Real-Time Polymerase-Kettenreaktion* (RQ-PCR, Wittwer *et al.*, 1989) wurde das Programm *AutoPrime* konstruiert. Es stellt ein Bindeglied zwischen der Sequenzdatenbank *Ensembl* (Hubbart *et al.*, 2002) und eines Programms zur Überprüfung der Primerqualität (*Primer3*, Rozen und Skaletsky 2000) dar. Es wurde in der Sprache Perl geschrieben und nutzt Bibliotheken der *Ensembl Perl-API* (Stabenau *et al.*, 2004). Die Ausgabe des Programmes kann wahlweise als HTML, Text oder XML formatiert werden. Die *eXtensible Markup Language* (XML, <http://www.w3.org/XML>)

ist wie HTML eine von der SGML abgeleitete hochgradig flexible Definitionssprache. Sie wurde für die elektronische Publikation von großen Datenmengen entwickelt und hat sich als Standard zum Austausch von Daten zwischen heterogenen Anwendungen und Geräten etabliert.

Das Programm Primer3 wurde von Rozen und Skaletsky am Whitehead-Institute und Howard Hughes Medical Institute (USA) entwickelt und ist frei erhältlich unter [http://frodo.wi.mit.edu/primer3/primer3\\_code.html](http://frodo.wi.mit.edu/primer3/primer3_code.html).

Die genutzten genomischen Sequenz-Wiederholungs-Bibliotheken wurden nach Antrag und Genehmigung durch J. Walichewicz vom Genetic Information Research Institute (Mountain View, USA) bezogen.

### 3.4. Das Analyseprogramm FACT

Um die Interpretation der Ergebnisse von Hochdurchsatzexperimenten wie DNA-Microarrays zu ermöglichen, wurde das System *Flexible Annotation And Correlation Tool* entwickelt. Es besteht aus einer Kern-Bibliothek mit Funktionen, welche in der Sprache Perl geschrieben ist, einer MySQL Datenbank und Zusatz-Modulen, welchen in Perl bzw. der statistischen Sprache R entwickelt wurden. Die Internet-Schnittstelle nutzt Perl-Skripte, welche dynamisch HTML und JavaScript generieren. Neben HTML kann eine Datenausgabe auch im Graphik-Format PNG (*Portable Network Graphics*-Format) oder als Text oder XML (s.o.) erfolgen.

Die Zusatz-Module verwenden Daten aus folgenden verschiedenen Datenquellen:

- Ensembl Datenbank (DB) Wellcome Trust and European Bioinformatics Institute (GB) (Hubbard *et al.*, 2002)
- IMAGE DB, Lawrence Livermore National Laboratory (Lennon *et al.*, 1996)
- Biological Biochemical Image DB, National Institute of Aging, NIH (USA) (Becker *et al.*, 2000)
- Mouse Genome DB, Jackson Laboratory (USA) (Blake *et al.*, 2003)
- GeneOntology DB, GeneOntology Consortium (GeneOntology Consortium, 2001)
- Cancer Genome Anatomy Projekt, National Cancer Institute, NIH (USA) (Strausberg *et al.*, 2000)
- LocusLink, National Institute of Health (USA) (Pruitt 2002)
- euGenes, University of Indiana (USA) (Gilbert, 2000)
- CloneBase, DKFZ Heidelberg (Kapitel 4.1.)

### 3.5. Microarray-Experimente

Die beschriebenen Experimente zur Untersuchung des *Non-Melanom* Hautkrebses am Modell der Chemikalien-behandelten Rückenhaut der Maus und weitere Microarray-Anwendungen wurden wie folgt durchgeführt (Wrobel, 2004 und Hummerich *et al.*, eingereicht).

#### Probenvorbereitung

Für die Experimente wurden zwei unterschiedliche Microarrays produziert und genutzt. Der erste enthielt sämtliche 20172 cDNA-Fragmente der murinen Klonsammlung *ArrayTAG*<sup>TM</sup> von LION Bioscience (Heidelberg). Der zweite bestand aus 15303 unterschiedlichen murinen cDNA-Fragmenten der Sammlung von National Institute of Aging (Bethesda, USA).

Die gewünschten Sequenzen wurden aus den cDNA-Klonen mit Hilfe der Polymerase-Kettenreaktion (PCR) amplifiziert. Dabei wurden 1 µl einer Übernachtskultur und Vektor-spezifische Primer benutzt. Die Reaktion fand in Primus HT Multiblock PCR-Systeme (MWG Biotech, Weiterstadt) mit folgender Zykleneinstellung statt: 2 min Denaturierung (94°C); 33 Zyklen mit je 15 sec Denaturierung (94°C), 30 sec Anlagerung (55°C), 50 sec Strangverlängerung (72°C); schließlich nochmals 5 min Synthese (72°C). Die DNA war dabei gelöst in 80 µl 10 mM Tris.HCl (ph 8,3 bei 25°C), 50mM KCl, 1,5 mM MgCl<sub>2</sub>, je 0,4 µM der Primer, 0,2 mM jedes dNTPs und 1,2 Units *Taq*-Polymerase (Applied Biosystems, Weiterstadt).

Qualität und Menge der Produkte wurden auf 1,2%igen *Ready-to-Run* Gelen mit 96 Taschen (Amersham Bioscience, Freiburg) überprüft.

Die cDNA wurden durch Alkoholfällung aufgereinigt und getrocknet.

#### „Spotting“

Mit Hilfe eines Roboters (*OmniGrid Microarrayer* von GeneMachines, USA) wurden die Proben auf ca. 100 Objektträgern (*Nexterion Slide E* von Schott Nexterion, USA) aufgebracht (*gespottet*). Die Arrays wurden mit einem *Chipwriter* (Virtek, Waterloo/Kanada) mit 24 Stahlnadeln (*Stealth SMP3 Spotting Pins* von Telechem, Sunnyvale, USA) (Sammlung von LION Bioscience), bzw. mit einem *OmniGrid Microarrayer* (GeneMachines, San Carlos, USA) mit 16 Nadeln (Sammlung vom

National Institute of Aging) gespottet. Die Fragmente wurden als Duplikate gespottet, ferner wurden Kontroll-Sequenzen des *Spot-Report*-Systems (Stratagene, Amsterdam/Niederlande) und Negativ-Kontrollen, wie murine C<sub>0</sub>t-1 DNA (Invitrogen, Groningen/Niederlande), tRNA der Hefe (Sigma-Aldrich, St.Louis, USA) und Poly-dA (Amersham Bioscience, Freiburg) aufgebracht. Durch Hitzebehandlung und UV-Bestrahlung wurden die Fragmente auf der Oberfläche fixiert.

### mRNA Vorbereitung

Die für die Microarray Versuche benötigte mRNA wurde zu verschiedenen Zeitpunkten jeweils aus Gewebeproben der dorsalen Rückenhaut der Mäuse extrahiert. (I) TPA-induzierte Rückenhaut (6h), (II) Papillome (10 Wochen), (III) SCC (~50 Wochen). Die entsprechende Kontroll-Haut von Mäusen aus dem gleichen Wurf wird zeitgleich extrahiert. Dazu wurde die Gesamt-RNA durch ein Trizol- (Invitrogen, Groningen/Niederlande) und RNEasy Midi-Kit (Qiagen, Hilden) -Protokoll extrahiert und aufgereinigt. Nach Qualitätsüberprüfung wurde die PolyA-RNA mittels des Oligotex-Systems (Qiagen, Hilden) gewonnen. Durch Omniscript Reverse Transkriptase (Qiagen, Hilden) mit Oligo-dNTPs und Cy3- bzw. Cy5-markierte dUTPs (Perkin Elmer, Weiterstadt) wurde fluoreszenzmarkierte cDNA generiert (Wrobel 2004).

### Hybridisierung

Die Hybridisierung der Proben fand über Nacht mit den vereinigten Cy3- und Cy5-Proben mit PolyA-RNA (Invitrogen, Groningen/Niederlande), tRNA (Sigma-Aldrich, St.Louis, USA) und C<sub>0</sub>t1-DNA (Invitrogen, Groningen/Niederlande) in *GeneTac* Hybridisierungsmaschinen (Genomic Solutions, Ann Arbor, USA) statt.

### Quantifizierung

Die Signale der hybridisierten Microarrays wurden mit Hilfe eines Scanners (*GenePix 4000B*, Axon Instruments, USA) und der zugehörigen Software (*GenePix Pro 4.0*, Axon Instruments, USA) quantifiziert. Die Vorverarbeitung der Daten (Filterung, Normalisierung) erfolgte mit Hilfe von Skripten in der Sprache R (Wrobel, 2004; Hummerich *et al.*, eingereicht).