

Entwicklung neuartiger Screening-Verfahren  
zur Mutationsdetektion in  
embryonalen Stammzellen der Maus

Von der Naturwissenschaftlichen Fakultät  
der Gottfried Wilhelm Leibniz Universität Hannover  
zur Erlangung des Grades

Doktor der Naturwissenschaften  
Dr. rer. nat.

genehmigte Dissertation  
von

Dipl.-Biol. Ruben Josef Rosenkranz  
geboren am 16. Mai 1979, in Jülich

2010

Referent: Prof. Dr. Christoph Peterhänsel

Koreferent: Prof. Dr. Helge Küster

Tag der Promotion: 22. Oktober 2010

## Zusammenfassung

In der post-genomischen Ära besteht eine der Hauptaufgaben der Wissenschaft in der Aufklärung der Funktion und dem Zusammenspiel vieler Gene. Die Maus besitzt hier als Modellorganismus einen besonderen Stellenwert, da eine Reihe von Methoden zu deren genetischen Veränderung zur Verfügung stehen. In Bezug auf die Analyse der Genfunktionen wäre eine Maus-Stammzellbibliothek ideal, die sämtliche Gene und Mutationstypen umfasst.

Eine preiswerte und schnelle Möglichkeit eine derartige Bibliothek zu schaffen ist die chemische Mutagenese embryonaler Stammzellen (ES) mit N-ethyl-N-nitrosoharnstoff (ENU). Die Behandlung der ES-Zellen mit ENU resultiert in Punktmutationen in der DNA, die zu stillen-, loss of function-, gain of function-Mutationen und Exon-Deletionen führen können. Diese Methode ist ungerichtet und bedarf daher im Nachhinein der Detektion der induzierten Mutationen. Bisher etablierte Screening-Verfahren sind zeitaufwendig und übertreffen den Arbeitsaufwand eines herkömmlichen Knockouts.

Ziel der Arbeit war es, ein neuartiges Gen-getriebenes Screening-Verfahren zu entwickeln, das eine effiziente Transkriptom-weite Detektion ENU induzierter Exon-Deletions-Mutanten erlaubt. Als Grundlage für die Arbeit wurden ES-Zellen mit ENU mutagenisiert, charakterisiert und Test-Bibliotheken angelegt. Es sind zwei Microarray Detektionsvarianten mittels Skip- bzw. Junction-Oligonukleotidsonden erarbeitet worden. Zur *in-silico* Synthese der Sonden wurde eine Software in der Programmiersprache Perl entwickelt. In 'Proof-of-Principle' Experimenten gelang es, die Umsetzbarkeit beider Varianten zu bestätigen und maßgebliche Parameter wie Oligolänge und die Analyse vereinigter Zell-Proben aufzuklären. Das kostengünstigere auf Skip-Oligonukleotidsonden basierende Verfahren wurde optimiert und an der Durchmusterung einer Testbibliothek erprobt. Die Daten zeigten, dass die Anzahl detektierbarer Mutationen dabei weit unter den Erwartungen lag. Die eingeschränkten Möglichkeiten, die Sequenz der Skip-Sonden zu optimieren, resultierten in Schwierigkeiten hinsichtlich Spezifität und Sensitivität des Arrays. Das entwickelte System schien daher für Hochdurchsatz Durchmusterungen von Stammzell-Bibliotheken nicht geeignet zu sein.

In einem zweiten Teil der Arbeit wurde erstmals erfolgreich das Transkriptom einer Maus ES-Zelllinie mittels Illuminas Solexa Sequenzierung analysiert und über speziell angefertigte Softwareprogramme das Potential der Technik zur Detektion von Exon-Spleiß-Mutationen gezeigt. Die geringe Sequenziertiefe, sprich Abdeckung des Transkriptoms, und die vergleichsweise hohen Kosten stehen einer Verwendung dieser Technik als alternativem Detektionsverfahren noch im Wege. Es ist jedoch abzusehen, dass im Zuge der technischen Weiterentwicklung dieses Verfahren an Attraktivität gewinnen wird.

## Abstract

In the post-genomic era, one of the main tasks of life science is to understand the functions and interactions of many genes. The mouse is particularly well suited to address questions concerning gene functions, because the mouse germ line is easy to manipulate. Ideally, a mutant resource covering all mouse genes and different mutation types would be available.

An inexpensive and fast method to generate such a library is the chemical mutagenesis of embryonic stem cells (ES) with N-ethyl-N-nitrosourea (ENU). At the molecular level ENU introduces point mutations into the genome, which can lead to silent-, loss of function-, gain of function-mutations and exon deletions. As ENU mutagenesis is undirected it requires a subsequent detection of induced mutations. Established screening procedures so far are time consuming and exceed the workload of conventional knockouts.

The aim of the study was to develop a novel gene-driven screening process, which effectively enables transcriptome-wide detection of ENU-induced exon deletion mutations. ES cells were mutagenized with ENU and characterized. A small archive of mutagenized ES cells was generated for the purpose of analysis. Two microarray detection variants have been elaborated based on skip-oligonukleotide probes and junction probes respectively. For the *in silico* synthesis of the probes we developed a software in the Perl programming language. In proof-of-principle experiments we succeeded in the confirmation of the feasibility of both microarray options and were able to clarify relevant parameters such as oligo-length and the analysis of pooled samples. The less cost-intensive variant based on the detection via skip-probes was further optimized and tested by screening a subset of the clone archive. The results showed that the number of detectable mutations was far below expectations. Limited opportunities to optimize the skip-sequence in general resulted in the difficulties regarding to specificity and sensitivity of the array. Therefore the developed system is not suitable for high throughput screenings of exon deletion mutation libraries.

Subsequently we were able to successfully analyze the transcriptome of a mouse ES cell line by Illumina Solexa sequencing. Furthermore, custom-made software programs helped us to demonstrate the potential of this technology for the detection of exon splicing mutations. However, at present the low sequencing depth and the comparatively high cost preclude the use of this technique as an alternative detection method of exon deletion mutations. Nevertheless with regard to the rapid technical development in this field of next generation sequencing it is predicted that this detection method might become an attractive alternative.

Schlagworte: Embryonale Stammzellen, ENU-Mutagenese, Durchmusterungsverfahren

Key words: embryonic stem cells, ENU mutagenesis, screening process

---

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung .....</b>	<b>9</b>
1.1	<i>Die Bedeutung des Modellorganismus Maus.....</i>	9
1.2	<i>Embryonale Stammzellen der Maus.....</i>	10
1.3	<i>Mutagenese der Maus.....</i>	12
1.3.1	Phänotyp-getriebene Ansätze .....	12
1.3.2	Gen-getriebene Ansätze .....	14
1.4	<i>Chemische Mutagenese von ES Zellen der Maus mittels ENU.....</i>	16
1.5	<i>Der Gen-getriebene Ansatz zur Mutationsdetektion in ENU mutierten ES-Zellen der Maus.....</i>	19
1.5.1	Mutationsfrequenzen .....	20
1.5.2	Durchmusterungsstrategien .....	20
1.6	<i>Ziele dieser Arbeit.....</i>	22
<b>2</b>	<b>Material und Methoden .....</b>	<b>23</b>
2.1	<i>Material .....</i>	23
2.1.1	Allgemeine Chemikalien .....	23
2.1.2	Allgemeine Verbrauchsmaterialien .....	23
2.1.3	Geräte und spezielle Materialien .....	23
2.1.4	Reaktionskits .....	25
2.1.5	Spezielle Chemikalien und Lösungen .....	25
2.1.6	Enzyme.....	26
2.1.7	Häufig verwendete Medien, Puffer und Lösungen.....	26
2.1.8	Maus-Zelllinien .....	28
2.1.9	Maus Gehirn-Gewebe .....	28
2.2	<i>Zellbiologische Methoden.....</i>	28
2.2.1	Ernten von Zellen .....	28
2.2.2	Zählen von Zellen und Zellkolonien .....	29
2.2.3	Einfrieren und Auftauen von Mauszellen.....	29
2.2.4	Embryonale Fibroblasten der Maus.....	29
2.2.5	Kultur embryonaler Stammzellen der Maus.....	31
2.2.6	Karyotypisierung embryonaler Stammzellen .....	32
2.2.7	Chemische Mutagenese embryonaler Stammzellen .....	33
2.2.8	Mutationsfrequenzen am <i>Hprt</i> Locus .....	33
2.3	<i>Molekularbiologische Methoden .....</i>	34
2.3.1	Agarose-Gelelektrophorese .....	34
2.3.2	Spektrophotometrie .....	35
2.3.3	Isolation von Gesamt-RNA .....	35

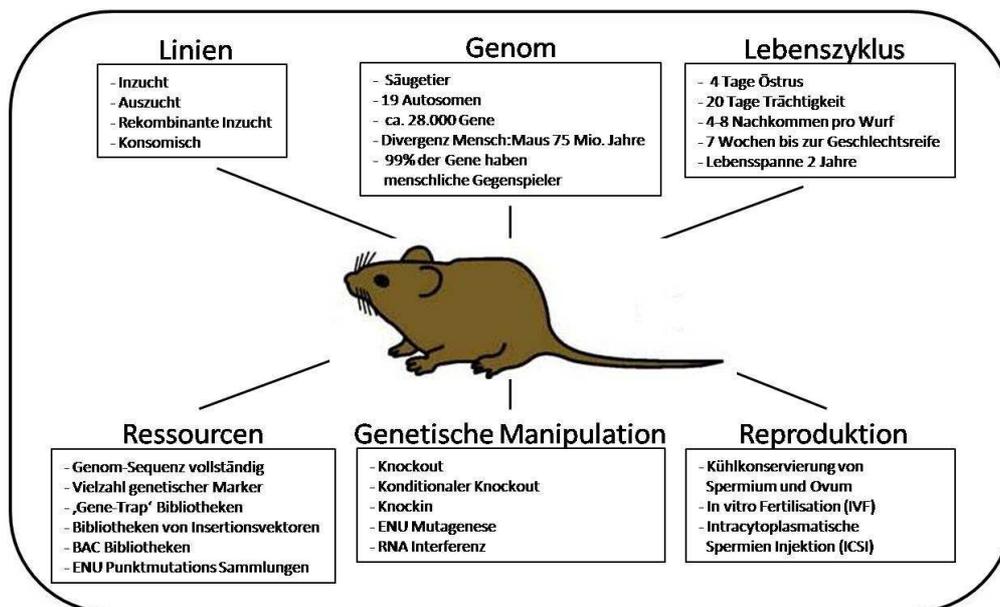
2.3.4	Isolation von messenger RNA.....	36
2.3.5	Synthese einzelsträngiger cDNA.....	37
2.3.6	Synthese doppelsträngiger cDNA .....	37
2.3.7	Polymerase-Kettenreaktion (PCR).....	38
2.3.8	Microarray-Analysen .....	39
2.3.9	Sequenzierverfahren.....	42
2.4	<i>Bioinformatische Software und Datenbanken.....</i>	44
2.4.1	Software .....	44
2.4.2	Spezielle Software.....	45
2.4.3	Datenbanken.....	45
<b>3</b>	<b>Versuchsbeschreibungen und Ergebnisse .....</b>	<b>46</b>
3.1	<i>ENU-Mutagenese.....</i>	46
3.1.1	Generierung einer Test-Bibliothek ENU mutierter ES-Zellen .....	46
3.1.2	ENU Mutationsspektren.....	47
3.1.3	Erwartete Häufigkeit von Exon-Deletionsmutanten.....	49
3.2	<i>Genexpressions Untersuchungen.....</i>	51
3.2.1	Genexpression in Wildtyp F1 ES-Zellen.....	51
3.2.2	Transkriptom-Analysen ENU mutierter F1 ES-Zellen.....	52
3.3	<i>Microarray basierte Detektion von Exon-Deletions Mutanten.....</i>	54
3.3.1	Allgemeine Vorüberlegungen zum Array Design .....	54
3.3.2	Entwicklung einer Software zur Erstellung von Skip- und Junction-Oligos.....	55
3.3.3	Technische Vorüberlegungen zum Microarray und zum Array Design.....	57
3.3.4	Machbarkeitsstudie zur Detektion von Exon-Deletionen.....	58
3.3.5	Pooling-Experimente mit Positiv-Kontrollen und ENU-mutierten ES-Zell-Klonen.....	63
3.3.6	Auswertung der ersten Pooling-Experimente hinsichtlich neuer Exon-Deletionen .....	67
3.3.7	PCR basierte Bestätigung möglicher Exon-Deletionen.....	69
3.3.8	Durchmusterung einer Test-Bibliothek ENU mutierter ES-Zellen .....	71
3.3.9	Neugestaltung des Microarray Aufbaus .....	75
3.4	<i>Detektion von Mutationen mittels Hochdurchsatzsequenzierung .....</i>	78
3.4.1	Aufbau des Experiments .....	79
3.4.2	Entwicklung einer Exon-Deletions Referenzdatenbank.....	79
3.4.3	Sequenzierung des ES-Zell Transkriptoms .....	82
3.4.4	Abdeckung des Transkriptoms und Fehlerquellen .....	83
3.4.5	Durchmusterung der Sequenzdatensätze nach Exon-Deletionen .....	87
<b>4</b>	<b>Diskussion.....</b>	<b>93</b>
4.1	<i>ENU-Mutagenese.....</i>	94
4.2	<i>Theoretische Detektionswahrscheinlichkeiten auf Transkript Ebene .....</i>	96

4.3	<i>Entwicklung eines Mikroarrays zur Detektion von Exon-Deletionen</i> .....	98
4.3.1	Die Wahl der Technik .....	98
4.3.2	Grundlagen der Micorarray-basierten Detektion.....	101
4.3.3	Untersuchungen zur Machbarkeit.....	104
4.3.4	Detektion und Bestätigung neuer Mutations Ereignisse.....	107
4.3.5	Umsetzung des entwickelten Verfahrens .....	108
4.3.6	Abschließende Optimierungsversuche .....	109
4.4	<i>Transkriptom Charakterisierung mittels Sequenzierung</i> .....	110
4.4.1	Sequenzierung des Transkriptoms muriner Stammzellen .....	112
4.4.2	Abgleich der Sequenzdaten gegen die Exon-Skip Datenbank db42.....	114
4.5	<i>Potential der Transkriptom Durchmusterungsstrategien</i> .....	116
<b>5</b>	<b>Literaturverzeichnis</b> .....	<b>118</b>
<b>6</b>	<b>Anhang</b> .....	<b>126</b>
6.1	<i>Abbildungsverzeichnis</i> .....	126
6.2	<i>Tabellenverzeichnis</i> .....	127
6.3	<i>Oligonukleotid-Primer</i> .....	128
6.4	<i>Abkürzungen</i> .....	130
<b>7</b>	<b>Lebenslauf des Verfassers</b> .....	<b>133</b>
<b>8</b>	<b>Publikationen</b> .....	<b>134</b>

# 1 Einleitung

## 1.1 Die Bedeutung des Modellorganismus Maus

Innerhalb des letzten Jahrhunderts hat sich die Maus (*Mus musculus*) zu dem herausragenden Modellorganismus für genetische Forschung an Säugetieren entwickelt (Brown et al., 2006). So war es auch das Genom der Maus, dessen Sequenz nach der Sequenzierung des menschlichen Genoms als nächstes Säugetiergenom entschlüsselt werden sollte (Lander et al., 2001; Mouse-Sequencing-Consortium 2002; Venter et al., 2001).



### Abbildung 1: Der Modellorganismus Maus

Die Abbildung zeigt einige herausragende Eigenschaften des Modellorganismus Maus (*Mus musculus*) mit besonderem Hinblick auf das genetische Modell Maus (verändert nach (Hacking 2008)).

Das große Interesse der Wissenschaft an der Maus liegt u.a. darin begründet, dass trotz 75 Millionen Jahren getrennter Evolution ihre anatomische und physiologische Komplexität große Ähnlichkeit zum Menschen aufweist. Zudem sind sich die Genome von Maus und Mensch ähnlich (40%-ige Identität der Sequenzen; (Mouse-Sequencing-Consortium 2002)). Auf Proteinebene lassen sich für mehr als 80% der Proteine der Maus Orthologe im Menschen identifizieren (Pennacchio 2003). Die Maus bietet sich daher als Modellsystem für Studien an, die zum besseren Verständnis der komplexen Funktionsweise des menschlichen Organismus und dessen Krankheiten beitragen (Nolan et al., 2002).

Im Vergleich zu anderen Säugetieren sprechen für die Verwendung als Modellorganismus Eigenschaften wie die einfache Zucht, die hohen Wurfgrößen und die kurze Generationszeit. Aufgrund ihrer geringen Größe ist die Haltung zudem platzsparend und damit kosteneffizient.

Weiterhin steht eine große Anzahl stabiler Inzuchtlinien zur Verfügung, die Untersuchungen an genotypisch einheitlichen Tieren ermöglichen und die Experimente verschiedener Laboratorien vergleichbar machen (Judd 2001).

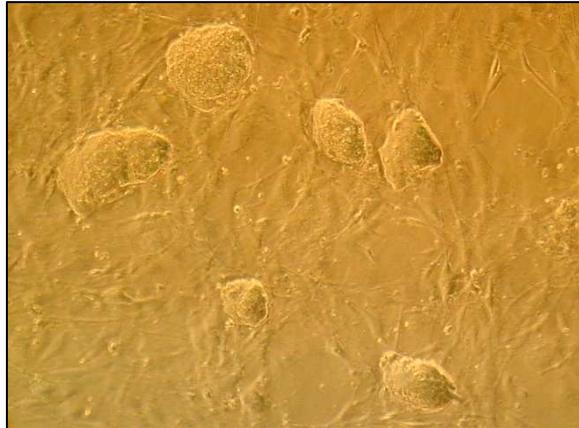
Die Beliebtheit der Maus als Modellorganismus der biologischen Forschung ist heute mit hoher Wahrscheinlichkeit auch darauf zurückzuführen, dass bis dato für kein anderes Wirbeltier so vielfältige Techniken zur Manipulation des Genoms etabliert wurden wie für die Maus (Beckers et al., 2002). An dieser Stelle seien hier beispielhaft die DNA Injektion in Pronuclei von Embryonen (Gordon et al., 1981) und die Infektion von Embryonen mit rekombinaten retroviralen (lentiviralen) Vektoren (Lois et al., 2002) genannt. Besonders hervorzuheben ist die Generierung chimärer Mäuse durch Manipulation der DNA embryonaler Stammzellen mittels homologer Rekombination (Thomas et al., 1987) und der anschließenden Mikroinjektion dieser Zellen in das Innere einer Blastozyste.

## **1.2 Embryonale Stammzellen der Maus**

Die ersten embryonalen Stammzellen (embryonic stem cells, ES Zellen) der Maus konnten 1981 aus der inneren Zellmasse (inner cell mass, ICM) von Maus Blastocysten isoliert und in Kultur gehalten werden. Dabei wurden Kultur Bedingungen verwendet, die bereits für die Kultur von embryonalen Karzinomzellen (embryonic carcinoma, EC cells) etabliert waren (Evans et al., 1981; Martin 1981). Die ES Zellen waren in der Lage, sich *in vitro* in verschiedenste Zelltypen zu differenzieren und nach Injektion in Mäuse Keimzelltumore (Teratome) hervorzurufen (Martin 1981). Kurz nach diesen grundlegenden Feststellungen konnte gezeigt werden, dass die ES Zellen nach Injektion in Blastocysten, anders als EC Zellen, in hohem Maße zu allen Geweben chimärer Mäuse, inklusive der Keimbahn, beitragen konnten (Bradley et al., 1984). Dies eröffnete einen praktischen Weg, Modifikationen in die Keimbahn von Mäusen und deren Nachkommen einzuschleusen. Als genetischer Hintergrund wurden mit wenigen Ausnahmen vornehmlich Linien wie 129, C57BL/6, DBA und deren Hybride zur Isolation embryonaler Stammzellen verwendet. Die meisten etablierten ES Zelllinien weisen den männlichen Genotypen (XY) auf, was insbesondere auf die Instabilität weiblicher ES-Zelllinien zurückzuführen ist (Zvetkova et al., 2005).

Die beiden markanten Eigenschaften von Stammzellen sind ihre Fähigkeit zur Selbsterneuerung und ihre Pluripotenz, d.h. die Fähigkeit Zellen aller somatischen Gewebe und der Keimbahn zu bilden. Eine funktionelle Definition embryonaler Stammzellen sollte daher zwei Nachweise beinhalten: Zum einen müssen die Zellen pluripotent sein, d.h. zur Bildung aller drei embryonaler Keimblätter (Ektoderm, Endoderm, Mesoderm) beitragen

können und zum anderen müssen sie in der Lage sein über einen längeren Zeitraum in Kultur die Pluripotenz aufrecht zu halten (Meissner et al., 2009).



**Abbildung 2: Lichtmikroskopische Aufnahme embryonaler F1 Stammzellkolonien**

Gezeigt sind sechs Kolonien embryonaler Stammzellen der Maus (F1-Stammzelllinie [2.1.8], Passage 18) auf einem 'feeder layer' aus CD1 embryonalen Fibroblasten 2 Tage nach dem Ausplattieren. Die rundliche kugelige Form ist eine typische morphologische Eigenschaft undifferenzierter Stammzellkolonien. Die Aufnahme ist eine 100-fache lichtmikroskopische Vergrößerung.

Für den Erhalt des ES Zell Phänotyps wurden in den letzten Jahren eine Reihe intrinsischer Faktoren wie z.B. die beiden Transkriptionsfaktoren Oct-4 und Nanog und extrinsische Faktoren wie z.B. LIF (leukemia inhibitory factor) und BMP (bone morphogenetic protein) identifiziert (Meissner et al., 2009; Yu et al., 2008). Während erstere oft als Marker für den pluripotenten ES-Zellstatus verwendet werden (Ohtsuka et al., 2008), werden letztere den Zellen mit dem Medium zur Verfügung gestellt. Das Cytokin LIF wird zum einen von einer Schicht mitotisch inaktivierter embryonaler Fibroblasten (MEF, mouse embryonic fibroblast), auf der ES Zellen in der Regel kultiviert werden (Abbildung 2), den Zellen zur Verfügung gestellt, oder es wird dem Medium als rekombinantes Protein beigefügt. LIF ist ein Cytokin der IL6 Familie welches den gp130 Rezeptor bindet und auf diesem Weg den JAK-STAT3 und Ras-MAPK Signalweg in ES Zellen der Maus aktiviert. Anhand von Studien konnte gezeigt werden, dass LIF *in vivo* keinen Einfluss auf die Entwicklung nimmt. Vielmehr scheint es eine Rolle bei der Aufrechterhaltung des Blastocysten bei verzögerter Einnistung des Prä-Implantationsembryos in den Uterus zu spielen (Nichols et al., 2001). BMP wird dem Medium als Bestandteil von fötalem Kälber Serum (fetal calf serum, FCS) zugefügt. Dieses Protein induziert die Expression von *Id* (inhibitor of differentiation) durch den Smad Signalweg und anderer Pro-Differenzierungs-Signalwege (Ohtsuka et al., 2008) und verhindert somit die Differenzierung der ES Zellen.

Embryonale Stammzellen der Maus spielen heute eine entscheidende Rolle in der funktionellen Genomik, d.h. in der Funktionsbeschreibung bisher uncharakterisierter Gene insbesondere im Zusammenhang mit Studien zur Entwicklung und zu Krankheiten. Möglich ist dies durch die bereits erwähnte Fähigkeit der ES-Zellen zur homologen Rekombination (Thomas et al., 1987). Zudem rückt neuerdings das potentielle therapeutische Potential der ES-Zellen vornehmlich in den Vordergrund. Dies wird durch Reprogrammierungs-Studien unterstützt in denen ES Zellen bzw. ES Zell ähnliche Zellen (iPS, induzierte pluripotente Stammzellen) durch nukleären Transfer und Transfektion bestimmter regulatorischer Schlüssel-Gene aus differenzierten Zellen regeneriert werden konnten (Gurdon et al., 2008; Jaenisch et al., 2008).

### **1.3 Mutagenese der Maus**

Ein fundamentales Ziel der heutigen biologischen Forschung des Post Genomischen Zeitalters besteht in der Zuordnung von Funktionen zu Genen, deren Sequenz und Expressionsverhalten weitestgehend bekannt sind (funktionelle Genomik). Dabei ist nicht nur die normale Funktion von Interesse, sondern insbesondere auch die Rolle der Gene bei der Entstehung menschlicher Krankheiten, die auf Mutationen beruhen (Oliver et al., 2007). Einer der effektivsten Wege die Funktion eines Gens zu studieren ist die *in vivo* Analyse der Phänotypen individueller Maus-Mutanten.

Die Rate spontaner Mutationen ist allerdings mit ca.  $1 \times 10^{-6}$ /Locus viel zu gering, um auf natürlichem Wege in kurzer Zeit eine ausreichende Anzahl interessanter Phänotypen erhalten zu können (Pretsch et al., 1994). Dies führte zu der Entwicklung einer Vielzahl von Strategien und Techniken zur Mutagenese der Maus.

Generell können vorwärts und rückwärts gerichtete Strategien unterschieden werden ('forward genetics', 'reverse genetics'). Bei den vorwärts gerichteten Strategien ist der Ausgangspunkt typischerweise ein bestimmter Phänotyp, gefolgt von der Charakterisierung des mutierten Gens. In den rückwärts gerichteten Ansätzen ist das Gen und dessen Sequenz bereits bekannt. Die Analyse der Funktion basiert dann auf der Veränderung des Gens (Zerstörung, Überexpression o.ä.) und der nachfolgenden Untersuchung des resultierenden Phänotyps. Man spricht bei den beiden Ansätzen daher auch von Phänotyp- bzw. Gen-getriebenen Ansätzen (Soewarto et al., 2009).

#### **1.3.1 Phänotyp-getriebene Ansätze**

Bei Phänotyp-getriebenen Ansätzen werden im Vorfeld keine Annahmen zu den genetischen Ursachen einer Erkrankung getroffen, sondern diese erst nach genauer Analyse des Phänotyps

ermittelt. Die Mutagenese erfolgt in der Regel über ungerichtete physikalische bzw. chemische Verfahren, die somit weitestgehend keiner Sequenzspezifität unterliegen und prinzipiell jedes Gen mutieren können. Zu den physikalischen Verfahren zählt beispielsweise die Röntgenbestrahlung ganzer Tiere (Muller 1927), die jedoch aufgrund der geringen Mutationseffizienz an Bedeutung verloren hat. Zudem fallen die Ausprägungen der erzielten Mutationen (in der Regel Umlagerungen und Deletionen) sehr unterschiedlich aus. Dennoch wird die Bestrahlung von Zellen heute noch zur Generierung großer chromosomaler Deletionen (mehrere Centimorgan) genutzt, die z.B. bei der Aufklärung der Konsequenzen segmentaler Haploidie oder der funktionellen Analyse bestimmter chromosomaler Bereiche von großem Nutzen sind (Schimenti et al., 2000; Thomas et al., 1998).

Von weitaus größerer Bedeutung als die physikalische Mutagenese ist die chemische Mutagenese, was in der höheren Effektivität und der weniger komplexen Ausprägung der induzierten Mutationen begründet liegt. Als Beispiele für chemische Mutagene seien hier N-Ethyl-N-nitrosoharnstoff (ENU), Triethylmelamin (TEM), Procarbazin, Chlorambucil, Ethylmethansulfonat (EMS), Cyclophosphamid und Diethylsulfat genannt. Ende der 70er Jahre erkannte dann Bill Russel, dass ENU (N-Ethyl-N-nitrosoharnstoff) das Agenz mit dem höchsten Mutagenisierungspotential für die Produktion beliebiger Genmutationen in der Maus ist (Russell et al., 1979). Männlichen adulten Mäusen (Generation G<sub>0</sub>) wird ENU mehrmals intraperitoneal injiziert, wodurch Punktmutationen in sämtlichen Körperzellen inklusive der prämeiotischen Spermatogonien hervorgerufen werden können (Justice et al., 2000). Nach der Verpaarung von Männchen (G<sub>0</sub>) mit mutierten prämeiotischen Spermatogonien sind die Nachkommen (G<sub>1</sub>) bereits heterozygot auf bestimmte Mutationen und können sofort auf dominante Mutationen untersucht werden (Rinchik 1991; Russell et al., 1982). Um auf rezessive Mutationen (Homozygote) screenen zu können, werden die G<sub>2</sub> Nachkommen miteinander oder mit der Ausgangslinie der Generation G<sub>1</sub> verpaart. Ist ein bestimmter Phänotyp feststellbar, muss im Anschluss daran der mutierte Locus mittels Kreuzungen unter Verwendung genetischer Marker eingegrenzt und mittels Positionsklonierung ermittelt werden. Die Generierung einer großen Anzahl verschiedener Phänotypen und die anschließende Charakterisierung der Mutationsereignisse auf molekularer Basis ist sehr aufwendig und wird daher nur in großen Mutageneseprogrammen durchgeführt (Hrabe de Angelis et al., 2000; Nolan et al., 2000; Rathkolb et al., 2000). Die Mutationsraten betragen ca. 0,4 bis 1 Punktmutationen pro Mbp, sodass bei einer Annahme von 1-3% codierender Sequenz 10-25 funktionelle Mutationen pro Genom erwartet werden können (Soewarto et al., 2009). Dieser Phänotyp-getriebene Ansatz hat hinsichtlich der Erforschung menschlicher

Krankheiten gegenüber den Gen-getriebenen Ansätzen, die auf der gezielten Veränderung bestimmter ausgewählter Gene basieren, den klaren Vorteil, dass keine Annahmen zur genetischen Ursache eines bestimmten Phänotyps im Vorfeld getroffen werden müssen. Ausgehend von einem entsprechenden Phänotyp können die genetischen Ursachen gezielt untersucht werden, die oft auf den verschiedensten Veränderungen von Allelen beruhen (Null Allele, Allele mit eingeschränktem Funktionsverlust oder veränderter Funktion), für die sich im Vorfeld nur schwer Annahmen treffen lassen (Acevedo-Arozena et al., 2008b).

### 1.3.2 Gen-getriebene Ansätze

Die Gen-getriebenen Ansätze der Mutagenese der Maus basieren in der Regel auf molekularbiologischen Methoden, mit denen spezielle Vektoren generiert werden, die dann in das Genom, im Wesentlichen von ES-Zellen, integriert werden. Hier unterscheidet man zwischen einer zielgerichteten und einer zufälligen Integration in das Genom. Die so *in vitro* veränderten embryonalen Stammzellen der Maus können nach genauer Charakterisierung zur Injektion in Blastocysten für die Generierung chimärer Tiere genutzt werden.

Die Methode der Genfallen (gene-trap) ist ein Beispiel für eine ungerichtete Integration eines Vektors in das Genom. Der gene-trap Vektor weist drei charakteristische Komponenten auf: Eine Spleiß-Akzeptor Stelle, ein promotorloses Reportergen (i.d.R. *lacZ*) und einen Selektionsmarker mit eigenem Promotor (z.B. das Neomycinresistenz-Gen Neomycin-Phosphotransferase, *neo*). Die Bestätigung der Integration des Vektors in das Genom erfolgt dann durch eine positive Selektion auf Neomycin resistente Zellen. Integriert der Vektor in ein Intron so wird mittels der integrierten Spleiß-Akzeptor Sequenz ein Fusionsprotein aus dem getroffenen Gen und dem Reporter-Gen gebildet, auf welches selektioniert werden kann. Die Integration in ein Intron bewirkt daher in der Regel einen Verlust der Proteinfunktion und kann so zur Generierung von Null-Mutanten genutzt werden. Diese Standardvariante der Gene-Trap Mutagenese hat den Nachteil, dass eine Selektion die Expression des Gens in ES-Zellen voraussetzt (Cecconi et al., 2000). In einem weiter entwickelten System wurde daher das Poly-Adenylierungssignal des unter konstitutiver Expression stehenden *neo* Gens gegen eine Spleiß-Akzeptor Stelle ausgetauscht. Das hat zur Folge, dass nur bei einer Integration innerhalb eines Gens Neomycin resistente Kolonien auftreten, da nur dann über die Fusion mit der Poly-Adenylierungssequenz des betroffenen Gens die Generierung eines funktionsfähigen Proteins möglich ist (Niwa et al., 1993). Zur anschließenden Identifizierung des getroffenen Gens wird in Gene-Trap Mutagenesen die Vektorsequenz als Matrize für eine 5'-RACE-PCR (rapid amplification of cDNA ends) genutzt. Da die Mutagenese ungerichtet stattfindet, ist dies keine Methode, mit der Forscher Mutanten für das sie speziell interessante

Gen generieren können. Vielmehr haben sich bestimmte Gruppen gebildet, die großangelegte Durchmusterungen (Screenings) durchführen und ES-Zell Bibliotheken anlegen, aus denen interessante Mutanten von Einzelnen erworben werden können (Araki et al., 2009; Friedel et al., 2007; Hansen et al., 2003; Skarnes et al., 2004).

Die wohl bekannteste und auch bedeutendste Gen-getriebene Strategie zur Geninaktivierung ist die des Gene-Targeting, d.h. der zielgerichteten Einführung einer Vektorsequenz in einen bestimmten genetischen Locus basierend auf homologer Rekombination in ES-Zellen. Dies wurde erstmalig am Beispiel des Hypoxanthin Phosphoribosyl Transferase Gens (*hprt*) der Maus erfolgreich durchgeführt (Doetschman et al., 1987; Thomas et al., 1987).

Der verwendete Vektor enthält dabei zwei der Insertionsstelle homologe Bereiche, die die auszutauschende Sequenz umschließen. Darin enthalten ist unter anderem ein Gen zur Positiv-Selektion auf die Integration (i.d.R. *neo*). Ausserhalb des Bereichs befindet sich in der Regel ein Negativ-Selektions Marker, über den Selektionsdruck auf die korrekte zielgerichtete homologe Rekombination ausgeübt werden kann (z.B. Thymidin Kinase Gen) (Mansour et al., 1988). Im einfachsten Fall, der Generierung einer Nullmutation kann so beispielsweise ein Exon des zu zerstörenden Gens gegen eine Neomycin Resistenz ausgetauscht werden. Neben der Generierung von Null Mutationen ('knock-outs', KO) können mit dieser Technologie über sogenannte 'knock-in' Strategien je nach Aufbau des Vektors eine Vielzahl anderer Modifikationen am Genom der Maus vorgenommen werden. Dazu zählen beispielsweise die Einführung von Deletionen, Substitutionen von Proteindomänen oder sogar die Einführung von Punktmutationen (Sung et al., 2004). Zudem lassen sich über Weiterentwicklungen unter Verwendung des Cre/LoxP Systems aus Bakteriophage P1 und des FLP/FRT Systems aus Hefe mittlerweile konditionale Geninaktivierungen durchführen (Gu et al., 1994; Vooijs et al., 1998). Solche Strategien können unter anderem auch dazu verwendet werden, die artifizielle Situation an der Integrationsstelle auf ein Minimum zu reduzieren.

Dennoch weist diese Technik den Nachteil auf, dass das Design der spezifischen Vektoren und die Selektion der korrekten Klone viel Zeit und Arbeitsaufwand erfordert, was insbesondere die Generierung verschiedener Allele für ein bestimmtes Gen erschwert. Doch sind es gerade die verschiedenen Allele eines Gens im Modell Maus, die als Grundlage für die Untersuchung menschlicher somatisch erworbener Krankheiten von Bedeutung sind (Oliver et al., 2007).

Zur Zeit existieren für ca. 50% der bekannten Gene der Maus Knockout-Modelle. Ziel eines Projektes mit dem Namen 'Knockout Mouse Project' (KOMP) des National Institute of Health

ist es in naher Zukunft eine nahezu vollständige öffentliche Bibliothek von Knockout Mutanten zu erstellen (Austin et al., 2004).

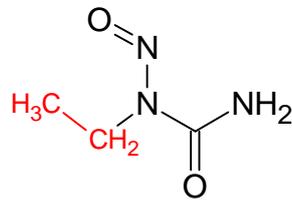
Eine weiterer bedeutender Ansatz zur Inaktivierung von Genen basiert auf der RNA Interferenz (Fire et al., 1998; Tuschl 2001). Lange doppelsträngige RNA Moleküle werden dabei von dem Enzym *DICER* in 21 bis 23 bp lange Fragmente zerschnitten (small interfering RNAs, siRNA). Ein Strang wird dann von einem weiteren Enzym unter Bildung des *Risc* Komplexes (RNAi silencing complex) aufgenommen, der RNA Moleküle an der komplementären Sequenz schneidet und somit die Translation dieses Gens mindert. Der Vorgang wird daher auch als post-transkriptionales Silencing bezeichnet und funktioniert bei fast allen höher entwickelten Lebewesen u.a. auch dem Säugetier (Hannon 2002). Diese biologische Antwort auf doppelsträngige RNA kann man sich als Werkzeug zur Manipulation der Genexpression und damit zur Generierung von Mutationen machen. Dazu führt man beispielsweise vektorvermittelt eine dem Zielgen komplementäre short hairpin RNA (shRNA) in das Genom einer ES-Zelle ein und bewirkt so über die Expression der shRNA die Herabregulation der Expression des Zielgens. Im Idealfall kann so ein dem einer Nullmutation entsprechender Phänotyp erzielt werden (Kunath et al., 2003). Die Methode bietet den Vorteil, dass durch geeignete Wahl des Promotors einfach konditionale und gewebespezifische Mutationen erreicht werden können. Der Nachteil der Technik ist in dem post-transkriptionellen Mechanismus an sich zu sehen, der nicht die Möglichkeit zur Generierung verschiedener Allele eines Gens bietet.

Der in dieser Arbeit verfolgte Gen-getriebene Ansatz beruht auf der chemischen Mutagenese von embryonalen Stammzellen der Maus mit dem chemischen Mutagen N-Ethyl-N-nitrosoharnstoff (ENU). Dieser Ansatz wird im nachfolgenden Absatz im Detail erläutert, doch sei hier bereits angemerkt, dass die Mutagenese an sich ungerichtet erfolgt, die Selektionstrategie jedoch Gen-getrieben erfolgen kann.

#### **1.4 Chemische Mutagenese von ES Zellen der Maus mittels ENU**

Anders als bei den weitverbreiteten Knockout Modellen bietet die Mutagenese mittels ENU den entscheidenden Vorteil, dass ENU Punktmutationen in die funktionellen Domänen von Genen einfügen kann. Dadurch können nicht nur Null-Mutationen generiert werden, sondern auch verschiedenste Allele eines Gens (Hypomorphe, hypermorphe oder neomorphe Mutationen und Spleiß Varianten) (Noveroske et al., 2000), die besser die Situation in menschlichen Patienten reflektieren, deren genetische Krankheiten bzw. Prädispositionen auf

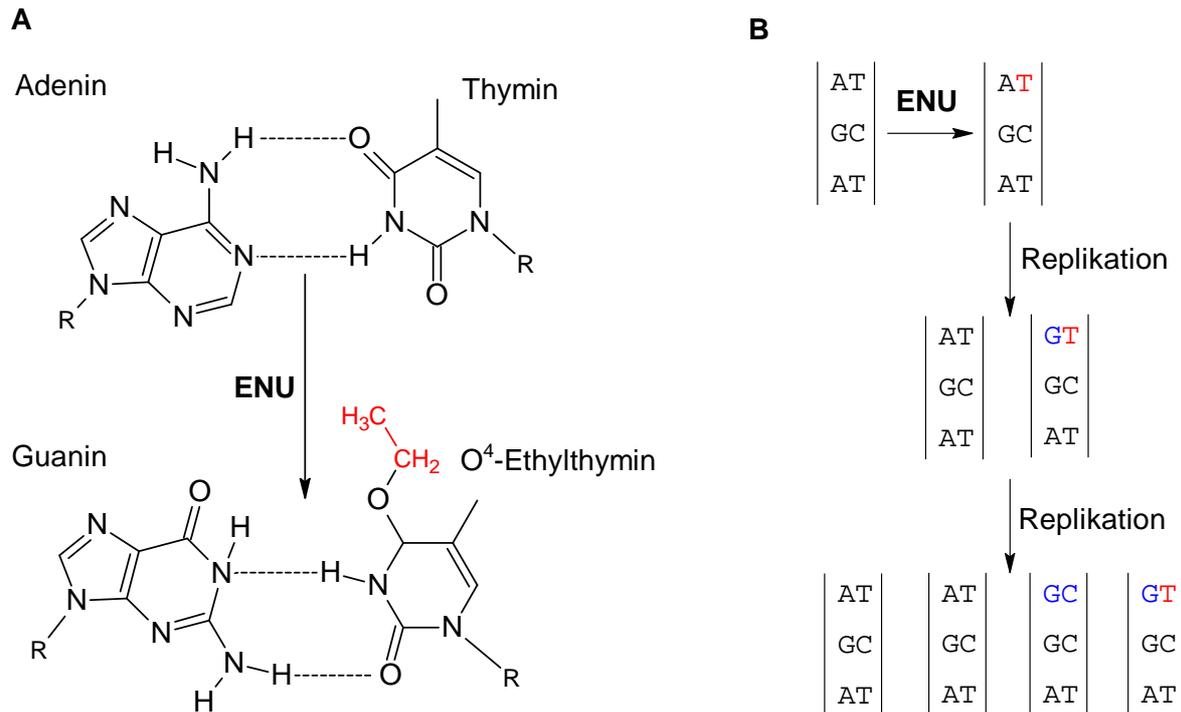
einzelnen Nukleotid Polymorphismen (SNP) beruhen (Soewarto et al., 2009). Die Mutagenese ist zudem nicht vektorvermittelt und hinterlässt keine artifiziellen Vektorsequenzen.



**Abbildung 3: Molekulare Struktur von N-Ethyl-N-nitrosoharnstoff (ENU)**

ENU ist ein synthetisches alkylierendes Agens, welches speziell zu Mutagenese genutzt wird. Die übertragbare Ethylgruppe ist rot hervorgehoben.

Bei ENU handelt es sich um ein synthetisches alkylierendes Agens, welches ursprünglich produziert wurde, um prinzipielle Mechanismen der Mutagenität zu erforschen (Abbildung 3). ENU wirkt direkt alkylierend auf Nukleinsäuren, d.h. eine Verstoffwechslung der Substanz im Vorfeld ist nicht notwendig (Noveroske et al., 2000). Die Ethylgruppe wird dabei über einen  $SN_1$ -Mechanismus direkt auf Sauerstoff- oder Stickstoffradikale übertragen, deren Vorkommen *in vitro* und *in vivo* bestätigt werden konnten. Hierzu zählen im Speziellen die  $N^1$ ,  $N^3$  und  $N^7$  Gruppen des Adenins, die  $O^6$ ,  $N^3$  und  $N^7$  Gruppen des Guanins, die  $O^3$ ,  $O^4$  und  $N^3$  Gruppen des Thymins und die  $O^3$  und  $N^3$  Gruppen des Cytosins (Shibuya et al., 1993). Die Zelle reagiert auf die DNA Addukte mittels zelleigener Reparatursysteme. Hierzu zählen beispielsweise die Nukleotid Exzisions Reparatur (NER) und die Reparatur mittels der  $O^6$ -Alkylguanin-DNA-alkyltransferase (AGT). Letzteres Enzym ist vermutlich dafür verantwortlich, dass generell seltener Mutationen an GC-Basenpaaren (~32-13%) als an AT-Basenpaarungen festgestellt werden konnten. Werden die Ethylgruppen jedoch z.B. aufgrund einer Sättigung der Reparatursysteme nicht entfernt, werden die betroffenen Basen im Zuge der nächsten DNA-Replikation nicht richtig erkannt und falsche Nukleotide eingebaut (Abbildung 4). Nach zwei Replikationsrunden hat sich ein einzelner Basenaustausch dann manifestiert und kann nicht mehr von den zelleigenen Reparatursystemen erkannt werden (Noveroske et al., 2000). Generell besteht eine direkte Abhängigkeit der eingesetzten ENU Mengen zu der erhaltenen Mutationsrate. Allerdings können zu hohe Dosen dazu führen, dass die Tiere die Mutagenese nicht überleben oder steril bleiben bzw. Stammzellen sterben oder ihre Keimbahngängigkeit verlieren (Chen et al., 2000; Justice et al., 2000; Munroe et al., 2000; Weber et al., 2000). Andererseits existiert auch ein Grenzwert unter dem keine Mutationen feststellbar sind, da die DNA-Reparatursysteme nicht gesättigt sind (Noveroske et al., 2000).



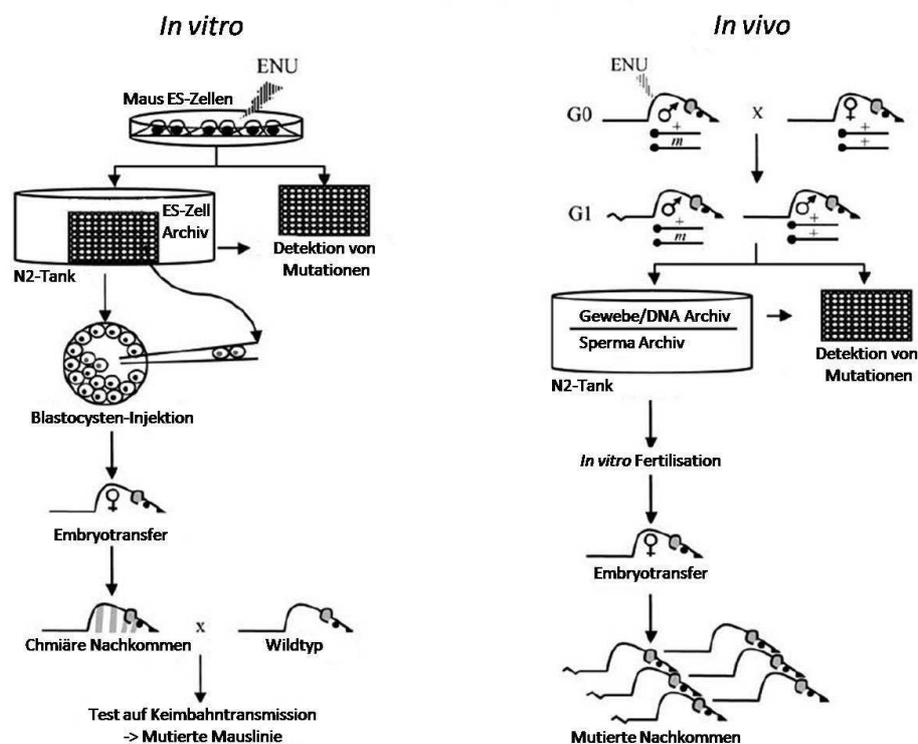
**Abbildung 4: Die mutagene Reaktion von ENU**

**A:** Bildung eines Ethyl-DNA Adduktes und Fehlpaarung von O<sup>4</sup>-Ethylthymin. R steht für den Rest der DNA. Die von ENU übertragene Ethylgruppe ist rot gekennzeichnet **B:** Mechanismus eines ENU induzierten Basen Austauschs. O<sup>4</sup>-Ethylthymin ist in rot dargestellt. Blaue Buchstaben heben die Basensubstitutionen hervor (Abb. in Anlehnung an Noveroske et al., 2000).

Prinzipiell sind 70 - 85% aller ENU induzierter Basen-Substitutionen A-T zu T-A Transversionen oder A-T zu G-C Transitionen. G-C zu C-G Transversionen werden in den wenigsten Fällen beobachtet. Die Basen-Substitutionen resultieren zu ca. 70% in nicht-synonymen Aminosäure-Substitutionen von denen ca. 65% Missense Mutationen und die restlichen Nonsense- bzw. Spleiß-Mutationen darstellen. Spleiß-Mutationen, gehen auf Punktmutationen auf genomischer Ebene zurück und resultieren in der Regel in der exakten Deletion eines Exons auf Transkriptebene. Sie machen mit ca. 20% einen erheblichen Anteil aller Mutationstypen aus sowohl in Phänotyp als auch in Gen-getriebenen Ansätzen (Barbaric et al., 2007; Chen et al., 2000; Greber et al., 2005; Justice et al., 1999). Obwohl ENU in der Regel Punktmutationen induziert, wurden in wenigen Fällen auch kleine Deletionen beobachtet (Barbaric et al., 2007; Shibuya et al., 1993).

## 1.5 Der Gen-getriebene Ansatz zur Mutationsdetektion in ENU mutierten ES-Zellen der Maus

Gen-getriebene Ansätze zur Detektion chemisch induzierter Mutationen wurden bereits an einer Vielzahl klassischer Modellorganismen wie *Arabidopsis thaliana*, *Ceanorhabditis elegans*, *Drosophila melanogaster*, dem Zebrafisch, der Ratte und der Maus erfolgreich durchgeführt (Greber et al., 2005). In der Maus werden bei der Mutagenese mit ENU zwei verschiedene Ansätze unterschieden (Abbildung 5): Zum einen werden Gen-getriebene Ansätze mit *in vivo* mutagenisierten Mäusen und zum anderen mit *in vitro* mutagenisierten ES-Zellen praktiziert (Acevedo-Arozena et al., 2008a). Letztere Variante wird in dieser Arbeit verfolgt und daher genauer ausgeführt.



**Abbildung 5:** Gen-getriebene Ansätze unter Verwendung von *in vitro* bzw. *in vivo* Mutagenesen der Maus. Dargestellt sind die prinzipiellen Vorgehensweisen bei Gen-getriebenen *in vitro* bzw. *in vivo* Mutagenesen der Maus mittels ENU. (Abb. in Anlehnung an Soewarto, 2009)

Das Ziel der ENU Mutagenese von ES Zellen der Maus ist generell die Generierung einer kryokonservierten Bibliothek von mutierten Stammzellklonen, welche die verschiedensten Serien von Allelen für jedes Gen von Interesse tragen (Becker et al., 2006; Chen et al., 2000). Zu Beginn der Arbeit lagen vier Veröffentlichungen zur ENU-Mutagenese von Stammzellen vor, die alle Genotyp-getriebene Ansätze verfolgten und erfolgreich zeigten, dass ES-Zellen nach der Mutagenese ihre Keimbahngängigkeit behalten und zur Generierung chimärer Tiere verwendet werden können (Chen et al., 2000; Greber et al., 2005; Munroe et al., 2000; Vivian

et al., 2002). In einigen wenigen Fällen sind auch Phänotyp-getriebene Ansätze mit ENU mutierten ES-Zellen durchgeführt worden, da in diesen Ansätzen grundlegende zelluläre Prozesse wie z.B. Reparatur der DNA, Apoptose oder Differenzierungsprozesse im Fokus standen, sodass keine Untersuchungen an lebenden Tieren zwingend erforderlich bzw. nicht möglich waren (Glenister et al., 2000; Marschall et al., 1999; Munroe et al., 2000).

### 1.5.1 Mutationsfrequenzen

Wichtiger Parameter bei der Mutagenese von ES-Zellen ist die Mutationsfrequenz. Eine zu hohe Frequenz kann mit dem Verlust der Keimbahngängigkeit einhergehen, während eine zu geringe Mutationsfrequenz die Wahrscheinlichkeit für eine beliebige Mutation deutlich verringert (Greber et al., 2005). Zur Ermittlung der Mutationsfrequenz verwendet man einen Negativ-Selektionstest, mit dem alle Klone mit funktionalen Mutationen des Gens der Hypoxanthin-guanin-phosphoribosyltransferase (*Hprt*) isoliert werden können. Wird das Purinanalog 6-Thioguanin (6-TG) dem Zellkultur-Medium zugesetzt bewirkt HPRT die Phosphorylierung des 6-TG, das dann in die DNA eingebaut wird. Dadurch werden Probleme bei der Replikation der DNA verursacht, die letztendlich den Tod der Zelle bewirken. Da die männlichen ES-Zellen hemizygot bezüglich des *Hprt*-Lokus sind (X-chromosomal), können diese im Falle einer funktionellen Mutation des *Hprt* Lokus die Selektion überleben (Hooper et al., 1987). Die Mutationsfrequenz am *Hprt* Lokus gibt daher die Anzahl resistenter ES-Zellklone unter der Gesamtzahl selektionierter Klone an und ist ein Maß für die Stärke der durchgeführten Mutagenese. Der bei *in vivo* Mutagenesen adulter Männchen verwendete Test zur Bestimmung der Mutationsfrequenz ist der spezifische Lokus-Test (SLT) (Barbaric et al., 2007), der entsprechend die Häufigkeit bestimmter Phänotypen widerspiegelt. Für die *in vitro* ENU Mutagenese von ES-Zellen konnte gezeigt werden, dass die maximal erreichbaren *Hprt* Mutationsfrequenzen mit denen der *in vivo* Mutagenesen adulter Männchen vergleichbar sind (ca. 1:1000) (Chen et al., 2000). Greber et al. (2005) ermittelten in ihren Experimenten unter den Gesichtspunkten einer ausreichend hohen Mutationsrate und der Erhaltung der Pluripotenz der Stammzellen eine optimale Mutationsfrequenz von 1:2500.

### 1.5.2 Durchmusterungsstrategien

In den Gen-getriebenen Ansätzen ist wie in den Phänotyp-getriebenen Ansätzen eine Durchmusterungsstrategie festzulegen. Dies betrifft zum einen den Aufbau der Bibliothek und zum anderen die Technik mit der diese nach interessanten Mutationen durchsucht werden soll. Für den Aufbau der Bibliothek sind generell zwei Verfahren beschrieben (Chen et al., 2000; Greber et al., 2005). Zum einen können die Zellen nach erfolgter Mutagenese so auf

Multiwell Platten verteilt werden, dass statistisch eine limitierte Anzahl verschiedener Klone pro well zu erwarten ist. Dabei ist allerdings von Beginn an sicherzustellen, dass die darauffolgende Durchmusterungstechnik sensitiv genug ist, eine Mutation in einer derartigen Verdünnung erfolgreich detektieren zu können. Die andere Alternative ist die Vereinzelnung der Klone durch manuelles Picken der Kolonien. Dies ist in Bezug auf die Detektionssicherheit die sichere Variante und erlaubt auch nachträgliches Durchmischen der Klone (Pooling), erfordert allerdings zweifelsfrei einen deutlich höheren Arbeitsaufwand, der ein Screening je nach Größe der Bibliothek bereits als unrentabel erscheinen lassen kann. Die Wahl der Größe der Bibliothek ist wiederum abhängig von der Mutationsfrequenz und damit der erwarteten Wahrscheinlichkeit für ein bestimmtes Mutationsereignis.

Für die gezielte Durchmusterung (Screening) der Bibliotheken (parallel angelegte DNA bzw. RNA Archive) nach Punktmutationen in interessanten Genen wurden bereits einige Hochdurchsatz-Technologien erfolgreich etabliert. Diese basieren in der Regel auf der Detektion von Punktmutationen mittels Heteroduplices. Dazu wird zu Beginn eine Amplifikation des interessanten Bereichs mittels PCR und eine anschließende Vermischung von Wildtyp- und putativen Mutanten PCR Produkten vorgenommen. Das PCR-Produkt Gemisch wird hitzedenaturiert und anschließend durch Kühlung renaturiert. Dadurch bilden sich neben Homoduplices auch Hetroduplices, die mittels verschiedener Verfahren einfach detektierbar sind (Colbert et al., 2001). Hierzu zählen insbesondere die physikalische Trennung von Hetero- und Homoduplices über die Temperatur Gradienten Elektrophorese (TGCE) (Li et al., 2002) oder die denaturierende HPLC (DHPLC, denat. high performance liquid chromatography) (Xiao et al., 2001). Ein weiteres Verfahren beruht auf der gezielten enzymatischen Spaltung von Heteroduplices mittels der Endonuklease CEL1 aus Sellerie und der anschließenden Gelelektrophoretischen Auftrennung der PCR-Produkte (Colbert et al., 2001). Die effektivste aber auch in Zeiten der Sequenzierungen der zweiten Generation noch nicht effizienteste und mit Abstand noch kostenintensivste Variante ist die direkte Sequenzierung interessanter Bereiche im Hochdurchsatzverfahren. Aufgrund der rapiden Entwicklung der Sequenzierverfahren der zweiten Generation ist allerdings davon auszugehen, dass es in naher Zukunft sowohl ökonomisch als technisch möglich sein wird, ganze Genome und damit Zellarchive kostengünstig parallel zu sequenzieren (Acevedo-Aroza et al., 2008a).

Für die gezielte Durchmusterung von RNA (cDNA) Bibliotheken nach interessanten Exon-Deletionen wurde von Greber et al. (2005) ein PCR Verfahren publiziert (Nested Exon skipping PCR), welches mittels sogenannter Exon-Skipping Primer in der Lage war, aus

hochgradig gepoolten cDNA Proben ENU mutierter ES-Zell Klone (100 Klone pro PCR-Reaktion) Exon-Deletionen erfolgreich zu detektieren. Die Exon-Skipping Primer waren dabei so aufgebaut, dass ihre Sequenz aus den stromauf- und stromabwärts gelegenen Exons des deletierten Exons zusammengesetzt war, d.h. die Sequenz des deletierten Exons wurde übersprungen (skipped). Mit dieser Technik gelang es aus einer Bibliothek von ca. 40.000 ENU mutierten ES-Zellklonen der Maus Exon-Deletionsmutanten der Exons 6 und 18 des Onkogens *Kit* zu isolieren und eine entsprechende Mausmutante mittels Blastocysten-Injektion zu generieren.

## **1.6 Ziele dieser Arbeit**

Ziel der Arbeit ist die Entwicklung neuartiger Screeningverfahren zur Mutationsdetektion in ENU mutierten embryonalen Stammzellen der Maus. Analog zu dem von Greber et al. (2005) publizierten Verfahren steht dabei die Detektion von Exon-Deletions Mutationen im Mittelpunkt, da das zu entwickelnde Verfahren bei Erfolg für den Aufbau eines öffentlichen Exon-Deletions Klon-Archivs genutzt werden soll. Die geplante Durchmusterungsstrategie weicht von den bisher beschriebenen Verfahren ab, weil sie ein paralleles Screening nach möglichst vielen putativen Exon-Deletionen in einer Vielzahl von Genen in einem Ansatz verfolgt. Mit Hinblick auf die Arbeitsintensität und die Kosten eines großangelegten Screenings soll die Sensitivität des Verfahrens dennoch derart ausgeprägt sein, dass die Analyse gepoolter Proben möglich ist.

Für die Entwicklung der Screening-Verfahren sind im Vorfeld grundlegende Arbeiten und Analysen erforderlich. Hierzu zählen insbesondere die Isolierung von Positiv-Kontrollen, die Generierung einer Bibliothek ENU mutierter embryonaler Stammzellen für Test-Screenings und die Analyse des Transkriptoms der verwendeten ES Zellen. Auf Basis der oben genannten Vorgaben soll im ersten Teil der Arbeit ein Microarray basiertes Verfahren entwickelt und hinsichtlich der Machbarkeit untersucht werden. Als alternatives Detektions-Verfahren wird in einem zweiten Teil der Arbeit eines der parallel zur Durchführung dieser Arbeit neuaufkommenden Sequenzierverfahren der zweiten Generation betrachtet.

Abschließend gilt es die Umsetzbarkeit beider Ansätze in großangelegten Screenings zur Erstellung eines Exon-Deletions ES-Zell-Archivs unter technischen, wissenschaftlichen und wirtschaftlichen Gesichtspunkten zu diskutieren.

## 2 Material und Methoden

### 2.1 Material

#### 2.1.1 Allgemeine Chemikalien

Alle allgemein gebräuchliche Chemikalien wiesen wenigstens den Reinheitsgrad *p.a.* auf und wurden von folgenden Anbietern bezogen: AppliChem (Darmstadt), BioRad Laboratories (München), Calbiochem (Bad Soden), Carl Roth (Karlsruhe), GE Healthcare Europe (Freiburg), Gibco BRL (Eggenstein), Invitex (Berlin), Invitrogen (Karlsruhe), MBI Fermentas (St. Leon-Rot), VWR (Darmstadt), Metabion (Martinsried), New England BioLabs (Frankfurt), PeqLab (Erlangen), Pharmacia (Freiburg), Promega (Madison, USA), QIAGEN (Hilden), Roche Applied Science (Mannheim), Serva (Heidelberg), Sigma (Taufkirchen), Worthington (Lakewood, NJ, USA).

#### 2.1.2 Allgemeine Verbrauchsmaterialien

Allgemein gebräuchliche Verbrauchsmaterialien wurden von folgenden Firmen bezogen: Applied Biosystems (Darmstadt), Beckman Coulter (Fullerton, USA), Biometra (Göttingen), BioRad Laboratories (München), Eppendorf (Hamburg), Gibco BRL (Eggenstein), Greiner (Solingen), Heraeus (Osterode), Millipore (Eschborn), MWG Biotech (München), Pharmacia (Freiburg), Serva (Heidelberg), Schott Glaswerke (Mainz), Sorvall (Bad Homburg), Whatman (Maidstone, England), Zinsser Analytic (Frankfurt).

#### 2.1.3 Geräte und spezielle Materialien

##### Agilent Microarrays, Scanner und Zubehör

Agilent Microarray Scanner AA	Agilent (Waldbronn)
Feature Extraction Software (V. 9.3.1)	Agilent (Waldbronn)
Hybridisierungskammer SureHyb	Agilent (Waldbronn)
Hybridisierungsofen	Agilent (Waldbronn)
Custom Gene Expression Microarray 4x44k	Agilent (Waldbronn)
Custom Gene Expression Microarray 8x15k	Agilent (Waldbronn)
Hybridisation Gasket Slides 4x44k	Agilent (Waldbronn)
Hybridisation Gasket Slides 8x15k	Agilent (Waldbronn)
Atlas Glass Hybridization Chamber	Clontech (Heidelberg)

### **Geldokumentationsgerät und Zubehör**

AlphaImager HP	Alpha Innotech Corporation (San Leandro, USA)
AlphaEase FC Software (V. 4.1.0)	Alpha Innotech Corporation (San Leandro, USA)

### **Gelelektrophorese-Apparaturen und Spannungswandler**

Agarosegel-Apparaturen	Wissenschaftliche Werkstätten MPI-MG (Berlin)
BioRad PowerPac 200	BioRad Laboratories (München)

### **Illumina Microarrays, Scanner und Zubehör**

Microarray Scanner, BeadStation 500X	Illumina (San Diego, USA)
MouseRef-8 v1.1 Bead Chip	Illumina (San Diego, USA)
BeadArray Reader Software	Illumina (San Diego, USA)
BeadStudio Software (V. 1.5.1.3)	Illumina (San Diego, USA)

### **Inkubatoren**

Innova CO-170	New Brunswick Scientific (Nürtingen)
Innova 44	New Brunswick Scientific (Nürtingen)

### **Kapillarsequenzierer**

ABI 3730 DNA Analyzer	Applied Biosystems (Darmstadt)
-----------------------	--------------------------------

### **Mikroskope**

Olympus CK2	Olympus (Hamburg)
Zeiss LSM 510 meta	Zeiss (Jena)

### **Photometer**

Ultraspec 2000	Pharmacia Biotech (Uppsala, Schweden)
Nanodrop ND 1000	Nanodrop Technologies (Wilmington, USA)

### **Solexa/Illumina Sequenzierer**

Illumina 1G Sequencing Plattform	Illumina (San Diego, USA)
Illumina analysis pipeline (V. 2.2.3)	Illumina (San Diego, USA)
Solexa SSB Kit	Illumina (San Diego, USA)

### **Sterilbank**

Cleanbench Herasafe Heraeus (Osterode)

### **Thermocycler und Thermoblöcke**

Peltier Thermal Cycler PTC-225 MJ Research (Waltham, USA)

Thermomixer 5436 Eppendorf (Hamburg)

### **Ultraschall-Bad**

UTR200 Hielscher Ultrasonics GmbH (Teltow)

### **Zählkammer**

Neubauer-Zählkammer Carl Roth GmbH (Karlsruhe)

### **Zentrifugen und Rotoren**

Centrifuge 5415-D Eppendorf (Hamburg)

Centrifuge 5810-R Eppendorf (Hamburg)

SpeedVac SC110 Savant Instruments (NY, USA)

#### **2.1.4 Reaktionskits**

BigDye Terminator v1.1 Cycle Sequencing Kit Applied Biosystems (Darmstadt)

Gene Expression Hybridization Kit Agilent (Waldbronn)

Gene Expression Wash Buffer Kit Agilent (Waldbronn)

Illumina Hybridisierungs- und Waschreagenzien Illumina (San Diego, USA)

Illumina TotalPrep RNA Amplification Kit Ambion (Foster City, USA)

Invisorb Spin DNA Extraction Kit Invitex (Berlin)

Low RNA Input Linear Amplification Kit PLUS, One-color Agilent (Waldbronn)

NucleoSpin Extract II Macherey und Nagel (Düren)

OligoTex mRNA mini Kit Qiagen (Hilden)

SuperScript double stranded cDNA synthesis Kit Invitrogen (Karlsruhe)

#### **2.1.5 Spezielle Chemikalien und Lösungen**

Cy3-Streptavidin Amersham Bioscience (Piscataway, USA)

DMEM mit (L-Glutamin) Pan-Biotech (Nürnberg)

Ethidiumbromid Sigma-Aldrich (Taufkirchen)

Fötales Rinderserum (FCS) Sigma-Aldrich (Taufkirchen)

Leukaemia inhibitory factor (LIF)	Francis Steward
Mitomycin C	Roche Applied Science (Mannheim)
N-Ethyl-N-nitrosoharnstoff (ENU)	Sigma-Aldrich (Taufkirchen)
Nichtessentielle Aminosäuren	Pan-Biotech (Nürnberg)
Penicillin/ Streptomycin Mix (100 x)	Pan-Biotech (Nürnberg)
TRIZol Reagenz	Invitrogen (Karlsruhe)
$\beta$ -Mercaptoethanol (500 x)	Pan-Biotech (Nürnberg)
PBS ohne Ca und Mg (1 x)	Pan-Biotech (Nürnberg)
Trypsin-EDTA (1 x)	Invitrogen (Karlsruhe)

### 2.1.6 Enzyme

DNase I	Worthington Biochemical Corporation (Lakewood, USA)
DNase I, RNase frei	Roche Applied Science (Mannheim)
<i>EcoRV</i>	New England Biolabs (Frankfurt)
GELase	Epicentre Biotechnologies (Madison, USA)
M-MLV Reverse Transcriptase	USB Corporation (Cleveland, USA)
Proteinase K	New England Biolabs (Frankfurt)
RNase Cocktail	Ambion (Foster City, USA)
RNase H	Ambion (Foster City, USA)
<i>Taq</i> -Polymerase	MPI-MG (Berlin)

### 2.1.7 Häufig verwendete Medien, Puffer und Lösungen

DNA-Extraktions-Puffer	Tris-HCl, pH 8,5	100 mM
	NaCl	100 mM
	EDTA, pH 8,0	10 mM
	Sodium-Dodecyl-Sulfat (SDS)	0,2% (w/v)
Ethyldiamintetraacetat (EDTA)	EDTA	0,5 M
	NaOH	auf pH 8,0
Ladepuffer-Rot	Kresolrot	1 mM
	Saccharose	60% (w/v)

PCR-Puffer (10x)	Tris-HCl, pH 8,5	500 mM
	(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	200 mM
	MgCl <sub>2</sub>	5 mM
	Tween 20	0,1% (v/v)
Tris-Acetat-EDTA-Puffer (TAE) (50x)	Tris	2 M
	Eisessig	5% (v/v)
	EDTA, pH 8,0	50 mM
Tris-EDTA-Puffer (TE)	Tris-HCl, pH 8,0	10 mM
	EDTA, pH 8	1 mM
DMEM ES-Medium	DMEM	
	FCS (hitzeinaktiviert, 50°C 45 min)	15% (v/v)
	Nichtessentielle Aminosäuren	1 x
	β-Mercaptoethanol	1 x
	Penicillin/ Streptomycin Mix	1 x
	LIF	1000 U/ml
DMEM Standardmedium	DMEM	
	FCS (hitzeinaktiviert, 50°C 45 min)	15% (v/v)
	Penicillin/ Streptomycin Mix	1 x
Einfriermedium	DMEM Standardmedium	
	DMSO	10% (v/v)
SSC (20x)	NaCl	3 M
	Na-Citrat, pH 7,0	0,3 M
Strip2-Puffer	Tris-HCl, pH 8,0	10 mM
	EDTA, pH 8	0,1 mM
Strip3-Puffer	Tris-HCl, pH 7,5	20 mM
	MgCl <sub>2</sub>	5 mM
	DTT	2 mM
	BSA	0,06 µg/ml
	RNaseH	5 U / ml

**Strip4-Puffer**

Tris-HCl, pH 8,0	1,5 mM
Triton-X-100	0,25‰ (v/v)
Tween 20	0,25% (v/v)
GuHCl	50 mM
Proteinase K	2 mg/l

**2.1.8 Maus-Zelllinien**

In dieser Arbeit wurden embryonale Stammzellen und embryonale Fibroblasten der Maus verwendet. Die embryonalen Stammzellen entstammten einem männlichen F1-Hybrid der beiden Inzuchtlinien C57BL/6J und 129Sv (Ausgangspassage 11 wurde freundlicherweise von Dr. Herrmann Bauer, MPI-MG zur Verfügung gestellt). Diese Zelllinie zeichnet sich durch eine höhere Keimbahngängigkeit im Vergleich zu den reinen Inzuchtlinien aus. Neben dieser F1-Zelllinie wurde auch die Linie CGR8 mit dem genetischen Hintergrund 129Sv für Testmutagenesen verwendet. Embryonale Fibroblasten wurden aus Embryonen des Auszuchtstamms CD1 präpariert, da dieser Stamm bekanntermaßen eine höhere Wurfgröße aufweist.

**2.1.9 Maus Gehirn-Gewebe**

Als unabhängige Referenzprobe für die Sequenzierung des Transkriptoms wurde mRNA aus dem Hirn einer Maus verwendet. Das Hirn wurde einer 12 Wochen alten männlichen C57BL/6J Maus entnommen, 2 x mit PBS gewaschen und bei -80°C aufbewahrt (Zur Verfügung gestellt von Dr. Ulrike Dohrmann).

**2.2 Zellbiologische Methoden**

Alle Zellkulturarbeiten wurden unter sterilen Bedingungen unter Verwendung steril verpackter Materialien und Lösungen durchgeführt. Grundlage für die angewandten Methoden bildete das Buch 'Manipulating the Mouse Embryo: A Laboratory Manual (Nagy et al., 2003). Soweit nicht anders angegeben wurden Zellkulturen bei 37°C, 5% CO<sub>2</sub> und 95% Luftfeuchtigkeit inkubiert. Medien und Lösungen wurden vor der Verwendung auf 37°C temperiert.

**2.2.1 Ernten von Zellen**

Zusammenhängende und an den Kulturschalen anhaftende Zellverbände wurden durch 'Trypsinieren' in Einzelzellsuspensionen aufgelöst: Das Kulturmedium wurde abgesaugt und die Kulturschale zweimal mit PBS gewaschen. Anschließend erfolgte so lange eine

Inkubation der Zellen mit einer Menge Trypsin-EDTA (2.1.5), die ausreichte den Boden der Schale zu benetzen, bis sich die Zellen von der Oberfläche lösten. Gestoppt wurde die Reaktion durch Zugabe von ca. 5 Volumen MEF Medium (2.1.7). Die Zellen wurden abschließend für 5 Minuten bei 250 x g pelletiert. Das Zellpellet wurde dann je nach weiterer Vorgehensweise (Expansion, Einfrieren) in dem entsprechenden Medium resuspendiert bzw. bis zur Extraktion von Nukleinsäuren auf  $-80^{\circ}\text{C}$  gelagert.

### **2.2.2 Zählen von Zellen und Zellkolonien**

Zellen wurden mit Hilfe einer Neubauer-Zählkammer (2.1.3) gezählt. ES-Zellen und MEFs waren dabei aufgrund der deutlich runderen Form und geringeren Größe der ES Zellen eindeutig zu unterscheiden. Stammzellen wurden routinemäßig während der Expansion und vor der Konservierung gezählt.

Stammzellkolonien auf '*feeder layern*' (2.2.4) wurden direkt auf der Kulturschale unter dem Mikroskop bei 40-facher Vergrößerung ausgezählt. Eine unter der Schale angebrachte Folie erleichterte dabei die Arbeit, indem sie es ermöglichte, über aufgedruckte Kreisausschnitte (1/2 bis 1/36) Teilbereiche einer Platte auszuzählen.

### **2.2.3 Einfrieren und Auftauen von Mauszellen**

Pelletierte Zellen wurden zur Konservierung in eiskaltem Einfriermedium resuspendiert und in einer vorgekühlten Styroporbox in einem  $-80^{\circ}\text{C}$  Gefrierschrank langsam heruntergekühlt, um einer Bildung von Eiskristallen und damit der Zerstörung von Zellen vorzubeugen. Der maximale Zelltitel betrug dabei  $5 \times 10^6$  Zellen pro ml. Nach 24 - 48h wurden die Zellen zur längerfristigen Lagerung in flüssigen Stickstoff überführt.

Das Auftauen der Zellen erfolgte durch Inkubation des Röhrchens im Wasserbad bei  $37^{\circ}\text{C}$ . Die aufgetaute Zellsuspension wurde schnellstmöglich mit 5-10 Volumen warmem MEF bzw. ES-MEF (2.1.7) Medium verdünnt und die Zellen anschließend pelletiert. Die Kontaktzeit der Zellen mit dem toxischen DMSO sollte so möglichst kurz gehalten werden. Das Zellpellet wurde dann in einer entsprechenden Menge Medium resuspendiert und auf das gewünschte Kulturschalenformat ausplattiert.

### **2.2.4 Embryonale Fibroblasten der Maus**

Embryonale Stammzellen der Maus wurden in dieser Arbeit auf einer Schicht mitotisch inaktiverter embryonaler Fibroblasten (MEF) kultiviert, da diese der spontanen Differenzierung der Stammzellen entgegenwirken. Diese Zellschicht wird auch als '*feeder layer*' bezeichnet.

#### **2.2.4.1 Isolation und Expansion embryonaler Fibroblasten**

Anders als ES-Zellen lassen sich embryonale Fibroblasten nicht unbegrenzt kultivieren, was eine Isolation frischer MEFs in bestimmten Zeitabständen erfordert. Schwangere Mäuse des Stamms CD1 (2.1.8) wurden am 14. Tag p. c. durch Genickbruch getötet und anschließend der Uterus entnommen. Nachdem Fruchtblase und Plazenta entfernt waren, wurden Kopf und dunkelrot gefärbte Organe des Embryos entfernt, da hier nur wenige Fibroblasten vorkommen. Das restliche Gewebe wurde im Anschluss mit einem Skallpell in wenig PBS sehr fein zerkleinert bis es sich pipettieren ließ. Die durch die Freisetzung von DNA bedingte Viskosität wurde durch Zugabe von 100 Units DNase I pro Embryo reduziert. Der Gewebeverband wurde dann durch Zugabe von 1 ml Trypsin EDTA pro Embryo und anschließender 15 minütiger Inkubation bei 37°C und leichtem Schwenken in einem 50 ml Falcon Röhrchen enzymatisch aufgeschlossen. Die Reaktion wurde durch Zugabe von 3 bis 5 Volumen warmem MEF Medium (2.1.7) gestoppt. Nach einer Inkubationsphase von ca. 5 Minuten hatten sich größere unverdaute Gewebepartikel am Boden des Reaktionsgefäßes abgesetzt. Der die Einzelzellen enthaltende Überstand wurde in ein frisches 50 ml Falcon überführt. Die Zellen wurden für 5 Minuten bei ca 10 x g pelletiert, der Überstand abgenommen, das Pellet in warmem MEF Medium resuspendiert und erneut pelletiert. Das Zellpellet wurde in so viel MEF Medium resuspendiert, dass drei Embryo-Äquivalente auf eine 15cm Kulturschale (ca. 25 ml) ausplattiert werden konnten. Nach ca. 16 Stunden wurde das Medium gewechselt. Bei den der Kulturschale anheftenden Zellen handelte es sich größtenteils um Fibroblasten, die als Passage 0 bezeichnet wurden.

Zur weiteren Expansion wurden konfluent bewachsene Kulturschalen trypsiniert, das resultierende Pellet im Verhältnis 1/3 zum vorherigen Zelltiter in frischem MEF Medium resuspendiert, erneut ausplattiert und die Fibroblasten bis zur Konfluenz kultiviert (Passage 1). Durch zwei weitere Expansionszyklen (Trypsinierung, 1/3 Verdünnung, Kultur bis zur Konfluenz) konnte die Ausgangsmenge (Passage 0) so auf das 27-fache (Passage 3) angehoben werden. Eine weitere Expansion wurde nicht vorgenommen, da höhere Passagen deutliche Einbußen der Wachstumsgeschwindigkeit zeigten und nicht mehr zu Generierung von '*feeder layern*' geeignet erschienen. Fibroblasten der Passage 3 wurden zur Verwendung als '*feeder layer*' mitotisch inaktiviert.

#### **2.2.4.2 Inaktivierung embryonaler Fibroblasten**

Die mitotische Inaktivierung der MEF erfolgte durch eine 2 stündige Behandlung der adherenten Zellen mit 10µg/ml des Spindelgiftes Mitomycin C (2.1.5) (Martin et al., 1975) in

MEF Medium. Die Fibroblasten wurden anschließend zwei Mal mit PBS gewaschen und trypsiniert. So gewonnene inaktivierte MEF wurden zur direkten Verwendung erneut ausplattiert oder zur späteren Verwendung eingefroren (2.2.3). Embryonale Stammzellen wurden allerdings erst 24 Stunden nach Ausplattieren der MEFs auf diesen kultiviert. Mit Hinblick auf die Qualität der Stammzellkultur sollte der Zelltitert zur Generierung von '*feeder layer*' so eingestellt werden, dass Fibroblasten die Kulturschale annähernd vollständig abdecken sich aber nicht gegenseitig überwachsen.

### **2.2.5 Kultur embryonaler Stammzellen der Maus**

Embryonale Stammzellen der hier verwendeten F1 ES Zelllinie (2.1.8) wurden auf '*feeder layern*' aus teilungsinaktivierten CD1 MEFs (2.2.4) kultiviert. Im Allgemeinen wurden 500 bis 2000 ES-Zellen pro cm<sup>2</sup> Kulturschalenoberfläche ausplattiert, die zu Koloniedichten von 50 bis 500 Kolonien pro cm<sup>2</sup> führten. Die Plattierungseffizienz war hauptsächlich abhängig von dem Zustand der Zellen, d.h. ob diese frisch aufgetaut oder bereits länger in Kultur gehalten wurden. Das Medium wurde in der Regel alle 24 bis 48 Stunden gewechselt. Soweit nicht anders angegeben wurden ES-Zellen maximal 4 Tage auf der gleichen Kulturschale kultiviert, um einer spontanen Differenzierung insbesondere großer ES-Zellkolonien entgegenzuwirken. Beim Umplattieren (Passagieren) und Einfrieren erfolgte keine Trennung der ES-Zellen von den Fibroblastenzellen, da deren Anteil an einer konfluent bewachsenen Kulturschale nur maximal 1% betrug. Es wurde viel Wert auf eine sehr gute Vereinzelung der Zellen während der Passagen gelegt, da ES-Zellen in ausplattierten Zellaggregaten dazu neigen spontan zu differenzieren. Die Anzahl der Passagen wurde fortlaufend vermerkt. Es wurde versucht die Passagenanzahl möglichst gering zu halten, da mit der Anzahl der Passagen die Gefahr einer karyotypischen Veränderung zunimmt (Eggan et al., 2002).

#### **2.2.5.1 Zellkultur in Multititerplatten**

Bei der Zellkultur in 96er Multititerplatten wurde grundsätzlich auf Zentrifugationsschritte verzichtet. Die Zellen waren somit nach der Passage in längerem Kontakt mit geringen Mengen Trypsin-EDTA. Daher wurde hier ein Mediumwechsel schon nach 6-12 Stunden durchgeführt.

Einzelne Positionen einer eingefrorenen Platte ließen sich durch die Zugabe von warmem Medium leicht herauspipettieren, während die anderen Positionen bei zügiger Vorgehensweise gefroren blieben.

### 2.2.5.2 Subklonieren embryonaler Stammzellkolonien

Die klonale Vervielfältigung einer einzelnen Stammzellkolonie erforderte zunächst eine geringe Koloniedichte ( $\text{max. } 10/\text{cm}^2$ ) auf der Kulturschale, um eine saubere Trennung einzelner Kolonien gewährleisten zu können. Weiterhin war eine gewisse Koloniegröße nötig, die in der Regel nach 6-8 Tagen Wachstum auf einer Schale erreicht wurde. Die Kulturschale wurde mit PBS gewaschen und so viel PBS in der Schale gelassen, dass es ausreichte den Boden zu benetzen. Unter 40-facher Vergrößerung wurde die gewünschte Kolonie mit einer feinen Pipettenspitze leicht von dem ‚feeder-layer‘ gelöst, ruckartig mit einem Volumen von 10  $\mu\text{l}$  PBS in die Spitze gesaugt und in 25  $\mu\text{l}$  warmes Trypsin-EDTA überführt. Nach wenigen Minuten ließen sich die Zellen der Kolonie vereinzeln, sodass die Reaktion durch Zugabe von ca. 220  $\mu\text{l}$  ES-DMEM gestoppt werden konnte. Ein Volumen von 250  $\mu\text{l}$  wurde dann zur weiteren Expansion der Stammzellen in eine mit ‚feeder layern‘ beschichtete Position einer 96er Multititerplatte pipettiert.

### 2.2.6 Karyotypisierung embryonaler Stammzellen

Stammzellen können im Laufe der Kultur karyotypische Abberationen aufweisen. Da in dieser Arbeit eine Bibliothek chemisch mutierter ES-Zellen angelegt werden sollte, wurden Zellen der für die Mutagenese verwendeten Passage (#14) im Vorfeld auf eine ordnungsgemäße Anzahl von Chromosomen untersucht.

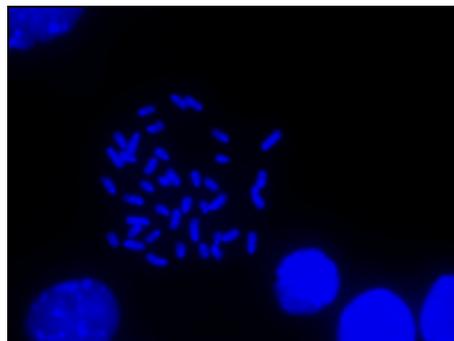


Abbildung 6: Fluoreszenzmikroskopische Aufnahme eines Karyogramms einer Maus ES-Zelle (F1)

Die Lyse der Zellen in einer hypotonischen KCl-Lösung und die Fixierung der Chromosomen wurde nach dem in dem Buch 'Manipulating the Mouse Embryo: A Laboratory Manual.' veröffentlichten Protokoll durchgeführt (Nagy et al., 2003). Die Chromosomen wurden mit einer 0,0001% (w/v) DAPI Lösung auf dem Objektträger angefärbt und durch Anregung des Farbstoffes bei 364 nm mittels eines Fluoreszenzmikroskopes sichtbar gemacht. Es wurden fünf Karyogramme erstellt, von denen alle einen normalen Karyotypen aufwiesen (Abbildung 6).

### 2.2.7 Chemische Mutagenese embryonaler Stammzellen

Zur chemischen Mutagenese embryonaler Stammzellen wurde das Mutagen Ethylnitrosoharnstoff (ENU) genutzt. ES-Zellen wurden dazu auf '*feeder layern*' bis zur Subkonfluenz kultiviert (2.2.5), trypsiniert und so in DMEM ES-Medium (2.1.7) resuspendiert, dass der Zelltitel zwischen 0,5 und 1 Mio Zellen pro ml Medium betrug (2.2.2). Im Allgemeinen wurden Mutagenesen mit mehr als  $3 \times 10^7$  Zellen durchgeführt, da man bei einer erwarteten Überlebensrate von ca. 5% (Greber et al., 2005) und einer minimalen Plattierungseffizienz von 10% Verluste von 99,5% aufzuweisen hat und ein Großteil der überlebenden Zellen zur Bestimmung der *Hprt*-Mutationsfrequenz (2.2.8) nötig ist.

Zur Mutagenese wurde frisch angesetztes ENU (Ausgangslösung: 40 mM ENU in PBS) bis zur gewünschten Endkonzentration (i.d.R. 2 mM) der Zellsuspension zugesetzt. Die Suspension wurde dann 90 Minuten bei 37°C leicht geschwenkt um eine gleichmäßige Verteilung des Mutagens zu gewährleisten. Anschließend wurden die Zellen abzentrifugiert, zweimal mit PBS gewaschen und in ES-Medium aufgenommen. Ein Teil der mutagenisierten ES-Zellen wurde sofort eingefroren (2.2.3), während ein anderer Teil zur Bestimmung der Überlebensrate und der Mutationsfrequenz genutzt wurde.

Zur Ermittlung der von der Plattierungseffizienz bereinigten Überlebensrate wurde ein Aliquot der Stammzellen vor der ENU-Zugabe entnommen und parallel bearbeitet. Nach erfolgter 90 minütiger Inkubation wurden dann gleiche Mengen ENU und nicht ENU behandelte Zellen auf Kulturschalen ausgesät. Nach drei Tagen Kultur wurden die Kolonien auf den Platten ausgezählt und verglichen. Das Verhältnis angewachsener unbehandelter Stammzellkolonien zu der Anzahl ausplattierter Zellen entspricht der Plattierungseffizienz, während das Verhältnis mutagenisierter Kolonien zu der Anzahl unbehandelter Kolonien die Überlebensrate widerspiegelt.

### 2.2.8 Mutationsfrequenzen am *Hprt* Locus

Der Erfolg und die Stärke einer ENU-Mutagenese wurde durch den *Hprt*-Selektionstest überprüft (Hooper et al., 1987). Dieser Test erlaubt es basierend auf einer negativen Selektion mit dem Purin-Analog 6-Thioguanin (6-TG) auf solche ES-Zellklone zu selektieren, die eine funktionelle Mutation des *Hprt*-Gens aufweisen. Der Anteil solcher Mutanten in einer Population chemisch mutagenisierter ES-Zellen entspricht der *Hprt*-Mutationsfrequenz und erlaubt somit eine Beurteilung des Mutationsereignisses.

Nach Ermittlung der Überlebensrate (2.2.7) wurde eine große Anzahl ( $1 \times 10^5$ ) überlebender ENU-mutierter ES-Zellen auf eine 15 cm Schale ohne '*feeder layer*' ausplattiert und kultiviert.

Vor Beginn der Selektion wurden die Stammzellkolonien zweimal passagiert und die Zellzahl dabei um eine 10er Potenz reduziert. Diese verlängerte Inkubationsphase (ca. 8-10 Tage) war nötig, damit noch vorhandenes HPRT-Enzym in den potentiellen Mutanten degradiert bzw. ausgedünnt werden konnte. Die Reduktion der Zellzahl mit jeder Passage sollte einer Anhäufung redundanter ES-Zellklone, bedingt durch das natürliche Wachstum der Kolonien, entgegenwirken. Mit der zweiten Passage wurden die Zellen auf zwei Kulturschalen aufgeteilt. Davon diente eine der Ermittlung der Plattierungseffizienz, d.h. es erfolgte keine Zugabe von 6-TG und die angewachsenen Kolonien wurden nach 3 Tagen Kultur ausgezählt, während die andere zur Selektion von *Hprt*-Funktionsmutanten genutzt wurde. Die Selektion startete einen Tag nach der zweiten Passage durch Zugabe von 10  $\mu$ M 6-TG in das Kulturmedium. Nach ca. 8 bis 10 Tagen Kultur wurden die 6-TG resistenten Kolonien ausgezählt und konnten je nach Bedarf subkloniert werden. Die Mutationsfrequenz entsprach dann der Anzahl 6-TG resistenter Kolonien dividiert durch die Anzahl angewachsener Kolonien auf der unbehandelten Kulturschale. Die Rate an spontan auftretenden *Hprt*-Funktionsmutaten wurde durch Mitführen einer Nullkontrolle, d.h. der Selektion auf 6-TG resistenten Kolonien innerhalb einer nicht-mutagenisierten ES-Zellpopulation, kontrolliert.

## 2.3 Molekularbiologische Methoden

### 2.3.1 Agarose-Gelelektrophorese

Die Agarose-Gelelektrophorese diente in dieser Arbeit der Analyse von Nukleinsäureproben. Die Moleküle werden dabei gemäß ihrer Wanderungsgeschwindigkeit im elektrischen Feld, welches an eine Agarose-Gelmatrix angelegt ist, aufgetrennt. Die Gele wurden nach Sambrook und Russell (Sambrook and Russell, 2001) hergestellt. Die Agarosekonzentration richtete sich dabei im Wesentlichen nach der Größe und Typ der zu analysierenden Moleküle (Tabelle 1). Die Elektrophorese fand in mit 1 x TAE (2.1.7) gefüllten Elektrophoresekammern bei 5 bis 15 V/cm für 30 - 90 min statt. Zur späteren Visualisierung der Nukleinsäuren enthielt das Gel Ethidiumbromid (EtBr; 2.1.5) welches in doppelsträngige Bereiche der Nukleinsäuremoleküle interkaliert und bei Anregung mit UV-Licht der Wellenlänge  $\lambda = 302$  nm fluoresziert. Die Dokumentation erfolgte über ein Geldokumentationssystem (2.1.3).

**Tabelle 1: Verwendete Agarosekonzentrationen in Agarosegelen**

<b>Molekül-Typ</b>	<b>Agarose (w/v)</b>
Gesamt-RNA	1,0%
genomische DNA	0,8%
Fragmente > 1000 bp	1,0%
Fragmente 300-1000 bp	1,5%
Fragmente < 300bp	2,0%

### 2.3.2 Spektrophotometrie

Für spektralphotometrische Untersuchungen wurde der Nanodrop ND 1000 mit der dazugehörigen Software nach Herstellerangaben verwendet. Die Konzentration von Nukleinsäureproben wurde über die Extinktion bei  $\lambda = 260$  nm ermittelt. Je nach Nukleinsäuretyp bietet die Software hier verschiedene Einstellungen, die eine direkte Umrechnung der Extinktion in eine Konzentrationsangabe ermöglicht (dsDNA: DNA50 entspricht 50  $\mu\text{g/ml}$  bei einer Extinktion von 1 gemessen mit einer konventionellen Messküvette mit einer Schichtdicke von 10mm; RNA-40  $\equiv$  40  $\mu\text{g/ml}$ ; ssDNA-33  $\equiv$  33  $\mu\text{g/ml}$ ). Das Verhältnis der Extinktionen bei Wellenlängen von 260 nm und 280 nm diente der Reinheitskontrolle der betrachteten Nukleinsäureproben. Für reine DNA-Lösungen wird ein Wert von  $A_{260}/A_{280} = 1,8$  erwartet, für reine RNA-Lösungen ein Wert von  $A_{260}/A_{280} = 2,0$ . Werte kleiner der erwarteten sind ein Anzeichen für eine Kontamination mit Proteinen oder Phenol, die eine starke Absorption nahe 280 nm aufweisen.

Das Nanodrop-Instrument wurde weiterhin dazu benutzt, die spezifische Cyanine-3-Aktivität (pmol/ $\mu\text{g}$  cRNA) in amplifizierten RNA-Proben zu bestimmen. Dieser Wert gibt Aufschluss darüber, ob bei der Amplifikation der RNA ausreichend Cy3 eingebaut wurde, um bei der nachfolgenden Hybridisierung auf Agilent-Microarrays entsprechende Signalintensitäten erreichen zu können. Die Extinktion bei  $\lambda = 550$  nm wurde von der Software zur Berechnung der vorhandenen Cy3-Konzentration genutzt. Die spezifische Aktivität wurde dann wie folgt berechnet (One-Color Microarray-Based Gene Expression Analysis Protokoll der Firma Agilent Technologies, [www.agilent.com](http://www.agilent.com)):

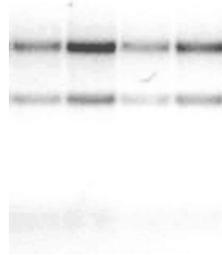
$$\text{Spezifische Aktivität Cy3 (pmol Cy3}/\mu\text{g cRNA)} = \frac{\text{Cy3-Konzentration (pmol}/\mu\text{l)}}{\text{cRNA-Konzentration (}\mu\text{g}/\mu\text{l)}}$$

Lag die Aktivität unter einem Wert von 9,0 pmol Cy3/ $\mu\text{g}$  cRNA wurde die Probe den Herstellerangaben folgend nicht für eine Hybridisierung genutzt.

### 2.3.3 Isolation von Gesamt-RNA

Analog zur Isolation genomischer DNA wurden ES-Zellen zunächst geerntet (2.2.1), gewaschen und pelletiert (1000 x g, 3 min). Maus Hirn wurde vorab unter Verwendung von flüssigem Stickstoff gemörstert und dann der Isolation unterzogen. Die Isolation von Gesamt-RNA erfolgte mit dem TRIzol-Reagenz (2.1.5) (Chomczynski et al., 1987) exakt nach dem vom Hersteller zur Verfügung gestellten Protokoll, weshalb hier auf eine detaillierte Beschreibung der Prozedur verzichtet wird. Die Aufreinigung basiert auf einer Extraktion mit einem Phenol-Chloroform-Gemischs und einer anschließenden Fällung mit 100% Isopropanol

(v/v). Zum sofortigen Schutz vor möglichen RNase-Kontaminationen enthält das Reagenz den RNase-Inhibitor Guanidinisothiocyanat. RNA wurde in dieser Arbeit durchweg in RNase freiem Wasser (Qiagen) gelöst.



**Abbildung 7: Beispielhafte Darstellung der Integritätskontrolle isolierter Gesamt-RNA**

Beispielhaft dargestellt ist die Ethidiumbromid vermittelte Fluoreszenz von Gesamt-RNA aus 4 ES-Zellen (F1) nach elektrophoretischer Auftrennung in einem 1%-igen Agarosegel. Es wurde jeweils 10% der Menge RNA aufgetragen, die aus einem Well einer 6er-Multitierplatte isoliert werden konnte (~ 0,5-3 µg RNA). Die Schärfe der beiden Banden der ribosomalen RNA und das Verhältnis (ca. 2:1) der 28S-RNA (obere Bande) zur 18S-RNA (untere Bande) deuten auf hohe Integrität der Proben hin.

Etwaige Kontaminationen der RNA-Proben mit genomischer DNA wurden durch einen anschließenden DNaseI-Verdau beseitigt. Der Verdau erfolgte in 1 x DNaseI Inkubations-Puffer (Roche) unter Zugabe von DNaseI (0.25 U/µg RNA) und anschließender Inkubation für 15 Minuten bei 37°C. Die Reaktion wurde gestoppt durch Zugabe von EDTA (pH 8,0) in einer Endkonzentration von 8 mM und einer 10 minütigen Inkubation bei 75°C. Die RNA wurde im Anschluss erneut mit Isopropanol gefällt und in Wasser gelöst. Gesamt-RNA, die für eine spätere Amplifikation mit dem 'Low RNA Input Linear Amplification Kit PLUS' der Firma Agilent gedacht war, wurde aufgrund einer Empfehlung in dem Handbuch des Kits nicht mit DNaseI behandelt.

Die Integrität der Gesamt-RNA wurde gelelektrophoretisch (2.3.1) anhand der Schärfe und der Intensitätsverhältnisse der 28S-RNA und 18S-RNA Banden überprüft. Die RNA-Konzentration wurde spektrophotometrisch (2.3.2) mit dem Nanodrop-Instrument ermittelt. RNA-Proben wurden bei -80°C gelagert.

### 2.3.4 Isolation von messenger RNA

Messenger RNA diente in dieser Arbeit als Ausgangsmaterial für die Synthese doppelsträngiger cDNA (ds-cDNA), welche nachfolgend in einem Hochdurchsatz-Sequenzierverfahren mittels Solexa-Technologie sequenziert wurde. Die Aufreinigung von mRNA aus Gesamt-RNA Proben erfolgte mit dem 'OligoTex mRNA mini Kit' (2.1.4) und basiert auf der Bindung der polyA Schwänze der mRNA Moleküle an oligodT Moleküle, die an eine Kieselgelmatrix gebunden sind. Mit Ausnahme der Elution erfolgte die mRNA

Extraktion entsprechend den Herstellerangaben. Als Ausgangsmaterial dienten 250 µg Gesamt-RNA. Um eine höhere Ausbeute zu erreichen wurde die mRNA abweichend vom Protokoll zweimal mit einem größeren Volumen von 100 µl 70°C heißem Puffer OEB (Quiagen) eluiert. Das Eluat wurde anschließend mit 0,1 Vol 3M Natrium-Acetat (pH 5,6) und 2,5 Vol 96% Ethanol (v/v) versetzt und die ausfallende mRNA durch Zentrifugation (14.000 x g , 30 min, 4°C) pelletiert. Das Pellet wurde zweimal mit 70% Ethanol gewaschen, in 9 µl RNase freiem Wasser (Qiagen) gelöst und in direktem Anschluss für die Synthese von ds-cDNA (2.3.6) genutzt.

### 2.3.5 Synthese einzelsträngiger cDNA

Die Synthese einzelsträngiger cDNA (cDNA) entspricht der reversen Transkription (RT) einer RNA-Matrize in einzelsträngige DNA ausgehend von einem an die RNA gebundenen Oligonukleotid-Primer. In dieser Arbeit wird durchweg ein oligodT<sub>18</sub>-Primer verwendet. Bei den cDNA Molekülen sollte es sich daher weitestgehend um Kopien eukaryotischer mRNA handeln, da nur diese einen Poly-A-Schwanz aufweisen.

Qualitätsgeprüfte RNA und oligodT<sub>18</sub>-Primer (Tabelle 2) wurden zunächst für 5 min auf 65°C erhitzt und anschließend sofort auf Eis abgekühlt. Diese Schritte dienten der Auflösung möglicher Sekundärstrukturen und der Verhinderung einer langsamen Rückfaltung. Im Anschluss wurden die restlichen Reaktionskomponenten (Tabelle 2) hinzugefügt und der Reaktionsansatz für 1h bei 37°C inkubiert. Eine Inaktivierung der Reversen Transkriptase erfolgte durch 10 minütige Inkubation des Ansatzes bei 70°C. Eine weitere Aufreinigung der cDNA erfolgte nicht. Sie konnte so als Matrize in der PCR (2.3.7) eingesetzt werden. Die cDNA-Proben wurden bei -20°C gelagert.

**Tabelle 2: Zusammensetzung einer Erststrang-cDNA-Synthese**

<b>Reaktionskomponente</b>	<b>Menge</b>
Gesamt-RNA	1 - 5 µg
oligodT <sub>18</sub> -Primer	20 ng
dNTPs	10 nmol
M-MLV Reaktionspuffer	1 x
M-MLV Reverse Transkriptase	100 U
H2O	ad 50 µl

### 2.3.6 Synthese doppelsträngiger cDNA

Doppelsträngige cDNA (ds-cDNA) wurde unter Verwendung des 'SuperScript double stranded cDNA synthesis Kit' (2.1.4) der Firma Invitrogen hergestellt. Dabei wurde exakt nach Herstellervorgaben gearbeitet. Als Ausgangsmaterial diente die wie unter Abschnitt

2.3.4 aufgereinigte mRNA. Prinzipiell basiert die Synthese auf folgenden Schritten: Zunächst wird ausgehend von einer mRNA-Matrize unter Verwendung eines oligodT<sub>18</sub>-Primers die Erststrang-cDNA analog zu Abschnitt 2.3.5 synthetisiert. RNaseH bewirkt in einem nächsten Teilschritt einen partiellen Abbau der RNA-Matrize, was dahingehend genutzt wird, dass diese nun als Primer für die Synthese des Zweitstrangs dienen können. Die Synthese wird von einer *E.coli* DNA-Polymerase durchgeführt, während eine Ligase die neu entstandenen DNA-Fragmente kovalent verknüpft.

Doppelsträngige cDNA wurde nach Herstellerangaben mit Phenol-Chloroform-Isoamylalkohol-Extraktion aufgereinigt, mittels Ethanol und Ammonium-Acetat präzipitiert und in TE gelöst. Die Lagerung erfolgte bei -20°C. Die DNA-Konzentration wurde spektrophotometrisch ermittelt und die Größenverteilung der cDNA-Fragmente auf einem Agarose-Gel überprüft.

### **2.3.7 Polymerase-Kettenreaktion (PCR)**

Die Polymerase-Kettenreaktion ist ein Verfahren welches *in vitro* die Vervielfältigung eines spezifischen DNA-Fragments bekannter Sequenz erlaubt (Mullis et al., 1986; Saiki et al., 1988). Prinzipiell läuft die Amplifikation in drei Schritten ab: Durch Hitze wird die DNA-Matrize in ihre Einzelstränge denaturiert. In einem zweiten Reaktionsschritt wird bei niedrigerer Temperatur eine Anlagerung kurzer synthetisch hergestellter Oligonukleotid-Primer an komplementäre Sequenzbereiche der Matrizen-DNA ermöglicht. Ausgehend von den 3'-OH Enden der Primer erfolgt dann die Synthese (Elongation) des jeweiligen komplementären DNA-Stranges mittels einer thermostabilen Polymerase. Die Wiederholung dieser Reaktionsschritte führt daher im Idealfall zu einer exponentiellen Vermehrung des durch die Primer begrenzten DNA-Fragmentes (Vermehrung der Anzahl der Ausgangsfragmente um den Faktor  $2^n$ ,  $n$  = Anzahl der Zyklen). Die Zusammensetzung eines Standard-PCR-Ansatzes ist in Tabelle 3 wiedergeben. Tabelle 4 gibt Auskunft über das in dieser Arbeit verwendete Standard-PCR-Thermoprofil. Soweit nicht anders angegeben erfolgte die Auswahl geeigneter Primersequenzen unter Zuhilfenahme der Primer3 Software (Rozen et al., 2000). Die Bezeichnung RT-PCR wird in dieser Arbeit als Synonym dafür benutzt, dass cDNA als Matrize in der PCR verwendet wurde. Der Erfolg der PCR wurde generell auf Agarosegelen überprüft.

**Tabelle 3: Zusammensetzung eines Standard-PCR-Ansatzes**

Die Primerkonzentrationen und die eingesetzte Menge DNA-Matrize variierten je nach Primerpaar und Beschaffenheit der Matrize (Qualität und Typ des Moleküls [cDNA, gDNA, PCR-Produkt]). Die verwendete Konzentration wurde jeweils in Vorversuchen ermittelt. I.d.R. wurden Primer mit Längen von 18 bis 22 bp verwendet.

<b>Komponente</b>	<b>Endkonzentration</b>
PCR-Puffer	1 x
dNTPs	400 µM
Primer 1	100 - 500 nM
Primer 2	100 - 500 nM
<i>Taq</i> -Polymerase	0,02 U/µl
DNA-Matrize	0,1 pg/µl - 10 ng/µl
Wasser	ad 50 µl

**Tabelle 4: Thermoprofil einer Standard-PCR**

Die für die Primer-Anlagerung gewählte Temperatur richtete sich nach den Beschaffenheiten der Primer ( $T_m$ -Wert) und wurde in Vorversuchen ermittelt. Je nach Länge der zu amplifizierenden Sequenz wurde die Elongationszeit variiert.

<b>Zyklen</b>	<b>Reaktionsschritt</b>	<b>Temperatur</b>	<b>Zeit</b>
1 x	Initialdenaturierung	95°C	5 min
34 x	Denaturierung	95°C	20 s
	Primer-Anlagerung	58-68°C	20 s
	Elongation	72°C	30 s /500 bp
1 x	Elongation	72°C	5 min

## 2.3.8 Microarray-Analysen

### 2.3.8.1 Illumina Genexpressions Microarrays

Die kommerziell von der Firma Illumina vertriebenen Maus-Genexpressions Microarrays wurden zur Analyse des Transkriptoms von Wildtyp und ENU-mutagenisierten embryonalen Stammzellen der Maus genutzt. Es wurde der MouseRef-8 v1.1 Bead Chip Array (2.1.3) genutzt, dessen Oligonukleotid-Zusammenstellung Genexpressionsanalysen der bis dato in die RefSeq-mRNA Datenbank (Pruitt et al., 2007) enthaltenen Transkripte erlaubte. Der komplette Microarray bestand aus jeweils acht Feldern mit identischer Oligo-Zusammenstellung, was eine parallele Analyse von acht Proben möglich machte.

RNA Isolation, DNase-Verdau und Qualitätskontrolle der Gesamt-RNA wurden wie bereits beschrieben durchgeführt (2.3.3). Für die Hybridisierung war im Vorfeld die Amplifikation der in der Gesamt-RNA enthaltenen mRNA und der Einbau Biotin markierter UTP Nukleotide (biotin-16-UTP) in diese copy-RNA (cRNA) nötig. Die lineare Amplifikation

basierte auf der von Eberwine (Eberwine 1996) entwickelten Strategie und wurde mit dem 'Illumina TotalPrep RNA Amplification Kit' (2.1.4) nach Herstellerangaben durchgeführt. Es wurden je 700 ng cRNA auf ein Feld hybridisiert. Die Hybridisierung, die Waschschritte, die Färbung durch Bindung von Cy3-Streptavidin an die Biotin markierten UTP und das Auslesen der Lichtsignale erfolgte exakt nach dem Handbuch des Herstellers (Illumina BeadStation 500x System Manual\_RevC, [www.illumina.com](http://www.illumina.com)). Es wurde der Scanner BeadStation 500x zusammen mit der BeadArray Reader Software verwendet (2.1.3). Zur Prozessierung und Analyse der Microarray-Daten diente die BeadStudio Software (Version 1.5.1.3, 2.1.3). Netto-Signalintensitäten wurden mit dieser Software zunächst durch Subtraktion des lokalen Hintergrunds eines jeden Oligo-Signals berechnet, die Signalintensitäten von Oligo-Replikaten (15-77 Positionen/RefSeq) gemittelt und die Daten gleichzeitig unter Verwendung der 'rank-invariant' Option (Kuhn et al., 2004) normalisiert. In dieser Arbeit wurde ein RefSeq-Gen (Pruitt et al., 2007) als transkribiert angesehen, wenn der 'detection-value' (Kuhn et al., 2004) einen Wert  $\geq 0,99$  aufwies. Streudiagramme (scatterplots) und das Bestimmtheitsmaß zum Vergleich des Transkriptoms zweier Proben wurden ebenfalls mit der BeadStudio Software erstellt.

### **2.3.8.2 Agilent Microarrays**

Teil dieser Arbeit war unter anderem die Detektion ENU-induzierter Spleiß-Mutationen in ES-Zellen der Maus. Zu diesem Zweck wurden sogenannte 'Oligonucleotide-Custom-Microarrays' der Firma Agilent genutzt. Wie das Wort 'custom' bereits andeutet, können Oligonukleotid-Sonden beliebiger Sequenz in Längen von 25 bis 60 bp nach Kundenwunsch von Agilent *in-situ* direkt auf der Glasoberfläche des Microarrays mittels sogenannter Ink-Jet Technologie synthetisiert werden. Neben den frei wählbaren Oligonukleotid-Sonden wurden von der Firma Agilent automatisch solche Oligonukleotide auf den Array aufgebracht, über die nach der Hybridisierung mittels der Feature Extraction Software (V. 9.3.1) eine Rasterung des Microarray-Bildes und damit eine genaue Zuordnung der Oligonukleotide stattfand. Des Weiteren wurden automatisch interne Kontroll-Oligonukleotide aufgebracht. Zu dem Amplifikationsansatz der zu hybridisierenden RNA-Proben wurden ein Mix von Kontroll-RNAs (One-color spike-ins) der Firma Agilent gegeben, die an diese speziellen Oligonukleotide binden sollten. Diese Kontroll-RNAs waren in verschiedenen Konzentrationen enthalten, sodass in der Auswertung kontrolliert werden konnte, ob eine lineare Amplifikation der RNA-Probe stattgefunden hat, in welchem Signalintensitäts-Bereich lineare Ergebnisse zu erwarten waren und ab welchen Intensitäten von Hintergrundsignal bzw. einer Signalsättigung ausgegangen werden konnte. Eine Übersicht zu den verwendeten

Microarray-Formaten gibt Tabelle 5. Das finale Design und die Zusammenstellung der Oligonukleotid-Proben erfolgte über die von der Firma Agilent zur Verfügung gestellte Web-Applikation 'eArray' (<https://earray.chem.agilent.com/earray/>). Es wurden 20 bp lange Agilent Standard Linker (Spacer) Oligonukleotidsequenzen an den 3'-Enden der Oligonukleotide vorgesehen. Dieser Linker hat die Aufgabe eines Abstandhalters zwischen Oberfläche des Arrays und der Oligonukleotidsequenz und verbessert dadurch die Hybridisierungseigenschaften des eigentlichen Oligos. Die Sequenz des Linkers ist synthetisch erstellt und bis dato in keiner Sequenzdatenbank eines Organismus verzeichnet.

**Tabelle 5: Verwendete Agilent One-Color Custom-Microarrays**

Die 8 x 15 K Microarrays waren erst ab Frühjahr 2007 verfügbar.

	<b>4 x 44 K</b>	<b>8 x 15 K</b>
Felder / Microarray	4	8
Gesamtzahl Oligos / Feld	45220	15744
Verfügbare Oligos / Feld	43803	15208
Agilent Kontroll-Oligos	1417	536

RNA Isolation und Qualitätskontrolle der Gesamt-RNA wurden wie bereits beschrieben durchgeführt (2.3.3). Von einem DNase-Verdau wurde in dem Handbuch (One-Color Microarray-Based Gene Expression Analysis Version 5.5, [www.agilent.com](http://www.agilent.com)) explizit abgeraten. Für die Hybridisierung war die Amplifikation der in der Gesamt-RNA enthaltenen mRNA und der Einbau markierter Nukleotide in diese cRNA (aRNA = amplified RNA) nötig. Das dazu verwendete 'Low RNA Input Linear Amplification Kit' (2.1.4) basierte auf der linearen Amplifikation von RNA nach Eberwine (1996). Im Gegensatz zu dem Kit der Firma Illumina (2.3.8.1) wurden mit diesem Kit Cyanine 3-CTP Nukleotide direkt in die aRNA eingebaut, sodass eine nachträgliche indirekte Färbung über die Bindung von Cy3-Streptavidin an Biotin markierte UTP überflüssig wurde. Die Amplifikation erfolgte strikt nach Hersteller-Protokoll unter Verwendung von 400 ng Gesamt-RNA. Der erfolgreiche Einbau der Cy3-CTP Nukleotide wurde spektrophotometrisch bestätigt (2.3.2). Die weitere Aufreinigung der aRNA, die 16-stündige Hybridisierung bei 65°C und 10 rpm im Hybridisierungsofen, die Waschschrte und das Scannen (Agilent Microarray Scanner AA, 2.1.3) wurden exakt nach Handbuch durchgeführt. Bei den Waschschrten wurde auf den optionalen Waschschrte mit Acetonitril verzichtet, da in Räumen mit Ozon-Werten kleiner 110 µg/m<sup>3</sup> gearbeitet werden konnte. Die Auswertung der Microarray-Bilder, d.h. die Zuordnung der Lichtsignale zu den entsprechenden Oligonukleotid-Sonden und die Transformation der analogen Lichtsignale in digitale Werte erfolgte mit der Feature Extraction Software (V. 9.3.1). Die Analyse der Signale und die genauere Beschreibung des

Microarray-Designs wird in dem Ergebnisteil vorgenommen, da diese Teile der in dieser Arbeit verfolgten Methodenentwicklung darstellen.

### **2.3.8.3 Wiederverwendung von Agilent-Microarrays**

Angesichts der Tatsache, dass die Kosten eines auf Microarrays basierenden Screeningverfahrens durch Wiederverwendung eines Arrays drastisch gesenkt werden können, wurde ein bereits etabliertes Protokoll (Hahnke et al., 2007) auf die Agilent-Microarrays angewandt. Diese Protokoll basiert auf einer alkalischen und enzymatischen Degradierung der an Oligonukleotid-Proben gebundenen RNA und einem Abbau restlicher RNase mittels Proteinase K Verdauung.

Ein bereits genutzter Microarray-Objekträger wurde zunächst für 10 Minuten bei Raumtemperatur in einem Standardfärbegefäß mit 0,05 N NaOH Lösung unter leichtem Schwenken inkubiert. Anschließend wurde der Objekträger kurz mit Strip2-Puffer (2.1.7) abgespült und für 5 Minuten bei RT unter leichtem Schwenken in einem Färbegefäß gewaschen. Nach einer Zentrifugation (15 x g, 1 min) in einem 50 ml Reaktionsgefäß wurde der trockene Array in eine Hybridisierungskammer (Atlas Glass Hybridization Chamber, 2.1.2) mit 1,5 ml 37°C warmem Strip3-Puffer (2.1.7) überführt und unter leichtem Schwenken im 37°C warmen Wasserbad inkubiert. Nach 1 Stunde erfolgte die Zugabe von 75 µl RNase-Cocktail (2.1.6). Nach einer weiteren Stunde wurde der Objekträger kurz in Strip2-Puffer gewaschen und in eine frische Hybridisierungskammer mit 2 ml Strip4-Puffer überführt (1h, 37°C). Danach wurde der Array für 3 Minuten bei 60°C mit 6 x SSC und für 5 Minuten bei RT mit 0,1 x SSC gewaschen und durch Zentrifugation (s.o.) getrocknet. Der Erfolg der Reinigung wurde durch einen abschließenden Scan des Microarrays nach Protokoll der Firma Agilent (2.3.8.2) überprüft. Arrays wurden maximal dreimal verwendet, dabei wurden keine Qualitäts-Einbußen festgestellt.

## **2.3.9 Sequenzierverfahren**

### **2.3.9.1 Cycle-Sequencing**

Die Sequenzierung von PCR-Produkten (2.3.7) erfolgte nach der von Sanger (Sanger et al., 1977) entwickelten Kettenabbruch Methode. Bei dieser als „Cycle-sequencing“ bekannten Methode findet im Vorfeld eine in-vitro-Amplifikation der DNA statt, die durch basenspezifische Kettenabbrüche gekennzeichnet ist.

PCR-Proben wurden im Vorfeld mit dem Invisorb Spin DNA Extraction Kit der Firma Invitex (2.1.4) nach Herstellerangaben aufgereinigt, um Primer und Pufferreste zu entfernen.

Die Amplifikation der PCR-Produkte und der Einbau der ddNTP erfolgte unter Verwendung des 'BigDye Terminator v1.1 Cycle Sequencing Kit' (2.1.4) mit dem in Tabelle 6 angegebenen Thermoprofil.

**Tabelle 6: Thermoprofil Cycle-Sequencing**

Die Primer-Anlagerungstemperatur richtete sich nach dem  $T_m$ -Wert des verwendeten Primers.

Zyklen	Reaktionsschritt	Temperatur	Zeit
1 x	Initialdenaturierung	96°C	1 min
25 x	Denaturierung	95°C	10 s
	Primer-Anlagerung	58-64°C	5 s
	Elongation	60°C	4 min

Der Ansatz (10 µl) setzte sich aus 3 ng PCR-Produkt / 100 bp, 20 pmol des zu verwendenden Primers und 2 µl BigDye Terminator Mix zusammen. Nach erfolgter Amplifikation wurde der Ansatz mittels Ethanol-fällung aufgereinigt, resuspendiert und auf den Kapillarsequenzierer ABI 3730 (2.1.3) aufgetragen. Es wurde die Software ABI Version 1.6.0 verwendet. Die Sequenzierung wurde von einer Service-Gruppe am MPI durchgeführt. Ein Analyse der Sequenzdaten erfolgte mit der Software BioEdit in der jeweils aktuellsten Version (<http://www.mbio.ncsu.edu/BioEdit/BioEdit.html>).

### 2.3.9.2 Illumina/Solexa-Sequencing

Illumina/Solexa-Sequenzierung (Bentley 2006) ist ein Sequenzierverfahren der zweiten Generation ('Sequencing-by-Synthesis') und erlaubt das parallele Sequenzieren mehrerer Millionen DNA-Moleküle aus einem Molekül-Gemisch. Prinzipiell erfolgt hier die Amplifikation einzelner DNA-Moleküle auf der Oberfläche einer Durchflussskammer der Illumina 1G Sequencing Plattform (2.1.3) gefolgt von der Sequenzierung mittels Fluorophor-markierter reversibler Kettenabbruchterminatoren. Pro Zyklus wird eine Base eines Ausgangsmoleküls sequenziert. Bis Frühjahr 2008 konnten 36 Zyklen vollzogen werden. Die Durchflussskammer besteht aus acht einzelnen Kammern, sodass acht Proben parallel analysiert werden können. Auf der Oberfläche jeder Kammer können mehrere Millionen Moleküle gebunden, amplifiziert und sequenziert werden. Eine graphische bzw. multimediale Präsentation der Methode ist auf den Webseiten des Herstellers zu finden (u.a. [http://www.illumina.com/systems/genome\\_analyzer.ilmn](http://www.illumina.com/systems/genome_analyzer.ilmn)).

Die Generation von DNA-Fragment Bibliotheken erfolgte unter Verwendung von 0,5 µg genomischer Maus-DNA bzw. Maus-ds-cDNA. Die DNA wurde für eine Stunde mit einem UTR200 Ultraschallgerät bei 100% Amplitude und 0,5 Zyklus Modus auf eine Größe von 300

bis 400 bp geschert. Adapterligation und Amplifikation wurden streng nach Angaben des Herstellers mit dem Solexa SSB Kit (2.1.3) durchgeführt. Im Speziellen wurden hier nach der Adapterligation Fragmente mit einer Größe von 120 bis 170 bp aus einem 2%-igen Agarosegel aufgereinigt und in der anschließenden Amplifikation eingesetzt. Die amplifizierten Proben wurden in einer Konzentration von 2 pM auf die Durchflusskammer geladen. Zur Generation der Daten (Sequenzen, Qualitätswerte) aus den entsprechenden Bilddateien wurde die Illumina-Pipeline Version 02.2.3 verwendet. Die Generation der DNA-Bibliothek, das Auftragen der Proben und die Steuerung der Illumina 1G Sequencing Plattform wurden freundlicherweise von Tatiana Borodina am MPI für Molekulare Genetik durchgeführt.

## 2.4 Bioinformatische Software und Datenbanken

### 2.4.1 Software

In dieser Arbeit wurde eine Reihe verschiedener Programme genutzt, die insbesondere die Arbeit mit DNA-Sequenzen ermöglichen und vereinfachen.

#### Tabelle 7: Software

Die Tabelle führt alle Programme auf, die in dieser Arbeit zur Analyse und Verarbeitung von Sequenzdaten genutzt wurden. Es wurde jeweils die aktuellste Version genutzt.

Name	Verwendungszweck	Referenz
BioEdit	DNA-Sequenzanalyse	<a href="http://www.mbio.ncsu.edu/BioEdit/BioEdit.html">http://www.mbio.ncsu.edu/BioEdit/BioEdit.html</a>
BioMart	Data-Mining	<a href="http://www.biomart.org/">http://www.biomart.org/</a>
BLAST	DNA-Sequenzhomologie Analysen	<a href="http://blast.ncbi.nlm.nih.gov/Blast.cgi">http://blast.ncbi.nlm.nih.gov/Blast.cgi</a>
ChemSketch	Zeichnen von Strukturformeln	<a href="http://www.acdlabs.com/chemsketch">http://www.acdlabs.com/chemsketch</a>
eArray	Design von Agilent Arrays	<a href="http://earray.chem.agilent.com/earray/">http://earray.chem.agilent.com/earray/</a>
ELAND	DNA-Sequenzhomologie Analysen von Solexa/Illumina Sequenzier Ergebnissen	Anthony J. Cox., Solexa Limited
FatiGO	Gen-Ontologie Analysen	<a href="http://www.fatigo.org/">http://www.fatigo.org/</a>
Primer3	Primerdesign	<a href="http://fokker.wi.mit.edu/primer3/input.htm">http://fokker.wi.mit.edu/primer3/input.htm</a>
UCSC Genome Browser	Data-Mining, DNA-Sequenzhomologie Analysen, BLAT	<a href="http://genome.ucsc.edu/">http://genome.ucsc.edu/</a>
Microsoft Office	Auswertung, Dokumentation, Grafiken	<a href="http://www.microsoft.com/">http://www.microsoft.com/</a>

Eine Zusammenstellung der genutzten kommerziell vertriebenen bzw. frei zugänglichen Software und deren allgemeine Verwendungszwecke gibt Tabelle 7. Etwaige Details zur

Nutzung werden bei der Besprechung der jeweiligen Ergebnisse gegeben. Geräte spezifische Software ist unter 2.1 aufgeführt.

## 2.4.2 Spezielle Software

Für spezielle Anwendungen zu denen keine Software verfügbar war, wurden Programme in der von Larry Wall 1987 entworfenen Skriptsprache 'Perl' der Version 5.x geschrieben. Perl ist eine leicht erlernbare interpretierte Programmiersprache, die weite Verbreitung in der Bioinformatik gefunden hat. Eine ihrer Stärken ist die einfache Verarbeitung von Texten mittels einer Vielzahl regulärer Ausdrücke. DNA-Sequenzen können im weitesten Sinne auch als Texte interpretiert werden, sodass sich Perl hier als Programmiersprache für DNA-Analysen anbietet.

Auf die in dieser Arbeit verwendeten Skripte wird im Ergebnisteil an entsprechender Stelle eingegangen, da sich ein Großteil der Algorithmen erst aus Ergebnissen vorangegangener Untersuchungen und Überlegungen entwickelt haben.

## 2.4.3 Datenbanken

In dieser Arbeit wurden verschiedene öffentlich verfügbare Nukleotid-Sequenz-Datenbanken genutzt. Die Sequenzen dienten beispielsweise als Referenz für die Suche nach Homologien oder als Basis für die *in silico* Oligo-Synthese.

**Tabelle 8: Nukleotid-Sequenz-Datenbanken**

Die Tabelle gibt einen Überblick über die verwendeten Sequenz-Datenbanken. Ist keine eindeutige Versionsnummer bzw. ein Zeitraum anstelle eines eindeutigen Download-Datums angegeben, so wurde bei den Analysen die jeweils aktuellste Version verwendet. Diese verwendeten Versionen können dann den Ergebnissen entnommen werden.

Sequenz-Datenbank	Version	Datum/ Zeitraum	Verwendungs zweck	Link
dbSNP126	NCBI Build 36	08.01.2008	SNP Abgleich	<a href="http://www.ncbi.nlm.nih.gov/">http://www.ncbi.nlm.nih.gov/</a>
Ensembl mouse	NCBI Build 34-37	2005-2008	Oligodesign, BioMart	<a href="http://www.ensembl.org/">http://www.ensembl.org/</a>
Maus ESTs	NCBI Build 37 (mm9)	27.11.2007	ELAND, BLAST	<a href="http://genome.ucsc.edu/">http://genome.ucsc.edu/</a>
Maus Genom (C57BL/6J)	NCBI Build 36 (mm8)	20.05.2007	ELAND, BLAST	<a href="http://genome.ucsc.edu/">http://genome.ucsc.edu/</a>
Maus Genom (Celera)		13.11.2007	ELAND, BLAST	<a href="http://www.ncbi.nlm.nih.gov/">http://www.ncbi.nlm.nih.gov/</a>
Maus RefSeq mRNA	NCBI Build 37 (mm9)	21.09.2007	ELAND, BLAST	<a href="http://genome.ucsc.edu/">http://genome.ucsc.edu/</a>
NCBI mouse collection	NCBI Build 34-37	2005-2008	BLAST (web-basiert)	<a href="http://www.ncbi.nlm.nih.gov/">http://www.ncbi.nlm.nih.gov/</a>

### 3 Versuchsbeschreibungen und Ergebnisse

In dieser Arbeit wurden Untersuchungen zu der Entwicklung neuartiger Screeningverfahren zur Detektion von Exon-Deletions Mutationen vorgenommen. Dazu wurde zunächst eine Bibliothek ENU mutierter Stammzellen erstellt, die das Ausgangsmaterial für die Durchmusterungsversuche darstellte. Nach der Charakterisierung von Positiv-Kontrollen und grundlegenden Voruntersuchungen wurde als erste Strategie der Detektion ein auf der Microarray Technologie basierendes Verfahren betrachtet. Ein weiterer Teil der Arbeit beschäftigte sich dann mit der Möglichkeit neuartige Sequenzierverfahren, hier im speziellen Illuminas Solexa Sequenzier Technologie, für die Analyse von Transkriptomen zu nutzen.

#### 3.1 ENU-Mutagenese

Zur Etablierung neuer Screeningverfahren zur Mutationsdetektion in embryonalen Stammzellen der Maus war es zunächst notwendig, Stammzellen mit ENU zu mutagenisieren und den Erfolg der Mutagenesen zu bestätigen, um so eine entsprechende Bibliothek mutierter Zellen für die folgenden Experimente zu erstellen. An einer kleinen Anzahl 6-Thioguanin resistenter Klone wurde das durch ENU-Mutagenese induzierte Mutationsspektrum durch Sequenzierung der *Hprt* cDNA analysiert.

##### 3.1.1 Generierung einer Test-Bibliothek ENU mutierter ES-Zellen

Ein entsprechendes ENU-Mutagenese-Protokoll war am MPI für Molekulare Genetik bereits etabliert (Greber et al., 2005). Es war daher auch bekannt, dass bei einer ENU-Konzentration von 2 mM und einer 90-minütigen Inkubationszeit *Hprt*-Mutationsfrequenzen (2.2.8) von ca.  $4 \times 10^{-4}$  (= 1 von 2500 überlebenden ES-Zellen trägt eine funktionelle Mutation im *Hprt*-Lokus) zu erwarten sind. Diese Frequenz hatte sich bereits als Standard in der Arbeitsgruppe bewährt, da einerseits ausreichend Mutationsereignisse in der DNA stattfanden, andererseits die ES-Zellen durch die Mutationsdosis nicht wesentlich in ihrer Fähigkeit der Keimbahngängigkeit beeinträchtigt wurden (Greber et al., 2005).

Die ersten beiden Mutagenesen (2.2.7) dienten der Kontrolle der Mutationsbedingungen und fanden daher unter Verwendung der Zelllinie CGR8 der Passage 16 (2.1.8) statt. Diese hat gegenüber der F1-Zelllinie (2.1.8) den Vorteil, dass sie eine höhere Wachstumsrate aufweist und ohne feeder-layer problemlos kultivierbar ist. Das vereinfacht die Kulturbedingungen und beschleunigte somit die ersten Voruntersuchungen zur Einstellung der Mutationsbedingungen. Der Nachteil dieser Zelllinie besteht in ihrer schwachen Keimbahngängigkeit, was im Fall dieser ersten Test-Experimente aber nicht von Bedeutung war, da nur mit der F1 Zelllinie

Bibliotheken aufgebaut werden sollten. In den ersten beiden Mutagenesen der CGR8 Zelllinie mit 2 mM ENU und 90-minütiger Inkubation konnten *Hprt*-Mutationsfrequenzen um 1/2500 ( $4,58 \times 10^{-4}$  (1/2184) bzw.  $3,60 \times 10^{-4}$  (1/2780)) erzielt werden (Tabelle 9). Da die Werte der Parameter 'Überlebensrate und Mutationsfrequenz', den Erwartungen entsprachen, wurden im Anschluss zwei weitere Mutagenesen mit der Zelllinie F1 (2.1.8) unter gleichen Bedingungen durchgeführt. Diese Mutagenesen lieferten auch zufrieden stellende Ergebnisse bezüglich der zu erzielenden Mutationsfrequenz von ca. 1/2500. Tabelle 9 zeigt die Ergebnisse der vier unabhängigen Mutagenesen.

**Tabelle 9: Charakteristische Parameter unabhängiger ENU-Mutagenesen**

Die Tabelle zeigt die Ergebnisse der im Text beschriebenen ENU Mutagenesen unter Verwendung von 2 mM ENU und einer Inkubationszeit der Zellen von 90 Minuten.

ENU-Mutagenese	Überlebensrate	<i>Hprt</i> -Mutationsfrequenz
#1 CGR8-1	5,5%	$4,58 \times 10^{-4}$ (1/2184)
#2 CGR8-2	6,8%	$3,60 \times 10^{-4}$ (1/2780)
#3 F1-1	6,3%	$4,82 \times 10^{-4}$ (1/2075)
#4 F1-2	3,9%	$3,45 \times 10^{-4}$ (1/2901)

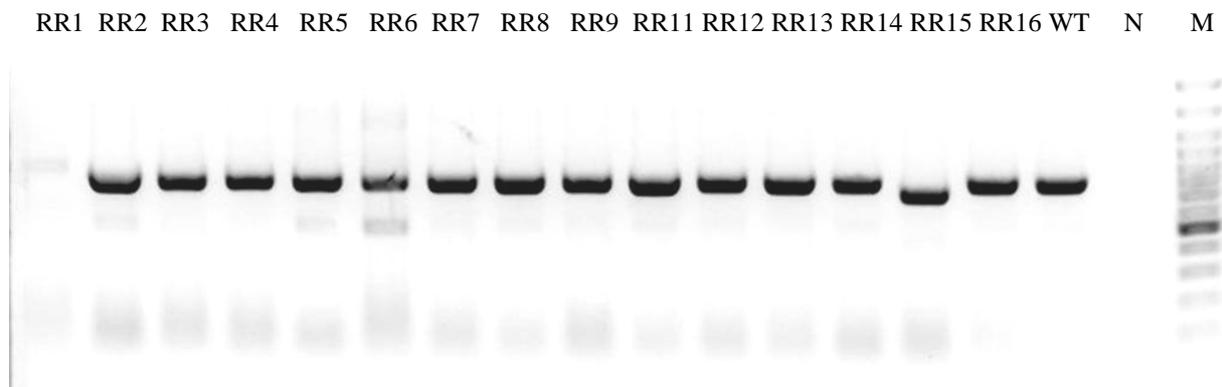
Kryokonservierte Zellen der ENU-Mutagenese #3 F1-1 wurden nach der Bestätigung der erfolgreichen Mutagenese zur Generierung einer kleinen Test-Bibliothek ausplattiert und ca. 200 Kolonien wurden subkloniert (im Folgenden mit RRx### bezeichnet) (2.2.5). Diese Klone wurden einzeln weiter expandiert und Aliquots zur DNA/RNA-Extraktion und zur Konservierung eingefroren (2.2.3).

### 3.1.2 ENU Mutationsspektren

Ein Ziel der Arbeit war, neuartige Screeningverfahren für die Detektion von Spleißmutationen in ES-Zellen zu entwickeln. Daher war es von Bedeutung, die Wahrscheinlichkeit für das Auftreten solcher Mutationen erneut zu bestimmen und mit den bis dato bekannten Angaben aus der Literatur (Chen et al., 2000; Greber et al., 2005; Justice et al., 1999) von ca. 20%-iger Wahrscheinlichkeit für Exon-Deletionen abzugleichen.

Hierzu wurde eine kleine Stichprobe von 25 6-TG resistenten Klonen aus der Bestimmung der Mutationsfrequenz der Mutagenese #3 F1-1 (Tabelle 9) hinsichtlich des Vorkommens von Exon-Deletionen auf Transkript-Ebene untersucht. RNA wurde aus den subklonierten ES-Zell Klonen extrahiert (2.3.3), DNaseI verdaut und mittels OligodT Priming in cDNA (2.3.5) umgeschrieben. Mittels RT-PCR (2.3.7) wurde dann ausgehend von 1 µl cDNA Matrize ein 797 bp großes Fragment bestehend aus dem offenen Leserahmen (ORF) und Teilen der 5' und 3' untranslatierten Bereiche (UTR) des *Hprt* Gens amplifiziert. Für die Amplifikation wurden

die in den UTR Bereichen platzierten Primer #Hprt\_UTR\_Fw und #Hprt\_UTR\_Rev verwendet. Abbildung 8 zeigt beispielhaft die gelelektrophoretische Auftrennung einiger analysierter *Hprt* PCR Produkte der untersuchten Funktionsmutanten. Hier ließ sich schon eindeutig erkennen, dass die Größe des PCR-Produktes der Mutante RR1 über und die des Amplifikates der Mutante RR15 unter der erwarteten Wildtyp Fragmentlänge von 797 bp lagen. Dies deutete auf Spleißmutationen hin. Auch das Produkt der Mutante RR21 zeigte eine geringere Größe in der Gelelektrophorese (Daten nicht gezeigt). Für die drei Klone RR10, RR17 und RR22 konnten keine *Hprt* PCR-Produkte amplifiziert werden.



**Abbildung 8: Gelelektrophoretische Auftrennung von *Hprt* RT-PCR Produkten**

Dargestellt ist die Ethidiumbromid vermittelte Fluoreszenz von RT-PCR Produkten nach elektrophoretischer Auftrennung in einem 1 % igen Agarosegel. Es wurden 5 µl eines Reaktionsansatzes aufgetragen. Die aufgetrennten DNA-Fragmente wurden mittels RT-PCR aus dem *Hprt* Transkript verschiedener *Hprt* Funktionsmutanten mit den Primern Hprt\_UTR\_Fw und Hprt\_UTR\_Rev generiert. WT = F1 Wildtyp Matrize; M = Größenstandard ‚GeneRuler™ 100 bp Plus DNA Ladder‘ 500 ng; N = Negativ-Kontrolle (ohne Matrizen DNA); RR = *Hprt* Funktionsmutante Nr.

Die PCR-Produkte wurden weiterhin säulenchromatographisch aufgereinigt und mittels Cycle-Sequencing analysiert (2.3.9.1). Die Sequenzierung erfolgte jeweils aus beiden Richtungen unter Verwendung oben genannter Hprt\_UTR Primer.

Die Ergebnisse der Sequenzierung sind tabellarisch in Tabelle 10 aufgeführt. Die zu Beginn angestrebte Klonzahl von 25 reduzierte sich auf 18 nichtredundante Klone, da für drei Klone kein PCR-Produkt amplifizierbar war und drei Mutationsereignisse redundant waren. In 11% der Fälle konnte demnach die Deletion eines Exons beobachtet werden. Den größten Anteil am Mutationsspektrum hatten die TA-AT Transversionen mit 44% aller Mutationen. In dem sequenzierten Bereich des *Hprt* Transkripts ließ sich für den Klon RR6 trotz 6-TG Resistenz keine Mutation erkennen.

**Tabelle 10: Mutationspektrum eines Sets von 22 6-Thioguanin resistenten ES-Zell Klonen**

Die Tabelle gibt einen Überblick über die mittels Sequenzierung ermittelten Mutationsereignisse in 22 6-TG resistenten ES-Zell Klonen der Mutagenese #3 F1-1 (Tabelle 9). Drei Mutationen aus dem gleichen Selektionsexperiment traten mehrfach auf. Dadurch reduzierte sich die Anzahl nichtredundanter Klone von 22 auf 18. Die Positionen beziehen sich auf das *Hprt* Gen, (NM\_013556).

Klon	Mutation	Position in CDS	Effekt
RR1	Insertion	Exon 6	68 bp (Intron 5) an Exon 6, Stop
RR 2	GC-TA Transversion	34	D -> Y
RR 3, RR 23	TA-GC Transversion	158	V -> G
RR 4	TA-AT Transversion	543	F -> L
RR 5	GC-TA Transversion	119	G -> V
RR 6	WT	-	-
RR 7, RR 20, RR 24	CG-TA Transition	527	P -> L
RR 8	TA-AT Transversion	299	I -> N
RR 9	TA-AT Transversion	389	V -> D
RR 11	TA-AT Transversion	563	V -> D
RR 12	TA-GC Transversion	407	I -> R
RR 13	GC-AT Transition	47	G -> D
RR 14	TA-AT Transversion	299	I -> N
RR 15	Deletion	Exon 8	Stop
RR 16	TA-AT Transversion	69	C -> Stop
RR 18, RR 25	TA-AT Transversion	389	V -> D
RR 19	TA-AT Transversion	23	V -> D
RR 21	Deletion	Exon 6	Stop

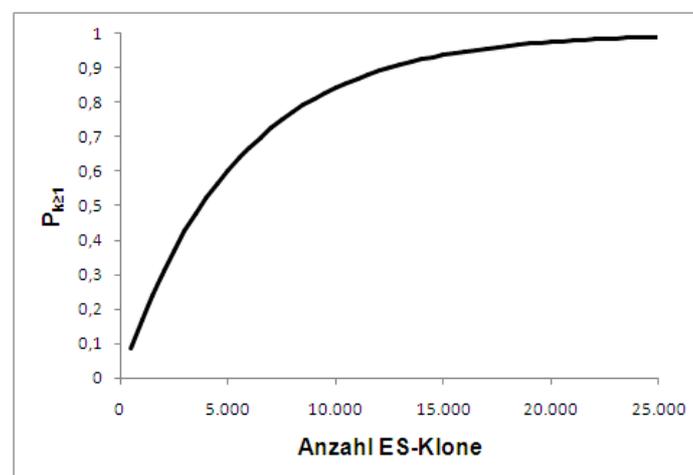
### 3.1.3 Erwartete Häufigkeit von Exon-Deletionsmutanten

Anhand der ermittelten Mutationsfrequenzen des *Hprt*-Lokus, der Wahrscheinlichkeit für das Auftreten von Exon-Deletion im *Hprt*-Transkript und der Anzahl exprimierter Gene erfolgte im Vorfeld eine Abschätzung der Häufigkeit mit der Exon-Deletionen anzutreffen sind. Hervorzuheben ist bei dieser Annäherung, dass es sich um eine reine Schätzung handelt, da einige Prämissen getroffen wurden. Eine Annahme dabei ist die, dass sich die für das *Hprt* Transkript ermittelten Werte weitestgehend auf den Durchschnitt der anderen Gene übertragen lassen. *Hprt* besteht aus 9 Exons. Für die durchschnittliche Anzahl an Exons in allen RefSeq-Transkripten der Maus wurde ein Wert von 9,7 ermittelt. Da es sich bei *Hprt* zudem um ein X-chromosomal lokalisiertes Gen handelt, wurde bei dieser Schätzung auch eine Verdopplung der Häufigkeit für das Auftreten von Deletionen für autosomal lokalisierte Gene angenommen (Autosomale Gene ca 95,8%, ENSEMBL 46; Faktor =  $F = 2 \times 0,958 = 1,916$ ). Weiterhin basiert die Ermittlung der *Hprt*-Mutationsfrequenz auf einer Selektion nach Funktionsmutanten. Für den Anteil von Exon-Deletionen an allen Mutationstypen wurde ein Wert von 20% angenommen, der denen in der Literatur angegebenen Werten größerer Stichproben von 19 bis 26% entspricht (Barbaric et al., 2007; Chen et al., 2000; Greber et al.,

2005; Justice et al., 1999; Noveroske et al., 2000). Dem im Rahmen dieser Arbeit ermittelten Anteil von 11% lagen lediglich zwei Exon-Deletionen zu Grunde.

<i>Hprt</i> Mutationsfrequenz der verwendeten Zelllinie F1-1:	$MF = 4,82 \times 10^{-4}$
Anteil Exon-Deletionen:	$P_E \approx 0,2$
Faktor Verdopplung autosomal:	$F \approx 1.916$
Hypothetische Exon-Deletionsfrequenz:	$MF_E = MF \times P_E \times F \approx 1.85 \times 10^{-4}$
Anzahl exprimierter Gene (vergl. 3.2):	$A_G \approx 7.475 \text{ Gene/Klon}$
Theoretisch detektierbare Exon-Deletionen = $MF_E \times A_G \approx$	<u>1,38 funktionelle Exon-Deletionen pro Klon</u> (1/0,72)

Diese Schätzung sagt aus, dass bei einer betrachteten Anzahl von 7.475 Genen eine Exon-Deletion in jedem 0,72ten Klon detektierbar sein sollte (1,38 Exon-Deletionen pro Klon). Da in dem ersten Teil dieser Arbeit allerdings nur solche Exons betrachtet werden, die nicht den ORF unterbrechen (ca. 31% aller Exons; berechnet anhand von RefSeq Sequenzen) verringert sich die Anzahl detektierbarer Exon-Deletionen auf eine Deletion in ca. 2,32 ENU mutierten Klonen. Groß angelegte Screenings müssen daher auf geschickt gewählten Durchmusterungsstrategien und sensitiven Methoden basieren, um das parallele Screening mehrerer Klone zu ermöglichen und somit wirtschaftlich sinnvoll zu sein.



**Abbildung 9: Detektions Wahrscheinlichkeit mindestens einer Exon-Deletion in einem bestimmten Gen in Abhängigkeit von der Anzahl untersuchter ES-Zell Klone**

Die Berechnung erfolgte mittels der Binominalverteilung unter Verwendung der im Text beschriebenen Annahmen und Faktoren:  $P_{k \geq 1} = P_0 = \binom{n}{k} * MF_E^k * (1 - MF_E)^{n-k}$ ,  $n$  = Anzahl betrachteter ENU mutierter ES-Zell Klone,  $k$  = Anzahl ENU mutierter ES-Zell Klone mit Exon-Deletions Mutation in einem bestimmten Gen,  $MF_E$  = hypothetische Exon-Deletionsfrequenz (s. Text),  $P$  = Detektionswahrscheinlichkeit in Abhängigkeit von  $k$  und  $n$ .

Mit Hinblick auf die künftig geplante Anlegung eines ES-Zell Archivs kann weiterhin unter Verwendung der Binominal-Verteilung berechnet werden, wie hoch die Wahrscheinlichkeit

ist, in einem Archiv bestimmter Größe ( $n$  ES-Zell Klone) mindestens eine Exon-Deletion ( $k \geq 1$ ) für ein bestimmtes Gen zu identifizieren. Dies kann nur unter oben beschriebenen Annahmen erfolgen, sodass es sich mehr um eine Annäherung handelt. Die Ergebnisse sind graphisch in Abbildung 9 dargestellt. Mit ungefähr 95%-iger Wahrscheinlichkeit kann mindestens eine Exon-Deletion in einem bestimmten Gen bereits ab einer ES-Zell Archiv Größe von ca. 16.500 Klonen identifiziert werden, mit 99%-iger Wahrscheinlichkeit ab einer Größe von ca. 25.000 Klonen.

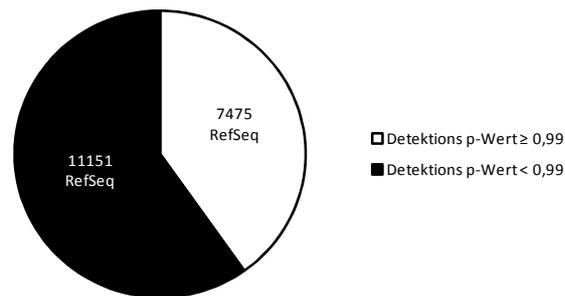
## 3.2 Genexpressions Untersuchungen

### 3.2.1 Genexpression in Wildtyp F1 ES-Zellen

Exon-Deletionen lassen sich am einfachsten und eindeutigsten durch Analyse der Transkripte des jeweiligen Gens identifizieren. Ein Ziel dieser Arbeit war es, das Transkriptom einer mutierten ES-Zelle global nach Exon-Deletions-Ereignissen durchforsten zu können. Daher machte es Sinn zunächst festzustellen, welche Gene in der verwendeten Zelllinie exprimiert waren, da nur bei diesen Genen der Nachweis einer Deletion auf Transkript-Ebene möglich ist.

Die bis dato zur Transkriptom-Analyse geeigneteste Methode war die Messung der Genexpression mittels der Microarray-Technologie. Hier wurde im speziellen die Microarray Plattform der Firma Illumina mit den kommerziell vertriebenen MouseRef-8 v1.1 Bead Chip Arrays genutzt (Kuhn et al., 2004). Damit konnte die Expression der im Maus RefSeq Datensatz (Pruitt et al., 2007) erhaltenen Transkripte analysiert werden.

RNA aus Wildtypzellen der F1 Stammzelllinie wurde isoliert, die Integrität bestätigt, nach Protokoll mit dem Illumina TotalPrep RNA Amplification Kit in aRNA umgeschrieben und auf den Bead Chip Array hybridisiert. Die Auswertung und Qualitätskontrolle erfolgte automatisiert über die Illumina BeadStudio Software unter Verwendung der unter 2.3.8.1 genannten Optionen. Der Array umfasste 23.775 durch RefSeq IDs eindeutig identifizierbare Oligonukleotide für 18.626 verschiedene RefSeq Gene. Einige Gene wurden durch verschiedene Oligonukleotide mehrfach repräsentiert, die in unterschiedlichen Bereichen des Transkripts platziert waren. Das Expressionslevel eines Gens war in dieser Untersuchung weitestgehend uninteressant. Von Bedeutung für Folgeexperimente war nur die Wahrscheinlichkeit, mit der nach diesem Experiment davon ausgegangen werden konnte, dass ein Gen in der untersuchten Zelllinie unter den gegebenen Kulturbedingungen exprimiert wird.



#### Abbildung 10: Anzahl exprimierter RefSeq Gene in embryonalen Stammzellen der F1 Linie

Dargestellt ist das Ergebnis einer Microarray Analyse mit dem Ziel die Gene herauszufinden, die mit 99 %-iger Sicherheit in den Stammzellen der Zelllinie F1 exprimiert sind. Hierzu wurde aRNA der F1 Zelllinie auf einem MouseRef-8 v1.1 Bead Chip Array (2.3.8.1) hybridisiert und die RefSeq Gene ermittelt, die einen Detektions p-Wert  $\geq 0,99$  aufwiesen. Insgesamt wurden 18.626 RefSeq Sequenzen berücksichtigt.

Für 8.796 RefSeq Oligonukleotide bzw. 7.475 RefSeq Gene wurde ein Detektions p-Wert (,detection p-value') (Kuhn et al., 2004) größer gleich 0,99 erreicht (Abbildung 10). Für diese Gene konnte durch diesen Algorithmus demnach mit mindestens 99%-iger Wahrscheinlichkeit eine fehlerfreie Detektion und somit Expression nachgewiesen werden. Sollte ein Gen, welches durch mehrere Oligonukleotide repräsentiert wurde, nur für ein Oligonukleotid einen Detektions p-Wert von 0,99 aufweisen, so wurde die Transkription des Gens dennoch als bestätigt angesehen.

### 3.2.2 Transkriptom-Analysen ENU mutierter F1 ES-Zellen

Der Einfluss einer ENU Mutagenese auf das Transkriptom von F1 ES Zellen wurde exemplarisch durch Microarray Analysen mit dem MouseRef-8 v1.1 Bead Chip an drei ENU Klonen betrachtet. Als Referenz Transkriptom diente ein biologisches Replikat einer RNA Probe des unter 3.2.1 verwendeten F1 Wildtyps.

Diese Untersuchung verfolgte hauptsächlich das Ziel ein grobes Verständnis davon zu bekommen, inwieweit die gewählten Mutagenese Bedingungen die globale Transkription beeinflussen. Weiterhin, mit Hinblick auf die zu einem späteren Zeitpunkt entwickelte Detektion von Exon-Deletion in diesen Klonen, könnten die Daten zudem dazu verwendet werden, den Einfluss einer solchen Deletion auf die Expression zu beurteilen.

Die RNA der Klone RRx000, RRx001 und RRx002 und F1 WT RNA wurde wie bereits unter 3.2.1 beschrieben, vorbereitet, hybridisiert und mittels der zugehörigen Software analysiert.

Der Einfluss der Mutagenese wurde durch zwei Untersuchungen in Zahlen gefasst. Zum einen wurde das Bestimmtheitsmaß  $R^2$  (Quadrat des Pearson'schen Korrelationskoeffizienten R) eines Datensatzes eines ENU-Klons und des Wildtyps als Maß für den linearen

Zusammenhang bestimmt und graphisch in Form von 'scatterplots' dargestellt. Zum anderen wurde die Gesamtzahl detektierbarer und somit als exprimiert angesehener Gene („detection p-value  $\geq 0,99$ ) bestimmt und aus diesen diejenigen herausgefiltert, deren Signalintensität im Klon um mehr als das 3-fache vom Wildtyp abwich. Zu Kontrollzwecken wurden auch die Ergebnisse der beiden Wildtyp Replikate aus den beiden verschiedenen Hybridisierungsereignissen verglichen.

**Tabelle 11: Einfluss der ENU Mutagenese auf das Transkriptom dreier ENU-Klone**

Die als Referenz bezeichnete Probe stammt aus einer anderen Hybridisierung (3.2.1) und ist ein biologisches RNA Replikat der F1 Wildtyp Zelle.

ES-Zelle	Detektierbare Refseq Gene (p $\geq 0,99$ )	R <sup>2</sup>	RefSeq Gene mit $\geq 3$ -fach höherem Signal im Vergleich zur Referenz	RefSeq Gene mit $\geq 3$ -fach niedrigerem Signal im Vergleich zur Referenz
F1 WT (Referenz)	7407	-	-	-
F1 WT (3.2.1)	7475	0,9858	2	1
RRx000	7371	0,9555	200	106
RRx001	7078	0,9703	11	25
RRx002	7234	0,9732	74	25

Die Ergebnisse zeigen einen Trend dahin, dass die Mutagenese einen negativen Einfluss auf die Anzahl detektierbarer RefSeq Gene und damit auf die Genexpression hat. In dem Klon RRx001 sind nur noch 95,6% der Gene der Wildtyp Referenz mit 99%-iger Wahrscheinlichkeit exprimiert. Allerdings wurden in der Voruntersuchung zur Genexpression in Wildtyp F1 Zellen (3.2.1) sogar 68 zusätzliche Transkripte detektiert. Aufgrund der geringen Stichprobengröße und der einzigen Referenz können hier keine Angaben zur Signifikanz der Aussage gemacht werden. Es lässt sich jedoch feststellen, dass das Bestimmtheitsmaß  $R^2 = 0,9858$  für den Vergleich der beiden Wildtyp Transkriptome größer ist als die  $R^2$ -Werte der Vergleiche mit den Mutanten. Das zeigt die größere lineare Abhängigkeit zwischen den beiden Wildtyp-Proben (biologische Replikate aus unterschiedlichen Hybridisierungen) im Vergleich zu der der Mutanten und der Wildtyp Probe (aus der gleichen Hybridisierung) und untermauert somit den Einfluss der Mutagenese auf die Expression einzelner Gene. Dies wird auch durch die höhere Anzahl der Gene mit mehr als 3-fach verstärkter bzw. abgeschwächter Expressionsstärke im Vergleich der Mutanten zur Referenz deutlich. Dennoch konnten im Vergleich der Wildtyp-Proben zwei Gene mit deutlich verstärkter (3,04- und 4,23-fach) und eins mit abgeschwächter (3,84-fach) Signalintensität ermittelt werden. Ein durch die Mutagenese bedingter Trend zur Über- bzw. Unterexpression ist anhand dieser drei Proben nicht belegbar; die Expression wird in beide Richtungen beeinflusst.

### 3.3 Microarray basierte Detektion von Exon-Deletions Mutanten

Mit dieser Arbeit wurde das Ziel verfolgt durch ENU Mutagenese induzierte Exon-Deletionen in embryonalen Stammzellen der Maus zu detektieren. Wie bereits beschrieben sollte die Detektion dabei global im Transkriptom erfolgen, d.h. es wurde nicht nach Mutationen in einzelnen speziellen Kandidaten-Genen gesucht, sondern es sollte das gesamte Transkriptom parallel betrachtet werden.

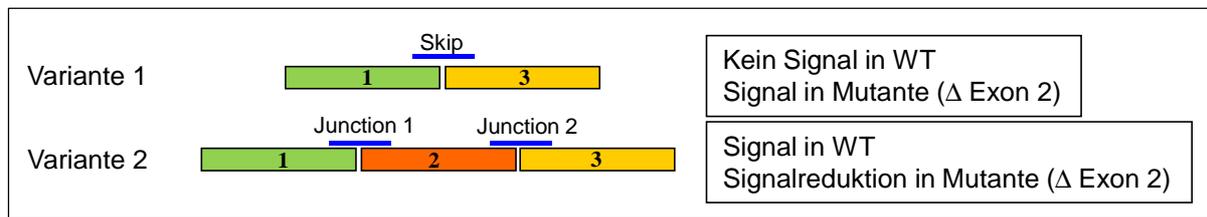
Hinsichtlich der Detektion gab es in dieser Studie allerdings eine wichtige Einschränkung. Es sollte nur nach solchen Exon-Deletionsereignissen gesucht werden, die nicht zu einem Abbruch des offenen Leserasters führen. Auf Transkript-Ebene betrachtet bedeutete dies, dass nur solche Exon-Deletionen von Interesse waren, bei denen die Basenanzahl im deletierten Exon ein Vielfaches von drei betrug. Für diese Restriktion gab es verschiedene Gründe: Zum einen sollten so reine Funktionsverlust-Mutanten (loss-of-function) ausgeschlossen werden, da sich diese ES-Zell-Klone durch andere genetische Mutationsverfahren günstiger und gezielter herstellen lassen (1.3). Es sollte eine Konzentration auf solche Mutanten erreicht werden, die aufgrund eines fehlenden Exons nur eine Zerstörung einzelner Proteindomänen aufweisen. Zum anderen war bekannt, dass mRNAs mit fehlerhaftem ORF über den Mechanismus des 'RNA mediated decay' (McGlincy et al., 2008; Wen et al., 2008) abgebaut werden können, was sie der angedachten Detektion auf Transkriptebene entziehen würde. Letztendlich könnte so der Anteil zu untersuchender Deletionsereignisse im Sinne der Technikentwicklung auf ein sinnvolles Maß reduziert werden. Der Anteil der Exons, deren Basenanzahl ein Vielfaches von drei beträgt, wurde mittels eines Skriptes auf 31% ermittelt (ENSEMBL 39).

Die zu Beginn der Arbeit aktuellste Methode zur Analyse der globalen Genexpression in Zellen und Geweben war die Bestimmung der relativen Expression anhand der Hybridisierung von RNA-Proben auf Oligonukleotid-Microarrays. Daher lag es nahe, ein auf dieser Technologie basierendes System zur globalen Detektion von Exon-Deletionen zu schaffen. Die dazu unternommenen Studien werden im Folgenden genauer erläutert.

#### 3.3.1 Allgemeine Vorüberlegungen zum Array Design

Grundsätzlich betrachtet musste die Detektion über das Design der auf dem Array synthetisierten, zur aRNA komplementären, Oligonukleotide erfolgen. Dafür wurden zwei Varianten in Erwägung gezogen, die schematisch in Abbildung 11 dargestellt sind (Pan et al., 2004). Die Oligonukleotid-Sequenz in Variante 1 sollte sich aus dem 3'-Ende des 5'-wärts gelegenen Exons des deletierten Exons und dem 5'-Ende des 3'-wärts gelegenen Exons

zusammensetzen. Bei einer zu detektierenden Deletion von beispielsweise Exon 2 sollte das Oligonukleotid damit am 5'-Bereich aus dem 3'-Bereich des Exon 1 und am 3'-Bereich aus dem 5'-Bereich des Exon 3 bestehen.



**Abbildung 11: Schematische Darstellung des Oligonukleotid-Aufbaus zur Detektion von Exon-Deletionen**  
Die Abbildung stellt schematisch die Positionierung der Oligonukleotide auf ihren komplementären Exonsequenzen dar. Die bunten Rechtecke verkörpern die Exons. Die Zahl in diesen gibt die Nummer des Exons im Wildtyp Transkript an. Die blauen Linien oberhalb der Exons repräsentieren die Oligonukleotide.

Für diese direkte Verknüpfung von Exon 1 und Exon 3 sollte daher auf dem Microarray nur ein Signal zu detektieren sein, wenn es sich um ein Deletionsereignis von Exon 2 handelt (ausgenommen natürliches alternatives Spleißen). In Variante 2 setzt sich die Oligonukleotid-Sequenz aus dem Übergang von benachbarten Sequenzen (stromauf oder stromabwärts) zum deletierten Exon zusammen. Für das gezeigte Beispiel ‚Junction 2‘ besteht der 5'-Bereich der Sequenz somit aus dem deletierten Exon 2 und der 3'-Bereich aus Exon 3. Hier wäre bei der Microarray-Hybridisierung bei einer Deletion von Exon 2 eine Signalreduktion im Vergleich zum WT-Referenzsignal für dieses Oligo zu erwarten. Variante 2 erfordert demnach 2 Oligonukleotid-Sonden zur eindeutigen Detektion einer Deletion. Diese Tatsache hätte keine Auswirkungen auf die Anzahl nötiger Oligonukleotid-Sonden und damit die Kosten, wenn alle Exons eines Transkripts betrachtet würden. Da in dieser Studie nur solche Exon-Deletionen selektiert werden sollten, die das offene Leseraster nicht unterbrechen, erfordert Variante 2 eine deutliche höhere Anzahl Sonden als Variante 1.

Oligonukleotid-Sonden basierend auf dem Design der Variante 1 werden im Folgenden ‚Skip-Oligos‘ genannt, während die der Variante 2 als ‚Junction-Oligos‘ bezeichnet werden.

### 3.3.2 Entwicklung einer Software zur Erstellung von Skip- und Junction-Oligos

Das Design der auf dem Oligonukleotid-Array zu synthetisierenden Skip- bzw. Junction-Oligo-Sequenzen musste automatisiert erfolgen können, da alleine weit über 30.000 Skip-Oligos für die in der F1 Zelllinie exprimierten Gene (7.475) gestaltet werden mussten. Es wurde daher ein Perl Skript entwickelt, welches ausgehend von einer Exonliste Skip- oder Junction-Oligos automatisiert erstellen konnte und damit die bioinformatische Grundlage für die Detektion von Exon-Deletionen bildete.

Datengrundlage für das Skript bildete eine Liste, welche alle Exons enthält, die in der F1-Stammzelllinie aufgrund vorrangiger Expressionsanalysen als exprimiert angesehen werden konnten. Diese Exon-Liste wurde unter Angabe der exprimierten Gene mittels der BioMart Plattform aus der ENSEMBL Datenbank heraus generiert (3.2.1 / z.B. ENSEMBL 46: 108.410 Exons von 295.164 in der Datenbank vorhandenen RefSeq Exons). Neben der Sequenz des Exons enthielt jeder Eintrag eine eindeutige Exon-Identifikationsnummer, die Ensembl-Gen Nummer, die Ensemble-Transkript Nummer und die Position des Exons innerhalb des Transkripts. Aufgrund der Tatsache, dass identische Exons in unterschiedlichen Transkripten eines Gens aufgrund der Identifikationsnummer zunächst als nicht-identisch betrachtet werden, enthält diese Liste einige Exons mehrfach. Dies erklärt die hohe Anzahl an Exons in der Liste im Vergleich zu der Anzahl von nur 7.475 Genen.

Das entwickelte Skript bestimmt zunächst die Sequenzlänge. Beträgt die Länge aus bereits genannten Gründen ein Vielfaches von drei, werden die flankierenden Exons anhand der Transkript ID und der Exonpositionsnummer ermittelt und analog der Beschreibung unter 3.3.1 zu Skip- oder Junction-Oligos vordefinierter Längen zusammengesetzt. Hier muss angemerkt werden, dass für das erste und letzte Exon eines Transkript grundsätzlich keine Skip-Oligonukleotide gestaltet werden konnten. Jedem so erstellten Oligonukleotid wies das Skript eine Identifikationsnummer zu, über die automatisiert eine eindeutige Zuordnung zum Oligo-Typ (Skip/Junction) und dem Deletionsereignis (Gen, Exon) in einer Datentabelle möglich war.

Da, wie bereits angemerkt, einige Gene durch mehrere Transkripte in der Datenbank repräsentiert sind und die Transkript ID als Schlüssel zur Identifikation zusammengehöriger Exons verwendet wird, kommt es vor, dass identische Oligos für das gleiche Gen, basierend auf unterschiedlichen Transkripten, gebildet werden. Das Skript sucht daher den produzierten Oligodatensatz nach Redundanzen ab und entfernt ein doppeltes Oligo gleicher Gene, damit keine Oligopositionen auf dem Array durch unnötige Redundanzen belegt werden. Dieser Weg mag umständlich erscheinen, doch garantiert dieser Weg über die Transkripte, dass alle bekannten Varianten alternativen Spleißens eines Gens berücksichtigt werden. Nicht alle Gene weisen ein Standard-Transkript (full-length Transkript) auf in dem sämtliche Exons vereint sind. Unter Verwendung der ENSEMBL 41 Datenbank wurden so beispielsweise 37.219 Skip-Oligonukleotide generiert. Die fehlerfreie Funktion des Skriptes wurde durch manuelle Kontrolle der erstellten Sequenzen von 50 zufällig ausgewählten Oligonukleotiden bestätigt.

---

Dieses Skript bildet auch das Basisgerüst für spätere Weiterentwicklungen (z.B. Analyse der Exonsequenzen nach repetitiven Sequenzen), die an entsprechender Stelle im Detail erläutert werden.

### 3.3.3 Technische Vorüberlegungen zum Microarray und zum Array Design

Eine wichtige Vorüberlegung war die zur technischen Realisierung der Detektion mittels Microarrays. Prinzipiell waren basierend auf 3.3.1 drei Detektionsvarianten denkbar: Die Detektion mittels Skip-Oligos (Variante 1), mittels Junction-Oligos (Variante 2) oder mittels einer Kombination beider Varianten. Darauf aufbauend war die maximale Anzahl Oligonukleotid-Sonden pro Array von großer Bedeutung, da sie die Anzahl parallel zu betrachtender Deletionsereignisse beeinflusste. Die Anzahl erforderlicher Skip-Oligonukleotide wurde im Vorfeld durch Verwendung des Perl Skriptes (3.3.2) auf ca. 38.000 berechnet, die der Junction-Oligos auf ca. 63.000.

Als technische Plattform für die Untersuchungen wurden die Custom Microarrays der Firma Agilent gewählt. Hier standen zu Beginn der Arbeit drei Einheiten mit Platz für 1 x 128.000, 2 x 65.000 bzw. 4 x 44.000 Nukleotid-Sonden zu Verfügung. Für jede der drei Detektionsvarianten war ein Microarray-Typ denkbar. Diese Oligonukleotid Microarrays boten zudem entscheidende Vorteile gegenüber Konkurrenz-Produkten. Zum einen wurde ein, bezogen auf die Oligonukleotidlänge (25-60 bp), frei zu gestaltender Array mit bereits vorhandener ‚Infrastruktur‘ geboten: Der Array beinhaltete Kontroll-Oligonukleotide zur Ermittlung der Hintergrundsignale und bot die Möglichkeit über mit-hybridisierte synthetische aRNA (‘spike-in’ Kontrollen) die dynamische Breite der Signalintensitäten zu erfassen. Die Auswertung der Kontrollen erfolgte automatisiert über die entsprechende Software. Dies vereinfachte somit auch den direkten Vergleich unabhängiger Hybridisierungsereignisse. Weiterhin stand mit der Web-basierten eArray Plattform (2.4.1) eine Oberfläche zur Verfügung, mit der der Nutzer seine Oligonukleotide auf dem Array automatisiert platzieren konnte. Ausschlaggebend war letztendlich auch ein entsprechend gutes Preis-Leistungs-Verhältnis und die einfache Handbarkeit der Arrays, beides Eigenschaften, die mit Hinblick auf künftig großangelegte Screening Experimente nicht außer Acht gelassen werden durften.

Eines der Haupt-Ziele dieser Arbeit bestand nun darin, diese Design-Varianten hinsichtlich ihrer praktischen Verwendbarkeit zu bewerten, um darauf aufbauend ein System zur Detektion von Exon-Deletionen mittels Micro-Arrays zu entwickeln, zu testen und zu optimieren. Die Entwicklung war technisch gesehen so geplant, dass aufbauend auf den Ergebnissen einfacher Array-Designs Machbarkeit und Schwachpunkte bzw. Stärken

---

analysiert und bewertet werden sollten, um diese Ergebnisse in das Design nachfolgender Arrays aufzunehmen.

### 3.3.4 Machbarkeitsstudie zur Detektion von Exon-Deletionen

Das 'Proof of Principle' Experiment soll zeigen, dass die im vorangegangenen Abschnitt theoretisch beschriebenen Oligonukleotid Zusammenstellungen der Varianten 1 und 2 bzw. eine Kombination beider Varianten zur Detektion von Exon-Deletionen mittels Oligonukleotid-Microarrays genutzt werden können.

Als biologisches Test-Material wurden das Transkriptom (RNA) der ENU Mutante RR15 und als Kontroll-Material das des F1 Wildtyps verwendet. Das Transkriptom der Mutante trägt eine Deletion des Exon 8 des *Hprt* Transkripts und eignete sich daher hervorragend als Positiv-Kontrolle für dieses Experiment. Es wurde hier bewusst kein synthetisches Konstrukt als Kontrolle verwendet, da das letztendliche Ziel in der Detektion von Deletionen aus einem Gemisch von RNA Molekülen, dem Transkriptom, bestand.

Den Hauptbestandteil des Experiments bildete die Planung und Durchführung des Oligo-Designs. Wie bereits erwähnt, war das Hauptziel zu testen, ob mit den theoretisch erarbeiteten Varianten eine Detektion von Deletionen, hier im Speziellen die Detektion einer Deletion des Exon 8 des *Hprt*-Transkripts, möglich ist. Darauf aufbauend sollte dann bewertet werden, welche der Varianten besser geeignet scheint bzw. ob eine Kombination beider Varianten größeren Erfolg verspricht. Nebenbei sollte bei erfolgreicher Detektion untersucht werden inwieweit die Länge der Oligonukleotide den Erfolg einer Detektion beeinflusst. Aus Publikationen (Castle et al., 2003; Chou et al., 2004; He et al., 2005) war bekannt, dass die Länge Einfluss auf Spezifität und Sensitivität nimmt. Je länger das Oligonukleotid desto höher die Signalintensität und damit die Aussicht auch schwach exprimierte Gene analysieren zu können. Gleichzeitig steigt mit der Länge des Oligonukleotids aber auch die Wahrscheinlichkeit für unspezifische Kreuz-Hybridisierungen mit komplementären Fragmenten anderer Transkripte. Es wurden daher Skip- und Junction-Oligos in einem von Agilent synthetisierbaren Bereich zwischen 30 und 60 Basenpaaren vorgesehen. Tabelle 12 gibt einen Überblick über die Oligonukleotid-Zusammenstellung des 'Proof-of-Principle' Arrays.

Neben den funktionell zu bewertenden Skip- und Junction-Oligos des *Hprt* Transkriptes für Exon 8 enthielt die Zusammenstellung alle anderen Skip- und Junction-Oligos des *Hprt*-Transkriptes. Diese sollten dazu dienen, den Einfluss der Sequenz auf das Hybridisierungsverhalten analysieren zu können. Als externe Negativ-Kontrolle, neben den bereits von Agilent fest vorgegebenen internen Kontrollen des Arrays, wurde eine Sequenz

aus Intron 7 des *Hprt* Transkriptes zufällig ausgewählt. Das Vorkommen der Sequenz konnte durch einen Blast (2.4) gegen die genomische Sequenz der Maus nur an entsprechender Stelle im Genom nachgewiesen werden. Weiterhin wurden keine Treffer in einem Blast (2.4) gegen die Maus EST-Datenbank erhalten.

**Tabelle 12: Oligonukleotid Zusammenstellung des Proof-of-Principle Microarrays**

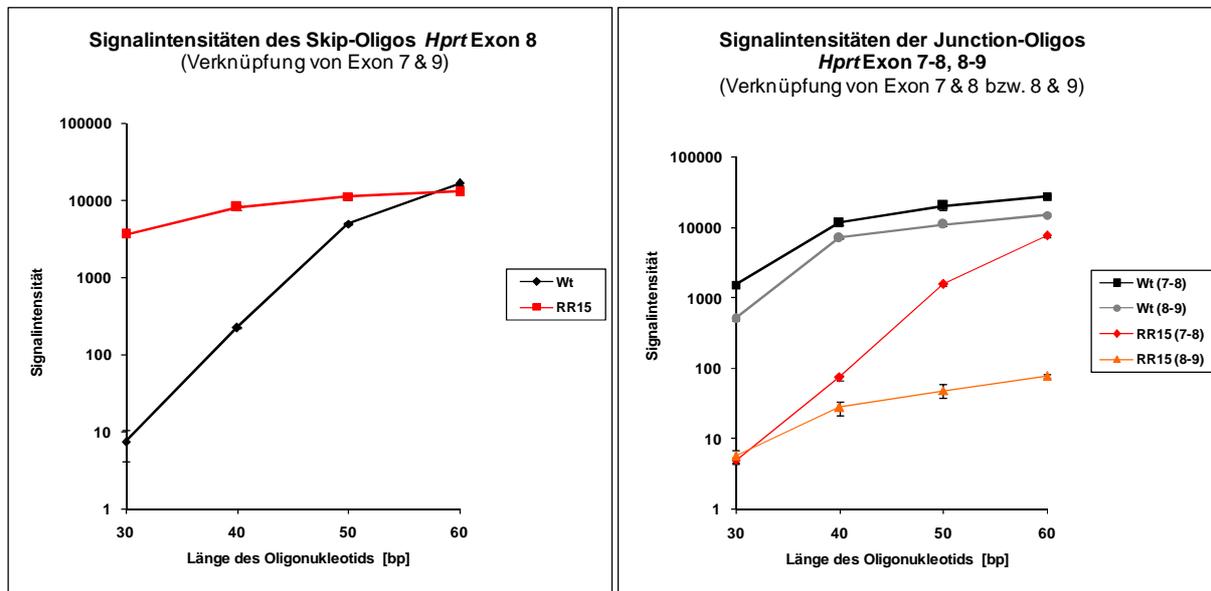
	Verknüpfung [Exon Nr.] / Sequenz	Anzahl verschiedener Oligos	Replikate	Längen [bp]
Skip-Oligos [Hprt]	1-3, 2-4, ..., 7-9	7	4	30, 40, 50, 60
Junction-Oligo [Hprt]	1-2, 2-3, ..., 8-9	8	4	30, 40, 50, 60
Negativ-Kontrolle [Hprt]	Intron 7 Base #1 bis Base #x (je nach Länge)	1	4	30, 40, 50, 60
„Auffüll Skip-oligos“	Zufällig ausgewählte Skip-Oligos exprimierter Gene	14515	1	30, 40, 60

Da der Array mit diesen 256 Oligonukleotiden (Replikate eingeschlossen) noch nicht einmal zu 1% gefüllt war, wurden die freien Positionen mit einem Set von 14.515 Skip-Oligos in drei verschiedenen Längen (30, 40, 50 bp) aufgefüllt. Die Oligonukleotide dafür wurden zufällig aus einem Set ausgewählt, welches im Vorfeld unter Verwendung des unter 3.3.2 beschriebenen Perl-Skripts generiert worden war (Datenbank ENSEMBL 41). Dieses Set von „Auffüll Skip-Oligos“ diente dem Zweck auf dem Array ein „realistisches“ Mileu zu schaffen, welches in künftigen Hybridisierungen Standard sein könnte.

### 3.3.4.1 Kontrolle der technischen Funktionalität des Hybridisierungsexperiments

Die Daten der ersten Hybridisierung wurden zunächst mit der Feature Extraction Software von Agilent normalisiert und ausgewertet (2.3.8.2). Die technischen Hybridisierungsbedingungen des Custom-Arrays waren korrekt, da die Rohdaten alle Tests passierten (Hintergrundsingale, Negativkontrollen, Positiv-Kontrollen, Rasterung). Weiterhin entsprach die mittels 'Spike-In' Kontrollen (Agilent) ermittelte dynamische Breite für einen linearen Zusammenhang von Signalintensität zu Probenkonzentration den Vorgaben Agilents (hier Signalintensitäten von  $10^{1,4}$  bis  $10^{5,3}$ ). Das bestätigte den Erfolg der vorangegangenen RNA Amplifizierung samt Labeling-Reaktion, sodass man davon ausgehen konnte, dass mögliche Problemursachen nicht in technischen Details zu suchen sind. Die Signalintensitäten der vier Oligoreplikate mit der Sequenz aus *Hprt* Intron 7, die als externe Negativ-Kontrollen fungierten, lagen auch deutlich im Bereich des Hintergrunds (ca.  $< 82$ ), während in der Wildtyp-Probe mit allen *Hprt* Exon-Junction-Oligos Signalintensitäten deutlich über dem

Hintergrund erzielt wurden. Diese Ergebnisse, die nicht mit Standard-Agilent Oligonukleotiden erhalten wurden, zeigten, dass die von Agilent ausgeführten Synthesen der Oligonukleotide funktionierten und die Zuordnung der einzelnen Signale im richtigen Raster erfolgt war (Negativ-Kontrollen).



**Abbildung 12: Signalintensitäten im Proof of Principle Experiment (Detektion von *Hprt* Deletion Exon 8)**

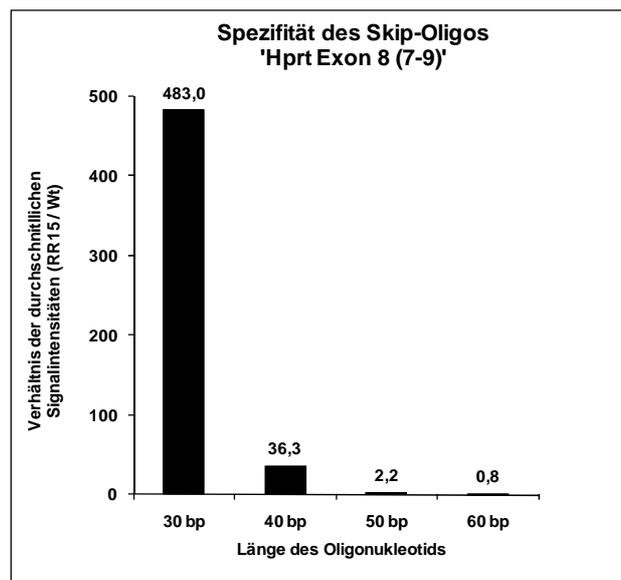
Dargestellt sind die Mittelwerte ( $n=4$ ) der mit einem Agilent Custom Array (4x40k) aufgezeichneten normalisierten Signalintensitäten für Oligonukleotide der beiden Detektionsvarianten in Abhängigkeit von Oligonukleotidlänge [bp] und hybridisierter Proben aRNA. Die Standardabweichung der Mittelwerte wird durch senkrechte Linien an den Messpunkten wiedergegeben. A = Detektionsvariante 1, Verwendung von Skip-Oligos. B = Detektionsvariante 2, Verwendung von Junction-Oligos. Wt = hybridisierte aRNA synthetisiert aus Wildtyp F1 RNA. RR15 = hybridisierte aRNA synthetisiert aus dem *Hprt*-Klon RR15 mit Deletion von Exon 8.

Nach erfolgreicher Bestätigung aller Kontrollen und damit der technischen Funktionalität des Arrays und der Hybridisierung wurden die Signalintensitäten der für die Detektion der Deletion des *Hprt* Exon 8 nötigen Oligonukleotide (Skip-Oligo: ‚Exon 8 (7-9)‘; Junction-Oligo: ‚Exon 7-8‘ und Exon ‚8-9‘) ausgewertet. Die Signalintensitäten sind graphisch dargestellt in Abbildung 12.

### 3.3.4.2 Detektion mittels Skip-Oligonukleotiden (Variante 1)

Mit dem Skip-Oligonukleotid *Hprt* Exon 8 lassen sich für alle Nukleotidlängen mit der Mutante RR15 deutliche Signale mit Intensitäten zwischen 3625 und 13074 erzielen, die in Abhängigkeit von der Oligonukleotidlänge zunehmen. Im Vergleich dazu liegt das Signal in der Wildtyp-Probe bei 30 bp langen Nukleotiden mit 7,51 noch deutlich im Intensitätsbereich des Hintergrundsignals, steigt dann bei Längen von 40 bp und 50 bp schnell über den Hintergrund hinaus an und erreicht bei 60 bp-langen Fragmenten bereits eine Signalstärke von 16616, vergleichbar mit der der Mutante.

Für die Signalintensitäten der 30 bp langen Fragmente kann man daher feststellen, dass dieses Ergebnis vollständig das aus den theoretischen Vorüberlegungen zur Detektionsvariante 1 erwartete Ergebnis widerspiegelt: Ein Signal kann nur mit der Probe RR15 und nicht mit dem Wildtyp erreicht werden, da die Verknüpfung von Exon 7 und Exon 9 nur in dem *Hprt* Transkript der Mutante und nicht im Wildtyp vorkommt. Es konnte somit gezeigt werden, dass sich die theoretischen Überlegungen zur Detektion von Exon-Deletionen mit Skip-Oligonukleotiden praktisch umsetzen lassen.



**Abbildung 13: Spezifität des Skip-Oligos ‚*Hprt* Exon 8‘ in Abhängigkeit von der Oligonukleotidlänge**

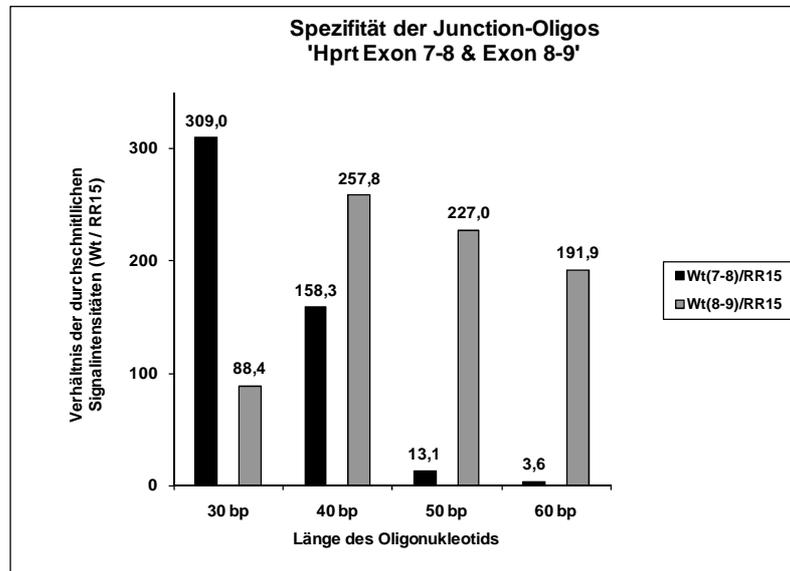
Dargestellt sind die Quotienten der Mittelwerte ( $n=4$ ) der mit einem Agilent Custom Array (4x40k) aufgezeichneten, normalisierten Signalintensitäten der aRNA Probe RR15 zur Probe Wildtyp für Oligonukleotide der Detektionsvariante 1 in Abhängigkeit von Oligonukleotidlänge [bp].

Mit steigender Oligonukleotidlänge nähern sich die Signalstärken von Mutante und Wildtyp allerdings so sehr an, dass keine Spezifität des Oligonukleotids mehr zu erkennen ist. Eine Detektion der Deletion ist mit Oligonukleotiden von 60 bp Länge nicht mehr möglich. Dies zeigt sich sehr gut in Abbildung 13 in der die Verhältnisse der Signalintensitäten von Mutante zu Wildtyp in Abhängigkeit der Oligonukleotidlängen dargestellt werden. Die Spezifität der Skip-Oligonukleotide nimmt mit zunehmender Länge exponentiell ab.

### 3.3.4.3 Detektion mittels Junction-Oligonukleotiden (Variante 2)

Die Detektion der Exon-Deletion mittels Junction-Oligos sollte basierend auf der geringeren Signalintensität der entsprechenden Oligonukleotide in der Mutanten-Probe im Vergleich zur Wildtypprobe erfolgen (autosomal lokalisierte Gene theoretisch 50% des Wildtyp Signals im linearen Signalbereich). Grundsätzlich lässt sich feststellen, dass die erreichten Signalstärken in der Wildtypprobe ausserhalb des Hintergrundsignal-Bereichs liegen und mit steigender

Oligonukleotidlänge erwartungsgemäß zunehmen. Die Funktionalität der aufgestellten Oligonukleotide ist damit belegt. Zusätzlich fiel auf, dass die Intensitäten der Oligonukleotide für das Junction-Oligo 7-8 durchweg über denen des Junction-Oligos 8-9 lagen. Dies kann in den unterschiedlichen GC-Gehalten begründet sein: 30mer 7-8 GC = 43%, 30mer 8-9 GC = 40%; 60mer 7-8 GC = 43%, 60mer 8-9 GC = 33%).



**Abbildung 14: Spezifität der Junction-Oligos ‚Hprt Exon 7-8 und Exon 8-9‘ in Abhängigkeit von der Oligonukleotidlänge**

Dargestellt sind die Quotienten der Mittelwerte ( $n=4$ ) der mit einem Agilent Custom Array (4x40k) aufgezeichneten normalisierten Signalintensitäten der aRNA Probe Wildtyp zur Probe RR15 für Oligonukleotide der Detektionsvariante 2 in Abhängigkeit von der Oligonukleotidlänge [bp].

Die Signalstärken, die für die beiden Oligonukleotidtypen mit der Probe der Mutante RR15 erreicht werden, befinden sich bei einer Oligonukleotidlänge von 30 bp im Bereich des Hintergrundsignals und steigen mit zunehmender Nukleotidlänge deutlich an. Für die Verbindung der Exons 7 und 8 werden bei einer Basenzahl von 60 beispielsweise nahezu die Intensitäten des Wildtyp Signals erreicht. Analog zu der Beobachtung, die mit dem Skip-Oligo mit der Wildtyp-RNA Probe gemacht worden sind, nimmt auch hier mit zunehmender Länge des Oligonukleotids die Spezifität deutlich ab. Deutlich zu erkennen ist diese Beobachtung in der graphischen Darstellung des Verhältnisses der Signalintensitäten von Wildtyp- zu Mutanten-Signal (Abbildung 14). Dass diese Anfälligkeit gegenüber Cross-Hybridisierungen jedoch sequenzabhängig ist, zeigt sich daran, dass die Spezifität des Junction-Oligos (8-9) deutlich schwächer ausfällt wie die des Oligos (7-8). Unklar ist warum; eine Erklärung dafür, dass die Spezifität des Junction-Oligos 8-9 bei 30 bp geringer ausfällt als bei den anderen Längen kann in der Berechnung der Spezifität liegen. Hier ist der Quotient von WT- zur Mutanten-Durchschnittssignalintensität berechnet worden. Da die Signalintensitäten der 30 bp langen Oligonukleotide (Junction 7-8 und 8-9) im Hintergrund

und damit ausserhalb des linearen Bereichs liegen eignen sich diese Werte nicht zur Bildung eines Verhältnisses. Für die Ergebnisse sind daher die absoluten Intensitäten ausschlaggebend.

#### **3.3.4.4 Zusammenfassende Auswertung des Proof of Principle Experiments**

Die Ergebnisse des Experiments zeigen anschaulich, dass eine Detektion von Exon-Deletionen mit beiden Detektionsvarianten möglich ist. Die Spezifität der Detektion ist umso höher je kürzer das verwendete Oligonukleotid auf dem Microarray ist. Die besten Ergebnisse wurden bei beiden Varianten mit Oligonukleotiden mit Längen von 30 bp erreicht. Oligonukleotidlängen ab 50 bp lieferten für beide Varianten unbefriedigende Ergebnisse. Höhere Signalintensitäten konnten jedoch mit längeren Nukleotiden erzielt werden, was von Vorteil bei Detektionen von Deletionen in schwach exprimierten Genen sein kann. Spezifität und Sensitivität stehen daher in beiden Varianten in indirektem Zusammenhang.

Im weiteren Verlauf der Arbeit sollte die Detektion basierend auf Skip-Oligonukleotiden (Detektionsvariante 1) weiter etabliert und getestet werden. Diese Entscheidung ist darin begründet, dass sich diese Methode rein auf die Detektion eines Signals in der Mutante beschränkt und ein eindeutiges Signal in Bezug auf das deletierte Exon liefert. Im Falle der Variante 2 sind für den eindeutigen Nachweis einer Deletion vergleichsweise 2 Sonden nötig. Anhand einer Sonde kann nicht entschieden werden, ob das stromauf- oder das stromabwärts gelegene Exon im Transkript fehlt (3.3.1). Daher ist die Variante 1 auch aus wirtschaftlichen Gesichtspunkten zu befürworten.

Weiterhin wurde die Oligonukleotidlänge von 30 bp zunächst als Standard für Folgeexperimente ausgewählt, da damit die spezifischsten Ergebnisse zu erwarten waren. Ob die mit dieser Länge erreichbaren Signalintensitäten auch bei schwach exprimierten Genen zur Detektion ausreichen, gilt es zu analysieren. Die Ergebnisse basieren auf dem stark exprimierten *Hprt* Transkript, welches nach den Expressionsanalysen in der Zelllinie F1 (3.2.1) an Position 476 der ‚Expressions-Hitliste‘ steht.

#### **3.3.5 Pooling-Experimente mit Positiv-Kontrollen und ENU-mutierten ES-Zell-Klonen**

Ausgehend von den in Abschnitt 3.3.4 beschriebenen Ergebnissen wurde ein neue Microarray-Struktur entwickelt, die die Detektion von Exon-Deletionen mittels 30 bp langen Skip-Oligonukleotiden ermöglichen sollte. Mit diesem Aufbau sollten folgende Analysen durchgeführt werden, die der Weiterentwicklung des Microarray-Designs dienen sollten:

- Ermittlung falsch-positiver Signale anhand von Wildtyp-Proben
- Bestätigung der Funktion des Micorarrays durch Detektion der Deletion von *Hprt* Exon 8 in der Positiv-Kontrolle RR15 (Reproduktion der Machbarkeitsstudie; 3.3.4)
- Test verschiedener Pooling-Verhältnisse anhand der Detektion bekannter Deletionen des *Hprt* Transkripts zweier Positiv-Kontroll-Klone (RR15:  $\Delta$ Exon8; BG15:  $\Delta$ Exon6) mit Hinblick auf die Kostenreduktion künftig angelegter Screenings (vergl. 3.1.3)

Die *in-silico* Synthese der notwendigen Oligonukleotide erfolgte mit der unter 3.3.2 entwickelten Software. Unter Verwendung der ENSEMBL 41 Maus Exon Datenbank und der unter 3.2.2 ermittelten 7475 Genen, die in beiden WT-Proben mit 99%-iger Wahrscheinlichkeit als exprimiert angesehen werden können, wurden Skip-Oligonukleotide für 37.219 Exon-Deletionen entworfen. Die Deletionen führen nicht zu einer Unterbrechung des ORF. Für weitere Analysen und zur Kontrolle enthielt der Array zudem sämtliche *Hprt* Skip-Oligonukleotide, die *Hprt* Junction-Oligonukleotide und die Oligonukleotid-Sonde komplementär zu der bereits beschriebenen Sequenz aus Intron 7 des *Hprt*-Gens (vergl. 3.3.4). Die Zusammenstellung der Oligonukleotide auf dem Microarray ist mit Ausnahme der von Agilent vorgegebenen Oligos in Tabelle 13 wiedergegeben.

**Tabelle 13: Oligonukleotid Zusammenstellung des Pooling-Experiments**

	Verknüpfung [Exon Nr.] / Sequenz	Anzahl verschiedener Oligos	Replikate	Länge [bp]
Skip-Oligos [Hprt]	1-3, 2-4, ..., 7-9	7	5	30
Junction-Oligo [Hprt]	1-2, 2-3, ..., 8-9	8	5	30
Negativ-Kontrolle [Hprt]	Intron 7 Base #1 bis Base #30	1	5	30
Screening Skip-Oligos	Alle Skip-Oligos der in der F1-Linie exprimierter Gene (7407)	37219	1	30

Die Vorbereitung der RNA-Proben für die Hybridisierung erfolgte wie unter Abschnitt 2.3.8.2 beschrieben. Für die Analyse zusammengestellter RNA-Proben verschiedener Mutanten wurde die Gesamt-RNA aller Mutanten vor der aRNA-Synthese anteilmäßig zusammengeführt. Einen Überblick über die hybridisierten aRNA-Proben gibt Tabelle 14.

Analog zu 3.3.4.1 wurden zunächst die Ergebnisse mit der Feature Extraction Software ausgewertet und normalisiert. Die technische Funktionalität der Hybridisierungen konnte bestätigt werden. Hier erfüllten alle Proben die von Agilent vorgegebenen Parameter. Auch die Signalintensitäten für die Negativ-Kontrolle (Sequenz aus Intron 7 *Hprt*) lagen im Bereich des Hintergrundsignals.

**Tabelle 14: Übersicht über die hybridisierten RNA-Proben/-Pools**

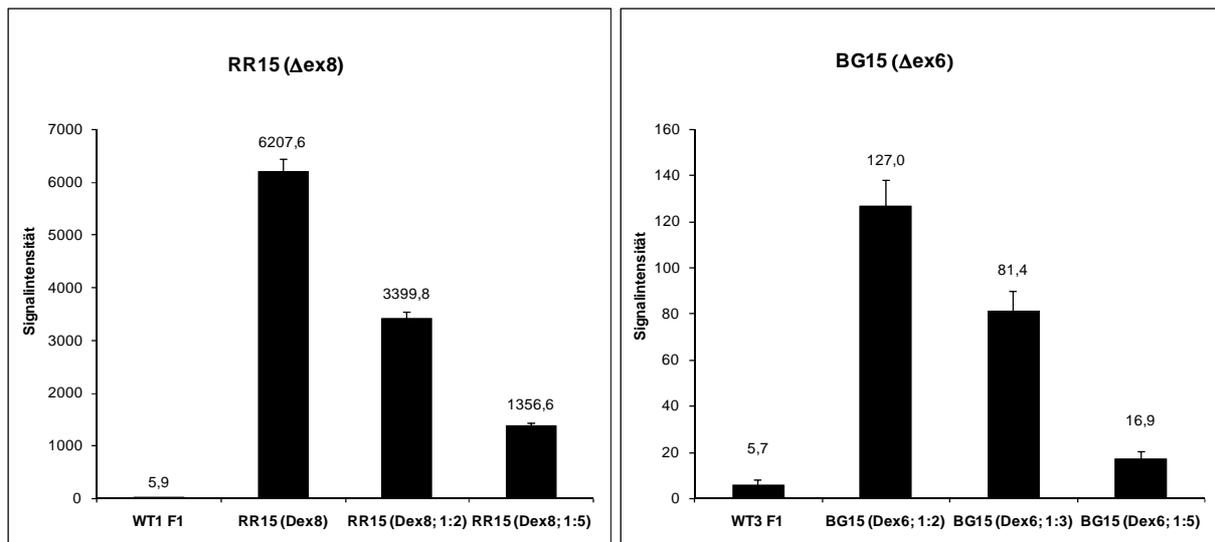
Wt<sub>#</sub> sind biologische Wildtyp Replikate; RR15 = Positiv-Kontroll RNA der Mutante RR15 mit Deletion von Hprt Exon 8; BG15 = Positiv-Kontroll RNA der Mutante BG15 mit Deletion von Hprt Exon 6; RRx### RNA aus ENU mutierten ES Zellklonen der Mutagenese F1-1 (3.1.1)

Microarray [#Objekträger.#Feld]	Klon-ID	Verdünnungsverhältnis einer Gesamt-RNA in der Probe
1.1	Wt <sub>1</sub> F1	Unverdünnt
1.2	RR15	Unverdünnt
1.3	RR15 + RRx004	1:2
1.4	Wt <sub>2</sub> F1	Unverdünnt
2.1	Wt <sub>3</sub> F1	Unverdünnt
2.2	BG15 + RRx004	1:2
2.3	BG15 + RRx005 + RRx006	1:3
2.4	RR15 + BG15 + RRx004 + RRx005 + RRx006	1:5

Um falsch-positive Signale (z.B. bedingt durch alternatives Spleissen) auszuschließen, wurden die Ergebnisse der drei unabhängigen Wildtyp Proben analysiert. Es wurden zunächst solche Sonden ermittelt, die Signalintensitäten von mehr als dem dreifachen des Hintergrundsignals aufwiesen. Diese Grenze wurde frei ausgewählt, da keine Angaben zu ähnlichen Untersuchungen vorlagen. Dadurch sollte garantiert werden, dass nicht solche Sonden als Falsch-Positiv eingestuft werden, die grundlegend schwache Signal aufgrund von leichten Cross-Hybridisierungen mit beispielsweise repetitiven Elementen erzeugen.

Anhand der Wildtyp Proben konnten so insgesamt 1023 der 37219 Oligonukleotid-Sonden ermittelt werden. Auch in den Proben der ENU-Mutanten gaben diese 1023 Sonden entsprechend deutliche Signale. Da es sich somit um falsch-positive Signale handelte, die nicht auf einer Exon-Deletion basierten, wurden diese bei der weiteren Analyse und Weiterentwicklungen nicht berücksichtigt. Eine Blast-Analyse der 1023 Sequenzen gegen dbEST ergab 100%-ige Treffer für 54 Sequenzen. Bei diesen Sequenzen handelte es sich daher um Spleiß-Alternativen. Weiterhin wies ein Großteil der restlichen Sequenzen (74%) einen GC-Gehalt größer 60% auf.

Zusätzlich konnte die Funktion des Array-Designs dadurch bestätigt werden, dass eine eindeutige Detektion der Deletion von *Hprt* Exon 8 aus der Mutante RR15 möglich war. Abbildung 15 zeigt die mit dem entsprechenden Oligonukleotid erzielten Signalintensitäten für Wt<sub>1</sub> F1 (5,9) und Mutante RR15 (6207,6). Entsprechend gute Intensitäten zeigte diese Skip *Hprt* Exon 8 Sonde auch bei der Hybridisierung gepoolter Proben (vergl. Tabelle 14). Die Signalintensität nimmt nahezu linear mit dem Verdünnungsverhältnis ab. Eine Detektion dieser Mutation aus einem 1:5 Pool war problemlos möglich.



**Abbildung 15: Signalintensitäten ausgewählter Sonden im Pooling Experiment**

Dargestellt sind die Mittelwerte ( $n=5$ ) der mit einem Agilent Custom Array (4x40k) aufgezeichneten normalisierten Signalintensitäten, die mit den Oligonukleotiden Skip Hprt Exon 8 (7-9) und Skip Hprt Exon 6 (5-7) in unterschiedlichen Proben (aRNA) erzielt wurden. Die analysierten Proben entsprechen den Ansätzen aus Tabelle 14.

Anders verhält sich die Detektion der *Hprt* Exon 6 Deletion über die entsprechende Skip-Sonde (Exon 5-7). Hier beträgt die Signalintensität in einer 1:2 Verdünnung mit 127,0 nur 3,7% der der oben genannten Sonde. Bei einer 1:5 Verdünnung liegt das Signal mit einer Intensität von 16,9 deutlich im Bereich des Hintergrunds (ca.  $< 55$ ). Im Vergleich zum Wildtyp-Signal der Sonde ist zwar ein Unterschied zu erkennen, doch ist fraglich, ob eine unbekannte Mutation aufgrund eines derart geringen Verhältnisses eindeutig zu erkennen wäre. Die geringe Signalintensität lässt sich vermutlich durch den vergleichsweise geringen GC-Gehalt der Sonde von nur 30% begründen. Die Sonde zur Detektion der Deletion von Exon 8 weist einen GC-Gehalt von 47% auf. Der GC-Gehalt der Sonde sollte daher bei der Bewertung möglicher Signale nicht ausser Acht gelassen werden.

Auf Basis dieser Beobachtung wurden hinsichtlich der Stringenz zwei Hybridisierungen bei 55°C und 60°C anstelle von 65°C durchgeführt, da geringere Hybridisierungstemperaturen auch von der Fa. ExonHit Therapeutics (Fehlbaum et al., 2005) bei kürzeren Sonden erfolgreich angewendet wurden. Doch ließen sich die bei diesen Temperaturen hybridisierten Arrays nicht auswerten. Zum einen konnte das Raster nicht richtig platziert werden, zum anderen machten unregelmäßig verteilte intensive Hintergrundsignale eine Auswertung unmöglich.

### 3.3.6 Auswertung der ersten Pooling-Experimente hinsichtlich neuer Exon-Deletionen

Die unter 3.3.5 nach allgemeinen Fragestellungen analysierten Proben wurden im Anschluss hinsichtlich der Detektion bisher unbekannter Exon-Deletionen untersucht. Hierzu wurde ein Excel Programm verwendet, welches ausgehend von den normalisierten Signalintensitäten zunächst den Quotient des Signals aus ENU mutierter Probe zu dem durchschnittlichen Signal der drei Wildtyp Proben der jeweiligen Sonde ermittelte. Unter Berücksichtigung der Verdünnung der ENU-Probe war ein Quotienten größer 1 damit ein Hinweis auf eine mögliche Exon-Deletion. Konsequenterweise hätten dann knapp 10% aller Oligonukleotid-Sonden weiter betrachtet werden müssen. Daher wurde hier eine weitere Schwelle eingeführt, die nur noch Quotienten größer 5 betrachtete. Diese Schwelle war frei gewählt, hätte aber beispielsweise eine Detektion der *Hprt* Exon 6 Deletion (Abbildung 15) zweifelsfrei erkennen lassen. Die Anzahl putativer Mutationen ließ sich dadurch erneut um den Faktor 10 reduzieren, d.h. durchschnittlich waren immer noch ca. 370 putative Exon-Deletionen pro Probe zu betrachten. Diese Zahl schien viel zu hoch und passte nicht annähernd zu der unter 3.1.3 beschriebenen Schätzung von ca. einer Exon-Deletion in 2,23 Klonen. Da jede ENU-Mutante in mindestens zwei Proben enthalten war, wurde das Programm dahingehend erweitert, dass ein Abgleich der Ergebnisse der Proben erfolgte und nur solche Exon-Skip Sonden selektiert wurden, die in beiden Proben entsprechend hohe Quotienten aufwiesen. Diese Filterung reduzierte die Anzahl möglicher neuer Mutationen auf 10 für die 5 analysierten ENU-Mutanten (Tabelle 14).

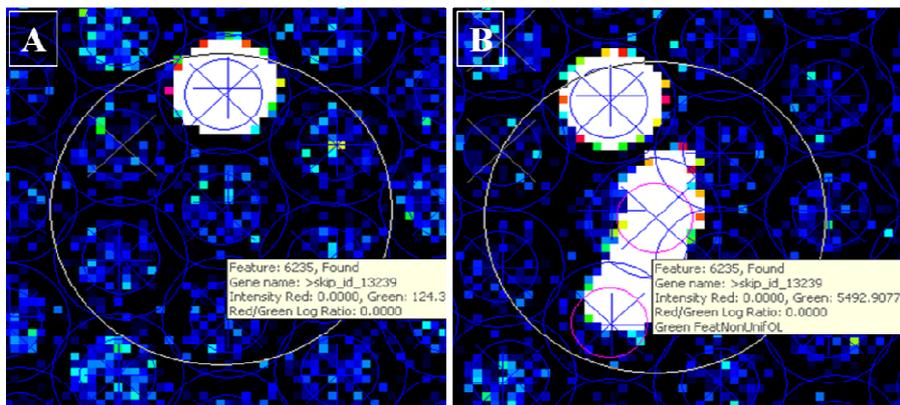
**Tabelle 15: Informationen zu putativen Exon-Deletionen**

MGI Symbol Kennungen, die mit einem Asteriks (\*) versehen sind, weisen auf das Vorhandensein bekannter Spleiß-Varianten für dieses Gen hin. Für die Mutanten (Mut) RRx005 und RRx006, hier mit 005 und 006 abgekürzt, war aufgrund des ausschließlich gemeinsamen Vorkommens in den Proben (Tabelle 14) keine eindeutige Zuordnung möglich. Die Signalintensitäten (Signalint.) sind jeweils für das niedrigste Poolingverhältnis (P) angegeben. Signalintensitäten des Wildtyps (Wt) sind das arithmetische Mittel aus drei biologischen Replikaten. Der Quotient (Q) berechnet sich wie folgt: Mut/Wt P

Sonden Nr.	Sonden Name	MGI Symbol	Ensembl Transkript ID	Exon Nr.	Signalint.		P	Q	Mutante
					Wt	Mut			
12507	skip_id_04399	Arnt*	ENSMUST00000015666	16	6,0	92,8	1:3	46,4	005+006
14032	skip_id_07731	Mgea6*	ENSMUST00000043421	17	6,0	181,8	1	30,3	RR15
25511	skip_id_06458	Stk10	ENSMUST00000020512	9	7,7	140,9	1	18,3	RR15
28085	skip_id_19803	Phf14*	ENSMUST00000090632	8	5,5	220,8	1:2	80,3	BG15
31117	skip_id_15875	Hat1*	ENSMUST00000028408	2	6,5	140,0	1	21,5	RR15
33902	skip_id_29462	Bub1b	ENSMUST00000038341	3	20,3	293,4	1:3	43,4	005+006
35035	skip_id_09425	Acin1*	ENSMUST00000022793	7	26,8	307,4	1:3	34,4	005+006
43902	skip_id_26133	Nup37	ENSMUST00000052355	8	10,0	891,2	1:3	267,4	005+006
13410	hprt_skip_30mer_07	Hprt	ENSMUST00000026723	8	5,9	6207,6	1	1052,1	RR15
11847	hprt_skip_30mer_05	Hprt	ENSMUST00000026723	6	5,7	127,0	1:2	44,6	BG15

Zwei dieser Mutationen waren die beiden bekannten Deletionen von Exon 8 und 6 des *Hprt* Transkripts. Tabelle 15 gibt Informationen zu den möglichen Mutationen. Für 5 der 8 Gene existieren Spleiß Isoformen, von denen allerdings keine den putativen Exon-Deletionen entspricht. Die mit den spezifischen Sonden erzielten Signalintensitäten liegen z.T. nur knapp über dem Bereich des Hintergrundsignals und erreichen nicht annähernd die Stärke des *Hprt* Exon 8 spezifischen Oligonukleotids. Dennoch sind deutliche Unterschiede von Wildtyp zu Mutante zu erkennen, u.a. auch bei dem sehr schwach exprimierten Gen *Stk10* (nur ca. 4,5% der Signalintensität des *Hprt* Transkriptes nach Illumina Expressionsmessungen; 3.2).

Bei genauerer Betrachtung der großen Anzahl falsch-positiver Werte vor dem Abgleich der Proben fiel auf, dass ein Großteil der gebildeten Quotienten auf Signalintensitäten basierte, die deutlich im Bereich des Hintergrund lagen. Dies zeigte, dass Werte im Hintergrundbereich (i.d.R. <50) und ausserhalb des linearen Bereichs (i.d.R. < 25), insbesondere im Vergleich zwischen verschiedenen Array-Feldern, nicht mit Sicherheit verglichen werden können. Weiterhin wurden bei 5 bis 50 Oligonukleotiden pro Array sehr hohe Signalintensitäten festgestellt, die auf Hybridisierungsartefakten basierten.



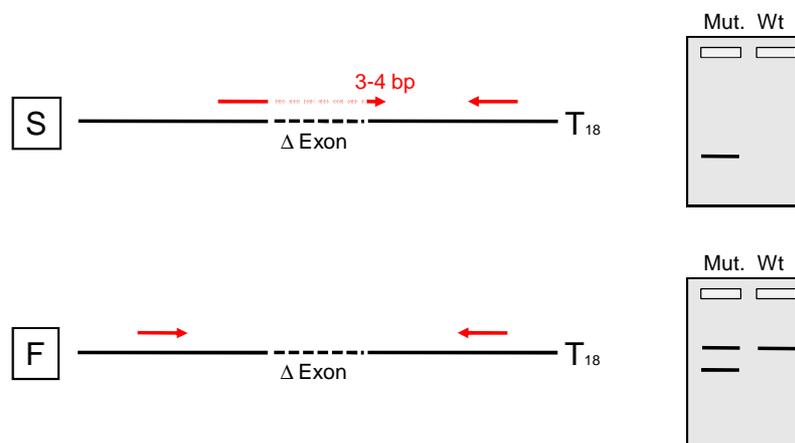
**Abbildung 16: Hybridisierungsartefakte**

Die Abbildung zeigt einen Weg der Entstehung falsch-positiver Signale aufgrund der Bildung von Hybridisierungsartefakten auf der Oberfläche des Microarrays. Die Aufnahme wurde erstellt mit der Feature Extraction Software und zeigt den Bereich um die Sonde Nr. 6235 (*skip\_id\_13239*). Blaue Kreise begrenzen Software basiert die einzelnen Positionen der Sonden. Die inneren blauen Kreise beschreiben den Bereich der in die Auswertung der Signalintensität einfließt. Die Signalintensität ist der besseren Anschaulichkeit eingefärbt und nimmt von weiß, über gelb, rot, grün, blau bis zu schwarz hin ab. A: Bereich um Sonde 6235 ohne Artefakt. B: Bereich um Sonde 6235 mit Artefakt. Dieses Artefakt wurde als solches von der Feature Extraction Software erkannt (vergl. Text im Bild).

Ein Beispiel für ein Artefakt gibt Abbildung 16. Basierend auf diesen Ergebnissen wurden in nachfolgenden Untersuchungen alle auffälligen Sonden, die von der Feature Extraction Software nicht automatisch als ‚nicht gleichmäßig‘ erkannt wurden, visuell überprüft.

### 3.3.7 PCR basierte Bestätigung möglicher Exon-Deletionen

Die Bestätigung möglicher Exon-Deletionen sollte mittels PCR auf Transkriptebene (cDNA) erfolgen. Dazu wurden zwei Primer-Systeme vorgesehen, die nachfolgend graphisch dargestellt sind (Abbildung 17). Analog zu den unter 3.3.1 beschriebenen Oligonukleotid-Sonden ist der Skip-Primer des Primersystems S aus Sequenzen der beiden Exons zusammengesetzt, die stromauf- (c) und stromabwärts (a) des deletierten Exons (b) liegen. Abweichend von den Skip Sonden sind nur die letzten drei bis vier Basen seines 3'-Endes komplementär zu dem 5'-Ende des Exons a. Im Falle einer Deletion von Exon b ist dieser Primer vollständig komplementär zu der Sequenz c-a, sodass ausgehend von diesem 3'-Ende eine Synthese stattfinden kann.



**Abbildung 17: Schematische Darstellung der PCR Systeme zur Bestätigung von Exon-Deletionen**

Die Abbildung zeigt schematisch die im Text beschriebene Vorgehensweise bei der Bestätigung von Exon-Deletionen mittels PCR. Die schwarzen Linien verkörpern das cDNA Molekül eines bestimmten Gens. Die gestrichelte schwarze Linie deutet eine Deletion der entsprechenden Exon-Sequenz an. Die roten Pfeile symbolisieren die Position der PCR Primer (Pfeil = 3'-Ende). Die grauen Rechtecke stehen symbolisch für Ausschnitte aus Agarosegelen. Im Falle einer erwarteten Amplifikation eines Produktes in der PCR mit dem entsprechenden Primersystem sind schwarze Linien eingezeichnet, die die Ethidiumbromid vermittelte Fluoreszenz der Produkte symbolisieren sollen. S = Detektion basierend auf einem Skip-Primer (links) und einem entsprechenden Rückwärts-Primer (rechts). F = Detektion basierend auf zwei die Deletion flankierenden Primern. Mut. = PCR-Produkte eines mutierten und eines Wildtyp Allels. Wt. = PCR-Produkt zweier Wildtyp Allele.

Liegt keine Deletion vor, so kann dieser Primer zwar mit dem größten Teil seiner Sequenz an Exon c binden, eine Synthese ausgehend von seinem 3'-Ende erfolgt dennoch nicht, da die letzten drei bis vier Basen nicht komplementär sind. Entsprechend war aus Vorversuchen bekannt, dass eine Übereinstimmung von nur drei bis vier Basen am 3'-Ende des Primers nicht ausreichen ein Fragment zu amplifizieren (Daten nicht gezeigt.). In einer PCR ausgehend von einer cDNA Matrize wird daher nur ein Produkt erwartet, wenn das untersuchte Transkript eine entsprechende Exon-Deletion aufweist.

Der Nachteil dieser Skip-Primer basierten PCR besteht darin, dass die Freiheiten zur Gestaltung des Skip-Primers ( $T_m$ , repetitive Sequenzen, etc.) aufgrund der vordefinierten

Position sehr eingeschränkt sind. Eine Optimierung der PCR-Bedingungen auf diesen Primer ist im Vorfeld nicht ohne größere Vorarbeiten möglich, da kein bestätigtes Matrizenmaterial zur Verfügung steht.

Im Primersystem F flankieren die beiden PCR Primer das deletierte Exon. Liegt keine Deletion vor, so erhält man in einer PCR nur Produkte von der Größe des Wildtyp-Fragments. Im Falle einer Deletion sind verschiedene Amplifikationsmuster denkbar, da neben dem mutierten Allel auch das Wildtyp Allel existiert. Zum einen kann neben dem Wildtyp Produkt ein um die Größe des deletierten Exons kleineres Deletions Produkt gebildet werden. Zum anderen kann sich das kleinere Produkt im Zuge der Amplifikation durchsetzen oder erst gar nicht bzw. nicht ausreichend amplifiziert werden, falls das mutierte Transkript beispielsweise durch ‚RNA mediated decay‘ im Vorfeld abgebaut wurde.



#### Abbildung 18: Bestätigung selektierter Exon-Deletionen mittels PCR

Die Abbildung zeigt die Ethidiumbromid vermittelte Fluoreszenz von PCR-Produkten nach gelelektrophoretischer Auftrennung in einem 1,5%-igen Agarosegel. Es wurden jeweils 10% eines PCR-Ansatzes aufgetragen. F = Primersystem bestehend aus zwei die Deletion flankierenden Primern; S = Primersystem aus einem Skip-Primer und einem Rückwärts-Primer; M = Größenstandard ‚GeneRuler™ 100 bp Plus DNA Ladder‘, 500 ng; - = Negativ-Kontrolle (ohne Matrizen DNA); Wt = Wildtyp F1 ES-Zell cDNA Matrize; 5 = RRx005 cDNA Matrize; 6 = RRx006 cDNA Matrize; RR15/BG15 = cDNA der Klone als Matrize.

Diese möglichen Amplifikationsmuster stellen bei dieser Methode den Nachteil dar. Da beide Methoden entsprechende Nachteile aufweisen, wurde eine parallele Anwendung beider Systeme beschlossen. Eine elegante Methode wäre die Verwendung sogenannter Poison-Primer (Edgley et al., 2002) gewesen. Nach einigen Vorversuchen stellte sich allerdings heraus, dass diese Methode einiges an Optimierungsarbeit voraussetzt und sich daher eher für großangelegte Screenings nach bestimmten wenigen Exon-Deletionen eignet, als für die Bestätigung einzelner Klone. Für alle zu bestätigenden Exon-Deletionen (Tabelle 15) wurden die beiden Primersysteme S und F erstellt (Primersequenzen in Anhang). Die cDNA Synthese erfolgte ausgehend von 1 µg Gesamt-RNA der ES-Zell Klone nach 2.3.5. Die PCR wurde wie in Abschnitt 2.3.7 beschrieben durchgeführt unter Verwendung 1% des jeweiligen cDNA

Ansatzes als DNA-Matrize. Die PCR-Produkte wurden anschließend gelelektrophoretisch aufgetrennt.

Für die Transkripte der Gene *Arnt* und *Stk10* konnten mit beiden Primersystemen auch nach diversen PCR Optimierungen (Tm, Primer-Konzentrationen, Matrize) und Austausch der Primer keine Produkte amplifiziert werden (Daten nicht gezeigt.). Beide Gene werden in F1-ES Zellen schwach exprimiert (4,9% [*Arnt*] bzw. 4,5% [*Stk10*] der Signalintensität des *Hprt* Transkriptes nach Illumina Expressionsmessungen; 3.2). Bestätigt werden konnten die Exon-Deletionen in den Transkripten der Gene *Mgea6* ( $\Delta$  Exon 17, 36 bp) und *Nup37* ( $\Delta$  Exon 8, 51 bp) über die Skip-Primer Systeme (Abbildung 18). Der Klon RRx006 war Träger der Exon-Deletion in *Nup37*. Die Deletion in *Mgea6* war auch anhand des kleineren Produktes der PCR mit den flankierenden Primern zu erkennen. Für die Transkripte der Gene *Acin1*, *Bub1b*, *Hat1* und *Phf14* konnten Produkte mit dem flankierenden Primersystem amplifiziert werden, doch waren hier keine Abweichungen vom Wildtyp-Amplifikationsmuster zu erkennen. Mit dem Primersystem S konnten für diese Transkripte keine Produkte amplifiziert werden. Alle amplifizierten Produkte entsprachen den im Vorfeld ermittelten Größen. Die Abbildung 18 zeigt repräsentativ für diese vier Gene das Amplifikationsmuster des *Phf14* Transkriptes.

### 3.3.8 Durchmusterung einer Test-Bibliothek ENU mutierter ES-Zellen

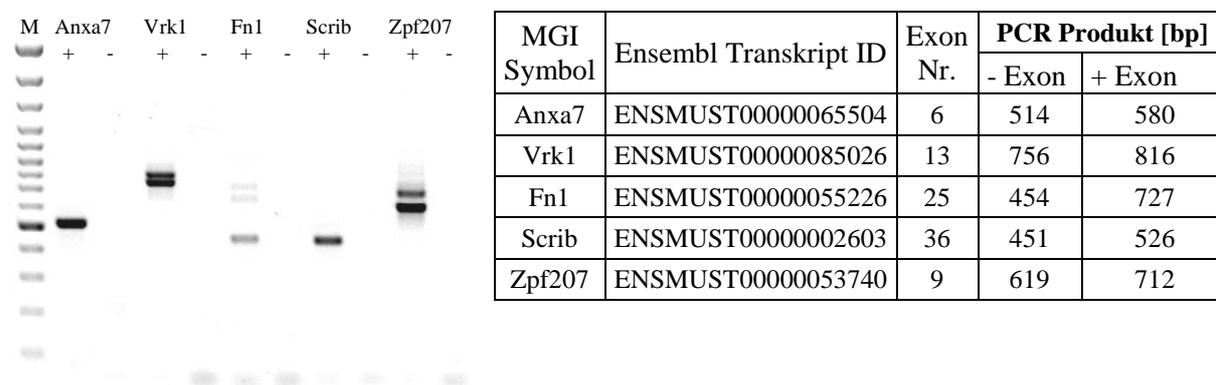
Basierend auf dem unter 3.3.5 erprobten Array-Design und den daraus gewonnenen Erkenntnissen wurde ein Array mit optimierter Oligonukleotid Zusammensetzung entworfen.

Es wurden folgende Veränderungen vorgenommen:

- Sonden mit falsch-positiven Signalen (1023 Oligos) wurden entfernt
- Aus den 54 Sonden der alternativ gespleißten Transkripte wurden 5 Spleiß-Ereignisse mittels PCR unter Verwendung der Primersysteme F nach Abbildung 17 bestätigt (Abbildung 19). Die entsprechenden Skip-Sonden wurden dann als interne Positiv-Kontrollen auf dem Array plaziert (5 Replikate).
- Zur Analyse der Expressionslevel und damit einer zusätzlichen internen Kontrolle zur Vergleichbarkeit der als Ausgangsmaterial verwendeten Gesamt-RNA Proben, wurden von 5 Haushaltsgenen (*Actb*, *B2m*, *Gabpd*, *Hprt*, *Ldh1*) die Oligonukleotid Sonden (50 bp) des Illumina Arrays übernommen (5 Replikate).

Mit dieser überarbeiteten Array-Struktur war es theoretisch möglich 36.196 Exon-Deletionen in 9363 Transkripten (7461 Genen) zu analysieren. Dieser erweiterte Aufbau der Sonden-Zusammenstellung wurde zunächst an drei biologischen Replikaten von aRNA des F1 ES-Zell Wildtyps und des Positiv-Kontroll Klons RR15 (1:5 Pool mit Wildtyp) getestet. Die

Ergebnisse dieser Hybridisierungen waren sehr zufrieden stellend, da alle technischen Parameter die Anforderungen der Feature Extraction Software erfüllten, die Intensitäten der Kontroll-Sonden (Skip-Oligos der alternativ gespleißten Gene, Housekeeping Gene) unter den vier Proben absolut vergleichbar waren und auch ein deutliches Signal für das Skip-Oligo *Hprt* Exon 8 (Mut/WT Signalverhältnis = 934) erzielt wurde.



#### Abbildung 19: Bestätigung alternativ gespleißter Exons mittels PCR

Die Abbildung zeigt die Ethidiumbromid vermittelte Fluoreszenz von PCR-Produkten nach gelelektrophoretischer Auftrennung in einem 1,5%-igen Agarosegel. Es wurden jeweils 20% eines PCR-Ansatzes aufgetragen. Aus den 54 Sonden alternativ gespleißter Exons wurden 6 Sonden ausgewählt, die als interne Positiv-Kontrollen auf Microarrays weiter Anwendung finden sollten. Zur Bestätigung dieser 6 Spleiß-Ereignisse wurden Primersysteme des Typs F nach Abbildung 17 erstellt und PCR mit Wt F1 ES-Zell cDNA als Matrize durchgeführt. Alle Ereignisse alternativen Spleißens konnten bestätigt werden, da Produkte für das kleinere alternativ gespleißte Transkript des jeweiligen Gens in allen Reaktionen amplifiziert wurden (vergl. nebenstehende Tabelle.). M = Größenstandard ‚GeneRuler™ 100 bp Plus DNA Ladder‘ 500 ng; - = Negativ-Kontrolle (ohne Matrizen DNA); + = Wildtyp F1 ES-Zell cDNA Matrize.

Für die Durchmusterung wurde ein Schema gewählt, welches das Vorkommen eines Klon in genau zwei Ansätzen mit jeweils 5 RNA Proben ermöglichte. Dieses 1:5 Pooling Schema (s.u.) ermöglichte, dass die Hybridisierungskosten auf 40% reduziert werden konnten. Weiterhin konnte die Auswertung analog zu 3.3.6 erfolgen, d.h. es fand ein Abgleich der Signalintensitäten statt. Es wurden nur die Sonden herausgefiltert, bei denen ein Klon in seinen beiden Ansätzen mit der gleichen Sonde einen erhöhten Quotient ( $> 5$ ; Mut/Wt Signal) aufwies. Insgesamt wurden 99 ENU-mutagenisierte Klone der Mutagenese #3 F1-1 (3.1.1) aufgeteilt auf 4 Gruppen a 25 Klone analysiert. Eine Gruppe enthielt als Kontrolle den Klon RRx006 mit der bestätigten Deletion in *Nup37*.

ID Ansatz	C1	C2	C3	C4	C5
R1	101	102	103	104	105
R2	106	107	108	109	110
R3	111	112	113	114	115
R4	116	117	118	119	120
R5	121	122	123	124	125

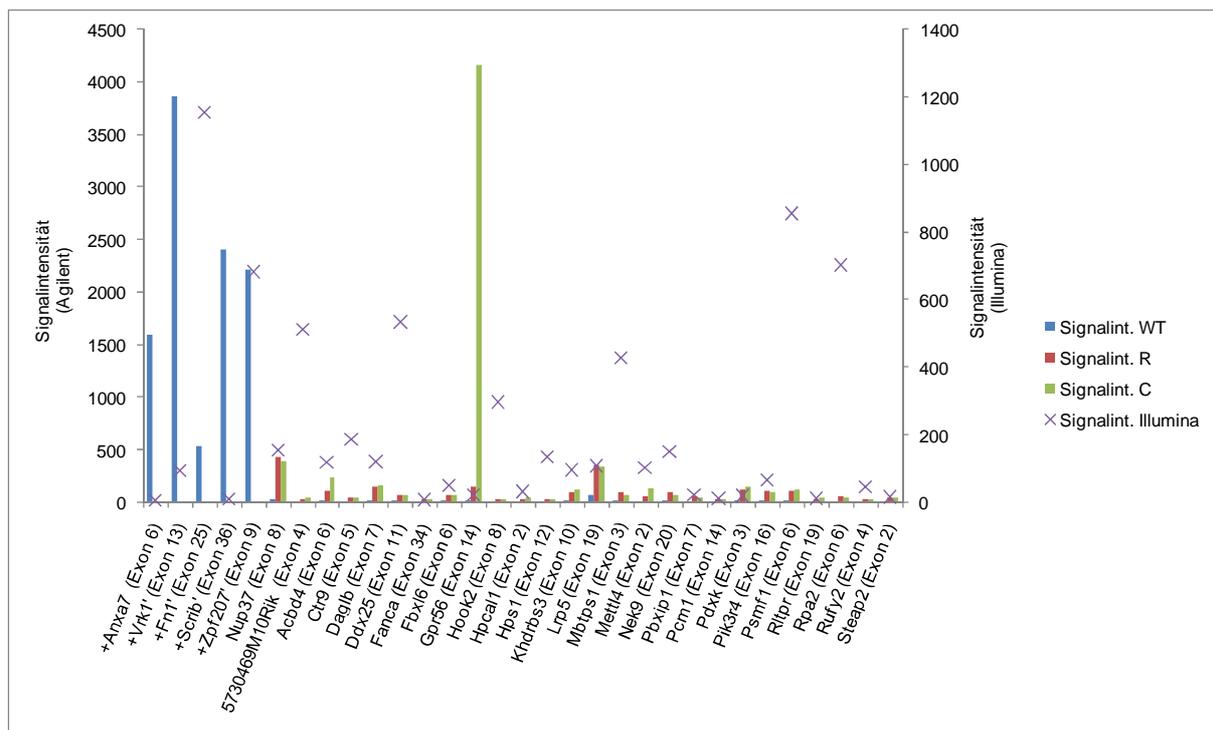
**Abbildung 20: Darstellung des genutzten Durchmusterungsschema einer Klon-Gruppe (25 Klone)**

Die Abbildung verdeutlicht beispielhaft den Aufbau einer der 4 Klon-Gruppen in dem Test-Experiment zur Durchmusterung einer Biobibliothek ENU mutierter ES-Zellen nach Exon-Deletions Mutationen. Jeder Hybridisierungsansatz (C1-C5 und R1-R5) enthält die RNA 5 verschiedener Klone. Jeder Klon ist in einem C und einem R Ansatz vertreten. Sonden, die in den Ergebnissen der Ansätze R4 und C3 erhöhte Signalintensitäten im Vergleich zum Wt Signal aufweisen, deuten beispielsweise auf eine entsprechende Deletion in der Mutante 118 hin.

Für jeden RNA Ampflikationsansatz von 5 RNA Proben wurde zunächst eine Stammlösung der RNA-Proben erstellt, in der die Proben mit gleichen Gesamt-RNA Mengen (w/v) vertreten waren. Von diesem Gemisch Gesamt-RNA wurden dann 400 ng in die unter 2.3.8.2 beschriebene Amplifikation zu aRNA eingesetzt. Alle weiteren Amplifikations- und Hybridisierungsschritte erfolgten streng nach Protokoll. Die erhaltenen Rohdaten wurden wie bereits beschrieben zunächst über die Feature Extraction Software normalisiert und analysiert. Zwei der 40 Hybridisierungen (R13, C18) wurden aufgrund unklarer Ergebnisse zur Linearität der Signalintensitäten verworfen und wiederholt. Eine Vergleichbarkeit der Signale der Haushaltsgene war danach für alle Proben gegeben. Zudem wurden mit allen Proben zufrieden stellende Signalintensitäten mit denen als interne Kontrollen genutzten Skip-Oligonukleotiden alternativ gespleißter Exons erhalten.

Insgesamt erfüllten 26 Skip-Sonden (Abbildung 21) die unter Abschnitt 3.3.6 beschriebenen Vorgaben des Auswerte-Algorithmus: Das Verhältnis der Signalintensitäten von Mutante zu Wildtyp musste in genau einem Ansatz R und genau einem Ansatz C (vergl. Abbildung 20) größer 5 sein. Von der Möglichkeit, dass 2 Klone die gleiche Mutation aufweisen, wurde hier aufgrund der geringen Wahrscheinlichkeit abgesehen. Erstaunlicherweise traten diese 26 putativen Mutationen in nur 19 der 99 untersuchten Klone auf. Eine der Exon-Deletionen war die Deletion des Exon 8 in *Nup37* (Klon RRx006), die als Positiv-Kontrolle eingesetzt wurde. Die erreichte Signalintensität lag allerdings unter der der vorherigen Hybridisierung (hier: 424,7 bzw. 385,3 (1:5), vorher: 651,8 (1:5); 3.3.6). Bei einer tieferen Betrachtung der Signalintensitäten die mit den Sonden der möglichen Exon-Deletionen erzielt wurden, fiel auf, dass von den 52 betrachteten Intensitäten 36 unter 100 Einheiten lagen (~ 69%), 24 davon sogar im Bereich des Hintergrundsignals (ca. < 50 Einheiten). Der höchste Wert wurde mit der *Nup37* Exon 8 spezifischen Sonde erzielt. Im Vergleich dazu betrug die

Signalintensitäten der internen Skip-Positiv-Kontrollen, die ebenfalls 30 bp lang sind, zwischen 524 und 4199 Einheiten in den Wildtyp-Proben. Die 50 bp lange Sonde des Haushaltsgens *Gapd* erreichte in den Wildtypproben sogar eine Durchschnittsintensität von 194.544 Einheiten. Für die Sonde spezifisch zur Deletion von *Grprf56* Exon 14 wurde in dem Ansatz C eine sehr hohe Intensität erzielt, deren Ursache unklar ist. Es handelte sich nicht um ein Hybridisierungsartefakt (vergl. Abbildung 16).



**Abbildung 21: Signalintensitäten ausgewählter Skip-Oligonukleotid Sonden**

Dargestellt sind die normalisierten Signalintensitäten (linke y-Achse) ausgewählter Skip-Oligonukleotide, die mit dem unter 3.3.7 beschriebenen Agilent Microarray aufgezeichnet wurden. Gezeigt sind die Intensitäten für die 5 internen Positiv-Kontroll Skip-Oligonukleotide (alternatives Spleißing; MGI Symbol durch ein vorgestelltes + markiert) und für die im Text beschriebenen 26 Sonden, deren Signale Hinweise auf Exon-Deletions-Mutation geben. Jede Sonde wird durch das MGI Symbol und das mögliche Deletions-Ereignis (z.B. Exon 6) beschrieben. Für alle Sonden ist die durchschnittliche Signalintensität (n=3) der Wildtyp-Proben angegeben (Positiv-Kontrollen n = 15, 3 Proben mit je 5 Replikaten). Für die 26 selektierten Sonden ist zusätzlich die Signalintensität in den beiden 1:5 Pooling-Ansätzen (R und C, n=1) dargestellt, in denen das Signal mind. 5 x größer als das des durchschnittlichen Wildtyp-Signals war (vergl. Text). Zu Vergleichszwecken ist auf der rechten y-Achse zusätzlich die normalisierte Signalintensität des jeweiligen Transkripts (Illumina Exoexpressions Array) aus der unter 3.2.1 durchgeführten Expressionsanalyse eingezeichnet.

Gegen die Aussagekraft der Ergebnisse spricht, dass für knapp 30% der Sonden mit Signalintensitäten kleiner 100 Einheiten auch in mindestens einem der anderen Ansätze einer Klon-Gruppe erhöhte Signalintensitätsverhältnisse größer als 4 (aber kleiner als 5) festgestellt werden konnten. Diese Schwankungen können darauf beruhen, dass sich die Intensitäten am unteren Ende des Bereichs linearer Signalintensitäten ( $\sim 10^{1.4}$  bis  $10^{5.3}$  Einheiten) bewegen.

Abbildung 21 zeigt zusätzlich die Signalintensitäten, die in der Expressionsanalyse von Wildtyp F1 ES-Zellen RNA mit einem Illumina Microarray für das entsprechende Transkript

erzielt wurden. Geht man davon aus, dass diese Signalintensitäten ein relatives Maß für die Expressionsstärke des Gens sind, so kann man keinen direkten Zusammenhang zu den Signalintensitäten der Skip-Oligonukleotide sehen. Im Vergleich zu *Hprt* (stärkste Illumina Signalintensität 2709 Einheiten) handelt es sich allerdings bei dem Großteil der betrachteten Transkripte um sehr schwach exprimierte Transkripte. Für 10 Transkripte (*Lrp5*, *Daglb*, *Gpr56*, *Pdxk*, *Acbd4*, *Psmf1*, *Pik3r4*, *Khdrbs3*, *Nek9*, *Mbtps1*) mit den eindeutigsten Skip-Oligo Signalen wurde der Nachweis der Deletion mittels PCR (vergl. 3.3.7) unternommen. Es konnte allerdings keine Deletion bestätigt werden (Daten nicht gezeigt).

### 3.3.9 Neugestaltung des Microarray Aufbaus

Die unzufrieden stellenden Ergebnisse der Durchmusterung des Test-Bibliothek ließen Zweifel an der praktischen Umsetzbarkeit des Detektionssystems und der Wirtschaftlichkeit zu. Es wurde daher beschlossen zunächst die Detektionsvariante mittels Skip-Oligonukleotiden beizubehalten aber basierend darauf ein neues Test-System zu entwickeln, welches mögliche Problemfaktoren verringert. Folgende technische Faktoren wurden anhand voriger Ergebnisse als problematisch eingestuft:

- Oligonukleotidsequenz (GC-Gehalt, Repeats)
- Sensitivität (Expressionsstärke, Länge der Oligonukleotide, zu hohe Verdünnung der Proben in Pools)

Aufgrund dieser Überlegungen wurde für das geplante Experiment entschieden zunächst auf Pooling zu verzichten und die Proben einzeln zu hybridisieren. Um die Mehrkosten für diesen Ansatz zu relativieren, wurde anstelle des 4 x 44 k Microarray Formats das von Agilent seit Sommer 2007 neu angebotene 8 x 15 k Format verwendet. Die Anzahl zu betrachtender Exon-Deletionen verringerte sich dadurch auf 15.208 Ereignisse. Zur Erhöhung der Signalintensitäten wurde die Länge der Sonden von bisher 30 bp auf 36 bp erhöht. Eine weitere Erhöhung der Länge wurde zur Erhaltung einer ausreichenden Spezifität (vergl. Abbildung 13) für nicht sinnvoll erachtet. Auf Basis der unter 3.3.2 entwickelten Software, der Liste der 7.475 exprimierten Gene und einer aktuellen Exon-Auflistung (BioMart, Ensembl 46) wurden so 31.503 Sonden für Exon-Deletionen generiert, die den ORF des Transkripts nicht unterbrechen. Das entwickelte Perl Skript wurde dahingehend erweitert, dass der GC-Gehalt der Sonden zwischen 35% und 65% betragen musste, nicht mehr als 4 Wiederholungen einzelner Nukleotide (z.B. AAAAA) vorkamen und mit der Sonde keine 100% Treffer in einem Blast gegen die Maus-EST Datenbank erzielt werden konnten (Extraktion falsch positiver Sonden). Dies reduzierte die Anzahl der Sonden auf 17.366. Um

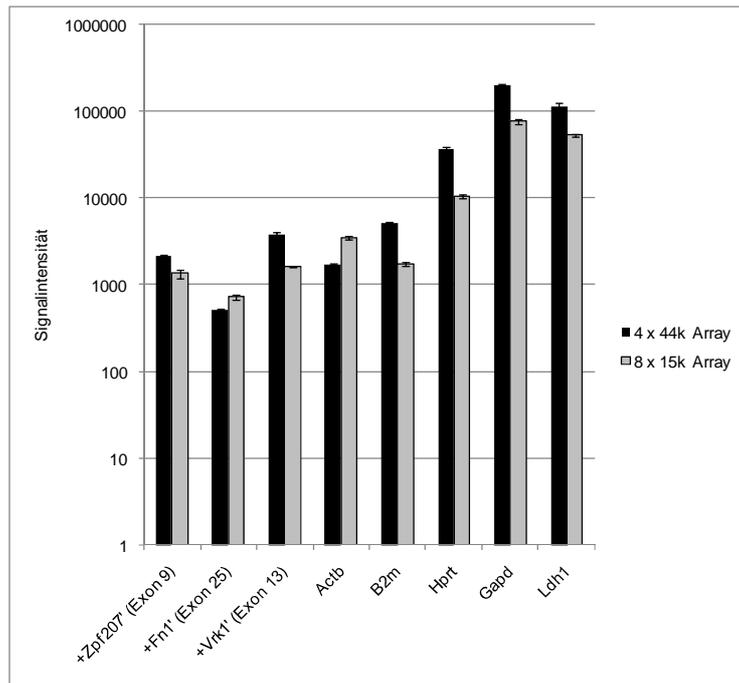
der maximal platzierbaren Anzahl von 15.208 Oligonukleotiden unter Berücksichtigung der Kontrollen gerecht zu werden, wurden die 2.227 Sonden der 17.366 entfernt, deren Transkripte die schwächsten Signale in der Expressionsanalyse (3.2.1) aufwiesen und die Anzahl der Positiv-Kontrollen reduziert. Den finalen Aufbau des Arrays zeigt Tabelle 16. Insgesamt konnten so 15.139 Deletions-Ereignisse in 4.698 Transkripten (3.885 Gene) parallel analysiert werden.

**Tabelle 16: Oligonukleotid Zusammenstellung des 8 x 15k Arrays**

	Verknüpfung [Exon Nr.] / Sequenz	Anzahl verschiedener Oligos	Replikate	Länge [bp]
Skip-Oligos [Hprt]	1-3, 2-4, ..., 7-9	7	3	36
Junction-Oligo [Hprt]	1-2, 2-3, ..., 8-9	8	3	36
Housekeeping Kontrollen	<i>Actb</i> , <i>B2m</i> , <i>Gapdh</i> , <i>Hprt</i> , <i>Ldh1</i> (vergl. 3.3.8)	5	3	50
Skip-Kontrollen (alt. gespleißt)	( <i>Fnl1</i> , Exon 25; <i>Vrk1</i> , Exon 13; <i>Zpf207</i> , Exon 9; vergl. 3.3.8)	3	3	36
Screening Skip-Oligos	Alle Skip-Oligos der in der F1-Linie exprimierter Gene (7407)	15.139	1	36

Für die Anzahl von 3.885 Genen kann man die Abschätzung nach 3.1.3 wiederholen und erhält für den 8 x 15 k Array einen Wert von einer erwarteten Detektion einer Exon-Deletion in 7,8 Klonen. Es wurden insgesamt 30 ENU mutagenisierte Einzelklone, u.a. der Klon RRx006 mit der Deletions-Mutation von Exon 8 in *Nup37*, und 2 biologische Wildtyp Replikate als Referenz hybridisiert. Die Vorbereitung der aRNA Proben und die Hybridisierung wurden nach Protokoll durchgeführt (2.3.3, 2.3.8.2). Die Auswertung und Normalisierung der erhaltenen Signale erfolgte mittels der Feature Extraction Software (2.1.3).

Zunächst wurde überprüft, inwieweit sich die Signalintensitäten der Skip-Oligonukleotid Positiv-Kontrollen durch die längeren Oligonukleotide von 36 bp im Vergleich zu der vorherigen Hybridisierung verändert haben. Erstaunlicherweise konnte eine Zunahme der Signalintensitäten nur für eine der drei Sonden beobachtet werden (+37%). Mit den beiden anderen Sonden wurden eine Abnahme der Intensitäten verzeichnet. Ähnliche Ergebnisse wurden mit den Sonden der Housekeeping Gene erzielt, deren Länge von 50 bp zwischen den alten und neuen Arrays nicht verändert wurde (Abbildung 22). Dies lässt den Schluss zu, dass zum einen die Intensitäten zwischen den verschiedenen Array-Formaten nicht vergleichbar sind und zum anderen, dass mit diesem Array hinsichtlich der Verbesserung der Sensitivität das Ziel generell nicht erreicht werden konnte.



**Abbildung 22: Signalintensitäten der Positiv-Kontrollen in 4 x 44 k und 8 x 15 k Arrays**

Dargestellt sind die durchschnittlichen Signalintensitäten der drei Positiv-Kontroll Skip-Oligonukleotide (*Zpf207*, *Fn1*, *Vrk1*) und der fünf Haushaltsgene (*Actb*, *B2m*, *Hprt*, *Gapdh*, *Ldh1*) die mit Wt F1 aRNA Proben erzielt wurden. (4 x 44 k: 3 Proben mit je 5 Oligo-Replikaten; 8 x 15 k: 2 Proben mit je 3 Oligo-Replikaten). Während die Oligonukleotide der Haushaltsgene in beiden Array-Formaten 50 bp lang sind, haben die Skip-Oligos auf dem 4 x 44 k Array eine Länge von 30 bp, auf dem 8 x 15 k Array eine Länge von 36 bp. Die Standardabweichung der Mittelwerte wird durch senkrechte Linien auf den Säulen dargestellt.

Die Auswertung hinsichtlich der Detektion neuer Exon-Deletionen erfolgte analog zu der bereits beschriebenen Verfahrensweise (3.3.6). Dabei passierten 22 Sonden die angewendeten Filter. Diese Zahl erschien bei einer erwarteten Anzahl von ca. 2-3 Deletionen relativ hoch. Da aus Kostengründen nur Einzelhybridisierungen durchgeführt wurden und somit kein Abgleich zwischen zwei Ansätzen mit gleichem Probenmaterial möglich war, konnten die Ergebnisse auch auf Hybridisierungsartefakten beruhen. Daher wurde hier besonderer Wert auf die visuelle Kontrolle der entsprechenden Signale gelegt. Bei 19 der 22 Signalpunkte konnten tatsächlich Unregelmäßigkeiten in der Struktur festgestellt werden (z.B. Signalpunkt nur zur Hälfte oder fleckig leuchtend; vergl. Abbildung 16), die von der Feature Extraction Software nicht als solche erkannt worden waren. Diese Signale wurden als Artefakte gewertet und nicht weiter betrachtet. Erwartungsgemäß konnte die bereits mehrfach bestätigte Deletion des Exon 8 im *Nup37* Transkript der Mutante RRx006 erneut detektiert werden. Für die entsprechende Sonde konnte im Vergleich zum vorherigen 4 x 44 k Array, auch unter Berücksichtigung der Pooling-Verhältnisse, hier eine deutliche Zunahme der Signalintensität ermittelt werden (ca. + 70%). Allerdings war auch eine Zunahme des Signals in der Wildtyp-Probe zu beobachten. Weiterhin konnten anhand der Hybridisierungsergebnisse eine Deletion des Exon 2 im Gen *Epb4.113* und eine Deletion des Exon 13 im Gen *Mod1* detektiert werden

(Tabelle 17). Der Versuch, diese beiden Mutationen mittels der beiden PCR-Systeme nach Abbildung 17 zu verifizieren, schlug fehl (Daten nicht gezeigt.). Für beide Transkripte wurden keine Produkte mit dem Skip-Primersystem erhalten. Mit den flankierenden Primern konnten nur die dem Wildtyp entsprechenden Fragmente amplifiziert werden.

**Tabelle 17: Informationen zu Deletions-Ereignissen des 8 x 15 k Arrays**

Signalintensitäten des Wildtyps (Wt) sind das arithmetische Mittel aus zwei biologischen Replikaten. Der Quotient (Q) berechnet sich wie folgt: Mut/Wt

Sonden Nr.	Sonden Name	MGI Symbol	Ensembl Transkript ID	Exon Nr.	Signalint.		Q	Mutante
					Wt	Mut		
1677	ag4_skip12303	Epb4.113	ENSMUST00000077529	2	3,9	675,9	171,3	RRx024
10378	ag4_skip29798	Mod1	ENSMUSE00000219622	13	5,1	119,1	23,2	RRx020
8913	ag4_skip02670	Nup37	ENSMUST00000052355	8	54,7	3435,1	62,8	RRx006

Zusammenfassend kann festgehalten werden, dass eine generelle Verbesserung der Sensitivität nicht erzielt werden konnte. Es wurde durch die erweiterte *in-silico* Synthese eine Konzentrierung auf erfolgsversprechendere Oligonukleotid-Sonden erzielt, allerdings wurde dafür eine Reduktion detektierbarer Deletionen in Kauf genommen. Doch würde diese Konzentrierung und die Verwendung von 8 x 15 k Arrays aber nahezu eine Halbierung der Kosten bedeuten, da bei gleichen Array-Kosten auf einem 8 x 15k Array 8 statt 4 Proben parallel analysiert werden können. Mit Blick auf die erzielten Signalintensitäten lässt sich festhalten, dass eine Detektion der beiden Deletions-Ereignisse in den Transkripten von *Epb4.113* und *Mod1* aus 1:5 Pools sehr unwahrscheinlich gewesen wäre. So wurden diese Ereignisse auch nicht mit dem unter 3.3.8 beschriebenen Aufbau detektiert.

An dieser Stelle wurde auf eine Weiterentwicklung des Detektionssystems aufgrund der genannten Schwächen verzichtet und stattdessen die zwischenzeitlich entwickelte Hochdurchsatzsequenzierung im Sinne der Detektion von Exon-Deletionen untersucht.

### 3.4 Detektion von Mutationen mittels Hochdurchsatzsequenzierung

Im Zuge der Resequenzierung ganzer Genome mittels Sequenzierverfahren der zweiten Generation (second generation sequencing) (Kim et al., 2007) und den ersten Versuchen diese Verfahren der Analyse ganzer Transkriptome zugänglich zu machen (Barbazuk et al., 2007; Cheung et al., 2006; Emrich et al., 2007; Weber et al., 2007), wurde in diesem Teil der Arbeit die Idee verfolgt, diese Techniken zur Detektion ENU induzierter Mutationen zu nutzen. Ein Vorteil einer auf Sequenzierung basierten Detektion zu der in den vorangehenden Abschnitten beschriebenen auf Microarray basierten Detektion wurde insbesondere in der Tatsache gesehen, dass keine Vorauswahl auf bestimmte Sequenzen erfolgen muss.

### 3.4.1 Aufbau des Experiments

Die in dieser Arbeit verfolgte Vorgehensweise zur Detektion von Exon-Deletionen mittels Second Generation Sequencing lässt sich einfach beschreiben: Doppelsträngige cDNA einer ENU mutagenisierten ES-Zelllinie wird mittels Solexa Sequenzierung (2.3.9.2) sequenziert. Die daraus resultierenden kurzen Sequenzfragmente (hier 27 bp) werden im Anschluss elektronisch mit einer Sequenz-Datenbank abgeglichen, die nur Sequenzfragmente möglicher Exon-Deletionen enthält. Vollständig homologe Sequenzen zwischen einem sequenzierten Fragment (nachfolgend als ‚Read‘ bezeichnet) und einer Sequenz der Datenbank sind dann ein eindeutiger Hinweis auf eine entsprechende Exon-Deletion in dieser Mutante, sofern alternatives Spleißen des Exons und Sequenzierfehler ausgeschlossen werden können.

Aus dieser Aufgabenstellung heraus wurden folgende Meilensteine festgelegt, die es für eine zielgerichtete Gestaltung und Auswertung des Detektion-Experiments zu betrachten galt:

1. Zunächst sollte untersucht werden, wie der Aufbau der Referenzdatenbank mit den Sequenzen möglicher Exon-Deletions Ereignisse zu gestalten ist, um eine schnelle und Ressourcen optimierte Analyse der zu erwartenden großen Datenmengen zu ermöglichen.
2. Es lagen bis dato keine Veröffentlichungen zur Analyse eines Transkriptoms, im Speziellen die Sequenzierung doppelsträngiger cDNA, mittels Solexa Sequenzierung vor. Es musste daher überprüft werden, ob eine DNA-Fragment Bibliothek, die ausgehend von ds-cDNA als Matrize mit dem Solexa SSB Kit (2.1.3) erstellt wurde, verwertbare Ergebnisse liefert.
3. Aufbauend auf diesen Ergebnissen musste geklärt werden, für wie viele Transkripte überhaupt eindeutige Sequenzen erhalten werden und wie hoch die Abdeckung der Transkripte mit den Reads ist. Existieren bestimmte technische Befangenheiten? Ist es zwingend erforderlich mit aufwendig normalisierten cDNA Bibliotheken zu arbeiten?
4. Abschließend sollte die Auswertung der Sequenzen hinsichtlich der Detektion von Exon-Deletionen beispielhaft anhand der ENU Mutante RRx006 (bestätigte Deletion von Exon 8 in *Nup37*) erfolgen.

### 3.4.2 Entwicklung einer Exon-Deletions Referenzdatenbank

Die Qualitätskontrolle und der Abgleich der durch Sequenzierung generierten 27 bp langen Sequenzfragmente sollte mittels ELAND (2.4.1) erfolgen. Bei ELAND handelt es sich um einen Algorithmus, der von Illumina speziell zum Abgleich der millionenfachen Reads mit Referenzsequenzen entwickelt wurde. Die Geschwindigkeit übertrifft BLAST um mehr als

das Hundertfache, dafür ist die Analyse aber auch deutlich einfacher: Es werden von der Software nur solche Sequenzen als homolog angesehen, die nicht mehr als zwei nicht-homologe Basen (nachfolgend als 'Mismatch' bezeichnet) zur Referenz aufweisen. Der Anwender erhält für jedes sequenzierte DNA Molekül bestimmte Rückgabewerte, die an dieser Stelle für das Verständnis späterer Analysen anhand der nachfolgenden Abbildung erläutert werden.

>I_1_505_815	GAAACACGAAATATTCACAAAGGA	U0	1	0	0	reflma_nm9_eland_transcriptome.txt	32473030	F	..	
>I_1_879_97	TTCTAGCTCTTTGCCAAGGTATGTTT	U1	0	1	0	reflma_nm9_eland_transcriptome.txt	32814590	F	..	226
>I_1_502_102	GTTAAACACAGGATATCAGGAGTTT	U0	1	0	0	reflma_nm9_eland_transcriptome.txt	15137983	F	..	
>I_1_895_106	GCTTCCACTTCAATCCCATCACCAT	U0	1	0	0	reflma_nm9_eland_transcriptome.txt	44637113	R	..	
>I_1_822_880	TTTACGTAATTAACCCGTTTATTCGAC	U1	0	1	0	reflma_nm9_eland_transcriptome.txt	16036029	R	..	27C
>I_1_753_877	GCAGAGTCTGAAATGAFAGATGAGGAG	U0	1	0	0	reflma_nm9_eland_transcriptome.txt	19856397	F	..	
>I_1_809_915	TATAGTTTCCCTCTTCTCTCTAATT	U0	1	0	0	reflma_nm9_eland_transcriptome.txt	54933489	R	..	
>I_1_939_125	GTTTACTTTTGTATATAAATCAGGA	U0	1	0	0	reflma_nm9_eland_transcriptome.txt	23232457	F	..	
>I_1_590_186	GTGAGTCTGCGCATCAGAGACTGAAA	U1	0	1	0	reflma_nm9_eland_transcriptome.txt	16035956	F	..	276
>I_1_789_908	GGACGAGACAAACATTTATCCGATC	U1	0	1	0	reflma_nm9_eland_transcriptome.txt	54464540	F	..	25C
>I_1_764_911	GGGACCCATCCCGCTTTTCTGTCA	U0	1	0	0	reflma_nm9_eland_transcriptome.txt	45524195	R	..	
>I_1_131_390	AACAGTGCCTTCTCGGAGTTTITTT	U1	0	1	0	reflma_nm9_eland_transcriptome.txt	35129737	F	..	276

### Abbildung 23: Allgemeine Erläuterung der ELAND Ergebnisse

Die Abbildung zeigt beispielhaft anhand eines Ausschnitt einer Datei den grundsätzlichen Aufbau der ELAND Resultate. **Spalte 1:** Identifikationsnummer des sequenzierten Fragments (Read); **Spalte 2:** Sequenz des Reads; **Spalte 3:** Art des Treffers gegen die Referenz [NM = no match, kein Treffer; QC = QC failure, keine Analyse aufgrund zu schlechter Qualität des Reads; RM = repeat masked (in dieser Arbeit nicht verwendet); U0, U1, U2 = unique match, eindeutiger Treffer in der Referenz DB mit keinem, einem oder zwei 'mismatches'; R0, R1, R2 = multiple exact matches, Read trifft mehrere Referenzsequenzen exakt mit keinem, einem oder zwei 'mismatches']; **Spalte 4:** Anzahl exakter Treffer; **Spalte 5:** Anzahl exakter Treffer mit einem 'mismatch'; **Spalte 6:** Anzahl exakter Treffer mit zwei 'mismatches'; **Spalte 7:** Name der Referenz-Datei; **Spalte 8:** Start-Position des Reads in der Referenzsequenz; **Spalte 9:** Richtung des Reads im Vergleich zur Referenz (F = Vorwärts, gleiche Sequenz; R = Rückwärts = Komplementäre Sequenz); **Spalte 10:** Interpretation von N in der Sequenz (.,' = unklar; ,D' = Deletion einer Base; ,I' = Insertion einer Base); **Spalte 11/12:** Position in dem Read und Typ in der Referenz des 'mismatch' in der ersten/zweiten Substitution.

Die Exon-Deletions Referenzdatenbank sollte alle möglichen Exon-Deletionen bekannter RefSeq Transkripte beinhalten. Die Erstellung dieser Sequenzen konnte mit dem bereits beschriebenen Skript zur *in-silico* Synthese von Skip-Oligonukleotiden erfolgen (3.3.2). Theoretisch kann man dies wie folgt formulieren: Für eine angedachte Deletion des Exons n aus Transkript X wird das 3'-Ende des Exons n-1 mit dem 5'-Ende des Exons n+1 zu gleichen Anteilen zu einer Exon-Deletionssequenz verknüpft.

Hier stellte sich die Frage, wie lang diese Sequenzen für eine optimale Analyse sein müssen: Bei 27 bp langen Reads durften die Referenzsequenzen nicht länger als 52 bp sein, um zu gewährleisten, dass ein Treffer eines Abgleichs von Read und Referenz auch aussagt, dass der Read mit mindestens einer Base über der Exongrenze liegt. Da ELAND allerdings bis zu 2 Mismatches pro Read zulässt, die auch beide am Ende des Reads liegen können, waren nur Referenzsequenzen mit Längen von maximal 48 bp denkbar. Ergebnisse, die mit diesen 48 bp langen Referenzsequenzen erzielt worden wären, hätten allerdings den Nachteil gehabt, dass die Wahrscheinlichkeit für einen sicheren Treffer aufgrund des Aufbaus der DNA aus 4 Basen nur 0,75 ( $p = 1 - 1/4^1$ ) betragen hätten. Die Wahrscheinlichkeit für einen sicheren Treffer

sollte für uns größer als 99% sein, sodass unter Berücksichtigung der 2 Mismatches noch 4 weitere Nukleotide über der Exongrenze liegen mussten ( $p = 1 - 1/4^4 = 0.996$ ). Für die Referenzsequenzen wurden daher Längen von 42 bp angenommen.

Diese Sequenzen wurden unter Verwendung aller Maus RefSeq Exonsequenzen (BioMart, Ensembl 46; 295.164 Exons aller Transkripte) mit dem unter 3.3.2 entwickelten Algorithmus generiert (165.931 Deletionssequenzen). Da die Transkripte eines Gens gleiche Exons enthalten, wurden redundante Deletionssequenzen nur einmal berücksichtigt (165.931 Deletionssequenzen). ELAND würde Treffer auf diese Sequenzen sonst als R einstufen, was zur Folge hätte, dass diese nicht mehr eindeutig zuzuordnen wären. Die korrekte Funktionalität des Skriptes zur Erstellung der db42 wurde durch die stichprobenartige Kontrolle von 20 zufällig gewählten Sequenzen bestätigt. Die Sequenz-Datenbank wurde als db42 bezeichnet.

Es muss an dieser Stelle angemerkt werden, dass es sich bei den Sequenzen in der db42 Datenbank natürlich nur um Exonsequenzen handelt, die *in silico* nach oben genanntem Muster zusammengestellt wurden. Auch wenn die Sequenzen hier als 'Exon-Deletionssequenzen' bezeichnet werden, deuten Treffer nicht zwangsläufig auf die mutationsbedingte Deletion des entsprechenden Exons hin. Zum größten Teil wird es sich bei den Treffern sogar um die Detektion alternativen Spleißens handeln. Es wird daher nachfolgend auch allgemein von der Detektion eines Exon-Skip Ereignisses gesprochen. Generell ist es unabdingbar neben der zu analysierenden Probe eines mutierten Klons auch die Probe einer Wildtypzelllinie zum Vergleich heranzuziehen. Zusätzlich kann ein Abgleich mit der Maus EST Datenbank erfolgen.

Da ELAND in der verwendeten Version noch nicht effizient mit mehreren FASTA Sequenzen in einer Datei (multiple Fasta Datei Format) arbeiten konnte und die Arbeit mit 165.931 Einzeldateien die Auswertung bedeutend verlangsamt hätte, wurde mittels eines Perl-Skriptes ein künstliches Genom aus diesen Sequenzen erstellt (transcriptome4eland.pl) (Rosenkranz et al., 2008). Dazu wurden Sequenzen durch 4 N (Basenpaare) von einander getrennt aneinandergesetzt und die jeweiligen Positionen und Informationen zur späteren Auswertung mittels eines anderen Skriptes in einer Datenbank vermerkt. Zur automatisierten Analyse der mittels ELAND erzeugten Datensätze (Abbildung 23) wurde ein Skript entwickelt (eland\_transcriptome\_analysis.pl) (Rosenkranz et al., 2008), welches für jede Referenzsequenz die Anzahl der passenden ‚Reads‘ ermittelte. Die Funktionalität der Skripte wurde anhand einer kleinen Anzahl von Datensätzen (50 Stück) bestätigt.

### 3.4.3 Sequenzierung des ES-Zell Transkriptoms

Die Sequenzierung des Transkriptoms erfolgte parallel für die Wildtyp Stammzelllinie F1 und den Klon RRx006 in einem Lauf, um absolut vergleichbare Bedingungen zu gewährleisten. Aus beiden Zelllinien wurde zunächst mittels TRIzol-Reagenz Gesamt-RNA isoliert, ein DNase-Verdau durchgeführt und die Integrität der RNA kontrolliert (2.3.3). Anschließend wurde ausgehend von 250 µg Gesamt-RNA eine Extraktion der mRNA wie unter 2.3.4 beschrieben mit dem OligoTex mRNA Mini Kit durchgeführt. Die mRNA-Fraktion wurde vollständig in die sich unmittelbar anschließende Synthese doppelsträngiger cDNA eingesetzt (2.3.6). Die cDNA wurde abschließend in 15 µl TE gelöst, die Konzentration spektrophotometrisch bestimmt und die Größenverteilung der Fragmente auf einem 1.5%-igen Agarosegel betrachtet. Die Ausbeute betrug für beide Proben ca. 1.5 µg ds-cDNA. Die Vorbereitung der Proben für die eigentliche Sequenzierreaktion, d.h. die Generierung von DNA-Bibliotheken erfolgte ausgehend von 0,5 µg ds-cDNA streng nach Protokoll (2.3.9.2). Die Sequenzierung auf der Illumina 1G Einheit dauerte 27 Zyklen, d.h. die resultierenden Sequenzen waren 27 bp lang. Die Auswertung der visuellen Rohdaten und damit die Generierung der Sequenz- und Qualitätsdaten wurde mit dem Goat Modul (Versionen: Firecrest v1.8.28 und Bustard v1.8.28) der ,Illumina Pipeline v02.2.5 durchgeführt.

Es galt zunächst festzustellen, ob die Sequenzierung der ds-cDNA-Bibliotheken erfolgreich war. Es wurde daher überprüft, ob ein erfolgreicher Abgleich mit einer Transkript-Datenbank möglich war. Dazu wurde analog zu 3.4.2 mit dem Skript transcriptome4eland.pl ein künstliches Genom aller RefSeq Sequenzen (NCBI build 37, 20.804 Transkripte) erstellt und mittels ELAND ein Abgleich mit den Sequenzierdaten durchgeführt. Die Auswertung des Abgleichs erfolgte mit dem Skript eland\_survey.pl, welches eine Zusammenfassung der ELAND Ergebnisse generiert. Die Ergebnisse zeigt Tabelle 18. An dieser Stelle muss angemerkt werden, dass es sich bei der verwendeten F1 Zelllinie um eine Kreuzung aus den Inzuchtlinien C57BL/6J und 129/Sv handelt, während die verwendeten Referenzsequenzen sich ausschließlich auf den C57BL/6J Hintergrund beziehen. Daher können in den Analysen einige Reads des 129/Sv Genotypen nicht auftauchen, wenn diese in einem Read mehr als 2 Basensubstitutionen zu C57BL/6J aufweisen (ELAND).

Die Ergebnisse zeigen, dass die gewählte Strategie der Proben-Vorbereitung die Sequenzierung des Transkriptoms erlaubt. Für 58,54% der F1 WT Reads (U+R) und 60,07% der RRx006 Reads (U+R) konnten Treffer in der RefSeq Datenbank erzielt werden. In der Wildtyp Probe wurden für 14.434 RefSeq und in der F1 ENU RRx006 Probe für 14.358 RefSeq der untersuchten 20.904 Transkripte eindeutige Reads (U0, U1, U2) ermittelt. Daraus

kann gefolgert werden, dass der Anteil transkribierter Gene ca. 69% in diesen Stammzelllinien beträgt und damit deutlich über dem Anteil liegt, der mit Microarray Analysen erfasst wird (ca. 40%, vergl 3.2.1).

**Tabelle 18: ELAND Abgleich der Sequenzdaten gegen die Maus RefSeq Datenbank**

Insgesamt wurden  $82,42 \times 10^6$  bp in der F1 Wildtyp-Probe und  $88,25 \times 10^6$  bp in der F1 RRx006 Probe sequenziert. Im Vergleich dazu enthielt die RefSeq Datenbank 20.904 Transkripte mit einer Gesamtanzahl von 57.556.007 bp. Die Einstufung der Reads ist in Abbildung 23 wiedergegeben.

Art der Reads	F1 Wildtyp		F1 RRx006	
	Anzahl Reads	%	Anzahl Reads	%
<b>Gesamt</b>	3.052.501	100	3.268.410	100
<b>NM</b>	1.153823	37,80	1.207.557	36,95
<b>QC</b>	111.651	3,66	97.589	2,99
<b>R0</b>	335.051	10,98	309.363	9,47
<b>R1</b>	75.216	2,46	63.035	1,93
<b>R2</b>	42.800	1,40	33.273	1,02
<b>U0</b>	999.929	32,76	1.230.297	37,64
<b>U1</b>	224.791	7,36	218.998	6,70
<b>U2</b>	109.240	3,58	108.298	3,31

Zu den hohen Anteilen von Reads ohne Treffer (NM) muss hier angemerkt werden, dass sich deren Anteil in weiteren tiefergehenden Analysen mit anderen Datenbanken (z.B. dbEST, 2.4.3) und der Verwendung anderer Algorithmen noch deutlich reduzieren lässt (Rosenkranz et al., 2008). Dies liegt insbesondere darin begründet, dass nicht alle Transkripte in der RefSeq Datenbank enthalten sind und der Homologie-Abgleich mittels ELAND, welcher nur zwei 'Mismatch-Basen' akzeptiert, sehr stringent ist.

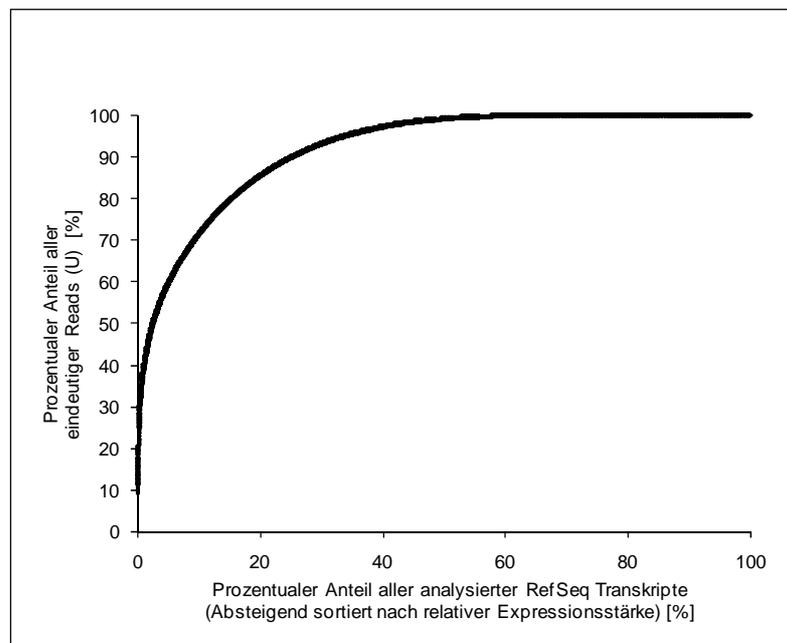
### 3.4.4 Abdeckung des Transkriptoms und Fehlerquellen

Ein weitere Frage, die für die Bewertung dieser Methode als Instrument zur genbasierten Detektion von Mutationen von entscheidender Bedeutung war, war die nach der Abdeckung des Transkriptoms. Unter der Annahme, dass die RefSeq Datenbank in ihrer Größe mit ca.  $57,56 \times 10^6$  bp der des vollständig transkribierten Transkriptoms entspricht, ist die erreichte Anzahl sequenzierter Basen von ca.  $79,40 \times 10^6$  (ohne QC Reads, Reads schlechter Qualität) prinzipiell ausreichend. Dies ist insbesondere dann der Fall, wenn man beachtet, dass im Falle der Wildtyp Probe für nur 14.434 der 20.904 untersuchten RefSeq eindeutige Reads (U0, U1, U2) und damit Expression nachgewiesen werden konnte (3.4.3).

Bedeutender für die Frage nach der Abdeckung ist die jedoch die Tatsache, dass die cDNA Moleküle der Transkripte aufgrund des unterschiedlich starken Expressionslevels der jeweiligen Gene der Sequenzierung in unterschiedlichen Anzahlen zur Verfügung stehen. Sortiert man die 14.434 RefSeq Transkripte (F1 WT) nach der Anzahl der U Reads (unique, eindeutige Treffer; vergl. Abbildung 23), so erkennt man bereits, dass die Abdeckung

deutlichen Schwankungen unterliegen muss. Als höchster Wert wurden 51.889 U Reads für eine RefSeq Sequenz und als niedrigster 1 U Read gemessen. Der Median betrug 23 eindeutige Reads pro Transkript, der Mittelwert jedoch 92,4 eindeutige Reads ungeachtet der Länge der Transkripte.

Um einen detaillierten Eindruck der Abdeckung gewinnen zu können, wurde zunächst die relative Expression der Transkripte bestimmt. Als Einheit wurde der Quotient aus der Anzahl von U-Reads für das jeweilige Transkript und der Länge des Transkripts in Basen gewählt, da ein eindeutiger Zusammenhang von Transkriptlänge zu der Anzahl von U reads festgestellt werden konnte (Rosenkranz et al., 2008). Die Entscheidung für die Arbeit mit dieser relativen Transkription wurde zusätzlich dadurch bestätigt, dass ein Vergleich der Transkriptionsstärken, die mittels Solexa Sequenzierung und mittels Illumina Microarrays ermittelt wurden, einen Pearson Korrelationskoeffizienten von 0,7 ergaben. Ähnliche Vergleiche unterschiedlicher Techniken ergaben deutlich schlechtere Koeffizienten (van Ruissen et al., 2005; Weber et al., 2007). Die hier erhaltenen Ergebnisse entsprachen sogar denen, die mit technischen Replikaten in den Vergleichen verschiedener Microarray Plattformen erzielt wurden (Kuo et al., 2006). Dadurch konnte gezeigt werden, dass sich mittels Solexa Sequenzierung die relative Häufigkeit von Transkripten und damit die Expressionstärke erfassen lassen (Rosenkranz et al., 2008).



**Abbildung 24: Abdeckung des Transkriptoms nach Sequenzierung nicht normalisierter cDNA**

Die Abbildung zeigt den prozentualen Anteil der eindeutigen Reads der cDNA der F1 Wildtyp Stammzelllinie, die auf den Anteil betrachteter Transkripte entfallen, wenn diese absteigend nach Expressionstärke sortiert werden. Insgesamt wurden 1.333.960 U Reads und 20.804 RefSeq Transkripte betrachtet. Für 14.434 (ca. 69%) RefSeq Transkripte konnten eindeutige homologe Sequenzen ermittelt werden.

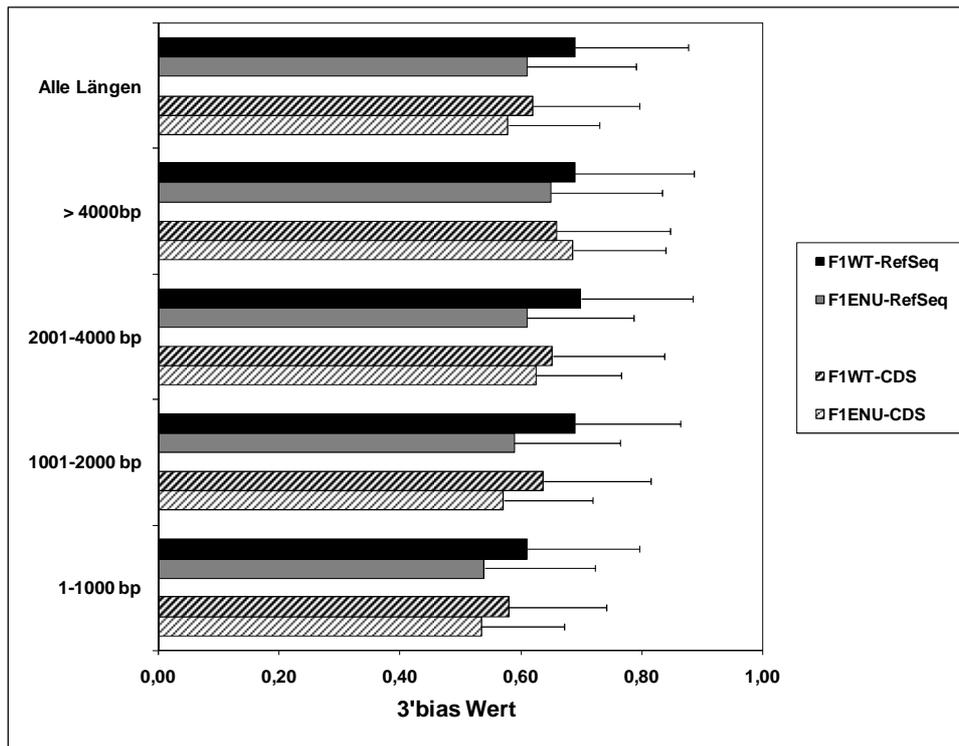
Für die Untersuchung der Abdeckung des Transkriptom wurden die Transkripte nach relativer Transkriptionsstärke (U Reads pro kb) absteigend sortiert und der prozentuale Anteil aller bis zu der angegebenen Transkriptionsstärke komplementären U Reads graphisch dargestellt. Repräsentativ für die beiden betrachteten Stammzellklone (Wildtyp und Mutante RRx006) sind die Ergebnisse der Wildtyp Sequenzierung in Abbildung 24 aufgeführt.

Die Abbildung zeigt anschaulich, dass eine Abdeckung des Transkriptom nicht annähernd erzielt werden konnte. Bereits 50% der eindeutigen U Reads fallen auf nur 2.6%, 75% der U Reads auf nur 12.2% der am stärksten transkribierten RefSeq Sequenzen. Folglich ist damit zu rechnen, dass bei schwach exprimierten Genen mit nur wenigen eindeutigen Reads eine Detektion von Exon-Deletion basierend auf dem unter 3.4.1 erläuterten Verfahren unter Verwendung nicht normalisierter cDNA nicht möglich ist.

Neben der Abdeckung des Transkriptom wurde weiterhin betrachtet, inwieweit eine Verteilung der einzelnen Reads innerhalb eines Transkripts gleichmäßig erfolgt. Aufgrund der Tatsache, dass für die Vorbereitung der ds-cDNA die RNA einer poly(A)<sup>+</sup> Selektion unterzogen wurde und die Synthese ausgehend von OligodT Primern stattfand, um den erheblichen Anteil anderer RNA Moleküle zu reduzieren, wurde mit einer bevorzugten Lage der Reads am 3'-Ende des Transkripts (3'-bias') gerechnet. Dieser Fehler sollte aufgrund der eingeschränkten Prozessivität der M-MLV Reversen Transkriptase insbesondere bei längeren Transkripten zum Tragen kommen. Zur Analyse dieser möglichen Fehlerquelle wurde die Position innerhalb des Transkripts (5'-Position auf der Transkriptsequenz) für jeden eindeutigen Read ermittelt und in Zahlen gefasst (Quotient aus Position in bp und Länge des Transkripts). Für jedes Transkript wurden diese Werte gemittelt, sodass ein Wert nahe 0 eine absoluten 5'-bias' und ein Wert nahe 1 einen 3'-bias' signalisierte. Um der Annahme Rechnung zu tragen, dass längere Transkripte einen verstärkten 3'-bias' aufweisen, wurden die Transkripte zusätzlich in Gruppen verschiedener Länge eingestuft. Die Untersuchung wurde sowohl mit RefSeq Sequenzen, die Teile der 3'- und 5'- untranslatierten Bereiche (UTR) enthalten, als auch mit den reinen codierenden Sequenzen (CDS) durchgeführt, um den Einfluss der UTR Bereiche erfassen zu können (Abbildung 25).

Die Ergebnisse des Experiments belegen die Annahme, dass eine verstärkte Anzahl der Reads vom 3'-Ende des Transkripts stammen. Dabei läßt sich erwartungsgemäß ein verstärkter Trend zum 3'-bias' bei längeren Transkripten beobachten. Unter Verwendung der CDS als Referenz fällt der 3'-bias' etwas geringer aus, sodass man auf eine leichte Anreicherung der Reads in der 3'-UTR schließen kann. Die Ergebnisse zeigen, dass die

Wahrscheinlichkeit für die Detektion einer putativen Exon-Deletion nach dem beschriebenen Verfahren im 5'-Bereich des Transkriptes gegenüber dem 3'-Bereich generell niedriger einzustufen ist. Doch wurde diese Untersuchung ungeachtet der relativen Expression des Gens und damit der Abdeckung des Transkriptes durchgeführt. Im Vergleich zur Abdeckung des Transkriptes ist diese Fehlerquelle mit Hinblick auf die Wahrscheinlichkeit der erfolgreichen Detektion einer Mutation allerdings als untergeordnet anzusehen.



**Abbildung 25: Untersuchung des 3'-bias' als mögliche Fehlerquelle**

Die Abbildung beschreibt die durchschnittliche Position aller eindeutigen Reads innerhalb eines Transkripts. Die 5'Anfangsposition auf der Referenzsequenz eines jeden Reads in bp wurde dividiert durch die Länge des Transkripts in bp. Anschließend wurde für jedes Transkript der Mittelwert der Read spezifischen Werte ermittelt. Dargestellt sind die Mittelwerte der Transkriptspezifischen Werte einer jeden Gruppe. 0 = absoluter 5'-bias, 1 = absoluter 3'-bias' **CDS**: Codierende Sequenz des jeweiligen Gens als Referenz; **RefSeq**: Refseq Transkript als Referenz.

Weiterhin wurde der GC-Gehalt als mögliche Einflussgröße auf die Verteilung der Reads in Betracht gezogen, da dieser bereits für SAGE und MPSS Untersuchungen beschrieben wurde (Margulies et al., 2005; Siddiqui et al., 2006). Dazu wurde der GC-Gehalt der Referenzsequenzen gegen den durchschnittlichen GC-Gehalt der komplementären Reads betrachtet. Die Referenzsequenzen wurden dabei entsprechend ihrer relativen Expressionsstärke als Maß für die Abdeckung gruppiert. Repräsentativ sind die mit der Wildtyp Linie erhaltenen Werte nachfolgend aufgeführt (Tabelle 1).

Generell konnte festgestellt werden, dass je höher die relative Expression des Transkriptes war, desto höher lag der durchschnittliche GC-Gehalt der Reads. Dieser Trend ist nur schwer

zu verstehen, doch ist sicher, dass diese Dimension der Einflussgröße im Vergleich zu der der Gesamtabdeckung zu vernachlässigen ist.

**Tabelle 19: Durchschnittlicher GC-Gehalt (%) von RefSeq und komplementären Reads**

	Anzahl RefSeq Transkripte	Anzahl Reads	Mittlerer GC- Gehalt (%) RefSeq	Mittlerer GC- Gehalt (%) Reads
Alle Reads		3052501		50,30
Alle U Reads	14.434	1.333.960	50,36	49,63
>1000 Reads/RefSeq	146	487.730	50,41	52,26
1000-101 Reads/RefSeq	2.356	548.963	50,21	48,66
100-11 Reads/RefSeq	6.685	277.076	50,44	47,41
10-1 Reads/RefSeq	5.247	20.191	50,34	43,21

Die Beeinflussung der Ergebnisse durch mögliche Kontaminationen mit genomischer DNA ist für die Detektion von Exon-Deletionsmutanten auf Basis der Exon-Deletionsdatenbank eher als gering einzustufen. Dennoch wurde diese Möglichkeit untersucht insbesondere mit Hinblick auf die Qualität der ermittelten relativen Expression der Gene. Hierzu wurde die Verteilung von Reads innerhalb zweier großer genomischer Bereiche untersucht, die keine Protein kodierenden Gene beinhalten (Gene-deserts). Für diese beiden Bereiche konnte gezeigt werden, dass Deletionen auf Chromosom 3 (1.511 kb) und auf Chromosom 19 (845 kb) zu phänotypisch unauffälligen Mäusen führen (Nobrega et al., 2004). Aus diesem Grund wurde ein ELAND Alignment des F1-WT Datensatzes gegen das Maus Genom durchgeführt und Reads innerhalb dieser Bereiche gezählt. Insgesamt wurden 40 Reads in dem Bereich auf Chromosom 3 und 19 Reads auf Chromosom 19 festgestellt. Hieraus errechnet sich eine durchschnittliche Kontamination mit genomischer DNA von einem Read pro 44,9 kb genomischer DNA.

### 3.4.5 Durchmusterung der Sequenzdatensätze nach Exon-Deletionen

Ungeachtet der Tatsache, dass die Abdeckung des Transkriptoms mit sequenzierten Fragmenten (Reads) generell zu gering erschien, um eine Auflösung zu erreichen, die es erlauben sollte nahezu alle Exon-Deletions Ereignisse aufzuspüren, wurden die vorhandenen Datensätze dahingehend analysiert. Für relativ stark exprimierte Gene mit hoher Abdeckung sollte weiterhin die Möglichkeit der Detektion gegeben sein.

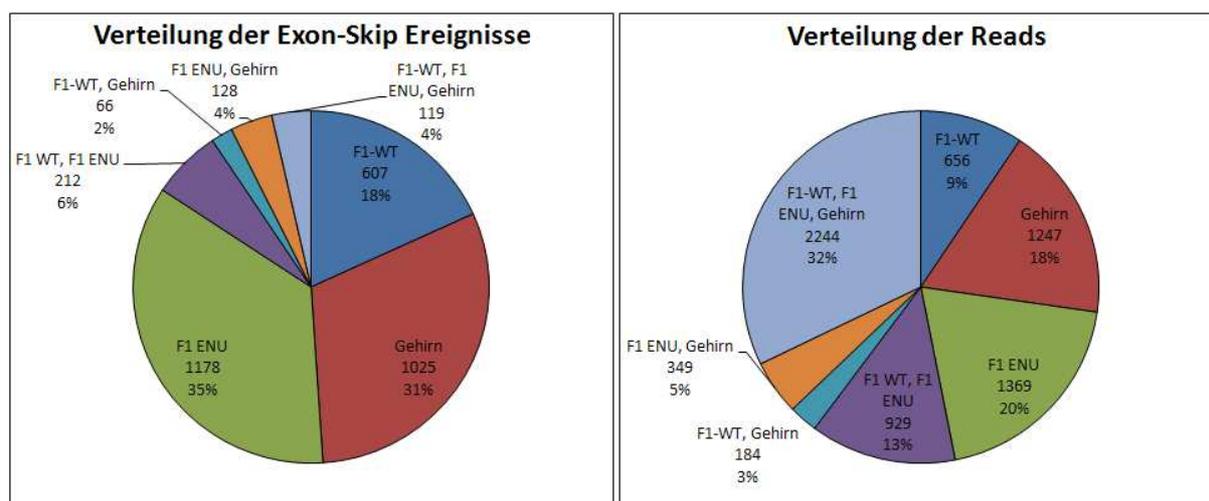
Zusätzlich zu den Datensätzen der beiden F1 Stammzellklone (Wildtyp und RRx006) wurde ein weiterer Datensatz analysiert. Dieser stammte aus der parallelen Sequenzierung einer ds-cDNA aus einer mRNA Probe des Gehirns einer 12 Wochen alten männlichen C57BL/6J Maus. Diese Probe wurde als unabhängige Referenzprobe in einer hier nicht weiter beschriebenen Analyse zur Vergleichbarkeit von Expressionsdaten, die mittels Microarray

Analysen und Sequenzierung gemessen wurden, genutzt (Rosenkranz et al., 2008). Aus diesem Grund wurden die Datensätze hier wiederum als unabhängige Referenz verwendet. Alle drei Datensätze wurden mittels ELAND gegen die db42 Datenbank analysiert (3.4.2). Die Rohergebnisse sind in der nachfolgenden Tabelle dargestellt (Tabelle 20). In die Auswertung konnten nur eindeutige U-Reads einbezogen werden, da Reads der Kategorie R nicht eindeutig zugeordnet werden können (vergl. Abbildung 23).

**Tabelle 20: ELAND Abgleich der Sequenzdaten gegen die db42 Datenbank**

Art der Reads	F1 Wildtyp		F1 ENU RRx006		C57BL/6J Gehirn	
	Anzahl Reads	%	Anzahl Reads	%	Anzahl Reads	%
<b>Gesamt</b>	3.052.501	100	3.268.410	100	2.839.833	100
<b>NM</b>	2.938.155	96,25	3.167.724	96,92	2.717.028	95,68
<b>QC</b>	112.323	3,68	97.767	2,99	120.700	4,25
<b>R0</b>	0	0,00	0	0,00	1	0,00
<b>R1</b>	5	0,00	3	0,00	2	0,00
<b>R2</b>	18	0,00	12	0,00	28	0,00
<b>U0</b>	306	0,01	451	0,01	465	0,01
<b>U1</b>	293	0,01	464	0,01	289	0,01
<b>U2</b>	1401	0,05	1989	0,06	1320	0,05

Die Gesamtanzahl eindeutiger Treffer betrug demnach für die Wildtyp Probe 2000 Reads, für die ENU Probe 2904 Reads und für die Referenz Gehirn Probe 2074 Reads. Da die Gesamtzahl nichts über die Anzahl der putativen Spleiß- bzw. Deletionsereignisse aussagt, wurde in einem weiteren Schritt ein Abgleich der drei Proben durchgeführt.



**Abbildung 26: Verteilung der möglichen Exon-Skip Ereignisse auf die Proben und die Reads**

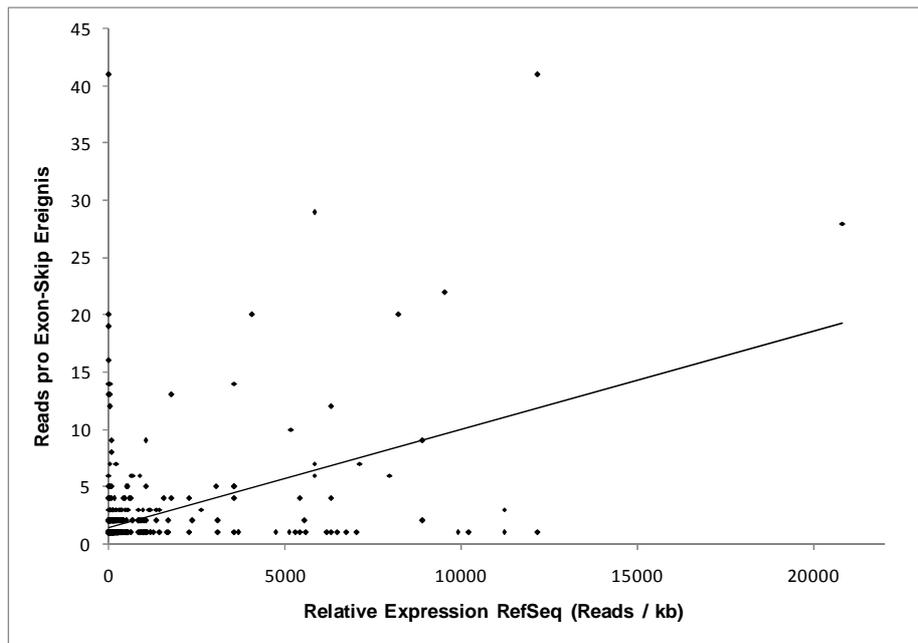
Das linke Diagramm zeigt die Verteilung der insgesamt 3335 detektierten verschiedenen Exon-Skip Ereignisse (ELAND Analyse gegen db42) auf die einzelnen Proben bzw. Gruppen von Proben (Anzahl und Anteil in %). Im rechten Diagramm werden die Anzahlen und Anteile der eindeutigen Reads, die auf die Gesamtheit der Skip-Ereignisse in der jeweiligen Probe bzw. Proben-Gruppe entfallen, dargestellt. In der Probe 'F1 ENU' sind so beispielsweise nur die exklusiv in dieser Probe erkannten Ereignisse zusammengefasst, in der Gruppe 'F1-WT, F1-ENU' alle Ereignisse, die in den beiden Proben 'F1-WT' und 'F1 ENU' aber nicht in der Probe 'Gehirn' auftreten. **F1-WT**: Probe der F1 Wildtyp ES Zelllinie; **F1 ENU**: Probe des F1 ENU RRxx006 ES-Zell Klons; **Gehirn**: Probe des C57BL/6J Gehirns.

Das linke Diagramm in Abbildung 26 zeigt, dass 84% aller 3.335 putativer Exon-Skip Ereignisse exklusiv in den einzelnen Proben ermittelt wurden. Ein zusätzlicher Abgleich mit BLAT gegen das Maus Genom bestätigte, dass es sich bei diesen Sequenzen nicht um Kontamination mit genomischer DNA handelte. 16% der Ereignisse sind in mehr als einer Probe feststellbar. Erwartungsgemäß ist hier die größte Übereinstimmung mit 212 Ereignissen (6%) bei den beiden F1 ES-Zell Proben zu erkennen. Weiterhin fällt auf, dass die Anteile an den Exon-Skip Ereignissen in der F1 ENU und der Gehirn Probe nahezu doppelt so hoch waren wie die in der F1 Wildtyp Probe. Letzteres steht im Einklang mit der Beobachtung, dass alternatives Spleißen in Gehirnzellen weit verbreitet ist (Stamm et al., 2005).

Der rechte Teil der Abbildung 26 stellt die Verteilung der eindeutigen Reads auf die nach Auftreten in den Proben gruppierten Skip-Ereignisse dar. Dabei zeigt sich eine eindeutige Umverteilung der Anteile im Vergleich zu der linken Darstellung. Die Anteile an Reads, die auf die Proben Gruppen F1 WT + Gehirn und F1 ENU + Gehirn fallen, sind als einzige kaum verändert. Die Gruppe, welche die Exon-Skip Ereignisse repräsentiert, die parallel in allen drei Proben zu erkennen sind, weist die höchste Anzahl an Reads (2.244) auf. Dies war in einem gewissen Rahmen zu erwarten, da hier Reads von drei Proben zusammengerechnet wurden. Dennoch ist die durchschnittliche Anzahl von 6,28 Reads pro Exon-Skip Ereignis und Klon deutlich höher als die in den einzelnen Proben und der Proben-Gruppe der ES-Zelllinien (F1 WT: 1,99 Reads; F1 ENU: 1,77 Reads; Gehirn: 1,55 Reads; F1 WT+ENU: 2,19 Reads). Die Ursache dafür liegt darin begründet, dass es wahrscheinlicher ist ein den drei Proben gemeinsames Spleiß-Ereignis zu ermitteln, wenn generell viele Reads pro Ereignis und Probe auftreten. Letzteres kann allgemein in der Expression des Gens und/oder in dem Vorkommen der Spleißvariante begründet liegen.

Um den Zusammenhang der Anzahl der Reads für ein spezifisches Ereignis und der relativen Expression in Zahlen zu fassen, wurde die Korrelation der Anzahl der Reads mit der unter 3.4.3 ermittelten relativen Expression der entsprechenden RefSeq untersucht. Für die Probe der F1 Wildtyp ES-Zelllinie betrug der Pearson'sche Korrelationskoeffizient  $R = 0,43$ , für die Probe der Mutante RRxx006  $R = 0,41$  und für die C57BL/6J Gehirnzell-Probe  $R = 0,48$ . Alle drei Korrelationskoeffizienten wiesen damit auf eine gewisse Korrelation hin (Abbildung 27). Für die Differenzierung zwischen einem Exon-Deletionsereignis und alternativem Spleißen in der ENU mutierten Probe RRx006 muss ein Vergleich mit der Wildtyp Probe F1-WT erfolgen. Doch stellt sich hier die Frage wann mit Sicherheit davon ausgegangen werden kann, dass es sich um ein Deletionsereignis handelt. Für die Mutante RRx006 lassen sich 1.178 exklusive Exon-Skip Ereignisse ermitteln, von denen allerdings 1.037 (88,0%) durch

nur einen einzigen Read repräsentiert werden. Es ist daher fraglich, ob in der Wildtyp Probe für dieses Ereignis nur aufgrund der Tatsache, dass es sich um eine schwach ausgeprägte Variante alternativen Spleißens handelt, kein Read erzeugt wurde. Statistische Test machten an dieser Stelle keinen Sinn, da die Stichprobengröße für die einzelnen Ereignisse nur sehr gering waren (1 Read, 1 Ereignis) und verschiedenste Größen Einfluss auf die Detektion eines Skip Ereignisses haben konnten.



**Abbildung 27: Korrelation zwischen der Anzahl Exon-Skip spezifischer Reads und relativer Expression**

In der Abbildung wird graphisch die Anzahl Exon-Skip spezifischer Reads zu der relativen Expression der zugehörigen RefSeq Sequenz (3.4.3) der F1-WT Probe in Beziehung gesetzt. Die Linie verkörpert die den Daten zugrunde liegende Trendlinie.

Zum eindeutigen Nachweis, ob es sich um eine Deletion oder alternatives Spleißen handelt, hätten mit allen exklusiven 1.178 Ereignissen der Probe F1 ENU weitere molekularbiologische Untersuchungen durchgeführt werden können, für deren Ausmaße diese grundlegende Betrachtung des Sequenzierverfahrens jedoch nicht ausgelegt war. Die andere Alternative war der Abgleich der betroffenen Exon-Skip Referenzsequenzen der db42 mit der EST Datenbank der Maus (dbEST). Dadurch wäre es möglich gewesen, bekannte alternative Spleiß-Ereignisse auszusortieren und so die Wahrscheinlichkeit für ein Exon-Deletionsereignis zu erhöhen. Doch liegen die Nachteile dieser Vorgehensweise auf der Hand: Zum einen kann es sich bei dem putativen Deletionsereignis immer noch um ein bisher nicht in dbEST erfasstes alternatives Spleißen eines Exons handeln. Zum anderen gehen durch die Selektion mögliche positive Exon-Deletions Kandidaten verloren. Für die 119 Skip-Ereignisse, für die in allen drei Proben Reads auftraten, konnte durch einen Abgleich gegen

dbEST mittels ELAND gezeigt werden, dass diese zu 100% in dbEST bereits erfasst waren. Bei einem Abgleich der exklusiven Ereignisse der Wildtyp ES-Zell Probe wurden für 581 der 607 Ereignisse Treffer in dbEST erzielt. Bei den restlichen 26 (4,3%) Ereignissen handelt es sich daher, abgesehen von Fehlern im Zuge des Spleißens der prä-mRNA und unter Berücksichtigung des Wildtyp Hintergrunds, um bisher in dbEST nicht erfasste Ereignisse alternativen Spleißens. Alle 26 Ereignisse wurden allerdings jeweils nur durch einen Read repräsentiert. Als Beispiel sei hier das Exon ENSMUSE00000102682 genannt, welches bis dato als konstitutives Exon in allen drei Transkripten des Gens *Slc36a1* unter ENSEMBL geführt wird. Erstaunlich dabei ist, dass die Entfernung dieses Exons (Länge 91 bp) zu einer Veränderung des Leserasters führt. Ein Abgleich der Sequenzen der exklusiven Ereignisse (1.178) der Mutante RRx006 gegen dbEST ergab Treffer für 1.124 Sequenzen, 54 (4,5%) waren unbekannt. Doch kann hier, wie bereits an anderer Stelle ausführlich erläutert wurde, keine Aussage dazu getroffen werden, ob diese Ereignisse Spleiß-Deletionen darstellen oder nicht. Für die Treffer der Mutante RRx006 bedeutet dies, dass trotz überzeugender Fakten, der Sequenz, und dem damit belegten natürlichen Vorkommen dieser bis dato in EST Datenbanken unbekannt Sequenz, experimentell im Vergleich gezeigt werden muss, dass es sich um ein auf die Mutation zurückzuführendes Ereignis handelt.

Aus diesem Grund wurden exemplarisch 13 (~25%) der 54 putativen Exon-Deletions Ereignisse in der Mutante RRx006 zufällig ausgewählt und mit Primersystemen analog zu 3.3.7 entsprechende PCRs durchgeführt.

**Tabelle 21: Ergebnisse der PCR zur Bestätigung möglicher Exon-Deletions Mutationen**

Primersystem F und Primersystem S vergl. Abbildung 17. Wt: Es wurde nur eine Bande in der erwarteten Größe des Wildtyp PCR-Produktes ermittelt; Mut: Es wurde nur eine Bande in der erwarteten Größe des mutierten PCR-Produktes ermittelt; Wt+Mut: Es wurden zwei Banden entsprechend den erwarteten Größen des Wildtyp und des kleineren mutierten PCR-Produktes ermittelt; '-': Kein PCR-Produkt mit dem Skip Primersystem erzielt; '+': PCR-Produkt der erwarteten Größe mit dem Skip Primersystem erzielt; \*Das Bandenmuster trat sowohl unter Verwendung von WT als von RRx006 cDNA auf. In den anderen Fällen konnte das Bandenmuster nur bei Verwendung von RRx006 cDNA als PCR Matrize erhalten werden.

Sonden Name	MGI Symbol	RefSeq ID	Exon Nr.	Primersystem F	Primersystem S
skip1.003700	Fn1	NM_010233	26	Wt	-
skip1.014138	Tpd5211	NM_009413	2	Wt	-
skip1.016954	Atp2b1	NM_026482	3	Wt	-
skip1.018778	Csrp2	NM_007792	5	Wt + Mut*	-
skip1.105065	Sec31b	NM_001033343	20	Wt + Mut	-
skip1.121294	Osbpl2	NM_144500	7	Wt	-
skip1.169341	Tpk1	NM_01386	6	Wt + Mut	+
skip1.177973	Ap2a2	NM_007459	4	Mut	+
skip1.197421	Cox4i1	NM_009941	4	Wt	-
skip1.203294	0610040D20Rik	NM_023153	5	Wt	-
skip1.211338	Rps271	NM_026467	2	Wt	-
skip1.212648	Ttc12	NM_172770	2	Wt + Mut	+
skip1.212923	Grit	NM_177379	10	Wt	-

Für 4 der 13 betrachteten möglichen Exon-Deletionsereignisse konnte mit dem Primersystem F in der PCR mit RRx006 cDNA als Matrize das für eine Mutation charakteristische Bandenmuster erhalten werden. Unter Verwendung von F1 Wt cDNA war in allen Fällen nur die Wildtyp Bande entsprechender Größe erkennbar. Dies bestätigte in diesen Fällen das Vorkommen einer ENU induzierten Exon-Deletion. In drei der vier Fälle konnte zudem in der Probe der Mutante und nicht in der Wildtyp-Probe der Nachweis mit dem Skip-Primersystem erfolgen. Für das Skip-Ereignis skip1.018778 wurden mit beiden Matrizen cDNAs (WT + ENU) und dem Primersystem F jeweils zwei Banden der erwarteten Größen erhalten. Dies zeigt, dass Exon 5 des Gens *Crsp2* alternativ gespleißt wird. Die deutlich geringere Intensität der Bande des kleineren ‚mutierten‘ Produktes (Bild nicht gezeigt) gab einen Hinweis auf wesentlich schwächere Expression dieser Variante.

Abschließend ist festzuhalten, dass die als Positiv-Kontrolle angedachte Deletion des Exon 8 des Gens *Nup37* in der Mutante RRx006 (3.3.6) nicht identifiziert werden konnte. Dies ist allerdings auch nicht verwunderlich, da für die 1288 bp lange RefSeq Sequenz des Transkripts *Nup37* insgesamt nur 10 eindeutige 27 bp lange Reads ermittelt werden konnten (Ergebnis der Untersuchung aus 3.4.3).

Grundsätzlich kann festgehalten werden, dass mit der Probe der Mutante RRx006 bei sonst gleichen Versuchsbedingungen deutlich mehr Treffer für Exon-Skip Ereignisse erzielt wurden als mit der Wildtyp Probe. Dennoch ist die statistische Sicherheit mit der sehr geringen Anzahl von Reads pro putativem Deletions-Ereignis viel zu gering, um von einer eindeutigen Deletion sprechen zu können.

## 4 Diskussion

Mit dieser Arbeit sollte der Versuch unternommen werden, neuartige Screeningverfahren zur Detektion von Exon-Deletionsmutationen in embryonalen Stammzellen der Maus zu entwickeln. Die Notwendigkeit für die Entwicklung eines neuartigen Verfahrens entstand aus der Motivation heraus, eine Bibliothek ENU mutierter Stammzellen der Maus generieren zu wollen, die für nahezu jedes beliebige Gen ES-Zell Klone mit Exon-Deletionsmutationen enthält. Diese ES-Zell Bibliothek sollte Forschern in den verschiedensten Bereichen biomedizinischer Forschung als Stammzellressource zur schnellen Generierung interessanter Mausmutanten zur Verfügung gestellt werden.

Exon-Deletionsmutanten können in vielen Feldern der Forschung als Modellsysteme genutzt werden. Strukturell betrachtet können Exon-Deletionen zu der Einführung vorzeitiger Stop-Codons, zur Veränderung der Primärstruktur von Proteinen und zur Veränderungen der 5' und 3'-UTR Bereiche führen. Daraus resultierende Nullmutanten könnten beispielsweise direkt zur Generierung einer Knockout Maus verwendet werden. Veränderungen in der Primärstruktur können durch die Deletion von Proteindomänen interessante Phänotypen hinsichtlich der Bindeeigenschaften, der intrazellulären Lokalisation, der enzymatischen Aktivität, der Stabilität und/oder der Glycolisierungs- und Phosphorylierungsmuster von Proteinen verursachen. Das sind Eigenschaften, die von hohem Interesse in der biologischen und pharmazeutischen Forschung sind, deren genaue Analyse in der Regel jedoch eine Zeitintensive Generierung von Modellsystemen voraussetzt (Stamm et al., 2005). Zudem werden mit fehlerhaftem Spleißen eine Vielzahl menschlicher Krankheiten in Verbindung gebracht (Faustino et al., 2003; Philips et al., 2000), für deren Untersuchung die Klone der angedachten Bibliothek eine wertvolle Ressource darstellen könnte.

Die Durchmusterung ENU mutierter Stammzellen zur Generierung einer derartigen Bibliothek setzt allerdings eine Technik voraus, die es erlaubt, im Hochdurchsatzverfahren tausende von Klonen auf mögliche Mutationen zu überprüfen. Von besonderer Bedeutung sind daher der Arbeitsaufwand und die Kosten, die ein derartiges Screening mit sich bringen würde.

Zu Beginn dieser Studie lag eine Arbeit vor in der mittels einer PCR basierten Technik im Hochdurchsatzverfahren eine Bibliothek von ca. 40.000 ENU mutierten ES-Zellen nach Exon-Deletionsmutanten durchmustert wurde (Greber et al., 2005). Diese Technik ist sehr gut geeignet um nach Mutationen in einem bestimmten Gen zu suchen. Die Durchmusterung einer vergleichbar großen Bibliothek nach beliebigen Deletionsmutanten im gesamten

Transkriptom würde allerdings trotz überschaubarer Materialkosten einen nicht tragbaren Arbeitsaufwand bedeuten (Etablierung der PCR-Systeme, Durchführung der PCR-Reaktionen). Aus diesem Grund galt es die Umsetzbarkeit eines Screeningverfahrens basierend auf der Microarray-Technologie zu untersuchen, da dieses Hybridisierungsverfahren die simultane Analyse tausender Sequenzen ermöglichte. Zum Ende der Arbeit rückten im Zuge der technischen Weiterentwicklung immer mehr preiswerte Hochdurchsatz-Sequenzierverfahren in den Fokus der Wissenschaft, die nach Auffassung vieler Wissenschaftler mit der Zeit Array-basierte Verfahren ablösen werden (Metzker 2009). Aus diesem Grund wurden die Grundlagen für eine Verwendung dieser Sequenzierungstechnik ebenfalls betrachtet.

#### 4.1 ENU-Mutagenese

Grundlage für die Entwicklung der Durchmusterungstechniken sollte eine Bibliothek ENU mutierter Stammzellen mit entsprechenden Positiv-Kontrollen bilden. Die Mutationsfrequenz der Test-Bibliothek sollte bereits der künftig zu durchmusternden Bibliothek entsprechen, damit das zu entwickelnde System auch problemlos übertragen werden kann. Für das künftige System wurden Mutationsfrequenzen (*Hprt*-Test, vergl. 1.5.1) von 1:2000 bis 1:2500 avisiert, da sich dieser Frequenzbereich in der Arbeitsgruppe bereits als geeignet erwiesen hatte (Greber et al., 2005). Generell sind mit ENU-Mutagenese Mutationsfrequenzen von 1:1000 und niedriger beschrieben worden (Chen et al., 2000; Vivian et al., 2002). Neben dem Vorteil der höheren Detektions-Wahrscheinlichkeit für ein beliebiges Mutationsereignis, haben höhere Frequenzen aber den Nachteil, dass das eigentliche Ziel, die Generierung von Mäusen aus den ES-Zellen über den Weg der Blastocysten-Injektion, deutlich erschwert wird. Dies zeigte sich beispielsweise an der Untersuchung von Vivian et al. (2002), die versuchten für die Gene *Smad2* und *Smad4* mittels ENU-Mutagenese Allele zu erzeugen, da die knock-out Modelle alle embryonal lethal waren. Von 20 ENU mutierten ES-Zell Klonen (Mutationsfrequenz  $< 10^{-3}$ ) mit nachgewiesener Mutation in einem der beiden Gene waren nur 13 karyotypisch normal. Von diesen 13 Klonen konnte wiederum nur mit 5 Klonen eine Keimbahntransmission erzielt werden. Als Ursache dafür kann die hohe Mutationsfrequenz angesehen werden. Zudem erhöht eine höhere Mutationsfrequenz auch die Anzahl der nötigen Rückkreuzungen zur Entfernung weiterer Mutationen nach erfolgreicher Keimbahntransmission.

Daher wurden zunächst Mutagenesen durchgeführt, mit dem Ziel die Mutationsbedingungen so weit kontrollieren zu können, dass die gewünschte Mutationsfrequenz im *Hprt*-Test um

1:2500 betrug. Der einfacheren Zellkulturbedingungen halber wurden diese Versuche mit der Zelllinie CGR8 durchgeführt (3.1.1). Die daraus gewonnenen Erfahrungen wurden auf Mutagenesen der Zelllinie F1 (C57BL/6Jx129Sv) übertragen, da diese Zelllinie für die Erstellung der Klon-Bibliothek vorgesehen worden war. Der heterogene genetische Hintergrund wurde in Kauf genommen, da mit dieser Zelllinie in dem Institut bereits sehr gute Ergebnisse bezüglich der Keimbahngängigkeit in knock-out Experimenten erzielt wurden. Zudem beträgt der Anteil an SNPs, die in der codierenden Sequenz von Genen lokalisiert sind, zwischen den beiden Inzuchtlinien nur ca. 1,5% aller SNPs (Adams et al., 2005). Insgesamt kann die Anzahl an SNPs in codierender Sequenz auf ca 64.000 geschätzt werden. Der Einfluss auf die sequenzbasierte Detektion von Exon-Deletions Mutationen nach den in dieser Arbeit verfolgten Verfahren wurde daher als gering eingestuft.

Bei der Mutagenese der F1-Zelllinie galt es herauszufinden, ob die Hybridzelllinie unter Verwendung gleicher Mutagenese-Bedingungen auch entsprechende Mutationsfrequenzen und Mutationsspektren aufweisen würde. Unterschiedliche Reaktionen auf ENU wurden von verschiedenen Laboren bei der Mutagenese ganzer Tiere verschiedener genetischer Hintergründe beobachtet (Justice et al., 2000). Die erzielten Überlebensraten, Mutationsfrequenzen (3.1.1) und Mutationsspektren (3.1.2) entsprachen allerdings denen der Inzucht Zelllinie CGR8 (129Sv) und den Angaben der Literatur (Barbaric et al., 2007; Chen et al., 2000). Jede groß angelegte Mutagenese lieferte zudem ausreichend Zellmaterial, um daraus später problemlos ES-Zell Archive mit mehr als 250.000 Klonen generieren zu können, ohne erneute Mutagenesen durchführen zu müssen.

Bei der Bestimmung der Mutationsspektren fiel auf, dass drei funktionelle Punktmutationen des *Hprt*-Lokus in mehreren Klonen auftraten. Die Redundanz ist auf die Passagen vor der eigentlichen 6-TG Selektion zurückzuführen, die nötig waren um noch vorhandenes HPRT Protein auszudünnen (2.2.8). Bei der Insertion in Klon RR1 war davon auszugehen, dass eine Punktmutation auf DNA-Level ein Spleißsignal erzeugt bzw. ein kryptisches Signal aktiviert hat. Der Klon RR6 war 6-TG resistent und die PCR lieferte eine schwach ausgeprägte zusätzliche Bande, ca. 200bp kleiner als die WT Bande, dennoch ließ sich keine Mutation auf Transkript Ebene feststellen. Eine DNA-Sequenzierung hätte hier sicher Aufschluss geben können, war aber nicht zielführend für das Projekt. Es konnten zwei Exon-Deletionsmutanten identifiziert werden, in denen exakt Exon 6 bzw. Exon 8 fehlte, was jeweils in der Ausbildung einer Nonsense Mutation resultierte. Dies bestätigte die Angaben der Literatur, in der ausnahmslos die basengenaue Deletionen vollständiger Exons beobachtet wurden (Barbaric et al., 2007). Diese Klone waren sehr willkommen, da sie als Kontroll-Basis zur Entwicklung

des Micorarray basierten Detektionssystems an dem Referenzgen *Hprt* eingesetzt werden konnten. Die Arbeit mit *in situ* erstellten Exon-Skip Konstrukten und den daraus resultierenden 'künstlichen' Reaktionsbedingungen konnte so verhindert werden.

Als Vorbereitung für kommende Test-Durchmusterungen zur Etablierung der Detektionsverfahren wurde eine kleine Test-Bibliothek ENU mutierter F1 ES-Zell Klone aus der Mutagenese F1-1 erstellt, da für diese Mutagenese die höchste Mutationsfrequenz von  $4,82 \times 10^{-4}$  am *Hprt*-Lokus ermittelt wurde. An dieser Stelle muss darauf hingewiesen werden, dass es sich bei den einzeln gepickten Kolonien der Test-Bibliothek nicht wirklich um reine Klone identischen genetischen Hintergrunds handelt, da immer nur ein Strang der Doppelhelix an einer bestimmten Position durch ENU ethyliert wird (Abbildung 4B) (Noveroske et al., 2000). Durch Trennung der Chromatiden in der Mitose entstehen so verschiedene Klone innerhalb einer Kolonie. Bei den ermittelten Überlebensraten von ca. 5% (Tabelle 9) ist allerdings nicht mit einer Verdopplung der Klonanzahl zu rechnen. Einen Einfluss auf die Ermittlung der Mutationsfrequenzen mittels des *Hprt*-Tests hat diese Tatsache jedoch nicht, da dabei die Zellen nach der Mutagenese mehreren Passagen unterzogen wurden.

## 4.2 Theoretische Detektionswahrscheinlichkeiten auf Transkript Ebene

Zu Beginn der Arbeit stand fest, dass die Durchmusterung auf Transkriptebeine stattfinden sollte: Es ist naheliegend Exon-Deletionen in der gespleißten mRNA zu detektieren, da sich diese erst auf Transkriptebeine manifestieren und der zu betrachtende Datensatz im Vergleich zur DNA deutlich reduziert ist (~95%). Zudem ist es aufgrund nur schwach konservierter Spleißsignalsequenzen und der Vielzahl *cis*-agierender Sequenz-Elemente (exonic/intronic splicing enhancers) mit ebenfalls relativ degenerierten Konsensussequenzen sehr schwierig auf genomischer Ebene gezielt Punktmutationen herauszufiltern, die zu einer Spleißmutationen führen könnten (Blencowe 2006). Ein bedeutender Nachteil der Transkript-basierten Detektion ist allerdings die Tatsache, dass nur Mutationen in transkribierten Genen identifizierbar sind. Damit war von Anfang an ausgeschlossen ein Archiv aufzubauen, welches Exon-Deletions Mutanten für nahezu jedes beliebige Gen enthalten sollte. Diese Restriktion wurde mit Hinblick auf die Komplexität der Detektion auf DNA-Ebene so akzeptiert.

Im Vorfeld der Methodenentwicklung musste theoretisch betrachtet werden, mit wie vielen Exon-Deletions Ereignissen pro Klon unter den gegebenen Randbedingungen (Mutationsfrequenz, Anzahl exprimierter Gene) zu rechnen war. Die Ermittlung dieser

Kenngröße war erforderlich, um im Zuge der Entwicklung des Detektionsverfahrens den Arbeitsaufwand und die Kosten für eine Erstellung eines Archivs und damit die Wirtschaftlichkeit der Vorgehensweise abwägen zu können.

Unter Verwendung einiger Annahmen wurde die Mutationsfrequenz für eine funktionelle Mutation am *Hprt*-Lokus, die sich einfach für jede Mutagenese mittels des beschriebenen 6-TG Resistenztests (1.5.1) ermitteln ließ, auf eine allgemeine Exon-Deletions-Mutationsfrequenz von ca.  $1.85 \times 10^{-4}$  Ereignissen pro Lokus übertragen. Zu den Annahmen gehörte die Tatsache, dass *Hprt* als durchschnittliches großes und damit repräsentatives Gen angesehen werden kann, sowohl in der Länge des Transkripts als auch in der Anzahl der Exons. Für Gene mit einer niedrigeren bzw. höheren Anzahl an Exons ist die Wahrscheinlichkeit für eine Spleißmutation erwiesenermaßen auch niedriger bzw. höher (Barbaric et al., 2007). Auf das gesamte Transkriptom gesehen (durchschnittlich 9,7 Exons/Gen), repräsentiert *Hprt* mit 9 Exons jedoch einen guten Durchschnitt. Weiterhin wurde eine Wahrscheinlichkeit von 20% für das Auftreten von Spleißmutationen aus der Literatur und nicht der ermittelte Wert von 11% angenommen. Die Entscheidung dafür lag in der höheren statistischen Sicherheit begründet, die durch die Vielzahl externer Daten geboten wurde. Hier muss angemerkt werden, dass die Anteile mit denen Spleißmutationen angegeben werden stark zwischen den Sequenz-basierten (ca. 7%) und Phänotyp-basierten Ansätzen (ca. 20%) variieren. Das liegt darin begründet, dass Spleißmutationen häufiger (i.d.R. ca 2/3 aller Exon-Deletionen) zu einer Verschiebung des Leserasters und damit zu einem Funktionsverlust führen als einfache Punktmutationen. Daher sind diese in Phänotyp-getriebenen Ansätzen überrepräsentiert (Barbaric et al., 2007). In Sequenz-basierten Durchmusterungen ist allerdings die Mutationsrate aufgrund der Einbeziehung nicht-funktionsrelevanter Punktmutationen auch deutlich höher anzusetzen. Beide Ansätze sind daher nicht miteinander vergleichbar. Da es sich bei dem *Hprt*-Test um einen Phänotyp-getriebene Selektion handelt, wird auch hier der entsprechende Wert verwendet.

Für die Abschätzung der theoretisch detektierbaren Spleißmutationen pro Klon und die spätere Auslegung des Detektionsverfahrens mussten neben der Mutationsfrequenz die Gene ermittelt werden, welche in der verwendeten F1 ES-Zelllinie exprimiert werden. Diese Analyse erfolgte mit einem Illumina Bead Chip Array. Maßgeblich für die Entscheidung, ob ein Gen als transkribiert angesehen wurde oder nicht, war der Detektionswert (detection-value, p) (Kuhn et al., 2004). In dieser Arbeit wurden nur solche Gene als transkribiert angesehen bei denen mindestens ein Oligonukleotid einen p-Wert größer als 0,99 aufwies, d.h. man konnte zu 99% sicher sein, dass das Gen transkribiert wird und das Signal nicht

falsch-positiv ist. Die Wahrscheinlichkeit wurde bewusst so hoch gesetzt, um für die weitere Analyse in dieser Hinsicht Unsicherheiten auszuschließen. Für insgesamt 7.475 der 18.626 RefSeq Gene galt  $p \geq 0,99$  in der Analyse einer Wildtyp F1 ES-Zell Probe (3.2.1). Es wurden demnach ca. 40% der Gene in der ES-Zelle als transkribiert angesehen. Dieser Wert ist vergleichbar mit einer unabhängigen *in-silico* Analyse von 249.200 EST Sequenzen in der 10.700 (ca. 36%) der betrachteten putativen 29.810 Gene in embryonalen Stammzellen bzw. frühen Embryostadien der Maus als transkribiert angesehen werden konnten (Sharov et al., 2003). Weitere Analysen mit ENU-mutierten F1-Zellklonen und biologischen Wildtyp Replikaten in dieser Arbeit zeigten in einer unabhängigen Hybridisierung vergleichbare Ergebnisse und untermauerten die Richtigkeit des Ergebnisses (3.2.2).

Unter Verwendung der Mutationsfrequenz und der Anzahl transkribierter und damit analysierbarer Gene ließ sich somit ein theoretischer Wert von ca. 1,38 Deletionen pro untersuchtem Klon abschätzen. Dieser Wert war von großer Bedeutung, um im Zuge der Methodenentwicklung einen Überblick über die zu erwartenden Kosten und damit die Wirtschaftlichkeit und Machbarkeit eines großangelegten Screenings zu erhalten.

### **4.3 Entwicklung eines Mikroarrays zur Detektion von Exon-Deletionen**

#### **4.3.1 Die Wahl der Technik**

Die Mikroarray-Technologie ist ein bewährtes Verfahren zur Messung der Gen-Expression in den verschiedensten Geweben und Zelltypen. Im Zuge des technischen Fortschritts erweiterte sich der Anwendungsbereich dieser Technologie auf eine Vielzahl anderer Gebiete wie z.B. auf die Analyse alternativen Spleißens, die Identifikation der Ziele RNA bindender Proteine, Chromatin-Immunopräzipitations-Studien (ChIP), Methylom-Analysen, die Detektion von Polymorphismen und sogar die Resequenzierung ganzer Genome (Mockler et al., 2005). Grundlage für diesen Entwicklungsfortschritt bildete insbesondere die Möglichkeit Arrays mit hohen Anzahlen von Sonden generieren zu können (high-density microarrays). Generell standen zu Beginn der Arbeit zwei Systeme zur Verfügung. Bei dem ersten System werden PCR-Produkte, Oligonukleotide oder klonierte DNA-Fragmente mechanisch auf die Oberfläche eines Arrays aufgebracht (ca. 10.000 bis 40.000 Positionen) (Skena et al., 1995). Das zweite System sieht eine direkte *in-situ* Synthese relativ kurzer Oligonukleotide (<100bp) auf der Array-Oberfläche vor (insbes. Fotolithographie und Ink-Jet Technologie), wie sie beispielsweise von den Firmen Affymetrix und Agilent durchgeführt wird (Hughes et al., 2001; Pease et al., 1994; Singh-Gasson et al., 1999). Diese Oligonukleotid Mikroarrays können Millionen verschiedener Oligonukleotid-Positionen mit jeweils mehreren tausend

Kopien des Oligonukleotids tragen und weisen zudem eine höhere Spezifität als klassische cDNA Arrays auf (Mockler et al., 2005). Während das Design neuer Mikroarrays bei der klassischen Fotolithographie neue Masken erfordert, ist die Ink-Jet Technologie in der Lage relativ unkompliziert und damit preisgünstig neue Oligonukleotid-Kompositionen in Längen zwischen 25 und 60 bp zu erstellen (Hughes et al., 2001).

Zu Beginn dieser Arbeit lagen bereits einige Arbeiten vor in denen die Mikroarray-Technologie zur Analyse alternativen Spleißens (AS) genutzt worden war. Da die zu detektierenden Exon-Deletionen nichts anderes als alternatives Spleißen ganzer Exons (Exon Kassetten Spleißen) darstellen, war es sinnvoll, aufbauend auf den veröffentlichten Ergebnissen, ein eigenes System zu entwickeln.

Der größte Anteil der publizierten Arbeiten verwendete dabei bereits Oligonukleotid-Mikroarrays. Hier zeigte sich, dass klassische kommerziell weitverbreitete Genexpression-Mikroarrays weniger geeignet waren, da sie nur wenige Oligonukleotide pro Gen aufweisen und diese oft in den 3'-UTR Bereichen platziert werden (Lee et al., 2004). Erste erfolgreiche Ergebnisse konnten mit dem Design Exon-spezifischer Sonden erzielt werden (Shoemaker et al., 2001). Hu et al. (2001) untersuchten alternatives Spleißen an den 3' Enden von 1.600 Ratten mRNAs indem sie 20 Oligonukleotid-Sonden in diese Bereiche platzierten (Hu et al., 2001). Jedoch war das Design des Mikroarrays noch suboptimal für diese Anwendung. Ebenso wurden sogenannte 'Tiling Arrays', bei denen in unterschiedlichen Abständen Oligonukleotide platziert wurden, für die Ermittlung von Spleiß-Stellen verwendet (Kampa et al., 2004). Die erste Entwicklung eines reinen Spleiß-spezifischen Mikroarrays wurde zur Analyse des konstitutiven Spleißens der Hefe entwickelt (Clark et al., 2002). Hier wurden zum ersten mal Exon-Junction Sonden verwendet, d.h. Sonden die Exon/Exon-Grenzen überspannen und deren Sequenz damit nur komplementär zu mRNA-Molekülen ist. Die Verwendung und Optimierung von Exon-Junction Sonden zur Analyse alternativen Spleißens wurde daraufhin von zwei Forschergruppen anhand des Spleißens der Gene *RBI* und *ANXA7* bzw. *CD44* und *TMP2* validiert (Castle et al., 2003; Wang et al., 2003). Die erste wahre groß angelegte Mikroarray-Studie zu alternativem Spleißen wurde 2003 veröffentlicht (Johnson et al., 2003). Die Autoren hatten ein Mikroarray Design basierend auf ca. 125.000 Exon-Junction Oligonukleotid-Sonden erstellt, das es theoretisch ermöglichte die Exon/Exon-Grenzen von 10.000 menschlichen RefSeq-Genen zu überprüfen und so Vorhersagen für putative AS Ereignisse zu erstellen. Die Ergebnisse gaben deutliche Hinweise für die Bedeutung alternativen Spleißens bei der Differenzierung verschiedener Gewebetypen. Basierend auf diesen guten Ergebnissen wurde in einer weiteren groß-angelegten Studie das

Design des Mikroarrays von einer anderen Gruppe übernommen und optimiert indem zusätzlich zu Exon-Junction Sonden auch Exon-spezifische Sonden hinzugenommen wurden, um das Vorkommen verschiedener Varianten genauer quantifizieren zu können (Pan et al., 2004). Auch wurden hier erstmalig Spleiß-spezifische Sonden verwendet, d.h. Oligonukleotide deren Sequenzen aus den flankierenden Exons eines alternativ gespleißten Exons zusammengesetzt sind (in dieser Arbeit als Exon-Skip Oligos bezeichnet).

Der bedeutenste Unterschied der Technik zur Detektion von Exon-Deletionen zu der des alternativen Spleißens musste in der Verwendung als Hochdurchsatz-Durchmusterungsverfahren gesehen werden. Während die Analyse alternativen Spleißens nur den Vergleich einer geringen Anzahl verschiedener Gewebe- bzw. Zelltypen verfolgte, sollte in dieser Arbeit der Grundstein für eine Technik zur Durchmusterung tausender von Klonen gelegt werden. Das bedeutete konkret, dass die Technik so einfach und damit so preisgünstig gehalten werden sollte wie möglich. Hinsichtlich der Wirtschaftlichkeit gab es zwei entscheidende Faktoren:

1. Anzahl der Oligonukleotide: Ziel war es nicht wie bei den bereits veröffentlichten Untersuchungen neben der reinen Detektion auch die Expressionstärken der Spleißvarianten zu untersuchen, sondern lediglich ein Ja/Nein-Signal zu erhalten, welches eine putative Exon-Deletion anzeigt. Daher stand die Anzahl notwendiger Oligonukleotide in direktem Zusammenhang zu einem Anstieg der Kosten. Betrachtet man die Array Kosten allein, bringt die Verdopplung der notwendigen Oligonukleotide eine Verdopplung der Materialkosten mit sich. Es wurde daher generell von Tiling Arrays aufgrund der Vielzahl benötigter Sonden abgesehen.
2. Pooling: Die Sensitivität sollte so hoch sein, dass ein Pooling von Proben ermöglicht wird. Allein die parallele Vorbereitung und Analyse zweier Proben resultiert in einer Kostenreduktion von 50%, ausgenommen des Arbeitsaufwands für die Isolation der RNAs.

Hinsichtlich der kostenrelevanten Eigenschaften wurden die in der Literatur bereits beschriebenen Verfahren betrachtet. So wurde generell davon abgesehen zweifarbige Mikrorray Analysen durchzuführen. Der ständige Vergleich gegen eine Wildtyp-Referenz verursacht nur zusätzliche Kosten für die Bereitstellung der Wildtyp Probe. Es wurde als ausreichend erachtet, einige Analysen auf einfarbigen Arrays mit Wildtyp-Proben durchzuführen, um falsch positive Signale (d.h. auch alternatives Spleißen) identifizieren zu

können und so Referenzwerte für den Vergleich mit den Ergebnissen ENU-mutierter Klone zu schaffen.

Das grundlegende experimentelle Design der Oligonukleotid Zusammenstellung des Micorarrays basierte auf den Ergebnissen von Pan et al. (2004), da deren Design Spleiß-spezifische Oligonukleotide bereits erfolgreich verwendete. Exon-spezifische Sonden wurden aus dem Ansatz von Pan ausgeklammert, da bei der Suche nach Exon-Deletionsmutanten kein Bedarf an einem quantitativem Vergleich der Transkriptvarianten bestand. Das System von Pan et al. (2004) wurde mit Hinblick auf die Kosten, die eine hohe Anzahl von Oligonukleotid-Sonden mit sich bringen, in dieser Arbeit in zwei kombinierbare Detektionsvarianten, der Detektion mittels Skip- und/oder mittels Junction-Sonden, differenziert (Abbildung 11). Für den Zweck der reinen Detektion einer Exon-Deletion schienen die Exon-Skip Oligonukleotidsonden am besten geeignet, da hier abgesehen von Kreuz-Hybridisierungen ein reines Ja/Nein Signal zu erwarten war und theoretisch nur ein Oligonukleotid pro Deletionsereignis nötig war. Diese Variante stellte daher bezüglich des Designs kostentechnisch die Minimalvariante dar. Nur geringfügig teurer war die Detektionsvariante mittels Junction-Sonden zu bewerten. Zwar sind theoretisch zwei Sonden nötig, um das fehlende Exon genau angeben zu können, doch relativiert sich diese Anzahl, wenn man das ganze Transkript betrachtet: Eine Junction-Sonde kann immer zur Detektion zweier Spleiß-Ereignisse genutzt werden. Der große Nachteil von Exon-Junction Sonden gegenüber Skip-Sonden besteht allerdings darin, dass im Falle einer Deletion nur eine Reduktion der Signalintensität zu erwarten ist und kein Ja/Nein Signal. Andererseits erlaubt diese Variante auch das Spleißen des ersten und letzten Exon zu überprüfen. Die Verwendung der Sondensysteme mit Blick auf die parallele Analyse von Proben (Pooling) ließ sich im Vorfeld schwer bewerten, da die Versuche in den publizierten Arbeiten zu alternativem Spleißen in verschiedenen Geweben natürlich nicht unternommen wurden. Prinzipiell war hier die Detektion mittels Skip-Oligonukleotiden besser zu bewerten, da Wildtyp Allele keine Signale erzeugen sollten wie es bei den Junction Sonden allerdings der Fall ist.

### **4.3.2 Grundlagen der Micorarray-basierten Detektion**

Für die praktische Umsetzung einer Detektion von Exon-Deletions Mutanten nach den genannten Verfahren galt es im Vorfeld zwei Vorraussetzungen zu betrachten: Zum einen die technische Umsetzung auf einer Array-Plattform und zum anderen die Verfahrensweise bei dem Design des Arrays, d.h. der Zusammenstellung der Sonden.

Wir entschieden uns zugunsten der von Rosetta Inpharmatics Inc. entwickelten Ink-Jet Oligonukleotid Microarrays der Firma Agilent (Hughes et al., 2001). Diese boten gegenüber

den Custom-Arrays anderer Hersteller (zu Beginn der Arbeit nur NimbleGen und Combimatrix) entscheidende Vorteile hinsichtlich der Flexibilität (insbes. frei wählbare Länge der Oligonukleotide auch innerhalb eines Arrays), der guten Standardisierung, der Handhabbarkeit und der Kontrolle der Arrays (interne Positiv- und Negativ-Kontrollen, Spike-In Kontrollen zur Ermittlung der dynamischen Breite der Signalintensitäten) (2.3.8.2 und 3.3.3). So konnten auch standardisierte Reagenzien zur Vorbereitung und Verarbeitung der Proben vom gleichen Hersteller erworben werden, die eine durchgehende Kontrolle der Analysen erlaubten. Zudem wurden bereits andere Studien, die sich mit der Verwendung von Junction-Sonden beschäftigten, auf dieser Plattform durchgeführt (Castle et al., 2003; Pan et al., 2004), sodass von einer einfachen Übertragung der Ergebnisse ausgegangen wurde. Bedeutenden Einfluss auf die Entscheidung für Agilent hatte auch der vergleichsweise geringe Preis von ca. 500 € pro Glasträger unabhängig davon welches Format, d.h. wie viele verschiedene Arrays auf der Oberfläche gewählt wurden. Die Kosten konnten weiter dadurch reduziert werden, dass die Arrays zweimal wiederverwendet werden können (Hahnke et al., 2007). Das Kit für die Probenvorbereitung kostete ca. 1.200 € für 10 Reaktionen, wobei auch hier die Kosten nahezu halbiert werden konnten, da eine RNA Amplifikation mit der Hälfte der Reagenzien ausreichend aRNA-Ausbeute für Hybridisierungen eines 4x44k bzw. 8x15k Feldes lieferte (Daten nicht gezeigt).

Für die Zusammenstellung der Sonden mussten die Sequenzen der einzelnen Exons vorliegen. Diese Information konnte aus ENSEMBL für alle transkribierten RefSeq Transkripte abgefragt werden. Auf dieser Datengrundlage wurde eine spezifische Software in Perl entwickelt, welche die Zusammenstellung von Junction- und/oder Skip-Sonden *in silico* ermöglichte (3.3.2).

Grundsätzlich bietet die Verwendung von Junction- bzw. Skip-Sonden nur wenig Spielraum hinsichtlich der sequenzbedingten biochemischen Eigenschaften der Oligonukleotide, da ihre Sequenz durch die Exongrenze festgelegt ist. Dies ist ein Nachteil an der Arbeit mit dieser Art von Sonden, da bei Microarray-Analysen generell Sonden mit vergleichbaren Schmelztemperaturen ( $T_m$ -Werte), geringen Tendenzen zur Bildung interner Sekundärstrukturen und hohem Informationsgehalt (wenige Wiederholungen, repeats) erfolgversprechendere Ergebnisse liefern (Srinivasan et al., 2005). Castle et al. (2003) analysierten im Detail anhand des Spleißens der humanen Gene *RBI* und *ANXA7* die Faktoren, welche den größten Einfluss auf die erfolgreiche Verwendung von Exon-Junction Oligonukleotid-Sonden haben sollten. Dies waren ihrer Auffassung nach die Länge des Oligonukleotids und die Positionierung über der Exon-Grenze. Dabei stellten sie fest, dass

kürzere Oligonukleotide mit Längen von 35 bis 40 Basen Länge und eine zentrierte Positionierung an der Exon-Grenze ( $\pm 5$  Basen) optimale Ergebnisse lieferten. Längere Sonden neigten aufgrund der langen homologen Bereiche zu alternativen Spleißvarianten (ca. 50% der Sonden) zu starken Kreuzhybridisierungen und daher zu falsch-positiven Ergebnissen. Eine Verschiebung der Position des Oligonukleotids um  $\pm 5$  Basen um die Exon-Grenze hatte hingegen kaum einen Einfluss auf die Spezifität, d.h. unter den Oligonukleotiden eines Ereignisses war in diesem Rahmen der Einfluss der Schmelztemperatur vernachlässigbar. Dennoch verwendeten Pan et al. (2004) ein Programm, welches die Exon-Junction Sonden so auf der Spleißstelle positionierte, dass beide Teile der Sonde je einen  $T_m$ -Wert von  $66^\circ\text{C} \pm 2^\circ\text{C}$  und die gesamte Sonde ebenfalls einen  $T_m$ -Wert von  $66^\circ\text{C} \pm 2^\circ\text{C}$  aufwies. Dies führte in Zusammenhang mit anderen Restriktionen (Vermeidung von einfachen Repeats wie CGG und CCC und internen Basenpaarungen) allerdings zu einer Reduktion der untersuchbaren Ereignisse alternativen Spleißens um ca. 45%. Aufgrund der Ergebnisse von Castle et al. (2003) und der Vermeidung drastischer Restriktionen hinsichtlich der Oligonukleotidqualität wurden in dieser Arbeit die Sonden zentriert angeordnet analog zu Johnson et al. (2004), d.h. sie bestanden grundsätzlich zu gleichen Teilen aus den beiden verknüpften Exon Sequenzen. Srinivasan (2005) stützt diese Vorgehensweise durch seine Aussage, dass die Verwendung des  $T_m$ -Wertes zur Abschätzung der Bindeeigenschaften gleich langer Sonden eher eine untergeordnete Rolle spielen sollte, da die Reaktion auf der Array-Oberfläche eine Binde- und keine Schmelzreaktion darstellt. Die Ausbildung einer Duplex wird seiner Meinung nach von deutlich mehr Faktoren beeinflusst als die temperaturbedingte Auftrennung der Duplex. Auch die von der Firma ExonHit Therapeutics in Zusammenarbeit mit Agilent zwischenzeitlich vertriebenen Spleiß-spezifischen Microarrays verfolgten die Detektion mittels Junction-Sonden, die zentriert positioniert sind (Fehlbaum et al., 2005). Für die ersten 'Proof-of-Principle Experimente' wurde daher im Vorfeld vollständig auf eine Qualitätskontrolle der Sequenzen hinsichtlich der Bindeeigenschaften zugunsten eines breiten Spektrums abgedeckter Spleißstellen verzichtet. Die entwickelte Software ließ allerdings die Möglichkeit offen, später Funktionen dieser Art zu integrieren. Die Länge der Sonden war frei wählbar, da trotz der umfangreichen Analyse von Castle et al. (2003) unter den publizierten Arbeiten kein Konsens bezüglich der Oligonukleotidlänge herrschte (Castle et al., 2003; Johnson et al., 2003; Pan et al., 2004; Religio et al., 2005) und dieses daher experimentell überprüft werden musste.

Für die Etablierung des Detektionssystems war zusätzlich eine Reduktion der parallel zu betrachtenden möglichen Detektionsereignisse angedacht (3.3), um die Kosten in den

Vorversuchen möglichst gering zu halten. Eine sinnvolle Reduktionsmöglichkeit wurde in der Variante gesehen, sich nur auf solche Exons zu konzentrieren, deren Deletion nicht zu einer Unterbrechung des offenen Leserasters führt. Proteine aus derart mutierten Transkripten könnten interessante Kandidaten für die Funktionsanalyse bestimmter Proteindomänen darstellen und unter Umständen wertvoller als reine Funktionsverlustmutanten sein, die auch einfacher mittels klassischer Knockout-Strategien generiert werden können (1.3). Exon-Deletionsmutationen, die aufgrund der Verschiebung des Leserasters zur Induktion eines frühzeitigen Stop-Codons führen und damit den gezielten Abbau des Transkripts (Nonsense-mediated mRNA decay, NMD) verursachen können, würden sich zudem einer Transkript-basierten Detektion entziehen (McGlinicy et al., 2008; Wen et al., 2008). Daher wurde diese Art der Restriktion in der Etablierung des Systems als durchaus sinnvoll erachtet und die Software dahingehend erweitert, dass Skip- bzw. Junction-Sonden nur für diese Exons erstellt wurden.

### 4.3.3 Untersuchungen zur Machbarkeit

Die Machbarkeit des angedachten Detektionsverfahrens galt es zunächst in der Praxis zu überprüfen und noch offene Parameter wie die Detektionsvariante, die Länge der Oligonukleotide und das Vermögen gepoolte Proben zu untersuchen zu klären. Für diese Parameter konnten der Literatur keine eindeutigen Aussagen bzw. Voruntersuchungen entnommen werden (Castle et al., 2003; Johnson et al., 2003; Pan et al., 2004; Religio et al., 2005).

Das unter 3.3.4 beschriebene Mikroarray-Design involvierte die Betrachtung der Detektionsvarianten bei verschiedenen Oligonukleotidlängen für das *Hprt* Transkript, da dafür Exon-Deletionsmutanten als Positivkontrollen zur Verfügung standen. Das Testsystem war sehr nah an das geplante Mikroarray Design angelehnt, um kein rein artifizielles System zu schaffen, das sich anschließend nicht problemlos übertragen lässt. So wurde das geplante Array-Format von 4x44k verwendet und der mit den Test-Oligonukleotiden bestückte Array wurde der Reaktionsumgebung wegen mit zufällig ausgewählten Skip-Oligonukleotiden aufgefüllt. Zudem wurde natürliches Probenmaterial, d.h. eine *Hprt* Exon 8 Deletionsmutante anstelle von klonierten Isoformen verwendet, um von Anfang an das biologische System zu simulieren.

Anhand der Untersuchung konnte gezeigt werden, dass prinzipiell eine Detektion der Deletion des Exon 8 des *Hprt* Transkripts über beide Oligonukleotid-Detektionsvarianten möglich ist. Entscheidenden Einfluss hatte dabei für beide Systeme allerdings die Länge der Oligonukleotide. Ab einer Länge von 50 bp war eine Detektion mittels Skip-Oligonukleotiden

nicht mehr möglich. Bei den Junction-Oligonukleotiden war die Detektion abhängig von der betrachteten Sonde. Dieses Ergebnis bezüglich der Länge stand in Einklang mit den zwischenzeitlich veröffentlichten Daten der Fa. ExonHit Therapeutics (Fehlbaum et al., 2005), die in Zusammenarbeit mit Agilent ebenfalls an einer Entwicklung einer Mikroarray basierten Analyse-Plattform für alternatives Spleißen auf Basis der von Pan et al. (2004) beschriebenen Vorgehensweise arbeiteten. ExonHit kam nach einem Vergleich von Oligonukleotidlängen zwischen 24 und 40 bp zu dem Schluss, dass die 24mere Oligonukleotide die höchste Spezifität aufweisen. Das galt sowohl für Skip- als auch für Junction- und Exon- spezifische Sonden. Trotz der hohen Spezifität der 24mere wurde durch Fehlbaum et al. (2005) beobachtet, dass sogar bei 24meren jeweils noch ca. 10% der Signalintensität einer Skip- bzw. Junction-Sonde durch Hybridisierung mit der nur zu 50% kompletären alternativen Spleißform bedingt waren. Srinivasan et al. (2005) benennen dieses Problem 'half-junction crosstalk' und sehen kürzere Sonden hinsichtlich der Spezifität biochemisch klar im Vorteil, da bei diesen i.d.R. die Differenz zwischen dem  $T_m$ -Wert des ganzen zu dem des halben Oligonukleotids deutlich größer ist als dies bei langen Oligonukleotiden der Fall ist. Da in dieser Arbeit 30mere die von Agilent minimal empfohlene Oligonukleotidlänge darstellten und mit diesen die spezifischsten Ergebnisse erzielt wurden, sollte diese Länge für weitere Untersuchungen vorerst Standard sein.

Die Entscheidung hinsichtlich des Detektionsverfahrens fiel auf die kostengünstigere Variante mittels Skip-Oligonukleotiden, da diese eine Positiv-Selektion der Mutationen ermöglichte und in dieser Arbeit im Vergleich der Signalstärken noch eine höhere Spezifität für 30mer-Oligonukleotide aufwies als die entsprechenden Junction-Oligonukleotide. Mit Hinblick auf angedachtes Pooling verschiedener Proben hat diese Variante den entscheidenden Vorteil, dass aufgrund der Positiv-Selektion nur qualitativ entscheidend ist, ob ein Signal vorhanden ist oder nicht. Bei der Variante mittels Junction-Sonden muss grundsätzlich quantitativ betrachtet werden, ob eine eindeutige Abnahme der Signalintensität für das Oligonukleotid erfasst werden kann. Diese Entscheidung wird erschwert mit steigender Anzahl von Proben in einem Pool, da die Intensität erwartungsgemäß gegen die maximale Wildtyp Intensität ansteigen wird.

Nachdem eine prinzipielle Machbarkeit mit den gesteckten Rahmenbedingungen gegeben war, galt es, die Kosten-Nutzen Parameter zu prüfen, d.h. die Möglichkeit Pools verschiedener Transkriptome mutierter ES-Zell-Klone zu untersuchen (4.3.2). Hierzu wurde das Array Design angepasst (30 bp Skip-Sonden, 3.3.5) und zwei *Hprt* Positiv-Kontroll-Klone mit Deletionen in Exon 8 und Exon 6 in verschiedenen Pools mit ENU-mutierten Klonen bis

zu einer 1:5 Verdünnung analysiert. In allen Verdünnungsstufen ließen sich die vorhandenen Deletionen ermitteln. Allerdings lag die Signalintensität der Exon 6 spezifischen Skip-Sonde bereits im Bereich des Hintergrundsignals wenn auch noch deutlich über dem des Wildtyps. Der Grund dafür wurde im geringen GC-Gehalt der Sonde von nur 30% gesehen. Dies zeigte auf, dass nicht nur Deletionen evtl. schwach exprimierter Gene durch das Raster fallen können, sondern auch Sonden mit schwächeren Bindeeigenschaften. Hybridisierungen bei niedrigeren Temperaturen bewirkten leider keinerlei Besserung dieser Situation. Im Gegenteil, die Arrays wurden dadurch nicht mehr auswertbar (zu hohe Hintergrundsignale). Restriktionen hinsichtlich des GC-Gehaltes bzw. des  $T_m$ -Wertes wie sie von Fehlbaum et al. (2005) und Pan et al. (2004) vorgenommen wurden, wurden als nicht zielführend erachtet. Unter den bereits gegebenen Restriktionen (keine Unterbrechung des ORF, Gen ist exprimiert) bot die Mikroarray-Plattform mit knapp 44.000 Oligonukleotidsonden genügend Freiraum, um auch solche Sonden aufzunehmen. Andererseits wurden jedoch 1.023 Sonden als falsch-positive gewertet, da diese auch mit den Wildtyp-Proben Signalintensitäten deutlich über dem Hintergrundsignal lagen und damit für nachfolgende Untersuchungen wertlos waren. Für nur 50 dieser Sonden konnten Treffer in der Maus EST Datenbank erzielt werden, was diese jedoch eindeutig als unbekannte, d.h. bis dato nicht in RefSeq vertretene Ereignisse alternativen Spleißens aufzeigte. Dadurch wurde jedoch indirekt auch die technische Funktionalität des Arrays bestätigt. Nahezu zwei Drittel der restlichen Sonden wiesen einen GC-Gehalt größer 60 % auf, was wiederum eine Restriktion wie von Pan et al. (2004) rechtfertigen würde. Doch konnten die Sonden in dieser Arbeit auch so auf empirischem Weg ausgeschlossen werden, was evtl. den Fehler minimierte zu stringent bei den Restriktionen vorzugehen.

Aus den Daten konnte insgesamt abgeleitet werden, dass eine Detektion von Exon-Deletionen mittels Skip-Oligonukleotid-Sonden möglich ist. Entscheidenden Einfluss hat dabei die Länge der Sonden. Auch aus gepoolten Proben ist ein Nachweis begrenzt möglich. Generell können allerdings die biochemischen Eigenschaften der Sonde und die Expressionsstärke des Transkripts negativen Einfluss auf die Detektion haben. Die Machbarkeit hinsichtlich der Wirtschaftlichkeit des Systems lässt sich erst nach Test-Studien beurteilen, da die Effizienz nur aus der Praxis heraus beurteilt werden kann. Bei Untersuchungen mit 1:5 verdünnten Proben in Pools beliefen sich die Materialkosten (Zeitpunkt 2008) auf ca. 35,00 € pro Probe, die in groß angelegten Durchmusterungen sicherlich noch reduziert werden könnten.

#### 4.3.4 Detektion und Bestätigung neuer Mutations Ereignisse

Anhand der Daten, die im Rahmen der Pooling Experimente für ENU-Mutanten generiert worden waren, sollte erstmalig die Detektion unbekannter Deletions-Mutationen erprobt werden. Prinzipiell zeigten alle die Sonden mögliche Kandidaten auf, deren Signalintensität über dem jeweiligen spezifischen Hintergrundsignal des Wildtyps lagen. Das traf für ca. 10% der Sonden zu. Daher wurden alle jene Signale herausgefiltert, die mindestens das Fünffache des Wildtypsignals betrug, wobei der Wert 'Fünffach' frei gewählt wurde, allerdings mit Berücksichtigung der Untersuchungen an den Positiv-Kontrollen. Nahezu 1% aller Sonden erfüllten diese Vorgabe, was eine detaillierte Betrachtung der Rohdaten und der dazugehörigen Bilder des Arrays unumgänglich machte. Dabei fiel auf, dass ein Großteil der erhöhten Signalintensitäten auf Hybridisierungsartefakten basierte. Bei diesen sehr kleinen Hybridisierungsartefakten, die z.T. exakt auf den Positionen der Sonden lagen und nur durch ihre nichtkreisrunde Ausprägung von tatsächlichen Signalen zu differenzieren waren, versagte die automatische Erkennung der Feature Extraction Software. Die Gründe dafür sind ungeklärt. Hersteller und Literatur konnten keine Klärung bringen. Generell wurde höchste Sorgfalt in Bezug auf die Reinheit der verwendeten Reagenzien und Reaktionsgefäße geübt und soweit es möglich war wurde nur mit spezieller 'Einmal-Plastikware' gearbeitet. Für mögliche positive Signale wurde daher grundsätzlich eine Kontrolle vorgenommen. Um falsch-positive Signale besser kontrollieren zu können, wurde zusätzlich die Tatsache, die auch in den Durchmusterungen Standard war, ausgenutzt, dass eine Probe grundsätzlich in zwei Pools enthalten war und somit in beiden ein entsprechendes Signal erkennbar sein musste. Dadurch wurde die Anzahl der Kandidaten auf 10 reduziert von denen 2 die beiden *Hprt* Positiv-Kontrollen waren, was die Vorgehensweise des Auswerte-Algorithmus bestätigte. Analog zu diversen anderen Veröffentlichungen, die sich mit der Detektion alternativen Spleißens auseinandersetzen (Fehlbaum et al., 2005; Johnson et al., 2003; Relogio et al., 2005), sollte eine Bestätigung auf Basis von RT-PCR erfolgen. Alternativ wäre beispielsweise eine Bestätigung mittels Northern-Blot möglich gewesen, doch standen die Einfachheit und der Zeitaufwand der Technik im Vordergrund, was bei größer angelegten Untersuchungen von enormer Bedeutung ist. Zusätzlich zu der Standard-Analyse mit flankierenden Primern um das deletierte Exon und dem anschließenden Vergleich der Produktgrößen wurde das von Greber et al. (2005) publizierte Verfahren angewandt. Dabei wurde ein Skip-Primer verwendet, sodass auch das PCR-System eine positive Selektion auf die Deletion bewirkt. Dies sollte in paralleler Verwendung zu dem flankierenden System eine Bestätigung ermöglichen, falls beispielsweise eine Herunterregulation des mutierten Allels

die Amplifikation gegenüber dem Wildtyp Allel erschwert (Greber et al., 2005). Über dieses Skip-System konnten zwei Mutationen bestätigt werden (3.3.7). Allerdings war nur bei der Mutation in *Mgea6* auch eine kleinere Bande mit dem flankierenden System zu erkennen. Für die Deletion in *Nup37* könnten oben getroffene Annahmen zur erschwerten Amplifikation des kürzeren Produkts aufgrund deutlich reduzierter Matrizenmengen zutreffen. Für zwei Transkripte war gar keine Amplifikation zu erzielen. Für die restlichen vier Produkte wurden nur Wildtyp-Produkte erhalten. Da hier Mutationen nicht mit Sicherheit ausgeschlossen werden können, müßten diese Kandidaten bei Interesse mit anderen Methoden (z.B. Northern-Blot, Klonierung) genauer untersucht werden. Die Skip-Primer sind in ihrer Effektivität schwer zu charakterisieren und optimieren, weil i.d.R. keine Positiv-Kontrollen zur Verfügung stehen und die biochemischen Charakteristika der Primer durch die festgesetzte Position auf dem Transkript schwer zu beeinflussen sind. Ein Problem, welches im Allgemeinen auch den Umgang mit Skip- und Junction-Probes erschwert und die Hauptschwierigkeit darstellt (Mockler et al., 2005). Dennoch stellt die RT-PCR aufgrund der Einfachheit, Schnelligkeit und des Preises des Verfahrens eine gute erste Kontrollreaktion dar.

#### 4.3.5 Umsetzung des entwickelten Verfahrens

Mit der Etablierung des Detektionsverfahrens mittels Skip-Oligonukleotiden wurde die Möglichkeit geschaffen, eine Test-Bibliothek von knapp 100 ENU-mutierten ES-Zellen parallel nach Deletionen zu durchmustern. Der erste Array wurde dahingehend optimiert, dass falsch-positive Klone generell entfernt wurden und 5 bestätigte alternative Spleißereignisse als interne Kontrollen aufgenommen wurden. Es wurden 40 Pools von jeweils 5 Klonen angelegt und durchgängig analysiert. Jeder Klon war in 2 Pools repräsentiert, sodass eine Kostenerparnis von ca. 40% gegenüber der einfachen Analyse eines Klons möglich war. Der Klon mit der Mutation in *Nup37* diente als Kontrolle der ganzen Durchmusterung. Aus den 100 analysierten Klonen ließen sich nach Durchlaufen des Auswerte-Algorithmus 26 mögliche Mutationen in 19 Klonen feststellen. Dies lag deutlich unter der theoretisch zu erwartenden Anzahl von ca. 43 Mutationsereignissen (1 Ereignis in 2,32 Klonen), konnte jedoch seine Ursache in den schlechten Hybridisierungseigenschaften bestimmter Skip-Sonden oder der schwachen Expression bestimmter Gene haben, die eine Detektion deutlich erschweren. Generell waren die erzielten Signalintensitäten der Skip-Oligonukleotide im Vergleich zu denen der internen Positiv-Kontrollen allerdings sehr niedrig. Sollten die entsprechenden Transkripte jedoch insgesamt stark exprimiert sein, könnte dies bedeuten, dass eine Detektion von Mutationen in schwach exprimierten Genen generell nicht möglich sein sollte, da deren Signale ausschließlich im Hintergrundbereich liegen könnten. Daher

wurde die Transkriptionsstärke der Transkripte mit den putativen Mutationen überprüft, was allerdings keinen direkten Zusammenhang zwischen Detektion und Expressionstärke erkenntlich machte. Die Aussagekraft der Hybridisierungsergebnisse war dennoch zweifelhaft, da sich viele der 26 Signale nahe des Hintergrunds bewegten. Zudem konnte von 10 ausgewählten möglichen Mutationsereignissen keines mittels der beiden PCR-Verfahren verifiziert werden. Die Ursachen dafür blieben unklar. Generell wurde anhand der Microarray-Daten davon ausgegangen, dass es sich bei den detektierten Signalen um mögliche Mutationen handelt, da deren Nachweis auf zwei unabhängigen Hybridisierungen basierte. Den prinzipiellen Erfolg der Durchmusterung mittels Microarray bestätigte zudem die Tatsache, dass die Mutation in *Nup37* und sämtliche Kontrollen (alternative Spleiß-Ereignisse) identifiziert werden konnten.

#### 4.3.6 Abschließende Optimierungsversuche

Die Schwachstellen des Systems wurden aufgrund der erzielten Ergebnisse weiterhin in der geringen Signalintensität der Sonden nahe des Hintergrundsignals gesehen und minderte damit die Glaubwürdigkeit des Systems. Maßgeblich trugen dazu auch die negativen Bestätigungsreaktionen mittels PCR bei, da diese für bekannte alternative Spleiß-Ereignisse problemlos funktionierten und auch in der Literatur als Standardnachweisverfahren genutzt wurden (Fehlbaum et al., 2005; Johnson et al., 2003; Religio et al., 2005).

Im Hinblick auf die generelle Diskussion über Sensitivität und Spezifität von Primern in Abhängigkeit von der Oligonukleotidlänge (Srinivasan et al., 2005) wurde daher ein Array-System mit längerer Oligonukleotidlänge von 36 bp zur Erhöhung der Signalintensität entworfen. Zudem wurden Restriktionen hinsichtlich des GC-Gehaltes und der Repeats analog zu den Arbeiten von Pan et al. (2004) und Fehlbaum et al. (2005) vorgenommen. Aus Kostengründen und der reduzierten Anzahl von Oligonukleotid-Sonden wurden dann Hybridisierungen von 30 einzelnen Klonen auf einem kleineren Array-Format durchgeführt (8x15k). Das Ziel der erhöhten Signalintensität konnte allerdings nicht erreicht werden, möglicherweise bedingt durch den Wechsel der Plattform. Es wurde auch keine erhöhte Anzahl von möglichen Mutationen festgestellt, sodass die Optimierungsversuche als gescheitert angesehen wurden. Generell wurden zwei zusätzliche Mutationsereignisse ermittelt, deren Nachweis aus Pools in der vorangehenden Durchmusterung vermutlich aufgrund der schwachen Signale nicht möglich war.

Letzendlich ließen die Zweifel an dem Detektionsverfahren, bedingt durch die geringen Signalintensitäten und die geringe Anzahl detektierter Mutationen, das System als nicht kontrollierbar und zu teuer für großangelegte Durchmusterungen erscheinen. Für die weitere

Optimierung des Systems hätten Sonden einzeln im Vorfeld im Detail hinsichtlich ihres Hybridisierungsverhaltens erprobt werden müssen, was enormen Arbeitsaufwand und finanzielle Mittel vorausgesetzt hätte, die im Rahmen der Arbeit nicht mehr aufzubringen waren. Abschließend kann noch diskutiert werden, ob eine parallele Verwendung von Skip- und Junction -Sonden erfolgsversprechender gewesen wäre. Der Vorteil hätte sicherlich in der Tatsache bestanden, dass aufgrund der negativen Selektion (Abnahme der Signalintensität bei Mutation) mittels Junction-Sonden die Information vorhanden gewesen wäre, ob eine Sonde überhaupt ein zufrieden stellendes Signal erzielt (Wildtyp Intensität). Diese Vorgehensweise hätte jedoch mehr als eine Verdreifachung der Kosten beinhaltet. Die Weiterentwicklung des Systems wurde daher eher in der Übertragung der Idee auf die neuauftkommenden Hochdurchsatzsequenzierverfahren gesehen, sodass eine grundlegende Betrachtung dieser Verfahrensweise angegangen wurde.

#### **4.4 Transkriptom Charakterisierung mittels Sequenzierung**

Für die globale Charakterisierung ganzer Transkriptome existieren zwei unterschiedliche Ansätze: Zum einen die Analyse mittels Hybridisierung und zum anderen die Untersuchung mittels Sequenzierung.

Die Hybridisierung modifizierter RNA auf Oligonukleotid-Microarrays erlaubt die Betrachtung einer Vielzahl verschiedener Gewebe, beschränkt sich dabei allerdings auf nur ausgewählte Sequenzen (Lennon et al., 1991; Southern 1992). Die "Whole-genome tiling Arrays" erlauben eine detailliertere Analyse des Transkriptoms konnten sich jedoch aufgrund der sehr hohen Kosten und der Komplexität der Datenauswertung nicht vollständig durchsetzen (Bertone et al., 2004; Cheng et al., 2005). Zudem eignen sich diese technisch nicht zur Analyse alternativen Spleißens. Arrays, die speziell zur Analyse alternativen Spleißens entworfen wurden, weisen grundsätzlich den Nachteil auf, dass sich nur bekannte Transkripte charakterisieren lassen und sich durch die eingeschränkten Möglichkeiten der Selektion optimaler Sonden oft Probleme hinsichtlich der Spezifität ergeben (Calarco et al., 2007; Pan et al., 2004). Diese Schwierigkeiten zeigen sich auch bei den in dieser Arbeit unternommenen Versuchen ein Array zur Detektion Mutagen-induzierter Exon-Deletionen zu etablieren, die in den vorangehenden Kapiteln beschrieben wurden. Die resultierenden Unsicherheiten der entwickelten Methode und die Schwierigkeiten der Analyse genauer Fehlerquellen sind diesem und den in der Literatur beschriebenen Verfahren gemein (Fehlbaum et al., 2005) und ließen die Analyse mittels Sequenzierung als Alternative in den Vordergrund treten. Hierzu trugen insbesondere die parallel zu dieser Arbeit stattfindenden

immensen technischen Weiterentwicklungen der Sequenzierverfahren zweiter Generation bei (second generation sequencing, next generation sequencing [NGS]).

Generell weisen Sequenz-basierte Verfahren gegenüber Array-basierten die Vorteile auf, dass keine Limitierung der Analyse durch die Vorauswahl der Sonden besteht, mit Hinblick auf die analogen Signale der Arrays den digitalen Signalen keine Grenzen gesetzt sind und keine komplexen Normalisierungen stattfinden müssen. Allerdings wiesen die bisherigen Verfahren wie z.B. der des EST (expressed sequence tag) Sequenzierens, SAGE (serial analysis of gene expression) (Velculescu et al., 1995) und MPSS (massively parallel signature sequencing) (Brenner et al., 2000) den Nachteil sehr hoher Kosten auf. Mittels SAGE und MPSS können alternative Spleiß-Ereignisse nicht gefunden werden; zum Nachweis solcher Ereignisse wurde bis dato die Analyse von EST-Datensätzen eingesetzt (Gupta et al., 2004). Neben den hohen Kosten sind EST-basierte Ansätze jedoch sehr anfällig für Fehler, die meist auf dem Klonierungsschritt basieren. Der Vorteil der neuartigen Sequenzierverfahren, wie z.B. Polony Sequencing (Shendure et al., 2005), SOLiD (Cloonan et al., 2008), 454 Pyrosequencing (Margulies et al., 2005) und die Illumina Solexa Technologie (Bentley 2006), liegt insbesondere in den großen Datenmengen begründet, die zu immer moderateren Kosten generiert werden können und in Zukunft wohl die Analyse eines humanen Genoms für nur 1.000\$ ermöglichen werden (Guryev et al., 2009).

Das Potential, diese Techniken zur Analyse ganzer Transkriptome (RNA-Seq, RNA short read high-through-put sequencing) einsetzen zu können, wurde zunächst anhand der Polony Multiplex Analyse an der Maus gezeigt (Kim et al., 2007). Weitere Transkriptom-Analysen auf der Basis einer Einzelnukleotidauflösung wurden an den Pflanzen *Arabidopsis thaliana*, *Medicago truncatula* und *Zea mays* erfolgreich durchgeführt (Cheung et al., 2006; Emrich et al., 2007; Weber et al., 2007) und zeigten so die Möglichkeit auf, die Technik auch zur Analyse des Transkriptoms mit Hinblick auf die Detektion von Exon-Deletionsmutationen zu nutzen. Aufgrund der höheren Anzahl kurzer Reads und der größeren Datenmenge (Audic et al., 1997) wurde in dieser Arbeit die Analyse mittels der Solexa Technologie gewählt. Da bis dato keine Protokolle und Veröffentlichungen zur Analyse von Transkriptomen mittels Illuminas Solexa Technologie vorlagen, mussten in einem 'Proof of Principle' Experiment zunächst die Machbarkeit erprobt und die Fehlerquellen aufgezeigt werden, bevor eine Auswertung der Daten hinsichtlich der Mutationsdetektion vorgenommen werden konnte. Die Suche nach potentiellen Exon-Deletionen konnte erst dann mittels eines Abgleichs zu einer Deletions-Referenzdatenbank erfolgen, die auf der für die Etablierung des Microarrays entworfenen Software zur Oligonukleotid-Synthese basierte.

#### 4.4.1 Sequenzierung des Transkriptoms muriner Stammzellen

Im Rahmen dieser Arbeit wurden drei Proben muriner RNA mittels RNA-Seq sequenziert. Dazu zählte RNA der verwendeten F1 ES-Zelllinie, RNA des ENU-mutagenisierten Klons RRxx006 mit bestätigter Exon 8 Deletionsmutation in *Nup37* und zu Kontrollzwecken extrahierte RNA aus dem Gehirn einer 12 Wochen alten männlichen Maus. Dabei wurden auf der Illumina 1G Plattform für alle drei Proben jeweils mehr als  $3 \times 10^6$  27 bp Reads erzeugt, entsprechend ca. 81 Mbp. Mehr als 58% der Reads der einzelnen Proben wiesen direkte Treffer in der RefSeq Datenbank auf und untermauerten damit die erfolgreiche Probenvorbereitung und Sequenzierung der RNA-Proben. Vergleichbare Werte sind in einer RNA-Seq Untersuchung des humanen Transkriptoms von HEK- und B-Zellen mittels Solexa Sequenzierung auf identischer Hardware erzielt worden. Hier konnten mit 50% der Reads gegen das Humane Genom und mit ca. 7% gegen eine Spleiß-Junction Datenbank Treffer erzielt werden (Sultan et al., 2008).

Von den untersuchten 20.904 RefSeq-Transkripten wurden in den ES-Zell Proben für ca. 69% der Referenzsequenzen mindestens ein eindeutiger Read verzeichnet, der die Expression des Transkriptes in der Zelllinie bestätigte. Mehr als 90% der Transkripte wurden allerdings durch zwei oder mehr Reads repräsentiert (Rosenkranz et al., 2008). Damit lag der Anteil transkribierter polyA Sequenzen um ca. 29% höher als es die vergleichbare Analyse mittels Illuminas RefSeq Microarray aufzeigte (3.2.1), was für die Sensitivität des Sequenzier-Verfahrens hinsichtlich der Ermittlung von Transkriptionsstärken spricht. Die Gruppe um Sultan et al. (2008) konnte ebenfalls 25% mehr transkribierte Gene mittels RNA-Seq identifizieren als in einer Microarray Analyse der gleichen RNA-Probe und zeigen, dass es sich dabei größtenteils um sehr schwach exprimierte Transkripte handelte, was die Sensitivität des Ref-Seq Verfahrens unterstreicht. Da statistisch gesehen die zufällige Verteilung eines Reads auf ein längeres Transkript wahrscheinlicher ist, wurde für die Berechnung der Transkriptionsstärken der Quotient aus Gesamtzahl eindeutiger Treffer des Transkriptes und dessen Länge in kb angenommen. So wurden beispielsweise für die Wildtyp Probe Werte zwischen 64.788,57 und 0,06 Reads/kb erzielt, die die weite dynamische Breite der Methode betonen. Für den direkten Vergleich der Expressionstärken verschiedener RNA-Seq Ansätze würde sich zusätzlich eine Einbeziehung der Gesamtzahl an Reads empfehlen, um so eine zusätzliche Normalisierung einzubringen. Weiterhin kann die Expressionstärke auch auf die Länge eines Exons bezogen werden, wie es von Sultan et al. (2008) berücksichtigt wurde, um über die Expressionstärke einzelner Exons detailliertere Informationen zu alternativem Spleißen zu erhalten. Diese Verfeinerungen waren in dieser Arbeit jedoch nicht erforderlich,

da das Ziel in der qualitativen Detektion von Exon-Deletionen gesehen wurde. Ein direkter Vergleich der Expressionsstärken der Transkripte in der F1 Wildtyp Probe, die mittels RNA-Seq und Microarray ermittelt wurden, zeigte einen Pearson'schen Korrelationskoeffizienten von 0,7. Eine ähnliche Studie, der Vergleich der Expressionstärken eines *Arabidopsis* Transkriptoms aus einem 454 Ref-Seq Ansatz mit einem ATH1 Array, ergab einen deutlich schlechteren Koeffizienten von 0,45 (Weber et al., 2007). Ein Wert von 0,7 ist sogar vergleichbar mit den Korrelationen, die in Vergleichen von Microarray-Plattformen mit technischen Replikaten berechnet wurden (Kuo et al., 2006). Diese Feststellung stützt damit die Funktionalität des hier beschriebenen RNA-Seq Ansatzes und unterstützt die Verwendung der normalisierten Expressionsstärken.

Eine bedeutende Voraussetzung für die Detektion von Exon-Deletionen bzw. Spleiß-Varianten mittels Sequenzierung ist eine ausreichend hohe Abdeckung des Transkriptoms (Richard et al., 2010). Einer Größe von ca  $57,56 \times 10^6$  bp des vollständigen Maus RefSeq Transkriptoms mit einer durchschnittlichen Expression von 69% aller Transkripte stehen mit den hier gezeigten Ergebnissen nur ca.  $36,02 \times 10^6$  bp eindeutig zuzuordnender Sequenzen (U-Reads) beispielsweise aus der Wildtyp-Probe gegenüber. Einer groben Abschätzung nach kann so mit einer Solexa Probe ('lane') unter den gegebenen Bedingungen nur eine einfache Abdeckung erzielt werden. Aufgrund der Verteilung der Reads nach der Expressionstärke der Transkripte ist allerdings bei wenigen Transkripten mit einer sehr hohen Abdeckung und bei vielen Transkripten mit einer sehr geringen Abdeckung zu rechnen. Diese Überlegung wird gestützt durch die hier gezeigte Studie zur Verteilung der Reads, die aufzeigt, dass 25% aller eindeutigen Reads auf nur 68 RefSeq Sequenzen fallen. Es war daher davon auszugehen, dass mit den hier sequenzierten Proben eine Detektion von Mutationen nur in stark exprimierten Genen möglich ist. Aufgrund der zufälligen Verteilung der 27 bp-Fragmenten in den Transkripten ist die zufällige Positionierung eines Reads auf einer Exon-Exon-Verbindung in schwach exprimierten Genen, die nur wenige Reads aufweisen relativ unwahrscheinlich. Eine Detektion mittels Oligonukleotid-Sonden auf einem Microarray ist hier wahrscheinlicher, sofern die Sonde die nötige Spezifität und Signalintensität leisten kann, da das Transkript vorrausgehenden Expressionanalysen nach in messbaren Mengen in voller Länge vorhanden sein sollte (ungeachtet der Option des mRNA mediated decay). Der zusätzlich messbare 3'-'bias' bedingt durch unvollständige reverse Transkription und anschließender zweifacher poly(A)<sup>+</sup> Selektion bewirkt zwar die Unterbewertung der Transkriptionsstärken in RefSeq Experimenten, ist aber eher zweitrangig für die angedachte Mutationsdetektion, so lange keine klare Abdeckung der Transkripte erreicht wird. Demnach unterstützen die hier

---

gezeigten Ergebnisse die Feststellungen von Weber et al. (2007), dass für Anwendungen wie z.B. der Ermittlung neuer Gene normalisierte cDNA Proben (Bogdanova et al., 2008) als Ausgangsmaterial notwendig sind.

#### **4.4.2 Abgleich der Sequenzdaten gegen die Exon-Skip Datenbank db42**

Die in dieser Arbeit generierten Sequenzdaten wiesen den Nachteil auf, dass keine vollständige Abdeckung des Transkriptoms erreicht wurde, sodass nicht alle putativen Exon-Skip Ereignisse detektierbar sein konnten. Richard et al. (2010) gehen davon aus, dass typische Sequenziertiefen, die nur ein bzw. zwei Solexa Reihen umfassen (ca. 8 Mio. Reads, Stand der Technik 2008), nur die Detektion von ca. der Hälfte aller Exon-Exon Junctions der transkribierten Gene erlaubt. Dennoch sollte insbesondere für stark exprimierte Gene eine ausreichende Abdeckung gewährleistet sein, was eine Auswertung der Daten zu Testzwecken durchaus legitimierte. Die Referenzdatenbank db42 enthielt sämtliche mögliche einfache Exon-Skip Ereignisse der zugrunde liegenden Maus RefSeq Sequenzdatenbank. Das Skript zur Generierung der Datenbank ließe sich jedoch problemlos auf das Fehlen mehrerer benachbarter Exons ausweiten, wie es für alternatives Spleißen durchaus üblich ist. In den bekannten Publikationen zu ENU-induzierten Exon-Deletionen beschränkte sich die Deletion jedoch durchweg auf ein einzelnes Exon basierend auf einer entsprechenden genomischen Punktmutation (Barbaric et al., 2007; Chen et al., 2000; Justice et al., 2000; Noveroske et al., 2000). In anderen Projekten, die sich vornehmlich mit der Detektion alternativer Spleißstellen und der Detektion unbekannter Isoformen beschäftigen, wurden in der Regel zunächst Abgleiche gegen das Genom durchgeführt und darauf aufbauend die resultierenden NM Reads gegen synthetische Spleiß-Junction Datenbanken untersucht (Sultan et al., 2008). Diese Vorgehensweise hat den Vorteil, dass auch Transkripte, die nicht in RefSeq enthalten sind, betrachtet werden und neben den Skip-Varianten auch sämtliche Junction-Varianten eines Gens quantifizierbar sind. Zwischenzeitlich wurden einige spezielle Software Programme publiziert, die erfolgreich anhand bestimmter Terminierungs-Codes von Introns und empirischen ermittelten Werten die Identifikation von Isoformen erlauben (Bryant et al., ; De Bona et al., 2008; Trapnell et al., 2009). Für die Analysen werden dabei lediglich die genomische Referenzsequenz und die RNA-Seq Datensätze herangezogen. Die in dieser Arbeit beschriebene Vorgehensweise hat den Nachteil, dass aufgrund der unvollständigen Grundlage der RefSeq Datenbank unbekannte Transkripte nicht erfasst werden und somit Skip-Ereignisse übersehen werden können. Der Vorteil liegt jedoch in der geringen Datenmenge der db42 und der klaren Differenzierung von Exon-Junction und Exon-Skip Ereignissen.

Die Länge der Referenzsequenzen in db42 wurde aus den unter 3.4.2 beschriebenen Gründen auf 42 bp festgelegt, um größtmögliche Sicherheit (>99,6%) in Bezug auf die Eindeutigkeit der Sequenz bei einem theoretisch angenommenen gleichem Vorkommen aller möglichen Nukleotidkombinationen zu schaffen. Insgesamt mussten so mindestens 6 Nukleotide eines 27 bp langen Reads über der Exon-Exon Grenze liegen. Die Fehleranfälligkeit des Solexa Verfahrens ist nicht zu unterschätzen. In Studien wurde gezeigt, dass die Fehlerrate zum 3'-Ende eines jeden Reads deutlich zunimmt. Während der Anteil fehlerhafter Nukleotideinbauten bis einschließlich Base 15 ca. 0,3% beträgt, nimmt der Anteil bis Base 27 auf mehr als 2% zu (Dohm et al., 2008). Diese Einschränkung, dass der Read mindestens 6 Basen über der Exon-Grenze liegen muss, ist die einzige Einschränkung, die in der Untersuchung vorgenommen wurde.

Anders als der Anzahl von Sonden bei der beschriebenen Microarray-Analyse ist der Anzahl zu betrachtender Referenzsequenzen kein Limit gesetzt, sodass keine Einschränkung hinsichtlich der Unterbrechung des Leserasters oder der Expression eines Transkripts vorgenommen werden muss. Diese Tatsache stellt den klaren Vorteil des Verfahrens gegenüber der Microarray Methode dar. Db42 enthält daher 165.931 Sequenzen, die sämtliche Exon-Skip Ereignisse der RefSeq Datenbank widerspiegeln. Ob es sich dabei um alternative Spleiß-Ereignisse handelt, wurde von vornherein offen gelassen, da bei weitem nicht alle Ereignisse für alle Zelltypen bzw. Gewebe bekannt sind, was generell einen Abgleich mit einer Wildtyp Referenz erfordert.

Der Abgleich der Reads der beiden F1 ES-Zell Proben und der Maus-Gehirn Referenzprobe führte zu jeweils mindestens 2000 eindeutigen Treffern gegen db42, die insgesamt 3.335 unterschiedliche Skip-Ereignisse repräsentierten. In einer parallel durchgeführten Studie mit drei Mausgeweben (Gehirn, Leber, Muskel) konnte bei einer im Vergleich zu der vorliegenden Arbeit ca. 4 mal höheren Abdeckung gezeigt werden, dass alternatives Spleißen in 3.462 Genen der Maus auftrat (Mortazavi et al., 2008). Dass es sich dabei nur um einen Bruchteil der Ereignisse alternativen Spleißens des Maus Transkriptoms handeln konnte, wurde von den Autoren unterstrichen und auch mit der unterschiedlich starken Abdeckung je nach Expressionstärke der Transkripte begründet. Die Verteilung der Ereignisse alternativen Spleißens waren wie in der hier vorgestellten Untersuchung auch sehr breit verteilt: Während die Gehirn-Probe 1250 exklusive Ereignisse aufwies, waren es 276 für das Muskelgewebe und 284 für die Leber-Probe. Trotz der höheren Abdeckung sind diese Werte vergleichbar mit denen unserer Studie, in denen die Gehirn Probe als zusätzliche Referenz 1025 exklusive Ereignisse und der Wildtyp der Stammzelllinie 607 aufwies.

Erstaunlich war, dass die Anzahl exklusiver Ereignisse in der Probe RRx006 um 94% höher lag als in der Wildtyp-Probe. Während man bei den 607 exklusiven Skip-Ereignissen in der Wildtyp Probe davon ausgehen konnte, dass es sich um alternatives Spleißen handeln musste, war die Beurteilung der 1179 Skip-Ereignisse in der ENU-mutagenisierten Probe schwierig. Eine abschließende Beurteilung wäre anhand der RNA-Seq Daten nur möglich gewesen, wenn man mit Sicherheit, d.h. bei einer vollständigen Abdeckung des Transkriptoms, keine derartigen Ereignisse in der Wildtyp Referenz feststellen konnte. Die Abdeckung war allerdings für eine solche Betrachtung viel zu gering, sodass man für den größten Teil der Exon-Skip-Ereignisse zunächst alternatives Spleißen als Ursache annehmen musste. Ein Abgleich der 1179 Ereignisse gegen die Maus-EST Datenbank, der eine Eingrenzung auf bisher absolut unbekannte Ereignisse ermöglichte, zeigte dass für 54 Skip-Ereignisse keine Treffer erzielt wurden. Das entsprach allerdings auch dem Anteil von 4,3% Skip-Ereignissen für die kein dbEST Treffer in der Wildtyp-Probe erreicht wurde. Solange keine ausreichenden Daten zu alternativem Spleißen in der dem Wildtyp entsprechenden Zelllinie vorliegen, wäre man gezwungen, für sämtliche exklusive Skip-Ereignisse einer Mutanten-Probe einen alternativen Nachweis durchzuführen. Während diese Arbeit für den Nachweis einer neuartigen Spleiß Isoform eines Gens durchaus gerechtfertigt erscheint, so ist der Arbeitsaufwand für den Nachweis einer Mutagen induzierten Exon-Deletion in einer Durchmusterung tausender mutierter Klone wirtschaftlich noch nicht zu vertreten. Dennoch wurde an einer kleinen Stichprobe von 13 der 54 Skip-Ereignisse eine Kontrolle mittels PCR durchgeführt und es konnten 4 zusätzliche Exon-Deletionsmutationen in dem Klon bestätigt werden. Dies zeigte, dass die Erwartungen hinsichtlich der beschriebenen Mutationsfrequenzen generell gehalten wurde bzw. diese sogar deutlich unterschätzt wurde, da nur ein Bruchteil des Transkriptoms sequenziert und davon nur ein Bruchteil genauer analysiert wurde. Die Abdeckung reichte jedoch nicht aus, die Deletion von Exon 8 in *Nup37* in dem Klon RRxx006 zu bestätigen.

#### **4.5 Potential der Transkriptom Durchmusterungsstrategien**

Mit dieser Arbeit wurde der Versuch unternommen, einen Gen-getriebenen Ansatz zur Durchmusterung ENU mutierter ES-Zelllinien nach Exon-Deletionsmutationen zu entwickeln. Das Augenmerk der Strategie lag dabei nicht auf der Durchmusterung nach Mutationen in ausgewählten Kandidaten-Genen, sondern vielmehr auf der parallelen Durchmusterung ganzer Transkriptome in großangelegten Studien. Der in dieser Arbeit entwickelte Microarray basierte auf einer Detektion mittels Exon-Skip Oligonukleotidsonden,

die die minimalste und damit auch kostengünstigste Variante der Transkriptom-weiten parallelen Durchmusterung darstellte. Der Array erlaubte rein wirtschaftlich betrachtet die Durchmusterung derart großer ES-Zell Klonbibliotheken, dass für jedes transkribierte Gen mindestens eine Exon-Deletionsmutation zur Verfügung gestellt werden könnte. Der Versuch, über optimale OligonukleotidAuswahl und Pooling die Kosten so gering zu halten, dass das System Zukunft hatte, resultierte in sehr hohen Anforderungen an die Spezifität der Sonden. Während in 'Proof of Principle' Experimenten gute Ergebnisse erzielt wurden, konnte das System in Test-Durchmusterungen nicht bestehen, da die Anzahl detektierter Exon-Deletionen weit unter den Erwartungen lagen. Das System wurde dadurch unwirtschaftlich. Probleme wie der begrenzte Spielraum bei der Wahl der Sonden und der daraus resultierende Kampf um die Spezifität bei ausreichender Signalintensität der Sonden ließen zudem nur geringfügige Optimierungsmöglichkeiten offen. Ähnliche parallel entwickelte Systeme nutzen deutlich mehr Oligonukleotidsonden zur Charakterisierung alternativen Spleißens und wurden auch intensiv in den letzten Jahren genutzt, weisen aber dennoch deutliche Einschränkungen auf, die einer Verwendung als etablierter Standardmethode im Wege stehen (Wang et al., 2010). Aufgrund der enormen Kosten ist eine Verwendung derartiger Microarrays in Hochdurchsatz-Durchmusterungen von Tausenden von Klonen vorerst nicht denkbar.

Rein technisch betrachtet kommen keine Zweifel auf, dass RNA-Seq das wohl derzeit größte Potential zur qualitativen als auch quantitativen Analyse alternativen Spleißens aufweist und damit durchaus als Option zur Detektion von Mutagen-induzierten Exon-Deletionen genutzt werden kann. Voraussetzung für die praktische Umsetzung in Durchmusterungen tausender Klone sind dennoch zum einen die vollständige Information über alternative Spleiß-Ereignisse in der Wildtyp-Referenz und zum anderen eine höhere und gleichmäßigere Abdeckung des Transkriptoms durch z.B. Verwendung normalisierter cDNA als Ausgangsmaterial bei gleichzeitig deutlich erhöhter Anzahl von Reads. Letztendlich sind die reinen Sequenzier-Kosten von ca. 800€ pro Probe in dieser Arbeit noch deutlich zu hoch für die Verwendung in Durchmusterungsverfahren. Hinsichtlich der Kosten pro bp wird allerdings eine deutliche Reduzierung in naher Zukunft aufgrund der rapiden Weiterentwicklung der Techniken erwartet (Guryev et al., 2009). Mit einem hohen Anstieg der pro Reaktion ermittelten Sequenzmengen können weiterhin auch Methoden wie die Multiplex Sequenzierung (Daines et al., 2009) zu deutlichen Kosteneinsparungen beitragen und RNA-Seq für Durchmusterungen interessant werden lassen.

## 5 Literaturverzeichnis

- Acevedo-Arozena, A., S. Wells, P. Potter, M. Kelly, R.D. Cox, and S.D. Brown.** 2008a. ENU mutagenesis, a way forward to understand gene function. *Annu Rev Genomics Hum Genet* **9**: 49-69.
- Acevedo-Arozena, A., S. Wells, P. Potter, M. Kelly, R.D. Cox, and S.D.M. Brown.** 2008b. ENU mutagenesis, a way forward to understand gene function. *Annu Rev Genomics Hum Genet* **9**: 49-69.
- Adams, D.J., E.T. Dermitzakis, T. Cox, J. Smith, R. Davies, R. Banerjee, J. Bonfield, J.C. Mullikin, Y.J. Chung, J. Rogers et al.** 2005. Complex haplotypes, copy number polymorphisms and coding variation in two recently divergent mouse strains. *Nat Genet* **37**: 532-536.
- Araki, M., K. Araki, and K. Yamamura.** 2009. International Gene Trap Project: towards gene-driven saturation mutagenesis in mice. *Curr Pharm Biotechnol* **10**: 221-229.
- Audic, S. and J.M. Claverie.** 1997. The significance of digital gene expression profiles. *Genome Res* **7**: 986-995.
- Austin, C.P., J.F. Battey, A. Bradley, M. Bucan, M. Capecchi, F.S. Collins, W.F. Dove, G. Duyk, S. Dymecki, J.T. Eppig et al.** 2004. The knockout mouse project. *Nat Genet* **36**: 921-924.
- Barbaric, I., S. Wells, A. Russ, and T.N. Dear.** 2007. Spectrum of ENU-induced mutations in phenotype-driven and gene-driven screens in the mouse. *Environ Mol Mutagen* **48**: 124-142.
- Barbazuk, W.B., S.J. Emrich, H.D. Chen, L. Li, and P.S. Schnable.** 2007. SNP discovery via 454 transcriptome sequencing. *Plant J* **51**: 910-918.
- Becker, S., M.H. de Angelis, and J. Beckers.** 2006. Use of chemical mutagenesis in mouse embryonic stem cells. *Methods Mol Biol* **329**: 397-407.
- Beckers, J. and M. Hrabe de Angelis.** 2002. Large-scale mutational analysis for the annotation of the mouse genome. *Curr Opin Chem Biol* **6**: 17-23.
- Bentley, D.R.** 2006. Whole-genome re-sequencing. *Curr Opin Genet Dev* **16**: 545-552.
- Bertone, P., V. Stolc, T.E. Royce, J.S. Rozowsky, A.E. Urban, X. Zhu, J.L. Rinn, W. Tongprasit, M. Samanta, S. Weissman et al.** 2004. Global identification of human transcribed sequences with genome tiling arrays. *Science* **306**: 2242-2246.
- Blencowe, B.J.** 2006. Alternative splicing: new insights from global analyses. *Cell* **126**: 37-47.
- Bogdanova, E.A., D.A. Shagin, and S.A. Lukyanov.** 2008. Normalization of full-length enriched cDNA. *Mol Biosyst* **4**: 205-212.
- Bradley, A., M. Evans, M.H. Kaufman, and E. Robertson.** 1984. Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. *Nature* **309**: 255-256.
- Brenner, S., M. Johnson, J. Bridgham, G. Golda, D.H. Lloyd, D. Johnson, S. Luo, S. McCurdy, M. Foy, M. Ewan et al.** 2000. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat Biotechnol* **18**: 630-634.
- Brown, S.D.M. and J.M. Hancock.** 2006. The mouse genome. *Genome Dyn* **2**: 33-45.
- Bryant, D.W., Jr., R. Shen, H.D. Priest, W.K. Wong, and T.C. Mockler.** Supersplat -- spliced RNA-seq alignment. *Bioinformatics*.
- Calarco, J.A., A.L. Saltzman, J.Y. Ip, and B.J. Blencowe.** 2007. Technologies for the global discovery and analysis of alternative splicing. *Adv Exp Med Biol* **623**: 64-84.
- Castle, J., P. Garrett-Engle, C.D. Armour, S.J. Duenwald, P.M. Loerch, M.R. Meyer, E.E. Schadt, R. Stoughton, M.L. Parrish, D.D. Shoemaker et al.** 2003.

- Optimization of oligonucleotide arrays and RNA amplification protocols for analysis of transcript structure and alternative splicing. *Genome Biol* **4**: R66.
- Cecconi, F. and B.I. Meyer.** 2000. Gene trap: a way to identify novel genes and unravel their biological function. *FEBS Lett* **480**: 63-71.
- Chen, Y., D. Yee, K. Dains, A. Chatterjee, J. Cavalcoli, E. Schneider, J. Om, R.P. Woychik, and T. Magnuson.** 2000. Genotype-based screen for ENU-induced mutations in mouse embryonic stem cells. *Nat Genet* **24**: 314-317.
- Cheng, J., P. Kapranov, J. Drenkow, S. Dike, S. Brubaker, S. Patel, J. Long, D. Stern, H. Tammana, G. Helt et al.** 2005. Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science* **308**: 1149-1154.
- Cheung, F., B.J. Haas, S.M.D. Goldberg, G.D. May, Y. Xiao, and C.D. Town.** 2006. Sequencing *Medicago truncatula* expressed sequenced tags using 454 Life Sciences technology. *BMC Genomics* **7**: 272.
- Chomczynski, P. and N. Sacchi.** 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* **162**: 156-159.
- Chou, C.-C., C.-H. Chen, T.-T. Lee, and K. Peck.** 2004. Optimization of probe length and the number of probes per gene for optimal microarray analysis of gene expression. *Nucleic Acids Res* **32**: e99.
- Clark, T.A., C.W. Sugnet, and M. Ares, Jr.** 2002. Genomewide analysis of mRNA processing in yeast using splicing-specific microarrays. *Science* **296**: 907-910.
- Cloonan, N., A.R. Forrest, G. Kolle, B.B. Gardiner, G.J. Faulkner, M.K. Brown, D.F. Taylor, A.L. Steptoe, S. Wani, G. Bethel et al.** 2008. Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat Methods* **5**: 613-619.
- Colbert, T., B.J. Till, R. Tompa, S. Reynolds, M.N. Steine, A.T. Yeung, C.M. McCallum, L. Comai, and S. Henikoff.** 2001. High-throughput screening for induced point mutations. *Plant Physiol* **126**: 480-484.
- Daines, B., H. Wang, Y. Li, Y. Han, R. Gibbs, and R. Chen.** 2009. High-throughput multiplex sequencing to discover copy number variants in *Drosophila*. *Genetics* **182**: 935-941.
- De Bona, F., S. Ossowski, K. Schneeberger, and G. Ratsch.** 2008. Optimal spliced alignments of short sequence reads. *Bioinformatics* **24**: i174-180.
- Doetschman, T., R.G. Gregg, N. Maeda, M.L. Hooper, D.W. Melton, S. Thompson, and O. Smithies.** 1987. Targetted correction of a mutant HPRT gene in mouse embryonic stem cells. *Nature* **330**: 576-578.
- Dohm, J.C., C. Lottaz, T. Borodina, and H. Himmelbauer.** 2008. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Res* **36**: e105.
- Eberwine, J.** 1996. Amplification of mRNA populations using aRNA generated from immobilized oligo(dT)-T7 primed cDNA. *Biotechniques* **20**: 584-591.
- Edgley, M., A. D'Souza, G. Moulder, S. McKay, B. Shen, E. Gilchrist, D. Moerman, and R. Barstead.** 2002. Improved detection of small deletions in complex pools of DNA. *Nucleic Acids Res* **30**: e52.
- Eggan, K., A. Rode, I. Jentsch, C. Samuel, T. Hennek, H. Tintrup, B. Zevnik, J. Erwin, J. Loring, L. Jackson-Grusby et al.** 2002. Male and female mice derived from the same embryonic stem cell clone by tetraploid embryo complementation. *Nat Biotechnol* **20**: 455-459.
- Emrich, S.J., W.B. Barbazuk, L. Li, and P.S. Schnable.** 2007. Gene discovery and annotation using LCM-454 transcriptome sequencing. *Genome Res* **17**: 69-73.
- Faustino, N.A. and T.A. Cooper.** 2003. Pre-mRNA splicing and human disease. *Genes Dev* **17**: 419-437.

- Fehlbaum, P., C. Guihal, L. Bracco, and O. Cochet.** 2005. A microarray configuration to quantify expression levels and relative abundance of splice variants. *Nucleic Acids Res* **33**: e47.
- Fire, A., S. Xu, M.K. Montgomery, S.A. Kostas, S.E. Driver, and C.C. Mello.** 1998. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**: 806-811.
- Friedel, R.H., C. Seisenberger, C. Kaloff, and W. Wurst.** 2007. EUCOMM--the European conditional mouse mutagenesis program. *Brief Funct Genomic Proteomic* **6**: 180-185.
- Glenister, P.H. and C.E. Thornton.** 2000. Cryoconservation--archiving for the future. *Mamm Genome* **11**: 565-571.
- Gordon, J.W. and F.H. Ruddle.** 1981. Integration and stable germ line transmission of genes injected into mouse pronuclei. *Science* **214**: 1244-1246.
- Greber, B., H. Lehrach, and H. Himmelbauer.** 2005. Mouse splice mutant generation from ENU-treated ES cells--a gene-driven approach. *Genomics* **85**: 557-562.
- Gu, H., J.D. Marth, P.C. Orban, H. Mossmann, and K. Rajewsky.** 1994. Deletion of a DNA polymerase beta gene segment in T cells using cell type-specific gene targeting. *Science* **265**: 103-106.
- Gurdon, J.B. and D.A. Melton.** 2008. Nuclear reprogramming in cells. *Science* **322**: 1811-1815.
- Guryev, V. and E. Cuppen.** 2009. Next-generation sequencing approaches in genetic rodent model systems to study functional effects of human genetic variation. *FEBS Lett* **583**: 1668-1673.
- Hacking, D.F.** 2008. 'Knock, and it shall be opened': knocking out and knocking in to reveal mechanisms of disease and novel therapies. *Early Hum Dev* **84**: 821-827.
- Hahnke, K., M. Jacobsen, A. Gruetzkau, J.R. Gruen, M. Koch, M. Emoto, T.F. Meyer, A. Walduck, S.H. Kaufmann, and H.J. Mollenkopf.** 2007. Striptease on glass: validation of an improved stripping procedure for in situ microarrays. *J Biotechnol* **128**: 1-13.
- Hannon, G.J.** 2002. RNA interference. *Nature* **418**: 244-251.
- Hansen, J., T. Floss, P. Van Sloun, E.M. Fuchtbauer, F. Vauti, H.H. Arnold, F. Schnutgen, W. Wurst, H. von Melchner, and P. Ruiz.** 2003. A large-scale, gene-driven mutagenesis approach for the functional analysis of the mouse genome. *Proc Natl Acad Sci U S A* **100**: 9918-9922.
- He, Z., L. Wu, M.W. Fields, and J. Zhou.** 2005. Use of microarrays with different probe sizes for monitoring gene expression. *Appl Environ Microbiol* **71**: 5154-5162.
- Hooper, M., K. Hardy, A. Handyside, S. Hunter, and M. Monk.** 1987. HPRT-deficient (Lesch-Nyhan) mouse embryos derived from germline colonization by cultured cells. *Nature* **326**: 292-295.
- Hrabe de Angelis, M.H., H. Flaswinkel, H. Fuchs, B. Rathkolb, D. Soewarto, S. Marschall, S. Heffner, W. Pargent, K. Wuensch, M. Jung et al.** 2000. Genome-wide, large-scale production of mutant mice by ENU mutagenesis. *Nat Genet* **25**: 444-447.
- Hu, G.K., S.J. Madore, B. Moldover, T. Jatko, D. Balaban, J. Thomas, and Y. Wang.** 2001. Predicting splice variant from DNA chip expression data. *Genome Res* **11**: 1237-1245.
- Hughes, T.R., M. Mao, A.R. Jones, J. Burchard, M.J. Marton, K.W. Shannon, S.M. Lefkowitz, M. Ziman, J.M. Schelter, M.R. Meyer et al.** 2001. Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nat Biotechnol* **19**: 342-347.
- Jaenisch, R. and R. Young.** 2008. Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. *Cell* **132**: 567-582.

- Johnson, J.M., J. Castle, P. Garrett-Engle, Z. Kan, P.M. Loerch, C.D. Armour, R. Santos, E.E. Schadt, R. Stoughton, and D.D. Shoemaker.** 2003. Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science* **302**: 2141-2144.
- Judd, B.** 2001. Experimental Organisms Used in Genetics. *Encyclopedia of Life Sciences*.
- Justice, M.J., D.A. Carpenter, J. Favor, A. Neuhauser-Klaus, M. Hrabe de Angelis, D. Soewarto, A. Moser, S. Cordes, D. Miller, V. Chapman et al.** 2000. Effects of ENU dosage on mouse strains. *Mamm Genome* **11**: 484-488.
- Justice, M.J., J.K. Noveroske, J.S. Weber, B. Zheng, and A. Bradley.** 1999. Mouse ENU mutagenesis. *Hum Mol Genet* **8**: 1955-1963.
- Kampa, D., J. Cheng, P. Kapranov, M. Yamanaka, S. Brubaker, S. Cawley, J. Drenkow, A. Piccolboni, S. Bekiranov, G. Helt et al.** 2004. Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22. *Genome Res* **14**: 331-342.
- Kim, J.B., G.J. Porreca, L. Song, S.C. Greenway, J.M. Gorham, G.M. Church, C.E. Seidman, and J.G. Seidman.** 2007. Polony multiplex analysis of gene expression (PMAGE) in mouse hypertrophic cardiomyopathy. *Science* **316**: 1481-1484.
- Kuhn, K., S.C. Baker, E. Chudin, M.H. Lieu, S. Oeser, H. Bennett, P. Rigault, D. Barker, T.K. McDaniel, and M.S. Chee.** 2004. A novel, high-performance random array platform for quantitative gene expression profiling. *Genome Res* **14**: 2347-2356.
- Kunath, T., G. Gish, H. Lickert, N. Jones, T. Pawson, and J. Rossant.** 2003. Transgenic RNA interference in ES cell-derived embryos recapitulates a genetic null phenotype. *Nat Biotechnol* **21**: 559-561.
- Kuo, W.P., F. Liu, J. Trimarchi, C. Punzo, M. Lombardi, J. Sarang, M.E. Whipple, M. Maysuria, K. Serikawa, S.Y. Lee et al.** 2006. A sequence-oriented comparison of gene expression measurements across different hybridization-based technologies. *Nat Biotechnol* **24**: 832-840.
- Lee, C. and M. Roy.** 2004. Analysis of alternative splicing with microarrays: successes and challenges. *Genome Biol* **5**: 231.
- Lennon, G.G. and H. Lehrach.** 1991. Hybridization analyses of arrayed cDNA libraries. *Trends Genet* **7**: 314-317.
- Li, Q., Z. Liu, H. Monroe, and C.T. Cuiat.** 2002. Integrated platform for detection of DNA sequence variants using capillary array electrophoresis. *Electrophoresis* **23**: 1499-1511.
- Mansour, S.L., K.R. Thomas, and M.R. Capecchi.** 1988. Disruption of the proto-oncogene int-2 in mouse embryo-derived stem cells: a general strategy for targeting mutations to non-selectable genes. *Nature* **336**: 348-352.
- Margulies, M., M. Egholm, W.E. Altman, S. Attiya, J.S. Bader, L.A. Bemben, J. Berka, M.S. Braverman, Y.-J. Chen, Z. Chen et al.** 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**: 376-380.
- Marschall, S., U. Huffstadt, R. Balling, and M. Hrabe de Angelis.** 1999. Reliable recovery of inbred mouse lines using cryopreserved spermatozoa. *Mamm Genome* **10**: 773-776.
- Martin, G.R.** 1981. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc Natl Acad Sci U S A* **78**: 7634-7638.
- Martin, G.R. and M.J. Evans.** 1975. Differentiation of clonal lines of teratocarcinoma cells: formation of embryoid bodies in vitro. *Proc Natl Acad Sci U S A* **72**: 1441-1445.
- McGlinchy, N.J. and C.W. Smith.** 2008. Alternative splicing resulting in nonsense-mediated mRNA decay: what is the meaning of nonsense? *Trends Biochem Sci* **33**: 385-393.
- Meissner, A., S. Eminli, and R. Jaenisch.** 2009. Derivation and manipulation of murine embryonic stem cells. *Methods Mol Biol* **482**: 3-19.

- Metzker, M.L.** 2009. Sequencing technologies - the next generation. *Nat Rev Genet* **11**: 31-46.
- Mockler, T.C., S. Chan, A. Sundaresan, H. Chen, S.E. Jacobsen, and J.R. Ecker.** 2005. Applications of DNA tiling arrays for whole-genome analysis. *Genomics* **85**: 1-15.
- Mortazavi, A., B.A. Williams, K. McCue, L. Schaeffer, and B. Wold.** 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* **5**: 621-628.
- Mouse-Sequencing-Consortium.** 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**: 520-562.
- Muller, H.J.** 1927. Artificial Transmutation of the Gene. *Science* **66**: 84-87.
- Mullis, K., F. Faloona, S. Scharf, R. Saiki, G. Horn, and H. Erlich.** 1986. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb Symp Quant Biol* **51 Pt 1**: 263-273.
- Munroe, R.J., R.A. Bergstrom, Q.Y. Zheng, B. Libby, R. Smith, S.W. John, K.J. Schimenti, V.L. Browning, and J.C. Schimenti.** 2000. Mouse mutants from chemically mutagenized embryonic stem cells. *Nat Genet* **24**: 318-321.
- Nagy, A., M. Gertsenstein, K. Vintersten, and R. Behringer.** 2003. *Manipulating the Mouse Embryo: A Laboratory Manual (Third Edition)*. Cold Spring Harbor Laboratory Press.
- Nichols, J., I. Chambers, T. Taga, and A. Smith.** 2001. Physiological rationale for responsiveness of mouse embryonic stem cells to gp130 cytokines. *Development* **128**: 2333-2339.
- Niwa, H., K. Araki, S. Kimura, S. Taniguchi, S. Wakasugi, and K. Yamamura.** 1993. An efficient gene-trap method using poly A trap vectors and characterization of gene-trap events. *J Biochem* **113**: 343-349.
- Nobrega, M.A., Y. Zhu, I. Plajzer-Frick, V. Afzal, and E.M. Rubin.** 2004. Megabase deletions of gene deserts result in viable mice. *Nature* **431**: 988-993.
- Nolan, P.M., A. Hugill, and R.D. Cox.** 2002. ENU mutagenesis in the mouse: application to human genetic disease. *Brief Funct Genomic Proteomic* **1**: 278-289.
- Nolan, P.M., J. Peters, M. Strivens, D. Rogers, J. Hagan, N. Spurr, I.C. Gray, L. Vizer, D. Brooker, E. Whitehill et al.** 2000. A systematic, genome-wide, phenotype-driven mutagenesis programme for gene function studies in the mouse. *Nat Genet* **25**: 440-443.
- Noveroske, J.K., J.S. Weber, and M.J. Justice.** 2000. The mutagenic action of N-ethyl-N-nitrosourea in the mouse. *Mamm Genome* **11**: 478-483.
- Ohtsuka, S. and S. Dalton.** 2008. Molecular and biological properties of pluripotent embryonic stem cells. *Gene Ther* **15**: 74-81.
- Oliver, P.L., E. Bitoun, and K.E. Davies.** 2007. Comparative genetic analysis: the utility of mouse genetic systems for studying human monogenic disease. *Mamm Genome* **18**: 412-424.
- Pan, Q., O. Shai, C. Misquitta, W. Zhang, A.L. Saltzman, N. Mohammad, T. Babak, H. Siu, T.R. Hughes, Q.D. Morris et al.** 2004. Revealing global regulatory features of mammalian alternative splicing using a quantitative microarray platform. *Mol Cell* **16**: 929-941.
- Pease, A.C., D. Solas, E.J. Sullivan, M.T. Cronin, C.P. Holmes, and S.P. Fodor.** 1994. Light-generated oligonucleotide arrays for rapid DNA sequence analysis. *Proc Natl Acad Sci U S A* **91**: 5022-5026.
- Pennacchio, L.A.** 2003. Insights from human/mouse genome comparisons. *Mamm Genome* **14**: 429-436.
- Philips, A.V. and T.A. Cooper.** 2000. RNA processing and human disease. *Cell Mol Life Sci* **57**: 235-249.

- Pretsch, W., J. Favor, W. Lehmacher, and A. Neuhauser-Klaus.** 1994. Estimates of the radiation-induced mutation frequencies to recessive visible, dominant cataract and enzyme-activity alleles in germ cells of AKR, BALB/c, DBA/2 and (102xC3H)F1 mice. *Mutagenesis* **9**: 289-294.
- Pruitt, K.D., T. Tatusova, and D.R. Maglott.** 2007. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* **35**: D61-65.
- Rathkolb, B., T. Decker, E. Fuchs, D. Soewarto, C. Fella, S. Heffner, W. Pargent, R. Wanke, R. Balling, M. Hrabe de Angelis et al.** 2000. The clinical-chemical screen in the Munich ENU Mouse Mutagenesis Project: screening for clinically relevant phenotypes. *Mamm Genome* **11**: 543-546.
- Religio, A., C. Ben-Dov, M. Baum, M. Ruggiu, C. Gemund, V. Benes, R.B. Darnell, and J. Valcarcel.** 2005. Alternative splicing microarrays reveal functional expression of neuron-specific regulators in Hodgkin lymphoma cells. *J Biol Chem* **280**: 4779-4784.
- Richard, H., M.H. Schulz, M. Sultan, A. Nurnberger, S. Schrunner, D. Balzereit, E. Dagand, A. Rasche, H. Lehrach, M. Vingron et al.** 2010. Prediction of alternative isoforms from exon expression levels in RNA-Seq experiments. *Nucleic Acids Res.*
- Rinchik, E.M.** 1991. Chemical mutagenesis and fine-structure functional analysis of the mouse genome. *Trends Genet* **7**: 15-21.
- Rosenkranz, R., T. Borodina, H. Lehrach, and H. Himmelbauer.** 2008. Characterizing the mouse ES cell transcriptome with Illumina sequencing. *Genomics* **92**: 187-194.
- Rozen, S. and H. Skaletsky.** 2000. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* **132**: 365-386.
- Russell, W.L., P.R. Hunsicker, D.A. Carpenter, C.V. Cornett, and G.M. Guinn.** 1982. Effect of dose fractionation on the ethylnitrosourea induction of specific-locus mutations in mouse spermatogonia. *Proc Natl Acad Sci U S A* **79**: 3592-3593.
- Russell, W.L., E.M. Kelly, P.R. Hunsicker, J.W. Bangham, S.C. Maddux, and E.L. Phipps.** 1979. Specific-locus test shows ethylnitrosourea to be the most potent mutagen in the mouse. *Proc Natl Acad Sci U S A* **76**: 5818-5819.
- Saiki, R.K., D.H. Gelfand, S. Stoffel, S.J. Scharf, R. Higuchi, G.T. Horn, K.B. Mullis, and H.A. Erlich.** 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**: 487-491.
- Sanger, F., S. Nicklen, and A.R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* **74**: 5463-5467.
- Schena, M., D. Shalon, R.W. Davis, and P.O. Brown.** 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**: 467-470.
- Schimenti, J.C., B.J. Libby, R.A. Bergstrom, L.A. Wilson, D. Naf, L.M. Tarantino, A. Alavizadeh, A. Lengeling, and M. Bucan.** 2000. Interdigitated deletion complexes on mouse chromosome 5 induced by irradiation of embryonic stem cells. *Genome Res* **10**: 1043-1050.
- Sharov, A.A., Y. Piao, R. Matoba, D.B. Dudekula, Y. Qian, V. VanBuren, G. Falco, P.R. Martin, C.A. Stagg, U.C. Bassey et al.** 2003. Transcriptome analysis of mouse stem cells and early embryos. *PLoS Biol* **1**: E74.
- Shendure, J., G.J. Porreca, N.B. Reppas, X. Lin, J.P. McCutcheon, A.M. Rosenbaum, M.D. Wang, K. Zhang, R.D. Mitra, and G.M. Church.** 2005. Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* **309**: 1728-1732.
- Shibuya, T. and K. Morimoto.** 1993. A review of the genotoxicity of 1-ethyl-1-nitrosourea. *Mutat Res* **297**: 3-38.
- Shoemaker, D.D., E.E. Schadt, C.D. Armour, Y.D. He, P. Garrett-Engele, P.D. McDonagh, P.M. Loerch, A. Leonardson, P.Y. Lum, G. Cavet et al.** 2001.

- Experimental annotation of the human genome using microarray technology. *Nature* **409**: 922-927.
- Siddiqui, A.S., A.D. Delaney, A. Schnerch, O.L. Griffith, S.J.M. Jones, and M.A. Marra.** 2006. Sequence biases in large scale gene expression profiling data. *Nucleic Acids Res* **34**: e83.
- Singh-Gasson, S., R.D. Green, Y. Yue, C. Nelson, F. Blattner, M.R. Sussman, and F. Cerrina.** 1999. Maskless fabrication of light-directed oligonucleotide microarrays using a digital micromirror array. *Nat Biotechnol* **17**: 974-978.
- Skarnes, W.C., H. von Melchner, W. Wurst, G. Hicks, A.S. Nord, T. Cox, S.G. Young, P. Ruiz, P. Soriano, M. Tessier-Lavigne et al.** 2004. A public gene trap resource for mouse functional genomics. *Nat Genet* **36**: 543-544.
- Soewarto, D., M. Klaften, and I. Rubio-Aliaga.** 2009. Features and strategies of ENU mouse mutagenesis. *Curr Pharm Biotechnol* **10**: 198-213.
- Southern, E.M.** 1992. Genome mapping: cDNA approaches. *Curr Opin Genet Dev* **2**: 412-416.
- Srinivasan, K., L. Shiue, J.D. Hayes, R. Centers, S. Fitzwater, R. Loewen, L.R. Edmondson, J. Bryant, M. Smith, C. Rommelfanger et al.** 2005. Detection and measurement of alternative splicing using splicing-sensitive microarrays. *Methods* **37**: 345-359.
- Stamm, S., S. Ben-Ari, I. Rafalska, Y. Tang, Z. Zhang, D. Toiber, T.A. Thanaraj, and H. Soreq.** 2005. Function of alternative splicing. *Gene* **344**: 1-20.
- Sultan, M., M.H. Schulz, H. Richard, A. Magen, A. Klingenhoff, M. Scherf, M. Seifert, T. Borodina, A. Soldatov, D. Parkhomchuk et al.** 2008. A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science* **321**: 956-960.
- Sung, Y.H., J. Song, and H.W. Lee.** 2004. Functional genomics approach using mice. *J Biochem Mol Biol* **37**: 122-132.
- Thomas, J.W., C. LaMantia, and T. Magnuson.** 1998. X-ray-induced mutations in mouse embryonic stem cells. *Proc Natl Acad Sci U S A* **95**: 1114-1119.
- Thomas, K.R. and M.R. Capecchi.** 1987. Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell* **51**: 503-512.
- Trapnell, C., L. Pachter, and S.L. Salzberg.** 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**: 1105-1111.
- Tuschl, T.** 2001. RNA interference and small interfering RNAs. *ChemBiochem* **2**: 239-245.
- van Ruissen, F., J.M. Ruijter, G.J. Schaaf, L. Asgharnegad, D.A. Zwijnenburg, M. Kool, and F. Baas.** 2005. Evaluation of the similarity of gene expression data estimated with SAGE and Affymetrix GeneChips. *BMC Genomics* **6**: 91.
- Velculescu, V.E., L. Zhang, B. Vogelstein, and K.W. Kinzler.** 1995. Serial analysis of gene expression. *Science* **270**: 484-487.
- Vivian, J.L., Y. Chen, D. Yee, E. Schneider, and T. Magnuson.** 2002. An allelic series of mutations in Smad2 and Smad4 identified in a genotype-based screen of N-ethyl-N-nitrosourea-mutagenized mouse embryonic stem cells. *Proc Natl Acad Sci U S A* **99**: 15542-15547.
- Vooijs, M., M. van der Valk, H. te Riele, and A. Berns.** 1998. Flp-mediated tissue-specific inactivation of the retinoblastoma tumor suppressor gene in the mouse. *Oncogene* **17**: 1-12.
- Wang, H., E. Hubbell, J.S. Hu, G. Mei, M. Cline, G. Lu, T. Clark, M.A. Siani-Rose, M. Ares, D.C. Kulp et al.** 2003. Gene structure-based splice variant deconvolution using a microarray platform. *Bioinformatics* **19 Suppl 1**: i315-322.
- Wang, L., Y. Xi, J. Yu, L. Dong, L. Yen, and W. Li.** 2010. A statistical method for the detection of alternative splicing using RNA-seq. *PLoS One* **5**: e8529.

- Weber, A.P.M., K.L. Weber, K. Carr, C. Wilkerson, and J.B. Ohlrogge.** 2007. Sampling the Arabidopsis transcriptome with massively parallel pyrosequencing. *Plant Physiol* **144**: 32-42.
- Weber, J.S., A. Salinger, and M.J. Justice.** 2000. Optimal N-ethyl-N-nitrosourea (ENU) doses for inbred mouse strains. *Genesis* **26**: 230-233.
- Wen, J. and S. Brogna.** 2008. Nonsense-mediated mRNA decay. *Biochem Soc Trans* **36**: 514-516.
- Xiao, W. and P.J. Oefner.** 2001. Denaturing high-performance liquid chromatography: A review. *Hum Mutat* **17**: 439-474.
- Yu, J. and J.A. Thomson.** 2008. Pluripotent stem cell lines. *Genes Dev* **22**: 1987-1997.
- Zvetkova, I., A. Apedaile, B. Ramsahoye, J.E. Mermoud, L.A. Crompton, R. John, R. Feil, and N. Brockdorff.** 2005. Global hypomethylation of the genome in XX embryonic stem cells. *Nat Genet* **37**: 1274-1279.

## 6 Anhang

### 6.1 Abbildungsverzeichnis

Abbildung 1: Der Modellorganismus Maus .....	9
Abbildung 2: Lichtmikroskopische Aufnahme embryonaler F1 Stammzellkolonien .....	11
Abbildung 3: Molekulare Struktur von N-Ethyl-N-nitrosoharnstoff (ENU) .....	17
Abbildung 4: Die Mutagene Reaktion von ENU .....	18
Abbildung 5: Gen-getriebene Ansätze unter Verwendung von <i>in vitro</i> bzw. <i>in vivo</i> Mutagenesen der Maus .....	19
Abbildung 6: Fluoreszenzmikroskopische Aufnahme eines Karyogramms einer Maus ES-Zelle (F1) .....	32
Abbildung 7: Beispielhafte Darstellung der Integritätskontrolle isolierter Gesamt-RNA .....	36
Abbildung 8: Gelelektrophoretische Auftrennung von <i>Hprt</i> RT-PCR Produkten .....	48
Abbildung 9: Detektions Wahrscheinlichkeit mindestens einer Exon-Deletion in einem bestimmten Gen in Abhängigkeit von der Anzahl untersuchter ES-Zell Klone .....	50
Abbildung 10: Anzahl exprimierter RefSeq Gene in embryonalen Stammzellen der F1 Linie .....	52
Abbildung 11: Schematische Darstellung des Oligonukleotid Aufbaus zur Detektion von Exon-Deletionen .....	55
Abbildung 12: Signalintensitäten im Proof of Principle Experiment (Detektion von <i>Hprt</i> Deletion Exon 8) .....	60
Abbildung 13: Spezifität des Skip-Oligos , <i>Hprt</i> Exon 8' in Abhängigkeit von der Oligonukleotidlänge .....	61
Abbildung 14: Spezifität der Junction-Oligos , <i>Hprt</i> Exon 7-8 und Exon 8-9' in Abhängigkeit von der Oligonukleotidlänge .....	62
Abbildung 15: Signalintensitäten ausgewählter Sonden im Pooling Experiment .....	66
Abbildung 16: Hybridisierungsartefakte .....	68
Abbildung 17: Schematische Darstellung der PCR Systeme zur Bestätigung von Exon- Deletionen .....	69
Abbildung 18: Bestätigung selektierter Exon-Deletionen mittels PCR .....	70
Abbildung 19: Bestätigung alternativ gesplinter Exons mittels PCR .....	72
Abbildung 20: Darstellung des genutzten Durchmusterungsschema einer Klon-Gruppe (25 Klone) .....	73
Abbildung 21: Signalintensitäten ausgewählter Skip-Oligonukleotid Sonden .....	74
Abbildung 22: Signalintensitäten der Positiv-Kontrollen in 4 x 44k und 8 x 15 k Arrays .....	77
Abbildung 23: Allgemeine Erläuterung der ELAND Ergebnisse .....	80
Abbildung 24: Abdeckung des Transkriptoms durch U Reads nach Sequenzierung nicht normalisierter cDNA .....	84
Abbildung 25: Untersuchung des 3'-bias als mögliche Fehlerquelle .....	86
Abbildung 26: Verteilung der möglichen Exon-Skip Ereignisse auf die Proben und die Reads .....	88
Abbildung 27: Korrelation zwischen der Anzahl Exon-Skip spezifischer Reads und relativer Expression .....	90

## 6.2 Tabellenverzeichnis

Tabelle 1: Verwendete Agarosekonzentrationen in Agarosegelen .....	34
Tabelle 2: Zusammensetzung einer Erststrang-cDNA-Synthese .....	37
Tabelle 3: Zusammensetzung eines Standard-PCR-Ansatzes.....	39
Tabelle 4: Thermoprofil einer Standard-PCR .....	39
Tabelle 5: Verwendete Agilent One-Color Custom-Microarrays .....	41
Tabelle 6: Thermoprofil Cycle-Sequencing .....	43
Tabelle 7: Kommerzielle Software .....	44
Tabelle 8: Nukleotid-Sequenz-Datenbanken .....	45
Tabelle 9: Charakteristische Parameter unabhängiger ENU-Mutagenesen .....	47
Tabelle 10: Mutationspektrum eines Sets von 22 6-Thioguanin resistenten ES-Zell Klonen .	49
Tabelle 11: Einfluss der ENU Mutagenese auf das Transkriptom dreier ENU-Klone .....	53
Tabelle 12: Oligonukleotid Zusammenstellung des Proof-of-Principle Microarrays .....	59
Tabelle 13: Oligonukleotid Zusammenstellung des Pooling-Experiments.....	64
Tabelle 14: Übersicht über die hybridisierten RNA-Proben/-Pools.....	65
Tabelle 15: Informationen zu putativen Exon-Deletionen.....	67
Tabelle 16: Oligonukleotid Zusammenstellung des 8 x 15k Arrays.....	76
Tabelle 17: Informationen zu Deletions-Ereignissen des 8 x 15k Arrays.....	78
Tabelle 18: ELAND Abgleich der Sequenzdaten gegen die Mouse RefSeq Datenbank.....	83
Tabelle 19: Durchschnittlicher GC-Gehalt (%) von RefSeq und komplementären Reads .....	87
Tabelle 20: ELAND Abgleich der Sequenzdaten gegen die db42 Datenbank.....	88
Tabelle 21: Ergebnisse der PCR zur Bestätigung möglicher Exon-Deletions Mutationen.....	91

### 6.3 Oligonukleotid-Primer

Name	Sequenz	Seite
Hprt_UTR_FW	5' -CTT ACC TCA CTG CTT TCC GG-3'	40
Hprt_UTR_REV	5' -CTG GCA ACA TCA ACA GGA CTC-3'	40
oligodt <sub>18</sub>	5' -TTT TTT TTT TTT TTT TTT-3'	-
Arnt_Fw	5' -ACC TCA CTT TGT GGT AGT CCA-3'	63
Arnt_Rev	5' -TGC CCT GCA GAA GAT TAG-3'	63
Arnt_Skip (Rev)	5' -CCA CAG GCT GGA CAG AAA CCT GGA-3'	63
Mgea6_Fw	5' -AGC CTT GGA AGG AGA ACG A-3'	63
Mgea6_Rev	5' -TGG TCC AGA AGG TTC TCA GG-3'	63
Mgea6_Skip (Rev)	5' -ATT CTT ACC AGG GGC AGG CTG-3'	63
Stk10_Fw	5' -ATC ACG CTG ATC GAG ATG G-3'	63
Stk10_Rev	5' -CAT CTG CTC CAG CTG CAG CTC-3'	63
Stk10_Skip (Rev)	5' -TGC AGT TTT GAG CCC GGA-3'	63
Phf14_Fw	5' -ATG GTG CCA AGG AGT GTA GC-3'	63
Phf14_Rev	5' -CGC CTT CCA TCC ACA CTA GA-3'	63
Phf14_Skip (Rev)	5' -AGG CAG CAG CAG AAG AGG CAA-3'	63
Hat1_Fw	5' -ACG CGT AAG TCC TTG CGG TC-3'	63
Hat1_Rev	5' -TGT CAT GTA GCC TAC GGT CGC-3'	63
Hat1_Skip (Fw)	5' -GCT CGG AAA TGG CGG TTG CT-3'	63
Bub1b_Fw	5' -GGT CAT GTC CAC ACT TCA GG-3'	63
Bub1b_Rev	5' -AGC CTC CTC CTC TTC ATT CC-3'	63
Bub1b_Skip (Rev)	5' -CTG CAG CAG CAG AAA CGA TA-3'	63
Acin1_Fw	5' -CAG CCG ACT CAT CCT CTA GC-3'	63
Acin1_Rev	5' -GCT GAC TTG GTC TGC AAC AG-3'	63
Acin1_Skip (Rev)	5' -CAA GCA GCA GCA GAA AAG GG-3'	63
Nup37_Fw	5' -ACA CCT GCA GGA TTT GGA AC-3'	63
Nup37_Rev	5' -TCT GTC ACC CAG AAC AAG AGC-3'	63
Nup37_Skip (Rev)	5' -GAT ATC ACT CGG TCC AGG TG-3'	63
Anxa7_Fw	5' -GGA TAT CCT CCT GCA GGT CA-3'	65
Anxa7_Rev	5' -CAA CAT CCA CAA TTG CTT GC-3'	65
Vrk1_Fw	5' -AAG CCA GGT GAG ATC GCT AA-3'	65
Vrk1_Rev	5' -GCT CTC AGA CAG GCC ATA CC-3'	65
Fnl_Fw	5' -ATG CAC CGA TTG TCA ACA GA-3'	65
Fnl_Rev	5' -AGC TCT GCA ACG TCC TCT TC-3'	65
Scrib_Fw	5' -GGC AGC CAG AAT GAA GTC TC-3'	65
Scrib_Rev	5' -TTG GGA TAA GGG AGG TGA TG-3'	65
Zpf207_Fw	5' -ACC TCT GAT GCC AGG TAT GC-3'	65
Zpf207_Rev	5' -GAT ATC ACT CGG TCC AGG TG-3'	65
Lrp5_Fw	5' -TGG ATC GAC CGC CAG CAG CAG-3'	68
Lrp5_Rev	5' -GAG CTC ATC ACG GAC TTG CTG-3'	68
Lrp5_Skip (Fw)	5' -CTG CTG ACT TGT GGT GCT GTC-3'	68
Dagbl_Fw	5' -CTC CAT CCT GAA GAC CAC AGG GCT-3'	68
Dagbl_Rev	5' -CGT ACC AAC AGC CAT GCA GC-3'	68
Dagbl_Skip (Fw)	5' -GAC CAT GTC TCT CCA GGG AAT-3'	68
Gpr56_Fw	5' -GAG CTT CCT GCT CAG CGA G-3'	68
Gpr56_Rev	5' -CAC TGT ACG TTA GTA GGA CAG-3'	68
Gpr56_Skip (Rev)	5' -GAG ATG GAG GTC CCT CTT GG-3'	68
Pdxk_Fw	5' -ATC CAG AGC CAT GTT GTC C-3'	68
Pdxk_Rev	5' -CAT ACA GTG CAG CAT GTC CAT-3'	68
Pdxk_Skip (Fw)	5' -GTT TTC AAA CCA CAC AGG TTA C-3'	68
Acbd4_Fw	5' -CCA GGC AGC AGT CAG CGT C-3'	68
Acbd4_Rev	5' -TCC AGG TCC AGG TCT CTG GGT-3'	68
Acbd4_Skip (Fw)	5' -GAT GCC TGG AAC AGG GTG ATC-3'	68
Psmf1_Fw	5' -ATG CCC TCC GGT ATG AGT C-3'	68
Psmf1_Rev	5' -CAC AGG TAC ATG TCA TCG TAG-3'	68
Psmf1_Skip (Rev)	5' -ATG GTC AGG GTT AGG TCC AA -3'	68
Pik3r4_Fw	5' -AGA TCT GGA ACA GCC AGA AGA-3'	68
Pik3r4_Rev	5' -CAT GTC TGA GCC AGC GGT CAG-3'	68
Pik3r4_Skip (Fw)	5' -GTG CTG GCT GTG CAT AGC TGT-3'	68
Khdrbs3_Fw	5' -GTG CTC ATT GAA GTG TTC GCA -3'	68

---

Khdrbs3_Rev	5'-AGT AGG CAT GAT GGT GAG GTG -3'	68
Khdrbs3_Skip (Fw)	5'-TCC TGT ATA CTG TTC AAA GAA ATG -3'	68
Nek9_Fw	5'-ATA GCA GTG GCC TGT CCA TC-3'	68
Nek9_Rev	5'-GAG CCT GCA ACC TGT CAA C-3'	68
Nek9_Skip (Rev)	5'-ATC CAG CCG AGG GGG CCA GG-3'	68
Mbtpsl_Fw	5'-ACT GAC TGT CTA CCT GTG GAA G-3'	68
Mbtpsl_Rev	5'-CAT CTG CCT GCA GTG TCT GG-3'	68
Mbtpsl_Skip (Fw)	5'-CAA GAT GCT CTC ACA GAA TAT-3'	68
Epb4.113_Fw	5'-ATG CAT GGC TTC GAC GAC TTG-3'	71
Epb4.113_Rev	5'-CAC TCA TTA GCA CAC CTG GC-3'	71
Epb4.113_Skip (Fw)	5'-CCT CAG AGG CCG TGA GGT CAT C-3'	71
Mod1_Fw	5'-TGG TTG TTA TGG CCA TGG AGA-3'	71
Mod1_Rev	5'-GTA GGA TCT GGT CAT AAT TAG-3'	71
Mod1_Skip (Rev)	5'-GTA TGC ATC TTG CAC AAT CTC AG-3'	71
skip1.003700_Fw	5'-CCC AGC TCA CTG ACC TAA GC-3'	84
skip1.003700_Rev	5'-TGA ATG AGT TGG CGG TGA TA-3'	84
skip1.003700_Skip (Fw)	5'-ACT GAC ACA GCA AAC GGA TC-3'	84
skip1.014138_Fw	5'-CCG CTA CAA GGA AGA GAT GG-3'	84
skip1.014138_Rev	5'-GTG TGC TGT GGA GTT CAG GA-3'	84
skip1.014138_Skip (Fw)	5'-AGG CAG AGT TAA TTC AGG TA-3'	84
skip1.016954_Fw	5'-CAA CTC AGT TGC GTA TAG CG-3'	84
skip1.016954_Rev	5'-TCT TCC AGA ACC TTC CAT CAC-3'	84
skip1.016954_Skip (Rev)	5'-CAG AAA CTT CTC CAC AAA CTT C-3'	84
skip1.018778-Fw	5'-CTC GGT CAG ACC TCG TTA GC-3'	84
skip1.018778-Rev	5'-TGG TTT TCA CTG GAG GAA GG-3'	84
skip1.018778-Skip (Fw)	5'-TCA TTG GAG CTG GGA AGG GT-3'	84
skip1.105065_Fw	5'-ATG CCT CCA GCA CCA ATT AC-3'	84
skip1.105065_Rev	5'-CGG GCA TAA AGG AAT GAG AA-3'	84
skip1.105065_Skip (Fw)	5'-GAG AAT GCA GCC TGC AGA AA-3'	84
skip1.121294_Fw	5'-ATC ACA GAG TAC ATG GAG CAT-3'	84
skip1.121294_Rev	5'-ACA TCA CCA GGC ACA TCA CTG-3'	84
skip1.121294_Skip (Fw)	5'-GAC ATA TGA GTT GAT CAG ACA-3'	84
skip1.169341_Fw	5'-CAA CCA CTT ATA TGA TCT CA-3'	84
skip1.169341_Rev	5'-TTG ATG GCC ATG GTC CAG AG-3'	84
skip1.169341_Skip (Fw)	5'-CTA CCT CCT CCA ACC CCA AAT-3'	84
skip1.177973_Fw	5'-TGA CAA GGC TCT TGA TGG CT-3'	84
skip1.177973_Rev	5'-TCT GTT GAT GCA GAC GTC AC-3'	84
skip1.177973_Skip (Fw)	5'-ACA CGG AAA AGC AGA TTA GAT-3'	84
skip1.197421_Fw	5'-TTG GCA AGA GAG CCA TTT CT-3'	84
skip1.197421_Rev	5'-AGC ATG GAC CAT TGG ATA CG-3'	84
skip1.197421_Skip (Fw)	5'-GAT GAG AAA GTT CAG TTG T-3'	84
skip1.203294_Fw	5'-GTC GCA TTG GTT CAT GAG TG-3'	84
skip1.203294_Rev	5'-GAT TCC CCG TTT CCC ATA TT-3'	84
skip1.203294_Skip (Fw)	5'-CGA TCC TCT CAC AGA CGA A-3'	84
skip1.211338_Fw	5'-AGC AGT TCG TCT CGG CGA CC-3'	84
skip1.211338_Rev	5'-CTT CAT AGC TCA TTA GTG TTG C-3'	84
skip1.211338_Skip (Fw)	5'-GGA GCC GTC ACC ATG CCC GTT G-3'	84
skip1.212648_Fw	5'-AGT CAC AGC CTG TCC TCA GCT-3'	84
skip1.212648_Rev	5'-TCC CCA AGC CCT CAC TGT AG-3'	84
skip1.212648_Skip (Fw)	5'-ATC CTG GGC ACT AAA GCC AGT-3'	84
skip1.212923_Fw	5'-TAG AGT CAG CTT GCT TCA GTG-3'	84
skip1.212923_Rev	5'-CAT TCT GAG CTA CAG TAT TG-3'	84
skip1.212923_Skip (Rev)	5'-GAA CAA CTT GGA ATC ACC TTT C-3'	84

---

## 6.4 Abkürzungen

129Sv	Inzuchtlinie der Maus
6-TG	6-Thioguanin
A <sub>230</sub>	Absorption des Lichts bei einer wellenlänge von 230 nm
AGT	O <sup>6</sup> -Alkylguanin-DNA-alkyltransferase
aRNA	Amplifizierte RNA
ATP	Adenosintriphosphat
BG	Bezeichnung bestimmter Stammzellklone
BLAST	<i>Basic Local Alignment Search Tool</i>
BLAT	<i>BLAST like alignment tool</i>
BMP	<i>Bone morphogenetic protein</i>
bp	Basenpaare
BSA	Rinderserumalbumin
CD1	Auszuchtlinie der Maus
cDNA	Komplementär DNA
CDS	Kodierende Sequenz eines Gens
CEL1	CEL1-Endonuklease aus Sellerie
CGR8	Inzuchtlinie der Maus
ChIP	<i>Chromatin immuno Precipitation</i>
cRNA	<i>Copy RNA</i>
CTP	Cytosintriphosphat
Cy3	Cyanin-3
D	Delta, Deletion
DAPI	4',6-Diamidin-2-phenylindol
db	Datenbank
db42	Exon-Deletions Datenbank
ddNTP	Didesoxyribonukleosid-Triphosphat
DICER	Helikase
DMEM	<i>Dulbecco's Modified Eagle's Medium</i>
DMSO	Dimethylsulfoxid
DNA	Desoxyribonukleinsäure
DNase	Desoxyribonuklease
ds-cDNA	doppelsträngige cDNA
dsDNA	doppelsträngige DNA
DTT	Dithiothreitol
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	Ethylendiamintetraessigsäure
EMS	Ethylmethansulfonsäure
ENU	N-Ethyl-N-nitrosoharnstoff
EST	<i>Expressed sequence tags</i>
ES-Zellen	Embryonale Stammzellen
EtBr	Ethidiumbromid
F1	Hybrid Stammzelllinie: Filialgeneration 1 (C57BL/6J x 129Sv)
FASTA	Textbasiertes Format zur Speicherung von Sequenzen

---

FCS	Fötale Rinderserum
G	Generation
g	Erdbeschleunigung
GTP	Guanintriphosphat
GuHCl	Guanidinhydrochlorid
HPLC	<i>High-performance liquid chromatography</i>
<i>Hprt</i>	Hypoxanthin Phosphoribosyl Transferase
Id	<i>Inhibitor of differentiation</i>
ID	Identifikationsnummer
IL6	Interleukin-6
iPS	Induzierte pluripotente Stammzellen
JAK-STAT3	Signalweg
KO	<i>Knockout</i>
LIF	<i>Leukemia Inhibitory Factor</i>
M	Molar
MEF	Embryonale Fibroblasten der Maus
MF	Mutationsfrequenz
MGI	Mouse Genome Informatics
M-MLV	<i>Moloney Murine Leukemia Virus</i>
MPSS	<i>Massively Parallel Signature Sequencing</i>
Mut	Mutante
NM	Solexa Read ohne Treffer in der entsprechenden Datenbank
ORF	Offenes Leseraster eines Gens
p	Wahrscheinlichkeit
p. c.	<i>post coitum</i>
PBS	<i>phosphate buffered saline</i>
PCR	Polymerase Kettenreaktion
pH	<i>Potentia Hydrogenii</i>
QC	Solexa Read schlechter Qualität
R	Pearson'schen Korrelationskoeffizient
R0, R1, R2	Solexa Read der mehrere Referenzsequenzen trifft
R <sup>2</sup>	Bestimmtheitsmaß
RACE	<i>Rapid Amplification of cDNA Ends</i>
Ras-MAPK	Signalweg
RefSeq	<i>NCBI Reference sequence</i> Datenbank
Risc	<i>RNAi silencing complex</i>
RNA	Ribonukleinsäure
Rnase	Ribonuklease
RNA-seq	Hochdurchsatz cDNA Sequenzierverfahren
rr	Bezeichnung bestimmter Stammzellklone
RR	Bezeichnung bestimmter Stammzellklone
RT	Reverse Transkription bzw. Reverse Transkriptase
SAGE	<i>Serial Analysis of Gene Expression</i>
shRNA	<i>Short hairpin RNA</i>
siRNA	<i>Small interfering RNA</i>

---

SN <sub>1</sub>	Nukleophile Substitution (monomolekularer Mechanismus)
SNP	Einzelnukleotid-Polymorphismen
SSC	<i>Standard Saline Citrate</i>
ssDNA	Einzelstrang DNA
TAE	Tris-Acetat-EDTA-Puffer
TE	Tris-EDTA-Puffer
TGCE	Temperatur Gradienten Elektrophorese
T <sub>m</sub>	Schmelztemperatur
Tris	Tris-(hydroxymethyl)-aminomethan
TTP	Thymidintriphosphat
U0, U1, U2	Solexa Read der exakt eine Referenzsequenz trifft
UTR	Untranslatierter Bereich einer mRNA
V.	Version
v/v	Volumen / Volumen
Vol	Volumen
w/v	Masse / Volumen
WT	Wildtyp

---

## 7 Lebenslauf des Verfassers

Ruben Josef Rosenkranz

geboren am 16.05.1979 in Jülich

- |                 |  |
|-----------------|--|
| 1989 - 1998     | Gymnasium Zitadelle in Jülich<br>Abschluß: Abitur  |
| 08/98-09/99     | Zivildienst in der Bahnhofsmision Köln   |
| 10/99-12/04     | Studium der Biologie an der RWTH Aachen<br>Schwerpunkt: Molekularbiologie<br>Abschluß: Diplom Biologe  |
| 01/2004-12/2004 | Diplomarbeit am Institut für Biologie I<br>(Molekulare Genetik) der RWTH-Aachen<br>Titel: Die physiologische Funktion der mitochondrialen<br>Glycolatdehydrogenase in <i>Arabidopsis thaliana</i> : Analyse<br>überexprimierender und mutierter Linien |
| 02/2005-02/2008 | Wissenschaftlicher Mitarbeiter am Max Planck Institut<br>für molekulare Genetik (Berlin), Abteilung Prof. Lehrach,<br>AG Himmelbauer   |
| 03/2008-06/2008 | Weiterbildung zum Klinischen Monitor/ Datenbankmanager,<br>PAREXEL Akademie (Berlin)   |
| Seit 07/2008    | Fachplaner Labortechnik,<br>dr. heinekamp Labor- und Institutsplanung Berlin GmbH  |

## 8 Publikationen

Rosenkranz Ruben, Borodina Tatiana, Lehrach Hans, Himmelbauer Heinz:  
Characterizing the mouse ES cell transcriptome with Illumina sequencing.  
*Genomics* 2008,92: 187-194

Niessen M, Thiruvedhi K, Rosenkranz R, Kebeish R, Hirsch H, Kreuzaler F, Peterhänsel C:  
Mitochondrial glycolate oxidation contributes to photorespiration in higher plants.  
*J Exp Bot* 2007, 58:2709-2715.

Kebeish R, Niessen M, Thiruvedhi K, Bari R, Hirsch H, Rosenkranz R, Stäbler N, Schönfeld B, Kreuzaler F, Peterhänsel C:  
Chloroplastic photorespiratory bypass increases photosynthesis and biomass production in *Arabidopsis thaliana*.  
*Nat Biotechnol* 2007, 25:593-5

## **Erklärung zur Dissertation**

gemäß §6(1) der Promotionsordnung der Naturwissenschaftlichen Fakultät der Gottfried Wilhelm Leibniz Universität Hannover für die Promotion zum Dr. rer. nat.

Hierdurch erkläre ich, dass ich meine Dissertation mit dem Titel "Entwicklung neuartiger Screening-Verfahren zur Mutationsdetektion in embryonalen Stammzellen der Maus" selbständig verfasst und die benutzten Hilfsmittel und Quellen sowie gegebenenfalls die zu Hilfeleistungen herangezogenen Institutionen vollständig angegeben habe.

Die Dissertation wurde nicht schon als Masterarbeit, Diplomarbeit oder andere Prüfungsarbeit verwendet.

Ruben Rosenkranz

## Ein herzlicher Dank

gebührt all denen, die zum Gelingen dieser Arbeit beigetragen haben.

Herrn Dr. Heinz Himmelbauer und Herrn Prof. Dr. Hans Lehrach danke ich für die Überlassung des Themas sowie für die Bereitstellung hervorragender Arbeitsbedingungen am Max-Planck-Institut für molekulare Genetik in Berlin.

Besonders danke ich Herrn Prof. Dr. Christoph Peterhänsel für die Unterstützung und Übernahme der Betreuung dieser Arbeit an der Gottfried Wilhelm Leibniz Universität Hannover. Auch möchte ich Prof. Dr. Helge Küster für die freundliche Bereitschaft zur Übernahme des zweiten Gutachtens danken.

Meinen Freunden und langjährigen Wegbegleitern im Labor Juliane Dohm, Ulrike Dohrmann, Cornelia Lange, Florian Mertes, Tobias Nolden, Robert Querfurth und Hans-Jörg Warnatz danke ich für ihre stete Hilfsbereitschaft, die zahlreichen Anregungen und Diskussionen und für die vielen schönen Stunden im und ausserhalb des Instituts.

Ebenfalls danke ich Steve Herrmann, Marion Klein, Ines Müller und Stefanie Palczewski für ihre tatkräftige, praktische Unterstützung.

Coco danke ich für ein wunderbares und erfülltes Leben.