

Aus der Medizinischen Klinik und Poliklinik V  
Klinik der Universität München  
Direktor: Prof. Dr. Jürgen Behr  
und  
aus dem Comprehensive Pneumology Center des  
Helmholtz-Zentrums München  
Direktor: Dr. Ali Önder Yildirim

***Identifizierung und Validierung von Referenzgenen  
zur relativen Quantifizierung der mRNA-Expression in neonatalem murinem  
Lungengewebe mittels RT-qPCR***

Dissertation  
zum Erwerb des Doktorgrades der Medizin  
an der Medizinischen Fakultät der  
Ludwig-Maximilians-Universität zu München

vorgelegt von

Hyunkyu Shin

aus

Busan, Südkorea

2023

Mit Genehmigung der Medizinischen Fakultät  
der Universität München

Berichterstatter: PD Dr. Anne Hilgendorff

Mitberichterstatter: Prof. Dr. Alexander Faußner  
PD Dr. Claudia Staab-Weijnitz

Mitbetreuung durch den  
promovierten Mitarbeiter: Dr. Stefan Dehmel

Dekan: Prof. Dr. med. Thomas Gudermann

Tag der mündlichen Prüfung: 30.03.2023

# Inhaltsverzeichnis

<b>Zusammenfassung</b> .....	<b>6</b>
<b>Abstract</b> .....	<b>7</b>
<b>Abbildungsverzeichnis</b> .....	<b>9</b>
<b>Tabellenverzeichnis</b> .....	<b>10</b>
<b>Abkürzungsverzeichnis</b> .....	<b>11</b>
<b>1 Einleitung</b> .....	<b>12</b>
<b>1.1 Genexpressionsforschung</b> .....	<b>12</b>
1.1.1 RT-qPCR für Genexpressionsanalyse.....	12
1.1.2 Relative Quantifizierung mittels eines Referenzgens bei der RT-qPCR .....	12
1.1.3 Anforderungen an zuverlässige Referenzgene.....	13
1.1.4 Einschränkung der konventionell verwendeten Referenzgene.....	13
<b>1.2 Neonatales Mausmodell der humanen Lungenerkrankung</b> .....	<b>14</b>
<b>1.3 Fragestellung</b> .....	<b>14</b>
<b>1.4 Zielsetzung</b> .....	<b>16</b>
<b>2 Prinzipien</b> .....	<b>17</b>
<b>2.1 Quantitative Echtzeit-PCR mit reverser Transkription</b> .....	<b>17</b>
2.1.1 Die klassische PCR .....	17
2.1.2 Vorteil der RT-qPCR gegenüber der klassischen PCR.....	19
2.1.3 Amplifikationsplot der RT-qPCR .....	20
2.1.4 Fluoreszenzfarbstoff für RT-qPCR.....	22
2.1.5 Bedeutung der RNA-Qualität in der RT-qPCR.....	26
2.1.5.1 RNA-Reinheit und -Integrität .....	26
2.1.5.2 Einfluss der RNA-Qualität auf die RT-qPCR.....	33
<b>2.2 Genevestigator</b> .....	<b>35</b>
<b>2.3 GeNorm und Normfinder</b> .....	<b>38</b>

<b>3</b>	<b>Methoden und Materialien</b>	<b>40</b>
<b>3.1</b>	<b>Vorauswahl potenzieller Referenzgenkandidaten</b>	<b>40</b>
3.1.1	Datenlage hinsichtlich der Referenzgene für die Mauslunge	40
3.1.2	Genevestigator	40
3.1.3	Kommerzielles Referenzgenpanel (Roche)	42
<b>3.2</b>	<b>Microarray-Transkriptom-Daten</b>	<b>42</b>
<b>3.3</b>	<b>Versuchstier und RNA-Extraktion aus den Lungengeweben</b>	<b>43</b>
<b>3.4</b>	<b>RNA-Qualitätskontrolle</b>	<b>44</b>
<b>3.5</b>	<b>Reverse Transkription</b>	<b>47</b>
<b>3.6</b>	<b>Quantitative Echtzeit-PCR</b>	<b>49</b>
<b>3.7</b>	<b>Interrun-Kalibrierung</b>	<b>53</b>
<b>3.8</b>	<b>GeNorm und Normfinder</b>	<b>55</b>
<b>3.9</b>	<b>Einzelzell-RNA-Sequenzierung</b>	<b>58</b>
<b>4</b>	<b>Ergebnisse</b>	<b>60</b>
<b>4.1</b>	<b>Vorauswahl der Referenzgenkandidaten</b>	<b>60</b>
4.1.1	Datenlage hinsichtlich der Referenzgene für die Mauslunge	60
4.1.2	Referenzgenvorschläge von Genevestigator	63
4.1.3	Roche kommerzielles Referenzgenpanel (RealTime Ready Assays)	64
4.1.4	Ausschluss einer Reihe von Kandidatengen	65
4.1.5	Erste Evaluierung in verfügbaren Microarray-Daten	66
4.1.6	Finale Referenzgenkandidaten	67
<b>4.2</b>	<b>RT-qPCR-Validierung für zuverlässige Referenzgene</b>	<b>69</b>
4.2.1	Zuverlässige Referenzgene unter physiologischen Bedingungen	69
4.2.2	Einfluss der neonatalen Lungenentwicklung auf die Expression von Referenzgenkandidaten	71
4.2.3	Einfluss des Geschlechts auf die Expression von Referenzgenkandidaten	73
4.2.4	Einfluss der RNA-Qualität auf die Expression von Referenzgenkandidaten	75
4.2.5	Einfluss der Hyperoxie-Exposition auf die Expression von Referenzgenkandidaten	77
<b>4.3</b>	<b>Einzelzell-RNA-Sequenzierung für zuverlässige Referenzgene</b>	<b>80</b>
<b>5</b>	<b>Diskussion</b>	<b>83</b>
	<b>Anhang</b>	<b>91</b>

<b>Literaturverzeichnis.....</b>	<b>97</b>
<b>Publikation.....</b>	<b>104</b>
<b>Mitwirkung.....</b>	<b>104</b>
<b>Danksagung .....</b>	<b>105</b>
<b>Affidavit.....</b>	<b>106</b>

## Zusammenfassung

Die quantitative Echtzeit-PCR (*real time quantitative polymerase chain reaction*, abgekürzt qPCR) ist eine moderne Technik für die Genexpressionsanalyse und ist in diversen Forschungsbereichen etabliert. Sie weist zahlreiche markante Vorteile auf, wie hohe Sensitivität und Spezifität, zuverlässige Reproduzierbarkeit sowie großer dynamischer Bereich mit der Fähigkeit, auch geringste Mengen an Genaktivität in Form von mRNA-Expression zuverlässig nachzuweisen.

Für die Interpretation der gewonnenen Rohdaten wird die sog. relative Quantifizierung bevorzugt eingesetzt, weil mit dieser leicht durchführbaren Methode die experimentell bedingten, unspezifischen Variationen in der Genexpression effektiv behoben werden können und dadurch nur echte Variationen der Genexpression, die das zu untersuchende, biologische Phänomen erklärt, detektiert werden können. Bei der relativen Quantifizierung wird die Expression der Zielgene zu der der Referenzgene "relativ" quantifiziert. Um zuverlässig als Referenzgene zu dienen, muss eine vernachlässigbar niedrige Expressionsvariabilität der Referenzgene vorausgesetzt werden, und zwar idealerweise unabhängig von biologischen, technischen und experimentellen Bedingungen.

Bei humanen Lungenerkrankungen werden oft Signalwege der Embryonalentwicklung reaktiviert. Daher bietet sich ein neonatales Mausmodell zur Untersuchung dieser Signalwege in murinem Lungengewebe an. Dabei ist die Auswahl geeigneter Referenzgene besonders herausfordernd, weil während des Reifungsprozesses der Lunge komplexe transkriptionelle und epigenetische Regulation diverser Gene stattfindet. Trotzdem liegen bisher nur begrenzt Daten über zuverlässige Referenzgene für Mauslungen während der neonatalen Entwicklungsphase und bei experimentell induzierter Lungenschädigung (z.B. durch Hyperoxie-Exposition) vor.

Im vorliegenden Projekt haben wir konventionell verwendete Referenzgene sowie neue Referenzgenkandidaten hinsichtlich ihrer Expressionsvariabilität in neonatalen Lungen von C57BL/6-Mäuse evaluiert, wobei biologische, technische und experimentelle Effekte berücksichtigt wurden. Im ersten Schritt wurden dafür Referenzgenkandidaten aus Fachartikeln zu Mauslungenstudien bzw. zu Evaluierungsstudien, die die Zuverlässigkeit von Referenzgenen untersucht haben, zusammengetragen. Außerdem wurden mittels der Analysesoftware Genevestigator bereits publizierte Microarray-Daten auf mögliche Referenzgenkandidaten hin analysiert. Daraus ergaben sich insgesamt 122 Referenzgenkandidaten. Davon wurde anschließend eine Reihe von Genen ausgeschlossen, die einerseits unsere vordefinierten Kriterien nicht erfüllten oder andererseits eine hohe Expressionsvariabilität in unseren eigenen Microarray-Daten bzw. in Genevestigator aufwiesen. Aus dieser Vorauswahl ergaben sich 24 Referenzgenkandidaten, die in RT-qPCR-Experimenten in Lungenproben von C57BL/6-Mäusen validiert wurden. Dabei wurden biologische (Entwicklungsstadium, Geschlecht), technische

(RNA-Qualität) sowie experimentelle (Hyperoxie-Exposition) Variablen berücksichtigt. Die gewonnenen Expressionsdaten, repräsentiert als unbearbeitete Cq-Werte, wurden mit Hilfe von zwei gängigen Analysealgorithmen (GeNorm und Normfinder) zur Identifizierung zuverlässiger Referenzgene analysiert. Die Expression der Referenzgenkandidaten in diversen pulmonalen Zelltypen wurde in Zusammenarbeit mit unserem Projektpartner Prof. Sucre (Vanderbilt University Medical Center, Nashville, USA) mittels Einzelzell-RNA-Sequenzierung bestimmt.

Kein einziger Referenzgenkandidat zeigte stabile Expression bei allen untersuchten Experimentalbedingungen, so dass kein einzelnes Gen als ubiquitär verwendbares und damit ideales Referenzgen eingesetzt werden könnte. Vielmehr erwiesen sich unterschiedliche Referenzgenkandidaten abhängig von den untersuchten biologischen, technischen und experimentellen Bedingungen und vom jeweiligen Algorithmus als am stabilsten exprimiert (siehe Abb. 5-1 Zusammenfassung der zuverlässigen Referenzgene in der Diskussion).

Die im Rahmen des vorliegenden Projektes identifizierten, stabil exprimierten Referenzgene werden zur akkuraten Interpretation zukünftiger RT-qPCR-Experimente in neonatalen Mäusen beitragen.

## Abstract

RT-qPCR is the established, state-of-the-art method for gene expression analysis in a variety of research areas. The technology has distinct advantages such as high sensitivity and specificity, reliable reproducibility and high dynamic range so as to detect transcripts at low copy numbers.

A so-called relative quantification is preferably applied for the interpretation of raw RT-qPCR data because the experiment-derived, unspecific variations of gene expression can be effectively removed through this easily applicable analysis method and the true variations of gene expression underlying the biologic phenomenon under investigation can be fully revealed. In the relative quantification, the expression of the gene of interest is "relatively" quantified to that of so-called reference genes. For a reliable performance, the reference genes need to possess negligible expression variability which should not be influenced by biological, technical and experimental conditions.

In human pulmonary diseases, the signal pathways of embryonic development are often activated and thus, the neonatal murine model provides various possibilities for investigating these signal pathways underlying human diseases. In the developing lungs of mice, the selection of appropriate reference genes is particularly challenging since complex transcriptional and epigenetic regulations of diverse genes occur. Despite the plethora of studies with neonatal lungs, there is only limited data available addressing reference gene expression for the developing lungs with experimentally induced lung injury, e.g., by hyperoxia exposure.

In the present study, commonly used and novel software-predicted reference gene candidates were evaluated for their expression variability in the lung of neonatal C57BL/6 mice while considering the effect of biologic, technical and experimental variables. In a first step, published studies for commonly used reference genes or evaluation studies for reliable reference genes were reviewed. In addition, publicly available microarray datasets were screened using the Genevestigator software, resulting in an initial selection of 122 candidates. A series of candidate genes was then excluded from further processing, if they were not matching pre-defined criteria or showed high expression variability in own microarray data derived from neonatal lung and data from Genevestigator. The 24 final reference gene candidates from this pre-selection step were evaluated by RT-qPCR in lung tissue samples of C57BL/6 mice under consideration of critical biological variables (developmental stage, sex) and technical conditions (RNA quality) as well as an experimental challenge (hyperoxia exposure). The obtained expression data were analyzed using GeNorm and Normfinder algorithms to identify reliable reference genes from the candidate gene set. In cooperation with Dr. Sucre (Vanderbilt University Medical Center, Nashville, USA), the cell-specific expression of the final reference gene candidates was analyzed using single-cell RNA-sequencing in the developing mouse lung.

None of the reference gene candidates showed stable gene expression throughout all investigated experimental conditions so that there was no ubiquitous ideal reference gene from our candidate genes, but rather, several different reference gene candidates were shown to be reliable reference genes depending on the investigated, biological, technical and experimental conditions and each analysis algorithm (see **Abb. 5-1** summarized reliable reference genes in the discussion section).

The identified reliable reference genes will provide valuable resource for accurate interpretation of future RT-qPCR experiments investigating neonatal mouse lung tissue.

# Abbildungsverzeichnis

Abb. 1-1 Varianz der Genexpression relativ zu Expressionslevel .....	15
Abb. 2-1 Der erste Zyklus der klassischen PCR .....	18
Abb. 2-2 Erste 3 Zyklen der konventionellen PCR.....	19
Abb. 2-3 Amplifikationsplot der RT-qPCR.....	20
Abb. 2-4 Funktionsweise von <i>SYBR Green I</i> .....	23
Abb. 2-5 Schmelzkurvenanalyse und Schmelzpeak.....	24
Abb. 2-6 FRET-Mechanismus.....	25
Abb. 2-7 Funktionsweise von TaqMan-Assay .....	26
Abb. 2-8 Binnenstruktur eines Analysechips und Analyseablauf.....	30
Abb. 2-9 Bioanalyser-Ergebnis .....	31
Abb. 2-10 Das Elektropherogramm .....	32
Abb. 2-11 Einfluss der RNA-Qualität auf Cq-Werte .....	34
Abb. 2-12 Benutzerschnittstelle von Genevestigator.....	36
Abb. 2-13 RefGene-Analysenergebnis .....	37
Abb. 3-1 Suchergebnis von Genevestigator .....	42
Abb. 3-2 Agilent RNA 6000 Nano Chip .....	45
Abb. 3-3 Chip-Priming-Station.....	46
Abb. 3-4 Importierung der Cq-Rohdaten in GenEx™.....	55
Abb. 3-5 GeNorm-Ergebnis.....	56
Abb. 3-6 Normfinder-Ergebnis .....	57
Abb. 4-1 Expression der Referenzgenkandidaten in drei Hauptzelltypen bei neonataler Lungenentwicklung von C57BL/6-Mäusen.....	81
Abb. 4-2 Expressionsprofil der Referenzgenkandidaten in pulmonalen Zelltypen.....	82
Abb. 5-1 Zuverlässige Referenzgene in der RT-qPCR-Validierung bezüglich untersuchten Experimentalbedingungen .....	85
Abb. 5-2 Eignung der 16 als zuverlässige Referenzgene identifizierten Kandidaten hinsichtlich der Analyseansätze/experimentellen Bedingungen.. .....	89

## Tabellenverzeichnis

Tab. 3-1 Anzahl der verfügbaren RNA-Proben .....	44
Tab. 3-2 Materialien mit Herstellerangaben.....	47
Tab. 3-3 Template-Primer-Mix .....	47
Tab. 3-4 RT-Mastermix.....	48
Tab. 3-5 Materialien mit Herstellerangaben.....	49
Tab. 3-6 Herstellung von PCR-Mastermix.....	49
Tab. 3-7 Programm für die quantitative Echtzeit-PCR.....	50
Tab. 3-8 Layout der RT-qPCR-Platten.....	52
Tab. 3-9 Berechnung des Kalibrierungsfaktors für die Interrun-Kalibrierung .....	54
Tab. 4-1 Konventionell verwendete Referenzgene für Lungengewebe der C57BL/6.....	60
Tab. 4-2 Referenzgenkandidaten aus Evaluierungsstudien.....	62
Tab. 4-3 Die von Genevestigator erfassten Studien mit Genexpressionsanalyse .....	63
Tab. 4-4 Genevestigator-Ergebnis .....	64
Tab. 4-5 Die 27 Gene aus dem Roche-Referenzgenpanel.....	65
Tab. 4-6 Die 46 potenziellen Referenzgenkandidaten mit Expressionsdaten aus unserer eigenen Transkriptom-Datenbank sowie Genevestigator.....	68
Tab. 4-7 GeNorm- und Normfinder-Ergebnis für physiologische Bedingung.....	71
Tab. 4-8 GeNorm-Ergebnis während der Lungenentwicklung.....	72
Tab. 4-9 Normfinder-Ergebnis während der Lungenentwicklung.....	73
Tab. 4-10 GeNorm-Ergebnis für Einfluss des Geschlechts.....	74
Tab. 4-11 Normfinder-Ergebnis für Einfluss des Geschlechts .....	75
Tab. 4-12 GeNorm-Ergebnis für Einfluss der RNA-Qualität.....	76
Tab. 4-13 Normfinder-Ergebnis für Einfluss der RNA-Qualität .....	77
Tab. 4-14 GeNorm-Ergebnis für Einfluss der Hyperoxie-Exposition.....	79
Tab. 4-15 Normfinder-Ergebnis für Einfluss der Hyperoxie-Exposition.....	80
Tab. 5-1 Intra- und Intergruppenvariationsanalyse von Normfinder .....	87

## Abkürzungsverzeichnis

ALI	<i>acute lung injury</i>
bp	Basenpaar
CF	<i>cystic fibrosis</i>
dNTP	Desoxyribonukleosidtriphosphate
ds	doppelsträngig
Exp.	Expression
GN	GeNorm
IRC	Interrun-Kalibrator
NF	Normfinder
kb	Kilobase
KO	Knockout
konz.	konzentriert
LPS	Lipopolysacchariden
nt	Nukleotide
PBS	<i>phosphate-buffered saline</i>
PCR	Polymerase-Kettenreaktion
PND	<i>Post natal day</i>
qPCR	quantitative Echtzeit-PCR
RG	Referenzgen(e)
RT	Reverse Transkription
RT-qPCR	Quantitative Echtzeit-PCR mit reverser Transkription
RTase(n)	Reverse Transkriptase(n)
scRNA-seq	<i>Single cell RNA sequencing</i>
SD	Standardabweichung
ss	einzelsträngig
Tg	transgen
VILI	<i>ventilator-induced lung injury</i>
wt	Wildtyp

# 1 Einleitung

## 1.1 Genexpressionsforschung

Seit der vollständigen Entschlüsselung des Humangenoms im Jahr 2003 haben biologisch-medizinische Forschungen auf Genebene erheblich an Bedeutung gewonnen. Dank diesem Erfolg beim humanen Genom-Projekt ist man nun in der Lage, spezifische Gene zu identifizieren, die für Suszeptibilität gegenüber bestimmten Erkrankungen sowie deren Pathogenese verantwortlich sind.<sup>1</sup> Folgerichtig kann die Genexpressionsforschung einen ausschlaggebenden Hinweis geben, ob ein Individuum einer bestimmten Erkrankung (vor allem Tumor- und chronischen Erkrankungen) gegenüber prädisponiert ist und ferner wie diese Erkrankung auf Genebene gezielt therapiert werden kann.<sup>2,3</sup> Außerdem bringt die Genexpressionsanalyse im Tiermodell der humanen Erkrankungen wertvolle Informationen über deren Pathogenese und somit auch mögliche Angriffspunkte zur Behandlung.

### 1.1.1 RT-qPCR für Genexpressionsanalyse

Eine der heutzutage am häufigsten eingesetzten Methoden bei der Genexpressionsanalyse ist die quantitative Echtzeit-qPCR (*real time quantitative polymerase chain reaction*, abgekürzt qPCR) in Kombination mit einer vorgeschalteten reversen Transkription (RT-qPCR). Diese Technik weist zahlreiche Vorteile wie hohe Sensitivität und Spezifität, verlässliche Reproduzierbarkeit und einen großen dynamischen Bereich auf. Sie kommt bevorzugt zum Einsatz, wenn das Expressionsprofil einer begrenzten Anzahl von Genen in zahlreichen Gewebe- oder Zellproben untersucht werden sollen und dazu nähere Information über die Target-Gene z.B. genaue Sequenz bereits vorliegt. Da mittels der RT-qPCR die Expression der Target-Gene (mRNA) präzise und rapide detektiert werden und diese Detektion in zahlreichen Proben gleichzeitig erfolgen kann, stellt die RT-qPCR auch eine geeignete Methode in der Molekulardiagnostik (Nachweis der krankheitserregenden Mikroorganismen) dar.

### 1.1.2 Relative Quantifizierung mittels eines Referenzgens bei der RT-qPCR

Variationen in der Genexpression, die von RT-qPCR-Experimenten festgestellt wurden, können prinzipiell zwei Ursachen haben. Die eine Variation der Genexpression erklärt die vorliegenden Phänotypen oder das zu untersuchende, biologische Phänomen, gilt somit als echte Variation der Genexpression. Die andere, unspezifische Variation der Genexpression wird durch unterschiedliche Gegebenheiten der Experimente selbst verursacht z.B. unterschiedliche Startmenge des Untersuchungsmaterials (unvermeidbare Pipettierungenauigkeiten), heterogene

Ausbeute der RNA-Extraktion und niedrige RNA-Qualität, inkonsistente Effizienzen bei der RT und qPCR usw. Somit kann diese Art der unspezifischen Variation in der Genexpression in jedem Zwischenschritt der RT-qPCR entstehen.<sup>4,5</sup> Eine bewährte Maßnahme zur Eindämmung dieser unspezifischen Variation ist Einsatz von Referenzgenen als interne Kontrolle.<sup>6,7</sup> Das Expressionslevel der Referenzgene wird zusammen mit Target-Genen im gleichen RT-qPCR-Lauf bestimmt und somit unterliegen sowohl die Referenzgene, als auch die Target-Gene derselben Experimentalbedingung. Wenn nun das Expressionslevel von Target-Genen zu dem der eingesetzten Referenzgene relativ quantifiziert wird, wird die unspezifische Variation der Genexpression, die durch Experimente selbst zustande kommt, effektiv aufgehoben.<sup>8</sup>

### 1.1.3 Anforderungen an zuverlässige Referenzgene

Um ihrer entscheidenden Funktion als interne Kontrolle bei der relativen Quantifizierung nachzukommen, sollten die Referenzgene einige Voraussetzungen erfüllen. Die Wichtigste ist ihre „stabile“ Genexpression, idealerweise unabhängig von Organen, Experimentalbedingungen und -behandlungen. Üblicherweise wird die Expressionsstabilität eines Gens anhand der Standardabweichung eines Parameters abgeschätzt, der das Expressionslevel der untersuchten Gene repräsentiert. Dieser Parameter für das Genexpressionslevel ist je nach eingesetzter Methode unterschiedlich definiert. Beispielsweise entspricht er bei Microarray-Experimenten der Signalintensität auf dem Analysechip und bei RT-qPCR dem sog. Cq-Wert (siehe dazu im Detail den Abschnitt 2.1.3). Im Endeffekt gilt folgende Aussage: Je kleiner die Standardabweichung des Parameters für das Genexpressionslevel in den untersuchten Gewebeproben ist, desto stabiler exprimiert ist das betreffende Gen.

### 1.1.4 Einschränkung der konventionell verwendeten Referenzgene

Früher wurden sog. Housekeeping-Gene als Referenzgene bei RT-qPCR-Experimenten eingesetzt, weil man damals von stabiler Expression dieser Housekeeping-Gene ausging. Diese Annahme beruhte darauf, dass die Housekeeping-Gene für Aufrechterhaltung der essentiellen Grundfunktion der Zellen verantwortlich sind und somit in einem Organismus ubiquitär konstant exprimiert sein sollten.<sup>9,10</sup> Tatsächlich waren laut Suzuki et al. Housekeeping-Gene wie *Gapdh*, *Actb*, *18S*- und *28S-rRNA* die am häufigsten eingesetzten Referenzgene (90% der bis 2000 publizierten Studien mit Genexpressionsanalyse in Säugetierzellen).<sup>3,11</sup>

Aber zahlreiche nachfolgende Studien konnten belegen, dass diese Housekeeping-Gene von verschiedenen experimentellen Behandlungen und unterschiedlichen Organen sowie Zellarten stark beeinflusst werden und somit deren Expressionsstabilität nicht als gegeben vorausgesetzt werden kann.<sup>6,11–25</sup> Daher wird heutzutage generell empfohlen, Referenzgene für den jeweiligen

Einsatzzweck zu evaluieren und entsprechend den vorliegenden experimentellen Bedingungen sorgfältig auszuwählen.<sup>3,15,26-28</sup>

## 1.2 Neonatales Mausmodell der humanen Lungenerkrankung

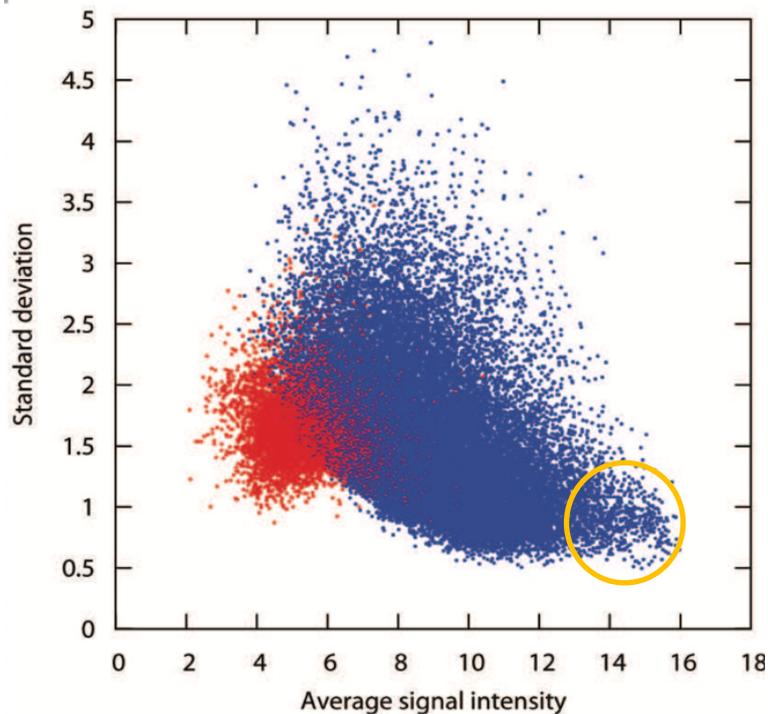
Die Maus ist die am häufigsten verwendete Versuchstierspezies zur Erforschung der Pathogenese humaner Erkrankungen, weil sie die folgenden Vorteile aufweist. Mäuse sind physiologisch und genetisch sehr ähnlich dem Menschen. So zeigen die beiden Spezies eine 99%ige Homologie ihrer Genome.<sup>29,30</sup> Aufgrund dessen kann jede Art der Genalteration, die einer humanen Erkrankung zugrunde liegt, durch hochentwickelte gentechnische Methoden zur Untersuchung in Mäuse eingeführt werden. Außerdem haben Mäuse einen beschleunigten Lebenszyklus im Vergleich zu Menschen. Aus diesem Grund kann man in einem Mausmodell der humanen Erkrankung in absehbarer Zeit Ergebnisse erzielen.

Mauslungen stellen ein geeignetes Gewebematerial zur Erforschung pulmonaler Erkrankungen dar, zumal in Mauslungen Tiermodelle der Lungenschädigung z.B. durch Hyperoxie-Exposition und gegebenenfalls mit gentechnischer Manipulation nachgebildet werden können. Darüber hinaus bewähren sich Mauslungen während der neonatalen Organentwicklung zur Untersuchung der neonatalen Pathogenese bzw. Prädisposition gegenüber verschiedener Lungenerkrankungen. Die neonatale Lungenentwicklung ist ein komplexer Vorgang, der mehrere Phasen (embryonale, pseudoglanduläre, kanalikuläre, sakkuläre und alveoläre) beinhaltet. Dabei spielt transkriptionelle und epigenetische Regulation diverser Gene für die Induzierung der phasengerechten morphologischen Ausbildung des Organs eine entscheidende Rolle. Außerdem obwohl bei der Lunge kein ständiger Zellaustausch (*cell turnover*) wie Darm oder Blut stattfindet, ist dieses Organ fähig, robust auf jegliche Organschädigung zu reagieren und dabei kommt der Expression bzw. Regulation diverser weiterer Gene in den Stamm-/Progenitorzellen große Bedeutung zu.<sup>31</sup>

## 1.3 Fragestellung

Hruz et al.<sup>32</sup> haben im Jahr 2011 eine große Anzahl von bis dahin publizierten Microarray-Daten von Menschen sowie Mäusen analysiert und konnten daraus unabhängig von untersuchten Spezies folgende Schlussfolgerungen ziehen: i) Es würden keine idealen Gene existieren, die in einem Organismus unabhängig von unterschiedlichen Organen und experimentellen Behandlungen ubiquitär stabil exprimiert sind (**Abb. 1-1**). Demzufolge scheint die früher angenommene Grundvoraussetzung für zuverlässige Referenzgene, nämlich universell konstante Genexpression, nicht erfüllbar zu sein. Daher sollten geeignete Referenzgene spezifisch für das

durchzuführende Experiment überprüft und die besten Referenzgene von den zur Verfügung stehenden eingesetzt werden. ii) Für jeden biologisch-experimentellen Kontext gäbe es Gene, deren Genexpressionsvarianz viel niedriger als die der konventionell verwendeten Referenzgene ist und die sich somit spezifisch für diesen biologisch-experimentellen Kontext besser eignen würden.



**Abb. 1-1 Varianz der Genexpression relativ zu Expressionslevel.** In der Grafik sind die SD-Werte der Genexpression aus 2.974 Maus-Microarrays gegen Expressionslevels (Mittelwert der Signalintensität) aufgetragen. Die blaue Wolke zeigt Probesets mit Present Calls in mindestens 10% der Microarrays und die rote Wolke diejenigen mit Absent Calls in mehr als 90% der Arrays. Zum einen zeigten die konventionell verwendeten Referenzgene (gelb eingekreiste Zone) zwar relativ niedrige SDs und generell hohe Expressionslevel, ca. 20% der Gene im Genom weisen aber genauso niedrige SDs auf (die blaue Wolke mit SDs <1 und Signalintensität <13). Zum anderen ist hieran zu erkennen, dass keine Gene existieren, deren SDs eindeutig niedriger als die der blauen Probenwolke liegen, damit sie als ideale Referenzgene verwendet werden könnten. Damit ist es eher unwahrscheinlich, dass die konventionell verwendeten Referenzgene schon die besten Referenzgene darstellen. *Grafik aus Hruz et al.*<sup>32</sup>

Bei der Untersuchung der aktuellen Datenlage wurde festgestellt, dass zuverlässige Referenzgene für neonatale Mauslungen bisher unzureichend evaluiert wurden. Insbesondere fehlen Studien, bei denen biologische, technische und experimentelle Faktoren bei der Auswahl der zuverlässigen Referenzgene für die neonatalen Mauslungen berücksichtigt wurden. Basierend auf der oben vorgestellten Erkenntnis von Hruz et al. wurde angenommen, dass neue zuverlässige Referenzgene spezifisch für diesen biologischen Kontext in genomweiter Suche nachzuweisen sind.

## 1.4 Zielsetzung

Im Rahmen des vorliegenden Projektes sollte die Expressionsstabilität von potenziellen Referenzgenen im Lungengewebe neonataler C57BL/6-Mäuse hinsichtlich biologischer (Entwicklungsstadien, Geschlecht), technischer (RNA-Qualität) und experimenteller (Hyperoxie-Exposition) Faktoren sowie für diverse pulmonale Zelltypen evaluiert werden.

## 2 Prinzipien

### 2.1 Quantitative Echtzeit-PCR mit reverser Transkription

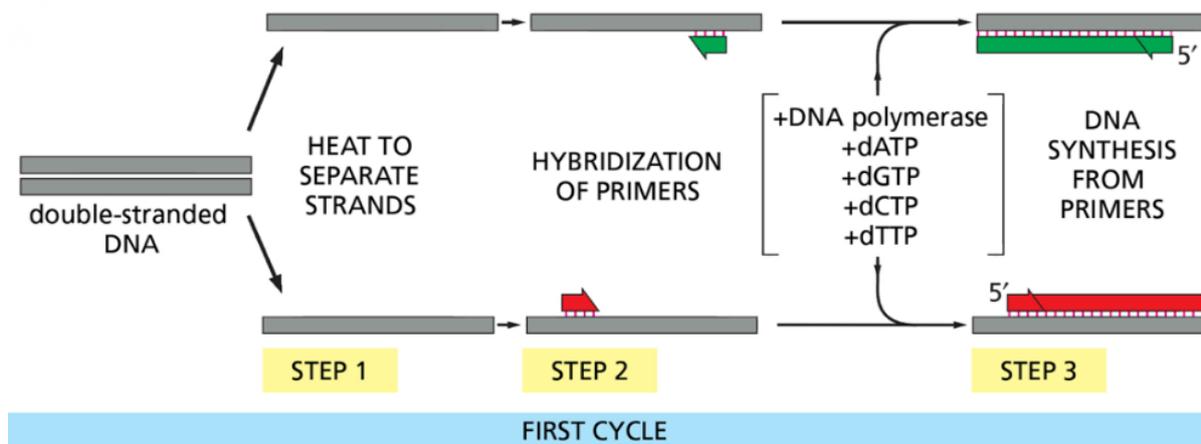
Die quantitative Echtzeit-PCR (*real time quantitative polymerase chain reaction*, abgekürzt qPCR) stellt eine Weiterentwicklung der klassischen Polymerase-Kettenreaktion (PCR) dar, die eine in kleiner Menge vorliegende Target-Sequenz aus DNA durch deren Vervielfältigung nachweist.

Heutzutage steht jedoch mRNA im Mittelpunkt der Genexpressionsanalyse, zumal sie als das Produkt der aktiven Gene (Transkript) die zu synthetisierenden Proteine näherungsweise repräsentiert. Die mRNA ist allerdings aufgrund von ubiquitär vorhandenen RNasen anfällig für Degradierung. Um eine degradierungsfreie Handhabung der mRNA in der Genexpressionsanalyse zu gewährleisten, wird die mRNA mit Hilfe der reversen Transkription in die komplementäre DNA (cDNA) umgeschrieben, welche im Vergleich zur mRNA widerstandsfähiger gegen Degradierung ist und dennoch alle Informationen über die protein-kodierende Sequenz der mRNA behält. Dazu kann die in der qPCR bevorzugt eingesetzte Taq-Polymerase nur DNA-Templates vervielfältigen, was im qPCR-Experiment eine vorgeschaltete reverse Transkription (RT) erforderlich macht, um aus den vorliegenden mRNA-Strängen cDNA-Versionen zu generieren. Somit ergibt sich auch die Abkürzung RT-qPCR für dieses Verfahren.<sup>33</sup>

Die RT-qPCR zeichnet sich durch mehrere Vorteile aus: Hohe Sensitivität und Spezifität, verlässliche Reproduzierbarkeit und großer dynamischer Bereich. Manchmal bleibt diese Methode aufgrund ihrer Besonderheit, kleinste Menge an Nukleinsäure verlässlich nachzuweisen, als einzig anwendbare Methode in der klinischen Diagnostik übrig.<sup>34</sup> Dank diesen zahlreichen Vorteilen kommt die RT-qPCR divers in Grundlagenforschung, medizinischer Diagnostik und Biotechnologie zum Einsatz.<sup>35</sup> Da die Funktionsweise der konventionellen PCR für das Verständnis des RT-qPCR-Prinzips grundlegend ist, wird zunächst auf die klassische PCR eingegangen.

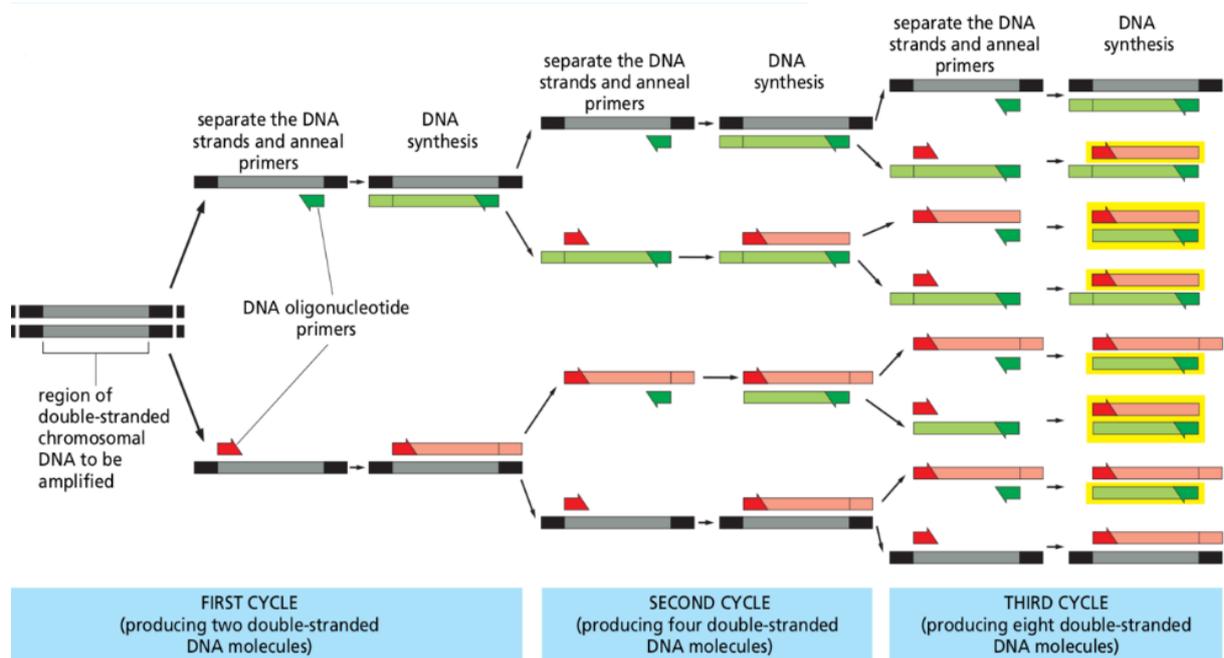
#### 2.1.1 Die klassische PCR

Bei der klassischen PCR wird eine bestimmte DNA-Sequenz milliardenfach vervielfältigt, indem dieses Genmaterial mit erforderlichen Reagenzien wie hitzeresistente DNA-Polymerase, Desoxyribonukleosidtriphosphate (dNTPs), Primer, Pufferlösung einer Reihe von Erhitzungs- und Abkühlungszyklen unterzogen wird. Ein Zyklus dieser thermischen Veränderung besteht aus drei Schritten: Denaturierung (*melting*), Primerhybridisierung (*annealing*) und Elongation (*extension*).



**Abb. 2-1 Der erste Zyklus der klassischen PCR.** Die drei Reaktionsschritte des 1. PCR-Zyklus, d.h. Denaturierung, Primerhybridisierung sowie Elongation sind schematisch dargestellt. Zwei Primer (Sense- und Antisense-Primer) für die jeweiligen Einzelstränge und die vier Bausteine der DNA (Desoxyribonukleotide) werden als Reaktionssubstrat zugegeben. Die Elongation erfolgt durch das Enzym DNA-Polymerase. Danach werden die erhaltenen Doppelstränge erneut durch Hitzeeinwirkung aufgeschmolzen und der Prozess wiederholt sich. Nach jedem Zyklus verdoppelt sich theoretisch die DNA-Menge. *Abbildung aus Alberts et al. S.473.* <sup>36</sup>

Im ersten Schritt wird das Reaktionsgefäß kurzzeitig (20-30 Sekunden) auf 94-98 °C erhitzt und dabei werden die beiden Stränge DNA durch Aufbruch der Wasserstoffbrückenbindungen voneinander getrennt (Denaturierung). Anschließend daran wird das Reaktionsgefäß schnell auf 65 °C abgekühlt, um die Rückbildung der Doppelstränge zu verhindern. Im zweiten Schritt wird das Reaktionsgefäß 20-40 Sekunden lang auf einer bestimmten Temperatur gehalten, die je nach Länge und Sequenz des Primers individuell ermittelt wird (meistens 55-65 °C). Dabei bindet der Primer an die 3'-Enden beider Template-Stränge (Primerhybridisierung, *annealing*). Der Primer ist eine kurze Sequenz aus Nukleotiden, die sich komplementär an ein Ende des Target-Gens anlagert und damit einer Initiierung der Polymerisation dieses spezifischen Target-Gens dient. Im letzten Schritt eines PCR-Zyklus wird das Reaktionsgefäß auf 72 °C aufgewärmt, so dass die DNA-Polymerase mit der Synthetisierung der neuen komplementären Stränge anfängt (Elongation/Extension). Die in diesem Schritt eingesetzte DNA-Polymerase stammt aus den in heißen Quellen lebenden thermophilen Bakterium *Thermus aquaticus* und deren hohe Thermostabilität vor allem im Denaturierungsschritt der PCR trug zur Erleichterung des PCR-Verfahrens bei.



**Abb. 2-2 Erste 3 Zyklen der konventionellen PCR** Hier sind erste 3 Zyklen der PCR dargestellt. Die bei jedem Zyklus gebildeten Tochter-DNA-Stränge werden im nächsten Schritt als Template verwendet (semikonservativ). Zu beachten ist, dass die PCR-Produkte mit der gesuchten DNA-Sequenz (Amplikons) erst nach dem 3. Zyklus erscheinen (gelb umrahmt). Davor sind sie noch flankierende Restsequenzen der Ausgangs-DNA vorhanden (*anchored products*). Wenn die PCR n-mal abgelaufen ist, sind (im Idealfall)  $2^n$  Tochter-DNA gebildet worden. *Abbildung aus Alberts et al. S.474*<sup>36</sup>

Wenn die Reaktionsbedingung optimal ist, verdoppelt sich das Target-Gen im Idealfall nach jedem abgelaufenen Zyklus und die neu synthetisierten Stränge dienen beim nächsten Zyklus als zusätzliche Templates für weitere Amplifikation (semikonservative Kettenreaktion). Ein PCR-Experiment besteht typischerweise aus 20- bis 40-Zyklen und der Erfolg der PCR d.h. Vorhandensein spezifischer Amplifikationsprodukte und deren Menge (semiquantitativ) wird nach dem Ende der gesamten PCR-Zyklen durch separates Nachweisexperiment z.B. Agarose-Gelelektrophorese validiert (Endpunkt-PCR).

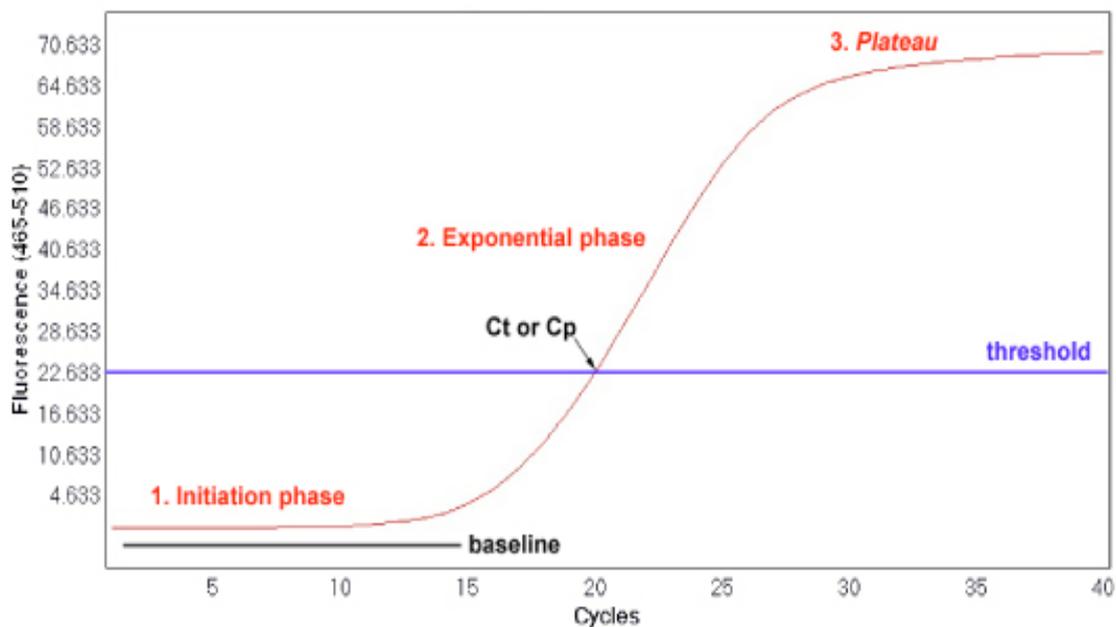
### 2.1.2 Vorteil der RT-qPCR gegenüber der klassischen PCR

Ein wesentlicher Unterschied zwischen der klassischen PCR ggf. mit vorgeschalteter reverser Transkription (RT-PCR) und der RT-qPCR ist, dass bei der Letzteren eine Fluoreszenzmarkierung der zu amplifizierenden Nukleinsäuresequenz eingeführt wird, deren Zunahme im Laufe der Reaktionen in Echtzeit erfasst werden kann (siehe dazu im Detail den Abschnitt 2.1.4). Die Auswertung der Signalkurve erlaubt ein noch schnelleres und exakteres Ergebnis mit Hilfe der RT-qPCR als bei der klassischen PCR (=semi-quantitativ) erzielt werden. Eine Schmelzkurvenanalyse am Ende der qPCR erlaubt Rückschlüsse auf den Ablauf der Reaktion und die Bildung der korrekten Reaktionsprodukte (Target-Amplifikation vs. Off-Target Amplifikation,

Primer-Dimere, etc.). In der klassischen Endpunkt-PCR fehlt diese Echtzeit-Information und der Erfolg der Vervielfältigung ist erst nach vollständig abgelaufenen Reaktionen zu beantworten. Durch Wegfall der Post-PCR-Evaluierung mittels Gelelektrophorese verkürzt sich die gesamte Experimentdauer bei der RT-qPCR und damit erhöht sich der Durchsatz der Experimente. Dazu können bei der RT-qPCR Probleme aufgrund von längerer Handhabung wegen des Post-PCR-Schrittes wie Kontamination und Manipulation der Amplifikationsprodukte vermieden werden.

### 2.1.3 Amplifikationsplot der RT-qPCR

Trägt man die abgelaufenen PCR-Zyklen auf der Abszissenachse und das Fluoreszenzsignal aus jeweiligen PCR-Zyklen auf der Ordinatenachse auf, erhält man einen Amplifikationsplot. Darin zeigen sich typischerweise drei Abschnitte: Baseline, exponentielle Phase und Plateauphase.



**Abb. 2-3 Amplifikationsplot der RT-qPCR.** Der Amplifikationsplot zeigt einen sigmoidalem Kurvenverlauf mit drei Abschnitten: Initiale Baseline, exponentielle Phase, Plateauphase. Am Schnittpunkt zwischen dem Schwellenwert und der Amplifikationskurve lässt sich der  $C_q$ -Wert ablesen. Innerhalb der exponentiellen Phase sollte es einen Wendepunkt geben, wo sich die Vervielfältigung des Amplikons aufgrund von nun stark verbrauchten Reaktionskomponenten (Primer, dNTP, usw.) zu verlangsamen beginnt. Die Reaktion läuft nicht mehr mit maximaler ressourcen-unbeschränkter Effizienz. *Grafik aus Rodriguez-Lázaro et al.*<sup>37</sup>

Als Baseline- oder Background-Signal sind Fluoreszenzsignale aus den initialen PCR-Zyklen (gewöhnlich bis zu 18. Zyklus) bezeichnet. Da in dieser Anfangsphase noch nicht genügend Amplifikationsprodukte vorliegen, um einen statistisch signifikanten Aufschwung des Signals zu erzeugen und auch das Instrument selbst nicht sensitiv genug für Detektion dieser schwachen Signale ist, verläuft die Signalkurve mit niedrigem Fluoreszenzsignal parallel zur Abszisse.

Dieser Baseline-Phase schließt sich eine exponentielle Phase mit steilem Aufstieg der Fluoreszenzsignale an. In dieser Phase vermehrt sich das Amplikon exponentiell aufgrund von reaktionsfördernden Bedingungen, wie beispielsweise noch überschüssig vorhandene Reaktanten sowie hocheffiziente DNA-Polymerase, begünstigte Primerhybridisierung bei weniger Konkurrenz zwischen Primern und noch in geringer Menge vorliegenden Amplifikationsprodukten usw.<sup>38</sup> All diese Faktoren tragen zur exponentiellen Vermehrung des Amplikons bei und die PCR-Effizienz beträgt hier fast 100%, also in jedem PCR-Zyklus verdoppelt sich das Amplikon.

Diese Besonderheiten der exponentiellen Phase ermöglichen eine präzise Abschätzung der Menge des Startmaterials, indem eine Analyse des sog. Cq-Wertes (*cycle of quantification*), auch Ct-Wert (*cycle threshold*) genannt, durchgeführt wird. Zur Ermittlung des Cq-Wertes wird zunächst ein sog. Schwellenwert (*Threshold line*) des Fluoreszenzsignals festgesetzt. Der Schwellenwert bezeichnet ein bestimmtes Niveau des Fluoreszenzsignals, über das ein statistisch signifikanter Signalanstieg über die Baseline angenommen wird und das bedeutet wiederum, dass Fluoreszenzsignale über diesem Schwellenwert tatsächlich aus dem sich exponentiell vermehrenden Amplikon stammen. Die RT-qPCR-eigene Software legt den Schwellenwert gewöhnlich auf ein Signalniveau mit 10-facher Standardabweichung vom Baseline-Fluoreszenzsignals fest. Der Cq-Wert entspricht letztendlich dem PCR-Zyklus am Schnittpunkt vom Schwellenwert und der Amplifikationskurve. Sobald der Cq-Wert frühzeitig im Laufe der RT-qPCR erfasst wird, kann das Vorhandensein des Zielgens in der untersuchten Probe bestätigt werden. Außerdem kann durch Erstellung einer Standardkurve anhand der ermittelten Cq-Werte die Start-Menge des Templates präzise abgeschätzt werden, zumal die Cq-Werte damit in einer guten Korrelation mit umgekehrter Proportionalität stehen, d.h. je höher die initiale DNA-Menge ist, desto kleiner wird der Cq-Wert und desto früher kommt der Anstieg des Fluoreszenzsignals.

Mit Fortschritt der Amplifikation werden die oben erwähnten reaktionsfördernden Faktoren langsam ausgeschöpft, so dass die Kurve allmählich in eine Plateauphase übergeht. Diese Verlangsamung der PCR-Reaktion durch Beschränkung der Reaktionsressourcen kann bereits innerhalb der exponentiellen Phase identifiziert werden, nämlich am Wendepunkt der Kurvensteigung (**Abb. 2-3** Pfeil). In der Plateauphase liegen die Reaktionsprodukte im Vergleich zu den Ausgangsverbindungen im Überschuss vor und die Aktivität der DNA-Polymerase ist durch den Anstieg der reaktionshemmenden Pyrophosphate vermindert. Überdies läuft der erste Schritt der PCR, die DNA-Denaturierung, in dieser Plateauphase nicht mehr effizient ab. All diese chemischen Gegebenheiten führen zu erheblicher Herabsetzung der PCR-Effizienz. Aus diesem Grund geht auch die umgekehrte Proportionalität zwischen Fluoreszenzsignal und initialer DNA-Menge verloren. Wenn eine Abschätzung der Startmenge der Template-DNA in dieser Plateauphase erfolgt, kann das Ergebnis stark variieren. Somit ist die Quantifizierung der

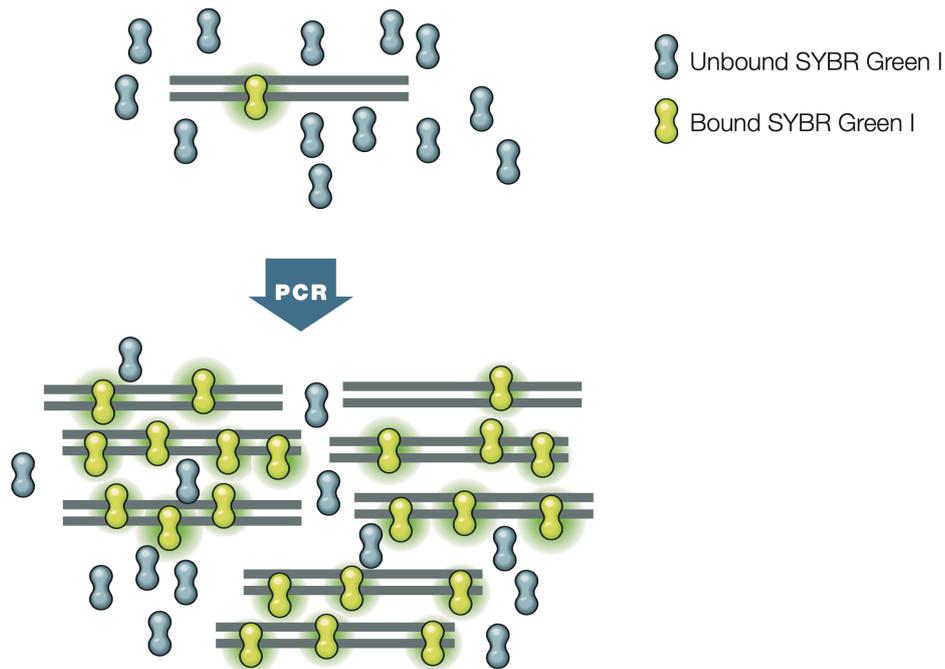
klassischen PCR mittels Gelelektrophorese am Endpunkt der PCR höchstens semiquantitativ – ein weiterer Vorteil der RT-qPCR.

#### 2.1.4 Fluoreszenzfarbstoff für RT-qPCR

Die Einführung von Fluoreszenzfarbstoffen, die in der Lage sind an DNA-Doppelstränge zu binden, war ein maßgeblicher Wendepunkt für die Weiterentwicklung von der klassischen zur RT-qPCR. Generell sind zwei Gruppen von Fluoreszenzfarbstoffen zu unterscheiden, nämlich sequenzunspezifische und -spezifische Fluoreszenzfarbstoffe.

Ein bekannter Vertreter der sequenzunspezifischen Gruppe der Fluoreszenzfarbstoffe ist *SYBR Green*. Dieser Fluoreszenzfarbstoff bindet an die kleine Furche der doppelsträngigen DNA (Interkalation) unspezifisch für deren Sequenz und sendet bei Anregung mit UV-Licht ein Fluoreszenzsignal aus. Der Fluoreszenzfarbstoff erfährt bei der Bindung an die DNA eine Konformationsänderung, die das Fluoreszenzsignal im Vergleich zur ungebundenen Version stark erhöht. Im Lauf der RT-qPCR-Zyklen wird immer mehr dsDNA gebildet und korrespondierend dazu erhöht sich das Fluoreszenzsignal und kann sich so bis zu 1000-fach verstärken.

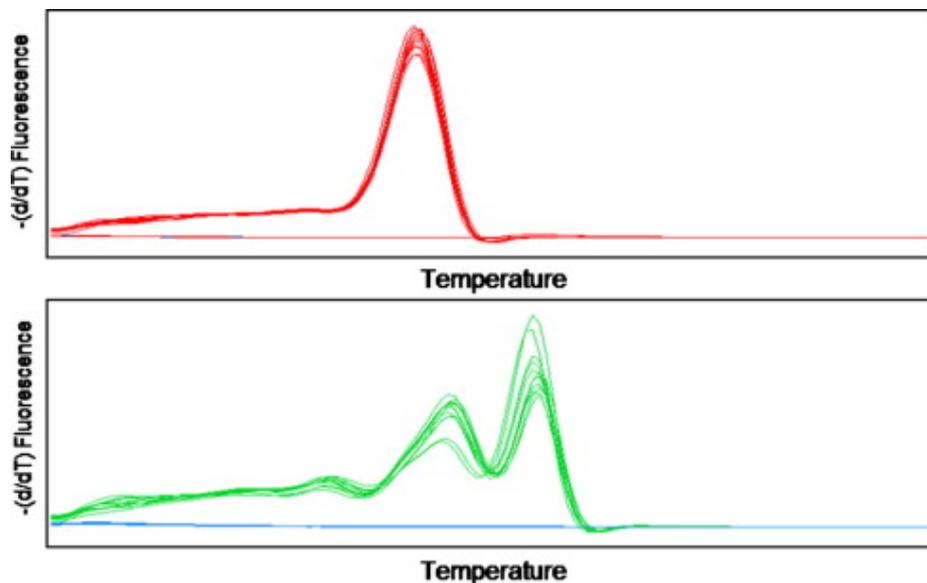
Obwohl die Sequenzunabhängigkeit dieser Farbstoffgruppe zu erleichtertem Experimentaufbau und Detektion multipler Gene beiträgt, bringt das auch einen wesentlichen Nachteil mit sich, und zwar die Tatsache, dass sie prinzipiell auch an irrelevante doppelsträngige Moleküle (z.B. Primer-Dimer und unspezifische PCR-Produkte) binden können. Das könnte unter Umständen zu Fehlinterpretationen wie z.B. einer zu hohen Expression des Zielgens führen.



**Abb. 2-4 Funktionsweise von SYBR Green I.** Der SYBR Green I-Farbstoff gibt im freien Zustand wenig Fluoreszenzsignal in die Umgebung ab. Nach Anbindung an dsDNA erhöht sich das Fluoreszenzsignal durch eine Konformationsänderung erheblich. Bei der Hybridisierungsphase (oberes Teilbild) mit noch wenigem an DNA gebundenem Farbstoff zeigt sich spärliche Signalintensität (grüner Kreis um den gebundenen Farbstoff herum). In der Polymerisations- und Elongationsphase (unteres Teilbild) wird jedoch die maximale Signalintensität erreicht. Gewöhnlich wird das Fluoreszenzsignal am Ende jedes Elongationsschrittes bei Wellenlänge 530nm gemessen. *Abbildung aus Real-Time PCR Applications Guide Biorad.*<sup>39</sup>

Um diesem Problem zu begegnen, kann am Ende jedes RT-qPCR-Experimentes mit sequenzunabhängigem Fluoreszenzfarbstoff eine sog. Schmelzkurve-Analyse durchgeführt. Sie dient zur Charakterisierung der PCR-Produkte und damit zur Identifizierung der Primer-Dimere oder anderer unspezifischer Produkte. Dafür wird ein zusätzlicher Zyklus nach der Beendigung der gesamten Amplifikationszyklen durchlaufen. Dabei wird die Temperatur langsam über die Schmelztemperatur der Primer (Primer- $T_m$ , z.B. 65°C) gesteigert und das Verhalten des Fluoreszenzsignals beobachtet. Bei langsamer Erhitzung auf diese Weise schmelzen die DNA-Doppelstränge und ihre Einzelstränge dissoziieren voneinander. Das Fluoreszenzsignal aus den dsDNA-Farbstoff-Komplexen lässt im Laufe der Erhitzung daher nach, weil der Fluoreszenzfarbstoff im freien Zustand viel weniger Signal abgibt. Wenn diese Veränderung des Fluoreszenzsignals gegen die Temperatur aufgetragen wird, erhält man die Schmelzkurve mit einem typischen steilen Abfall des Signals an der Schmelztemperatur ( $T_m$ ), an der die DNA-Moleküle jeweils zur Hälfte als Doppel- und Einzelstränge vorliegen. Um den steilen Signalabfall an der Schmelztemperatur, d.h. den Wendepunkt (Veränderung der Steigung) der Schmelzkurve visuell besser darzustellen, wird die Kurve als erste negative Ableitung dieser Schmelzkurve gegen die Temperatur ( $-dF/dT$ ) dargestellt. Wenn lediglich das gewünschte

Amplifikationsprodukt in der abgelaufenen RT-qPCR vervielfältigt wurde, dann sollte sich ein Schmelzpeak aus der Schmelzkurvenanalyse ergeben (**Abb. 2-5**).



**Abb. 2-5 Schmelzkurvenanalyse und Schmelzpeak.** In der Grafik ist die Veränderung der Signalintensität gegen Temperatur aufgetragen. Man sieht typischerweise nur einen Peak an der Schmelztemperatur ( $T_m$ ), wenn unspezifische PCR-Produkte ausgeschlossen sind (rote Kurve). Bei Amplifikation der unspezifischen PCR-Produkte (z.B. Primer-Dimer) lassen sich multiple Peaks (grüne Kurve). Grafik aus D'haene et al.<sup>40</sup>

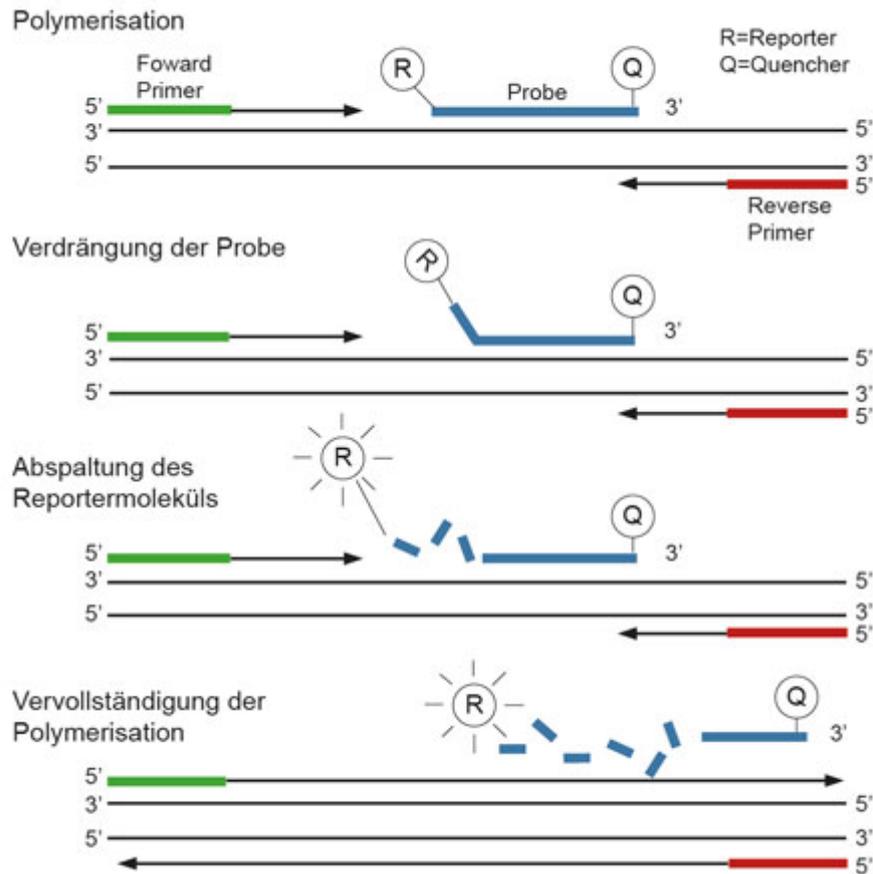
Die zweite Gruppe von Fluoreszenzfarbstoffen bindet an das Target-Amplikon sequenzspezifisch mit Hilfe komplementärer Oligonukleotid-Sonden. Obwohl für diese Gruppe je nach Hersteller zahlreiche Varianten verfügbar sind, ist das zugrunde liegende Prinzip gleich, der sog. FRET-Mechanismus (**F**örster-**R**esonanzenergietransfer). Dieser physikalische Mechanismus beschreibt die Energieübertragung zwischen zwei lichtempfindlichen Molekülen (Chromophore). Demnach wird die Energie eines angeregten Farbstoffs (Donor) auf den zweiten Farbstoff (Akzeptor) durch nicht-radioaktive, intermolekulare Kräfte (Dipol-Dipol-Wechselwirkungen) übertragen und dabei ist die Effizienz dieser Energieübertragung umgekehrt proportional zur sechsten Potenz des intermolekularen Abstandes.



**Abb. 2-6 FRET-Mechanismus.** Förster (Fluoreszenz)-Resonanzenergietransfer beschreibt einen physikalischen Prozess, bei dem die Energieübertragung vom Donor auf den Akzeptor von ihrer Nähe zueinander abhängig erfolgt. Die linke Bildhälfte zeigt den eben wirkenden FRET-Prozess. Hier wird die durch Exzitation aufgetretene Energie des Donors (grün) vom Akzeptor (rot) gleich übernommen, weil sich die Beiden in der Nähe befinden. Rechts ist eine Energieabgabe in die Umgebung als Fluoreszenzsignal abzulesen, wegen großer Entfernung beider Chromophore. *Abbildung aus Real-time PCR handbook, Lifetechnologies.*<sup>41</sup>

Basierend auf diesem Mechanismus wird die Position der beiden Komponenten des Fluoreszenzfarbstoffes (Donor/Reporter und Akzeptor/Quencher) so gestaltet, dass sie im von DNA dissoziierten Zustand sehr nah aneinander stehen und somit bei wirkendem FRET-Mechanismus keine Emission des Signals zustande kommt (*quenched*). Aber wenn der Abstand zwischen beiden Farbstoffmolekülen durch Amplifikation der Zielsequenz zunimmt, dann bleibt der FRET-Mechanismus aus und bei Anregung des Donors die Energie als Fluoreszenzsignal in die Umgebung abgegeben. Dieses Signal wird durch RT-qPCR-Instrument registriert.

Dieses Prinzip wird bei den Hydrolyse- (TaqMan-) Assays ausgenutzt, die in der vorliegenden Arbeit angewendet wurde. Dabei wird zunächst eine Sonde aus Oligonukleotiden ausgewählt, die eine komplementäre Sequenz zur Target-Sequenz besitzt und somit dafür spezifisch ist. Diese Sonde wird dann an deren beiden Enden mit Fluoreszenzfarbstoff markiert, nämlich am 5'-Ende mit einem Fluoreszein-Derivat als Donor/Reporter z.B. 6-Carboxy-Fluorescein (FAM) und am 3'-Ende mit einem Rhodamin-Derivat als Akzeptor/Quencher z.B. 6-Carboxy-tetramethyl-rhodamin (TAMRA). Da das Fluoreszenzsignal aus dem Reporter bei intaktem Zustand der an der Zielsequenz hybridisierten Sonde durch den Quencher effektiv unterdrückt wird (FRET-Mechanismus), ist dabei kein Fluoreszenzsignal detektierbar. Bei der Extensionsphase wird das 5'-Ende der Sonde durch die Taq-Polymerase verdrängt. Dabei entsteht eine Y-förmige Sekundärstruktur des Sonde-Zielsequenz-Komplexes, wo die 5'-3'-Exonuklease der Taq-Polymerase angreifen kann. Dadurch wird die Sonde hydrolytisch zerlegt und der Abstand zwischen dem Reporter und dem Quencher wird vergrößert. Nun kann das Fluoreszenzsignal nicht mehr nach FRET-Mechanismus unterdrückt werden, sondern wird in die Umgebung abgegeben, was durch das qPCR-Gerät detektiert wird.



**Abb. 2-7 Funktionsweise von TaqMan-Assay.** In den obersten Doppelsträngen ist eine hybridisierte TaqMan-Sonde zu erkennen, an welcher 5'-Ende ein Reporter und an 3'-Ende ein Quencher versehen sind. Hier wird das Fluoreszenzsignal aus dem Reporter aufgrund von räumlicher Nähe zum Quencher effektiv unterdrückt. Bei der Extensionsphase der qPCR wird jedoch das den Reporter tragende 5'-Ende der Sonde durch die Polymerase verdrängt und durch die 5'-3'-Exonukleaseaktivität der Polymerase zerlegt. Nun wird das Fluoreszenzsignal aus dem Reporter aufgrund von fehlendem FRET-Mechanismus in die Umgebung abgegeben und von qPCR-Gerät detektiert. *Abbildung aus der Website von MVZ Martinried.*<sup>42</sup>

## 2.1.5 Bedeutung der RNA-Qualität in der RT-qPCR

### 2.1.5.1 RNA-Reinheit und -Integrität

Die Bedeutung einer hohen RNA-Qualität für verlässliches RT-qPCR-Ergebnis wurde bereits durch zahlreiche Studien belegt.<sup>43-45</sup> Aufgrund des ubiquitären Vorkommens RNA-abbauender Enzyme, sog. Ribonukleasen (kurz RNasen), sind RNA-Moleküle trotz ihrer thermodynamischen Stabilität sehr anfällig für Degradierung.<sup>45</sup> In RT-qPCR-Experimenten kann sich die RNA-Degradierung in jedem Schritt ereignen, z.B. bei unsachgemäßer Gewebeentnahme oder RNA-Extraktion, Lagerung bei ungünstigen Bedingungen, inadäquater Handhabung der RNA-Proben usw.<sup>46,47</sup> Aus diesen Gründen sollten bereits vor Beginn der Experimente Maßnahmen getroffen werden, die eine Degradation der RNA verhindern, wie z.B. adäquate Probenlagerung bei -20 °C

oder -80 °C, saubere und sterile Handhabung im Experiment zur Vermeidung der RNase-Wirkung ggf. mit Verwendung eines RNase-Inhibitors.

Die RNA-Qualität wird generell hinsichtlich zweier Aspekte überprüft, nämlich Reinheit und Integrität der RNA-Proben. Die RNA-Reinheit (*purity of RNA*) besagt, wie weit die extrahierten RNA-Proben frei von Verunreinigungen sind. Somit wird die hohe RNA-Reinheit durch Vermeidung von Kontamination mit genomischer DNA, Proteinen, organischen Verbindungen erreicht, die sich bei unsachgemäß durchgeführter RNA-Extraktion beimischen könnten. Außerdem trägt eine Beseitigung verschiedener organischer Substanzen mit hemmender Wirkung auf die reverse Transkription bzw. auf die RT-qPCR auch zur Steigerung der RNA-Qualität bei. Zu diesen inhibitorischen Faktoren zählen z.B. Hämoglobin, Fett, Glykogen, Zellkomponenten, Kalziumionen, DNA-bindende Proteine, Gallensäure, IgG. Durch Beeinträchtigung der RT-qPCR-Effizienz können sich diese Substanzen besonders negativ auf RT-qPCR-Experimente auswirken.<sup>45,48</sup>

Die konventionelle Methode zur Bestimmung der RNA-Reinheit stellt die spektrophotometrische Messung der Probe bei verschiedenen Wellenlängen dar. Die Absorption von Licht bei einer Wellenlänge von 260nm ist spezifisch für Nukleinsäuren aller Arten, d.h. sowohl RNA als auch DNA, und bei 280nm für Proteine, inklusive die mit aromatischen Aminosäureresten (Tryptophan, Histidin, Phenylalanin und Tyrosin). Folglich weist ein niedriger  $A_{260}/A_{280}$ -Quotient auf Kontamination mit Proteinen hin. Bei einem  $A_{260}/A_{280}$ -Quotient von 1,8-2,2 wird es von einer akzeptabel reinen RNA ausgegangen.

Eine Reihe von organischen Substanzen z.B. Guanidiniumthiocyanat und Phenol nehmen UV-Licht bei 230nm auf. Aufgrund dessen kann anhand des  $A_{260}/A_{230}$ -Quotienten überprüft werden, ob solche organischen Verbindungen, die üblicherweise zum Zellaufschluss während der RNA-Isolierung dienen, in der Probe zurückgeblieben sind. Reine RNA weist einen  $A_{260}/A_{230}$ -Quotienten von mindestens 1,7 auf.

Der wesentliche Vorteil der spektralphotometrischen Analyse besteht darin, dass sehr geringe RNA Volumen (0,5-2,0 µl) zur gleichzeitigen Bestimmung der RNA-Reinheit und -Konzentration (Lambert-Beer-Gesetz) ausreichen. Nachteilig ist jedoch die relativ niedrige Spezifität und Sensitivität dieser Methode vor allem bei Proben, die in sehr niedriger Konzentration (unter 100 ng/µl) vorliegen.<sup>6,49</sup> Ein hoher  $A_{260}/A_{280}$ -Quotient kann anstatt durch hohe RNA-Reinheit, auch durch eine Kontamination mit genomischer DNA verursacht werden, weil verschiedene Arten von Nukleinsäuren (ds- oder ssDNA, RNA usw.) bei einer Wellenlänge von 260nm Licht absorbieren und folglich spektralphotometrisch nicht voneinander unterschieden werden können. Zur RNA-Integrität kann die Methodik keine Aussagen treffen, da stark degradierte RNA selbst problemlos UV-Licht absorbiert und damit unter Umständen eine hohe RNA-Reinheit vortäuschen kann.

Die RNA-Integrität beschreibt die Intaktheit der isolierten RNA-Moleküle. Zur Integritätskontrolle einer RNA-Probe würde am besten der mRNA-Anteil der RNA-Probe genommen, weil die mRNA als eigentliche Zielsequenz im Mittelpunkt der Genexpressionsanalyse steht und sich somit den entscheidenden Faktor für den Erfolg der RT-qPCR darstellt. In Wirklichkeit ist eine Überprüfung der mRNA-Integrität kaum möglich, weil die mRNA nur 1-3% der Gesamt-RNA ausmacht und infolgedessen selbst mit der sensitivsten Methode schwer nachzuweisen ist. Im Gegensatz dazu nimmt ribosomale RNA (rRNA) den überwiegenden Anteil der Gesamt-RNA (bis zu 90%) ein. Daher ist die Integrität der rRNA als Proxy für die Qualität der Gesamt-RNA, einschließlich des mRNA-Anteils, anzusehen.<sup>46,50</sup> In der eukaryotischen Zelle bilden 28S- und 18S-rRNA-Moleküle den Hauptanteil der rRNA. Da die Größe der 28S- und 18S-rRNA jeweils ca. 5kb und 2kb beträgt, sollte das Größenverhältnis der beiden rRNA-Untereinheiten bei vollständiger Intaktheit idealerweise ca. 2.5 sein. Generell geht man aber bereits ab einem rRNA-Größenverhältnis über 2 von einer hohen RNA-Integrität aus. Auf Basis dieser Grundlage wurde traditionell die Integrität der 28S- und 18S-rRNA zum Zweck der Qualitätskontrolle einer RNA-Probe herangezogen, unter der Annahme, dass sie die Integrität der mRNA vollständig vertreten würde.

Die klassische Methode zur Bestimmung der RNA-Integrität basierend auf dem 28S-/18S-rRNA-Verhältnis ist die Gelelektrophorese mit Anwendung von Agarose- oder Acrylamid-Gel. Die Funktionsweise dieser Technik beruht auf der Eigenschaft der RNA-Moleküle, aufgrund ihrer negativen Ladung (Phosphat-Gruppe) im elektrischen Feld zum positiven Pol zu wandern, wobei Fragmente mit großer Molekülmasse langsamer als solche mit kleiner Molekülmasse wandern. Durch Fluoreszenzfarbstoffe wie Ethidiumbromid oder *SYBR Green* (sensitiver als Ethidiumbromid) werden die nach ihrer Masse getrennten RNA-Fragmente als Banden im UV-Licht sichtbar gemacht und durch Vergleich ihrer Laufweite mit Referenzbanden (Leiter, Größenmarker) können die RNA-Fragmente weiter differenziert werden, d.h. 28S-, 18S-rRNA. Die Interpretation des Ergebnisses aus der Gelelektrophorese erfolgt semiquantitativ, indem 28S- und 18S-rRNA-Banden hinsichtlich ihrer Färbeintensität inspiziert werden. Wenn das Verhältnis der Fluoreszenzintensität aus 28S- und 18S-rRNA über 2:1 beträgt, kann es von einer hohen RNA-Integrität ausgegangen werden. Zusätzliche Banden und/oder verschmierte Spuren im Agarosegel deuten auf RNA-Degradierung hin.

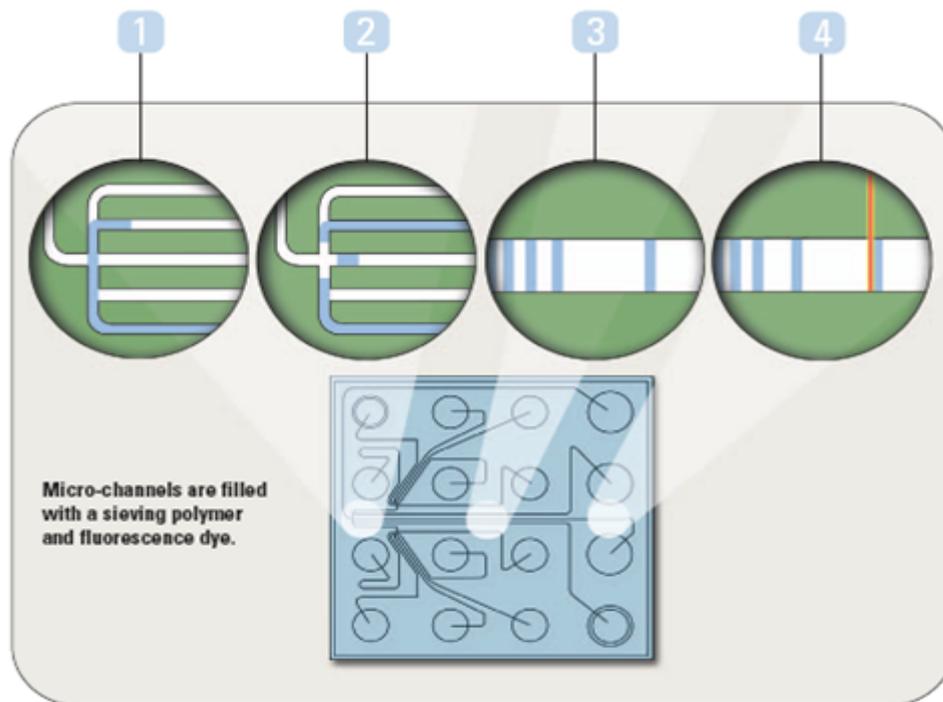
Obgleich die Gelelektrophorese ein kostengünstiges Verfahren ist, ist deren Durchführung aufgrund von erforderlicher manueller Vorarbeit und geringer Automatisierungsmöglichkeit zeitaufwändig (Experimentdauer bis zu zwei Stunden).<sup>51</sup> Trotz diesem Aufwand bei Durchführung besitzt die Information aus der Gelelektrophorese relativ geringe Aussagekraft, denn die Ergebnisse der Gelelektrophorese beruhen vollständig auf subjektiven und semiquantitativen Interpretation der Banden auf dem Gel und gerade aus diesem Grund kann eine

geringgradige RNA-Degradierung nicht sicher erfasst werden.<sup>52</sup> Bei fehlender Möglichkeit zur Standardisierung sind außerdem die Ergebnisse aus unterschiedlichen Laboratorien nicht miteinander vergleichbar. Die Gelelektrophorese benötigt darüber hinaus relativ große Mengen an Untersuchungsmaterial (0,5-2µg Gesamt-RNA), um zuverlässige Datenauswertung durch bessere Visualisierung zu erzielen<sup>52,53</sup> und dieser höhere Materialverbrauch ist unerwünscht bei klinischer Diagnostik, bei der das Untersuchungsmaterial oft nur begrenzt verfügbar ist.<sup>26,53</sup>

Diese Schwächen des klassischen Verfahrens führten zur Entwicklung einer neuen Methode zur RNA-Qualitätsprüfung mit Standardisierungsmöglichkeit, und zwar der Ende der 90er Jahre eingeführten innovativen Technik mit Mikrofluid-Prinzip. Der Agilent 2100 Bioanalyzer war das erste auf dem Markt erhältliche Gerät mit dieser Technik und ist heute der Goldstandard zur RNA-Qualitätskontrolle.<sup>54</sup>

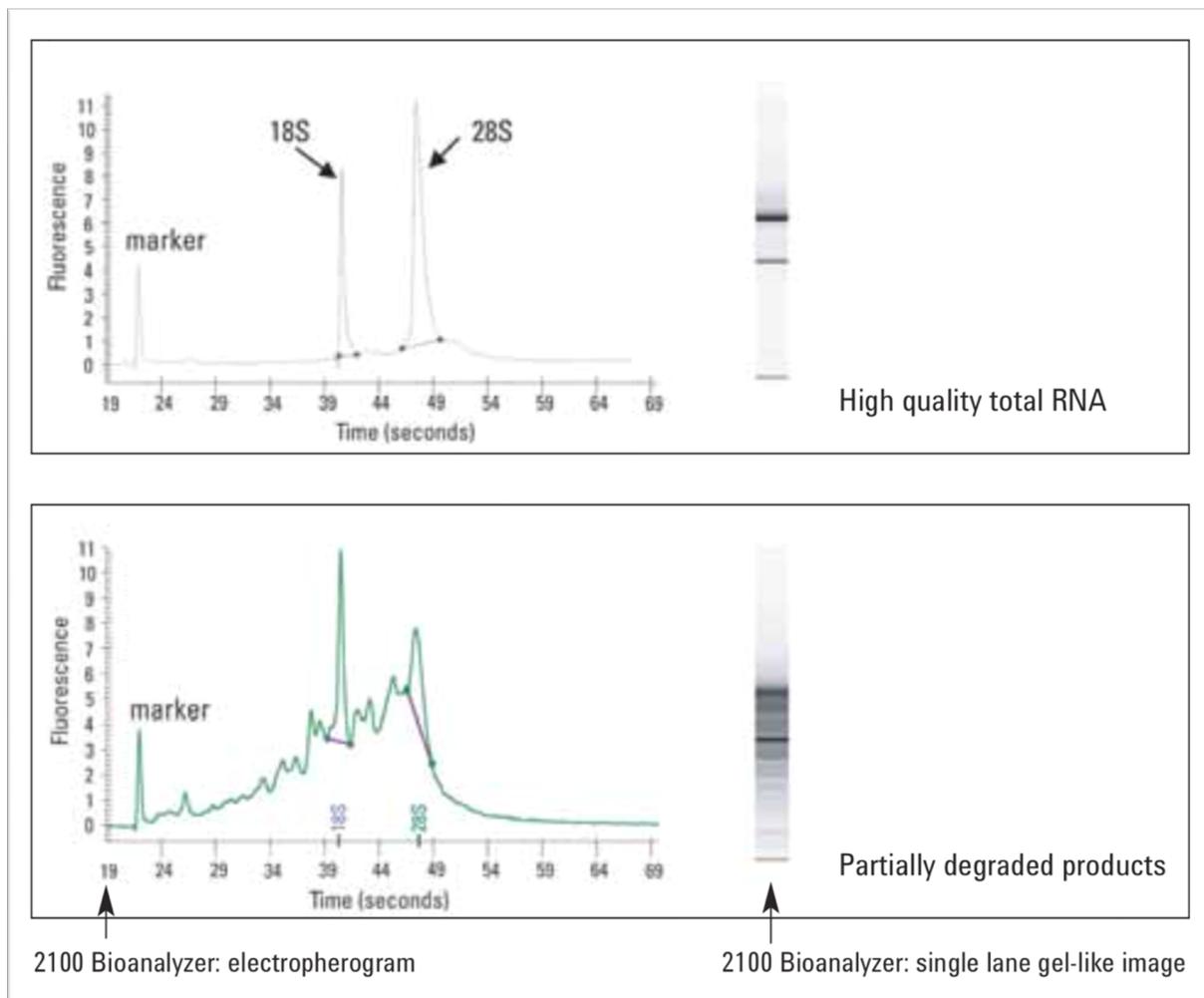
Die Funktionsweise der Probenanalyse im Bioanalyzer gleicht der der klassischen Gelelektrophorese. Aber durch Miniaturisierung der nötigen Instrumente der klassischen Gelelektrophorese gestattet der Bioanalyzer eine verbesserte Präzision, verringertes Kontaminationsrisiko, verkürzte Analysezeiten (ca. 30 min) und deutlich geringeren Probenverbrauch (<200pg Gesamt-RNA). Außerdem ermöglicht die Automatisierung und Zusammenführung der aufwändigen Arbeitsabläufe der klassischen Gelelektrophorese, nicht nur die Reproduzierbarkeit zu erhöhen, sondern auch die gesamte Experimentdauer zu verkürzen. Mit Hilfe digitaler Ausgabeformate können die Experimentergebnisse zu jeder Zeit weiterverarbeitet und die Ergebnisse von Labor zu Labor verglichen werden.

In einem Analysechip des Bioanalyzers können bis zu 12 RNA-Proben in einem Lauf analysiert werden. Die in die Probengefäße gegebenen RNA-Proben werden über die Gel-gefüllten Mikrokanäle in Richtung Trennkanal gezogen. Durch angelegte elektrische Spannung werden die eingespritzten RNA-Proben gemäß ihrer Molekülmasse im Trennkanal aufgetrennt. Die einzelnen Komponenten der RNA-Probe werden mit Hilfe eines Fluoreszenzfarbstoffs durch Laser-induzierte Fluoreszenz (LIF) detektiert.<sup>55-57</sup>



**Abb. 2-8 Binnenstruktur eines Analysechips und Analyseablauf.** In der unteren Bildhälfte ist ein Analysechip mit 16 Reaktionsgefäßen sowie miteinander zusammengeschlossenen Mikrokanälen gezeichnet. Diese Mikrokanäle sind mit Gel als molekulares Sieb für Nukleinsäure oder Proteinen sowie Fluoreszenzfarbstoff aufgefüllt. Die in die Probengefäßen eingefüllten RNA-Proben werden durch elektrische Spannung zum Trennkanal bewegt (Schritt 1 und 2, in der oberen Bildhälfte) und dort werden einzelne Komponente der RNA gemäß ihrer Molekülmasse elektrophoretisch getrennt (Schritt 3). Die getrennten RNA-Komponenten werden durch das Signal aus der Fluoreszenzmarkierung detektiert (Schritt 4). *Abbildung aus Agilent 2100 Bioanalyzer Eine Plattform - zahlreiche Möglichkeiten für die Analyse von DNA, RNA, Proteinen und Zellen, Agilent Technologies, S.3.*<sup>58</sup>

Das Analyseergebnis des Bioanalyzers wird als sog. Elektropherogramm mit gelähnlicher Abbildung und Datentabelle angezeigt. Auf dem Elektropherogramm ist das Fluoreszenzsignal der unterschiedlichen RNA-Untereinheiten gegen Zeit aufgetragen. Da die 28S- und 18S-rRNA die größten Anteile einer RNA-Probe ausmachen, sind diese beiden RNA-Untereinheiten als große Peaks auf dem Elektropherogramm zu sehen. Dabei tritt der 28S-Peak etwas später als der 18S-Peak in Erscheinung, weil die 28S-rRNA gemäß ihrer Molekülmasse langsamer als die 18S-rRNA im elektrischen Feld wandert. Im Falle einer RNA-Degradierung werden diese beiden größten Peaks des Fluoreszenzsignals zuerst betroffen und beginnen entsprechend abzusinken, weil die fluoreszenz-interkalierenden Stellen in der RNA durch die Degradierung geschädigt werden.<sup>45</sup> Dazu steigt die Signalintensität der Fast- und Inter-Region durch anfallende degradierte RNA-Fragmente an. Diese beiden Phänomene verschieben die Signalverteilung auf dem Elektropherogramm insgesamt nach links, also zu kleineren RNA-Untereinheiten (**Abb. 2-9**).<sup>54</sup>

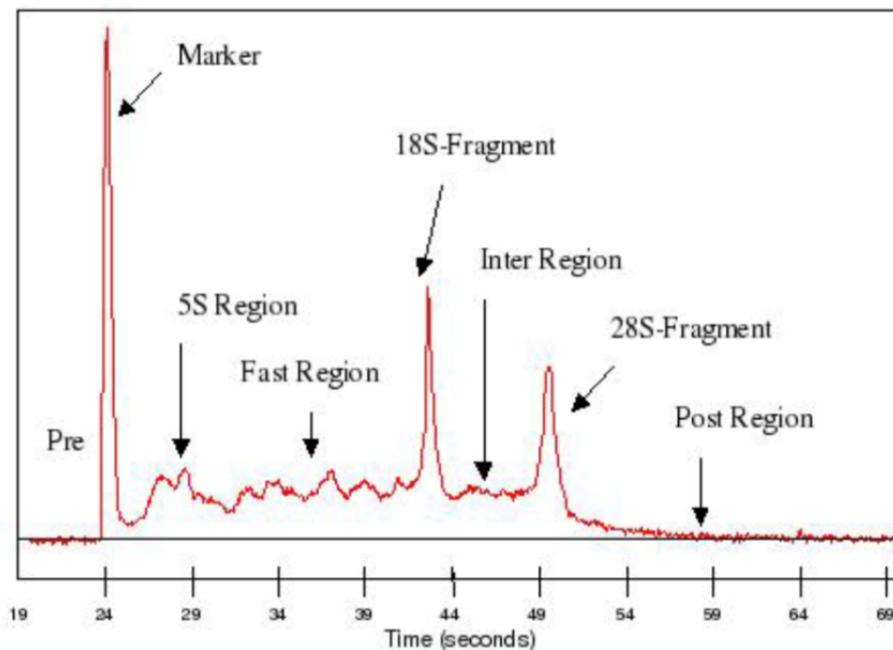


**Abb. 2-9 Bioanalyzer-Ergebnis.** Das Elektropherogramm einer RNA-Probe mit hoher Qualität (oben) weist zwei eindeutig erkennbare Peaks für die 18S- und 28S-rRNA auf und dementsprechend sind auf der gel-ähnlichen Abbildung (rechts daneben) zwei klare Banden zu sehen. Unten erkennt man ein weiteres Elektropherogramm mit verringertem 28S-Peak sowie Erhöhung des Baseline-Signals zwischen den 18S- und 28S-Fragmenten (Inter-Region) bei einer partiell degradierten RNA-Probe. Auf der gel-ähnlichen Abbildung (rechts daneben) sind entsprechend stark verschmierte Zonen erkennbar. *Abbildung aus Agilent 2100 Bioanalyzer Application Compendium S.67.<sup>59</sup>*

Obwohl das 28S/18S-rRNA-Verhältnis weiterhin einen wichtigen Bestandteil der neuen, auf dem Mikrofluid-Prinzip basierten Methode darstellt, reicht sie allein nicht aus, um die Integrität einer RNA-Probe zu überprüfen. Das liegt an der Anfälligkeit der 28S-rRNA für Degradierung. Die 28S-rRNA ist ein großes Molekül im Vergleich zu 18S-rRNA<sup>45</sup> und sie besitzt reichlich Sekundär- und Tertiärstrukturen. Des Weiteren enthält die 28S-rRNA AU-reiche Sequenzen (sog. *hidden break*) und aufgrund dessen durchläuft die 28S-rRNA einen bislang nicht vollständig abgeklärten Prozess und wird dabei in kleinere rRNAs zerlegt.<sup>54</sup> All diese Faktoren führen zu einer höheren Anfälligkeit der 28S-rRNA gegenüber Degradierung im Vergleich zu 18S-rRNA und folglich neigt das 28S/18S-rRNA-Verhältnis dazu, trotz aller Vorsichtsmaßnahmen während der RNA-Vorbereitung a priori niedrig zu sein, was im ersten Blick auf niederqualitative RNA hinweisen

würde. Außerdem mag eine Abnahme des 28S/18S-rRNA-Verhältnisses durch unspezifische Schädigung der RNA bedingt sein, z.B. unsachgemäße Handhabung der RNA-Probe, massive Apoptose oder Nekrose der Gewebe.<sup>45,47</sup> Basierend auf diesen Beobachtungen bezüglich 28S/18S-rRNA-Verhältnis behaupten bereits mehrere Studien, dass das 28S/18S-rRNA-Verhältnis die mRNA-Integrität nicht vollständig abbilden würde und somit ein niedriges 28S/18S-rRNA-Verhältnis nicht zwangsläufig auf eine niedrige mRNA-Qualität hindeuten muss.<sup>11,44,45,52,54,61</sup>

Zur Behebung dieser Problematik und Standardisierung der Ergebnisse aus der RNA-Integritätsbestimmung wurde im Rahmen einer Studie ein statistisch-mathematischer Algorithmus entwickelt, durch den eine sog. RNA-Integritätsnummer (RIN) für die zu untersuchende RNA-Probe berechnet wird.<sup>62</sup> Durch diesen neuen Algorithmus sind nun einheitliche, objektive und reproduzierbare Ergebnisse der RNA-Integritätskontrolle erhältlich und infolgedessen wurde ein Experimentator-unabhängiger Vergleich der Ergebnisse ermöglicht.



**Abb. 2-10 Das Elektropherogramm.** Auf einem Elektropherogramm lassen sich insgesamt neun Regionen des Signals identifizieren und ihre Form und Höhe werden bei der Berechnung des RIN-Wertes berücksichtigt (die sog. Precursor-Region kurz nach 28S-Fragment ist hier nicht dargestellt). Da bei der RNA-Integritätskontrolle alle Signale aus der RNA-Probe überprüft werden, bringt der Bioanalyser aussagekräftigere Ergebnisse im Vergleich zur klassischen Gelelektrophorese. *Abbildung aus Schroeder et al.*<sup>54</sup>

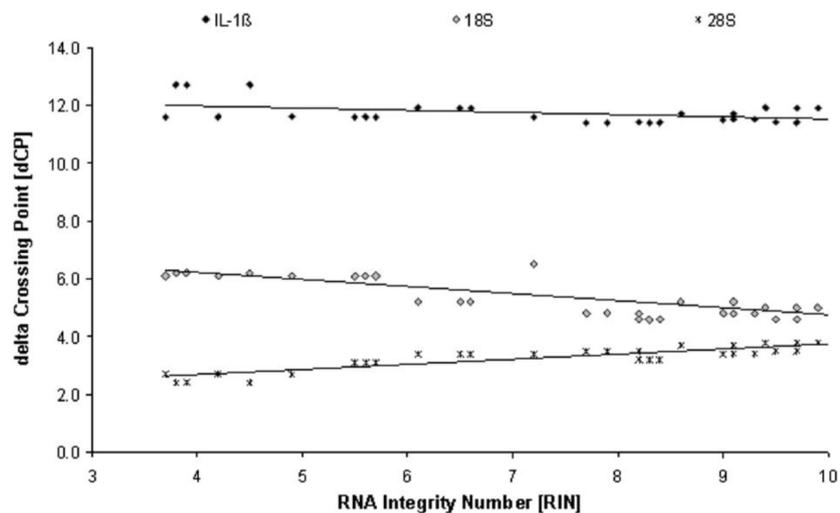
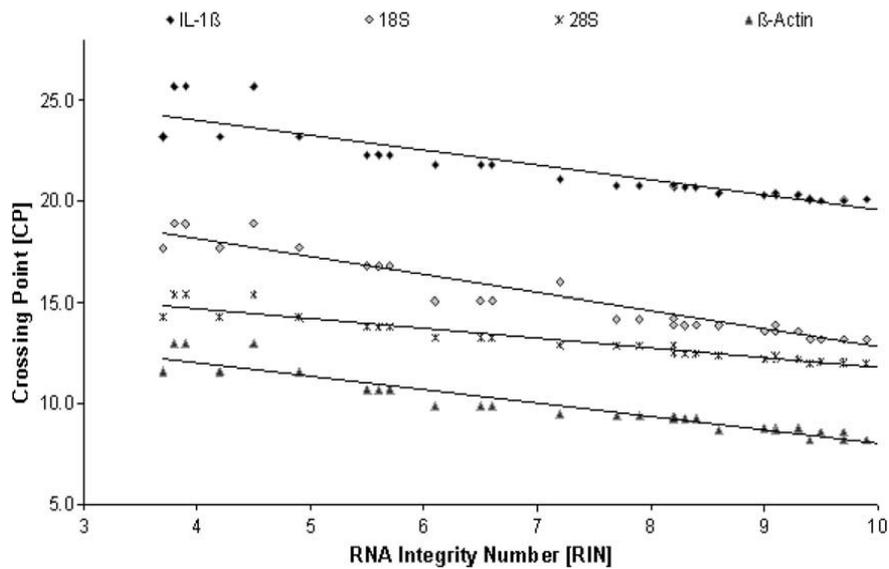
Für eine umfassende Beurteilung der RNA-Integrität werden im sog. RIN (RNA Integrity Number) -Algorithmus nicht nur die 28S- und 18S-rRNA, sondern auch alle sonstigen Komponenten einer RNA-Probe betrachtet, die auf dem Elektropherogramm als verschiedene Regionen des Signals sichtbar werden (**Abb. 2-10**). Diese unterschiedlichen Regionen des Signals haben jeweils eigene Bedeutung, z.B. ist die Höhe des 28S-Peaks ein geeigneter Parameter für eine beginnende

Degradierung der RNA-Probe, weil die 28S-rRNA im Vergleich zu anderen Komponenten der RNA-Probe v.a. 18S-rRNA anfälliger für Degradierungsprozesse ist. Die Fast-Region spiegelt wider, wie weit der Degradierungsprozess fortgeschritten ist. Auf diese Weise werden alle Regionen des Signals auf dem Elektropherogram sowie weitere Eigenschaften des Signals, die durch maschinelles Lernen extrahiert wurden, bei der Festlegung der RIN-Werte miteinbezogen. Dadurch wird die Bestimmung der RNA-Integrität noch präziser und aussagekräftiger als bei der klassischen Methode, die lediglich auf dem Verhältnis der rRNA-Untereinheiten beruht.

Im Rahmen des RIN-Systems wird die Integrität einer RNA-Probe mit einem Wert von 1 bis 10 bewertet, wobei ein RIN-Wert von 1 eine vollständig degradierte RNA-Probe und ein RIN-Wert von 10 eine vollkommen intakte RNA-Probe bezeichnet. Bei einem RIN-Wert über 5 kann von einer guten hohen RNA-Integrität ausgegangen werden und bei RIN-Werten über 8 von einer perfekt für Downstream-Experimente - wie RT-RT-qPCR oder Microarrays - geeigneten RNA-Probe RT-qPCR.<sup>43,45</sup>

#### 2.1.5.2 Einfluss der RNA-Qualität auf die RT-qPCR

Wie bereits erwähnt, stellt eine hohe RNA-Qualität eine unabdingbare Grundlage für die akkurate Quantifizierung der Genexpression dar. Fleige und Pfaffl demonstrierten im Jahr 2006, dass eine Korrelation zwischen RNA-Qualität (RIN-Wert) und Genexpression (Cq-Wert) in konkreten experimentellen Daten vorliegt.<sup>47,63</sup> Dabei wurden RNA-Proben aus unterschiedlichen Geweben und Zelllinien künstlich degradiert, um ein breites Spektrum von Proben mit unterschiedlichen RIN-Werten zu erhalten. In diesen RNA-Proben wurden die Cq-Werte der mRNA von vier Genen mit unterschiedlichem Expressionslevel durch RT-qPCR bestimmt. Daraus wurde eine hochsignifikante, umgekehrt proportionale Korrelation zwischen RIN- und Cq-Werten ( $p < 0,01$ ) gezeigt, d.h. je höher die RNA-Qualität war, desto kleiner waren die Cq-Werte. Daraus wurde gefolgert, dass die RNA-Qualität tatsächlich Einfluss auf das RT-qPCR-Ergebnis hat. Dieser Effekt konnte - zumindest teilweise - durch den Einsatz geeigneter Referenzgene effektiv kompensiert werden (**Abb. 2-11**).



**Abb. 2-11 Einfluss der RNA-Qualität (RIN-Wert) auf Cq-Werte (beim Einsatz geeigneter Referenzgene).** Oben: RNA-Qualität (RIN-Wert) verhält sich umgekehrt proportional zu Cq-Werten. Je höher die RNA-Qualität der Proben war, desto kleiner wurden die Cq-Werte. Diese Korrelation gilt auch für Gene mit unterschiedlichem Expressionsniveau wie IL-1 $\beta$ , 18S- und 28S-rRNA,  $\beta$ -Actin. Unten: Bemerkenswerterweise konnte dieser Effekt stark reduziert werden, wenn die Cq-Werte der Zielgene relativ zu den Cq-Werten geeigneter Referenzgenen quantifiziert wurden ( $dCq = Cq(\text{Target}) - Cq(\text{Referenzgen})$ ). Cp-Wert (*crossing point*) in der Abbildung ist eine alternative Bezeichnung für den Cq-Wert. *Abbildung aus Fleige et al.*<sup>45</sup>

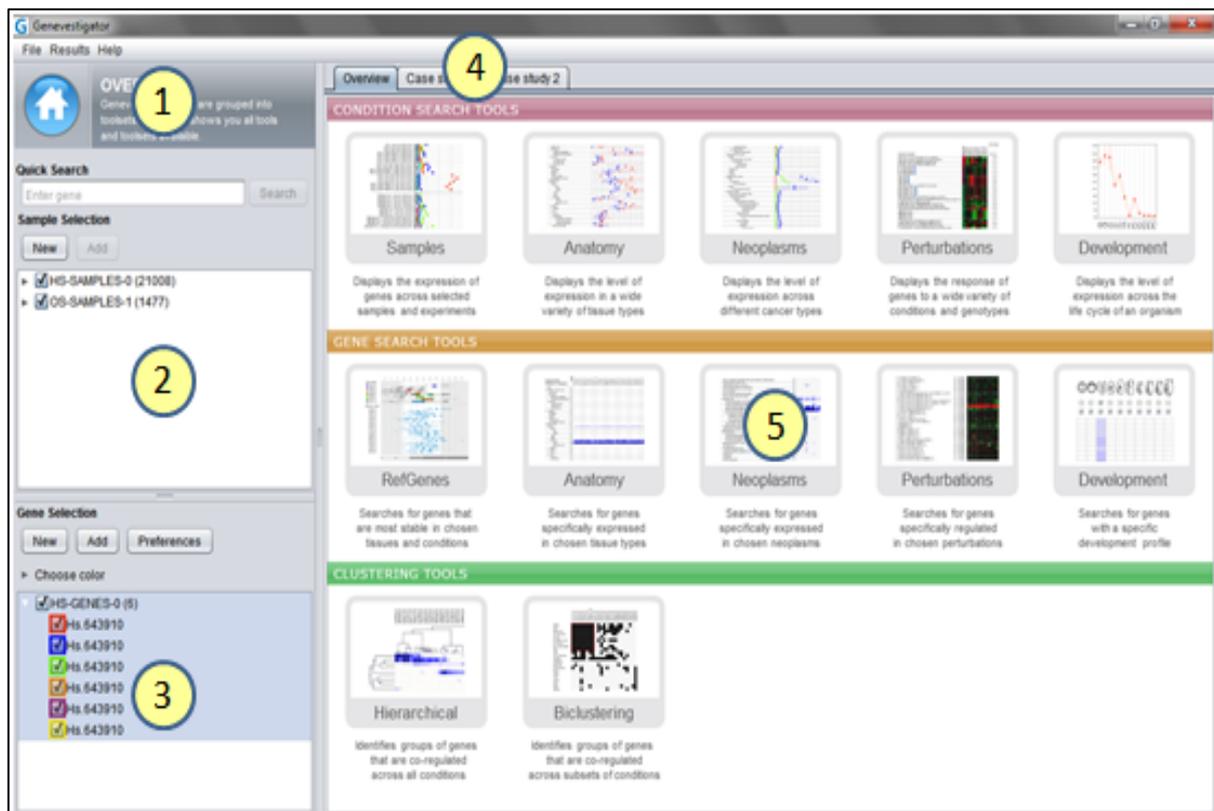
Anders als bei Microarray-Experimenten können bei der RT-qPCR auch noch degradierte Fragmente einer RNA-Probe einwandfrei vervielfältigt werden. So konnten bei Genexpressionsanalysen in RT-qPCR-Studien mit degradierten RNA-Proben gute Ergebnisse erzielt werden, die mit Ergebnissen aus intakten RNA-Proben vergleichbar waren, solange ribosomale Banden bei der RNA-Qualitätskontrolle nachweisbar waren.<sup>39,56</sup> Aus diesem Grund

muss ein sinnvolles Ergebnis nicht unbedingt ausgeschlossen werden, auch wenn RNA-Proben einmal aus unerwarteten Gründen degradiert sind. In diesem Fall ist eine relative Quantifizierung mit bestgeeigneten Referenzgenen zu empfehlen, zumal dadurch, wie oben beschrieben, der Effekt der niedrigen RNA-Qualität klein gehalten werden kann.

## 2.2 Genevestigator

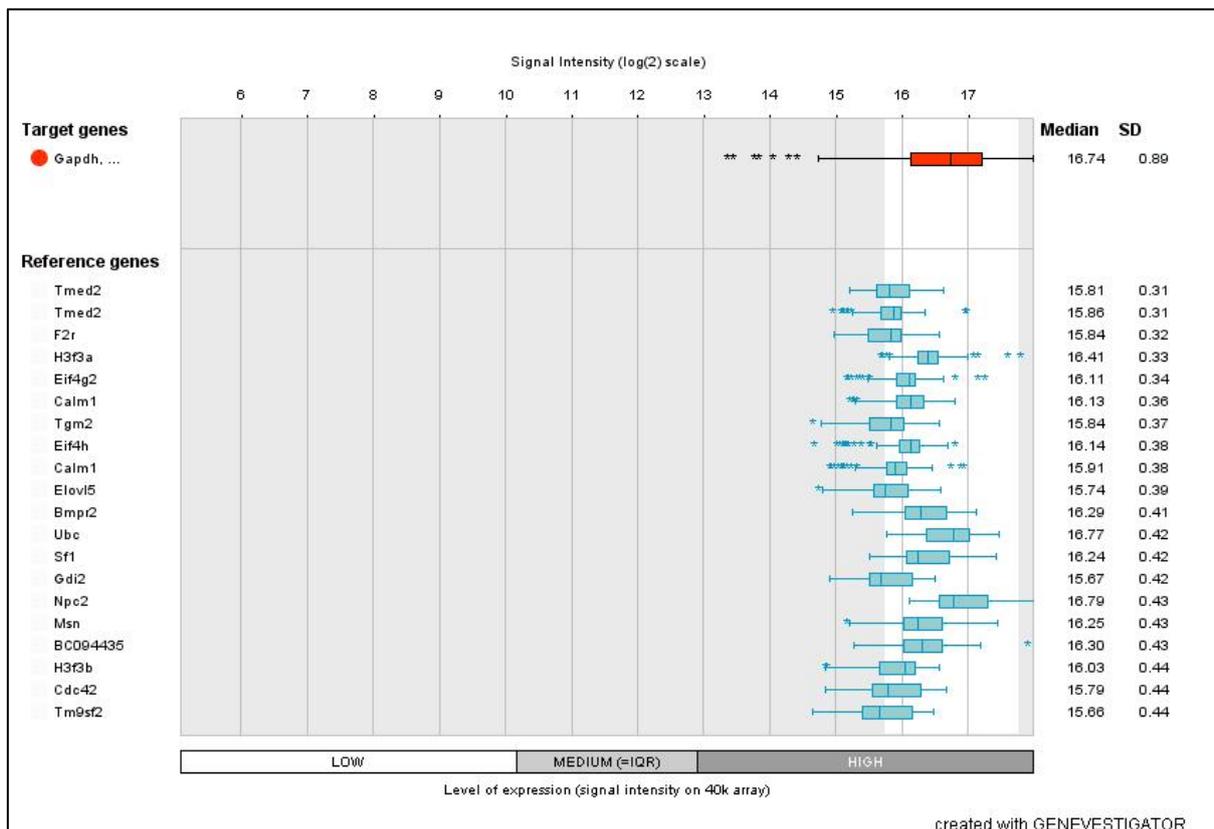
Genevestigator<sup>®</sup> ist eine web-basierte Software mit diversen leistungsfähigen Subtools zur Analyse bereits publizierter Genexpressionsdaten. Die riesige Microarray-Datenbank von Genevestigator ermöglicht Metaanalysen der verfügbaren Genexpressionsdaten insbesondere hinsichtlich spezifischer Versuchsbedingungen, die dem Studiendesign des Experimentators entsprechen. Zu diesem Zwecke wurden im Rahmen des Genevestigator-Projektes die verfügbaren Microarrays aus bekannten Genom-Datenspeichern wie GEO, ArrayExpress, SRA sorgfältig überprüft und die Informationen der zugehörigen Studien annotiert. Zur Vergleichbarkeit der gesammelten Daten wurde die Qualität des Proben Sets aus den Microarrays gründlich kontrolliert und deren Ergebnisse wurden gemäß eigener Methoden normalisiert.

Genevestigator kommt in folgenden drei Fragestellungen zur Anwendung. Zum einen, kann das „*Condition search toolset*“ eingesetzt werden, um bestimmte Versuchsbedingungen herauszufinden, welche die Expression von Zielgenen regulieren. Zum anderen kann mit Hilfe des „*Gene search toolset*“ eine Reihe von Genen identifiziert werden, die unter gegebenen Versuchsbedingungen stabil exprimiert werden und daher als Referenzgene dienen könnten. Des Weiteren erlaubt das „*Similarity search toolset*“, bei einer gegebenen Versuchsbedingung Expressionsprofile verschiedener Gene zu untersuchen und so koexprimierte Gene zu identifizieren.



**Abb. 2-12 Benutzerschnittstelle von Genevestigator.** (1) Home-Button, (2) "Sample selection"-Panel zur Auswahl der Microarray-Proben-Sets für gegebene Experimentalbedingungen, (3) "Gene selection"-Panel zur Auswahl der Target- bzw. Referenzgene, deren Expressionsstabilität mit der durch Genevestigator vorgeschlagenen Gene zu vergleichen ist, (4) Tabs für Wechsel zwischen geöffneten Tools (5) Panel für Toolsets. *Abbildung aus Genevestigator User Guide.*<sup>65</sup>

Das Analysetool *RefGenes* gehört zum „Gene search toolset“. Dieses Programm analysiert die Microarray-Daten mit den vom Experimentator vorgegebenen Experimentalbedingungen, um stabil exprimierte Referenzgene zur gewählten Bedingung zu identifizieren. Dabei berechnet *RefGenes* die Standardabweichung (SD) des Expressionslevels (log<sub>2</sub>-skalierte Signalintensität aus Microarray-Chips) von genomweit in Frage kommenden Genen. Anhand dieser SD werden die 20 am stabilsten exprimierten Gene mit dieser spezifischen Experimentalbedingung angezeigt. Da die Suche nach stabilen Referenzgenen in *RefGenes* genomweit erfolgt, werden dabei nicht nur die bereits herkömmlich verwendeten Referenzgene gefunden, sondern auch neue Gene als Referenzgene vorgeschlagen. Die durch *RefGenes* vorgeschlagenen Referenzgene waren laut Evaluierung von Hruz et al. den konventionell verwendeten Referenzgenen als interne Kontrolle weit überlegen, vor allem in den Experimentalbedingungen, die zur Suche nach zuverlässigen Referenzgenen vor Beginn der *RefGenes*-Analyse vorgegeben waren.<sup>66</sup>



**Abb. 2-13 RefGene-Analysenergebnis.** In diesem Beispielergebnis sind die 20 besten Referenzgene für eine vorgegebene Experimentalbedingung dargestellt. Als Parameter für Expressionslevel sowie -stabilität werden jeweils mediane Signalintensität und Standardabweichung herangezogen und in einem Box-Whisker-Plot dargestellt. Der Expressionsbereich, innerhalb dessen zuverlässige Referenzgene gesucht werden sollen, kann dabei vorgegeben werden (im Beispiel 15,8-17,8 gewählt). Außer dieser Funktion zur Suche nach stabil exprimierten Referenzgenen kann das Expressionsprofil beliebig auswählbarer Gene in der gleichen Experimentalbedingung ermittelt werden. Hier wurde zum Vergleich das Expressionsprofil von *Gapdh*, ein bekanntes, konventionell verwendetes Referenzgen, ausgewählt (roter Balken oben). Dieses Gen zeigt ein hohes Expressionslevel, jedoch ist seine SD im Vergleich zu darunter stehenden, neu identifizierten Referenzgenkandidaten zu hoch, um als stabiles Referenzgen für diese Experimentalbedingung in Betracht zu kommen. *Screenshot aus Genevestigator.*

Es ist empfehlenswert, dass Referenzgene außer ihrer stabilen Genexpression auch ein mit Zielgenen vergleichbares Expressionsniveau aufweisen, damit Target- und Referenzgene mit gleicher Effizienz vervielfältigt werden können. Dadurch erhöht sich die Sensitivität, auch geringe Veränderungen der Genexpression zu erfassen.<sup>11,24,50,67,68</sup> Wie in einer genomweiten Evaluierungsstudie gezeigt wurde, sind konventionell verwendete Referenzgene (sog. Housekeeping-Gene) oft viel stärker als die interessierenden (Ziel-)Gene exprimiert. Wenn solche stark exprimierten Gene als interne Kontrolle eingesetzt werden, können die PCR-Reaktionen für Target- und Referenzgene stark voneinander abweichende Reaktionseffizienzen aufweisen, was letztendlich zu einem signifikanten Bias bei der relativen Quantifizierung führen kann.<sup>66</sup> In der RefGenes-Analyse kann man auch Expressionslevel für die zu suchenden Referenzgene vorgeben,

sodass geeignete Referenzgene mit zu den Zielgenen vergleichbaren Expressionsleveln identifiziert werden können (**Abb. 2-13**).

### 2.3 GeNorm und Normfinder

Zur Identifizierung stabil exprimierter Referenzgene wurden in der vorliegenden Arbeit die zwei gängigsten Analysealgorithmen, GeNorm und Normfinder, eingesetzt, um Rohdaten aus RT-qPCR-Experimenten auszuwerten. Bei beiden Analysealgorithmen entsprechen stabil exprimierte Gene denjenigen mit niedrigen Stabilitätswerten, bei GeNorm als M-Wert und bei Normfinder als SV-Wert (*stability value*) bezeichnet. Beiden Analysemethoden liegen unterschiedliche mathematische Vorgehensweisen zugrunde.

Der GeNorm-Algorithmus stützt sich auf die Logik, dass das Verhältnis der Genexpression zweier idealer Referenzgene immer gleich bleiben sollte, unabhängig von untersuchten RNA-Proben, Experimentalbedingungen und Zelltypen, denn konstante Genexpression ist die Grundvoraussetzung für ideale Referenzgene. Jegliche Schwankungen dieses Expressionsverhältnisses zweier Gene weist darauf hin, dass eins oder beide Gene nicht konstant exprimiert sind.<sup>3</sup> Beruhend auf diesem Prinzip ermittelt die GeNorm-Analyse den sog. M-Wert, der hier als Parameter für die Genstabilität dient. Je kleiner der M-Wert ist, desto stabiler exprimiert ist das betreffende Gen. Zur Berechnung dieses M-Wertes wurde in GeNorm ein paarweiser Vergleich der Genexpression (*pairwise comparison of gene expression*) eingeführt, in dem das oben erwähnte Prinzip verkörpert ist. Im paarweisen Vergleich wird jedes Kandidatengen mit allen anderen Kandidatengenen hinsichtlich der Genexpression 1:1 verglichen. Dabei wird für ein bestimmtes Paar zweier Kandidatengenen, z.B. Gen A und B, das Verhältnis ihrer log<sub>2</sub>-skalierten Expressionslevel in allen Gewebeproben bestimmt. Die Standardabweichung von allen sich ergebenden Expressionsverhältnissen wird als paarweise Variation für das Genpaar A-B definiert. Die Berechnung der paarweisen Variationen wird für alle anderen Genpaare in gleicher Weise fortgesetzt (A-C, A-D,...). Der M-Wert für das Gen A entspricht letztendlich dem Mittelwert aller ermittelten paarweisen Variationen bezogen auf das Gen A.

Auf diese Weise werden M-Werte von allen Kandidatengenen berechnet und immer das eine Kandidatengen mit dem höchsten M-Wert (d.h. mit der niedrigsten Genstabilität) wird von der weiteren Analyserunde ausgeschlossen. Die Analysen werden so lange fortgesetzt, bis letztendlich zwei Kandidatengene, beide mit dem niedrigsten M-Wert, übrigbleiben. Diese zwei Gene entsprechen den besten Referenzgene von GeNorm.

Die Abschätzung der Genexpressionsstabilität im Normfinder-Algorithmus erfolgt anhand eines mathematischen Modells, in dem sowohl Intragruppen-, als auch Intergruppenvariationen berücksichtigt sind (Aufgrund von Komplexität dieses mathematischen Modells wurde hier auf

dessen Darstellung verzichtet. im Detail siehe Andersen et al.<sup>5</sup>). Diese zwei Arten der Variation in Genexpressionsdaten werden dann in einen Stabilitätswert (SV-Wert) zusammengeführt und, wie bei GeNorm, gilt die Aussage, je kleiner der SV-Wert, desto stabiler ist das betreffende Gen exprimiert.

Mehrere Studien empfehlen, für akkurate relative Quantifizierung mehr als ein Referenzgen einzusetzen. Korrespondierend dazu berechnen GeNorm und Normfinder die beste Anzahl der simultan einzusetzenden Referenzgene. GeNorm durch Ermittlung des geometrischen Mittelwertes der relativen Genexpression der Kandidatengene und Normfinder durch Berechnung der sog. akkumulierten SD von Kandidatengenen. Aus beiden Analysealgorithmen ergab sich, dass ein simultaner Einsatz von zwei besten Referenzgenen bereits zielführend ist, denn dadurch wird die größte Verbesserung der Referenzgenleistung bereits erreicht.<sup>69</sup> Eine Anwendung von einer größeren Anzahl der Referenzgene ist aus Zeit- und Kostengründen praktisch unvertretbar.

Die zwei vorgestellten Analysealgorithmen besitzen jeweils eigene Stärken und Schwächen bei der Identifizierung stabil exprimierter Referenzgene aufgrund ihrer unterschiedlichen statistisch-mathematischen Vorgehensweise. So ist z.B. GeNorm nicht zur Analyse ko-regulierter Genen (z.B. gleiche Funktionsgruppe) geeignet, weil diese Gene mit immer gleichbleibenden Expressionsverhältnissen fälschlicherweise als stabil exprimierte Referenzgene identifiziert werden würden. Liegen mehrere Behandlungsgruppen vor, ist Normfinder besser geeignet, weil dabei nicht nur Intra- sondern auch Intergruppenvariationen der Genexpression zur Analyse der zuverlässigen Referenzgene berücksichtigt werden. Dieses Feature von Normfinder war in der vorliegenden Arbeit mit mehreren vergleichenden Experimentalgruppen besonders vorteilhaft und deswegen könnte man auf die Ergebnisse aus Normfinder mehr Bedeutung als die von GeNorm beilegen. Als ein untergeordneter Nachteil des Normfinder-Algorithmus wäre zu erwähnen, dass er für Validität der Ergebnisse eine minimale Anzahl der Kandidatengene und Proben pro Gruppe benötigt. So sollten die zu analysierenden Kandidatengene mindestens drei und Gewebeproben pro Gruppe mindestens acht sein, wobei je mehr zu bestimmende Gene oder Gewebeproben sind, desto aussagekräftiger sind die Ergebnisse.<sup>70</sup>

## 3 Methoden und Materialien

### 3.1 Vorauswahl potenzieller Referenzgenkandidaten

Da es aufgrund der erheblichen Anzahl der im Mausgenom vorhandenen Gene (ca. 30 000 Protein-kodierende Gene) praktisch unmöglich wäre, alle einzelnen Gene auf ihre Eignung als Referenzgen in einer bestimmten experimentellen Situation zu überprüfen (zeitliche und Kostengründe), haben wir im ersten Schritt eine Vorauswahl potenzieller Referenzgenkandidaten aus den vier folgenden Quellen getroffen:

- a) Studien mit Einsatz von konventionell verwendeten Referenzgenen,
- b) Evaluierungsstudien, die stabil exprimierte Referenzgene untersuchen,
- c) Web-basierte Analysesoftware für publizierte Microarray-Daten (Genevestigator),
- d) Kommerzielles Referenzgenpanel (Roche).

#### 3.1.1 Datenlage hinsichtlich der Referenzgene für die Mauslunge

Um einen Überblick über die aktuelle Datenlage hinsichtlich bisher verwendeter Referenzgene für neonatale Mauslungen zu erhalten und herauszufinden, haben wir zuerst die frei zugängliche Internet-Datenbank PubMed (<https://pubmed.ncbi.nlm.nih.gov/>) nach Studien durchsucht, die Referenzgene für dieses Gewebe in RT-qPCR-Experimenten eingesetzt oder evaluiert haben. Dabei wurden Suchbegriffe wie „reference gene, murine lung, C57BL/6, RT-qPCR, hyperoxia, neonatal“ einzeln oder in Kombination verwendet. Die so identifizierten Studien ließen sich in zwei Gruppen kategorisieren. Zur ersten Gruppe gehören Studien, in denen eine RT-qPCR als experimentelle Methode durchgeführt wurde und herkömmlich verwendete Referenzgene im Rahmen der relativen Quantifizierung zum Einsatz kamen. Die zweite Gruppe umfasst Studien, die die Zuverlässigkeit von Referenzgenen in murinen Geweben evaluierten und dabei auch neue zuverlässigere Referenzgene als die herkömmlich verwendeten vorschlugen.

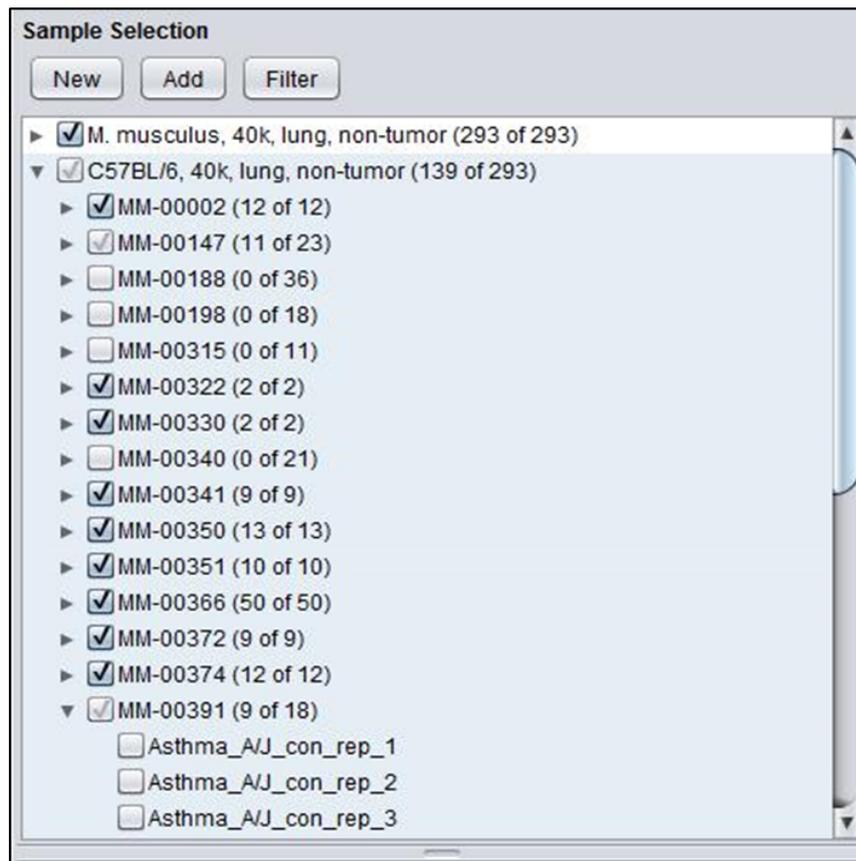
#### 3.1.2 Genevestigator

Zunächst wurden von der hohen Anzahl in Genevestigator verfügbarer Microarray-Daten nur diejenigen Microarray-Experimente herausgefiltert, die in den Lungengeweben von C57BL/6-Mäusen durchgeführt wurden. Es gibt zwei Wahlmöglichkeiten für die Microarray-Plattform, nämlich Affymetrix Mouse Genom 430 2.0 Array und U74Av2 Array. Diese unterschiedlichen Plattformen unterscheiden sich voneinander durch Anzahl der in einem Lauf nachweisbaren Transkripte (40.000 Transkripte für die erstere Plattform und 12.000 für die letztere) sowie Erscheinungsdatum der Plattform im Markt (Das Mouse Genom 430 2.0 Array ist die neue

Version). Da es im Mouse Genom U74Av2 Array keine Microarray-Proben gab, die zu unserem experimentellen Setting (Lunge, C57BL/6-Mäuse) passen, wurde in unserer vorliegenden Arbeit nur das Mouse Genom 430 3.0 Array berücksichtigt.

Aus der Suche mit den Bedingungen (Lungengewebe von C57BL/6-Mäusen, Affymetrix Mouse Genom 430 2.0 Array) ergaben sich letztendlich 139 Microarray-Proben aus 11 Microarray-Experimenten (10 Studien), in denen die Proben mit unterschiedlichen Genotypen z.B. *Cftr*<sup>-/-</sup>, *Gadd45a*<sup>-/-</sup>, *Pbef*<sup>-/-</sup> sowie diversen experimentellen Behandlungen z.B. Exposition gegenüber Zigarettenrauch, Sensibilisierung gegen Ovalbumin, Ventilations-induzierte Lungenschädigung usw. eingeschlossen sind (**Abb. 3-1**).

Mit diesen Microarray-Proben-Sets (n=139) führte Genevestigator die Analyse aus und zeigte die besten Referenzgene mit Angabe der Signalintensität für Genexpression (log<sub>2</sub>-skaliert) und deren Standardabweichung (vgl. **Abb. 2-13**). Die Analyse wurde in drei Gruppen mit unterschiedlich definiertem Genexpressionsniveau durchgeführt, nämlich in Hoch-, Mittel- und Niedrig-Gruppe, deren log<sub>2</sub>-skalierte Signalintensität jeweils >13,00 und 10,00-13,00 sowie <10,00 betrug. Im Endeffekt ergaben sich aus diesem Schritt insgesamt 60 Referenzgenevorschläge (je 20 Referenzgene für die drei Analysegruppen mit unterschiedlichem Expressionsniveau) und somit wurde das gesamte Expressionsniveau gleichmäßig berücksichtigt.



**Abb. 3-1** Suchergebnis von Genevestigator mit Anzahl der gefundenen Microarray-Proben. In der Abbildung sind die von Genevestigator gefundenen Studien (MM-Nummer) sowie die dazugehörigen Microarray-Proben zu erkennen, die zu der vorgegebenen Experimentalbedingung (C57BL/6, Lunge, 40k-Microarray, Non-Tumor) passen. Außerdem sind die annotierten Informationen der jeweiligen Microarray-Proben ersichtlich. Beispielsweise für die Studie MM-00391 sieht man 9 von insgesamt 18 Microarray-Proben, die zur Analyse ausgewählt sind. In der eingeblendeten Liste sind drei Proben aus A/J-Mäusen erkennbar, die bei fehlenden Hacken nicht zur Analyse eingeschlossen sind, weil sie für unser experimentelles Setting irrelevant sind. *Screenshot aus Genevestigator.*

### 3.1.3 Kommerzielles Referenzgenpanel (Roche)

Bei der Vorauswahl der Referenzgenkandidaten wurden die Gene aus dem Referenzgenpanel von Roche berücksichtigt, die aufgrund ihres häufigen Einsatzes kommerziell etabliert wurden.

## 3.2 Microarray-Transkriptom-Daten

Microarray-Experimente wurden von unserem Kooperationspartner Universitätsklinikum Gießen durchgeführt (Tierversuchsantrag-Nr. 117-10) und die Daten daraus wurden zur weiteren Interpretation im vorliegenden Projekt bereitgestellt. Im Folgenden ist die Durchführung der Microarray-Experimente kurz erläutert.

Lungengewebe wurde von 5-7 Tage alten C57BL/6-Mäusen gewonnen, die unter Normoxie-Bedingung ( $FiO_2=0,21$ ) spontan atmeten. Es gab drei biologische Replikate und jedes biologische

Replikate wurden in zwei technischen Replikaten untersucht. Die entnommenen Gewebe wurden direkt in flüssigem Stickstoff eingefroren und aufbewahrt. Extraktion der Gesamt-RNA erfolgte mittels miRNeasy Mini Kit (Qiagen) nach Homogenisierung der Lungengewebe gemäß der Anleitung des Herstellers. Die extrahierte Gesamt-RNA wurde einer cRNA-Synthese, -fragmentierung und endlich einer Hybridisierung auf CodeLink 10K Mouse Array mittels CodeLink Expression Assay Kit (GE Healthcare) laut der Anleitung des Herstellers unterzogen.

Die Signale aus Punkten (*spots*) auf Microarray-Images wurden mittels CodeLink System Software 5.0.0.31312 quantifiziert, die lokal hintergrundkorrigierte Roh- sowie medianzentrierte, innerhalb der Objektträger normalisierte Daten generierte. Für weitere Analysen wurden nur die letzteren Daten bestehend aus 10 181 Proben Sets und sechs Gewebeproben angewendet. Die durch die Microarray-Proben-Sets repräsentierten Gene wurden mit Hilfe von BioLite Package (Bioconductor) mit der Bibliothek „mwgcod.db“ für CodeLink Mouse Whole Genome arrays annotiert. Eine Qualitätskontrolle der gewonnenen Expressionsdaten wurde mittels eines automatisierten Workflows durchgeführt. Dieser Prozess beinhaltet Rausnahme der Kontrollgene, derjenigen Gene mit niedrigen *Flag*-Werten sowie der Proben Sets mit hohem Anteil ( $\geq 50\%$ ) an fehlenden Messwerten. Nach der Qualitätskontrolle blieben insgesamt 3 651 Microarray-Proben Sets. Für Replikate wurde am Ende jeweils deren Mittelwert genommen. Die Datenlücken wurden durch sequenzielle-KNN-Imputation (*k-nearest neighbor algorithm*) in Anwendung von SeqKnn Vers. 1.0 ersetzt. Die ersetzten Datensets wurden mittels normalizeQuantiles von Limma Package quantilnormalisiert. Die endgültigen Daten wurden zum Schluss log<sub>2</sub>-skaliert dargestellt.

Alle Microarray-Daten wurden in Gene Expression Omnibus von NCBI (<https://www.ncbi.nlm.nih.gov/geo/>) hochgeladen und sind unter GEO Series Accession Number von GSE189505 zugänglich.

### 3.3 Versuchstier und RNA-Extraktion aus den Lungengeweben

Die Gesamt-RNA aus den murinen Lungengeweben für RT-qPCR-Validierung wurde von unserem Kooperationspartner im Max-Planck-Institut für Herz- und Lungenforschung, Bad Nauheim (AG Prof. Morty, Tierversuchsantrag-Nr. B2/277) bereitgestellt. Alle Lungengewebe stammen aus C57BL/6J-Mäusen, die mit zeitlich geplanter Trächtigkeit (Paarung von Weibchen im gleichen Zyklusstadium) gezüchtet wurden, um Versuchstiere im gleichen Alter zu erhalten. Vor der Resektion wurden die Lungen zum Auswaschen des Blutes mit PBS (phosphate buffered saline) perfundiert und das Geschlecht der Tiere wurde durch Nachweis des Y-Chromosoms mittels RT-qPCR bestätigt. Die RNA-Extraktion erfolgte mittels peqGOLD Total RNA Kit von Peqlab mit homogenisierten Lungengeweben aus zwei Experimentalgruppen, nämlich Normoxie-Gruppe ( $\text{FiO}_2=0,21$ ) und der Hyperoxie ausgesetzte Gruppe ( $\text{FiO}_2=0,8$ ).

Die extrahierten RNA-Proben wurden in zwei Lieferungen von unserem Kooperationspartner zu uns eingesandt. Die RNA-Qualität der zuerst eingetroffenen Gruppe (Gruppe 1) mit 76 RNA-Proben mittels Agilent 2100 Bioanalyzer (Abschnitt 3.4) bestimmt und anschließend wurden jeweils sechs RNA-Proben (Tab.A1) mit der besten RNA-Qualität aus allen Alters- und Behandlungsgruppen in die RT-qPCR-Validierung zugeführt (n=60). Von der später eingetroffenen Gruppe (Gruppe 2) mit 28 RNA-Proben wurden alle RNA-Proben außer zwei mit zu niedriger RNA-Konzentration nach RNA-Qualitätskontrolle in die RT-qPCR-Validierung zugeführt (n=26).

**Tab. 3-1 Anzahl der verfügbaren RNA-Proben.** Die RNA-Proben aus Mauslungen mit unterschiedlichen Alters- und Experimentalgruppen wurden von unserem Kooperationspartner Universitätsklinikum Gießen bereitgestellt. Die Gruppe 2 wurde zur Erhöhung der Gesamtanzahl der RNA-Proben zu einem späteren Zeitpunkt ergänzt. Von der Gruppe 1 wurden jeweils sechs RNA-Proben mit der besten RNA-Qualität in den Alters- und Behandlungsgruppen und von der Gruppe 2 alle RNA-Proben außer zwei mit zu niedriger RNA-Konzentration zur RT-qPCR-Validierung ausgewählt. Die Anzahl der verfügbaren RNA-Proben in Klammern und davon die der für RT-qPCR-Validierung ausgewählten Proben ohne Klammern.

		PND2.5	PND5.5	PND14.5	PND21.5	PND28.5
Gruppe 1	Normoxie	6 (8)	6 (9)	6 (10)	6 (6)	6 (6)
	Hyperoxie	6 (9)	6 (8)	6 (8)	6 (6)	6 (6)
Gruppe 2	Normoxie	-	2 (2)	6 (6)	6 (6)	6 (6)
	Hyperoxie	-	- (2)	2 (2)	4 (4)	-

### 3.4 RNA-Qualitätskontrolle

Die RNA-Qualitätskontrolle wurde mittels Agilent 2100 Bioanalyzer (Agilent Technologies) unter Anwendung des Agilent RNA 6000 Nano Kits (Agilent Technologies) durchgeführt. Alle eventuell gefrorenen Reagenzien wurden vor 30 Minuten der Anwendung in der Raumtemperatur aufgetaut. Zur Vorbereitung des Gels wurden zuerst 550µl der Gel-Matrix in einem Spin-Filter mittels Mikrozentrifuge 10 Minuten zentrifugiert (ca. 1500g), um eventuell beigemischte Partikel zu beseitigen. 65µl der Aliquote der zentrifugierten Gel-Matrix wurde zur weiteren Nutzung in einem RNase-freien Eppendorf-Röhrchen aufbewahrt.

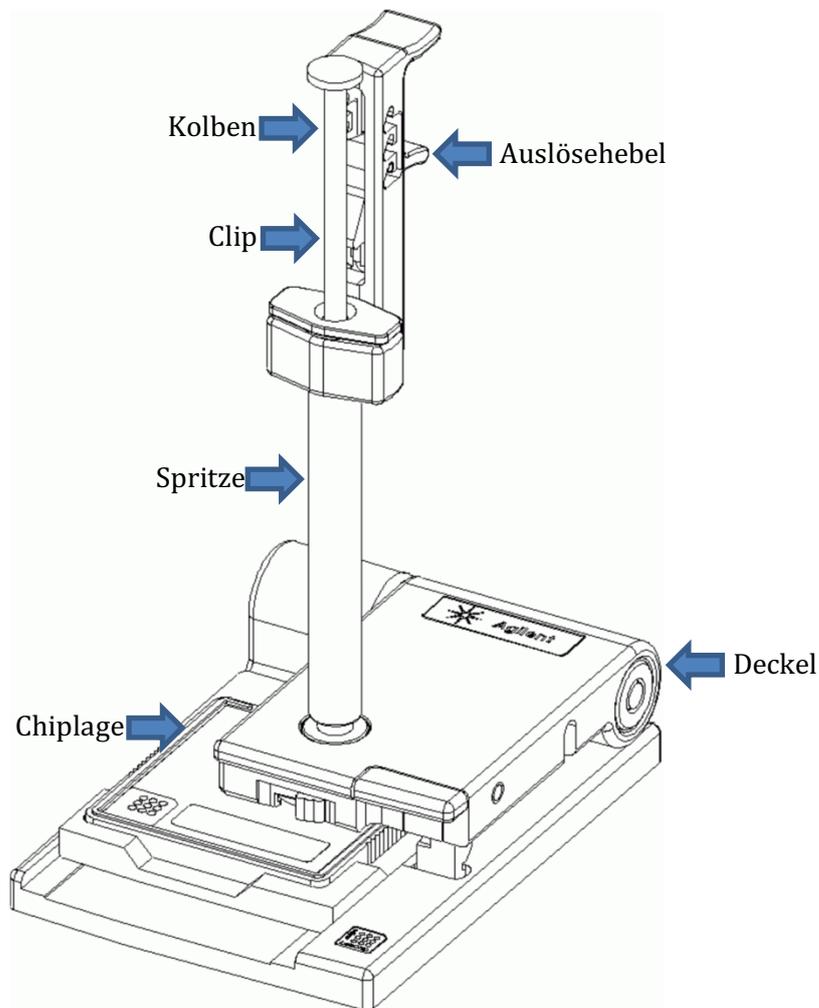
Im nächsten Schritt soll ein Gel-Farbstoff-Mix hergestellt werden. Dafür wurde zunächst das Farbstoffkonzentrat 10 Sekunden mit einem Vortexgerät gemischt und davon wurde 1µl in die 65µl Aliquote der filtrierten Gel-Matrix zugesetzt. Der erstellte Gel-Farbstoff-Mix wurde kurz gevortext und bei 13.000g für 10 Minuten abzentrifugiert.

Danach folgte eine Beladung des Analysechips und begann mit Reinpipettieren des 9,0µl Gel-Farbstoff-Mixes ins Probengefäß mit der Markierung „G“ (Abb. 3-2). Der zugegebene Gel-Farbstoff-Mix soll die Mikrokanäle im Chip komplett lückenlos auffüllen und zu diesem Zweck kommt die Chip-Priming-Station zum Einsatz.



**Abb. 3-2 Agilent RNA 6000 Nano Chip.** Es sind insgesamt 16 Reaktionsgefäße in einem Chip verfügbar, von denen 12 Reaktionsgefäße für Probenanalyse vorgesehen sind. In vier Reaktionsgefäßen werden der Gel-Farbstoff-Mix sowie der Leiter zugesetzt (siehe Markierung auf dem Chip). *Abbildung aus Agilent RNA 6000 Nano Kit Guide.*

Die Chip-Priming-Station besteht aus einer Grundplatte, in die der Chip eingelegt wird, dem Deckel und der diesem Deckel anschließenden Spritze. Vor ihrer Anwendung im Schritt der Chip-Beladung muss die Dichtung der Verbindung zwischen Chip und Spritze entsprechend der Anweisung kontrolliert werden. Dafür wurde der Kolben der Spritze bis 1,0ml Position aufgezogen und ein leerer Chip wurde auf die Grundplatte gelegt. Nach dem Zuschließen des Deckels wurde der Kolben runterdrückt, bis dessen hinterer Teil unter den Klipp einrastet. Nach Abwarten von 5 Sekunden wurde der Auslösehebel hochgezogen, damit der Kolben der Spritze wieder freigegeben wird. Wenn dabei der Kolben innerhalb einer Sekunde bis 0,3ml zurückspringt, ist die Abdichtung gesichert.



**Abb. 3-3 Chip-Priming-Station.** Alle Mikrokanäle im Chip müssen vor Analyse mit Gel als molekulares Sieb aufgefüllt werden, was durch die Chip-Priming-Station mit Hilfe ihrer Druck-ausübenden Spritze. *Abbildung aus Agilent Using Your Chip Priming Station.*

Anschließend wurden die Chips für Analysen beladen. Dafür wurde nach Zugabe des 9,0µl Gel-Farbstoff-Mixes die Chip-Priming-Station zugeschlossen und der in diesem Zeitpunkt bei 1ml befindliche Kolben wurde runtergedrückt, bis er durch den Klipp festgehalten wird. Nach Abwarten von 30 Sekunden wurde der Kolben wieder freigegeben. Nach Kontrolle, dass der Kolben dabei mindestens bis 0,3ml Markierung zurücktritt wurden nochmal 5 Sekunden abgewartet. Danach wurde der Kolben langsam bis 1ml zurückgezogen. Die Chip-Priming-Station geöffnet und den beiden restlichen Probengefäßen mit „G“-Markierung jeweils 9,0µl Gel-Farbstoff-Mix zugesetzt.

Anschließend wurde der Fluoreszenz-Marker jeweils 5µl in die 12 Probengefäße sowie ins Gefäß für die Leiter zugegeben. Dann wurde 1µl der RNA-Leiter ins Gefäß mit entsprechendem Zeichen und jeweils 1µl der maximal 12 RNA-Proben wurden in die Probengefäße pipettiert. Zur Auflösung der möglichen sekundären RNA-Struktur wurde hier die RNA-Probe bei 70°C 2

Minuten lang erhitzt. Am Schluss wurde der vorbereitete Chip im Chip-Vortexer bei 2400rpm für eine Minute vortext und darauffolgend wurde die Analyse im Agilent 2100 Bioanalyser gestartet.

### 3.5 Reverse Transkription

**Tab. 3-2 Materialien mit Herstellerangaben.**

Mastercycler® Instrument	Eppendorf
Transcriptor First Strand cDNA-Synthese Kit	Roche Applied Science
- Reverse Transkriptase	
- Reaktionspuffer (Tris/HCl, KCl, MgCl <sub>2</sub> )	
- RNase Inhibitor	
- dNTP Mix	
- verankerter Oligo(dT) <sub>18</sub> -Primer	
- Random Hexamer Primer	
- PCR Grade Wasser	

Die reverse Transkription wurde mittels Transcriptor First Strand cDNA-Synthese Kit unter Einhaltung der Herstelleranweisung durchgeführt. Die in diesem Kit angewendete reverse Transkriptase (RTase) ist eine rekombinante RTase, die in *E. coli* exprimiert wird. Dieses Enzym beinhaltet folgende Enzymaktivitäten: RNA-abhängige DNA-Polymerase, DNA-abhängige DNA-Polymerase, unwinding-Aktivität und RNase-H-Aktivität.

**Tab. 3-3 Template-Primer-Mix.**

Komponente	Finale Konzentration	Volumen
Gesamt-RNA	1µg Gesamt-RNA	variabel
Random Hexamer Primer, 600pmol/µl	60µM	2µl
PCR Grade Wasser	zu einem Endvolumen =13µl	variabel
Endvolumen		13µl

Alle gefrorenen Reagenzien wurden vor der Anwendung in Raumtemperatur aufgetaut, zentrifugiert und anschließend in ein Eisbad bereitgestellt. Zunächst wurde der Template-Primer-Mix zum Endvolumen von 13µl hergestellt (**Tab. 3-3**). Anschließend wurde der Template-Primer-Mix zur Denaturierung der sekundären RNA-Struktur bei 65°C für 10 Minuten erhitzt und unmittelbar danach im Eisbad abgekühlt. Darauffolgend wurde ein RT-Mastermix hergestellt, indem die in der untenstehenden Tabelle aufgelisteten Reagenzien in der Reihenfolge zugegeben und durch vorsichtiges Auf- und Abpipettieren gut vermischt wurden. Zur Ansammlung der Probe auf dem Boden des Röhrchens wurde der erstellte RT-Mastermix ohne Vortexen kurz zentrifugiert.

**Tab. 3-4 RT-Mastermix.**

Komponente	Finale Konzentration	Volumen
Reaktionspuffer, 5x konzentriert.	1x (8mM MgCl <sub>2</sub> )	4µl
RNase Inhibitor, 40U/µl	20U	0.5µl
DNTP Mix, 10mM each	1mM each	2µl
Reverse Transkriptase, 20U/µl	10U	0.5µl
Endvolumen		7µl

Dieser RT-Mastermix wurde dann in den bereitgestellten Template-Primer-Mix zum Endvolumen von 20µl zugesetzt und durch Pipettieren vorsichtig durchgemischt und nachfolgend abzentrifugiert. Das Reaktionsgefäß mit RNA-Proben und allen Reagenzien wurde dann in den Thermocycler (Mastercycler Eppendorf) hineingestellt und die reverse Transkription wurde gestartet. Das Inkubationsprogramm ist je nach verwendeten Primern und Länge der Target-mRNA individuell eingestellt. Für die vorliegende reverse Transkription mit Random-Hexamer-Primern von 600 pmol/µl sowie Target-mRNAs bis zu 4kb wurde die folgende Programmierung ausgewählt.

- 10 Minuten bei 25°C
- 30 Minuten bei 55°C
- 5 Minuten bei 85°C (Zur Inaktivierung der reversen Transkriptase)

Durch Überführung des Reaktionsgefäßes ins Eisbad wurde die RT beendet. Darauf folgend wurde die hergestellte cDNA mit sterilem Wasser im Verhältnis 1:10 verdünnt und bei -20°C zur späteren Benutzung bewahrt.

### 3.6 Quantitative Echtzeit-PCR

**Tab. 3-5 Materialien mit Herstellerangaben.**

LightCycler® 96 Instrument	Roche Applied Science
FastStart Essential DNA Probes Master	Roche Applied Science
- Taq-DNA-Polymerase	
- Reaktionspuffer	
- dNTP	
- MgCl <sub>2</sub>	
- PCR-Grade-Wasser	
RealTime Ready Single Assays für 24 Referenzgenkandidaten	Roche Applied Science
LightCycler® 480 Multiwell Plates 96	Roche Applied Science
LightCycler® 480 Sealing Foils	Roche Applied Science

Die quantitative Echtzeit-PCR wurde mittels FastStart Essential DNA Probes Master mit Hydrolyse- (TaqMan-) Sonden im LightCycler® 96 Instrument durchgeführt. Für die Assays der jeweiligen Referenzgenkandidaten wurden RealTime Ready Single Assays angewendet (Tab. 3-5). Alle Reagenzien wurden vor Anwendung in Raumtemperatur aufgetaut und in einer Mikrozentrifuge kurz abzentrifugiert. Nach vorsichtigem Durchmischen durch Auf- und Abpipettieren wurden die Reagenzien in einem Eisbad bis zur weiteren Verwendung aufbewahrt. Zunächst wurde ein PCR-Mastermix vorbereitet, indem die in der

Tab. 3-6 aufgelisteten Reagenzien in der angegebenen Reihenfolge in ein Reaktionsgefäß gegeben und durch Auf- und Abpipettieren vorsichtig durchgemischt wurden. Dabei wurde ein Master-Mix für eine oder mehrere Platten angesetzt. Der Primer-Probe-Mix enthielt die für das jeweilige Gen spezifischen Primer- und SONDENSEQUENZEN und variierte entsprechend abhängig von den zu bestimmenden Referenzgenkandidaten (Tab. 4-6).

**Tab. 3-6 Herstellung von PCR-Mastermix.**

Komponente	Volumen
PCR Grade Wasser	1.5µl
Primer Probe Mix (10x konzentriert)	1µl
Master Mix (2x konzentriert)	5µl
Endvolumen	7.5µl

Als nächstes wurden jeweils 7,5 µl des vorbereiteten PCR-Mastermixes in die Vertiefungen (Wells) einer LightCycler® 480 Multiwell Plate 96 entsprechend dem Plattenlayout (

Tab. 3-8) vorgelegt. Danach wurden jeweils 2,5µl der zu untersuchenden, 1:10 verdünnten cDNA-Proben aus Lungengewebe zugesetzt, die wie im Abschnitt 3.5 ausgeführt durch reverse

Transkription gewonnen worden waren. Für die Non-Template-Kontrolle (NTC) wurde eine entsprechende Menge an PCR Grade Wasser (Roche Applied Science) anstelle der cDNA-Proben zugegeben. In die Reaktionsgefäße für Interrun-Kalibrator (IRC) wurde die vorab hergestellte, universelle cDNA-Probe (Details siehe Abschnitt 3.7) zugesetzt und hier wurde über alle Platten hinweg das Expressionsniveau von *Tbp* bestimmt. Die so vorbereiteten RT-qPCR-Platten wurden mit einer Verschlussfolie zum Schutz der Proben vor Verdunstung versehen.

**Tab. 3-7 Programm für die quantitative Echtzeit-PCR.**

	Temperatur (°C)	Rampe (°C/s)	Dauer (s)
Vorinkubation	95	4,4	600
2-Schritt-Amplifikation (45x)	95	4,4	10
	60	2,2	30
Signalakquisition	72	-	1
Kühlung	40	2,2	30

Kurz vor RT-qPCR-Start wurde die Platte bei 1500g für 2 Minuten zentrifugiert, um die Reagenzien zum Boden des Reaktionsgefäßes zu bringen. Nach Hineinstellen der Platte in den LightCycler® 96 (Roche Life Science) wurde die RT-qPCR mit dem oben tabellarisch dargestellten Programm gestartet (**Tab. 3-7**). Die Signalmessung bei jedem Amplifikationszyklus erfolgt 1 Sekunde lang bei 72°C. Nach dem Ende der gesamten Amplifikation wurde für 30 Sekunden auf 40°C gekühlt.

Die hier verwendete Taq-DNA-Polymerase ist an einer blockierenden chemischen Gruppe gebunden und so ist ihre Aktivität bis zu 75 °C gehemmt. Durch eine vorgeschaltete Vorinkubation (*pre-incubation*) im Rahmen einer Heißstart-PCR (*hot start PCR*) wird sie aktiviert. Da der Primer bei dieser hohen Temperatur nicht an die DNA hybridisieren kann und die DNA-Polymerase selbst noch inaktiv ist, so wird die Amplifikation unspezifischer Produkte ebenfalls vermieden.

Der Assay eines Referenzgenkandidaten erfolgte aufgrund von einer großen Anzahl der zu bestimmenden cDNA-Proben über drei RT-qPCR-Platten (Platte A, B, C). Die Platte A enthält die cDNA-Proben mit Nummer 1 bis 44 (PND2.5 Nox, PND2.5 Hyx, PND 5.5 Nox, PND 5.5 Hyx, PND 14.5 Nox) und die Platte B die mit Nummer 45 bis 76 (PND 14.5 Hyx, PND 21.5Nox, PND 21.5 Hyx, PND 28.5 Nox, PND 28.5 Hyx). In der Platte C wurden die später ergänzten cDNA-Proben mit Nummer 79 bis 104 (PND 5.5 Nox, PND 14.5 Hyx, PND 14.5 Nox, PND 21.5 Hyx, PND 21.5 Nox, PND 28.5 Nox) untersucht. Alle cDNA-Proben wurden als Duplikat eingesetzt. Jede Platte beinhaltet zusätzlich eine NTC und einen IRC jeweils in Triplikat.

Somit wurden für den vollständigen Assay eines Referenzgenkandidaten insgesamt 181 Reaktionen ausgeführt (jeweils 60 Reaktionen für Platte A und B + 52 Reaktionen für Platte C + 9 NTCs). Für *Tbp* wurden aufgrund dessen Anwendung als Interrun-Kalibrator zusätzlich 216

Reaktionen durchgeführt (3 IRC/Platte x 72 Platten insgesamt). Für die gesamten Assays aller Referenzgenkandidaten wurden 72 Platten (24 Referenzgenkandidaten x 3 Platten) eingesetzt.

**Tab. 3-8 Layout der RT-qPCR-Platten.** Für den vollständigen Assay eines Referenzgenkandidaten in gesamten cDNA-Proben wurden drei RT-qPCR-Platten (Platte A, B, C) verwendet. Die Nummern auf den Stellen der Reaktionsgefäße entsprechen den cDNA-Nummern und NTC sowie IRC, jeweils Nicht-Template-Kontrolle und Interrun-Kalibrator. Die cDNA-Nr. 79 und 80 wurden aus der Gruppe (PND5.5 Nox) gewonnen, cDNA 81 und 82 aus PND14.5 Hyx, cDNA 83-88 aus PND14.5 Nox, cDNA 89-92 aus PND21.5 Hyx, cDNA93-98 aus PND21.5 Nox, cDNA 99-104 aus PND28.5 Nox

Platte A	1	2	3	4	5	6	7	8	9	10	11	12	Gruppe
A	1	1	3	3	4	4	5	5	6	6	8	8	PND2.5 Nox
B	10	10	11	11	12	12	13	13	15	15	16	16	PND2.5 Hyx
C	18	18	19	19	20	20	22	22	25	25	26	26	PND5.5 Nox
D	27	27	28	28	29	29	30	30	31	31	32	32	PND5.5 Hyx
E	35	35	37	37	39	39	40	40	43	43	44	44	PND14.5 Nox
F	NTC	NTC	NTC										
G													
H										IRC	IRC	IRC	

Platte B	1	2	3	4	5	6	7	8	9	10	11	12	Gruppe
A	45	45	46	46	47	47	48	48	50	50	51	51	PND14.5 Hyx
B	53	53	54	54	55	55	56	56	57	57	58	58	PND21.5 Nox
C	59	59	60	60	61	61	62	62	63	63	64	64	PND21.5 Hyx
D	65	65	66	66	67	67	68	68	69	69	70	70	PND28.5 Nox
E	71	71	72	72	73	73	74	74	75	75	76	76	PND28.5 Hyx
F	NTC	NTC	NTC										
G													
H										IRC	IRC	IRC	

Platte C	1	2	3	4	5	6	7	8	9	10	11	12	Gruppe
A	79	79	80	80	81	81	82	82	83	83	84	84	
B	85	85	86	86	87	87	88	88	89	89	90	90	
C	91	91	92	92	93	93	94	94	95	95	96	96	
D	97	97	98	98	99	99	100	100	101	101	102	102	
E	103	103	104	104									
F	NTC	NTC	NTC										
G													
H										IRC	IRC	IRC	

### 3.7 Interrun-Kalibrierung

Da Genexpressionsanalysen heutzutage oft multiple, zu bestimmende Target-Gene und zahlreiche Gewebeproben beinhalten, ist es praktisch unvermeidlich, dass ein qPCR-Experiment in mehr als einem Lauf mit mehr als einer qPCR-Platte erfolgt. Dabei können technische (Lauf zu Lauf) Variationen entstehen, die zur akkuraten Quantifizierung der Target-Gene ausgeglichen werden müssen.<sup>71</sup>Zur Interrun-Kalibrierung der im Rahme der vorliegenden Arbeit durchgeführten RT-qPCR-Experimente wurde zunächst eine universelle RNA-Probe hergestellt, die als Interrun-Kalibrator (IRC) diente. Dafür wurden von den 10 Alters- und Behandlungsgruppen jeweils die RNA-Proben mit bester RNA-Qualität ausgewählt, die in einem Eppendorf-Gefäß durchgemischt wurden (je 13µl von insgesamt 31 Proben. Siehe Anhang **Tab.A1**). Der so vorbereitete IRC wurde nach reverser Transkription in jede einzusetzende RT-qPCR-Platte als Triplikat hinzugegeben und darin wurde die Expression von *Tbp* bestimmt. Da hier der IRC und das darin zu bestimmende Gen in allen RT-qPCR-Platten immer identisch bleiben, sollten die daraus resultierenden Cq-Werte nach RT-qPCR logischerweise unverändert bleiben. Jegliche Variation der gemessenen Cq-Werte für *Tbp* sollte durch unterschiedliche RT-qPCR-Läufe bedingt sein und muss entsprechend korrigiert werden. Dafür wurde die folgende Formel aus einer früheren Studie verwendet.<sup>72</sup>

$$Cq_{norm} = Cq_{GOI} - \frac{1}{m} \sum_{j=1}^m \left( Cq_{IPCj} - \frac{1}{n} \sum_{i=1}^n Cq_{IPCi} \right)$$

Dabei bezeichnet  $Cq_{norm}$  den interrun-kalibrierten (normalisierten) Cq-Wert,  $Cq_{GOI}$  den rohen Cq-Wert des zu detektierenden Gens (*gene of interest*),  $Cq_{IPCj}$  den Cq-Wert des j-sten IRC-Replikates in einer RT-qPCR-Platte (*IPC* steht für *interplate calibration*),  $m$  die Anzahl aller IRC-Replikate in einer RT-qPCR-Platte und  $n$  die gesamte Anzahl von IRCs aller RT-qPCR-Platten. In Worten beschrieben, entspricht der interrun-kalibrierte Cq-Wert dem rohen Cq-Wert, abgezogen mit der mittleren Abweichung der IRC-Replikate der betreffenden RT-qPCR-Platte vom arithmetischen Mittel der gesamten IRCs. Der konkrete Berechnungsprozess für das vorliegende Experiment siehe **Tab. 3-9**.

**Tab. 3-9 Berechnung des Kalibrierungsfaktors für die Interrun-Kalibrierung.** Zur Interrun-Kalibrierung wurden Cq-Werte für *Tbp* in jeder Platte als Triplikate vorliegenden IRC bestimmt (die erste Tabelle unten. Beachte, dass die Gennamen in der ersten Zeile nur für richtige Zuordnung der Platten stehen! Alle Cq-Werte sind für *Tbp*!) Anschließend wurde der Cq-Mittelwert der gesamten IRCs von allen RT-qPCR-Platten berechnet (Cq-Mittelwert 27,46). Dann wurde dieser Cq-Mittelwert von allen rohen Cq-Werten der IRCs abgezogen (Cq-Abweichung der IRCs). Daraufaufgehend wurde der Mittelwert aller Cq-Abweichungen in jeweiligen RT-qPCR-Platten, d.h. von drei Replikaten berechnet und die hieraus ergebenden Werte entsprechen sog. Kalibrierungsfaktor (CF). Schließlich wurde von jedem rohen Cq-Wert dieser platten- sowie gen-spezifischer CF abgezogen, um letztendlich interrunkalibrierte Cq-Werte zu erhalten.

**1. The Cq values of all IRCs**

		<i>Tbp</i>	<i>Rexo5</i>	<i>Auh</i>	<i>Csnk1a1</i>	<i>Dolpp1</i>	<i>Efhd2</i>	<i>Eif3a</i>	<i>Eif4g2</i>	<i>Elac2</i>	<i>Grn</i>	<i>Kdm3b</i>	<i>Maea</i>	<i>Mpv17l</i>	<i>Nupl2</i>	<i>Pcsk7</i>	<i>Puf60</i>	<i>Tjrc</i>	<i>Tmed2</i>	<i>Rpl4</i>	<i>Tnks2</i>	<i>Wars2</i>	<i>Zc3h11a</i>	<i>Psm4</i>
Plate A	Rep1	26,27	26,58	26,72	26,36	26,74	26,76	26,53	26,45	26,61	26,49	26,35	25,87	25,41	26,68	26,63	26,69	26,78	26,59	26,48	27,33	27,54	27,26	27,33
	Rep2	25,14	26,69	26,94	26,59	26,89	26,87	26,68	26,49	26,60	26,78	26,66	26,54	26,63	26,75	26,72	26,80	27,10	27,08	25,61	26,92	27,39	27,04	27,56
	Rep3	26,24	26,64	26,78	26,83	26,96	26,78	26,52	26,70	26,59	26,60	26,62	26,64	26,77	26,77	26,81	26,98	27,18	26,85	26,30	27,29	27,26	27,30	27,56
Plate B	Rep1	26,44	26,57	26,60	26,61	26,64	26,45	26,01	26,53	26,65	26,42	26,38	24,71	26,03	26,61	26,65	26,72	26,78	26,65	26,42	27,42	26,27	27,22	27,43
	Rep2	26,35	26,45	26,59	27,20	26,65	26,50	27,04	27,04	26,76	26,58	26,38	26,24	26,71	26,69	26,75	26,92	27,00	27,09	26,75	26,60	26,77	27,27	27,24
	Rep3	26,40	26,75	26,90	26,90	27,02	26,56	25,73	26,58	26,63	26,47	26,72	26,60	26,61	26,91	26,94	26,96	26,98	27,02	25,85	27,40	27,27	27,39	27,50
Plate C	Rep1	28,54	27,91	28,62	29,07	29,69	28,96	29,62	30,37	30,05	30,11	30,17	28,86	28,87	28,61	28,65	28,57	28,29	28,20	28,27	28,28	28,13	28,76	28,65
	Rep2	28,67	28,26	28,75	28,69	28,08	29,14	30,03	30,30	30,11	30,13	30,48	29,12	28,90	28,61	28,59	28,72	28,32	28,56	28,77	28,39	28,30	28,76	28,64
	Rep3	28,61	29,03	29,01	29,33	28,57	29,48	30,10	30,10	30,10	30,48	30,28	28,92	29,17	28,33	28,30	26,70	28,24	28,33	28,67	28,37	28,54	28,59	28,66

**2. The mean of all Cq values** 27,46

**3. The deviation of Cq values (each original Cq value - the mean)**

		<i>Tbp</i>	<i>Rexo5</i>	<i>Auh</i>	<i>Csnk1a1</i>	<i>Dolpp1</i>	<i>Efhd2</i>	<i>Eif3a</i>	<i>Eif4g2</i>	<i>Elac2</i>	<i>Grn</i>	<i>Kdm3b</i>	<i>Maea</i>	<i>Mpv17l</i>	<i>Nupl2</i>	<i>Pcsk7</i>	<i>Puf60</i>	<i>Tjrc</i>	<i>Tmed2</i>	<i>Rpl4</i>	<i>Tnks2</i>	<i>Wars2</i>	<i>Zc3h11a</i>	<i>Psm4</i>
Plate A	Rep1	-1,19	-0,88	-0,74	-1,10	-0,72	-0,70	-0,93	-1,01	-0,85	-0,97	-1,11	-1,59	-2,05	-0,78	-0,83	-0,77	-0,68	-0,87	-0,98	-0,13	0,08	-0,20	-0,13
	Rep2	-2,32	-0,77	-0,52	-0,87	-0,57	-0,59	-0,78	-0,97	-0,86	-0,68	-0,80	-0,92	-0,83	-0,71	-0,74	-0,66	-0,36	-0,38	-1,85	-0,54	-0,07	-0,42	0,10
	Rep3	-1,22	-0,82	-0,68	-0,63	-0,50	-0,68	-0,94	-0,76	-0,87	-0,86	-0,84	-0,82	-0,69	-0,69	-0,65	-0,48	-0,28	-0,61	-1,16	-0,17	-0,20	-0,16	0,10
Plate B	Rep1	-1,02	-0,89	-0,86	-0,85	-0,82	-1,01	-1,45	-0,93	-0,81	-1,04	-1,08	-2,75	-1,43	-0,85	-0,81	-0,74	-0,68	-0,81	-1,04	-0,04	-1,19	-0,24	-0,03
	Rep2	-1,11	-1,01	-0,87	-0,26	-0,81	-0,96	-0,42	-0,42	-0,70	-0,88	-1,08	-1,22	-0,75	-0,77	-0,71	-0,54	-0,46	-0,37	-0,71	-0,86	-0,69	-0,19	-0,22
	Rep3	-1,06	-0,71	-0,56	-0,56	-0,44	-0,90	-1,73	-0,88	-0,83	-0,99	-0,74	-0,86	-0,85	-0,55	-0,52	-0,50	-0,48	-0,44	-1,61	-0,06	-0,19	-0,07	0,04
Plate C	Rep1	1,08	0,45	1,16	1,61	2,23	1,50	2,16	2,91	2,59	2,65	2,71	1,40	1,41	1,15	1,19	1,11	0,83	0,74	0,81	0,82	0,67	1,30	1,19
	Rep2	1,21	0,80	1,29	1,23	0,62	1,68	2,57	2,84	2,65	2,67	3,02	1,66	1,44	1,15	1,13	1,26	0,86	1,10	1,31	0,93	0,84	1,30	1,18
	Rep3	1,15	1,57	1,55	1,87	1,11	2,02	2,64	2,64	2,64	3,02	2,82	1,46	1,71	0,87	0,84	-0,76	0,78	0,87	1,21	0,91	1,08	1,13	1,20

**4. The mean deviation of each triplicate (calibration factor, CF)**

		<i>Tbp</i>	<i>Rexo5</i>	<i>Auh</i>	<i>Csnk1a1</i>	<i>Dolpp1</i>	<i>Efhd2</i>	<i>Eif3a</i>	<i>Eif4g2</i>	<i>Elac2</i>	<i>Grn</i>	<i>Kdm3b</i>	<i>Maea</i>	<i>Mpv17l</i>	<i>Nupl2</i>	<i>Pcsk7</i>	<i>Puf60</i>	<i>Tjrc</i>	<i>Tmed2</i>	<i>Rpl4</i>	<i>Tnks2</i>	<i>Wars2</i>	<i>Zc3h11a</i>	<i>Psm4</i>
CF	Plate A	-1,57	-0,82	-0,64	-0,86	-0,59	-0,65	-0,88	-0,91	-0,86	-0,83	-0,91	-1,11	-1,19	-0,72	-0,74	-0,63	-0,44	-0,62	-1,33	-0,28	-0,06	-0,26	0,03
	Plate B	-1,06	-0,87	-0,76	-0,55	-0,69	-0,95	-1,20	-0,74	-0,78	-0,97	-0,96	-1,61	-1,01	-0,72	-0,68	-0,59	-0,54	-0,54	-1,12	-0,32	-0,69	-0,16	-0,07
	Plate C	1,15	0,94	1,34	1,57	1,32	1,74	2,46	2,80	2,63	2,78	2,85	1,51	1,52	1,06	1,06	0,54	0,83	0,91	1,11	0,89	0,87	1,25	1,19

**5. The CF was subtracted from original Cq values of the assayed gene, interrunkalibration completed.**

### 3.8 GeNorm und Normfinder

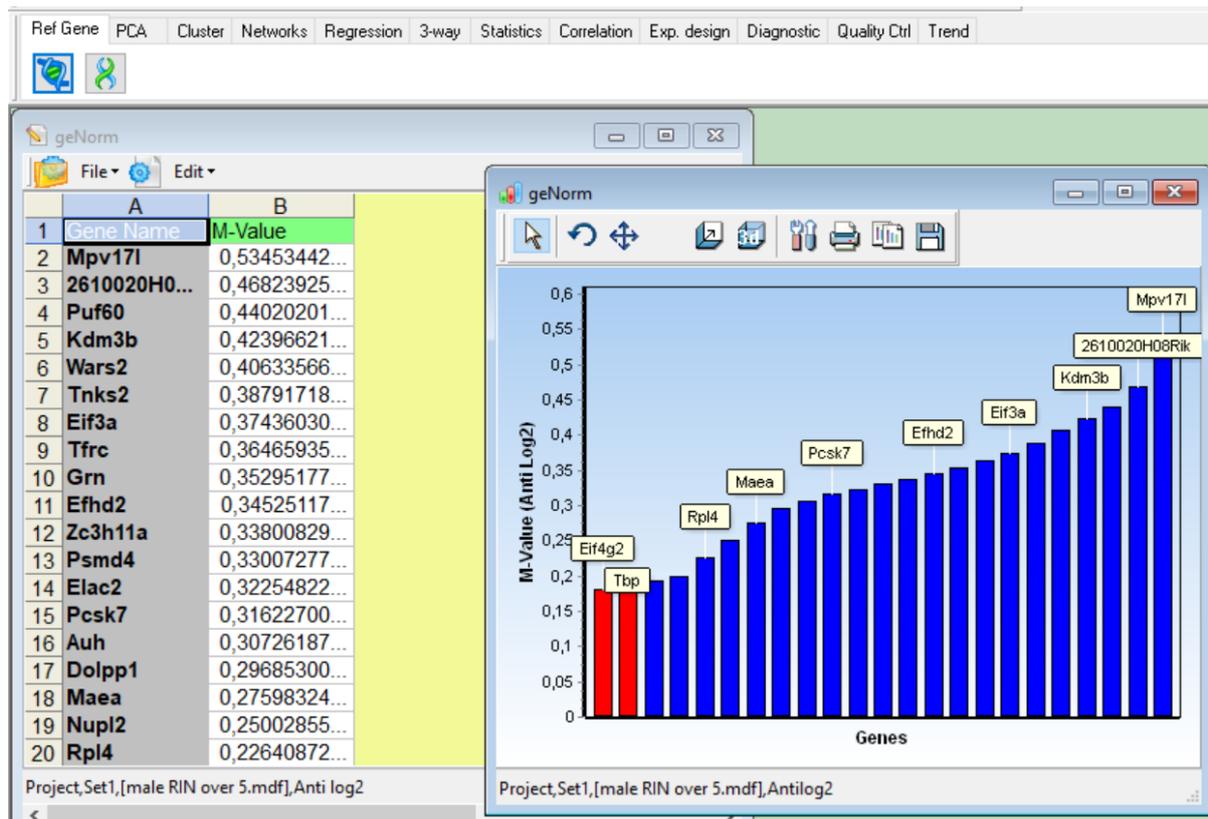
Zur Auswertung der Cq-Rohdaten aus der RT-qPCR-Validierung für die vorgewählten Referenzgenkandidaten wurden die zwei gängigsten Auswertelgorithmen, nämlich GeNorm und Normfinder, eingesetzt, welche in der Software für RT-qPCR-Datenanalyse, nämlich GenEx™ 7.0 (Bio MCC), integriert sind. GenEx™ bietet verschiedene Importmöglichkeiten für Cq-Rohdaten. Im vorliegenden Experiment wurden die Cq-Rohdaten aus der RT-qPCR-geräteigenen Software (Roche) zur Überprüfung der deskriptiven Statistik zuerst in Excel übertragen und danach zur eigentlichen Auswertung durch GeNorm und Normfinder in GenEx™ importiert (**Abb. 3-4**).

	A	B	C	D	E	F	G
1	CDNA Nr	Tbp	2610020H0...	Auh	Csnk1a1	Dolpp1	Efhd2
2	3	27,0769323...	27,2335990...	26,3219323...	21,3369323...	28,7719323...	26,0219323...
3	8	26,6019323...	26,7385990...	25,6919323...	21,6769323...	27,9519323...	25,1919323...
4	11	27,7869323...	27,8785990...	27,3819323...	22,2269323...	29,8569323...	26,9769323...
5	12	28,8419323...	28,7135990...	28,3669323...	23,2169323...	31,0119323...	28,3619323...
6	18	28,4619323...	28,0235990...	28,1769323...	22,6519323...	30,4319323...	27,7469323...
7	19	27,9519323...	27,8185990...	27,5969323...	22,2619323...	29,9869323...	27,1719323...
8	20	28,5819323...	28,4935990...	27,9069323...	22,8619323...	30,6119323...	27,8319323...
9	22	28,5219323...	28,4435990...	27,9169323...	22,7869323...	30,4769323...	27,5769323...
10	32	27,6769323...	27,7935990...	27,5519323...	22,1719323...	29,5819323...	27,0769323...
11	45	27,9335990...	29,0352657...	28,2435990...	22,3919323...	30,9402657...	27,7619323...
12	47	27,5985990...	29,0152657...	27,9435990...	22,1219323...	30,1052657...	27,3619323...
13	48	28,1335990...	29,3152657...	28,2335990...	22,4069323...	30,3852657...	27,6969323...
14	50	27,8635990...	29,0552657...	28,1335990...	22,1619323...	30,4252657...	27,6519323...
15	51	27,6635990...	28,8202657...	27,5435990...	22,1119323...	29,9352657...	25,8719323...
16	60	26,9985990...	28,6102657...	27,0035990...	21,2369323...	29,2852657...	26,5269323...
17	61	27,6585990...	28,8752657...	27,6485990...	21,8819323...	30,0352657...	27,0119323...
18	63	27,0985990...	28,6852657...	27,2735990...	21,5269323...	29,5002657...	26,5019323...
19	64	26,9385990...	28,3402657...	26,8735990...	21,1769323...	29,0302657...	26,0769323...
20	73	26,4985990...	28,7402657...	26,2785990...	21,0519323...	28,6252657...	25,6419323...
21	74	26,6085990...	28,7652657...	26,5635990...	21,2019323...	28,7702657...	26,0769323...
22	76	25,8785990...	28,1802657...	25,9585990...	20,6619323...	27,9902657...	24,9769323...

Source: C:\Users\espre\Downloads\male RIN over 5.xlsx

**Abb. 3-4 Importierung der Cq-Rohdaten in GenEx™.** Unter "Tools" auf der Menüleiste wurde "Data manager" geöffnet und anschließend in "File-Open" die Cq-Rohdaten im Excel-Dateiformat geladen. Hier sind beispielsweise die importierten Rohdaten der Gruppe "Male, RIN>5" dargestellt. Außerdem sind die Symbole für GeNorm und Normfinder in der Funktionsleiste "Ref Gene" zu erkennen. *Bildschirmfoto aus GenEx™.*

Nach Importierung der Cq-Rohdaten wurden GeNorm- und Normfinder-Analyse mit Anklicken des entsprechenden Symbols auf der Funktionsleiste "Ref Gene" gestartet. Das Ergebnis der jeweiligen Analyse-Algorithmen wird sowohl tabellarisch mit berechneten Stabilitätswerten als auch grafisch dargestellt (**Abb. 3-5, Abb. 3-6**). Mit GeNorm werden die zwei am stabilsten exprimierten Gene in der ausgewerteten Analysegruppe identifiziert und Normfinder das eine Gen. Die restlichen mituntersuchten Referenzgenkandidaten werden ebenfalls entsprechend den errechneten Stabilitätswerten in eine Rangordnung gebracht.



**Abb. 3-5 GeNorm-Ergebnis.** Als Beispiel ist hier ein Analyseergebnis von GeNorm dargestellt. Es kommt sowohl tabellarisch mit genauen Stabilitätswerten als auch grafisch als Balkendiagramm zur Darstellung. *Bildschirmfoto aus GenEx™*

	A	B	C	D	E
1	Gene Name	SD	Acc. SD	SD	0,1891
2	Maea	0,1891	0,1891		
3	Rpl4	0,2111	0,1417	Best Gene	Maea
4	Auh	0,2118	0,1179		
5	Eif4g2	0,2198	0,1041		
6	Eif3a	0,2249	0,0947		
7	Zc3h11a	0,2273	0,0875		
8	Tbp	0,2403	0,0825		
9	Tfrc	0,2427	0,0783		
10	Csnk1a1	0,2589	0,0753		
11	Dolpp1	0,2678	0,0729		
12	Tmed2	0,2731	0,0708		
13	Nupl2	0,2947	0,0694		
14	Tnks2	0,3094	0,0683		
15	Pcsk7	0,3892	0,0692		
16	Psm4	0,401	0,0699		
17	Efhd2	0,4073	0,0703		
18	Elac2	0,4233	0,0707		
19	Wars2	0,4358	0,071		
20	Puf60	0,4507	0,0714		
21	Kdm3b	0,4984	0,0722		
22	Gm	0,5084	0,0729		
23	2610020H0...	0,7015	0,0766		

**Abb. 3-6 Normfinder-Ergebnis.** Hier ist beispielsweise ein Analyseergebnis von Normfinder zu sehen. Die Normfinder-SV-Werte (Spalte "SD") als Parameter für Expressionsstabilität der jeweiligen Referenzgenkandidaten sind dargestellt. Demnach ist *Maea* das am stabilsten exprimierte Kandidatengen mit einem Normfinder-SV-Wert von 0,1891 in der ausgewerteten Analysegruppe. Normfinder berechnet außerdem sog. akkumulierte Standardabweichung (Spalte "Acc. SD"), die zur Ermittlung einer optimalen Anzahl der simultan einsetzbaren Referenzgene berechnet wird. Im oben vorgestellten Fall wäre die optimale Anzahl der Referenzgene 14 (Ab dem 15. Gen steigt die akkumulierte SD wieder an), jedoch eine Anwendung so großer Anzahl der Referenzgene ist aus Zeit- und Kostengründen praktisch unrealistisch (dazu im Detail siehe Abschnitt 2.3). *Bildschirmfoto aus GenEx™*

Zur Dateninterpretation wurden die verfügbaren Lungenproben in folgende Analysegruppen unterteilt.

- Normoxie & PND2.5-28.5 & RIN>5 (n=17)
- Normoxie & RIN>5 & PND2.5 (n=6) oder PND5.5 (n=7) oder PND15.5-28.5 (n=4)
- Normoxie & RIN>5 & männliche (n=21) oder weibliche (n=20) Mäuse
- Hyperoxie & PND2.5-28.5 & RIN>5 (n=36)
- Normoxie & PND2.5-28.5 & RIN<5 (n=33)

Basierend auf diesen Analysegruppen mit unterschiedlichen Experimentalbedingungen wurden GeNorm bzw. Normfinder abhängig von Analysezielen folgendermaßen angewendet.

1. Zuverlässige Referenzgene für eine bestimmte Experimentalbedingung: Dafür wurde die betreffende Analysegruppe z.B. physiologische Bedingung (Normoxie & PND2.5-28.5 & RIN>5) durch GeNorm bzw. Normfinder analysiert. Anhand ergebnen Stabilitätswerten wurden die Referenzgenkandidaten in eine Rangordnung gebracht, wobei nur diejenigen

Kandidatengene mit einem GeNorm-M- bzw. Normfinder-SV-Wert unter 0,5 als akzeptabel stabil exprimierte Referenzgene betrachtet wurden.<sup>73,74</sup>

2. Zuverlässige Referenzgene unter Einfluss einer experimentellen Variable, d.h. Entwicklungsstadien, Geschlecht, Hyperoxie-Exposition, RNA-Qualität: Dafür wurde die Expressionsstabilität der Referenzgenkandidaten zwischen zwei (ggf. mehr als zwei) Analysegruppen verglichen, in denen außer der zu untersuchenden experimentellen Variable alle sonstigen konstant gehalten wurden. Beispielsweise zur Suche nach stabilen Referenzgenen unter Hyperoxie-Exposition wurden die Analysegruppen jeweils aus der Normoxie- und Hyperoxie-Bedingung verglichen, in beiden die sonstigen experimentellen Variablen, d.h. Entwicklungsstadien und RNA-Qualität konstant gehalten wurden.

Der Vergleich zwischen Analysegruppen erfolgte in zwei Vorgehensweisen. Erstens durch Vergleich des mittleren Stabilitätsrankings jeweiliger Referenzgenkandidaten in den zu vergleichenden Analysegruppen. Zweitens durch Intra- und Intergruppenvariationsanalyse von Normfinder vonstatten.

Zur Berechnung der mittleren Stabilitätsrankings wurde zunächst GeNorm- bzw. Normfinder-Analyse für jeweilige zu vergleichende Analysegruppen durchgeführt und anhand ergebnen Stabilitätswerten wurden die Referenzgenkandidaten in eine Rangordnung gebracht und entsprechend Stabilitätsranking innerhalb der jeweiligen Analysegruppe vergeben. Als nächstes wurde mittleres Stabilitätsranking der jeweiligen Referenzgenkandidaten in den zu vergleichenden Analysegruppen berechnet, und zwar nur für diejenigen Referenzgenkandidaten, die in allen zu vergleichenden Analysegruppen einen GeNorm-M- bzw. Normfinder-Stabilitätswert unter 0,5 (akzeptabler Stabilitätswert s.o.) aufwiesen. Zum Schluss wurde der Referenzgenkandidat mit dem höchsten mittleren Stabilitätsranking als das beste Referenzgen bezeichnet.

Die Intra- und Intergruppenvariationsanalyse von Normfinder wurde wie oben erläutert ausgeführt. Als Ergebnis ergaben sich zum einen das eine beste Referenzgen und zum anderen die beste Kombination zweier Referenzgene hinsichtlich den zu vergleichenden Analysegruppen.

### 3.9 Einzelzell-RNA-Sequenzierung

Die Einzelzell-RNA-Sequenzierung der Referenzgenkandidaten in diversen pulmonalen Zelltypen wurde freundlicherweise von unserem Kooperationspartner Pädiatrie, Vanderbilt University Medical Center, USA (AG Prof. Sucre) durchgeführt und die ergebnen Daten wurden uns zur weiteren Interpretation zur Verfügung gestellt. Unten ist die Durchführung des Experimentes kurz zusammengefasst.

Die Daten aus Einzelzell-RNA-Sequenzierung in Lungenproben aus den neonatalen Mäusen am PND0, 3, 5, 7 sowie 14 wurden hinsichtlich der Expression der Referenzgenkandidaten in diversen pulmonalen Zelltypen analysiert, wie bereits in einer früheren Arbeit beschrieben wurde.<sup>75</sup> Dieser Datensatz wurde von der zellulären Suspension aus Mauslungen generiert, in der Ter119-positive Blutzellen sowie CD45-positive Immunzellen depletiert wurden. Die Sequenzierung erfolgte mittels 10x Genomics Chromium Plattform. Variation in Genexpression von den Referenzgenkandidaten wurde anhand Standardabweichung in den diversen Zellen beurteilt und die Gene mit der niedrigsten Standardabweichung wurden als Gene mit der geringsten Expressionsvariabilität ausgewertet.

## 4 Ergebnisse

### 4.1 Vorauswahl der Referenzgenkandidaten

#### 4.1.1 Datenlage hinsichtlich der Referenzgene für die Mauslunge

Bei der Untersuchung der aktuellen Datenlage hinsichtlich der Referenzgene wurden zuerst sechs unabhängige Genexpressionstudien<sup>76-81</sup> identifiziert, in denen ein oder mehrere von konventionell verwendeten Referenzgenen für RT-qPCR-Experimente mit Lungengewebe von C57BL/6-Mäusen zum Einsatz kamen (**Tab. 4-1**). Die Altersspanne der darin verwendeten Mäuse reichte von 3 Tagen bis 26 Monate. Neonatale Mäuse mit Alter von bis zu einem Monat nach Geburt wurden lediglich in einer<sup>78</sup> von den sechs Studien verwendet. Neben Wildtyp-Mäusen wiesen die in den Studien untersuchten Mäuse unterschiedliche Genotypen (transgene und/oder knockout Allele) für die Gene *Tph1*, *Ucp2/3*, *Il-4*, *Stat6*, *SPC-IL-18* auf. Außerdem wurden die Versuchsmäuse entsprechend dem jeweiligen Studienziel unterschiedlich experimentell behandelt (Kalorienrestriktion, Behandlung mit Bleomycin, Hypoxie-Exposition). In diesen Studien wurden insgesamt sieben konventionell verwendete Referenzgene, nämlich *Actb*, *B2m*, *Gapdh*, *Gusb*, *Hmbs*, *Oaz1*, *Pgk1* identifiziert, wobei *Gapdh* in zwei voneinander unabhängigen Studien als Referenzgen eingesetzt wurde. Die Eignung dieser Gene als Referenzgen wurde bemerkenswerter Weise in den betreffenden Studien nicht evaluiert.

**Tab. 4-1 Konventionell verwendete Referenzgene für Lungengewebe der C57BL/6-Mäuse.** *Gapdh* wurde in zwei Studien als Referenzgen verwendet. G3P: Glycerinaldehyd-3-Phosphat, WT: Wildtyp, TG: transgen. T: Tage, W: Wochen, M: Monate

Symbol	Entrez	Vollständiger Name	Informationen über die untersuchten Mäuse			
			Geschlecht	Alter	Genotyp außer WT	Referenzen
<i>Oaz1</i>	18245	Ornithin-Decarboxylase-Antizym 1	beides	3,5-5 M	<i>Tph1</i> (-/-)	Rothman et al. <sup>80</sup>
<i>Gusb</i>	110006	$\beta$ -Glucuronidase	f	8, 19, 26 M	<i>Ucp2/3</i> TG	McDonald et al. <sup>81</sup>
<i>B2m</i>	12010	$\beta$ 2-Mikroglobulin	n/a	6-8 W	-	Tzouveleakis et al. <sup>79</sup>
<i>Hmbs</i>	15288	Hydroxymethylbilan-Synthase	f	3-21 T	-	Kitowska et al. <sup>78</sup>
<i>Gapdh</i>	14433	G3P-Dehydrogenase				
<i>Actb</i>	11461	$\beta$ -Aktin	m	8-12 W	<i>Il-4, Stat6</i> (-/-)	Yamaji et al. <sup>76</sup>
<i>Pgk1</i>	18655	Phosphoglyceratkinase 1				
<i>Gapdh</i>	14433	G3P-Dehydrogenase	f	5, 9, 13 W	<i>SPC-IL-18</i> TG	Sakazaki et al. <sup>77</sup>

In der weiteren Recherche in *PubMed* wurden vier Studien gefunden, im Rahmen derer die Eignung konventionell verwendeter Referenzgene überprüft und Vorschläge für zuverlässige neue Referenzgene gemacht wurden (**Tab. 4-2**).<sup>23,82-84</sup> Die Altersspanne der untersuchten Mäuse

in dieser Art von Studien reichte von 8-15 Wochen, d.h. die Lungenentwicklung war bereits abgeschlossen. Die Mäuse wurden außerdem unterschiedlich genetisch oder experimentell behandelt z.B. *p50/p105 (Nfkb1)*-Knockout, Hyperoxie-Exposition. Eine von den vier Studien stellte sich als eine Metaanalyse von 2.543 publizierten Microarray-Experimenten mit Mäusen dar, ohne genauere Angaben zum Alter der Mäuse zu machen.<sup>23</sup>

In den genannten vier Studien wurden insgesamt 30 Referenzgenkandidaten gefunden und davon sind einige Gene wie *B2m*, *Gapdh*, *Actb* bereits als konventionell verwendete Referenzgene bekannt. Diese Gene zeigten in der untersuchten Experimentalbedingung der jeweiligen Studie eine stabile Genexpression. *Rpl13a* erwies sich in zwei unabhängigen Studien als gutes Referenzgen.

**Tab. 4-2 Referenzgenkandidaten aus Evaluierungsstudien.** Die oberen zwei Studien evaluierten konventionell verwendeten Referenzgene in bestimmten Experimentalbedingungen (Hyperoxie-Exposition, induzierte Lungenentzündung) und die aufgeführten Gene erwiesen sich jeweils in den Experimentalbedingungen als bewährte Referenzgene mit der niedrigsten Alteration des Expressionslevels. Die unteren zwei Studien erzielten neuartige verlässliche Referenzgene durch Evaluation herausfinden. Daraus resultierten die aufgeführten Gene mit höchster Expressionsstabilität und sie waren außerdem ubiquitär in verschiedenen Geweben nachweisbar. \* In der betreffenden Studie wurde eigentlich *Eif3f* vorgeschlagen. G3P: Glycerinaldehyd-3-Phosphat.

Symbol	Entrez	Genname	Studienaufbau
<i>B2m</i>	12010	β2-Mikroglobulin	
<i>Rplp2</i>	67186	Saures ribosomales 60S-Protein P2	8 Wochen alte, männliche C57BL/6J-Mäuse mit Hyperoxie-Exposition: 4 Gene mit minimaler Alteration von mRNA-Level. <sup>82</sup>
<i>Gapdh</i>	100042025	G3P-Dehydrogenase	
<i>Ipo8</i>	320727	Importin 8	
<i>Actb</i>	11461	β-Aktin	10-12 Wochen alte, <i>p50/p105</i> -defiziente Mäuse mit induzierter Lungenentzündung: 4 Gene mit stabiler Expression unter dieser Bedingung. <sup>83</sup>
<i>Gusb</i>	110006	β-Glucuronidase	
<i>Rpl13a</i>	22121	60S-ribosomales protein, L13a	
<i>Ppia</i>	268373	Peptidylprolylisomerase A	
<i>Eef2</i>	13629	Eukaryotischer Elongationsfaktor 2	12-15 Wochen alte C57BL/6-Mäuse mit Genexpressionsuntersuchung mittels SAGE: 9 Gene mit konstanter Expression in allen untersuchten Geweben. <sup>84</sup>
<i>Rpl37</i>	67281	Ribosomales Protein L37	
<i>Leng8</i>	232798	Leukocyte receptor cluster member 8	
<i>Heatr3</i>	234549	Heat repeat containing 3	
<i>Eif3a*</i>	66085	Eukaryotischer Translationsinitiationsfaktor 3a	
<i>Puf60</i>	67959	Poly(U)-binding-splicing factor	
<i>Rpl38</i>	67671	Ribosomales Protein L38	
<i>Chmp2a</i>	295670	Charged multivesicular body protein 2a	
<i>Psm4</i>	19185	26S-Proteasome-Nicht-ATPase-regulatorische Untereinheit 4	
<i>Rps29</i>	20090	Ribosomales Protein S29	
<i>Rpl4</i>	67891	Ribosomales Protein L4	
<i>Oaz1</i>	18245	Ornithin-Decarboxylase-Antizyme 1	
<i>Rpl13a</i>	22121	Ribosomales Protein L13a	
<i>Rpl6</i>	19988	Ribosomales Protein L6	
<i>Srp14</i>	20813	Signal recognition particle 14 kDa	
<i>Rpl24</i>	68193	Ribosomales Protein L24	
<i>Rpl27</i>	19942	Ribosomales Protein L27	
<i>Rps13</i>	68052	Ribosomales Protein S13	
<i>Rpl9</i>	20005	Ribosomales Protein L9	
<i>Rpl22</i>	19934	Ribosomales Protein L22	
<i>Rps16</i>	20055	Ribosomales Protein S16	
<i>Rps12</i>	20042	Ribosomales Protein S12	
<i>Rps20</i>	67427	Ribosomales Protein S20	

#### 4.1.2 Referenzenvorschläge von Genevestigator

Für die vorgegebene Experimentalbedingung d.h. Lunge der C57BL/6-Mäuse wurden in Genevestigator insgesamt zehn publizierte Studien mit Microarray-Experimenten gefunden, welche eine genomweite Expressionsanalyse bei unterschiedlichen experimentellen Bedingungen mit Lungengewebe von C57BL/6-Mäuse beschäftigt haben.<sup>10,85-93</sup> Die untersuchten Mäuse waren überwiegend Wildtyp, jedoch drei gentechnisch modifizierte Mauslinien (*Cf*<sup>-/-</sup>, *Pbef*<sup>-/-</sup>, *Gadd45a*<sup>-/-</sup>) waren auch eingeschlossen. Die Altersspanne der Mäuse reichte von 2 bis 22 Wochen und somit war die Lungenentwicklung bei allen Studien außer einer<sup>88</sup> bereits vollendet. Außerdem wurden in diesen Studien im Sinne eines Mausmodells einer humanen Erkrankung unterschiedliche experimentelle Behandlungen ausgeführt, um deren Effekt festzustellen (**Tab. 4-3**).

**Tab. 4-3 Die von Genevestigator erfassten Studien mit Genexpressionsanalyse.** Für die vorgegebene Experimentalbedingung, d.h. Lunge der C57BL/6-Mäuse wurden insgesamt zehn Genexpressionsstudien mittels Microarray identifiziert. ALI: *acute lung injury*, CF: Cystische Fibrose, GPCR: G-protein-gekoppelter Rezeptor, VILI: *ventilator-induced lung injury*, LPS: lipopolysaccharide, PBS: *phosphate-buffered saline*.

Studie	Studienaufbau	Informationen über Mäuse		
		Genotyp	Alter	Geschlecht
Jacobson et al. <sup>85</sup>	Effekt von Simvastatin im Mausmodell von ALI	wt	8-10	m
Haston et al. <sup>94</sup>	CF und Non-CF-Mäuse mit unterschiedlichen Stämmen	wt, <i>Cf</i> <sup>-/-</sup>	12	beides
Thorrez et al. <sup>10</sup>	Expressionsanalyse der für ribosomale Proteine kodierenden Gene	wt	10-12	beides
Lattin et al. <sup>87</sup>	Expressionsanalyse der für GPCR kodierenden Gene in murinen Makrophagen	wt	8-10	beides
McGrath et al. <sup>88</sup>	Effekt der Exposition zu Zigarettenrauch auf neonatale Lungen	wt	2, 8	n/a
Hong et al. <sup>89</sup>	Erforschung der mechanistischen Beteiligung von <i>Pbef</i> in VILI	wt, <i>Pbef</i> <sup>-/-</sup>	8-22	m
Kozul et al. <sup>90</sup>	Exposition zu Hochdosis-Arsen in Trinkwasser oder Nahrungsmittel	wt	6-7	m
Wray et al. <sup>91</sup>	Genexpressionsanalyse bei kurzzeitiger VILI	wt	adult	m
Meyer et al. <sup>92</sup>	Genexpressionsanalyse bei VILI in <i>Gadd45a</i> <sup>-/-</sup> -Mäuse	wt, <i>Gadd45a</i> <sup>-/-</sup>	8-12	m
Camateros et al. <sup>93</sup>	Expressionsanalyse des allergie-assoziierten Transkriptom nach Behandlung mit Resiquimod	wt	16	n/a

Alle 139 erfassten Microarray-Proben aus diesen zehn Studien wurden durch Genevestigator analysiert und daraus ergaben sich die 20 besten Gene mit hoher Expressionsstabilität für jeweils drei Expressionslevels (Hoch, Mittel, Niedrig). Für *Tmed2* aus der hohen Expressionsgruppe waren drei verschiedene Probe Sets auf dem Affymetrix Microarray vorhanden (1436452\_x\_at,

1455968\_x\_at, 1436451\_a\_at). Somit wurden in diesem Schritt insgesamt 58 Referenzgenkandidaten identifiziert (**Tab. 4-4**).

**Tab. 4-4 Genevestigator-Ergebnis.** Berücksichtigt man das mehrfache Vorhandensein von *Tmed2* aufgrund mehrfach vorhandener Microarray-Probe Sets resultierten aus diesem Schritt insgesamt 58 Referenzgenkandidaten. Exp. level: Expressionslevel, SD: Standardabweichung.

Hoch			Mittel			Niedrig		
Gene	Exp. level	SD	Gene	Exp. level	SD	Gene	Exp. level	SD
<i>Fkbp1a</i>	15,43	0,24	<i>Dolpp1</i>	11,63	0,16	<i>Trim37</i>	9,51	0,22
<i>Eif4h</i>	15,77	0,25	<i>Mpv17l</i>	10,72	0,18	<i>Zfp182</i>	9,77	0,22
<i>Tmed2</i> (1436452_x_at)	15,29	0,25	<i>Elac2</i>	11,27	0,19	<i>Mcp1</i>	9,55	0,24
<i>Efh2</i>	15,31	0,27	<i>Kdm3b</i>	12,63	0,19	<i>Auh</i>	9,31	0,24
<i>Zc3h11a</i>	15,15	0,28	<i>1435789_x_at</i>	11,58	0,19	<i>Cep57l1</i>	10,01	0,25
<i>Sec62</i>	15,18	0,30	<i>Maskbp3</i>	12,36	0,20	<i>Caprin2</i>	9,06	0,26
<i>Atp6v0c</i>	15,03	0,31	<i>Rab23</i>	11,28	0,20	<i>Wars2</i>	9,73	0,26
<i>Tmed2</i> (1455968_x_at)	15,81	0,31	<i>Plekhm3</i>	12,01	0,20	<i>Ing3</i>	9,77	0,26
<i>Tmed2</i> (1436451_a_at)	15,86	0,32	<i>Pcsk7</i>	12,00	0,20	<i>Nupl2</i>	9,32	0,27
<i>F2r</i>	15,84	0,32	<i>Metap1</i>	12,14	0,20	<i>Il15ra</i>	9,97	0,27
<i>Myadm</i>	15,24	0,33	<i>Zfp592</i>	11,19	0,20	<i>Slc35a5</i>	9,28	0,27
<i>Hnrnp1</i>	15,15	0,33	<i>Gabpb2</i>	12,58	0,20	<i>Crnk1</i>	9,63	0,28
<i>Csnk1a1</i>	15,15	0,33	<i>Hdgfrp2</i>	10,33	0,21	<i>Rhot1</i>	9,51	0,28
<i>Grn</i>	15,35	0,33	<i>Maea</i>	12,91	0,21	<i>1110035M17Rik</i>	9,21	0,28
<i>H3f3a</i>	16,42	0,34	<i>Gtf2h3</i>	11,07	0,21	<i>Nek4</i>	9,9	0,28
<i>Eif4g2</i>	16,11	0,34	<i>Nup214</i>	12,23	0,21	<i>Zxda</i>	8,33	0,28
<i>Tnks2</i>	15,12	0,34	<i>Heatr7a</i>	12,19	0,21	<i>Nudt6</i>	9,65	0,28
<i>Ywhaq</i>	15,25	0,34	<i>Zfp251</i>	11,38	0,21	<i>LOC106740</i>	9,13	0,28
<i>Canx</i>	15,04	0,35	<i>Ints5</i>	11,39	0,21	<i>Rexo5</i>	9,83	0,28
<i>Tomm20</i>	15,49	0,35	<i>Rrp1b</i>	10,99	0,22	<i>1446253_at</i>	9,7	0,28

#### 4.1.3 Roche kommerzielles Referenzgenpanel (RealTime Ready Assays)

Die 27 Gene, die aufgrund ihres häufigen Einsatzes als kommerzielles Referenzgenpanel von Roche etabliert sind, wurden in unsere Kandidatengenliste berücksichtigt (**Tab. 4-5**). Dabei waren einige Gene z.B. *Actb*, *B2m*, *Gapdh*, *Gusb*, *Hmbs* bereits in anderen Quellen für Kandidatengene identifiziert.

**Tab. 4-5 Die 27 Gene aus dem Roche-Referenzgenpanel.** Diese Gene wurden aufgrund ihres häufigen Einsatzes als Referenzgene als kommerzielles Panel etabliert.

Symbol	Entrez	Vollständiger Genname
<i>Actb</i>	11461	β-Aktin
<i>Alas2</i>	11656	Delta-Aminolevulinat-Synthase 2
<i>B2m</i>	12010	β2-Mikroglobulin
<i>Cyp2d13</i>	68444	Cytochrom P450, Familie 2, Subfamilie d, Polypeptid 13
<i>Fxyd6</i>	59095	FXYP domain-containing ion transport regulator 6
<i>Fxyd7</i>	57780	FXYP domain-containing ion transport regulator 7
<i>G6pdx</i>	14381	Glukose-6-Phosphat-Dehydrogenase X
<i>Gapdh</i>	100042025	Glycerinaldehyde-3-Phosphat-Dehydrogenase
<i>Gusb</i>	110006	β-Glucuronidase
<i>H2-Ea-ps</i>	100504404	Histocompatibility 2, class II antigen E alpha, pseudogene
<i>Hbb-b1</i>	15129	Hämoglobin, Untereinheit beta 1
<i>Hmbs</i>	15288	Hydroxymethylbilansynthase
<i>Hprt</i>	15452	Hypoxanthin-Guanin-Phosphoribosyl-Transferase
<i>Hsp90ab1</i>	15516	Hitzeshockprotein 90 kDa, ab1
<i>Ipo8</i>	320727	Importin 8
<i>Mx2</i>	17858	Myxovirus (influenza virus) resistance 2
<i>Oas1b</i>	23961	2'-5'-Oligoadenylatsynthetase 1b
<i>Pgk1</i>	18655	Phosphoglyceratkinase 1
<i>Pla2g2a</i>	18780	Phospholipase A2, Gruppe IIA
<i>Ppia</i>	268373	Peptidylprolylisomerase A
<i>Rpl13a</i>	22121	Ribosomales Protein L13a
<i>Rplp0</i>	11837	Ribosomales Protein, large, P0
<i>Sdha</i>	66945	Succinat-Dehydrogenase-Komplex, Untereinheit A, Flavoprotein
<i>Tbp</i>	21374	TATA-Bindeprotein
<i>Tfrc</i>	22042	Transferrin-Rezeptor
<i>Trpt1</i>	107328	tRNA-Phosphotransferase 1
<i>Ywhaz</i>	22631	Tyrosin-3-/Tryptophan-5-Monooxygenase-Aktivierungsprotein, zeta

#### 4.1.4 Ausschluss einer Reihe von Kandidatengen

Aus den geschilderten Quellen für die Auswahl der Referenzgenkandidaten wurden insgesamt 122 potenzielle Referenzgenkandidaten identifiziert und in einer Genliste (**Tab.A2**) zusammengestellt. Davon wurden 27 Gene auf Grund folgender Kriterien ausgeschlossen:

- Gene, die in mehr als einer Quelle identifiziert wurden (Überlappende Quellen, n=14).
- Gene, bei denen Pseudogene bekannt sind (n=12). Das Pseudogen bezeichnet eine DNA-Sequenz, die einem echten Gen ähnlich aufgebaut ist, jedoch nicht in der Lage ist, ein funktionierendes Protein zu kodieren. Das Vorhandensein des Pseudogens ist besonders

problematisch bei RT-qPCR-Experimenten, weil der Primer sowohl an das zu detektierende Gen, als auch an dessen Pseudogenvariante anlagern kann, was letztendlich zu einer fehlerhaften Quantifizierung des Zielgens führen würde.<sup>95,96</sup> Das Vorhandensein von Pseudogenen für den jeweiligen Referenzgenkandidaten wurde durch Suchanfrage auf [www.informatics.jax.org](http://www.informatics.jax.org) überprüft.

- Gene, die ribosomale Proteine kodieren (n=10). Diese Gengruppe zeigte in einer früheren Evaluierungstudie<sup>10</sup> generell instabile Expression und wurde somit zur Auswahl von Referenzgenen nicht in Betracht gezogen. Eine Ausnahme bildet *Rpl4*, das zur Bestätigung der Befunde der oben genannten Studie in unsere RT-qPCR-Validierung eingeschlossen wurde.
- Das Kandidatengen *Hbb-b1* kodiert eine Untereinheit des Hämoglobins, das dem Sauerstofftransport dient. *Hbb-b1* wurde als Referenzgenkandidat von vorne herein ausgeschlossen, da wir für dieses Gen annehmen, dass es in unserem Anwendungsfall (Hyperoxie) nicht stabil exprimiert wird.

#### 4.1.5 Erste Evaluierung in verfügbaren Microarray-Daten

Nach Ausschluss der 27 Kandidatengene anhand der oben genannten Kriterien lagen insgesamt 95 potenzielle Referenzgenkandidaten im weiteren Auswahlprozess vor. Im nächsten Schritt wurde die Expression dieser 95 Kandidatengene mit Hilfe von Transkriptom-Daten eines eigenen Microarray-Experiments im Lungengewebe neonataler Mäuse (PND5-7, Normoxie) evaluiert. Dabei wurde die Expression der 46 von den 95 Kandidatengenen bestätigt. Die 49 Kandidatengene waren entweder in den Microarray-Daten nicht exprimiert oder auf der verwendeten Microarray-Plattform (CodeLink UniSet Mouse I Bioarray 10K) nicht enthalten. Die Standardabweichung (SD) der Signalintensität aus den Microarray-Daten wurde als Parameter für Expressionsstabilität betrachtet.

Die Expressionsstabilität dieser 46 Kandidatengene wurde außerdem in Lungen der größtenteils adulten C57BL/6-Mäuse (lediglich eine von insgesamt 10 Studien mit neonatalen Mäusen) aus der Genevestigator-Analyse (Abschnitt 3.1.2) überprüft, die ebenfalls mit SD-Werten der Signalintensität repräsentiert war. Die SD-Werte aus Genevestigator reichten von 0,16-0,92.

Auf diese Weise wurden für jedes der 46 Kandidatengene zwei SD-Werte (einmal aus den eigenen Microarray-Daten (Abschnitt 3.2) und einmal aus Genevestigator (Abschnitt 3.1.2) erhalten und zu einer Overall-SD zusammengeführt, um die Expressionsstabilität dieser Gene in den sowohl neonatalen als auch adulten Mauslungen abzuschätzen. Die Overall-SD kam durch die untenstehende Formel zustande.

$$SD_{overall} = \sqrt{\frac{SD_{ref}^2 + SD_{array}^2}{2}}$$

Dabei bezeichnet  $SD_{ref}$  die SD-Werte aus Genevestigator (Subtool "RefGenes") und  $SD_{array}$  die SD-Werte aus unseren eigenen Microarray-Daten. Somit entspricht die Varianz für die Overall-SD dem Mittelwert von beiden Varianzen aus unseren eigenen Microarray-Daten sowie Genevestigator.

#### 4.1.6 Finale Referenzgenkandidaten

Anhand der Overall-SD wurden die 46 potenziellen Referenzgenkandidaten in einer Rangliste nach absteigender Genstabilität (ansteigende Overall-SD) sortiert (**Tab. 4-6**). Die 20 Kandidatengene mit der niedrigsten Overall-SD (0,118-0,242) wurden zur weiteren Validierung in RT-qPCR-Experimenten ausgewählt. Da 18 dieser 20 Gene auf Vorschlägen aus der Genevestigator-Analyse beruhten, wurden vier weitere Kandidatengene von den Rängen 21-46 des Stabilitätsrankings ausgewählt, um somit andere Datenquellen (publizierte Studien über Referenzgene und kommerzielles Referenzgenpanel) für die Kandidaten-Validierung zu berücksichtigen. Auf diese Weise wurden somit 24 Gene als finale Referenzgenkandidaten für die RT-qPCR-Validierung ausgewählt.

**Tab. 4-6 Die 46 potenziellen Referenzgenkandidaten mit Expressionsdaten aus unserer eigenen Transkriptom-Datenbank sowie Genevestigator.** Anhand der Overall-SD, die beide Microarray-Datensätze aus neonatalen (unsere eigenen Microarray-Daten) und adulten (Genevestigator) Mauslungen wiedergibt, wurden die besten 20 Gene sowie vier weitere Gene aus dem Stabilitätsranking 21-46 als finale Referenzgenkandidaten ausgewählt (vgl. fett geschriebene Gennamen). Damit wurden unterschiedliche Datenquellen sowie unterschiedliches Expressionslevel berücksichtigt. \**Fkbp1a* konnte in 6 von 60 Lungenproben bei der RT-qPCR nicht nachgewiesen werden und wurde daher von der weiteren Analyse mittels GeNorm und Normfinder ausgeschlossen (vgl. Abschnitt 4.2). Für das Expressionslevel (Exp.level) stehen H für Hoch, M für Mittel, N für Niedrig.

Gename	Entrez	Datenquelle	Exp. level	SD <sub>ref</sub>	SD <sub>array</sub>	SD <sub>overall</sub>	Ranking
<b><i>Dolpp1</i></b>	57170	c	M	0,16	0,05	0,118	1
<b><i>Mpv17l</i></b>	93734	c	M	0,18	0,03	0,129	2
<b><i>Elac2</i></b>	68626	c	M	0,19	0,05	0,139	3
<b><i>Pcsk7</i></b>	18554	c	M	0,20	0,03	0,143	4
<b><i>Kdm3b</i></b>	277250	c	M	0,19	0,07	0,143	5
<b><i>Maea</i></b>	59003	c	M	0,21	0,05	0,152	6
<b><i>Fkbp1a*</i></b>	14225	c	H	0,24	0,06	0,174	7
<b><i>Wars2</i></b>	70560	c	M	0,24	0,08	0,179	8
<b><i>Auh</i></b>	11992	c	N	0,24	0,08	0,180	9
<b><i>Efhd2</i></b>	27984	c	H	0,27	0,02	0,192	10
<b><i>Zc3h11a</i></b>	70579	c	H	0,28	0,02	0,198	11
<b><i>Rexo5</i></b>	434234	c	N	0,28	0,03	0,199	12
<b><i>Nupl2</i></b>	231042	c	N	0,27	0,10	0,204	13
<b><i>Tmed2</i></b>	56334	c	H	0,31	0,02	0,220	14
<b><i>Grn</i></b>	14824	c	H	0,33	0,02	0,234	15
<b><i>Eif3a</i></b>	13669	b	N	0,32	0,09	0,235	16
<b><i>Tbp</i></b>	21374	d	M	0,33	0,04	0,235	17
<b><i>Csnk1a1</i></b>	93687	c	H	0,33	0,08	0,241	18
<b><i>Eif4g2</i></b>	13690	c	H	0,34	0,03	0,241	19
<b><i>Tnks2</i></b>	74493	c	H	0,34	0,04	0,242	20
<b><i>F2r</i></b>	14062	c	H	0,32	0,13	0,243	21
<b><i>Rab23</i></b>	19335	c	M	0,34	0,10	0,251	22
<b><i>Myadm</i></b>	50918	c	H	0,33	0,15	0,256	23
<b><i>Rrp1b</i></b>	72462	c	M	0,37	0,07	0,266	24
<b><i>Hprt</i></b>	15452	d	H	0,37	0,07	0,267	25
<b><i>Canx</i></b>	12330	c	H	0,35	0,15	0,268	26
<b><i>Eif4h</i></b>	22384	c	H	0,39	0,04	0,277	27
<b><i>Slc35a5</i></b>	74102	c	H	0,41	0,03	0,291	28
<b><i>Atp6v0c</i></b>	11984	c	H	0,45	0,06	0,321	29
<b><i>Ywhaz</i></b>	22631	d	H	0,45	0,06	0,321	30
<b><i>Mx2</i></b>	17858	d	N	0,43	0,17	0,328	31
<b><i>Hdgfrp2</i></b>	15193	c	M	0,52	0,00	0,368	32

Genname	Entrez	Datenquelle	Exp. level	SD <sub>ref</sub>	SD <sub>array</sub>	SD <sub>overall</sub>	Ranking
<b>Puf60</b>	67959	b	H	0,54	0,01	0,382	33
<b>Tfrc</b>	22042	d	H	0,51	0,21	0,390	34
<i>Tomm20</i>	67952	c	H	0,56	0,04	0,397	35
<b>Psm4</b>	19185	b	H	0,62	0,00	0,438	36
<i>Sdha</i>	66945	d	H	0,69	0,06	0,490	37
<i>G6pdx</i>	14381	d	H	0,71	0,04	0,503	38
<i>Rhot1</i>	59040	c	H	0,53	0,48	0,504	39
<i>Ing3</i>	71777	c	M	0,78	0,05	0,553	40
<i>Gusb</i>	110006	a, b, d	M	0,79	0,01	0,559	41
<i>Pla2g2a</i>	18780	d	N	0,80	0,03	0,566	42
<i>Actb</i>	11461	a, b, d	H	0,83	0,04	0,588	43
<b>Rpl4</b>	67891	b	H	0,85	0,02	0,601	44
<i>Gapdh</i>	14433	a, b, d	H	0,88	0,05	0,623	45
<i>Hsp90ab1</i>	15516	d	H	0,92	0,47	0,730	46

## 4.2 RT-qPCR-Validierung für zuverlässige Referenzgene

Bei der RT-qPCR-Validierung fiel *Fkbp1a* dadurch auf, dass es in 6 von 60 Lungenproben (Platte A und B) nicht detektiert werden konnte (Cq-Werte nicht bestimmbar). Aufgrund von dieser Nicht-Detektierbarkeit in den RT-qPCR-Experimenten wurde *Fkbp1a* von den weiteren Datenanalysen mittels GeNorm und Normfinder ausgeschlossen. *Psm4* wies im kleinen Anteil der untersuchten Lungenproben (13 von insgesamt 86 Lungenproben) ebenfalls sehr niedriges Expressionslevel auf, so dass dieses Gen in ähnlicher Experimentalbedingung wie der Vorliegenden kein bewährtes Referenzgen sein könnte, obwohl dieses Gen bis zu Datenanalysen mittels GeNorm und Normfinder berücksichtigt wurde. Folglich bestehen weiterhin 23 finale Referenzgenkandidaten.

Zur Bestimmung des Expressionsprofils der finalen Referenzgenkandidaten wurde eine Reihe von RT-qPCR-Experimenten durchgeführt. Darauffolgend erfolgten Analysen der Cq-Rohdaten mittels GeNorm und Normfinder zur Identifizierung der zuverlässigen Referenzgene hinsichtlich biologischer, technischer und experimenteller Faktoren, d.h. Entwicklungsstadien, Geschlecht, RNA-Qualität sowie Hyperoxie-Exposition.

### 4.2.1 Zuverlässige Referenzgene unter physiologischen Bedingungen

Zur Identifizierung zuverlässiger Referenzgene unter physiologischer Bedingung wurde die Analysegruppe (FiO<sub>2</sub>=0,21, PND2.5-28.5, RIN>5; n=17; cDNA-Nr. 1, 3-6, 8, 18-20, 22, 25, 79, 80,

83, 94, 97, 104) jeweils durch GeNorm und Normfinder analysiert. Eine Übersicht der verwendeten RNAs bzw. cDNAs ist in Anhang A (**Tab.A1**) zu finden.

Mit GeNorm wurden insgesamt sieben Gene aus den 23 Referenzgenkandidaten identifiziert, die jeweils einen GeNorm-M-Wert von <0,5 aufwiesen. Dabei stellen *Nupl2* und *Rpl4* die Referenzgene stabilster Expression mit einem M-Wert von 0,280 dar, gefolgt von *Csnk1a1*, *Maea*, *Eif3a*, *Elac2* und *Dolpp1* in der Reihenfolge zunehmender M-Werte und somit abnehmender Expressionsstabilität.

Mit Normfinder wurden unter den gleichen Experimentalbedingungen insgesamt acht Gene aus den 23 Referenzgenkandidaten identifiziert, die jeweils einen Normfinder-SV-Wert von <0,5 aufwiesen. Dabei wurde *Maea* als das am stabilsten exprimierte Referenzgen mit einem SV-Wert von 0,205 gefunden, gefolgt von *Csnk1a1*, *Nupl2*, *Rpl4*, *Zc3h11a*, *Eif3a*, *Pcsk7* und *Dolpp1* in der Reihenfolge zunehmender SV-Werte und somit abnehmender Expressionsstabilität (**Tab. 4-7**).

**Tab. 4-7 GeNorm- und Normfinder-Ergebnis für physiologische Bedingung.** Die Stabilitätswerte für die Referenzgenkandidaten unter physiologischer Bedingung (FiO<sub>2</sub>=0,21, PND2.5-28.5, RIN>5; n=17) sind dargestellt. Die Stabilitätswerte <0,5 sind in Fettdruck hervorgehoben.

Genname	GeNorm		Normfinder	
	M-Wert	Ranking	SV-Wert	Ranking
<i>Nupl2</i>	<b>0,280</b>	1	<b>0,311</b>	3
<i>Rpl4</i>	<b>0,280</b>	1	<b>0,329</b>	4
<i>Csnk1a1</i>	<b>0,375</b>	3	<b>0,278</b>	2
<i>Maea</i>	<b>0,401</b>	4	<b>0,205</b>	1
<i>Eif3a</i>	<b>0,431</b>	5	<b>0,409</b>	6
<i>Elac2</i>	<b>0,468</b>	6	0,527	9
<i>Dolpp1</i>	<b>0,498</b>	7	<b>0,496</b>	8
<i>Pcsk7</i>	0,513	8	<b>0,476</b>	7
<i>Zc3h11a</i>	0,528	9	<b>0,368</b>	5
<i>Auh</i>	0,541	10	0,553	10
<i>Kdm3b</i>	0,555	11	0,608	11
<i>Psmc4</i>	0,568	12	0,616	12
<i>Efh2</i>	0,589	13	0,810	15
<i>Tbp</i>	0,637	14	0,729	13
<i>Wars2</i>	0,676	15	0,774	14
<i>Rexo5</i>	0,714	16	0,868	17
<i>Tnks2</i>	0,748	17	0,850	16
<i>Tfrc</i>	0,778	18	0,897	18
<i>Tmed2</i>	0,806	19	1,003	20
<i>Eif4g2</i>	0,836	20	0,993	19
<i>Puf60</i>	0,884	21	1,341	21
<i>Grn</i>	0,937	22	1,452	22
<i>Mpv17l</i>	1,053	23	2,444	23

#### 4.2.2 Einfluss der neonatalen Lungenentwicklung auf die Expression von Referenzgenkandidaten

Zur Identifizierung stabil exprimierter Referenzgene während der neonatalen Lungenentwicklung wurden die Analysegruppen PND2.5 (n=6; cDNA-Nr. 1, 3-6, 8), PND5.5 (n=7; cDNA-Nr. 18-20, 22, 25, 79, 80) und PND14.5-28.5 (n=4; cDNA-Nr. 83, 94, 97, 104) unter physiologischen Experimentalbedingungen (FiO<sub>2</sub>=0,21) und guter RNA-Qualität (RIN>5) jeweils durch GeNorm und Normfinder analysiert.

Mit GeNorm wurden neun Gene aus den 23 Referenzgenkandidaten identifiziert, die in allen drei Altersgruppen einen GeNorm-M-Wert von <0,5 aufwiesen. Von diesen neun Genen stellte sich

*Elac2* als das am stabilsten exprimierte Referenzgen mit dem niedrigsten Mittelwert der drei Rankings für die Einzelgruppen heraus (**Tab. 4-8**).

Mit Normfinder wurden fünf Gene aus den 23 Referenzgenkandidaten identifiziert, die in allen drei Altersgruppen einen Normfinder-SV-Wert von <0,5 aufwiesen. Von diesen fünf Genen zeichneten sich *Maea* und *Nupl2* als die am stabilsten exprimierten Referenzgene mit dem niedrigsten Mittelwert der drei Rankings für die Einzelgruppen aus (**Tab. 4-9**).

**Tab. 4-8 GeNorm-Ergebnis zur Referenzgenexpression während der Lungenentwicklung.** Die M-Werte für die Referenzgenkandidaten zu unterschiedlichen Zeitpunkten der neonatalen Entwicklung, d.h. PND2.5 (n=6), PND5.5 (n=7), PND14.5-28.5 (n=3) ansonsten unter physiologischer Bedingung (FiO<sub>2</sub>=0,21) und mit guter RNA-Qualität (RIN>5) sind dargestellt. Die Stabilitätswerte <0,5 sowie die Gene mit einem M-Wert <0,5 in allen Gruppen sind in Fettdruck hervorgehoben. Das am stabilsten exprimierte Referenzgen war *Elac2* mit M-Werten <0,5 in allen drei Gruppen und dem niedrigsten Mittelwert der drei Stabilitätsrankings für die Einzelgruppen (grau hervorgehoben).

Genname	PND2.5		PND5.5		PND14.5-28.5		Ranking Mittelwert
	M-Wert	Ranking	M-Wert	Ranking	M-Wert	Ranking	
<i>Elac2</i>	<b>0,061</b>	1	<b>0,284</b>	4	<b>0,303</b>	7	4
<i>Nupl2</i>	<b>0,141</b>	5	<b>0,314</b>	6	<b>0,221</b>	4	5
<i>Rpl4</i>	<b>0,222</b>	10	<b>0,36</b>	8	<b>0,152</b>	1	6
<i>Pcsk7</i>	<b>0,061</b>	1	<b>0,132</b>	1	0,514	19	7
<i>Psm4</i>	<b>0,278</b>	14	<b>0,334</b>	7	<b>0,189</b>	3	8
<i>Dolpp1</i>	<b>0,125</b>	4	<b>0,132</b>	1	0,558	21	9
<i>Kdm3b</i>	<b>0,165</b>	6	<b>0,253</b>	3	0,5	18	9
<i>Maea</i>	<b>0,292</b>	15	<b>0,412</b>	11	<b>0,239</b>	5	10
<i>Grn</i>	<b>0,097</b>	3	0,814	21	<b>0,376</b>	9	11
<i>Tmed2</i>	<b>0,248</b>	12	0,723	19	<b>0,152</b>	1	11
<i>Auh</i>	<b>0,204</b>	8	<b>0,463</b>	13	<b>0,48</b>	16	12
<i>Csnk1a1</i>	<b>0,317</b>	17	<b>0,382</b>	9	<b>0,42</b>	11	12
<i>Eif4g2</i>	<b>0,236</b>	11	0,619	17	<b>0,347</b>	8	12
<i>Wars2</i>	<b>0,189</b>	7	0,585	16	<b>0,459</b>	14	12
<i>Efhd2</i>	<b>0,265</b>	13	<b>0,295</b>	5	0,532	20	13
<i>Tbp</i>	<b>0,215</b>	9	0,551	15	<b>0,469</b>	15	13
<b>Zc3h11a</b>	<b>0,331</b>	18	<b>0,4</b>	10	<b>0,432</b>	12	13
<i>Rexo5</i>	<b>0,343</b>	19	0,505	14	<b>0,401</b>	10	14
<i>Tfrc</i>	<b>0,304</b>	16	0,771	20	<b>0,261</b>	6	14
<i>Eif3a</i>	<b>0,355</b>	20	<b>0,429</b>	12	<b>0,49</b>	17	16
<i>Tnks2</i>	<b>0,372</b>	21	0,665	18	<b>0,444</b>	13	17
<i>Puf60</i>	<b>0,414</b>	22	0,881	22	0,586	22	22
<i>Mpv17l</i>	0,509	23	0,966	23	0,737	23	23

**Tab. 4-9 Normfinder-Ergebnis zur Referenzgenexpression während der Lungenentwicklung.** Die Normfinder-SV-Werte für die gleiche Experimentalbedingung wie GeNorm oben sind dargestellt. Die SV-Werte <0,5 sowie die Gene mit einem SV-Wert <0,5 in allen drei Gruppen sind in Fettdruck hervorgehoben. Die am stabilsten exprimierten Referenzgene waren *Maea* und *Nupl2* mit SV-Werten <0,5 in allen drei Gruppen und dem niedrigsten Mittelwert der drei Stabilitätsrankings für die Einzelgruppen (grau hervorgehoben).

Genname	PND2.5		PND5.5		PND14.5-28.5		Ranking Mittelwert
	M-Wert	Ranking	M-Wert	Ranking	M-Wert	Ranking	
<b><i>Maea</i></b>	<b>0,366</b>	15	<b>0,106</b>	2	<b>0,086</b>	1	6
<b><i>Nupl2</i></b>	<b>0,216</b>	8	<b>0,395</b>	6	<b>0,254</b>	4	6
<b><i>Csnk1a1</i></b>	<b>0,217</b>	9	<b>0,275</b>	3	<b>0,314</b>	8	7
<i>Eif4g2</i>	0,053	1	1,051	18	<b>0,208</b>	3	7
<b><i>Rexo5</i></b>	<b>0,259</b>	11	<b>0,381</b>	5	<b>0,336</b>	9	8
<i>Rpl4</i>	<b>0,104</b>	5	0,522	11	<b>0,278</b>	7	8
<b><i>Zc3h11a</i></b>	<b>0,355</b>	14	<b>0,106</b>	1	<b>0,365</b>	10	8
<i>Wars2</i>	<b>0,089</b>	3	0,545	12	<b>0,462</b>	14	10
<i>Eif3a</i>	<b>0,271</b>	12	0,505	9	<b>0,418</b>	12	11
<i>Elac2</i>	<b>0,376</b>	16	0,596	14	<b>0,125</b>	2	11
<i>Psm4</i>	0,509	21	<b>0,475</b>	7	<b>0,258</b>	5	11
<i>Tbp</i>	<b>0,09</b>	4	0,567	13	0,515	16	11
<i>Tmed2</i>	<b>0,053</b>	2	1,084	19	<b>0,426</b>	13	11
<i>Tfrc</i>	<b>0,168</b>	6	1,104	20	<b>0,387</b>	11	12
<i>Auh</i>	<b>0,201</b>	7	0,755	16	0,541	17	13
<i>Grn</i>	<b>0,336</b>	13	1,363	21	<b>0,27</b>	6	13
<i>Kdm3b</i>	<b>0,257</b>	10	0,519	10	0,578	18	13
<i>Dolpp1</i>	<b>0,403</b>	18	<b>0,366</b>	4	0,808	22	15
<i>Pcsk7</i>	<b>0,39</b>	17	<b>0,492</b>	8	0,704	20	15
<i>Tnks2</i>	<b>0,412</b>	19	0,762	17	<b>0,474</b>	15	17
<i>Efh2</i>	0,506	20	0,717	15	0,665	19	18
<i>Puf60</i>	0,811	22	1,537	22	0,73	21	22
<i>Mpv17l</i>	1,815	23	1,902	23	2,871	23	23

Die Intra- und Intergruppenvariationsanalyse von Normfinder zeigte *Maea* als das am stabilsten exprimierte Referenzgen mit einer neu-berechneten SD von 0,2207 sowie *Maea* und *Nupl2* als die beste Kombination der Referenzgene mit einer neu-berechneten SD von 0,1492.

#### 4.2.3 Einfluss des Geschlechts auf die Expression von Referenzgenkandidaten

Zur Identifizierung stabil exprimierter Referenzgene hinsichtlich des Geschlechtes wurden die Analysegruppen mit männlichen (n=21; cDNA-Nr. 3, 8, 11, 12, 18-20, 22, 32, 45, 47, 48, 50, 51, 60, 61, 63, 64, 73, 74, 76) bzw. weiblichen (n=20; cDNA-Nr. 1, 4-6, 10, 13, 15, 16, 25, 27-31, 46, 59, 62,

71, 72, 75) Mäusen aus ansonsten physiologischer Bedingung (FiO<sub>2</sub>=0,21) und guter RNA-Qualität (RIN>5) jeweils durch GeNorm und Normfinder analysiert.

Mit GeNorm wurden 16 Gene aus den 23 Referenzgenkandidaten identifiziert, die in beiden Geschlechtsgruppen jeweils einen GeNorm-M-Wert von <0,5 aufwiesen. Von diesen 16 Genen stellten sich *Tbp* und *Tmed2* als die am stabilsten exprimierten Referenzgene mit dem niedrigsten Mittelwert der zwei Rankings für die Einzelgruppen heraus (**Tab. 4-10**).

Mit Normfinder wurden 14 Gene aus den 23 Referenzgenkandidaten detektiert, die in beiden Geschlechtsgruppen einen Normfinder-SV-Wert von <0,5 zeigten. Von diesen 14 Genen wurde *Eif4g2* als das am stabilsten exprimierte Referenzgen mit dem niedrigsten Mittelwert der zwei Rankings für die Einzelgruppen erkannt (**Tab. 4-11**).

**Tab. 4-10 GeNorm-Ergebnis für Einfluss des Geschlechts.** Die M-Werte der Referenzgenkandidaten für unterschiedliches Geschlecht, d.h. männliche (n=21) und weibliche (n=20) Mäuse aus ansonsten physiologischer Bedingung (FiO<sub>2</sub>=0,21, RIN>5) sind dargestellt. Die M-Werte <0,5 sowie die Gene mit einem M-Wert <0,5 in beiden Geschlechtsgruppen sind in Fettdruck hervorgehoben. Die besten Referenzgene waren *Tbp* und *Tmed2* mit M-Werten <0,5 in beiden Gruppen und dem niedrigsten Mittelwert der zwei Stabilitätsrankings für die Einzelgruppen (grau hervorgehoben).

Genname	Männlich		Weiblich		Ranking Mittelwert
	M-Wert	Ranking	M-Wert	Ranking	
<b><i>Tbp</i></b>	<b>0,182</b>	1	<b>0,255</b>	3	2
<b><i>Tmed2</i></b>	<b>0,193</b>	3	<b>0,193</b>	1	2
<i>Csnk1a1</i>	<b>0,199</b>	4	<b>0,193</b>	1	2,5
<i>Eif4g2</i>	<b>0,182</b>	1	<b>0,27</b>	4	2,5
<i>Nupl2</i>	<b>0,25</b>	6	<b>0,288</b>	5	5,5
<i>Rpl4</i>	<b>0,226</b>	5	<b>0,349</b>	7	6
<i>Maea</i>	<b>0,276</b>	7	<b>0,37</b>	8	7,5
<i>Dolpp1</i>	<b>0,297</b>	8	<b>0,388</b>	9	8,5
<i>Auh</i>	<b>0,307</b>	9	<b>0,438</b>	12	10,5
<i>Elac2</i>	<b>0,323</b>	11	<b>0,405</b>	10	10,5
<i>Tfrc</i>	<b>0,365</b>	16	<b>0,324</b>	6	11
<i>Zc3h11a</i>	<b>0,338</b>	13	<b>0,422</b>	11	12
<i>Pcsk7</i>	<b>0,316</b>	10	0,52	18	14
<i>Efhd2</i>	<b>0,345</b>	14	<b>0,481</b>	15	14,5
<i>Grn</i>	<b>0,353</b>	15	<b>0,468</b>	14	14,5
<i>Psmc4</i>	<b>0,33</b>	12	0,533	19	15,5
<i>Wars2</i>	<b>0,406</b>	19	<b>0,451</b>	13	16
<i>Eif3a</i>	<b>0,374</b>	17	<b>0,494</b>	16	16,5
<i>Kdm3b</i>	<b>0,424</b>	20	0,506	17	18,5
<i>Tnks2</i>	<b>0,388</b>	18	0,554	20	19
<i>Puf60</i>	<b>0,44</b>	21	0,632	22	21,5
<i>Rexo5</i>	<b>0,468</b>	22	0,584	21	21,5
<i>Mpv17l</i>	0,535	23	0,735	23	23

**Tab. 4-11 Normfinder-Ergebnis für Einfluss des Geschlechts.** Die Normfinder-SV-Werte der Referenzgenkandidaten für die gleiche Experimentalbedingung wie GeNorm oben sind dargestellt. Die SV-Werte <0,5 sowie die Gene mit einem SV-Wert <0,5 in beiden Geschlechtsgruppen sind in Fettdruck hervorgehoben. Das beste Referenzgen war *Eif4g2* mit M-Werten <0,5 in beiden Gruppen und dem niedrigsten Mittelwert der zwei Stabilitätsrankings für die Einzelgruppen (grau hervorgehoben).

Genname	Männlich		Weiblich		Ranking Mittelwert
	SV-Wert	Ranking	SV-Wert	Ranking	
<b><i>Eif4g2</i></b>	<b>0,22</b>	4	<b>0,211</b>	1	2,5
<b><i>Rpl4</i></b>	<b>0,211</b>	2	<b>0,261</b>	4	3
<b><i>Maea</i></b>	<b>0,189</b>	1	<b>0,284</b>	6	3,5
<b><i>Tbp</i></b>	<b>0,24</b>	7	<b>0,214</b>	3	5
<b><i>Tfrc</i></b>	<b>0,243</b>	8	<b>0,214</b>	2	5
<b><i>Eif3a</i></b>	<b>0,225</b>	5	<b>0,365</b>	11	8
<b><i>Tmed2</i></b>	<b>0,273</b>	11	<b>0,269</b>	5	8
<b><i>Zc3h11a</i></b>	<b>0,227</b>	6	<b>0,35</b>	10	8
<b><i>Auh</i></b>	<b>0,212</b>	3	<b>0,456</b>	14	8,5
<b><i>Csnk1a1</i></b>	<b>0,259</b>	9	<b>0,322</b>	8	8,5
<b><i>Dolpp1</i></b>	<b>0,268</b>	10	<b>0,333</b>	9	9,5
<b><i>Nupl2</i></b>	<b>0,295</b>	12	<b>0,296</b>	7	9,5
<b><i>Elac2</i></b>	<b>0,423</b>	17	<b>0,455</b>	13	15
<b><i>Tnks2</i></b>	<b>0,309</b>	13	0,57	17	15
<b><i>Wars2</i></b>	<b>0,436</b>	18	<b>0,428</b>	12	15
<b><i>Pcsk7</i></b>	<b>0,389</b>	14	0,668	19	16,5
<b><i>Efhd2</i></b>	<b>0,407</b>	16	0,574	18	17
<b><i>Kdm3b</i></b>	<b>0,498</b>	20	0,508	15	17,5
<b><i>Psm4</i></b>	<b>0,401</b>	15	0,679	20	17,5
<b><i>Grn</i></b>	0,508	21	0,559	16	18,5
<b><i>Puf60</i></b>	<b>0,451</b>	19	1,033	22	20,5
<b><i>Rexo5</i></b>	0,702	22	0,831	21	21,5
<b><i>Mpv17i</i></b>	1,368	23	2,064	23	23

Die Intra- und Intergruppenvariationsanalyse von Normfinder zeigte *Eif4g2* als das am stabilsten exprimierte Referenzgen mit einer neu-berechneten SD von 0,0476 sowie *Tbp* und *Eif4g2* als die beste Kombination der Referenzgene mit einer neu-berechneten SD von 0,0346.

#### 4.2.4 Einfluss der RNA-Qualität auf die Expression von Referenzgenkandidaten

Zur Identifizierung stabil exprimierter Referenzgene in Proben mit niedriger und hoher RNA-Qualität wurden die Analysegruppen mit RIN>5 (n=17; cDNA-Nr. 1, 3-6, 8, 18-20, 22, 25, 79, 80, 83, 94, 97, 104) und RIN<5 (n=33; cDNA-Nr. 26, 35, 37, 39, 40, 43, 44, 53-58, 65-70, 84-88, 93, 95, 96, 98, 99, 100-103) aus ansonsten physiologischer Bedingung (FiO<sub>2</sub>=0,21, PND2.5-28.5) jeweils durch GeNorm bzw. Normfinder analysiert.

In der GeNorm-Auswertung gab es kein einziges Gen aus den 23 Referenzgenkandidaten, das in beiden RNA-Qualitätsgruppen einen GeNorm-M-Wert von <0,5 zeigte, sodass daraus kein zuverlässiges Referenzgen ergab (**Tab. 4-12**). Mit Normfinder wurden drei Gene aus den 23 Referenzgenkandidaten identifiziert, die jeweils einen Normfinder-SV-Wert von <0,5 in beiden RNA-Qualitätsgruppen aufwiesen. Davon erwies sich *Maea* als das am stabilsten exprimierte Referenzgen mit dem niedrigsten Mittelwert der zwei Rankings für die Einzelgruppen (**Tab. 4-13**).

**Tab. 4-12 GeNorm-Ergebnis für Einfluss der RNA-Qualität.** Die M-Werte der Referenzgenkandidaten für unterschiedliche RNA-Qualität, d.h. RIN>5 (n=17) und RIN<5 (n=33) aus ansonsten physiologischer Bedingung (FiO<sub>2</sub>=0,21, PND2.5-28.5) sind dargestellt. Die M-Werte <0,5 sind in Fettdruck hervorgehoben. Aus der GeNorm-Analyse ergab sich kein einziges Kandidatengen, das in beiden Analysegruppen einem M-Wert <0,5 zeigte, somit kein verlässliches Referenzgen in diesem Experimentalsetting.

Genname	RIN<5		RIN>5		Ranking Mittelwert
	M-Wert	Ranking	M-Wert	Ranking	
<i>Eif3a</i>	0,541	3	<b>0,431</b>	5	4
<i>Nupl2</i>	0,766	8	<b>0,28</b>	1	4,5
<i>Maea</i>	0,669	6	<b>0,401</b>	4	5
<i>Dolpp1</i>	0,64	5	<b>0,498</b>	7	6
<i>Kdm3b</i>	<b>0,469</b>	1	0,555	11	6
<i>Csnk1a1</i>	0,908	11	<b>0,375</b>	3	7
<i>Efhd2</i>	0,585	4	0,589	13	8,5
<i>Pcsk7</i>	0,868	10	0,513	8	9
<i>Zc3h11a</i>	0,827	9	0,528	9	9
<i>Psm4</i>	0,704	7	0,568	12	9,5
<i>Rpl4</i>	1,32	19	<b>0,28</b>	1	10
<i>Elac2</i>	1,099	15	<b>0,468</b>	6	10,5
<i>Eif4g2</i>	<b>0,469</b>	1	0,836	20	10,5
<i>Rexo5</i>	0,971	12	0,714	16	14
<i>Tbp</i>	1,138	16	0,637	14	15
<i>Tnks2</i>	1,059	14	0,748	17	15,5
<i>Tfrc</i>	1,016	13	0,778	18	15,5
<i>Auh</i>	1,537	22	0,541	10	16
<i>Wars2</i>	1,384	20	0,676	15	17,5
<i>Tmed2</i>	1,196	17	0,806	19	18
<i>Grn</i>	1,259	18	0,937	22	20
<i>Puf60</i>	1,453	21	0,884	21	21
<i>Mpv17l</i>	1,619	23	1,053	23	23

**Tab. 4-13 Normfinder-Ergebnis für Einfluss der RNA-Qualität.** Die Normfinder-SV-Werte der Referenzgenkandidaten für die gleiche Experimentalbedingung wie GeNorm oben sind dargestellt. Die SV-Werte <0,5 sowie die Gene mit einem SV-Wert <0,5 in beiden Gruppen der RNA-Qualität sind in Fettdruck hervorgehoben. Das beste Referenzgen war *Maea* mit M-Werten <0,5 in beiden Gruppen und dem niedrigsten Mittelwert der zwei Stabilitätsrankings für die Einzelgruppen (grau hervorgehoben).

Genname	RIN<5		RIN>5		Ranking Mittelwert
	SV-Wert	Ranking	SV-Wert	Ranking	
<b><i>Maea</i></b>	<b>0,468</b>	3	<b>0,205</b>	1	2
<b><i>Zc3h11a</i></b>	<b>0,297</b>	1	<b>0,368</b>	5	3
<i>Csnk1a1</i>	0,817	6	<b>0,278</b>	2	4
<b><i>Pcsk7</i></b>	<b>0,372</b>	2	<b>0,476</b>	7	4,5
<i>Nupl2</i>	0,845	7	<b>0,311</b>	3	5
<i>Eif3a</i>	0,752	5	<b>0,409</b>	6	5,5
<i>Dolpp1</i>	0,611	4	<b>0,496</b>	8	6
<i>Kdm3b</i>	0,986	10	0,608	11	10,5
<i>Rpl4</i>	1,449	18	<b>0,329</b>	4	11
<i>Efhd2</i>	0,97	8	0,81	15	11,5
<i>Elac2</i>	1,157	14	0,527	9	11,5
<i>Psm4</i>	0,99	11	0,616	12	11,5
<i>Tfrc</i>	0,979	9	0,897	18	13,5
<i>Tbp</i>	1,159	15	0,729	13	14
<i>Rexo5</i>	1,003	12	0,868	17	14,5
<i>Tnks2</i>	1,144	13	0,85	16	14,5
<i>Auh</i>	2,262	22	0,553	10	16
<i>Wars2</i>	1,675	19	0,774	14	16,5
<i>Eif4g2</i>	1,176	16	0,993	19	17,5
<i>Tmed2</i>	1,298	17	1,003	20	18,5
<i>Grn</i>	1,824	20	1,452	22	21
<i>Puf60</i>	1,92	21	1,341	21	21
<i>Mpv17l</i>	2,32	23	2,444	23	23

Die Intra- und Intergruppenvariationsanalyse von Normfinder zeigte *Pcsk7* als das am stabilsten exprimierte Referenzgen mit einer neu-berechneten SD von 0,1832 sowie *Rpl4* und *Zc3h11a* als die beste Kombination der Referenzgene mit einer neu-berechneten SD von 0,1460.

#### 4.2.5 Einfluss der Hyperoxie-Exposition auf die Expression von Referenzgenkandidaten

Um zuerst stabil exprimierte Referenzgene unter Hyperoxie-Bedingung selbst herauszufinden, wurde die hyperoxie-exponierte Gruppe (FiO<sub>2</sub>=0,8, PND2.5-28.5, RIN>5; n=36; cDNA-Nr. 10-13, 15, 16, 27-32, 45-48, 50, 51, 59-64, 81, 82, 89-92) jeweils durch GeNorm bzw. Normfinder analysiert.

Mit GeNorm wurden acht Gene aus den 23 Referenzgenkandidaten identifiziert, die jeweils einen GeNorm-M-Wert von  $<0,5$  aufwiesen. Davon zeichneten *Zc3h11a* und *Auh* durch die stabilste Genexpression mit einem M-Wert von 0,299 aus, gefolgt von *Rpl4*, *Dolpp1*, *Csnk1a1*, *Maea*, *Pcsk7* und *Elac2* in der Reihenfolge zunehmender M-Werten und somit abnehmender Genstabilität (**Tab. 4-14**). Mit Normfinder wurden in den gleichen Experimentalbedingungen ebenfalls acht Gene aus den 23 Referenzgenkandidaten detektiert, die jeweils einen Normfinder-SV-Wert von  $<0,5$  aufwiesen. Dabei stellte *Rpl4* als das am stabilsten exprimierte Referenzgen mit einem SV-Wert von 0,235 hervor, gefolgt von *Zc3h11a*, *Auh*, *Maea*, *Dolpp1*, *Csnk1a1*, *Nupl2* und *Eif3a* in der Reihenfolge zunehmender SV-Werten und somit abnehmender Genstabilität (**Tab. 4-15**).

In der darauffolgenden Analyserunde wurden das GeNorm- bzw. Normfinder-Ergebnis unter Hyperoxie-Bedingung mit dem unter physiologischer Bedingung verglichen, damit der Einfluss der Hyperoxie-Exposition auf Expressionsstabilität der 23 Referenzgenkandidaten untersucht werden kann. Mit GeNorm wurden fünf Gene aus den 23 Referenzgenkandidaten erkannt, deren GeNorm-M-Werte in beiden Experimentalbedingungen, d.h. physiologischer und hyperoxie-exponierter Bedingung  $<0,5$  betragen. Davon erwies sich *Rpl4* als das am stabilsten exprimierte Referenzgen mit dem niedrigsten Mittelwert der zwei Rankings für die Einzelgruppen (**Tab. 4-14**). Mit Normfinder wurden sieben Gene aus den 23 Referenzgenkandidaten detektiert, die jeweils einen Normfinder-SV-Wert von  $<0,5$  in beiden Experimentalgruppen aufwiesen. Davon stellten *Maea* und *Rpl4* als die am stabilsten exprimierten Referenzgene mit dem niedrigsten Mittelwert der zwei Rankings für die Einzelgruppen dar (**Tab. 4-15**).

**Tab. 4-14 GeNorm-Ergebnis für Einfluss der Hyperoxie-Exposition.** Die M-Werte der Referenzgenkandidaten für physiologische (n=17) und Hyperoxie-Gruppe (n=36) Gruppe aus ansonsten konstant gehaltener Bedingung (PND2.5-28.5, RIN>5) sind dargestellt. Die M-Werte <0,5 sowie die Gene mit einem M-Wert <0,5 in beiden Gruppen sind in Fettdruck hervorgehoben. Die besten Referenzgene war *Rpl4* mit M-Werten <0,5 in beiden Gruppen und dem niedrigsten Mittelwert der zwei Stabilitätsrankings für die Einzelgruppen (grau hervorgehoben).

Genname	Hyperoxie		Normoxie		Ranking Mittelwert
	M-Wert	Ranking	M-Wert	Ranking	
<b><i>Rpl4</i></b>	<b>0,34</b>	3	<b>0,28</b>	1	2
<b><i>Csnk1a1</i></b>	<b>0,407</b>	5	<b>0,375</b>	3	4
<i>Zc3h11a</i>	<b>0,299</b>	1	0,528	9	5
<b><i>Maea</i></b>	<b>0,436</b>	6	<b>0,401</b>	4	5
<i>Nupl2</i>	0,51	9	<b>0,28</b>	1	5
<i>Auh</i>	<b>0,299</b>	1	0,541	10	5,5
<b><i>Dolpp1</i></b>	<b>0,372</b>	4	<b>0,498</b>	7	5,5
<b><i>Elac2</i></b>	<b>0,496</b>	8	<b>0,468</b>	6	7
<i>Pcsk7</i>	<b>0,474</b>	7	0,513	8	7,5
<i>Eif3a</i>	0,535	10	<b>0,431</b>	5	7,5
<i>Psmc4</i>	0,562	11	0,568	12	11,5
<i>Efh2</i>	0,601	13	0,589	13	13
<i>Wars2</i>	0,631	14	0,676	15	14,5
<i>Tbp</i>	0,669	15	0,637	14	14,5
<i>Kdm3b</i>	0,745	18	0,555	11	14,5
<i>Eif4g2</i>	0,583	12	0,836	20	16
<i>Tnks2</i>	0,7	16	0,748	17	16,5
<i>Tfrc</i>	0,722	17	0,778	18	17,5
<i>Rexo5</i>	0,773	19	0,714	16	17,5
<i>Tmed2</i>	0,797	20	0,806	19	19,5
<i>Puf60</i>	0,843	21	0,884	21	21
<i>Grn</i>	0,89	22	0,937	22	22
<i>Mpv17l</i>	0,95	23	1,053	23	23

**Tab. 4-15 Normfinder-Ergebnis für Einfluss der Hyperoxie-Exposition.** Die SV-Werte der Referenzgenkandidaten für die gleiche Bedingung wie GeNorm oben sind dargestellt. Die SV-Werte <0,5 sowie die Gene mit einem SV-Wert <0,5 in beiden Gruppen sind in Fettdruck hervorgehoben. Die besten Referenzgene waren *Maea* und *Rpl4* mit SV-Werten <0,5 in beiden Gruppen und dem niedrigsten Mittelwert der zwei Stabilitätsrankings für die Einzelgruppen (grau hervorgehoben).

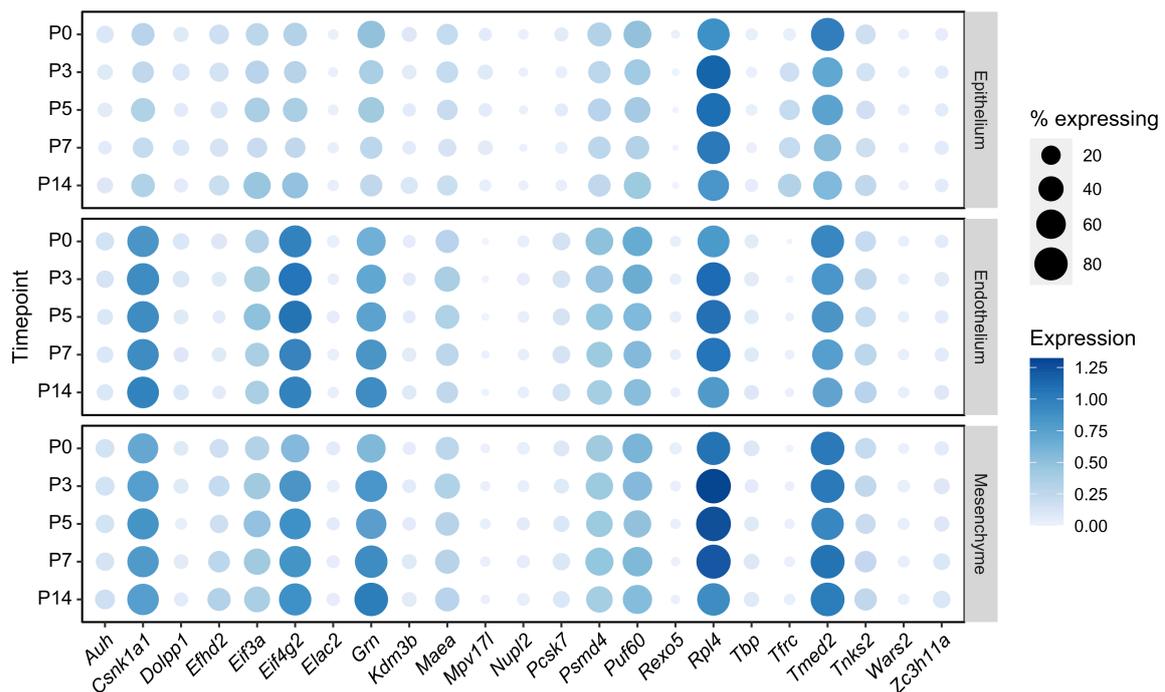
Genname	Hyperoxie		Normoxie		Ranking Mittelwert
	SV-Wert	Ranking	SV-Wert	Ranking	
<b><i>Rpl4</i></b>	<b>0,235</b>	1	<b>0,329</b>	4	2,5
<b><i>Maea</i></b>	<b>0,314</b>	4	<b>0,205</b>	1	2,5
<i>Zc3h11a</i>	<b>0,253</b>	2	<b>0,368</b>	5	3,5
<b><i>Csnk1a1</i></b>	<b>0,333</b>	6	<b>0,278</b>	2	4
<b><i>Nupl2</i></b>	<b>0,465</b>	7	<b>0,311</b>	3	5
<i>Auh</i>	<b>0,301</b>	3	0,553	10	6,5
<b><i>Dolpp1</i></b>	<b>0,322</b>	5	<b>0,496</b>	8	6,5
<b><i>Eif3a</i></b>	<b>0,476</b>	8	<b>0,409</b>	6	7
<i>Pcsk7</i>	0,529	10	<b>0,476</b>	7	8,5
<i>Elac2</i>	0,523	9	0,527	9	9
<i>Wars2</i>	0,547	11	0,774	14	12,5
<i>Tbp</i>	0,726	14	0,729	13	13,5
<i>Tnks2</i>	0,698	12	0,85	16	14
<i>Psmc4</i>	0,743	16	0,616	12	14
<i>Kdm3b</i>	0,872	18	0,608	11	14,5
<i>Eif4g2</i>	0,718	13	0,993	19	16
<i>Efh2</i>	0,815	17	0,81	15	16
<i>Tfrc</i>	0,735	15	0,897	18	16,5
<i>Rexo5</i>	0,884	19	0,868	17	18
<i>Tmed2</i>	0,913	20	1,003	20	20
<i>Puf60</i>	1,224	21	1,341	21	21
<i>Grn</i>	1,39	22	1,452	22	22
<i>Mpv17l</i>	1,517	23	2,444	23	23

Die Intra- und Intergruppenvariationsanalyse von Normfinder zeigte *Rpl4* als das am stabilsten exprimierte Referenzgen mit einer neu-berechneten SD von 0,1004 sowie *Maea* und *Zc3h11a* als die beste Kombination der Referenzgene mit einer neu-berechneten SD von 0,0711.

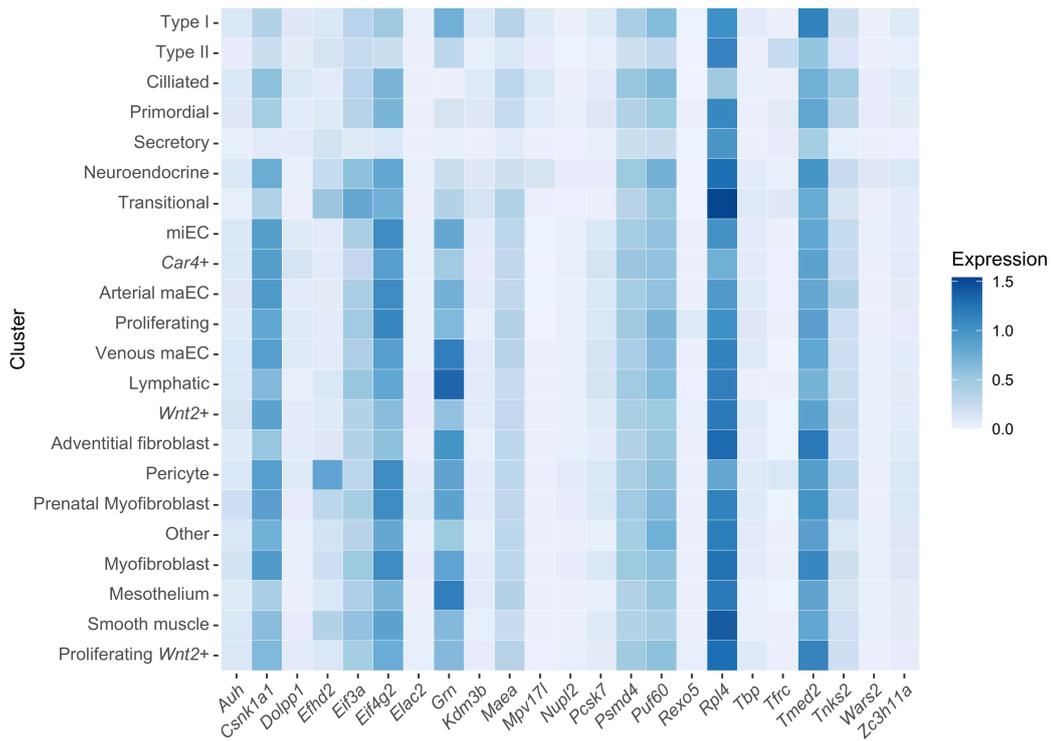
### 4.3 Einzelzell-RNA-Sequenzierung für zuverlässige Referenzgene

Expression der Referenzgenkandidaten wurde mittels Einzelzell-RNA-Sequenzierung (scRNA-Seq) in diversen pulmonalen Zelltypen bei der neonatalen Lungenentwicklung evaluiert. Genexpressionsanalyse in drei wichtigsten Zelltypen (Epithel, Endothel, Mesenchym) bei der neonatalen Lungenentwicklung (PND0-14) identifizierte *Csnk1a1*, *Eif4g2*, *Grn*, *Rpl4* und *Tmed2* als

Referenzgenkandidaten mit hohem Expressionslevel, wobei die ersteren drei Kandidatengene in Epithelzellen etwas schwächer exprimiert waren (**Abb. 4-1**). Ferner wurde die hohe Expression von diesen fünf Referenzgenkandidaten auch in weiteren diversen Zelltypen bestätigt (**Abb. 4-2**). Die Variation in Genexpression der Referenzgenkandidaten in scRNA-Seq-Datensatz wurde mittels SCTransform-Normalisierung berechnet und in eine Rangordnung gebracht. Dabei erwiesen sich *Rexo5*, *Mpv17l*, *Wars2*, *Nupl2* und *Elac2* als die am stabilsten exprimierten Referenzgenkandidaten mit niedrigen Expressionsvariationen, jedoch waren deren Expressionslevel niedrig im Vergleich zu anderen Referenzgenkandidaten (**Abb. 4-1**).



**Abb. 4-1 Expression der Referenzgenkandidaten in drei Hauptzelltypen bei neonataler Lungenentwicklung von C57BL/6-Mäusen.** Man erkennt, dass *Csnk1a1*, *Eif4g2*, *Grn*, *Rpl4* und *Tmed2* in diesem Experimentalsetting hoch exprimiert sind, wobei die ersteren drei Gene in Epithelzellen etwas geringer exprimiert sind. Die Größe des Kreises bezeichnet den Anteil der Zellen mit einem Expressionslevel oberhalb der Nachweisgrenze und die Farbintensität entspricht dem durchschnittlichen Expressionslevel. P: postnataler Tag. Grafik mit freundlicher Genehmigung vom Kooperationspartner Prof. Sucre, Pädiatrie, Vanderbilt University Medical Center, USA.



**Abb. 4-2 Expressionsprofil der Referenzgenkandidaten in weiteren pulmonalen Zelltypen.** Durchschnittliches Expressionslevel in diversen Zelltypen der Mauslunge in PND0-14. Je dunkler, desto höher das Expressionsniveau. Grafik mit freundlicher Genehmigung vom Kooperationspartner Prof. Sucre, Pädiatrie, Vanderbilt University Medical Center, USA.

## 5 Diskussion

Mit zunehmender Bedeutung an biologisch-medizinischen Forschungen auf Genebene spielt die quantitative Echtzeit-PCR (qPCR) als eine der am häufigsten eingesetzten Methoden für Genexpressionsanalyse – also bei der Bestimmung der Aktivität von Genen – eine große Rolle. Mit ihren zahlreichen Vorteilen ist die qPCR mit vorgeschalteter reverser Transkription (RT-qPCR) eine bewährte Methode, wenn das Expressionsprofil einer begrenzten Anzahl von Genen in zahlreichen Proben bestimmt werden soll.

Die relative Quantifizierung mit Hilfe von Referenzgenen als interne Kontrolle ist eine bevorzugte Methode zur Analyse und Interpretation der RT-qPCR-Rohdaten. Durch Einbezug dieser internen Kontrolle können unspezifische Variationen der Genexpression, bedingt z.B. durch unvermeidbare Unterschiede bei der Probenextraktion, RNA-Aufbereitung, cDNA-Herstellung und Durchführung der RT-qPCR an sich, effektiv ausgeglichen werden. Die zu diesem Zweck verwendeten Referenzgene weisen idealerweise eine stabile Genexpression, d.h. möglichst niedrige Standardabweichung des Expressionslevels, in den zu untersuchenden Proben unter den gegebenen experimentellen Bedingungen aufweisen.

Früher wurde angestrebt, ideale Referenzgene mit ubiquitär konstanter Genexpression, unabhängig von untersuchten Organen und Experimentalbehandlungen zu verwenden. In dem Sinne wurden sog. Housekeeping-Gene bevorzugt als Referenzgene eingesetzt, weil sie für essenzielle Grundfunktion der Zellen verantwortlich sind und somit davon ausgegangen wurde, dass sie ubiquitär stabil exprimiert sein müssten. Weiterführende Studien belegten allerdings, dass die Expressionsstabilität dieser Gruppe von Genen unter bestimmten experimentellen Bedingungen doch einer starken Regulation unterliegen kann und ihr Einsatz daher stark vom experimentellen Setting abhängt. Außerdem konnte aus solchen Studien eine gemeinsame Schlussfolgerung abgeleitet werden, dass es keine ubiquitär stabil exprimierten idealen Referenzgene existieren würden.<sup>66</sup>

Mausmodelle werden bevorzugt bei der Erforschung von humanen Erkrankungen hinsichtlich deren Pathogenese und Therapiemöglichkeiten eingesetzt, weil Mäuse außer der genetischen Ähnlichkeit mit Menschen gentechnische Modifizierungsmöglichkeiten anbieten und einen kurzen Lebenszyklus aufweisen, sodass Forschungsergebnisse mit Mäusen in relativ kurzer Zeit erzielt werden können.

In dem Sinne sind Mauslungen ein geeignetes Material zur Erforschung humaner Lungenerkrankungen. Für das Lungengewebe vor allem aus neonatalen Mäusen ist die Auswahl geeigneter Referenzgene besonders herausfordernd, weil diverse Gene während der neonatalen Lungenenwicklung unterschiedlich reguliert und exprimiert werden. Hinzu kommen noch unterschiedliche Effekte durch Lungenschädigungen im jeweils verwendeten Mausmodell.

Im vorliegenden Projekt wurden stabil exprimierte Referenzgene für die Lunge neonataler Mäuse zur Anwendung bei der RT-qPCR identifiziert und validiert. Dabei wurden biologische (Entwicklungsstadien, Geschlecht), technische (RNA-Qualität) und experimentelle (Hyperoxie-Exposition) Faktoren sowie diverse pulmonale Zelltypen berücksichtigt.

Dafür wurden zuerst geeignete Referenzgenkandidaten aus publizierten RT-qPCR-Studien und Evaluierungsstudien zu Referenzgenen sowie publizierte Microarray-Daten (Genevestigator) vorausgewählt. Dabei wurde festgestellt, dass in den genannten Datenquellen Referenzgene aus neonatalen Mäusen (bis zu einem Monat nach Geburt) unterrepräsentiert sind. Diese mangelhafte Datenlage für neonatale Mäuse unterstreicht einerseits die Notwendigkeit unseres Projektes, zuverlässige Referenzgene für neonatale Mäuse zu identifizieren und zu validieren, aber stellte andererseits eine Einschränkung bei der Vorauswahl potenziell geeigneter Kandidatengene für neonatale Mäuse dar. Somit stammen unsere Referenzgenkandidaten in der Vorauswahl größtenteils aus adulten Mäusen. Aus diesem Grund wurden unsere eigenen Microarray-Daten aus neonatalen Mäusen im Alter von PND5-7 herangezogen, um in der Vorauswahl auch über für neonatale Mäuse gut geeignete Referenzgenkandidaten zu verfügen.

Zur Identifizierung stabil exprimierter Referenzgene bei unterschiedlichen experimentellen Bedingungen wurden die zwei gängigsten Analysealgorithmen, GeNorm und Normfinder, angewendet. In den durchgeführten Datenanalysen mittels dieser Analysealgorithmen gab es kein einziges Gen von den Referenzgenkandidaten, das bei allen Experimentalbedingungen eine stabile Expression zeigte und somit als das universelle ideale Referenzgen dienen könnte. Vielmehr konnten zuverlässige Referenzgene spezifisch für die jeweiligen untersuchten Experimentalbedingungen identifiziert werden (**Abb. 5-1**).



**Abb. 5-1 Zuverlässige Referenzgene in der RT-qPCR-Validierung bezüglich untersuchten Experimentalbedingungen.** Die fett hervorgehobenen Gene wurden in beiden Analysealgorithmen jeweils für denselben Analyseansatz gleichzeitig als zuverlässige Referenzgene identifiziert, während die farbig hervorgehobenen Gene entweder nur in der GeNorm- (grün) oder nur in der Normfinder-Analyse (rot) als am stabilsten exprimierte Gene erkannt wurden. *Pcsk7* (blau) war zwar in beiden Analysealgorithmen das am stabilsten exprimierte Referenzgen, jedoch jeweils in unterschiedlichen Analyseansätzen.

In den Einzelgruppenanalysen, d.h. nur in der physiologischen oder hyperoxischen Bedingungen zur Suche nach stabil exprimierten Referenzgenkandidaten brachten GeNorm und Normfinder vergleichbare Ergebnisse, so dass für diese Analysegruppen jeweils sechs Gene von beiden Analysealgorithmen gemeinsam als stabil exprimierte Referenzgene identifiziert wurden. Diese Referenzgene waren für die physiologische Gruppe *Csnk1a1*, *Dolpp1*, *Eif3a*, *Maea*, *Nupl2*, *Rpl4* und für die hyperoxische Gruppe *Auh*, *Csnk1a1*, *Dolpp1*, *Maea*, *Rpl4*, *Zc3h11a* (**Abb. 5-1** fett hervorgehobene Gene jeweils im roten und blauen Kreis). Davon waren *Nupl2*, *Rpl4* (GeNorm) und *Maea* (Normfinder) für die physiologische Gruppe die am stabilsten exprimierten Referenzgene, wohingegen *Auh*, *Zc3h11a* (GeNorm) und *Rpl4* (Normfinder) in der hyperoxischen Gruppe die

stabilste Expression zeigten. Da diese Gene im jeweiligen experimentellen Setting von zwei unabhängigen und auf unterschiedlichen Prinzipien beruhenden Analysealgorithmen gleichzeitig als stabil exprimierte Referenzgene anerkannt wurden, sind sie als zuverlässige Referenzgene zu interpretieren.

In den vergleichenden Analysen ergaben sich anhand Stabilitätsrankings zwischen unterschiedlichen Experimentalgruppen, d.h. Entwicklungsstadien, Geschlecht, RNA-Qualität, Hyperoxie-Exposition lieferten GeNorm und Normfinder zum Teil voneinander abweichende Ergebnisse. So ergaben sich *Elac2* (GeNorm) und *Maea*, *Nupl2* (Normfinder) als die am stabilsten exprimierten Referenzgene hinsichtlich unterschiedlichen Entwicklungsstadien (**Abb. 5-1** die Gene im orangefarbenen Kreis). In zwei unterschiedlichen Geschlechtern waren *Tbp*, *Tmed2* (GeNorm) und *Eif4g2* (Normfinder) am stabilsten exprimiert (**Abb. 5-1** die Gene im grauen Kreis). Hinsichtlich der unterschiedlicher RNA-Qualität konnte GeNorm kein einziges Gen detektieren, das in beiden RNA-Qualitätsgruppen einen akzeptablen M-Wert von  $<0,5$  zeigte und somit waren alle Referenzgenkandidaten nicht geeignet als zuverlässige Referenzgene für dieses experimentelle Setting. Dennoch identifizierte Normfinder in der gleichen experimentellen Bedingung *Maea* als das am stabilsten exprimierte Referenzgen (**Abb. 5-1** die Gene im gelben Kreis). Im Vergleich zwischen der physiologischen und hyperoxischen Gruppe zeigten *Rpl4* (GeNorm) und *Maea*, *Rpl4* (Normfinder) die stabilsten Genexpression und damit zeigten GeNorm und Normfinder hier vergleichbare Ergebnisse (**Abb. 5-1** die Gene im grünen Kreis).

Die Intra- und Intergruppenvariationsanalysen von Normfinder brachten für die jeweiligen Analyseansätze zum Gruppenvergleich konkordante Ergebnisse wie die auf Stabilitätsrankings basierten Analysen mittels Normfinder (**Tab. 5-1**). Lediglich aus der Analyse der unterschiedlichen RNA-Qualitätsgruppen ergaben sich inkonsistente Ergebnisse. So wäre *Maea* der am stabilsten exprimierte Referenzgenkandidat laut der auf Stabilitätsranking basierten Analyse, während die Intra- und Intergruppenvariationsanalyse *Pcsk7* als das am stabilsten exprimierte Gen, *Rpl4* und *Zc3h11a* als die beste Kombination identifizierte. Diese widersprüchlichen Ergebnisse innerhalb der Normfinder-Analyse sowie der weiter oben erwähnte Fehlschlag der GeNorm-Analyse hinsichtlich der unterschiedlichen RNA-Qualitätsgruppen könnten auf einen maßgebenden Einfluss der niedrigen RNA-Qualität ( $RIN < 5$ ) auf Expressionsstabilität der Referenzgenkandidaten hinweisen. Außerdem unterstützen diese Ergebnisse die allgemein bekannte Tatsache, dass eine hohe RNA-Qualität vor Beginn der RT-qPCR-Experimente gewährleistet sein soll.

**Tab. 5-1 Intra- und Intergruppenvariationsanalyse von Normfinder.** Das Ergebnis für jeweilige Analyseansätze sind vergleichbar wie die auf Stabilitätsranking basierte Analyse von Normfinder außer dem Analyseansatz hinsichtlich der RNA-Qualität. RG steht für Referenzgen(e).

Analyseansätze	Normfinder Intra- und Intergruppenvariationsanalyse	
	Das am stabilsten exprimierte RG	Die beste RG-Kombination
Entwicklungsstadien	<i>Maea</i>	<i>Maea and Nupl2</i>
Geschlecht	<i>Eif4g2</i>	<i>Eif4g2 and Tbp</i>
RNA-Qualität	<i>Pcsk7</i>	<i>Rpl4 and Zc3h11a</i>
Hyperoxia-Exposition vs physiologische Bedingung	<i>Rpl4</i>	<i>Maea and Zc3h11a</i>

In der RT-qPCR-Validierung gab es mehrere bemerkenswerte Ergebnisse. Das *Tbp* aus dem Roche-Referenzgenpanel in der Vorauswahl ist eins der gut bekannten, konventionell verwendeten Referenzgene, dessen Expressionsstabilität als Housekeeping-Gen in früheren Studien in Frage gestellt wurde.<sup>97,98</sup> Dieses Gen zeigte allerdings in unserer Analyse, zumindest hinsichtlich der unterschiedlichen Geschlechter die beste Expressionsstabilität, so dass es in diesem experimentellen Setting durchaus noch als zuverlässiges Referenzgen eingesetzt werden kann. Aus dieser Feststellung kann ferner erschlossen werden, dass sich die konventionell verwendeten Referenzgene, deren Expressionsstabilität generell bezweifelt wurden, unter bestimmten Experimentalbedingungen noch als zuverlässige Referenzgene einsetzen lassen.

Für das *Rpl4*, eins der für die ribosomalen Proteine kodierenden Gene, gab es in der Vorauswahl zwei einander widersprechende Evaluierungsstudien. Die eine Studie<sup>99</sup> mit der Metaanalyse von publizierten Microarrays unterstützt den Einsatz von *Rpl4* als Referenzgen, weil dieses Gen laut statistischen Kriterien wie CV (*coefficient of variation*), MFC (*maximum fold change*) zur Beurteilung der Expressionsstabilität eines der stabil exprimierten Kandidatengene darstellte. Eine andere, etwas später erschienene Evaluierungsstudie<sup>10</sup> hat genau diese statistischen Parameter zur Einschätzung der Genexpressionsstabilität in Frage gestellt und von Anwendung der für ribosomale Proteine kodierenden Gene als Referenzgene abgeraten, wobei Expressionsschwankungen dieser Gengruppe in unterschiedlichen Gewebearten betont wurden. In unserer RT-qPCR-Validierung zeigte *Rpl4*, welches als ein Vertreter für die ribosomale Proteine kodierenden Gene repräsentativ untersucht wurde, in mehreren Analyseansätzen niedrige Expressionsvariabilität und damit Eignung als Referenzgen unter bestimmten experimentellen Bedingungen. Diese widersprüchlichen Ergebnisse könnten durch die unterschiedlichen experimentellen Bedingungen zwischen der vorliegenden Arbeit und der oben genannten Evaluierungsstudie<sup>10</sup> begründet sein z.B. Analyse der Gewebeart (in Lunge in der vorliegenden Arbeit vs. in 22 unterschiedlichen Gewebearten, Analyse in neonatalen vs. adulte Mäuse).

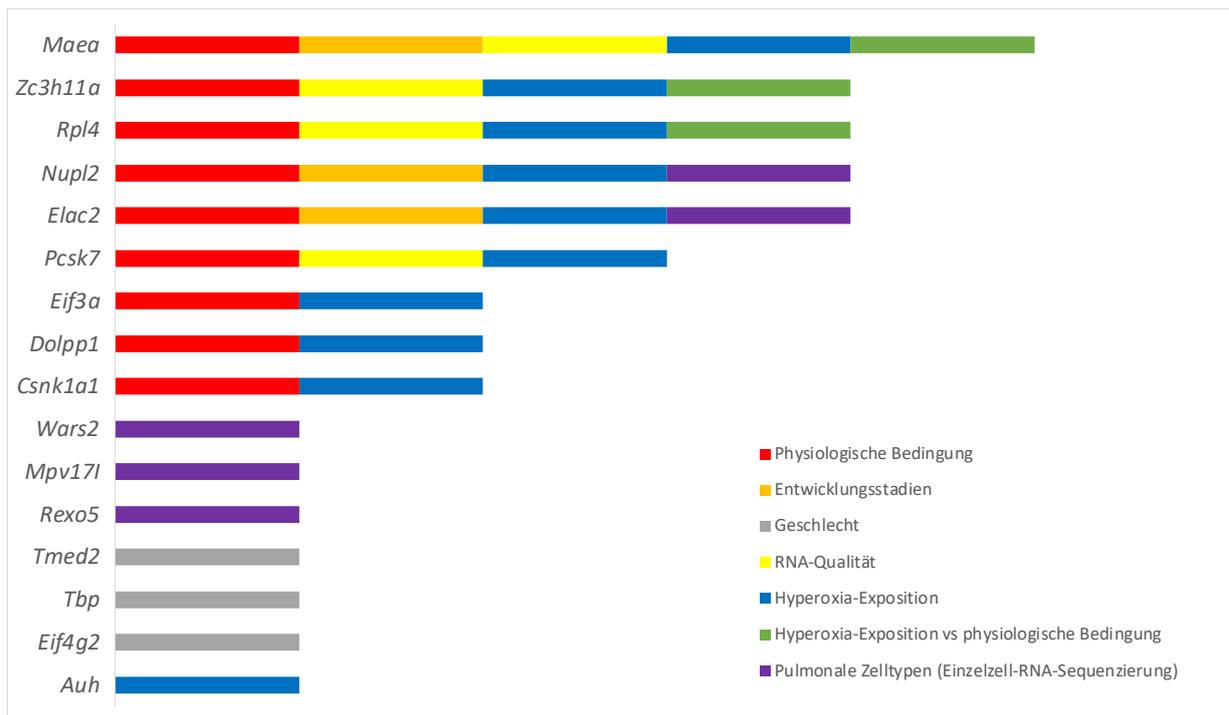
Zwischen Ergebnissen für Expressionsstabilität der Referenzgenkandidaten jeweils aus der Vorauswahl und der RT-qPCR-Validierung bestand erhebliche Diskrepanz. Als ein extremes Beispiel ist *Mpv17l* zu nennen. Dieses Gen war laut der Vorhersage unserer Vorauswahl der zweitbeste Referenzgenkandidat hinsichtlich seiner Expressionsstabilität (Overall-SD), jedoch dieses Gen zeigte unabhängig von Analyseansätzen immer die niedrigste Expressionsstabilität in der RT-qPCR-Validierung. Diese erhebliche Diskrepanz der Ergebnisse aus der Vorauswahl und der RT-qPCR-Validierung könnte dadurch erklärbar sein, dass die Vorauswahl grundsätzlich auf Microarray-Daten beruht, wobei die Validierung mittels RT-qPCR erfolgt ist. Außerdem könnten die in der Vorauswahl berücksichtigten adulten Mäuse auf diese Diskrepanz begründet haben, da die RT-qPCR-Validierung ausschließlich in neonatalen Mäusen (PND2.5-28.5) durchgeführt wurde.

Zwei von den initial ausgewählten, 24 Referenzgenkandidaten, nämlich *Fkbp1a* und *Psm2* waren von ihrem Expressionsprofil her bereits gleich nach der RT-qPCR auffällig. So war *Fkbp1a* in 6 von 60 Lungenproben gar nicht detektierbar und *Psm2* zeigte in 13 von 86 Lungenproben sehr niedrige Expression mit hohen Cq-Werten, sodass *Fkbp1a* von der weiteren Datenauswertung mittels GeNorm und Normfinder vorab rausgenommen wurde und von der Anwendung von *Psm2* als Referenzgen abzuraten gewesen wäre, wenn dieses Gen letztendlich zu einem unserer zuverlässigen Referenzgene gehört hätte. Weshalb diese zwei Gene in der RT-qPCR so niedrige Expression gezeigt haben, könnte an den experimentellen Bedingungen im Rahmen dieser Arbeit gelegen haben, die möglicherweise nicht optimal für diese beiden Gene gewesen sind. Außerdem da die sechs RNA-Proben, in denen die Expression von *Fkbp1a* nicht detektierbar war, alle eine niedrige RNA-Qualität von RIN<5 aufwiesen, wäre ein Einfluss der niedrigen RNA-Qualität auf die Expression von *Fkbp1a* möglich.

Bei der separat durchgeführten Einzelzell-RNA-Sequenzierung wurden *Rexo5*, *Mpv17l*, *Wars2*, *Nupl2*, *Elac2* hinsichtlich diverser pulmonaler Zelltypen als die am stabilsten exprimierten Referenzgene identifiziert (von *Rexo5* nach *Elac2* mit abnehmender Expressionsstabilität). Dieses Ergebnis ist zum Teil inkonsistent mit dem Ergebnis aus der RT-qPCR-Validierung. So gehörten *Rexo5*, *Mpv17l*, *Wars2* in allen Analyseansätzen der RT-qPCR-Validierung zu den eher instabil exprimierten Genen; das höchste Ranking der Expressionsstabilität, das diese drei Kandidatengene jemals erreicht haben, war das 8. (*Wars2* in der Normfinder-Analyse für Entwicklungsstadien). *Mpv17l* war sogar in allen Analyseansätzen immer am instabilsten exprimiert. Dennoch sollte eine Verwendung dieser fünf Gene in Erwägung gezogen werden, wenn die relative Quantifizierung zellspezifisch erfolgen soll.

Zusammenfassend wurden 16 von den 23 Referenzgenkandidaten in der RT-qPCR-Validierung oder der Einzelzell-RNA-Sequenzierung mindestens einmal als stabil exprimierte Referenzgene

identifiziert. Diese 16 Gene waren *Auh*, *Csnk1a1*, *Dolpp1*, *Eif3a*, *Eif4g2*, *Elac2*, *Maea*, *Mpv17l*, *Nupl2*, *Pcsk7*, *Rexo5*, *Rpl4*, *Tbp*, *Tmed2*, *Wars2*, *Zc3h11a* (Abb. 5-2).



**Abb. 5-2 Eignung der 16 als zuverlässige Referenzgene identifizierten Kandidaten hinsichtlich der Analyseansätze/experimentellen Bedingungen.** *Maea* wurde am häufigsten in unterschiedlichen Analyseansätzen als eines der zuverlässigen Referenzgene identifiziert.

Der überwiegende Anteil (13 von 16 Referenzgenen) dieser zuverlässigen Referenzgene stammt aus der Genevestigator-Analyse im Vorauswahlschritt und zwei Gene aus Evaluierungsstudien über zuverlässige Referenzgene sowie ein Gen aus dem Roche-Referenzgenpanel, was die Robustheit vom Genevestigator-Algorithmus zur Suche nach stabil exprimierten Referenzgenkandidaten unterstreicht. Von den Referenzgenkandidaten, die in früheren qPCR-Studien bereits als konventionell verwendete Referenzgene eingesetzt wurden (sieben Kandidatengene aus der Quelle a im Abschnitt 3.1), erwies sich in unserer RT-qPCR-Validierung letztendlich kein einziges Gen stabil exprimiert, was auf Schwäche der konventionell verwendeten Referenzgene hinweist.

Die 16 zuverlässige Referenzgene besaßen außerdem unterschiedliches Expressionsniveau. So sind *Csnk1a1*, *Eif3a*, *Eif4g2*, *Rpl4*, *Tmed2*, *Zc3h11a* einer hohen Expressionsgruppe (mediane, log<sub>2</sub>-skalierte Signalintensität von Genevestigator in der Lunge der C57BL/6-Mäuse = 13,90-17,39) zuzuordnen und *Elac2*, *Dolpp1*, *Maea*, *Mpv17l*, *Pcsk7*, *Tbp*, *Wars2* einer mittleren Expressionsgruppe (mediane, log<sub>2</sub>-skalierte Signalintensität = 10,05-12,91) sowie *Auh*, *Nupl2*, *Rexo5* einer niedrigen Expressionsgruppe (mediane, log<sub>2</sub>-skalierte Signalintensität = 8,78-9,32). Da ähnliche Expressionsniveaus von Referenzgenen und Target-Genen die Sensitivität für Detektion geringer Genexpressionsalteration erhöhen,<sup>68</sup> sollten dieses unterschiedliche

Expressionsniveau für die im Rahmen dieser Arbeit identifizierten, zuverlässigen Referenzgene entsprechend dem Expressionsniveau der Target-Genen in zukünftigen Studien berücksichtigt werden.

Weiterführende Literaturrecherche deckte die Rolle von unseren 16 zuverlässigen Referenzgenen in unterschiedlichen Erkrankungen auf. So belegten frühere Studien, dass *Elac2*, *Csnk1a1*, *Eif3a*, *Eif4g2*, *Tmed2*, *Mpv17l* an der Pathogenese der benignen und malignen Lungenerkrankungen mit unterschiedlicher Funktion beteiligt sind,<sup>100-104</sup> jedoch die Aufgabe dieser Gene in neonatalen Lungen mit oder ohne Erkrankung ist weitgehend nicht erforscht.

Die im Rahmen dieses Projektes identifizierten, zuverlässigen Referenzgene für neonatale Mauslungen unter Berücksichtigung der biologischen, technischen und experimentellen Variablen werden für zukünftige Genexpressionsstudien bei adäquater Auswahl der geeigneten Referenzgene beitragen und können außerdem einen Einblick in bislang unbekannte Gene in Zusammenhang mit neonatalen Lungenerkrankungen bieten.

## Anhang

**Tab.A1 Übersicht der RNA/cDNA-Proben aus Lungengewebe von C57BL/6-Mäusen.** Nox und Hyx stehen jeweils für Normoxie und Hyperoxie, d.h. die Mäuse kommen jeweils aus physiologischer (FiO<sub>2</sub>=0,21) oder hypoxischer (FiO<sub>2</sub>=0,80) Bedingung. Das „x“ in der Spalte "IRC" bedeutet, dass die betreffende Probe zur Herstellung des Interrun-Kalibrators (IRC, siehe Abschnitt 3.7) verwendet wurden.

cDNA#	Beschreibung	Geschlecht	IRC	RIN	cDNA#	Beschreibung	Geschlecht	IRC	RIN
1	PND2.5 Nox	f	x	7,0	60	PND21.5 Hyx	m	x	6,1
3	PND2.5 Nox	m	x	8,1	61	PND21.5 Hyx	m	-	5,7
4	PND2.5 Nox	f	x	7,8	62	PND21.5 Hyx	f	x	6,4
5	PND2.5 Nox	f	x	7,4	63	PND21.5 Hyx	m	x	7,1
6	PND2.5 Nox	f	x	8,0	64	PND21.5 Hyx	m	x	6,2
8	PND2.5 Nox	m	x	6,2	65	PND28.5 Nox	m	-	n/a
10	PND2.5 Hyx	f	x	6,4	66	PND28.5 Nox	f	-	n/a
11	PND2.5 Hyx	m	x	6,7	67	PND28.5 Nox	f	-	n/a
12	PND2.5 Hyx	m	x	6,2	68	PND28.5 Nox	m	-	4,5
13	PND2.5 Hyx	f	x	7,1	69	PND28.5 Nox	f	-	n/a
15	PND2.5 Hyx	f	x	6,8	70	PND28.5 Nox	m	-	3,9
16	PND2.5 Hyx	f	x	7,6	71	PND28.5 Hyx	f	-	8,2
18	PND5.5 Nox	m	x	6,4	72	PND28.5 Hyx	f	-	8,2
19	PND5.5 Nox	m	x	7,7	73	PND28.5 Hyx	m	-	8,1
20	PND5.5 Nox	m	x	6,2	74	PND28.5 Hyx	m	-	7,8
22	PND5.5 Nox	m	x	6,4	75	PND28.5 Hyx	f	-	8,3
25	PND5.5 Nox	f	-	5,7	76	PND28.5 Hyx	m	-	9,2
26	PND5.5 Nox	m	-	3,7	79	PND5.5 Nox_add	n/a	-	5,6
27	PND5.5 Hyx	f	x	7,1	80	PND5.5 Nox_add	n/a	-	5,3
28	PND5.5 Hyx	f	x	7,0	81	PND14.5 Hyx_add	n/a	-	7,5
29	PND5.5 Hyx	f	x	5,4	82	PND14.5 Hyx_add	n/a	-	5,4
30	PND5.5 Hyx	f	x	7,8	83	PND14.5 Nox_add	n/a	-	5,6
31	PND5.5 Hyx	f	x	6,6	84	PND14.5 Nox_add	n/a	-	n/a
32	PND5.5 Hyx	m	x	6,6	85	PND14.5 Nox_add	n/a	-	n/a
35	PND14.5 Nox	f	-	5,4	86	PND14.5 Nox_add	n/a	-	5,9
37	PND14.5 Nox	m	-	3,1	87	PND14.5 Nox_add	n/a	-	n/a
39	PND14.5 Nox	m	-	3,3	88	PND14.5 Nox_add	n/a	-	4,1
40	PND14.5 Nox	f	-	3,7	89	PND21.5 Hyx_add	n/a	-	9,2
43	PND14.5 Nox	m	-	n/a	90	PND21.5 Hyx_add	n/a	-	8,8
44	PND14.5 Nox	m	-	n/a	91	PND21.5 Hyx_add	n/a	-	9,1
45	PND14.5 Hyx	m	x	7,5	92	PND21.5 Hyx_add	n/a	-	9,1
46	PND14.5 Hyx	f	-	5,5	93	PND21.5 Nox_add	n/a	-	n/a
47	PND14.5 Hyx	m	x	7,7	94	PND21.5 Nox_add	n/a	-	7,7
48	PND14.5 Hyx	m	x	7,5	95	PND21.5 Nox_add	n/a	-	6,1
50	PND14.5 Hyx	m	x	7,9	96	PND21.5 Nox_add	n/a	-	n/a
51	PND14.5 Hyx	m	x	7,0	97	PND21.5 Nox_add	n/a	-	7,8

<b>cDNA#</b>	<b>Beschreibung</b>	<b>Geschlecht</b>	<b>IRC</b>	<b>RIN</b>	<b>cDNA#</b>	<b>Beschreibung</b>	<b>Geschlecht</b>	<b>IRC</b>	<b>RIN</b>
53	PND21.5 Nox	f	-	n/a	98	PND21.5 Nox_add	n/a	-	6,2
54	PND21.5 Nox	f	-	n/a	99	PND28.5 Nox_add	n/a	-	3,9
55	PND21.5 Nox	f	-	2,7	100	PND28.5 Nox_add	n/a	-	6,0
56	PND21.5 Nox	m	-	n/a	101	PND28.5 Nox_add	n/a	-	n/a
57	PND21.5 Nox	f	-	3,2	102	PND28.5 Nox_add	n/a	-	n/a
58	PND21.5 Nox	m	-	4,3	103	PND28.5 Nox_add	n/a	-	n/a
59	PND21.5 Hyx	f	-	5,6	104	PND28.5 Nox_add	n/a	-	5,5

**Tab.A2 Die 122 potenzielle Referenzgenkandidaten in der Vorauswahl.** Die initialen 122 Referenzgenkandidaten wurden aus vier voneinander unabhängigen Datenquellen, nämlich Studien mit Einsatz von konventionell verwendeten Referenzgenen (Quelle a), Evaluierungsstudien über Eignung der konventionell verwendeten Referenzgene mit Vorschlag neuer geeigneter Referenzgene (Quelle b), Analyse der öffentlich verfügbaren Microarray-Daten mittels Genevestigator (Quelle c) and Roche Referenzgenpanel mit häufig eingesetzten Referenzgenen (Quelle d). Anschließend wurden 76 Gene mit den in der Tabelle dargestellten Gründen von der RT-qPCR-Validierung rausgenommen. Die Kandidatengene, die als "nicht detektierbar in unseren eigenen Microarray-Daten" bezeichnet sind (35 Gene), waren einfach nicht in der verwendeten Micorarray-Plattform (CodeLink UniSet Mouse I Bioarray 10K) zur Expressionsbestimmung eingeschlossen, jedoch für die 24 Gene davon wurden jeweils andere Gene in der gleichen Genfamilie untersucht (jeweils ohne Geneexpression).

Index	Gene	Entrez	Quelle	Referenz	Exklusionskriterien
1	<i>Actb</i>	11461	b	Stoeger et al. <sup>97</sup>	überlappende Quellen
2	<i>B2m</i>	12010	b	Shimada et al. <sup>105</sup>	überlappende Quellen
3	<i>Gapdh</i>	14433	b	Shimada et al.	überlappende Quellen
4	<i>Gusb</i>	110006	b	Stoeger et al.	überlappende Quellen
5	<i>Oaz1</i>	18245	b	De Jonge et al. <sup>99</sup>	überlappende Quellen
6	<i>Actb</i>	11461	d	Roche Referenzgenpanel	überlappende Quellen
7	<i>B2m</i>	12010	d	Roche Referenzgenpanel	überlappende Quellen
8	<i>Gapdh</i>	14433	d	Roche Referenzgenpanel	überlappende Quellen
9	<i>Gusb</i>	110006	d	Roche Referenzgenpanel	überlappende Quellen
10	<i>Hmbs</i>	15288	d	Roche Referenzgenpanel	überlappende Quellen
11	<i>Pgk1</i>	18655	d	Roche Referenzgenpanel	überlappende Quellen
12	<i>Ipo8</i>	320727	d	Roche Referenzgenpanel	überlappende Quellen
13	<i>Ppia</i>	268373	d	Roche Referenzgenpanel	überlappende Quellen
14	<i>Rpl13a</i>	22121	d	Roche Referenzgenpanel	überlappende Quellen
15	<i>Oaz1</i>	18245	a	Rothman et al. <sup>80</sup> , De Jonge et al.	bekannte Pseudogene
16	<i>Eef2</i>	13629	b	Kouadjo et al. <sup>106</sup>	bekannte Pseudogene
17	<i>Rpl22</i>	19934	b	De Jonge et al.	für ribosomales Protein kodierende Gene, <sup>10</sup> bekannte Pseudogenene
18	<i>Rpl27</i>	19942	b	De Jonge et al.	für ribosomales Protein kodierende Gene, bekannte Pseudogenene
19	<i>Rpl37</i>	67281	b	Shimada et al., Kouadjo et al.	für ribosomales Protein kodierende Gene, bekannte Pseudogenene
20	<i>Rpl38</i>	67671	b	Kouadjo et al.	für ribosomales Protein kodierende Gene, bekannte Pseudogenene
21	<i>Rpl9</i>	20005	b	De Jonge et al.	für ribosomales Protein kodierende Gene, bekannte Pseudogenene
22	<i>Rplp2</i>	67186	b	Shimada et al.	für ribosomales Protein kodierende Gene, bekannte Pseudogenene
23	<i>Rps12</i>	20042	b	De Jonge et al.	für ribosomales Protein kodierende Gene, bekannte Pseudogenene
24	<i>Rps13</i>	68052	b	De Jonge et al.	für ribosomales Protein kodierende Gene, bekannte Pseudogenene
25	<i>Rps16</i>	20055	b	De Jonge et al.	für ribosomales Protein kodierende Gene, bekannte Pseudogenene
26	<i>Rps20</i>	67427	b	De Jonge et al.	für ribosomales Protein kodierende Gene, bekannte Pseudogenene
27	<i>Hbb-b1</i>	15129	d	Roche Referenzgenpanel	erwartete Expressionsalteration unter Hyperoxie-Bedingung
28	<i>B2m</i>	12010	a	Tzouveleakis et al. <sup>79</sup>	nicht exprimiert in unseren eigenen Microarray-Daten
29	<i>Hmbs</i>	15288	a	Kitowska et al. <sup>78</sup>	nicht exprimiert in unseren eigenen Microarray-Daten
30	<i>Pgk1</i>	18655	a	Yamaji et al.	nicht exprimiert in unseren eigenen Microarray-Daten
31	<i>Chmp2a</i>	295670	b	Kouadjo et al.	nicht exprimiert in unseren eigenen Microarray-Daten
32	<i>Heatr3</i>	234549	b	Kouadjo et al.	nicht detektierbar in unseren eigenen Microarray-Daten
33	<i>Ipo8</i>	320727	b	Shimada et al.	nicht detektierbar in unseren eigenen Microarray-Daten
34	<i>Leng8</i>	232798	b	Kouadjo et al.	nicht detektierbar in unseren eigenen Microarray-Daten

Index	Gene	Entrez	Quelle	Referenz	Exklusionskriterien
35	<i>Ppia</i>	268373	b	Stoeger et al.	nicht exprimiert in unseren eigenen Microarray-Daten
36	<i>Rpl24</i>	68193	b	De Jonge et al.	nicht detektierbar in unseren eigenen Microarray-Daten
37	<i>Rpl6</i>	19988	b	De Jonge et al.	nicht exprimiert in unseren eigenen Microarray-Daten
38	<i>Rps29</i>	20090	b	De Jonge et al.	nicht exprimiert in unseren eigenen Microarray-Daten
39	<i>Srp14</i>	20813	b	De Jonge et al.	nicht detektierbar in unseren eigenen Microarray-Daten
40	<i>Rpl13a</i>	22121	b	Stoeger et al., De Jonge et al.	nicht exprimiert in unseren eigenen Microarray-Daten
41	<i>Heatr7a</i>	223658	c	Genevestigator	nicht detektierbar in unseren eigenen Microarray-Daten
42	<i>Maskbp3</i>	71783	c	Genevestigator	nicht detektierbar in unseren eigenen Microarray-Daten
43	<i>Zfp182</i>	319535	c	Genevestigator	nicht detektierbar in unseren eigenen Microarray-Daten
44	<i>1446253_at</i>	n/a	c	Genevestigator	nicht detektierbar in unseren eigenen Microarray-Daten
45	<i>Cep57l1</i>	103268	c	Genevestigator	nicht detektierbar in unseren eigenen Microarray-Daten
46	<i>Il15ra</i>	16169	c	Genevestigator	nicht detektierbar in unseren eigenen Microarray-Daten
47	<i>Ints5</i>	109077	c	Genevestigator	nicht detektierbar in unseren eigenen Microarray-Daten
48	<i>LOC106740</i>	106740	c	Genevestigator	nicht detektierbar in unseren eigenen Microarray-Daten
49	<i>Metap1</i>	75624	c	Genevestigator	nicht detektierbar in unseren eigenen Microarray-Daten
50	<i>Nek4</i>	23955	c	Genevestigator	nicht exprimiert in unseren eigenen Microarray-Daten
51	<i>Nup214</i>	227720	c	Genevestigator	nicht detektierbar in unseren eigenen Microarray-Daten
52	<i>Plekha3</i>	241075	c	Genevestigator	nicht detektierbar in unseren eigenen Microarray-Daten
53	<i>Sec62</i>	69276	c	Genevestigator	nicht detektierbar in unseren eigenen Microarray-Daten
54	<i>Ywhaq</i>	22630	c	Genevestigator	nicht detektierbar in unseren eigenen Microarray-Daten
55	<i>Zfp592</i>	233410	c	Genevestigator	nicht detektierbar in unseren eigenen Microarray-Daten
56	<i>Zxda</i>	668166	c	Genevestigator	nicht exprimiert in unseren eigenen Microarray-Daten
57	<i>Gabpb2</i>	213054	c	Genevestigator	nicht detektierbar in unseren eigenen Microarray-Daten
58	<i>Mcp1</i>	244329	c	Genevestigator	nicht detektierbar in unseren eigenen Microarray-Daten
59	<i>Nudt6</i>	229228	c	Genevestigator	nicht detektierbar in unseren eigenen Microarray-Daten
60	<i>Trim37</i>	68729	c	Genevestigator	nicht detektierbar in unseren eigenen Microarray-Daten
61	<i>Zfp251</i>	71591	c	Genevestigator	nicht detektierbar in unseren eigenen Microarray-Daten
62	<i>1110035M17Rik</i>	432396	c	Genevestigator	nicht detektierbar in unseren eigenen Microarray-Daten
63	<i>1435789_x_at</i>	n/a	c	Genevestigator	nicht detektierbar in unseren eigenen Microarray-Daten
64	<i>Caprin2</i>	232560	c	Genevestigator	nicht detektierbar in unseren eigenen Microarray-Daten
65	<i>Crnk1</i>	66877	c	Genevestigator	nicht exprimiert in unseren eigenen Microarray-Daten
66	<i>Gtf2h3</i>	209357	c	Genevestigator	nicht detektierbar in unseren eigenen Microarray-Daten
67	<i>H3f3a</i>	15078	c	Genevestigator	nicht detektierbar in unseren eigenen Microarray-Daten
68	<i>Hnrnp1</i>	59013	c	Genevestigator	nicht detektierbar in unseren eigenen Microarray-Daten
69	<i>Rplp0</i>	11837	d	Roche Referenzgenpanel	nicht detektierbar in unseren eigenen Microarray-Daten
70	<i>Trpt1</i>	107328	d	Roche Referenzgenpanel	nicht detektierbar in unseren eigenen Microarray-Daten
71	<i>Alas2</i>	11656	d	Roche Referenzgenpanel	nicht detektierbar in unseren eigenen Microarray-Daten
72	<i>Cyp2d13</i>	68444	d	Roche Referenzgenpanel	nicht exprimiert in unseren eigenen Microarray-Daten
73	<i>Fxyd6</i>	59095	d	Roche Referenzgenpanel	nicht exprimiert in unseren eigenen Microarray-Daten
74	<i>Fxyd7</i>	57780	d	Roche Referenzgenpanel	nicht detektierbar in unseren eigenen Microarray-Daten
75	<i>H2-Ea-ps</i>	100504404	d	Roche Referenzgenpanel	nicht exprimiert in unseren eigenen Microarray-Daten

Index	Gene	Entrez	Quelle	Referenz	Exklusionskriterien
76	<i>Oas1b</i>	23961	d	Roche Referenzgenpanel	nicht detektierbar in unseren eigenen Microarray-Daten
77	<i>Actb</i>	11461	a	Yamaji et al.	-
78	<i>Gapdh</i>	14433	a	Sakazaki et al, Yamaji et al.	-
79	<i>Hsp90ab1</i>	15516	d	Roche Referenzgenpanel	-
80	<i>Pla2g2a</i>	18780	d	Roche Referenzgenpanel	-
81	<i>Gusb</i>	110006	a	McDonald et al. <sup>81</sup>	-
82	<i>Rhot1</i>	59040	c	Genevestigator	-
83	<i>Ing3</i>	71777	c	Genevestigator	-
84	<i>G6pdx</i>	14381	d	Roche Referenzgenpanel	-
85	<i>Sdha</i>	66945	d	Roche Referenzgenpanel	-
86	<i>Hdgfrp2</i>	15193	c	Genevestigator	-
87	<i>Tomm20</i>	67952	c	Genevestigator	-
88	<i>Mx2</i>	17858	d	Roche Referenzgenpanel	-
89	<i>Atp6v0c</i>	11984	c	Genevestigator	-
90	<i>Ywhaz</i>	22631	d	Roche Referenzgenpanel	-
91	<i>Eif4h</i>	22384	c	Genevestigator	-
92	<i>Slc35a5</i>	74102	c	Genevestigator	-
93	<i>Canx</i>	12330	c	Genevestigator	-
94	<i>Myadm</i>	50918	c	Genevestigator	-
95	<i>Rrp1b</i>	72462	c	Genevestigator	-
96	<i>Hprt</i>	15452	d	Roche Referenzgenpanel	-
97	<i>Rab23</i>	19335	c	Genevestigator	-
98	<i>F2r</i>	14062	c	Genevestigator	-
99	<i>Rpl4</i>	67891	b	De Jonge et al.	-
100	<i>Psm4</i>	19185	b	Kouadjo et al.	-
101	<i>Tfrc</i>	22042	d	Roche Referenzgenpanel	-
102	<i>Puf60</i>	67959	b	Kouadjo et al.	-
103	<i>Csnk1a1</i>	93687	c	Genevestigator	-
104	<i>Tnks2</i>	74493	c	Genevestigator	-
105	<i>Eif4g2</i>	13690	c	Genevestigator	-
106	<i>Eif3a</i>	13669	b	Kouadjo et al.	-
107	<i>Tbp</i>	21374	d	Roche Referenzgenpanel	-
108	<i>Grn</i>	14824	c	Genevestigator	-
109	<i>Tmed2</i>	56334	c	Genevestigator	-
110	<i>Nupl2</i>	231042	c	Genevestigator	-
111	<i>Rexo5</i>	434234	c	Genevestigator	-
112	<i>Zc3h11a</i>	70579	c	Genevestigator	-
113	<i>Efh2</i>	27984	c	Genevestigator	-
114	<i>Auh</i>	11992	c	Genevestigator	-
115	<i>Wars2</i>	70560	c	Genevestigator	-
116	<i>Fkbp1a</i>	14225	c	Genevestigator	-

Index	Gene	Entrez	Quelle	Referenz	Exklusionskriterien
117	<i>Maea</i>	59003	c	Genevestigator	-
118	<i>Kdm3b</i>	277250	c	Genevestigator	-
119	<i>Elac2</i>	68626	c	Genevestigator	-
120	<i>Pcsk7</i>	18554	c	Genevestigator	-
121	<i>Mpv17l</i>	93734	c	Genevestigator	-
122	<i>Dolpp1</i>	57170	c	Genevestigator	-

## Literaturverzeichnis

1. Weeraratna, A. T., Nagel, J. E., De Mello-Coelho, V. & Taub, D. D. Gene expression profiling: From microarrays to medicine. *Journal of Clinical Immunology* **24**, 213–224 (2004).
2. Ståhlberg, A., Håkansson, J., Xian, X., Semb, H. & Kubista, M. Properties of the Reverse Transcription Reaction in mRNA Quantification. *Clinical Chemistry* **50**, 509–515 (2004).
3. Vandesompele, J. *et al.* Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* **3**, RESEARCH0034 (2002).
4. Vandesompele, J., Kubista, M. & Pfaffl, M. M. W. Reference Gene Validation Software for Improved Normalization. *Real-time PCR: Current Technology and Applications* **4**, 47–64 (2009).
5. Andersen, C. L., Jensen, J. L. & Ørntoft, T. F. Normalization of real-time quantitative reverse transcription-PCR data: A model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Research* **64**, 5245–5250 (2004).
6. Dheda, K. *et al.* Validation of housekeeping genes for normalizing RNA expression in real-time PCR. *Biotechniques* **37**, 112–119 (2004).
7. Karge, W. H., Schaefer, E. J. & Ordovas, J. M. Quantification of mRNA by polymerase chain reaction (PCR) using an internal standard and a nonradioactive detection method. *Methods Mol Biol* **110**, 43–61 (1998).
8. Radonić, A. *et al.* Guideline to reference gene selection for quantitative real-time PCR. *Biochem Biophys Res Commun* **313**, 856–62 (2004).
9. Bas, a, Forsberg, G., Hammarström, S. & Hammarström, M.-L. Utility of the housekeeping genes 18S rRNA, beta-actin and glyceraldehyde-3-phosphate-dehydrogenase for normalization in real-time quantitative reverse transcriptase-polymerase chain reaction analysis of gene expression in human T lymphocytes. *Scand J Immunol* **59**, 566–573 (2004).
10. Thorrez, L. *et al.* Using ribosomal protein genes as reference: A tale of caution. *PLoS ONE* **3**, (2008).
11. Suzuki, T., Higgins, P. J. & Crawford, D. R. Control selection for RNA quantitation. *Biotechniques* **29**, 332–337 (2000).
12. Kouadjo, K. E., Nishida, Y., Cadrin-Girard, J. F., Yoshioka, M. & St-Amand, J. Housekeeping and tissue-specific genes in mouse tissues. *BMC Genomics* **8**, 127 (2007).
13. Frericks, M. & Esser, C. A toolbox of novel murine house-keeping genes identified by meta-analysis of large scale gene expression profiles. *Biochim Biophys Acta* **1779**, 830–837 (2008).
14. Zhang, X., Ding, L. & Sandford, A. J. Selection of reference genes for gene expression studies in human neutrophils by real-time PCR. *BMC Molecular Biology* **6**, 1–7 (2005).

15. Radonić, A. *et al.* Guideline to reference gene selection for quantitative real-time PCR. *Biochemical and Biophysical Research Communications* **313**, 856–862 (2004).
16. Thorrez, L. *et al.* Using ribosomal protein genes as reference: A tale of caution. *PLoS ONE* **3**, (2008).
17. Huggett, J., Dheda, K., Bustin, S. & Zumla, A. Real-time RT-PCR normalisation; strategies and considerations. *Genes and Immunity* **6**, 279–284 (2005).
18. Thellin, O. *et al.* Housekeeping genes as internal standards: Use and limits. *Journal of Biotechnology* **75**, 291–295 (1999).
19. Bémeur, C. *et al.* Decreased  $\beta$ -actin mRNA expression in hyperglycemic focal cerebral ischemia in the rat. *Neuroscience Letters* **357**, 211–214 (2004).
20. Schmid, H. *et al.* Validation of endogenous controls for gene expression analysis in microdissected human renal biopsies. *Kidney International* **64**, 356–360 (2003).
21. Tricarico, C. *et al.* Quantitative real-time reverse transcription polymerase chain reaction: normalization to rRNA or single housekeeping genes is inappropriate for human tissue biopsies• 1. *Anal Biochem* **309**, 293–300 (2002).
22. St-Pierre, J., Grégoire, J. C. & Vaillancourt, C. A simple method to assess group difference in RT-qPCR reference gene selection using GeNorm: The case of the placental sex. *Scientific Reports* **7**, 0–7 (2017).
23. de Jonge, H. J. M. *et al.* Evidence Based Selection of Housekeeping Genes. *PLoS ONE* **2**, e898 (2007).
24. Gong, H. *et al.* Evaluation of candidate reference genes for RT-qPCR studies in three metabolism related tissues of mice after caloric restriction. *Scientific Reports* **6**, 1–12 (2016).
25. Tatsumi, K. *et al.* Reference gene selection for real-time RT-PCR in regenerating mouse livers. *Biochemical and Biophysical Research Communications* **374**, 106–110 (2008).
26. Taylor, S., Wakem, M., Dijkman, G., Alsarraj, M. & Nguyen, M. A practical approach to RT-qPCR—Publishing data that conform to the MIQE guidelines. *Methods* **50**, S1–S5 (2010).
27. Huggett, J., Dheda, K., Bustin, S. & Zumla, A. Real-time RT-PCR normalisation; strategies and considerations. *Genes and Immunity* **6**, 279–284 (2005).
28. Foquet, B. & Song, H. There is no magic bullet: The importance of testing reference gene stability in RT-qPCR experiments across multiple closely related species. *PeerJ* **8**, (2020).
29. Perlman, R. L. Mouse Models of Human Disease: An Evolutionary Perspective. *Evolution, Medicine, and Public Health* eow014 (2016) doi:10.1093/emph/eow014.
30. Rosenthal, N. & Brown, S. The mouse ascending: perspectives for human-disease models. **9**, 993–999 (2007).
31. Herriges, M. & Morrissey, E. E. Lung development: Orchestrating the generation and regeneration of a complex organ. *Development (Cambridge)* **141**, 502–513 (2014).

32. Hruz, T. *et al.* RefGenes: Identification of reliable and condition specific reference genes for RT-qPCR data normalization. *BMC Genomics* **12**, (2011).
33. Adams, G. A beginner's guide to RT-PCR, qPCR and RT-qPCR. *Biochemist* **42**, (2020).
34. Huggett, J., Dheda, K., Bustin, S. & Zumla, A. Real-time RT-PCR normalisation; strategies and considerations. *Genes and Immunity* **6**, 279–284 (2005).
35. Bustin, S. A. Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): Trends and problems. *Journal of Molecular Endocrinology* **29**, 23–39 (2002).
36. Alberts, B., Johnson, A., Walter, P., Lewis, J. & Raff, M. *Molecular Biology of the Cell*. (Norton & Company, 2014).
37. Rodriguez-Lázaro, D. *Real-Time PCR in food science: Current technology and applications*. (Caister Academic Press, 2013).
38. Life technologies. Real-time PCR handbook. *Applied Biosystems* 1–70 (2014) doi:CO320850812.
39. Biorad. *Real-Time PCR Applications Guide*. (2006).
40. D'haene, B., Vandesompele, J. & Hellemans, J. Accurate and objective copy number profiling using real-time quantitative PCR. *Methods* vol. 50 262–270 Preprint at <https://doi.org/10.1016/j.jymeth.2009.12.007> (2010).
41. Lifetechnologies. *Real-time PCR handbook*. (2012).
42. Dr. Klein, Dr. Rost & Kollegen. Real-Time PCR. <https://www.medizinische-genetik.de/diagnostik/allgemeine-informationen/methoden/real-time-pcr>.
43. Derveaux, S., Vandesompele, J. & Hellemans, J. How to do successful gene expression analysis using real-time PCR. *Methods* **50**, 227–230 (2010).
44. Mueller, O. & Schroeder, A. RNA Integrity Number ( RIN ) – Standardization of RNA Quality Control Application. *Nano* 1–8 (2004) doi:10.1101/gr.189621.115.7.
45. Fleige, S. & Pfaffl, M. W. RNA integrity and the effect on the real-time qRT-PCR performance. *Molecular Aspects of Medicine* **27**, 126–139 (2006).
46. Vermeulen, J. *et al.* Measurable impact of RNA quality on gene expression results from quantitative PCR. *Nucleic Acids Research* **39**, (2011).
47. Pérez-Novo, C. A. *et al.* Impact of RNA quality on reference gene expression stability. *Biotechniques* **39**, 52–56 (2005).
48. Bustin, S. A. & Nolan, T. Pitfalls of quantitative real-time reverse-transcription polymerase chain reaction. *J Biomol Tech* **15**, 155–66 (2004).
49. Bustin, S. A. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *Journal of Molecular* **25**, 169–93 (2000).
50. Szabo, A. *et al.* Statistical modeling for selecting housekeeper genes. *Genome Biology* **5**, (2004).

51. Mueller, O. *et al.* A microfluidic system for high-speed reproducible DNA sizing and quantitation. *Electrophoresis* **21**, 128–134 (2000).
52. Imbeaud, S. *et al.* Towards standardization of RNA quality assessment using user-independent classifiers of microcapillary electrophoresis traces. *Nucleic Acids Research* **33**, 1–12 (2005).
53. Bustin, S. a & Nolan, T. Pitfalls of Quantitative Real-time PCR. *Journal of Biomolecular Techniques* **15**, 155–166 (2004).
54. Schroeder, A. *et al.* The RIN: An RNA integrity number for assigning integrity values to RNA measurements. *BMC Molecular Biology* **7**, 1–14 (2006).
55. Plattform, E. Agilent 2100 Bioanalyzer Angewandte Lab-On-A-Chip-Technologie - der Agilent 2100 Bioanalyzer Automatisierte , schnelle Analyse mit hervorragender Datenqualität für die analytische Biochemie. *Design*.
56. Agilent Technologies. Agilent RNA 6000 Nano Kit Guide. 2006 1–28 (2006).
57. Agilent-Technologies. Agilent 2100 Bioanalyzer 2100 Expert User's Guide. *Agilent Technologies* 1–366 (2005).
58. *Agilent 2100 Bioanalyzer Eine Plattform-zahlreiche Möglichkeiten für die Analyse von DNA, RNA, Proteinen und Zellen.* (2007).
59. *Agilent 2100 Bioanalyzer Application Compendium DNA, RNA, Protein and cell analysis.* (2011).
60. Melen, G. J., Pesce, C. G., Rossi, M. S. & Kornblihtt, A. R. Novel processing in a mammalian nuclear 28S pre-rRNA: Tissue-specific elimination of an “intron” bearing a hidden break site. *EMBO Journal* **18**, 3107–3118 (1999).
61. Marx, V. RNA Quality: Defining the Good, the Bad & the Ugly. *Genomics & Proteomics* **4**, 14 (2004).
62. Schroeder, A. *et al.* The RIN: An RNA integrity number for assigning integrity values to RNA measurements. *BMC Molecular Biology* **7**, 1–14 (2006).
63. Fleige, S. & Pfaffl, M. Einfluss der RNA-Integrität auf die quantitative real-time RT-PCR. *Laborwelt; 5: 4* 12–14 (2007).
64. Schoor, O. *et al.* Moderate degradation does not preclude microarray analysis of small amounts of RNA. *Biotechniques* **35**, 1192–1201 (2003).
65. Nebion. Genevestigator user manual. Preprint at (2021).
66. Hruz, T. *et al.* RefGenes: Identification of reliable and condition specific reference genes for RT-qPCR data normalization. *BMC Genomics* **12**, 156 (2011).
67. Coulson, D. T. R. *et al.* Identification of valid reference genes for the normalization of RT qPCR gene expression data in human brain tissue. *BMC Molecular Biology* **9**, 1–11 (2008).
68. Silver, N., Best, S., Jiang, J. & Thein, S. L. Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR. *BMC Molecular Biology* **7**, (2006).

69. Abbott, L. F., Regehr, W. G. & Ab, M. A. GenEx User Guide. *Nature* **431**, 1–20 (2004).
70. Andersen, C. L., Jensen, J. L. & Ørntoft, T. F. Normalization of real-time quantitative reverse transcription-PCR data: A model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Research* **64**, 5245–5250 (2004).
71. Biogazelle. Inter-run Calibration.
72. Ståhlberg, A., Rusnakova, V., Forootan, A., Anderova, M. & Kubista, M. RT-qPCR work-flow for single-cell data analysis. *Methods* (2013) doi:10.1016/j.ymeth.2012.09.007.
73. Hellemans, J., Mortier, G., De Paepe, A., Speleman, F. & Vandesompele, J. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol* **8**, R19 (2007).
74. Ramhøj, L., Axelstad, M. & Svingen, T. Validation of endogenous reference genes in rat cerebral cortex for RT-qPCR analyses in developmental toxicity studies. *PeerJ* **2019**, (2019).
75. Negretti, N. M. *et al.* A single-cell atlas of mouse lung development. *Development* **148**, (2021).
76. Yamaji-Kegan, K. *et al.* Hypoxia-induced mitogenic factor (HIMF/FIZZ1/RELMalpha) increases lung inflammation and activates pulmonary microvascular endothelial cells via an IL-4-dependent mechanism. *J Immunol* **185**, 5539–5548 (2010).
77. Sakazaki, Y. *et al.* Overexpression of chitinase 3-Like 1/YKL-40 in Lung-Specific IL-18-transgenic mice, smokers and COPD. *PLoS ONE* **6**, 3–10 (2011).
78. Kitowska, K. *et al.* Functional role and species-specific contribution of arginases in pulmonary fibrosis. *American Journal of Physiology - Lung Cellular and Molecular Physiology* **294**, (2008).
79. Tzouveleakis, A. *et al.* Down-regulation of the inhibitor of growth family member 4 (ING4) in different forms of pulmonary fibrosis. *Respiratory Research* **10**, (2009).
80. Rothman, R. B. *et al.* Altered gene expression in pulmonary tissue of tryptophan hydroxylase-1 knockout mice: Implications for pulmonary arterial hypertension. *PLoS ONE* **6**, (2011).
81. McDonald, R. B. *et al.* Characterization of survival and phenotype throughout the life span in UCP2/UCP3 genetically altered mice. *Experimental Gerontology* **43**, (2008).
82. Shimada, I., Matsui, K., Iida, R., Tsubota, E. & Matsuki, T. Time course of housekeeping gene expression changes in diffuse alveolar damage induced by hyperoxia exposure in mice. *Legal Medicine* **11**, S151–S154 (2009).
83. Yin, R., Tian, F., Frankenberger, B., de Angelis, M. H. & Stoeger, T. Selection and evaluation of stable housekeeping genes for gene expression normalization in carbon nanoparticle-induced acute pulmonary inflammation in mice. *Biochemical and Biophysical Research Communications* **399**, 531–536 (2010).
84. Kouadjo, K. E., Nishida, Y., Cadrin-Girard, J. F., Yoshioka, M. & St-Amand, J. Housekeeping and tissue-specific genes in mouse tissues. *BMC Genomics* **8**, (2007).

85. Jacobson, J. R. *et al.* Simvastatin attenuates vascular leak and inflammation in murine inflammatory lung injury. *American Journal of Physiology - Lung Cellular and Molecular Physiology* **288**, (2005).
86. Haston, C. K., Tomko, T. G., Godin, N., Kerckhoff, L. & Hallett, M. T. Murine candidate bleomycin induced pulmonary fibrosis susceptibility genes identified by gene expression and sequence analysis of linkage regions. *J Med Genet* **42**, 464–473 (2005).
87. Lattin, J. E. *et al.* Expression analysis of G Protein-Coupled Receptors in mouse macrophages. *Immunome Research* **4**, 1–13 (2008).
88. McGrath-Morrow, S. *et al.* Impaired lung homeostasis in neonatal mice exposed to cigarette smoke. *American Journal of Respiratory Cell and Molecular Biology* **38**, 393–400 (2008).
89. Hong, S. B. *et al.* Essential role of pre-B-cell colony enhancing factor in ventilator-induced lung injury. *American Journal of Respiratory and Critical Care Medicine* **178**, 605–617 (2008).
90. Kozul, C. D. *et al.* Chronic exposure to arsenic in the drinking water alters the expression of immune response genes in mouse lung. *Environmental Health Perspectives* **117**, 1108–1115 (2009).
91. Wray, C. *et al.* Claudin-4 augments alveolar epithelial barrier function and is induced in acute lung injury. *Am J Physiol Lung Cell Mol Physiol* **297**, L219–L227 (2009).
92. Meyer, N. J. *et al.* GADD45a is a novel candidate gene in inflammatory lung injury via influences on Akt signaling. *The FASEB journal: official publication of the Federation of American Societies for Experimental Biology* **23**, 1325–1337 (2009).
93. Camateros, P. *et al.* Modulation of the allergic asthma transcriptome following resiquimod treatment. *Physiol Genomics* **38**, 303–318 (2009).
94. Haston, C. K., Cory, S., Lafontaine, L., Dorion, G. & Hallett, M. T. Strain-dependent pulmonary gene expression profiles of a cystic fibrosis mouse model. *Physiol Genomics* **25**, 336–345 (2006).
95. Chen, S. M., Ma, K. Y. & Zeng, J. Pseudogene: Lessons from PCR bias, identification and resurrection. *Molecular Biology Reports* **38**, 3709–3715 (2011).
96. Balaji, S. & Vanniarajan, A. Implication of Pseudo Reference Genes in Normalization of Data from Reverse Transcription-Quantitative PCR. *Gene* **757**, (2020).
97. Yin, R., Tian, F., Frankenberger, B., de Angelis, M. H. & Stoeger, T. Selection and evaluation of stable housekeeping genes for gene expression normalization in carbon nanoparticle-induced acute pulmonary inflammation in mice. *Biochemical and Biophysical Research Communications* **399**, 531–536 (2010).
98. Radonić, A. *et al.* Guideline to reference gene selection for quantitative real-time PCR. *Biochem Biophys Res Commun* **313**, 856–62 (2004).
99. de Jonge, H. J. M. *et al.* Evidence based selection of housekeeping genes. *PLoS ONE* **2**, (2007).

100. Chen, Y. C. *et al.* Epigenome-wide association study on asthma and chronic obstructive pulmonary disease overlap reveals aberrant DNA methylations related to clinical phenotypes. *Scientific Reports* **11**, (2021).
101. Kim, S. K., Jung, S. M., Park, K. S. & Kim, K. J. Integrative analysis of lung molecular signatures reveals key drivers of idiopathic pulmonary fibrosis. *BMC Pulmonary Medicine* **21**, (2021).
102. Xie, D., Zhou, Y. & Luo, X. Amorphous silica nanoparticles induce tumorigenesis via regulating ATP5H/ SOD1-related oxidative stress, oxidative phosphorylation and EIF4G2/PABPC1-associated translational initiation. *PeerJ* **2019**, (2019).
103. Feng, L., Cheng, P., Feng, Z. & Zhang, X. Transmembrane p24 trafficking protein 2 regulates inflammation through the TLR4/NF- $\kappa$ B signaling pathway in lung adenocarcinoma. *World Journal of Surgical Oncology* **20**, (2022).
104. Pouwels, S. D. *et al.* Genetic variance is associated with susceptibility for cigarette smoke-induced DAMP release in mice. *American Journal of Physiology - Lung Cellular and Molecular Physiology* **313**, (2017).
105. Shimada, I., Matsui, K., Iida, R., Tsubota, E. & Matsuki, T. Time course of housekeeping gene expression changes in diffuse alveolar damage induced by hyperoxia exposure in mice. *Legal Medicine* **11**, S151–S154 (2009).
106. Kouadjo, K. E., Nishida, Y., Cadrin-Girard, J. F., Yoshioka, M. & St-Amand, J. Housekeeping and tissue-specific genes in mouse tissues. *BMC Genomics* **8**, 127 (2007).

## Publikation

Diese Arbeit ist aktuell in *Scientific Reports* bei Nature publiziert und öffentlich zugänglich.

**scientific reports**



**OPEN** **Reference genes for the developing mouse lung under consideration of biological, technical and experimental confounders**

H. Shin<sup>1</sup>, R. E. Morty<sup>2,3</sup>, J. M. Sucre<sup>4</sup>, N. M. Negretti<sup>4</sup>, M. Markmann<sup>5</sup>, H. Hossain<sup>6</sup>, S. Krauss-Etschmann<sup>1,8,9</sup>, S. Dehmel<sup>1,10</sup> & A. Hilgendorff<sup>1,7</sup>✉

For gene expression analysis, the raw data obtained from RT-qPCR are preferably normalized to reference genes, which should be constantly expressed regardless of experimental conditions. Selection of reference genes is particularly challenging for the developing lung because of the complex transcriptional and epigenetic regulation of genes during organ maturation and injury repair. To date, there are only limited experimental data addressing reliable reference genes for this biological circumstance. In this study, we evaluated reference genes for the lung in neonatal C57BL/6 mice under consideration of biological, technical and experimental conditions. For that, we thoroughly selected candidates from commonly used reference genes side-by-side with novel ones by analyzing publicly available microarray datasets. We performed RT-qPCR of the selected candidate genes and analyzed their expression variability using GeNorm and Normfinder. Cell-specific expression of the candidate genes was analyzed using our own single-cell RNA-sequencing data from the developing mouse lung. Depending on the investigated conditions, i.e., developmental stages, sex, RNA quality, experimental condition (hyperoxia) and cell types, distinct candidate genes demonstrated stable expression confirming their eligibility as reliable reference genes. Our results provide valuable information for the selection of proper reference genes in studies investigating the neonatal mouse lung.

## Mitwirkung

Dankenswerterweise haben mehrere Wissenschaftler aus unterschiedlichen Institutionen zur Vervollständigung der vorliegenden Arbeit beigetragen. Prof. Dr. Morty, ehemals am Max-Planck-Institut für Herz- und Lungenforschung in Bad Nauheim (aktuell Translational Lung Research Center Heidelberg), hat RNA-Proben aus Mauslungen mit unterschiedlichen biologischen Merkmalen inklusive Hyperoxie-Exposition zur RT-qPCR-Validierung unserer Referenzgenkandidaten bereitgestellt. Prof. Dr. Hossain und Dr. Markmann, beide ehemals am Universitätsklinikum Gießen, haben das Microarray-Experiment in neonatalen Mauslungen zur Etablierung unserer eigenen Transkriptom-Daten durchgeführt und die Ergebnisse analysiert. Prof. Dr. Sucre und Dr. Negretti von der Vanderbilt University Medical Center, Nashville, USA haben die Expression unserer Referenzgenkandidaten in diversen pulmonalen Zelltypen mittels Einzelzell-RNA-Sequenzierung evaluiert. PD Dr. Hilgendorff und Dr. Dehmel sowie Prof. Dr.

Krauss-Etschmann haben dieses Projekt konzipiert und die Experimente beaufsichtigt. Außerdem haben sie bei der Interpretation der Ergebnisse und der Publikation der Arbeit wesentlich beigetragen.

## Danksagung

Mein erster Dank gilt Frau Prof. Dr. Susanne Krauss-Etschmann, die mich ursprünglich als ihren Doktoranden angenommen und so mir eine Möglichkeit geboten hat, in einem der renommierten Forschungszentren Deutschlands an einem Forschungsprojekt teilnehmen zu dürfen. Susanne hat auf diese Weise mein großes Interesse an Grundlagenforschung erweckt und mich damals trotz aller Schwierigkeiten im Studium zur Durchsetzung motiviert.

Aus meinem tiefsten Herzen möchte ich mich besonders bei Frau PD Dr. Anne Hilgendorff und Herrn Dr. Stefan Dehmel bedanken. Anne hat nach dem Weggang von Susanne aus dem Institut die Betreuung meiner Doktorarbeit bedingungslos übernommen und nun die Arbeit vertrauenswürdig bis zum Ende geführt. Sie hat mein Projekt ausgezeichnet betreut, indem sie immer wertvolle Zeit genommen hat, um mit meinen Fragestellungen auseinanderzusetzen. Vor allem durch zahlreiche Diskussionen während der Publikation hat Anne mir wissenschaftliche Denkweise beigetragen, was zukünftig für meine Forschungskarriere in der Pathologie bestimmt hilfreich wird. Außerdem hat sie durch ihre vielseitigen Forschungsideen aus langjähriger Erfahrung und Rekrutierung von verschiedenen Forschern unser Projekt auf höheres publizierbares Niveau angehoben.

Ich weiß nicht, wie ich am besten mein Dankeschön an Stefan äußern soll. Stefan hat mich von Beginn an mit unendlicher Geduld begleitet und mir in jedem kritischen Punkt meiner Doktorarbeit wertvolle Ratschläge gegeben. Darüber hinaus hat er mich in frustrierenden Momenten immer wieder zum Vorantreiben ermutigt. Ich kann ihm überhaupt nicht genug danken, dass er dazu für Komplettierung meiner Doktorarbeit nicht nur inhaltlich, als auch sprachlich sehr viel Zeit genommen hat.

Wie bereits unter dem gesonderten Abschnitt oben erwähnt, haben verschiedene Forscher aus unterschiedlichen Instituten zur Bereicherung und Vervollständigung dieses Projektes beigetragen. Ich möchte zuerst bei Herrn Prof. Dr. Hamid Hossain und Frau Dr. Melanie Markmann meinen herzlichen Dank aussprechen. Sie haben sich um Durchführung, Auswertung und Publizierung unserer eigenen Microarray-Daten aus neonatalen Mäusen bemüht und sich mir immer gerne zur Verfügung gestellt, um auf meine Fragen sehr ausführlich und nett zu beantworten. Ich darf nie vergessen, bei unseren Kollegen in USA, Prof. Dr. Jennifer Sucre und Dr. Nicholas Negretti meinen innigsten Dank auszusprechen. Ihre Daten aus der Einzelzell-RNA-

Sequenzierung haben die Qualität meines Projektes besonders erhoben. Ohne Herr Prof. Dr. Rory Morty wäre meine Arbeit gar nicht entstanden und so hat er zu meinem großen Dank die RNA-Proben aus dem Lungengewebe der zu validierenden Mäuse bereitgestellt.

Des Weiteren bin ich sehr dankbar, dass ich meine Dissertation am Comprehensive Pneumology Center des Helmholtz Zentrums München durchführen durfte.

Ich bedanke mich ganz herzlich bei meinen Eltern, die zur finanziellen und psychischen Unterstützung meines Medizinstudiums in Deutschland wirklich alle Möglichkeiten in Korea ausgeschöpft haben. Ich danke meiner Frau für ihr Dabeisein und meinen drei Kindern, die mich einerseits keinen einzigen Absatz dieser Arbeit ununterbrochen schreiben ließen, aber andererseits mir immer wieder neue Kraft zum Weitermachen gegeben haben.

Im Moment erinnere ich mich wieder an die unzähligen schlaflosen Nächte wegen unabsehbaren Abschlusses dieser Arbeit und meine Gebete an Gott in jenen Nächten. Aber ich bedanke mich aufrichtig, dass ich zur rechten Zeit meinen letzten Satz dieser Doktorarbeit schreiben kann.



LUDWIG-  
MAXIMILIANS-  
UNIVERSITÄT  
MÜNCHEN

Promotionsbüro  
Medizinische Fakultät



## Eidesstattliche Versicherung

Shin, Hyunkyu

Name, Vorname

Ich erkläre hiermit an Eides statt,

dass ich die vorliegende Dissertation mit dem Titel

**Identifizierung und Validierung von Referenzgenen zur relativen Quantifizierung der mRNA-Expression in neonatalem murinem Lungengewebe mittels RT-qPCR**

selbständig verfasst, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

Ich erkläre des Weiteren, dass die hier vorgelegte Dissertation nicht in gleicher oder in ähnlicher Form bei einer anderen Stelle zur Erlangung eines akademischen Grades eingereicht wurde.

Stuttgart, den 01.08.2022

Ort, Datum

Hyunkyu Shin

Unterschrift Doktorandin bzw. Doktorand