

# Identifizierung von Kandidatengenvarianten der Eutergesundheit beim Milchrind

Dissertation

zur

Erlangung des akademischen Grades

doctor rerum naturalium (Dr. rer. nat.)

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität Rostock



vorgelegt von

Mathias Ernst, geb. am 03.07. 1978 in Bad Langensalza

aus Rostock

**Gutachter:**

1. Gutachter

Prof. Dr. Manfred Schwerin

Leibniz-Institut für Nutztierbiologie (FBN), Dummerstorf

2. Gutachter

Prof. Dr. Renate Horn

Universität Rostock, Mathematisch-Naturwissenschaftliche Fakultät, Institut für  
Pflanzen-genetik

3. Gutachter

Prof. Dr. Georg Erhardt

Justus-Liebig-Universität Giessen, Fachbereich Agrarwissenschaften, Ökotoxikologie  
und Umweltmanagement, Institut für Tierzucht und Haustiergenetik

Datum der Einreichung: 26.10.2012

Datum der Verteidigung: 27.05.2013

# Inhaltsverzeichnis

<b>Einleitung</b>	<b>1</b>
<b>I Literaturübersicht</b>	<b>3</b>
<b>1 Eutergesundheit/Mastitis</b>	<b>3</b>
1.1 Definition, Klassifikation und Ökonomie . . . . .	3
1.2 Pathogene . . . . .	3
1.3 Euteranatomie und Mastitis . . . . .	4
1.4 Zucht auf Eutergesundheit . . . . .	6
1.5 Immunologische Mechanismen der Eutergesundheit . . . . .	9
1.5.1 Immunologische Besonderheiten in Wiederkäuern . . . . .	9
1.5.2 Die Bedeutung von Mechanismen der angeborenen, nichtadaptiven Immunität für die Eutergesundheit . . . . .	9
1.5.3 Die Bedeutung von Mechanismen der adaptiven Immunität für die Eutergesundheit . . . . .	10
B-Lymphocyten . . . . .	10
T-Lymphocyten . . . . .	12
<b>2 Ansätze der molekularen Genomanalyse</b>	<b>13</b>
2.1 Experimentelle Ansätze der strukturellen Genomanalyse . . . . .	13
2.1.1 Kartierung merkmalsassoziierter Genomregionen . . . . .	13
2.1.2 Identifizierung positioneller Kandidatengene (vergleichende Genkartierung) . . . . .	16
2.1.3 Der BTA18-SCS-QTL . . . . .	17
2.1.4 Identifizierung merkmalsassoziierter Genvarianten in potentiellen funktionellen Kandidatengenen . . . . .	19
2.2 Experimentelle Ansätze der funktionellen Genomanalyse . . . . .	21
2.2.1 Holistische mRNA-Untersuchungen - die DNA-Microarray-Technologie	21
2.2.2 DNA-Microarrayanalysen als Werkzeug zur Dissektion genetischer Resistenz . . . . .	24
2.3 Integrative Ansätze zur Identifizierung potentieller funktioneller Kandidatengene . . . . .	25
2.3.1 Funktionelle Annotation . . . . .	25
2.3.2 Expressionsdaten . . . . .	27
2.4 QTL und Kandidatengene der Eutergesundheit . . . . .	28
<b>II Material und Methoden</b>	<b>31</b>
<b>1 Ein bioinformatischer heuristischer Ansatz zur Ermittlung einer Rangfolge potentieller funktioneller Kandidatengene</b>	<b>31</b>
1.1 Ermittlung und Beschaffung ( <i>Retrieval</i> ) relevanter Publikationen für die positionellen Kandidatengene . . . . .	31
1.2 Ermittlung des datenbankbasierten Rangkoeffizienten . . . . .	35
1.2.1 MEDLINE-Format . . . . .	35

1.2.2	Reguläre Ausdrücke . . . . .	36
1.2.3	Der datenbankbasierte Rangkoeffizient . . . . .	37
1.3	Der kombinierte Datenbank/Transkriptom-Ansatz: Integration der Datensätze	39
<b>2</b>	<b>Statistische Auswertung experimenteller Daten</b>	<b>40</b>
2.1	Hardy-Weinberg-Gleichgewicht . . . . .	40
2.2	Kopplungsungleichgewicht . . . . .	40
2.3	Euklidische Distanz . . . . .	40
2.4	SNP-Assoziationsuntersuchungen . . . . .	41
<b>3</b>	<b>Verwendete Software</b>	<b>44</b>
3.1	Perl . . . . .	44
3.2	R . . . . .	44
3.3	Emboss . . . . .	45
3.4	BLAST . . . . .	45
<b>4</b>	<b>Materialien und Tiere</b>	<b>46</b>
4.1	Bullen mit extrem differenter Verteilung des SCS-Schätzwertes MA-BLUP .	46
4.2	Bullen mit Heterozygosie am BTA18-SCS-QTL . . . . .	46
4.3	Auswahl der Stichprobe für die Transkriptomstudie des Verbundprojektes M.A.S.-Net . . . . .	46
4.4	Die für die Assoziationsuntersuchungen genutzte DH-Halbgeschwisterpopulation	49
<b>5</b>	<b>Labormethoden</b>	<b>50</b>
5.1	Polymerasekettenreaktion . . . . .	50
5.2	Agarosegelelektrophorese . . . . .	50
5.3	Bestimmung der DNA-Konzentration . . . . .	52
5.4	DNA-Sequenzierung nach Sanger . . . . .	52
5.5	Pyrosequenzierung <sup>TM</sup> . . . . .	53
<b>III</b>	<b>Ergebnisse</b>	<b>54</b>
<b>1</b>	<b>Identifizierung von potentiellen funktionellen positionellen Kandidatengen im BTA18-SCS-QTL mittels eines heuristischen bioinformatischen Ansatzes</b>	<b>54</b>
1.1	Identifizierung positioneller Kandidatengene des BTA18-SCS-QTL mittels der vergleichenden Genkarten . . . . .	54
1.2	Die Anwendung des heuristischen Datenbankansatzes zur Identifizierung po- tentieller funktioneller Kandidatengene . . . . .	56
1.2.1	Ermittlung und Beschaffung ( <i>Retrieval</i> ) relevanter Literatur . . . . .	56
1.2.2	Die Berechnung des datenbankbasierten Rangkoeffizienten auf Grund- lage der ermittelten Referenzliteratur . . . . .	58
1.3	Der kombinierte Ansatz: Integration von Datenbankansatz und Transkriptom- studie . . . . .	61
1.3.1	BTA18-SCS-QTL-spezifische Ergebnisse der M.A.S.-Net Transkriptom- studie . . . . .	61
1.3.2	Integration der QTL-spezifischen Transkriptomresultate in den heuri- stischen datenbankbasierten Ansatz . . . . .	62

<b>2</b>	<b>Identifizierung und vergleichende Sequenzierung von SNPs in den potentiellen funktionellen Kandidatengen</b>	<b>68</b>
2.1	Genomische Darstellung der SNPs . . . . .	68
2.1.1	SNPs im Gen <i>CD22</i> . . . . .	70
2.1.2	SNPs im Gen <i>CD79A</i> . . . . .	71
2.1.3	SNPs in den Genen <i>HCST und TYROBP</i> . . . . .	73
2.2	Selektive Poolgenotypisierung . . . . .	75
2.3	Konkordanztest . . . . .	78
2.4	Fazit des mehrstufigen Auswahlverfahrens zur Bestimmung potentiell merkmalsassoziierter SNPs . . . . .	80
<b>3</b>	<b>Merkmalsassoziationen</b>	<b>81</b>
3.1	Genotypisierung . . . . .	81
3.2	Modellierung . . . . .	83
3.2.1	<i>CD22</i> . . . . .	85
3.2.2	<i>CD79A</i> . . . . .	87
3.2.3	<i>HCST</i> . . . . .	90
<b>IV</b>	<b>Diskussion</b>	<b>91</b>
1	Der bioinformatische Ansatz	92
2	Die ausgewählten Kandidatengene	98
3	Identifizierung potentiell merkmalsassoziierter SNPs	100
<b>V</b>	<b>Zusammenfassung</b>	<b>108</b>
<b>VI</b>	<b>Appendix</b>	<b>I</b>
<b>VII</b>	<b>Literaturverzeichnis</b>	<b>XXI</b>

## Tabellenverzeichnis

1	Klassifikation von Störungen der Eutergesundheit bzw. Mastitiden nach den Kriterien der DVG (DVG, 2002) . . . . .	4
2	Überblick über die vier von Ron & Weller (2007) aufgestellten Selektionskriterien für potentielle funktionelle Kandidatengene quantitativer Merkmale beim Nutztier auf . . . . .	26
3	Beschreibung der Klassen innerhalb des objektorientierten Ansatzes, die für die Behandlung der beim heuristischen datenbankbasierten Ansatz verwendeten Daten definiert wurden . . . . .	31
4	Die Datenbank Gene2PubMed listet Referenzliteratur für Gene . . . . .	32
5	Die in den EUtils zusammengefassten Services . . . . .	33
6	Die für die realisierten EFetch-Anfragen verwendeten Parameter . . . . .	34
7	Die in dieser Arbeit für die Implementierung des heuristischen datenbankbasierten Ansatzes genutzten Metasymbole in Regulären Ausdrücken . . . . .	36
8	Merkmale der in der FUGATO M.A.S.-Net-Transkriptomstudie untersuchten Färsengruppen QQ und qq . . . . .	48
9	Genomische PCR: Zusammensetzung der Reaktionsansätze . . . . .	51
10	Genomische PCR: Thermocyclerprogramm . . . . .	51
11	Thermocyclerprogramm für die Sequenzierung von DNA-Fragmenten nach der Sanger-Methode (Sanger <i>et al.</i> , 1977) . . . . .	52
12	Die Ausdehnung der Intervalle des BTA18-SCS-QTL und der syntenen Region auf HSA19 . . . . .	55
13	Die Anzahl der Publikationen, die für die positionellen Kandidatengene in der Datenbank Gene2PubMed identifiziert wurden . . . . .	56
14	Die für die Spezifizierung der immunologischen Hypothese gewählten Regulären Ausdrücke . . . . .	58
15	Die wahrscheinlichsten potentiellen funktionellen Kandidatengene der Eutergesundheit im BTA18-SCS-QTL gemäß dem heuristischen datenbankbasierten Ansatz . . . . .	60
16	Die Zahl differentiell exprimierter Probesets unter den 212 Probesets, welche 142 positionelle Kandidatengene repräsentieren . . . . .	63
17	Kontingenztafel des Abgleichs der EntrezGeneID-Annotationen von 254 positionellen Kandidatengenen mit der Annotation des verwendeten bovinen <i>GeneChip</i> <sup>®</sup> . . . . .	63
18	Die wahrscheinlichsten potentiellen funktionellen Kandidatengene der Eutergesundheit im BTA18-SCS-QTL gemäß dem kombinierten Datenbank/Transkriptom-Ansatz . . . . .	67
19	Ausdehnung und genomische Lokalisation (im bovinen Build 3.1) der vier positionellen und potentiellen funktionellen Kandidatengene <i>CD22</i> , <i>CD79A</i> , <i>HCST</i> und <i>TYROBP</i> des BTA18-SCS-QTL . . . . .	69
20	Namen und Lokalisation identifizierter SNPs im Gen <i>CD22</i> . . . . .	70
21	Namen und Lokalisation identifizierter SNPs im Gen <i>CD79A</i> . . . . .	72
22	Namen und Lokalisation identifizierter SNPs in den Genen <i>HCST</i> und <i>TYROBP</i> . . . . .	74
23	Relative Allelverhältnisse von 16 in den ausgewählten potentiellen funktionellen Kandidatengenen identifizierten SNPs in zwei genetisch extrem divergenten DNA-Pools . . . . .	76

24	Konkordanztest an Kandidaten-SNPs in Bullen mit heterozygoten Allelen des BTA18-SCS-QTL: Zusammenfassung . . . . .	78
25	Konkordanztest an Kandidatengen-SNPs in Bullen mit heterozygoten Allelen des BTA18-SCS-QTL: Einzelresultate . . . . .	79
26	Zusammenfassung der Ergebnisse der Typisierung von sechs SNPs im DH-Halbgeschwisterdesign . . . . .	81
27	Signifikante Assoziationen von SNPs der Gene <i>CD22</i> und <i>HCST</i> mit Konformations- bzw. Milchleistungsmerkmalen . . . . .	86
28	Signifikante Assoziationen von SNPs im Gen <i>CD79A</i> mit Merkmalen der Eutergesundheit (SCS) und Konformation . . . . .	89
29	Differentiell exprimierte Probesets für Kandidatengene des BTA18-SCS-QTL	I
30	Resultate der M.A.S.-Net Transkriptomstudie: Expressionswerte . . . . .	III
31	Potentielle Kandidatengene der bovinen Eutergesundheit nach Ogorevc <i>et al.</i> (2009) . . . . .	VIII
32	Merkmalsbeeinflussende Regionen (QTL) für die Merkmale SCC und SCS in verschiedenen Milchviehrassen . . . . .	IX
33	Übersicht über die verwendeten Perlmodule . . . . .	X
34	Genomische Primer für PCR und Sequenzierung von SNPs in potentiellen funktionellen Kandidatengenen . . . . .	XI
35	DNA-Pyrosequenzierung: Genomische Primer . . . . .	XII
36	Phänotypen der typisierten DH-Halbgeschwisterpopulation . . . . .	XIII
37	Die Verteilung der für die Assoziationsstudie an sechs SNPs typisierten DH-Bullen auf die sechs Halbgeschwisterfamilien und paternale Verwandtschaftsverhältnisse . . . . .	XIV
38	Übersicht über die Ergebnisse der Assoziation von sechs SNPs mit dem SCS in den ersten drei Laktationsperioden sowie mehreren Konformations- und Milchleistungsphänotypen . . . . .	XV
39	Promotorelemente vom Typ $\gamma$ -IRE wurden in verschiedenen Publikationen beschrieben und teilweise funktional charakterisiert. Ein $\gamma$ -IRE wird durch den Promotor-SNP CD79A_-182:A>G disruptiert. . . . .	XVI
40	Verschiedene Parameter der Beschaffenheit von Euter und Fundament . . . .	XVII

## Abbildungsverzeichnis

1	Darstellung der Anatomie eines Milchvieheuters (Sagittalschnitt) . . . . .	5
2	Der BTA18-SCS-QTL in seinen Kartierungsstadien . . . . .	18
3	EUtills: Beispielcode für die Übermittlung einer Datenbankabfrage . . . . .	34
4	Pseudocode der Prozedur zur Ermittlung des Rangkoeffizienten DbRK auf Grundlage des heuristischen datenbankbasierten Ansatzes . . . . .	38
5	Box: Statistische Modelle für die SNP-Assoziationen . . . . .	42
6	Selektive Poolsequenzierung: Zuchtwerte der Extremwertpools . . . . .	47
7	Selektionsstrategie der vergleichenden Transkriptomstudie des Verbundprojektes M.A.S.-Net . . . . .	48
8	Konservierung humaner positioneller Kandidatengene im Rind . . . . .	55
9	Plots der Rangkoeffizienten gegen die Anzahl von Referenzpublikationen bzw. die Anzahl von <i>matches</i> für 129 positionelle Kandidatengene . . . . .	59
10	Boxplots der Rangdifferenz zwischen dem kombinierten Datenbank/Transkriptom-Ansatz und dem heuristischen datenbankbasiertem Ansatz für 177 Probesets, gruppiert nach Expressionsstatus . . . . .	64
11	Scatterplots der Rangdifferenz zwischen dem kombinierten Datenbank/Transkriptom-Ansatz und dem heuristischen datenbankbasiertem Ansatz für 37 Probesets . . . . .	65
12	CD22: Struktur und SNPs . . . . .	71
13	CD79A: Struktur und SNPs . . . . .	72
14	Der genomische Kontext des Promotor-SNP CD79A_-182:A>G . . . . .	73
15	HCST und TYROBP: Struktur und SNPs . . . . .	74
16	Beispiele für Sequenzchromatogramme, die in der selektiven Poolsequenzierung gewonnen und für die Abschätzung von Allelfrequenzen verwendet wurden . . . . .	77
17	Linkage Disequilibrium-Plot der sechs in der DH-Halbgeschwisterpopulation typisierten SNPs . . . . .	82
18	Cluster- <i>Heatmap</i> der Resultate der SNP-Assoziationsberechnungen nach dem Einzel-SNP-Modell . . . . .	84
19	Relative Schätzwertgrößen für die Effekte der homozygoten Genotypen der <i>CD79A</i> -SNPs . . . . .	88
20	Die Frequenzverteilung von Genen, denen gemäß Hintermair (2007) mehrere Probesets auf dem bovinen <i>GeneChip</i> <sup>®</sup> zugeordnet werden, in Abhängigkeit von der Anzahl der Probesets pro Gen . . . . .	XIX
21	Beispiel für das MEDLINE-Textformat von in der PubMed indizierten Publikationen . . . . .	XX



## Abkürzungen

---

%	Prozent
°C	Grad Celsius
A	Adenin
Abb.	Abbildung
ABCG2	<i>ATP-binding cassette, sub-family G (WHITE), member 2</i>
Add. GE	additiv genetischer Effekt
AFLP	<i>amplified fragment-length polymorphism</i>
ANOVA	<i>analysis of variance</i>
APC	<i>antigen presenting cell</i>
ATP	Adenosintriphosphat
BAC	<i>bacterial artificial chromosome</i>
BCL3	<i>B-cell lymphoma 3</i>
BCR	B-Zell-Rezeptor
BLAST	<i>basic local alignment search tool</i>
BLNK	<i>B-cell linker</i>
BLUP	<i>best linear unbiased predictor</i>
BMBF	Bundesministerium für Bildung und Forschung
BoLA	<i>bovine leukocyte antigen (syn. MHC)</i>
bp	<i>basepair</i>
BSG	<i>bidirectional selective genotyping</i>
BTA	Bos taurus autosome
Btau X.Y	bovine Genomsequenz in Version X.Y
C	Cytosin
Ca	Calcium
CD	<i>cluster of differentiation</i>
cDNA	<i>complementary DNA</i>
CDS	<i>coding DNA sequence</i>
CEACAM	<i>carcinoembryonic antigen-related cell adhesion molecule</i>
ChIP	Chromatin-Immunopräzipitation
cm	centiMeter
cM	centiMorgan
COCA	<i>Coordinative Component Analysis</i>
CPAN	<i>Comprehensive Perl Archive Network</i>
CRAN	<i>Comprehensive R Archive Network</i>
δ	delta
DAP	DNAX-activation protein
DbRk	Datenbank-Rangkoeffizient
DbTRK	Datenbank/Transkriptom-Rangkoeffizient
dbSNP	SNP-Referenzdatenbank des NCBI
dd	didesoxy
DGAT1	<i>diacylglycerol O-acyltransferase 1</i>
DH	Deutsche Holstein
DHV	Deutscher Holstein Verband e.V.
DNA	desoxyribonucleic acid
DVG	Deutsche Veterinärmedizinische Gesellschaft

---

---

DYD	<i>daughter yield deviation</i>
DZ	Drüsenzisterne
ε	Restfehler
ebd.	ebenda
EDC	<i>effective daughter contribution</i>
EDTA	Ethylendiamintetraessigsäure
EMBOSS	<i>European Molecular Biology Open Software Suite</i>
EMSA	<i>Electrophoretic Mobility Shift Assay</i>
Ets	E-twenty six
€	Euro
EUutils	<i>Entrez programming utilities</i>
e.V.	eingetragener Verein
FC	<i>fold change</i>
FcR	Immunglobulinrezeptor
F <sub>con</sub>	F-Statistik nach konditioneller Methode
FEZL	<i>forebrain embryonic zinc finger like</i>
fwd.	<i>forward</i>
FISH	<i>fluorescence in-situ hybridization</i>
FR	Fürstenbergsche Rosette
FUGATO	Funktionelle Genomanalyse im tierischen Organismus
γ	Gamma
γ-IRE	Interferon-γ <i>responsive element</i>
g	Gramm
G	Guanin
GC	Anteil von G und C an einer Nukleotidsequenz
GDD	<i>granddaughter design</i>
GH	<i>growth hormone</i> , als Name eines Genes
GRB2	<i>growth factor receptor-bound protein 2</i>
GSD	<i>gene symbol disambiguation</i>
GWAS	<i>genomewide association study</i>
h <sup>2</sup>	Heritabilität
H <sub>2</sub> O	Wasser
HCST	<i>hematopoetic cell signal transducer</i>
HNRNPK	<i>heterogeneous nuclear ribonucleoprotein K</i>
HoA	<i>Hash of Arrays</i>
HSA	Homo sapiens autosome
HTML	<i>hypertext markup language</i>
HWG	Hardy-Weinberg-Gesetz/Gleichgewicht
IBD	<i>identical by descent</i>
ID	Identifikationsnummer
IFN	Interferon
Ig	Immunglobulin
IL	Interleukin
IMI	<i>intramammal infections</i>
IRF2BP1	<i>interferon regulatory factor 2 binding protein 1</i>
IT(A/I)M	<i>immune tyrosin-based (activation/inhibition) motif</i>
IUPAC	<i>International Union of Pure and Applied Chemistry</i>

---

---

Kap.	Kapitel
Kb	Kilobasen
KIR	<i>killer immunoglobulin-like receptor</i>
KLK	Kallikrein
KLR	<i>killer cell lectin-like receptor</i>
l	Liter
LA	<i>linkage analysis</i>
LAIR	<i>leukocyte-associated immunoglobulin-like receptor</i>
LD	<i>linkage disequilibrium</i>
LILR	<i>leukocyte immunoglobulin-like receptor</i>
Limma	<i>linear models for microarray data</i>
LKT	Leukotoxin
LN	Lymphknoten
LOD	<i>logarithm of odds</i>
LOESS	<i>locally estimated scatterplot smoothing</i>
log	Logarithmus
LRT	<i>likelihood ratio test</i>
M	molar (mM: milliMolar, nM: nanoMolar)
MA(S)	<i>marker assisted (selection)</i>
MAF	<i>minor allele frequency</i>
MAQC	<i>MicroArray Quality Control Consortium</i>
Mb	Megabasen
MEC	<i>mammary epithelial cell</i>
MeSH	<i>Medical Subject Headings</i>
MHC	<i>major histocompatibility complex</i>
MIAME	<i>minimum information about a microarray experiment</i>
min	Minute
miRNA	microRNA
mRNA	<i>messenger RNA</i>
n	Stichprobenumfang
NAb	<i>natural antibody</i>
NCBI	<i>National Center for Biotechnology Information</i>
NF- $\kappa$ B	<i>nuclear factor kappa-light-chain-enhancer of activated B-cells</i>
NK	Natürliche Killerzelle
NTP	Nukleosidtriphosphat
OMIM	<i>Online Mendelian Inheritance in Man</i>
p	Irrtumswahrscheinlichkeit
PA	Parenchym
PAMP	<i>pathogen associated molecular pattern</i>
PCR	<i>polymerase chain reaction</i>
PCRE	<i>perl compatible regular expressions</i>
Perl	<i>practical extraction and reporting language</i>
PI3K	<i>phosphatidylinositol-4,5-bisphosphate 3-kinase</i>
PLC $\gamma$ 2	phospholipase C, gamma 2
PLIER	<i>Probe Logarithmic Intensity Error</i>
PMID	pubmed identifier
PPi	Pyrophosphat

---

---

PSA	<i>prostate specific antigen</i>
PVRL2	<i>poliovirus receptor-related 2</i>
Q	Allel des BTA18-SCS-QTL für erhöhte Mastitisresistenz
q	Allel des BTA18-SCS-QTL für verminderte Mastitisresistenz
QT(L/N)	<i>quantitative trait (locus/nucleotide)</i>
QTDT	<i>(quantitative) transmission disequilibrium test</i>
RELB	<i>v-rel avian reticuloendotheliosis viral oncogene homolog B</i>
REML	<i>residual maximum likelihood</i>
rev.	<i>reverse</i>
ROS	<i>reactive oxygen species</i>
RNA	ribonucleic acid
rs (+ Ziffern)	bezeichnet die Referenznummer eines SNP
RZ	Relativer Zuchtwert
RZS	Relativer Zuchtwert Zellzahl
$\sigma$ , s	Standardabweichung
$s_r$	Residuenstandardfehler
S.	Staphylococcus
s.o. / s.u.	siehe oben / siehe unten
Sc.	Streptococcus
SCC	<i>Somatic Cell Count</i>
SCS	<i>Somatic Cell Score</i>
SDP	<i>selective DNA pooling</i> (selektive Poolsequenzierung)
sec	Sekunde
SIGLEC	<i>sialyl binding ig-like lectin</i>
SHC	<i>Src homology 2 domain containing</i>
SHP-1	<i>SH2-domain-containing protein tyrosine phosphatase 1</i>
SK	Strichkanal
SNP	<i>single nucleotide polymorphism</i> , Einzelbasenaustausch
SPP1	<i>secreted phosphoprotein 1</i> (=Osteopontin)
SQL	<i>structured query language</i>
ss	<i>single stranded</i>
ss (+ Ziffern)	bezeichnet eine Neuübermittlung an <i>dbSNP</i>
ssh	<i>secure shell</i>
syn.	synonym
T	Thymin
Tab.	Tabelle
TAE	Tris-Acetat-EDTA
TF	Transkriptionsfaktor
TGFB	<i>transforming growth factor beta</i>
TLR	<i>toll-like receptor</i>
$T_m$	Schmelztemperatur
TREM	<i>triggering receptor expressed on myeloid cells</i>
TYROBP	<i>TYRO protein tyrosine kinase-binding protein</i>
U	unit
u.a.	unter anderem
UID	<i>unique ID</i>
URL	<i>uniform resource locator</i>

---

---

USG	<i>unidirectional selective genotyping</i>
UV	ultraviolett
vgl.	vergleiche
VIT	Vereinigte Informationssysteme Tierzucht
XML	<i>extensible markup language</i>
z.B.	zum Beispiel
$\mu$	Mittelwert
$\infty$	unendlich

# Einleitung

Intramammäre Euterinfektionen, d.h. Mastitiden, stellen für die Milchviehwirtschaft eine immense Belastung dar, und zwar sowohl ökonomisch durch den verminderten Milchertrag und dessen Folgen (Halasa *et al.*, 2007) als auch für das Wohlbefinden der Kuh. Die für lange Zeit vorherrschende Fokussierung auf erhöhte Milchleistung und -qualität als Zuchtziel beim Milchrind hat diese Problematik verschärft. Durch verbesserte Managementmethoden, z.B. in der Hygiene, konnte eine Verschiebung des Erregerspektrums erreicht, die Mastitiden aber nicht entscheidend zurückgedrängt werden; ebenso sind Forschungen in Richtung einer Mastitisimpfung bis dato erfolglos geblieben. Daher ist die Identifizierung von Kandidatengen und Genvarianten, die die Mastitisempfindlichkeit beeinflussen, von größtem Interesse.

Dem stehen jedoch erhebliche Schwierigkeiten entgegen: Zum einen ist man bei Mastitisresistenzstudien auf die Analyse eines genetisch korrelierten Surrogatmerkmals, dem vom Gehalt somatischer Zellen in der Milch abgeleiteten *Somatic Cell Score*, SCS, angewiesen. Für dieses sehr polygene Merkmal wurden merkmalsbeeinflussende Regionen (*quantitative trait loci*, QTL) auf nahezu allen bovinen Autosomen (BTA) identifiziert (Hu *et al.*, 2007).

QTL-Kartierungen haben eine inhärent niedrige Auflösung; dementsprechend sind für einen QTL gegebenenfalls hunderte positionelle Kandidatengene in Betracht zu ziehen. Feinkartierungen von QTL (Meuwissen & Goddard, 2000) können deren Auflösung um den Preis erhöhter Komplexität der Analyse erheblich vergrößern (z.B. Sahana *et al.* (2008); insbesondere im Fall von QTL in genomischen Regionen mit großer Gendichte ist dieser Ansatz im Sinne der Reduktion auf wenige, experimentell überprüfbare Kandidatengene allerdings nicht hinreichend.

Gefordert sind hier bioinformatische Methoden zur Priorisierung von Kandidatengen nach funktionellen Kriterien. Ein quantitativer Ansatz hierfür kann darüber hinaus die Integration von Ergebnissen aus Hochdurchsatz-*omics*-Studien ermöglichen, die in zunehmendem Maße zur Dissektion genetischer Resistenzen, auch beim Milchrind, herangezogen werden.

Ist, idealerweise gut begründet auf Basis solcher integrativer Ansätze, eine Auswahl potentieller funktioneller Kandidatengene getroffen, besteht die nächste Hürde im konkreten Nachweis einer Assoziation von genomischen Variationen, meist Einzelbasenaustausche (*single nucleotide polymorphisms*, SNPs), dieser Gene mit dem gewünschten Merkmal, z.B. dem SCS. Die speziellen Verhältnisse bei großen Nutztieren (großes Generationsintervall, keine oder kaum transgene Tiere bzw. Inzuchtlinien verfügbar, Populationsstrukturen) machen es notwendig, eine Variation mehreren Tests mit jeweils verschiedenem Design und untersuchten Hypothesen zu unterziehen (Mackay, 2001; Ron & Weller, 2007); erst die Kombination unabhängiger Indizien, die aus diesen Tests abgeleitet werden, erlaubt die Evaluierung einer Merkmalsassoziation. Ein solch rigoroser Nachweis ist bisher beim Milchrind nur in wenigen Fällen gelungen; Beispiele sind *DGAT1* (Grisart *et al.*, 2002), *ABCG2* (Cohen-Zinder

*et al.*, 2005; Seroussi, 2009) und *GH* (Blott *et al.*, 2003) für Leistungsmerkmale und *FEZL* (Sugimoto *et al.*, 2006, 2011) für Eutergesundheit.

In der hier vorgestellten Arbeit im Rahmen des BMBF-geförderten Verbundprojektes M.A.S.-Net wird der Versuch unternommen, potentielle funktionelle Kandidatengene und -varianten eines mehrfach bestätigten (Kühn *et al.*, 2003; Schulman *et al.*, 2004; Xu *et al.*, 2006) und feinkartierten (Baes *et al.*, 2009) SCS-QTL auf Rinderautosom (BTA) 18 zu identifizieren. Dabei sollen beide dargestellten thematischen Ebenen adressiert werden. Zunächst wird eine skriptgesteuerte bioinformatische Methode zur quantitativen Priorisierung positioneller Kandidatengene entwickelt und auf die Gene des QTL angewendet. Sie beruht auf der Auswertung standardisierter *Medical Subject Headings* (*MeSH*) Annotationen wissenschaftlicher Referenzpublikationen und deren Abgleich mit Sets von Stichworten, die die begründete Hypothesen über die immunologische Natur genetischer Resistenzmechanismen gegen Mastitis abbilden. Die Integration dieses Ansatzes mit einer vergleichenden Transkriptomstudie, die sich durch ein besonderes und im Sinne der Dissektion genetischer Mastitisresistenz aussagekräftiges Design auszeichnet, bildet ein weiteres Thema.

Im anschließenden Teil der Arbeit werden die durch Anwendung des integrativen Ansatzes identifizierten potentiellen funktionelle Kandidatengene auf SNPs gescreent und diese in drei konsekutiven, in nichtüberlappenden Populationen der Milchviehrasse Deutsche Holstein durchgeführten Tests, nämlich einer von der selektiven Poolsequenzierung, SDP (Darvasi & Soller, 1994), abgeleiteten Sequenzierung von Bullen mit entgegengesetzt extremen Zuchtwerten für Eutergesundheit, einem Konkordanztest an QTL-heterozygoten Bullen (vgl. Seroussi, 2009) und einer Assoziationsstudie an einer großen Halbgeschwisterpopulation auf Assoziationen mit Eutergesundheits- und Konformationsmerkmalen untersucht. Obwohl bei der Interpretation der Resultate aufgrund des Auswahlverfahrens der SNPs ein potentieller Selektionsbias zu berücksichtigen ist, werden im Sinne von Ron & Weller (2007) Indizien präsentiert, die auf eine Beteiligung von SNPs im Gen *CD79A* (bzw. mit diesen gekoppelte SNPs) an genetischen Mechanismen der Mastitisresistenz hindeuten.

## Teil I

# Literaturübersicht

## 1 Eutergesundheit/Mastitis

### 1.1 Definition, Klassifikation und Ökonomie

Unter dem Begriff der Mastitiden wird ein Spektrum von Infektionen der Milchdrüsen (*intramammal infections, IMI*) in der Gesamtheit ihrer milchbildenden, speichernden und ableitenden Abschnitte subsumiert (Wendt *et al.*, 1994). Sie sind sehr divers in ihrer Ätiologie und Manifestation; letztere erstreckt sich von phänotypisch unauffälligen subklinischen Fällen bis hin zur akuten klinischen Mastitis. Der gemeinsame Nenner der Mastitiden ist ein erhöhter Gehalt an somatischen Zellen in der Milch (*Somatic Cell Count, SCC*). Tabelle 1 gibt einen Überblick über die Klassifikation der Mastitiden, bei der der SCC eine wichtige Rolle spielt, gemäß der Definition durch die Deutsche Veterinärmedizinische Gesellschaft (DVG).

Zusammengenommen gehören die Mastitiden zu den Milchvieherkrankungen mit dem höchsten Schadenspotential. In einer vergleichenden Analyse der Produktionskosten verschiedener Milchvieherkrankungen in England führen Kossaibati & Esslemont (1997) die Mastitis mit einem Anteil von 38% der Gesundheitskosten an erster Stelle, vor den Klauenerkrankungen mit 27%. Nach Roosen *et al.* (2004) steht sie nach den Fruchtbarkeitsstörungen an zweiter Stelle unter den Ursachen für vorzeitige Schlachtung deutscher Kühe.

Über die Ursachen der durch Mastitis verursachten Kosten gibt es in der Literatur unterschiedliche Angaben, da hier viele direkte bzw. indirekte Kostenfaktoren zu berücksichtigen sind. In einer Metaanalyse von seit 1990 erschienenen Publikationen zu dieser Problematik werden u.a. Medikation, verworfene Milch, Veterinärservice, Arbeitsaufwand, (verminderte) Produktqualität, Materialinvestitionen, Diagnostik, Folgeerkrankungen und Merzung zu variablen Anteilen als Faktoren genannt (Halasa *et al.*, 2007). Den Hauptanteil, beispielsweise 56-68% der Verluste durch klinische Mastitis in einer Studie durch Wolfova *et al.* (2006), macht jedoch die Verminderung der Milchproduktion aus, vgl. dazu auch Hillerton *et al.* (1992) und McInerney *et al.* (1990).

### 1.2 Pathogene

Hauptverursacher boviner Mastitiden sind Bakterien, die nach der Ätiologie der durch sie verursachten Erkrankungen in kontagiöse euterassoziierte und umweltassoziierte Pathogene unterschieden werden (Smith & Hogan, 1995).

Zu den ersteren, die gut an das Eutermilieu angepasst sind und vor allem subklinische Mastitiden verursachen, zählen *Streptococcus agalactiae* (der Erreger des Gelben Galttes),



**Tabelle 1:** Klassifikation von Störungen der Eutergesundheit bzw. Mastitiden nach den Kriterien der DVG (DVG, 2002). Zur Definition des SCC siehe Text bzw. Formel 1.

---

• <b>Normale Sekretion:</b>	keine pathologischen Veränderungen; $SCC \leq 100.000$
• <b>Latente Infektion:</b>	Pathogene nachgewiesen; $SCC \leq 100.000$
• <b>Unspezifische Mastitis:</b>	keine Pathogene nachgewiesen; aber Symptome einer Entzündung (Rötung, Schwellung, erhöhte Temperatur, Schmerzen); $SCC > 100.000$
• <b>Mastitis:</b>	Pathogene nachgewiesen, $SCC > 100.000$
▶ <b>Subklinische M.:</b>	keine äußerlichen Symptome
▶ <b>Klinische M.:</b>	
- <b>akut:</b>	Entzündungssymptome liegen vor; makroskopische Veränderungen der Milch
- <b>subakut:</b>	Entzündungssymptome können fehlen; makroskopische Veränderungen der Milch, besonders im Vorgemelk
▶ <b>Chronische M.:</b>	zunehmende Proliferation von fibrinösem Gewebe; klinischer oder subklinischer Verlauf; keine Entzündungssymptome; makroskopische Veränderungen der Milch (nicht immer)

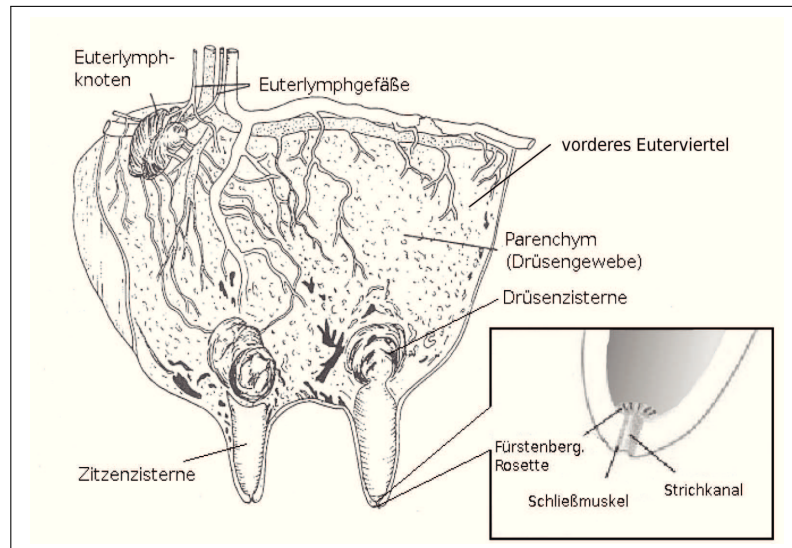
---

*Sc. dysgalactiae* und *Staphylococcus aureus*. Durch konsequente Einhaltung eines 5-Punkte-Mastitiskontrollprogrammes, dessen Kernelemente Hygiene und Antibiotikagabe sind, konnten diese Pathogene in den letzten Jahrzehnten spürbar zurückgedrängt werden (Dodd, 1983; Myllys *et al.*, 1994; Piepers *et al.*, 2007). Dies führte zu einer Verschiebung des Erregerspektrums hin zu umweltassoziierten Keimen (Myllys *et al.*, 1994), die nativ die Umgebung der Kuh sowie beispielsweise deren Gastrointestinalsystem (Darmbakterien) besiedeln. Dringen sie jedoch in die Milchdrüsen ein, so lösen sie dort klinische Mastitiden mit hoher Pathogenität und Mortalität aus. Für den größten Teil (60-68% in Shpigel *et al.*, 1998) der klinischen Mastitiden sind gram-negative Koliforme der Gattungen *Escherichia*, *Klebsiella* und *Enterobacter* verantwortlich (Menzies *et al.*, 1995), ein kleinerer Teil wird durch *S. uberis* verursacht (Shpigel *et al.*, 1998).

Neben diesen klassischen Erregern werden gelegentlich *Mycoplasma spp.*, verschiedene Hefen und Algen in Eutergewebe und Milchproben nachgewiesen (Costa *et al.*, 1998).

### 1.3 Euteranatomie und Mastitis

Die Milchdrüse ist eine modifizierte apokrine Schweißdrüse, die als zusammengesetzte tubulöse Drüse mit verzweigten, alveolären Endstücken (Stapeldrüse) beschrieben wird (Liebich *et al.*, 2004). Beim Rind besteht sie aus vier Milchdrüseneinheiten (Euterviertel) mit getrennten Hohlraumsystemen und Drüsenbezirken, die in jeweils einer einzelnen Zitze münden. Einen Überblick über die Eutermorphologie gibt Abbildung 1.



**Abbildung 1:** Schematischer Sagittalschnitt durch ein voll ausgebildetes Milchvieheuter mit vorderem und hinterem linken Euterviertel sowie Detaildarstellung des terminalen Bereiches einer Zitze.

Abbildung modifiziert nach einer Darstellung in Frandson (1986)<sup>1</sup>.

Die Invasion von Pathogenen, die zur Mastitis führt, erfolgt in den meisten Fällen über die Zitzenöffnung (galaktogen). Die Pathogene passieren dann zunächst den 1-4 mm im Durchmesser betragenden Strichkanal, dessen verhorntes Plattenepithel Keratin bildet. Die physischen und bakteriostatischen Eigenschaften des Keratin stellen eine wichtige Barriere für Pathogene dar (Capuco *et al.*, 1992). Während der Trockenstehphase bildet das Keratin einen undurchlässigen Pfropf.

Am Übergang des Strichkanals in das Lumen der Zitzenzisterne liegt, charakterisiert durch eine sternförmige Ansammlung von Längsfalten der Schleimhaut, die Fürstenbergsche Rosette. Diese bietet durch ihre vergrößerte Oberfläche einerseits einen Angriffspunkt für eindringende Pathogene, hat aber vermutlich als lymphoide Struktur auch eine protektive Funktion (Hibbitt *et al.*, 1996). So lassen sich in diesem Bereich Aggregationen von Zellen des Immunsystems, insbesondere Lymphocyten, feststellen; näheres dazu in Kap. I.1.5.2.

Die Bedeutung der Fürstenbergschen Rosette für die Immunabwehr im infizierten Euter wurde jüngst durch eine Transkriptomstudie unterstrichen (Rinaldi *et al.*, 2010).

In einer relativ aktuellen Arbeit (Düvel, 2010) wurde gezeigt, dass der Aufbau der Fürstenbergschen Rosette altersabhängigen Veränderungen unterliegt: Während in Färsen ein dichtes, geschlossenes Epithel und im subepithelialen Bereich viele Immunzellen und relativ wenige Kollagenfasern gefunden wurden, war die Gewebestruktur in uni- und multiparen Kühen deutlich aufgelockert mit höherer Kollagendichte in multiparen Kühen. Unklar ist,

<sup>1</sup>Bildquelle: [http://agriculture.kzntl.gov.za/publications/production\\_guidelines/dairying\\_in\\_natal/dairy6\\_1.htm](http://agriculture.kzntl.gov.za/publications/production_guidelines/dairying_in_natal/dairy6_1.htm)

ob und wie dies in Zusammenhang mit der höheren Inzidenz von Euterinfektionen in älteren Kühen in Verbindung steht (Oltenacu & Ekesbo, 1994).

Die Drüsenzisternen sind mit einem zweischichtigen Epithel ausgekleidet, das basal aus einer Schicht flacher Zellen und apikal aus einer Schicht hochprismatischer Zellen besteht.

Jede Zisterne wird aus mehreren Milchgängen gespeist, die sich wiederum in ein Geflecht von Lobuli verästeln. In diesen findet die Milchsynthese durch spezialisierte exokrine Epithelzellen statt, die wiederum von Myoepithelzellen (Korbzellen) umgeben sind. Zusammen bilden sie die Alveoli, von denen jeder Lobulus mehrere Hundert enthalten kann. Aus den Alveoli erfolgt der Milchausstoß in die Milchgänge durch Kontraktion der Korbzellen, die durch das Hormon Oxytocin induziert wird. Dieses als Euterparenchym bezeichnete funktionale Gewebe ist in das Stroma eingebettet, welches aus fibroblastischem Bindegewebe und Fettzellen (Adipocyten) besteht und neben seiner mechanischen Stützungsfunktion auch als Quelle von Lipidreserven, extrazellulärer Matrix und Wachstumsregulatoren dient (Hovey *et al.*, 1999).

Zwischen Wiederkäuern, Nagern und Primaten bestehen substantielle Unterschiede in der Morphologie der Milchdrüsen, so dass eine Übertragung diesbezüglicher Erkenntnisse, die bei Maus und Mensch gewonnen wurden, auf das Rind (und Schaf) nicht ohne weiteres möglich ist (Hovey *et al.*, 1999; Connor *et al.*, 2007).

## 1.4 Zucht auf Eutergesundheit

Die der Milchviehzucht zugrundeliegenden Zuchtziele waren lange einseitig auf die Erhöhung der Milchleistung ausgerichtet. Obwohl dies zu einer dramatischen Erhöhung der pro Kuh und Laktation zu erzielenden Milchmenge geführt hat (Oltenacu & Algers, 2005), ging es mit einer Verschlechterung funktionaler Parameter wie Fruchtbarkeit und Eutergesundheit einher (Ingvarsen *et al.*, 2003). Die genetische Korrelation zwischen der Milchleistung einer Laktation (305 Tage) und dem Auftreten von Mastitis liegt bei 0.37 im Kanadischen Holstein (Uribe *et al.*, 1995) und 0.46 im Finnischen Ayrshire (Poso & Mantysaari, 1996). Eine biologische Interpretation geben Rauw *et al.* (1998) anhand der *Ressource Allocation Theory* (Goddard & Beilharz, 1977; Beilharz *et al.*, 1993; Rauw, 2008), derzufolge die gezielte Zucht auf einzelne Merkmale (z.B. Milchproduktion), das Optimum der Anpassung<sup>2</sup> für die Zuchtobjekte in Richtung des Zuchtzieles verschiebt, wodurch größere Anteile der begrenzten Ressourcen an das Zuchtziel zugewiesen (alloziert) werden müssen, die nicht mehr als Puffer für die Abwehr von Stressfaktoren dienen können.

Zunehmend wurden funktionale Merkmale wie Fruchtbarkeit, Langlebigkeit und Gesundheit als Zuchtziele in die Zuchtprogramme aufgenommen (Kadarmideen & Simm, 2002; Phi-

---

<sup>2</sup>engl: *fitness*

lipsson & Lindhe, 2003), mit besonderem Fokus auf der Resistenz gegen Mastitis als einem der wichtigsten Gesundheitsparameter (s.o.). In Norwegen war dieser Ansatz bei der Zurückdrängung der klinischen Mastitis erfolgreich (Osteras *et al.*, 2007). Der ökonomische Ertrag züchterischer Selektion beim Rind kann durch Berücksichtigung von Gesundheitsparametern in der Zuchtwertschätzung erheblich gesteigert werden (Oltenacu & Broom, 2010).

Der Gehalt somatischer Zellen in der Milch (SCC) wurde bereits als phänotypisches Diagnosekriterium der Mastitis genannt; berechnet wird der SCC gemäß Formel 1.

Den Hauptanteil der somatischen Zellen der Milch stellen im physiologischen Normalzustand Leukocyten (weiße Blutkörperchen), unter denen wiederum die Lymphocyten dominieren (Rivas *et al.*, 2001); abgestoßene Epithelzellen bilden den Rest. Die Zusammensetzung des Leukocytenanteils ändert sich mit einer Euterinfektion; dann steigt der Anteil von Neutrophilen Granulocyten drastisch an (Kelly *et al.*, 2000; Lindmark-Mansson *et al.*, 2006).

Durch logarithmische Transformation wird aus dem SCC der *Somatic Cell Score* (SCS, Formel 2) berechnet. Dieses Merkmal hat eine relativ geringe Heritabilität,  $h^2$  bis 0,15 (Emanuelson *et al.*, 1988) bzw. 0,14 (Carlen *et al.*, 2004), die gleichwohl wesentlich höher als die der klinischen Mastitis (maximal 0,03 nach Carlen *et al.*, 2004) ist, und eine moderate bis hohe genetische Korrelation mit letzterer von 0,6 - 0,8 (Emanuelson *et al.*, 1988) bzw. 0,7 (Carlen *et al.*, 2004) aufweist.

Formeln zur Berechnung der Parameter SCC und SCS für den Gehalt somatischer Zellen in der Milch nach Shook (1993):

$$\text{Somatic Cell Count (SCC)} = \frac{\text{Somatische Zellen}}{\text{ml Milch}} \quad (1)$$

$$\text{Somatic Cell Score (SCS)} = \log_2\left(\frac{\text{SCC}}{100000}\right) + 3 \quad (2)$$

Der SCS wird daher in Zuchtprogrammen, die die Eutergesundheit berücksichtigen, als Surrogatmerkmal für das Auftreten klinischer Mastitis eingesetzt (Philipsson *et al.*, 1995; Kadarmideen & Pryce, 2001; Mark *et al.*, 2002; Carlen *et al.*, 2004). Zuchtziel ist dabei ein verminderter SCS.

Dass Selektion auf einen niedrigen SCS durch die resultierende verbesserte Eutergesundheit positive Auswirkungen auf die Dauer des produktiven Lebens von Milchvieh und damit die Profitabilität hat, ist mehrfach gezeigt worden (Sewalem *et al.*, 2006; Miller *et al.*, 2009). Verschiedene Beobachtungen deuten jedoch darauf hin, dass eine zu erfolgreiche Selektion auf niedrigen SCS den gegenteiligen Effekt einer erhöhten Rate von Euterinfektionen durch die

Kompromittierung des Immunsystems hervorbringt (Suriyasathaporn *et al.*, 2000; Beaudreau *et al.*, 2002; Oltenacu & Broom, 2010); dies ist jedoch umstritten.

Bis vor einigen Jahren war es bei der Zuchtwertschätzung üblich, Werte für den SCS, die zu verschiedenen Zeiten während einer Laktationsperiode ermittelt wurden, als wiederholte Messungen desselben Merkmales mit einer genetischen Korrelation von 1 zu betrachten (Reents *et al.*, 1995). Untersuchungen deuten darauf hin, dass die Heritabilität des SCS während einer Laktation Schwankungen unterworfen ist (z.B. Haile-Mariam *et al.*, 2001), so dass einzelne Testtagesmessungen des SCS über eine Laktation hinweg in neueren *Random Regression* Modellen als distinkte korrelierte Merkmale eingesetzt werden (u.a. Mrode & Swanson, 2003).

Die paarweisen genetischen Korrelationen für den SCS in den ersten drei Laktationsperioden werden mit  $\approx 0,8 - 0,95$  angegeben (Haile-Mariam *et al.*, 2001; Carlen *et al.*, 2004).

Relative Zuchtwerte (RZ), etwa der Relative Zuchtwert Somatische Zellzahl (RZS), werden ermittelt, indem die Verteilung des jeweiligen Zuchtwertes, hier also der für SCS, innerhalb einer Zuchtbullenkohorte in eine Normalverteilung mit Mittelwert  $\mu = 100$  und Standardabweichung  $\sigma = 12$  transformiert wird. Die Skala wird dabei umgekehrt, d.h. anders als beim SCS selbst steht beim RZS ein höherer Wert für einen favorisierten Zuchtwert.

In der aktuellen Formulierung der Zuchtwertschätzung beim Deutschen Holstein wird der RZS mit 7% im Gesamtzuchtwert berücksichtigt. Die Eutergesundheit geht jedoch als Hilfsmerkmal ebenfalls in die funktionale Nutzungsdauer, die mit 20% gewichtet wird, ein.

Die Erfassung der Phänotypen von Milchrindern, also z.B. Gesundheits- und Leistungsmerkmalen, erfolgt in Deutschland zentral durch die Vereinigten Informationssysteme Tierzucht (VIT) in Verden. Ebenfalls durch diese Institution werden aus den aufgenommenen Merkmalen von Bullentöchtern die entsprechenden Zuchtwerte der Bullen berechnet, beispielsweise in Form der mittleren Töchterabweichung (*Daughter Yield Deviation*, DYD). So wird etwa die von einem Bullen vererbte Zahl somatischer Zellen der Milch (DYD-SCS) durch ein *Random-Regression*-Testtagesmodell (eine Verfeinerung gegenüber dem vorher angewandten *Fixed-Regression*-Modell), das die Töchterbeobachtungen um maternale und Umwelteffekte korrigiert, berechnet (Liu *et al.*, 2004).

## 1.5 Immunologische Mechanismen der Eutergesundheit

### 1.5.1 Immunologische Besonderheiten in Wiederkäuern

Das Immunsystem von Wiederkäuern (*Ruminantia*) weist gegenüber dem von Primaten oder Nagern einige Besonderheiten, auch mit direktem Bezug zu den Milchdrüsen, auf. So wird etwa das immunologische Repertoire der Milchdrüse stärker als in anderen Spezies, wo es eher dem von mukösen Oberflächen ähnelt, durch das periphere Immunsystem geprägt (Sheldrake *et al.*, 1988; Harp *et al.*, 1988, 2004). Andere Eigenheiten betreffen den Anteil regulatorischer  $CD4^+CD8^- \gamma\delta$ -T-Lymphocyten (Hein & Mackay, 1991; Vaccarelli *et al.*, 2008), die Anfälligkeit für Leukotoxin (LKT) aufgrund von Mutationen ihres  $\beta 2$ -Integrings (Zecchinon *et al.*, 2005) oder die passive Immunisierung der Nachkommen ausschließlich über das Kolostrum (Jaster, 2005).

Für das Studium der immunologischen Verhältnisse sind demnach andere Spezies der Wiederkäuer (*Ruminantia*) besonders geeignet, beispielsweise das Schaf (*Ovis aries*) (Mavrogiani *et al.*, 2005).

### 1.5.2 Die Bedeutung von Mechanismen der angeborenen, nichtadaptiven Immunität für die Eutergesundheit

Von großer Bedeutung für die Pathogenabwehr im Euter sind die Zellen der angeborenen Immunität. Diese besitzen keine hochvariablen Antigenrezeptoren wie die B- und T-Lymphocyten. Vielmehr verfügen sie über keimbahncodierte Rezeptoren für ein konserviertes Spektrum an molekularen Mustern auf der Zelloberfläche von Pathogenen (*pathogen associated molecular pattern*, PAMP). Beispiele sind die Mitglieder der *Toll-like-Rezeptor* (TLR)-Familie.

Unter den zellulären Komponenten der angeborenen Immunabwehr sind v.a. die neutrophilen Granulozyten und die Makrophagen zu nennen. Erstere sind zwar im gesunden Euter nur in relativ geringer Zahl vertreten, können aber im Verlauf einer Euterinfektion bis zu 90% der Gesamtleukozytenpopulation des Euters ausmachen (Paape *et al.*, 2003). Sie üben dann multiple Funktionen aus, z.B. phagozytieren sie Pathogene, produzieren bakterizide Peptide (z.B. Defensine) und wirken inflammatorisch durch Sekretion cytotoxischer Substanzen, z.B. ROS (*reactive oxygen species*). Da diese immunpathologisch auch das Eutergewebe schädigen, ist die rechtzeitige Eliminierung exzessiver neutrophiler Granulozyten durch Apoptose ein wichtiger Schritt bei der Rückbildung einer Entzündungsreaktion (Buret, 2010).

Makrophagen wirken ebenfalls phagozytisch, wobei die diejenigen des Euters weniger effizient zu sein scheinen als Neutrophile (Sordillo, 2005). Die primäre Funktion von Makrophagen im Falle einer Pathogeninvasion des Euters ist die Sekretion von Cytokinen (z.B. Interferonen) und Chemokinen (z.B. IL-8), Eikosanoiden (z.B. Prostaglandine und Leukotriene) und

Akute-Phase-Proteinen (Bruckmaier, 2005). Ihre Rolle bei der direkten Pathogenabwehr wird jedoch infrage gestellt (Rainard & Riollot, 2006). Daneben können sie als antigen-präsentierende Zellen (APC) an der Induktion einer adaptiven Immunantwort mitwirken.

Natürliche Killerzellen (NK-Zellen) können potentiell auf mehrfache Weise in die Abwehr von Euterinfektionen involviert sein, wobei zu diesem Thema noch kaum Untersuchungen vorliegen. Eine aktuelle Übersicht zu NK-Zellen des Rindes allgemein bieten Boysen & Storset (2009). NK-Zellen besitzen Rezeptoren, z.B. der Familien der *killer immunoglobulin-like receptors* (KIR<sup>3</sup>) und Ly49, die aberrante Expressionsmuster klassischer MHC-I-Moleküle, wie sie zum Beispiel als Folge einer Virusinfektion auftreten, erkennen. Die Liganden anderer NK-Rezeptoren (etwa NKG2D) sind nichtklassische MHC-I-Moleküle, die stressinduziert exprimiert werden. Schließlich verfügen NK-Zellen auch über Rezeptoren für Immunglobuline (z.B. Fc $\gamma$ RIII (CD16)), die effektiv opsonisierte Pathogene erkennen. Die Bindung NK-Rezeptor-Ligand hat dabei gewöhnlich die Zerstörung der Zielzelle (Perforin- oder Fas-vermittelt) zur Folge (Lunemann *et al.*, 2009). Darüber hinaus wurde für NK-Zellen von Mensch (Krensky, 2000) und Rind (Sordillo *et al.*, 2005) ebenfalls die Expression bakterizider Peptide (Granulysin) nachgewiesen.

Eine wichtige Rolle bei der Detektion und frühen Abwehr von Euterpathogenen scheinen auch die Milchdrüsenepithelzellen (*mammary epithelial cells*, MEC), für die eine Hochregulierung einer Vielzahl immunrelevanter Gene, darunter Cytokine, Chemokine, Pathogenrezeptoren (u.a. TLRs) und bakterizider Peptide ( $\beta$ -Defensine) nach PAMP-Stimulation gezeigt wurde (z.B. Pareek *et al.*, 2005; Günther *et al.*, 2009), wahrzunehmen.

Vaskuläre Endothelzellen des bovinen Euters sekretieren *in vitro* nach Stimulation mit *E.coli*-Endotoxin PAF (*platelet activating factor*), IL-1 $\beta$ , IL8 und ICAM1 (Corl *et al.*, 2008).

### 1.5.3 Die Bedeutung von Mechanismen der adaptiven Immunität für die Eutergesundheit

#### B-Lymphocyten und Immunglobuline

Die Dynamik von B-Lymphocyten und Immunglobulinen im Euter ist laktationsabhängig. Plasmazellen, also terminal differenzierte B-Lymphocyten, die Immunglobuline sezernieren, wurden im laktierenden Euter in der terminalen Region des Strichkanals (Nickerson *et al.*, 1984) und am Übergang zwischen Zitenzisterne und Strichkanal, d.h. im Bereich der Fürstenbergschen Rosette (Abb. 1) gefunden (Collins *et al.*, 1986).

Nach experimenteller Inokulation des Strichkanals von Schafen mit dem Leukotoxinsekretierenden Bakterium *Mannheimia haemolytica* konnten Mavrogianni *et al.* (2005) im

---

<sup>3</sup>Der Locus dieser Genfamilie liegt im distalen Bereich von BTA18.

der Fürstenbergschen Rosette entsprechenden Bereich die Entstehung von lymphoiden Follikeln beobachten. Es wurde gezeigt, dass diese Follikel mit nicht terminal ausdifferenzierten B-Lymphocyten, Makrophagen und  $\gamma\delta$ -T-Lymphocyten angereichert sind (Fragkou *et al.*, 2010); dies deutet auf eine lokale adaptive humorale Immunantwort in der Zitze und zusammen mit einer Anreicherung von B-Lymphocyten und CD4<sup>+</sup>-T-Helferzellen in den Lymphknoten infizierter Euter auf eine protektive Rolle adaptiver Immunmechanismen gegen Euterpathogene (ebd.). In einer Feldstudie wurden derartige lymphoide Follikel ebenfalls in den Zitzen klinisch gesunder Schafe gefunden (Mavrogianni *et al.*, 2007).

Die Induktion lymphoider Follikel als lokale adaptive Immunantwort mit teils pathologischem, teils protektivem Charakter wird in der Literatur für solche Gewebe beschrieben, die, ähnlich wie die Zitzenöffnung, wiederholt mit Pathogenen oder Antigenen in Kontakt kommen, d.h. einem erhöhten/wiederholten Pathogen- und Antigendruck ausgesetzt sind (Carragher *et al.*, 2008).

Protektive Bedeutung kommt den lymphoiden Follikeln besonders zu, wenn die Kapazitäten der angeborenen Immunabwehr eingeschränkt sind (Fragkou *et al.*, 2010), beispielsweise durch die Aktivitätsminderung von Neutrophilen. Letzteres ist ein wesentliches Phänomen der periparen Immundepression, d.h. der Phase einer verminderten Immunkompetenz der Milchkuh während der Zeit der Abkalbung (Mallard *et al.*, 1998; Morris *et al.*, 2009; Piepers *et al.*, 2009).

Über den Anteil der B-Lymphocyten an der Lymphocytenpopulation der Milch gibt es in der Literatur uneinheitliche Angaben. In Abhängigkeit vom Zeitpunkt während der Laktation, den verwendeten B-Zell-Markern und anderen Faktoren schwanken die Angaben zwischen  $\approx$  3-8% (Duhamel *et al.*, 1987; Harp *et al.*, 2004), und 20-25% (Concha *et al.*, 1978; Park *et al.*, 1992).

Einigkeit herrscht jedoch darüber, dass in der Milch B-Lymphocyten a) den weit geringeren Teil der gesamten Lymphocytenpopulation ausmachen und b) dieser Anteil weit niedriger ist als zur gleichen Zeit im peripheren Blut. Harp *et al.* (2004) fanden Hinweise für eine verstärkte Migration von B-Lymphocyten aus der Peripherie ins Euter in der periparen Periode.

Mit fortschreitendem Alter der Milchkuh sinkt die Zahl der im peripheren Blut zirkulierenden B-Lymphocyten (Ohtsuka *et al.*, 2010). Ob dies mit der durch Oltenacu & Ekesbo (1994) konstatierten erhöhten Mastitisinzidenz in älteren Milchkuhen in Verbindung steht, bleibt zu untersuchen.

Immunglobuline, die sezernierte Form des B-Zell-Rezeptors, sind für den Schutz des Euters vor Infektionen von großer Bedeutung (Sordillo *et al.*, 2007).

Eine Subpopulation von B-Lymphocyten (hauptsächlich CD5<sup>+</sup>) produziert natürliche Antikörper (*natural antibodies*, NAb), bei denen es sich um keimbahncodierte, möglicherweise



polyspezifische Immunglobuline handelt und die Elemente der angeborenen und der adaptiven Immunität in sich vereinen. Van Kneysel *et al.* (2007) beschreiben das Vorkommen von NAb in der Milch laktierender Kühe während der zweiten bis neunten Woche der Laktation; die Autoren postulieren eine Kausalität zwischen der Verminderung von NAb im Zuge einer negativen Energiebilanz nach der Abkalbung und einer reduzierten Immunkompetenz, die mit einer erhöhten Mastitisanfälligkeit einhergeht.

Die Serumantikörper-Reaktivität gegen Ovalbumin als Maß für die Immunkapazität nutzten Wagter *et al.* (2000) für die Klassifizierung weiblicher Holstein-Rinder und stellten eine Korrelation dieses Parameters mit dem Auftreten einer periparen (d.h. einer während der Abkalbperiode, am Beginn der Laktation) auftretenden Mastitis fest. Es wird vermutet, dass ein niedriger Serumantikörperspiegel durch den verstärkten Transport von Immunglobulinen in die Milchdrüsen bedingt wird. Eine inverse Korrelation zwischen natürlichen Antikörpern in Milch und Serum finden z.B. van Kneysel *et al.* (2007). Dagegen ist nach Wagter *et al.* (2000) eine allgemein höhere Antikörperkapazität mit einem erhöhten Transfer von Immunglobulinen in die Milchdrüsen positiv korreliert. Die Serumantikörperkapazität gegen Ovalbumin ist partiell erblich ( $h^2 = 0,32$  (peripar) bis  $> 0,5$  (Wagter *et al.*, 2000) bzw.  $0,48$  (Burton *et al.*, 1989)). Leitner *et al.* (2008) fanden eine negative Korrelation zwischen dem Titer von maternalem IgG im Kolostrum und der Inzidenz bakterieller Infektionen bei Kühen in der frühen Laktation. Die Zahl peripherer B-Lymphocyten ist in multiparen (d.h. in bereits mehrfach trächtigen) Kühen im Vergleich zu erst- oder zweitkalbenden Kühen verringert. Ohtsuka *et al.* (2009) sehen darin eine mögliche Erklärung für die erhöhte Inzidenz von Mastitis in älteren Milchkühen, die u.a. von Grohn *et al.* (1998) beschrieben wurde.

## T-Lymphocyten

Die Zusammensetzung der T-Lymphocytenpopulation im Euter unterliegt ebenfalls laktationsabhängigen, darüber hinaus auch altersabhängigen Schwankungen. In der Trockenstehphase dominieren  $CD4^+$ - gegenüber  $CD8^+$ - T-Lymphocyten im Verhältnis  $4,4 \pm 2,2$ . Während der Laktation liegt dieses Verhältnis bei  $0,47 \pm 0,36$ , d.h., hier überwiegen die  $CD8^+$ -T-Lymphocyten (Asai *et al.*, 2000). Während der ersten drei Wochen der Laktation liegt der Anteil von  $CD8^+$ -T-Lymphocyten an der Gesamtlymphocytenpopulation in der Milch signifikant über dem im Blut (Harp *et al.*, 2004).

Der Anteil der  $CD8^+$  T-Lymphocyten an der Gesamtpopulation der  $\alpha\beta$ -T-Lymphocyten, die im Kolostrum gefunden wird, sinkt mit fortschreitendem Alter; selbiges gilt für die Zahl peripherer zirkulierender  $CD4^+$ -T-Lymphocyten (Ohtsuka *et al.*, 2010).

$CD8^+$ -T-Lymphocyten erkennen Komplexe aus MHC-I-Molekülen und endogenen Peptidantigenen, die auf der Oberfläche ihrer Zielzellen exprimiert werden; nach ihrer Aktivierung

entfalten sie cytotoxische Aktivität. Ihre Anreicherung am Euterepithel lässt eine Rolle dieser Zellen bei der Beseitigung infizierter oder beschädigter Epithelzellen und somit bei der Erhaltung der Integrität des Euterepithels vermuten (Taylor *et al.*, 1994; Asai *et al.*, 2000).

CD4<sup>+</sup>-T-Lymphocyten, die auch als T-Helferzellen bezeichnet werden, werden dagegen durch antigenpräsentierende Zellen (*antigen presenting cells*, APC), zu denen Makrophagen, B-Lymphocyten und dendritische Zellen gezählt werden und die exogene Peptidantigene auf zelloberflächengebundenen MHC-II-Komplexen präsentieren, aktiviert. Aktivierte T-Helferzellen wiederum stimulieren unter anderem die Proliferation antigenspezifischer B-Lymphocyten, u.a. durch Sekretion von IL-4.

## 2 Ansätze der molekularen Genomanalyse

### 2.1 Experimentelle Ansätze der strukturellen Genomanalyse

#### 2.1.1 Kartierung merkmalsassoziierter Genomregionen

Eine Methode, Kandidatengene für ein polygenes, quantitatives Merkmal wie Milchleistung oder Eutergesundheit zu finden, besteht darin, zunächst genomische Regionen, die dieses Merkmal beeinflussen (*quantitative trait locus/loci*, QTL) einzugrenzen. Dies kann beispielsweise mittels einer Kopplungsanalyse (*linkage analysis*, LA) erfolgen. Beim Nutztier (d.h. für *Outbred*-Populationen) wurden verschiedene Versuchsdesigns genutzt; beim Hausrind, das von seiner Populationsstruktur her besonders dafür geeignet ist, hat sich das von Weller *et al.* (1990) vorgeschlagene *granddaughter design*, GDD, etabliert. Hierbei werden zunächst Bullen selektiert, die für die gewählten genetischen Marker segregieren. Deren Söhne (d.h. Halbgeschwister) werden an diesen Markern genotypisiert, während die Aufnahme der phänotypischen Merkmale an den Töchtern der Söhne erfolgt (zur Berechnung von Bullenzuchtwerten aus phänotypischen Beobachtungen der Bullentöchter siehe Kap. I.1.4).

Die kombinierte Informativität benachbarter Marker wird für die Kartierung genutzt, indem z.B. bei der Methode der *multimarker regression* die Übertragungswahrscheinlichkeit (*transmission probabilities*) von Haplotypen solcher Marker in einer Varianzanalyse (ANOVA) als unabhängige Variable zur Erklärung der Phänotypunterschiede eingesetzt wird (Knott *et al.*, 1996).

Erste gesamtgenomische QTL-Kartierungen beim Milchrind mit der GDD-Methode stellten Georges *et al.* (1995) für Milchproduktionsmerkmale und Zhang *et al.* (1998) für Milchleistungs- und Eutergesundheitsmerkmale vor. Seither wurden für den Merkmalskomplex Eutergesundheit auf nahezu allen Autosomen des Rindes (*Bos Taurus autosome*, BTA) QTL identifiziert. Einen Überblick über beschriebene QTL für das Eutergesundheitsmerkmal

SCS, gibt Tabelle 32 (im Anhang). Wenn verschiedene Studien QTL in derselben genomischen Region beschreiben, stellt sich das Problem, ob es sich dabei nur um einen QTL mit möglicherweise pleiotropen Effekten oder um separate QTL handelt. Metaanalytische Verfahren können hier Klarheit verschaffen (Hayes & Goddard, 2001; Khatkar *et al.*, 2004). Mittlerweile stehen auch Softwarelösungen zu diesem Zweck zur Verfügung (Arcade *et al.*, 2004; Veyrieras *et al.*, 2007).

Das Auflösungsvermögen von QTL-Kartierungen mittels Kopplungsanalyse ist niedrig; es liegt im Bereich von bis zu mehreren Dutzend Centimorgan (cM) mit Hunderten bis Tausenden positionellen Kandidatengenomen. Durch die Verwendung dichter Markerkarten kann es nicht über einen bestimmten Grad hinaus erhöht werden, da die Kopplungsanalyse nur Rekombinationsereignisse berücksichtigt, die in der typisierten Population (die zwei bis drei Generationen umfasst) selbst auftreten und die zwischen eng benachbarten Markern innerhalb dieser Zeitspanne sehr selten sind.

Eine Aufhebung dieser Beschränkung kann durch Kartierungsmethoden, die Kopplungsungleichgewichte (*linkage disequilibrium*, LD) zur Etablierung von Haplotypstrukturen zwischen genetischen Markern im Bereich des QTL nutzen (LD-Mapping), erreicht werden, da diese alle Rekombinationsereignisse seit der QTL-kausalen Mutation berücksichtigen (Darvasi *et al.*, 1993).

Derartige Methoden wurden zunächst für diskrete Merkmale (z.B. kranke vs. gesunde Individuen) entwickelt und angewendet (Hastbacka *et al.*, 1992; Houwen *et al.*, 1994; Terwilliger, 1995). Es wird dabei versucht, innerhalb von Familienstrukturen einzelne Haplotypen zu identifizieren, die auf einen Haplotyp eines gemeinsamen Vorfahren zurückgehen (d.h., die *identical by descent*, IBD, sind) und in einer Merkmalsklasse (z.B. erkrankten Individuen) exklusiv oder gehäuft auftreten. Je stärker eine solche Beziehung hervortritt, desto enger sind die den Haplotyp konstituierenden Marker mit der kausalen Mutation gekoppelt und segregieren zusammen mit dieser und desto enger ist folglich ihre physische Nähe.

Durch Riquet *et al.* (1999) wurde dieser Ansatz in einem zweistufigen experimentellen Protokoll erfolgreich für ein quantitatives Merkmal, die Feinkartierung eines QTL für Milchproduktion auf BTA14, adaptiert.

Ein formelles statistisches Verfahren, das Kopplungsanalyse und LD-Analyse für die QTL-Feinkartierung kombiniert, wurde durch T. Meuwissen und Kollegen eingeführt (Meuwissen & Goddard, 2000; Meuwissen *et al.*, 2002). Die dafür ideale Haplotypstruktur besteht aus Haplotypen, die jeweils 4-6 Marker beinhalten (Grapes *et al.*, 2006). Bei diesem iterativen Ansatz wird pro Durchlauf in ein Intervall zwischen zwei Markern ein putativer QTL platziert. Unter dieser Annahme tragen Haplotypen der jeweils flankierenden Marker, die *identical by*

*descent* (IBD) sind, wahrscheinlich identische QTL-Allele, d.h. die erwarteten Korrelationen zwischen den Effekten identischer Haplotypen sind hoch. Ob die sich daraus ergebende Kovarianzstruktur der Haplotypeneffekte die reale Kovarianzstruktur der Phänotypdaten abbildet, d.h. ob der reale QTL sich tatsächlich im oder nahe bei dem jeweiligen Markerintervall befindet, kann durch einen *likelihood ratio*-Test (LRT) eines gemischten linearen Modells, das diese Kovarianzstruktur in einem Faktor reflektiert, gegen ein entsprechendes Nullmodell statistisch formal getestet werden.

Durch derartige LA/LD-Feinkartierung etablierte Assoziationen zwischen Markerhaplotypen und QTL-Allelen besitzen populationsübergreifende Gültigkeit und müssen, anders als solche durch einfache LA-Kartierung, nicht neu in jeder Familie validiert werden (Khatkar *et al.*, 2004).

Die LA/LD-Methode wurde sowohl durch Simulationsstudien plausibilisiert und validiert (Abdallah *et al.*, 2003; Uleberg & Meuwissen, 2007) als auch real für die Feinkartierung z.B. von QTL für Milchleistung auf BTA6 (Olsen *et al.*, 2004) und BTA14 (Farnir *et al.*, 2002) oder Mastitis auf BTA11 (Schulman *et al.*, 2009) eingesetzt.

Als Alternativen für die QTL-Feinkartierung sei hier der vom *transmission disequilibrium test*, TDT (Spielman *et al.*, 1993) abgeleitete Quantitative TDT, QTDT, (Abecasis *et al.*, 2000), der auf der Analyse von Nuklearfamilien (d.h. Eltern plus je 1 Nachkomme) beruht<sup>4</sup> erwähnt.

Die Entwicklung von Hochdurchsatztechnologie für die Genotypisierung von SNPs eröffnet neue Möglichkeiten für die Kartierung merkmalsbeeinflussender Regionen. Auf dem zwölften QTL-MAS-Workshop 2008 in Uppsala, Schweden, wurde ein simulierter Datensatz<sup>5</sup> vorgestellt, an dem verschiedene Ansätze der QTL-Detektierung und Feinkartierung getestet werden sollten. Durch Crooks *et al.* (2009) wurden sechs derartige Studien (Bink & van Eeuwijk, 2009; Cleveland & Deeb, 2009; Heuven *et al.*, 2009; Tarres *et al.*, 2009; Lam *et al.*, 2009; Ledur *et al.*, 2009) ausgewertet; die Ergebnisse sind generell ermutigend, gleichwohl konnten spürbare Unterschiede in statistischer Teststärke und Genauigkeit der QTL-Kartierung beobachtet werden, so dass hier durchaus Verbesserungspotential gesehen wird (Crooks *et al.*, 2009). So konnte, abhängig von der gewählten Methode, die wahre Position des simulierten QTL bis auf 1-2 cM genau bestimmt werden; folglich ist auch mit diesen Methoden die Erwartung nicht plausibel, einen QTL bis auf das Niveau des einzelnen Gens auflösen zu können.

Eine weitere Studie am selben Datensatz mithilfe des QTDT stammt von Simianer & Pimentel (2009).

---

<sup>4</sup>Ein Review zu familienbasierten Kartierungsmethoden bieten Ewens *et al.* (2008)

<sup>5</sup>6 Chromosomen, 6000 equidistante SNPs, 50 bialele QTL

Eine praktische Anwendung, in der  $\approx 4850$  SNPs des *Affymetrix MegAllele GeneChip Bovine Mapping 10K SNP array* zur Kartierung von QTL für bovine Produktionsmerkmale in einem gemischten Design (GDD + selektive Genotypisierung) nutzte, fand moderate Übereinstimmungen der QTL mit bereits in der Literatur bekannten (Daetwyler *et al.*, 2008).

Mit einer neuen technischen Generation von SNP-Arrays (*Illumina Bovine SNP50 Bead-Chip*, Matukumalli *et al.*, 2009) können  $\approx 50.000$  bovine SNPs simultan analysiert werden. Eine genomweite Assoziationsstudie (*genome wide association analysis*, GWAS) mit Fokus auf einer Reihe boviner Leistungsmerkmale wurde 2009 publiziert (Cole *et al.*, 2009); weitere folgten, z.B. Cole *et al.* (2011).

### 2.1.2 Identifizierung positioneller Kandidatengene (vergleichende Genkartierung)

Auch nach einer Feinkartierung ist für einen gegebenen QTL mit bis zu mehreren hundert positionellen Kandidatengenen für die kausale Mutation zu rechnen. Mittlerweile liegen für das Rind Kopplungskarten (*linkage maps*, Snelling *et al.*, 2005; Arias *et al.*, 2009), physikalische Karten, v.a. auf Grundlage von BAC (*Bacterial Artificial Chromosome*)-Sequenzen (Snelling *et al.*, 2007) und die gesamte Genomsequenz (der großteils die Sequenzierung eines Referenztieres der Fleischrindrasse Hereford, *L1 Dominette* zugrunde liegt) in der aktualisierten Version 4.0 (Elsik *et al.*, 2009) vor. Zum Zeitpunkt vieler QTL-Studien beim Rind waren diese Informationen zum bovinen Genom nicht oder nur unzureichend verfügbar, so dass zur Identifizierung positioneller Kandidatengene auf Synteniebeziehungen zwischen Rinder-QTL und chromosomalen Regionen des Menschen, für den jederzeit eine fortgeschrittenere Genomannotation verfügbar war, zurückgegriffen werden musste (z.B. Goldammer *et al.*, 2004; Connor *et al.*, 2006; Weikard *et al.*, 2006; Ron *et al.*, 2007). Aufgrund der Tatsache, dass die evolutionäre Linie der Nagetiere eine im Vergleich zu Rind und Mensch doppelte Rate an chromosomalen Strangbrüchen aufweist (Murphy *et al.*, 2005), ist die Maus für solche Zwecke weniger geeignet (Hayes *et al.*, 2003).

Gesamtgenomische komparative Genkarten des Rindes, die auf der *Radiation Hybrid-Technologie* (Cox *et al.*, 1990) und der Sequenzierung von EST (*expressed sequence tags*) beruhen, weisen 768 (Band *et al.*, 2000), 1463 (Everts-van der Wind *et al.*, 2004) bzw. 1716 (Itoh *et al.*, 2005) Ankerpunkte im humanen Genom auf; dies entspricht  $\approx 45\%$ ,  $65\%$  bzw.  $72\%$  Abdeckung. Die ersten beiden der genannten Karten bilden das bovine IL-TX (Illinois-Texas) RH<sub>5000</sub>-Panel; durch Einbeziehung von Informationen aus BAC-End-Sequenzierungen konnte es auf einen Abdeckungsgrad von  $91\%$  des humanen Genoms (die theoretische Obergrenze liegt nach Everts-van der Wind *et al.* (2005) bei  $94\%$ ) erweitert werden (Larkin *et al.*, 2003; Everts-van der Wind *et al.*, 2005).

Einen alternativen, FISH (*Fluorescence In Situ Hybridization*)-basierten komparativen Ansatz zwischen Rind, Mensch und Maus mit 91 %iger Abdeckung des humanen Genoms präsentierten Schibler *et al.* (2006).

### 2.1.3 Der BTA18-SCS-QTL

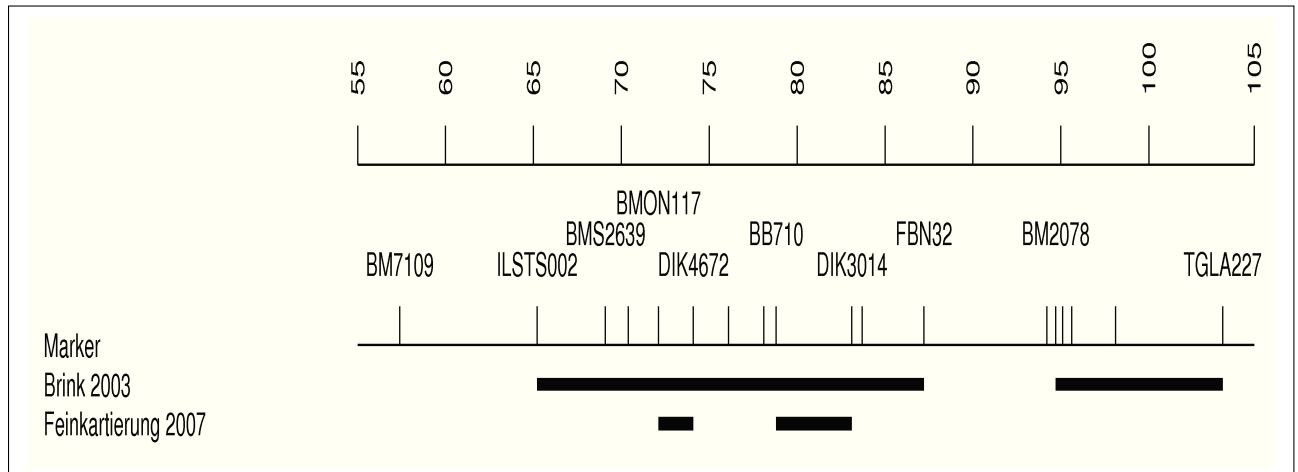
Wie aus Tabelle 32 (im Anhang) hervorgeht, wurden für das mit Mastitis korrelierte Merkmal SCS von verschiedenen Autoren merkmalsbeeinflussende Regionen auf BTA18 identifiziert. Eine Kartierung durch Kopplungsanalyse (LA) in einem GDD, Kap. I.2.1.1) wurde durch Brink (2003) vorgestellt und durch Xu *et al.* (2006) weitgehend bestätigt. Diese Analysen platzierten den QTL im distalen Bereich von BTA18, wobei das 95%-Konfidenzintervall LA-typisch sehr groß ist. In Brink (2003) umfasst es  $\approx 40$  cM von insgesamt 105 cM des BTA18, wobei es in zwei Teilintervalle zerfällt (Abb. 2); das distal gelegene und durch das Markerpaar BM2078 und TGLA227 begrenzte weist einen höheren LOD (*logarithm of odds*)-Score und damit eine erhöhte Wahrscheinlichkeit für die kausale Mutation auf. Eine der Zielstellungen des Verbundprojektes M.A.S.-Net war es, diesen QTL durch Feinkartierung höher aufzulösen, was mittels der kombinierten LA/LD-Methode nach Meuwissen & Goddard (2000) in einer aus einem GDD abgeleiteten Population, über intermediäre Projektstufen in der Studie von Baes *et al.* (2009) resultierte. Im Mai 2007 war eine dieser Projektstufen Gegenstand eines M.A.S.-Net-Symposiums; dort zeigte sich, dass das proximale Intervall von Brink (2003) seinerseits zwei durch die Marker BMON117 und DIK4672 bzw. BB710 und DIK3014 begrenzte Teilintervalle enthält, die im folgenden mit A (bzw. 5) und B (bzw. 9) bezeichnet werden und für die die Annahme einer Kopplung mit dem QTL die beobachtete phänotypische Varianz am besten erklärt (Abb. 2).

Der distale Bereich von BTA18 weist eine Syntenie mit dem langen Arm des humanen Chromosoms 19 (HSA19q<sup>6</sup>) auf, wie schon früh durch Zoo-FISH (Solinas-Toldo *et al.*, 1995) und Hybridisierung in somatischen Zellen (Gao & Womack, 1997) festgestellt wurde. Parallel mit der Etablierung der gesamtgenomischen komparativen Karten, vgl. Kap. I.2.1.2, wurde die BTA18 $\leftrightarrow$ HSA19 Karte durch z.B. RH-Mapping (Goldammer *et al.*, 2002; Mömke *et al.*, 2005) und FISH mit spezifischen BAC-Sequenzen (Brunner *et al.*, 2003) verfeinert und potentielle Strangbruchpunkte identifiziert. Eine durch Everts-van der Wind *et al.* (2004) angenommene Inversion auf dem bovinen Chromosom gegenüber dem humanen konnte dabei durch Mömke *et al.* (2005) nicht bestätigt werden.

HSA19 hat Eigenschaften, die es von den restlichen humanen Chromosomen abheben. Verglichen mit dem Gesamtgenom weist es eine mehr als doppelte Gendichte auf, nämlich  $\approx 1400$  proteincodierende Loci mit  $\approx 2300$  verschiedenen Transkripten, die bei einer physikalischen Länge von  $\approx 55$  Megabasen (Mb) ca.  $26 \frac{\text{Genen}}{\text{Mb}}$  und einer genetischen Länge von ca. 2,1

---

<sup>6</sup>HSA: *Homo sapiens autosome*; HSA19 wird fast völlig von der Cytobande 13 dominiert  $\Rightarrow$  HSA19q13



**Abbildung 2:** Der BTA18-SCS-QTL in seinen Kartierungsstadien. Dargestellt ist der distale Bereich von BTA18 im Bereich von 38-66 Mb. Das oberste Panel gibt eine Skalierung in cM, das darunter liegende eine Auswahl genetischer Marker dieser Region. Im zweiten Panel von unten sind die durch Brink (2003) kartierten Teilintervalle des dortigen 95%-Konfidenzintervalles (insgesamt  $\approx 40$  cM) ausgewiesen. Durch eine LA/LD-Analyse im Rahmen von FUGATO M.A.S.-Net wurden innerhalb des proximalen dieser Teilbereiche zwei separate Intervalle mit maximaler Wahrscheinlichkeit für die kausale Mutation identifiziert. Diese werden durch die Markerpaare BMON117 und DIK4672 bzw. BB710 und DIK3014 begrenzt und sind im unteren Panel dargestellt; sie haben eine Ausdehnung von 2,0 (Intervall A, links) bzw. 4,3 (B, rechts) cM; vgl. Tab. 12).

$\frac{\text{cM}}{\text{Mb}}$  entsprechen (Grimwood *et al.*, 2004). Auch der Anteil von Exons an der Gesamtsequenz ist gegenüber dem genomischen Durchschnitt stark erhöht (6,4% vs. 1,5%, ebd.).

Diese erhöhte Gendichte hängt mit der Präsenz mitgliederreicher Gencluster zusammen. Beispielsweise sind die meisten und größten Cluster von Transkriptionsfaktoren des C2H2-Zinkfinger-Typs auf HSA19 lokalisiert, darunter der Cluster “19.12” mit  $\approx 40$  Genen in der Cytobande 19q13.4 (Tadepally *et al.*, 2008).

Von großer immunologischer Bedeutung sind die v.a. von Natürlichen Killerzellen (NK-Zellen) exprimierten immunglobulin-ähnlichen Rezeptoren der *LILR*-, *LAIR*- und *KIR*-Familien die als Resultate von Genduplikationen in einem großen Cluster innerhalb der Cytobande 19q13.4 organisiert sind (Hsu *et al.*, 2002). Cluster von *KIR*-Genen mit vergleichbaren Funktionen finden sich auch beim Rind auf BTA18q (Storset *et al.*, 2003), nicht aber bei Nagetieren, wo deren Rolle als NK-Rezeptoren durch Lektine der *Ly49*-Familie übernommen wird (Boysen & Storset (2009), ein Fall von konvergenter Evolution und Argument für die Ausbeutung der Mensch $\leftrightarrow$ Rind-Syntenie).

Zu nennen sind weiterhin Cluster von Apolipoproteinen (Klos *et al.*, 2008), der *Cyp2*-Familie von P450-Cytochromen (Hoffman *et al.*, 2001) und  $\gamma$ -Untereinheiten von  $\text{Ca}^{2+}$ -Kanälen (*CACNG*-Gene, Burgess *et al.*, 2001).

Serinproteasen der Kallikrein-Familie (KLK) auf 19q13.4 werden in Zusammenhang mit der Entstehung von (hormonabhängigen) Tumoren, u.a. von Prostata (KLK3 ist auch als

Tumormarker PSA, *prostate specific antigen*, bekannt) und Eierstöcken (Clements *et al.*, 2004; Bayani *et al.*, 2011) gebracht.

Ein relativ junges Forschungsgebiet sind die microRNAs (miRNAs), d.h. kurze, einzelsträngige, nichttranslatierte RNAs, die durch Prozessierung längerer Transkripte entstehen und eine Rolle in der posttranskriptionellen Regulation der Expression von bis zu  $\approx 60\%$  der Säugergene, darunter vielen in immunrelevanten Prozessen, spielen (Wessner *et al.*, 2010). HSA19q13.4 beherbergt den mit mehr als 50 verschiedenen miRNAs größten bis dato beim Menschen identifizierten Cluster dieser Gene<sup>7</sup> (Bentwich *et al.*, 2005).

#### 2.1.4 Identifizierung merkmalsassoziierter Genvarianten in potentiellen funktionellen Kandidatengen

Die phänotypische Variation eines Merkmales, die beispielsweise einem QTL zugrunde liegt, wird durch DNA-Sequenzpolymorphismen, in diesem Fall zumeist SNPs, die in kodierenden oder regulatorischen Bereichen von an der Merkmalsausprägung beteiligten Genen liegen, verursacht.

#### Selektive Poolsequenzierung

Als ressourcenschonende Alternative zur Einzelsequenzierung bei der Kartierung potentiell merkmalsassoziierter Genvarianten beim Nutztier wurde die selektive Poolsequenzierung (*selective DNA pooling*, SDP) vorgeschlagen (Darvasi & Soller, 1994; Lipkin *et al.*, 1998).

Dieses Verfahren besteht aus zwei Stufen: Zunächst untersucht man die Verteilung des interessierenden Merkmales, wobei es sich beispielsweise um phänotypische Beobachtungen oder genetische Zuchtwerte handeln kann, in der zur Verfügung stehenden Population. Für die Sequenzierung werden nun nur die Tiere selektiert, die sich in den Enden<sup>8</sup> dieser Verteilung befinden. SDP wurde durch Darvasi & Soller (1994) unter der Annahme, dass jeder QTL nur einen sehr geringen Teil der phänotypischen Varianz erklärt und die Varianz innerhalb jeder Genotypklasse etwa der Gesamtvarianz entspricht, theoretisch begründet. Gallais *et al.* (2007) zeigten, dass die Methode auch unter der Annahme von *major* QTL gut funktioniert.

Ob nur eines (*unidirectional selective genotyping*, USG) oder beide Enden (*bidirectional selective genotyping*, BSG) der Verteilung berücksichtigt werden, kann von ihrem Typ abhängig gemacht werden; im Falle von Normalverteilung sollte BSG der Vorzug gegeben werden (Gallais *et al.*, 2007; Sen *et al.*, 2009).

SDP ist besonders effizient, wenn die Kosten für die Phänotypisierung der Tiere gegenüber denen der Genotypisierung gering ausfallen (Gallais *et al.*, 2007); da in der Rin-

---

<sup>7</sup>Da dieser Cluster primatenspezifisch ist, lässt sich hier über die Syntenie kein direkter Vergleich zum Rind ziehen

<sup>8</sup>oder Schwänzen, engl. *tails*, daher die Alternativbezeichnung *tail analysis*



derzucht Phänotypen regulär und in großem Umfang erhoben werden und quasi *a priori* verfügbar sind, trifft dies hier in hohem Maße zu.

Eine weitere Kostensenkung wird dadurch erreicht, dass die DNA der selektierten Tiere nicht separat genotypisiert wird, sondern in einem (oder, in einer konzeptionellen Erweiterung des Ansatzes<sup>9</sup>, mehreren (Korol *et al.*, 2007)) DNA-Pool(s) zusammengefasst. Durch Sequenzierung dieser Pools können die Allelfrequenzen bi- oder multialleler Marker verglichen und auch numerisch präzise bestimmt werden (Mariasegaram *et al.*, 2007), jedoch keine individuellen Genotypen.

Allelfrequenzverschiebungen zwischen Extremwertpools, die nach dem Prinzip des SDP generiert wurden, haben zur Identifizierung von SNPs in bovinen Kandidatengenomen, die mit Merkmalen der Eutergesundheit (z.B. Sharma *et al.*, 2006) bzw. Milchleistung (z.B. Skelding *et al.*, 2010) assoziiert sind, beigetragen.

## Heterozygosie am QTL

Um die kausale Mutation eines QTL für das Merkmal der Rate von Zwillingsgeburten beim Rind zu finden, sequenzierten Kim *et al.* (2009) zunächst einen Bullen, der für diesen QTL (auf BTA5) heterozygote Allele aufwies. Dahinter stand die Erwartung, dass SNPs, die mit dem QTL gekoppelt (*linked*) sind und in enger räumlicher Beziehung mit ihm stehen, in QTL-heterozygoten Individuen ebenfalls als Heterozygote in Erscheinung treten. Dies wird auch gelegentlich als Validierungskriterium für bereits postulierte Kandidaten-SNPs herangezogen und dann als Konkordanztest bezeichnet (Ron & Weller, 2007; Seroussi, 2009).

## Pyrosequenzierung

Mittels der Methode der Pyrosequenzierung kann DNA schnell, zuverlässig und kostengünstig (z.B. werden keine markierten Primer benötigt) sequenziert werden, auch und gerade im Hochdurchsatzbetrieb (Diggle & Clarke, 2004). Sie wurde durch Ronaghi *et al.* (1996) eingeführt und beruht auf folgendem Prinzip: Zunächst werden einzelsträngige DNA-Fragmente (ssDNA) der interessierenden genomischen Region erzeugt (z.B. durch PCR); an diesem Template erfolgt in der eigentlichen Reaktion eine DNA-Synthese durch eine Exonuklease-negative ( $\text{exo}^-$ ) DNA-Polymerase. Bei jedem Einbau eines Nukleotids in den DNA-Strang wird durch Hydrolyse des dNTP-Moleküls ein Pyrophosphatmolekül (PPi) freigesetzt, welches durch das Enzym ATP-Sulfurylase für die Umwandlung von Adenosin-5'-Phosphosulfat (APS) in ATP benötigt wird. Das erzeugte ATP dient schließlich als Substrat einer Luziferase (aus dem Glühwürmchen *Photinus pyralis*), welche die Umwandlung von Luziferin

---

<sup>9</sup>bezeichnet als *fractioned pool design*, FPD

zu Oxyluziferin katalysiert, wobei proportional zur Menge des produzierten ATP Photonen emittiert und durch eine Kamera detektiert werden. Da sich zu jedem Zeitpunkt nur eines der vier möglichen dNTPs (wobei dATP durch  $\alpha$ -Thio-dATP ersetzt wird, da letzteres von der Luziferase nicht verarbeitet werden kann) im Reaktionsansatz befindet, kann aus dem zeitlichen Verlauf der Lichtemissionen direkt auf die zugrunde liegende DNA-Sequenz geschlossen werden.

Für die Hochdurchsatzgenotypisierung von SNPs ist die Pyrosequenzierung gut etabliert und validiert (Nordfors *et al.*, 2002; Zhou *et al.*, 2006). Als Technologieplattform wird dabei häufig, so auch in dieser Arbeit, das System PSQ<sup>TM</sup>HS96A (neu: PSQ<sup>TM</sup>96MA) der Fa. Qiagen eingesetzt.

## 2.2 Experimentelle Ansätze der funktionellen Genomanalyse

Die Entwicklung von Hochdurchsatztechnologien und ihre Anwendung in der Biologie hat Hocquette *et al.* (2007) zufolge zu einem vierfachen Paradigmenwechsel geführt: Erstens können nun molekulare Signaturen (anstelle einzelner Schlüssel-moleküle) untersucht werden; zweitens wird die hypothesenbasierte durch hypothesengenerierende Forschung komplementiert; drittens können die Struktur-Funktions-Wechselwirkungen eines Genoms in ihrer Gesamtheit erfasst werden und viertens können integrative Ansätze das Wissen über Genfunktionen in diversen physiologischen Prozessen zusammenführen (Hocquette *et al.*, 2007).

Von den auf verschiedenen Ebenen der Biologie angesiedelten Ansätzen der funktionellen Genomanalyse (Metabolomik, Proteomik, Transkriptomik) wird im folgenden lediglich letztere näher beleuchtet, da Ergebnisse einer Transkriptomanalyse hier verwertet wurden.

### 2.2.1 Holistische mRNA-Untersuchungen - die DNA-Microarray-Technologie

Das Grundprinzip aller DNA-Microarrayverfahren ist ein reverser *Northern Blot*, d.h., eine cDNA-Bibliothek wird gegen ein auf einer festen Oberfläche verankerte DNA-Sonden hybridisiert. Durch Miniaturisierung können auf einer Fläche von wenigen cm<sup>2</sup> dabei mehrere Zehntausend verschiedene Sonden untergebracht werden, deren Hybridisierungsstatus durch hochauflösende Detektionsverfahren (z.B. Laser) eindeutig bestimmt werden kann (Skena *et al.*, 1995). Nach der Art und Synthese der fixierten DNA-Sonden (cDNA bzw. Oligonukleotide) werden mehrere Typen unterschieden.

Die Literatur verzeichnet etliche Beispiele für die Konstruktion boviner cDNA-Microarrays, z.B. immunspezifische (Donaldson *et al.*, 2005; Watkins *et al.*, 2008), zelltypspezifische (Jensen *et al.*, 2006; Burton *et al.*, 2001), gewebespezifische (Lehnert *et al.*, 2004) oder gesamtgenomische (Suchyta *et al.*, 2003).

Eine andere Methode nutzt photolithographische Verfahren, um sequenzspezifische Oligonukleotide *de novo* auf einer Oberfläche zu synthetisieren (McGall *et al.*, 1996). Dieses Verfahren wird hauptsächlich kommerziell von Anbietern wie *Affymetrix* zur Herstellung standardisierter Oligonukleotidmicroarrays eingesetzt, z.B. der *GeneChip*<sup>®</sup>-Plattform.

Auf DNA-Microarrays dieser Plattform ist jedes transkribierte Gen durch eine Reihe von Probesets repräsentiert, wobei die Zahl von Probesets pro Gen recht variabel ist. Die Sequenzen dieser Probesets sind aus Gründen der RNA-Stabilität meist vom 3'-Ende eines Transkriptes abgeleitet. Oft können die einzelnen Probesets eines Gens als Replikate betrachtet werden, sie können aber beispielsweise auch Splicevarianten repräsentieren.

Jedes Probeset besteht wiederum aus einer Kollektion etwa 25 bp langer Nukleotidoligomere (den *probes*), die gleichmäßig über den *GeneChip*<sup>®</sup> verteilt sind. Die Hälfte der *probes* eines Probesets hat die exakte komplementäre Nukleotidsequenz ihrer Zielsequenz (*perfect match*), die andere Hälfte unterscheidet sich an genau einer Position (*mismatch*); vgl. dazu Liu *et al.* (2003).

Hinsichtlich der Beurteilung von Qualität und Reproduzierbarkeit der mit Microarraytechniken erzielten Resultate sprechen Yauk & Berndt (2007) von zwei Phasen. Die erste, etwa von 2000 - 2003, ist gekennzeichnet durch sehr hohe Erwartungen an diese neuen Technologien, gepaart mit Ernüchterung aufgrund von Vergleichsstudien, die zu skeptischen Ausblicken führte (z.B. Hardiman, 2004). Studien aus dem Zeitraum von 2004 - 2006 ergeben ein positiveres Bild. Während in der ersten Phase 8 Studien mit eher positivem 5 Studien mit eher negativem Befund gegenüberstanden, waren unter den in Phase 2 ausgewerteten Studien 29 eher positive, darunter der Bericht des *MicroArray Quality Control Consortium* (MAQC, Shi *et al.*, 2006), und nur 3 eher negative.

In aktuelleren Publikationen, die sich beispielsweise mit der Reproduzierbarkeit von Expressionsprofilen von Monocyten/Makrophagen (Maouche *et al.*, 2008) oder miRNAs (Yauk *et al.*, 2010) bzw. der Übertragbarkeit von mit maschinellen Lernverfahren gewonnenen Klassifikatoren und Signaturgenen (Fan *et al.*, 2010) über verschiedene DNA-Microarrayplattformen hinweg befassen, wurde die Verlässlichkeit dieser Technologie bestätigt; in Klebanov & Yakovlev (2007) wurde gezeigt, dass die Variabilität technischer Replikate als Störgröße in einem Referenzdatensatz einen vernachlässigbaren Einfluss hat.

Unter den Gründen für das vergleichsweise schlechtere Abschneiden der älteren Untersuchungen führen Yauk & Berndt (2007) plattformbedingte Faktoren wie eine mangelhafte Annotation der DNA-Microarrays, Kontaminationen von cDNA-Bibliotheken als Grundlage von cDNA-Arrays, mangelnde Erfahrung in der Handhabung der neuen Technologie, Unkenntnis umweltbedingter Störeinflüsse, beispielsweise durch Ozon (Fare *et al.*, 2003) auf die Fluoreszenzfarbstoffe sowie suboptimales experimentelles Design und Analysemethoden auf.

Zur höheren laborübergreifenden Vergleichbarkeit von DNA-Microarraystudien hat auch

ein etablierter Standard für die zusammen mit den Ergebnissen zu publizierenden experimenteller Rahmendaten, die *Minimum Information About a Microarray Experiment (MIAME)*, (Brazma *et al.*, 2001), beigetragen.

Die wesentlichen Parameter in der Auswertung von DNA-Microarraystudien sind die statistische Signifikanz einer differentiellen Expression eines Gens und deren Quantifizierung, z.B. der *fold change* im Fall von diskreten Untersuchungsgruppen oder ein Regressionskoeffizient bei kontinuierlichen Merkmalen. Meist werden die Gene dabei als statistisch voneinander unabhängige Größen angesehen und  $n = \text{Anzahl der Gene}$  unabhängige Hypothesentests durchgeführt, beispielsweise mit einer T-Statistik, B-Statistik<sup>10</sup>, U-Statistik oder ANOVA. Es wurden Ansätze beschrieben, durch genübergreifende Methoden die Verlässlichkeit der genweisen Hypothesentests zu steigern. Smyth (2004) beschreibt eine Methode, durch Einbeziehung genübergreifender (Varianz-)Informationen mittels linearer Modellierung und empirischer Bayes-Verfahren die Parameter der genweisen T-Tests mit erhöhter Präzision zu bestimmen; der vorgestellte moderierte T-Test kann daher weniger konservativ kalkuliert werden. Dieser und ähnlich gelagerte Ansätze sind strikt genweisen statistischen Tests überlegen (u.a. Jeffery *et al.*, 2006).

Die statistische Signifikanz einer Differenz erlaubt *per se* noch keine Aussage über eine biologische Signifikanz, da prinzipiell arbiträr kleine Expressionsunterschiede bei genügend kleinen Intragruppenvarianzen als statistisch signifikant deklariert werden können. Frühe veröffentlichte DNA-Microarraystudien verzichteten auf dieses Kriterium und berücksichtigten nur die Größe des Expressionsunterschiedes (*fold change*) (z.B. de Risi *et al.*, 1996). Einige aktuellere Untersuchungen wie Fuscoe *et al.* (2007) oder Guo *et al.* (2006) plädieren ebenfalls für die Präferenz des *fold change* bei der Aufstellung von Listen differentiell exprimierter Gene. Dagegen argumentieren Maouche *et al.* (2008) dafür, das Kriterium des *fold change* um eine anhand eines liberalen Schwellenwertes getroffene Aussage zur statistischen Signifikanz zu ergänzen.

---

<sup>10</sup>B ist der (natürliche) Logarithmus der Wahrscheinlichkeitsquote, mit der das Gen tatsächlich differentiell exprimiert ist

### 2.2.2 DNA-Microarrayanalysen als Werkzeug zur Dissektion genetischer Resistenz

DNA-Microarrayanalysen haben sich bereits mehrfach als Hilfsmittel bei der Analyse genetisch determinierter differentieller Merkmalsausprägungen bewährt und dies gleichermaßen in Pflanzen (Wang & Nettleton, 2008), Tiermodellen für Autoimmunerkrankungen (Mattapallil *et al.*, 2008) oder Darmkrebs (Guda *et al.*, 2003) sowie im Nutztier, hier vor allem im Hinblick auf die Pathogenresistenz. Beispiele hierfür sind Parasiteninfektionen bei Schaf (Keane *et al.*, 2006; Andronicos *et al.*, 2010) und Rind (Glass & Jensen, 2007; Jensen *et al.*, 2008; Araujo *et al.*, 2009), allgemeine Resistenz (Nino-Soto *et al.*, 2008) oder als klassisches Beispiel die Marek-Krankheit des Huhnes (Liu *et al.*, 2001).

Der gemeinsame Nenner dieser Studien ist, dass sie Expressionsprofile von Studienobjekten mit differentiellem genetischen Hintergrund vergleichen. Erhebliche Unterschiede bestehen jedoch in der Beschaffenheit desselben bzw. in der Methode seiner Deduktion. In Modelltieren wie Maus und Ratte wird dabei häufig auf Inzuchtlinien zurückgegriffen. Bei Nutztieren liegt der Fall komplizierter und es wurden hier verschiedene Ansätze verfolgt. Einige Studien vergleichen Expressionsprofile von Vertretern verschiedener Rassen mit bekannten divergenten Phänotypausprägungen. So untersuchten etwa Glass & Jensen (2007) und Jensen *et al.* (2008) Expressionsprofile der Rasse Sahiwal von *Bos indicus*, die vergleichsweise resistent gegen den Parasiten *Theileria annulata* ist, im Vergleich mit Holsteinrindern, und Verschoor *et al.* (2010) verglichen Holstein- und Jerseyrinder hinsichtlich subklinischer Paratuberkuloseinfektionen; diese Krankheit hat in letzterer Rasse eine besonders hohe Inzidenz.

Andere Studien wie die von Andronicos *et al.* (2010) zur Nematodenresistenz des Schafes oder Araujo *et al.* (2009) zu der des Rindes nutzen experimentelle Populationen, die für das interessierende Merkmal segregieren. Nino-Soto *et al.* (2008) teilten Holsteinrinder aufgrund phänotypischer Parameter nach immunogener Stimulation retrospektiv in Gruppen mit hoher, mittlerer und niedriger Immunreaktivität ein und zogen aus Expressionsunterschieden zwischen diesen Rückschlüsse auf genetische Faktoren der Krankheitsresistenz.

Für die Aussagekraft von DNA-Microarrayexperimenten hinsichtlich genetischer Faktoren ist es von Vorteil, wenn in die Auswahl der Studienobjekte Erkenntnisse zur genetischen Architektur einfließen (de Koning *et al.*, 2005). Zu dieser Thematik gibt es vielfältige Literatur aus dem Bereich der genetischen Genomik (*genetical genomics*, Jansen & Nap, 2001), wo selektives Transkriptionsprofiling eingesetzt wird, um bei maximaler Erhaltung der Aussagekraft den Umfang der benötigten Microarrayanalysen zu minimieren. In einer aktuellen Vergleichsstudie hat sich dabei die Selektionsstrategie der maximalen genetischen Dissimilarität (Jin *et al.*, 2004) gegenüber zum Beispiel rekombinations- oder phänotypbasierten Strategien als überlegen erwiesen (Cardoso *et al.*, 2008). Unter den von Jin *et al.* (2004) diskutierten Simulationsansätzen wiederum weist die Fokussierung auf genetische Dissimila-

rität innerhalb einer begrenzten genomischen Region, beispielsweise einem QTL, gegenüber gesamtgenomischer Dissimilarität die beste Bewertung auf.

Wählt man für eine *genetical genomics*-Studie Versuchstiere aus, die für einen interessierenden QTL entgegengesetzt homozygote Allele aufweisen, so erhöht sich die Teststärke einer solchen Analyse auf das Niveau rekombinanter Inzuchtlinien, die generell nur in Modelltieren verfügbar sind (de Koning *et al.*, 2005).

Eine weitere Erwägung betrifft den physiologischen Zustand der untersuchten Tiere. Als auf die Analyse von Krankheitsresistenzen ausgelegte Erweiterung des ursprünglichen Konzeptes von Jansen & Nap (2001) schlagen de Koning *et al.* (2005) vor, solche Untersuchungen an immunstimulierten Tieren vorzunehmen, um aus gegebenenfalls divergierenden Immunreaktionen Schlüsse auf genetische Resistenzfaktoren zu ziehen. Letztere wurden allerdings in einer Studie von Diez-Tascon *et al.* (2005) zur Nematodenresistenz des Schafes stark von den Auswirkungen der Infektion überlagert. Um diesen *Confounding*-Effekt zu eliminieren, arbeiteten in einer Nachfolgestudie Keane *et al.* (2006) folglich nur mit nichtinfizierten Schafen. Auf ähnliche Probleme stießen bereits früher Schwerin *et al.* (2003) bei der Analyse der Mastitisresistenz des Rindes und analoge Schlussfolgerungen führten zur Konzeption von M.A.S.-Net.

## 2.3 Integrative Ansätze zur Identifizierung potentieller funktionaler Kandidatengene

QTL-Kartierungen haben, abhängig von der zugrunde liegenden Markerdichte eine Auflösung im ein- bis unteren zweistelligen cM-Bereich. Genomabschnitte dieser Größe umfassen bis zu mehrere Hundert positionelle Kandidatengene (Wayne & McIntyre, 2002). Ron & Weller (2007) schlugen daher vor, für die Identifizierung von potentiellen funktionellen Kandidatengenen und darauf folgend von merkmalsbeeinflussenden Nukleotiden (*quantitative trait nucleotide*, *QTN*) mehrstufige Protokolle unter Einbeziehung mehrerer einzelner Indizien zu implementieren. Konkret stellen sie vier Kriterien auf, die in Tabelle 2 erläutert sind.

### 2.3.1 Funktionelle Annotation

Eines dieser Indizien ist die Einordnung der positionellen Kandidatengene aufgrund ihrer bekannten physiologischen Rolle. Diese kann aus dem Corpus der für das jeweilige Gen vorliegenden wissenschaftlichen Literatur extrahiert werden (Natarajan & Ganapathy, 2007). Im Kontext von Krankheit und Resistenz ist hier die Datenbank OMIM (*Online Mendelian Inheritance in Man*, Hamosh *et al.*, 2002) von Bedeutung, die das Ziel verfolgt, durch manuelle Kuratierung der wissenschaftlichen Literatur Beziehungen zwischen humanen Genen und Erkrankungen mit genetischer Komponente herzustellen und die mittlerweile einen Umfang von 18.000 detaillierten Datensätzen aufweist (Nakazato *et al.*, 2009). Computerbasierte

**Tabelle 2:** Überblick über die vier von Ron & Weller (2007) aufgestellten Selektionskriterien, die idealerweise auf ein Gen zutreffen sollten, um es als ein potentielles funktionelles Kandidatengen für einen QTL beim Nutztier in Betracht zu ziehen, wobei von den zum Zeitpunkt der Publikation anerkannten Kandidatengenen keines alle vier erfüllte.

Kriterium	Das Gen ...
1	... hat eine bekannte physiologische Funktion für den Phänotyp von Interesse
2	... wurde experimentell in anderen Spezies durch z.B. <i>Knock Out</i> -Studien, Mutationen oder Transgene als merkmalsrelevant bestätigt
3	... wird vorzugsweise (-orig.: <i>preferentially</i> -) in merkmalsrelevanten Geweben exprimiert
4	... wird vorzugsweise während merkmalsrelevanter (ontogenetischen) Entwicklungsphasen exprimiert

Lösungen wie das von Bajdik *et al.* (2005) vorgestellte Programm *CGMIM* streben die automatisierte Extraktion von Informationen aus den in OMIM vorgehaltenen Datensätzen zum Zweck der automatisierten funktionellen Genannotation an. Dieser Ansatz stößt jedoch an Grenzen, da die Informationen in der OMIM in natürlicher Sprache präsentiert werden, die textanalytischen Verfahren nur schwer zugänglich ist.

Nakazato *et al.* (2009) plädieren daher für die Nutzung der *Medical Subject Headings (MeSH)* für funktionelles *Profiling* von Genen. Dieses System wurde 1960 eingeführt; eine Darstellung der historischen Entwicklung findet sich bei Coletti & Bleich (2001). Es stellt nach einer größeren Umstrukturierung (Johnston *et al.*, 1998) heute ein aus Deskriptoren, Konzepten und Termen aufgebautes hierarchisches, standardisiertes biomedizinisches Vokabular bereit (Nelson *et al.*, 2004). Die mehr als 24.000 Terme in 11 Hierarchieebenen werden für die manuelle Annotation von in der wissenschaftlichen Literaturdatenbank MEDLINE indizierten Publikationen durch Arbeitsgruppen des *National Center for Biotechnology Information* (NCBI) verwendet. Stevenson *et al.* (2008) sowie Natarajan & Ganapathy (2007) bestätigen die hohe Qualität der MeSH-Annotation von MEDLINE-indizierten Artikeln durch deren erfolgreichen Einsatz in kompetitiven Versuchen zu *Word Sense Disambiguation* bzw. funktioneller Annotation von Gensets.

### 2.3.2 Expressionsdaten

Neben der physiologischen Funktion eines Gens sind für seine Klassifizierung als potentielles funktionelles Kandidatengen nach Ron & Weller (2007) auch andere Faktoren zu berücksichtigen. An vorderer Stelle wird dabei der Expressionsstatus des Gens in merkmalsrelevanten Geweben, wie sie zum Beispiel in Ansätzen der funktionellen Genomanalyse evaluiert werden kann (s.o.) genannt. Für die Zusammenführung von biologischem Hintergrundwissen und Genexpressionsstudien gibt es in der Literatur verschiedene Ansätze.

#### Filterstrategien

Hier wird zunächst ein Set an Kriterien definiert, welche dann als Filter für Kandidatengene genutzt werden, wobei nur Gene in der Intersektion aller Filter für die experimentelle Analyse berücksichtigt werden. Eine direkte Umsetzung der Vorschläge von Ron & Weller (2007) bieten dabei Ron *et al.* (2007); dort wurden Expressionsprofile von verschiedenen Stadien der Entwicklung des murinen Milchdrüsengewebes als Filter für bovine positionelle und funktionelle Kandidatengene für Milchleistungsparameter genutzt. Analog gingen Lu *et al.* (2008) bei der Identifizierung valider Kandidatengene für murines Vorderhirngewicht sowie Wayne & McIntyre (2002) auf der Suche nach Kandidatengenen für Ovariolenanzahl bei der Fruchtfliege *Drosophila* vor. Methodisch können hier auch bioinformatische Werkzeuge eingeordnet werden, welche vorgegebene Listen differentiell exprimierter Gene auf eine Überrepräsentation bestimmter physiologischer Kategorien untersuchen. Ein Beispiel, das für diesen Zweck das MeSH-Vokabular nutzt, ist *MeSHer* (Djebbari *et al.*, 2005).

Problematisch an einer solchen Herangehensweise kann jedoch sein, dass die für einen gegebenen Phänotyp relevanten Gene keineswegs die mit den größten Expressionsunterschieden sein müssen; tatsächlich können irrelevante, von Miklos & Maleszka (2004) als *innocent bystander* apostrophierte Expressionsunterschiede diese überlagern. Gleichzeitig können geringe Expressionsunterschiede, z.B. innerhalb von Signaltransduktionskaskaden, welche von den Methoden der DNA-Microarrayauswertung teils nicht als statistisch signifikant erkannt werden und die in den beschriebenen Ansätzen daher nicht berücksichtigt werden (können), großen Einfluss auf die differente Ausprägung von Phänotypen haben.

#### Globale Ansätze

Globale Ansätze versuchen daher, das vorhandene biologische Hintergrundwissen in die Formulierung der bei der Auswertung von DNA-Microarrayexperimenten zu testenden Hypothesen einfließen zu lassen. Methoden wie *globaltest* (Goeman *et al.*, 2004) und *GlobalAncova* (Hummel *et al.*, 2008) setzen die Erkenntnis, dass Genprodukte in Netzwerken agieren (Leung & Cavalieri, 2003) um, indem sie die Hypothese testen, ob die Gene in einem *a priori*



definierten Netzwerk ein koordiniertes Expressionsmuster zeigen, ohne dass die beteiligten Gene im einzelnen statistisch signifikant differentiell exprimiert sein müssen.

Einen anderen Weg der Integration der Biologie in die DNA-Microarrayauswertung, ohne die Auflösung auf der Ebene des einzelnen differentiell exprimierten Gens aufzugeben, beschreiben Morrison *et al.* (2005). Sie modifizierten Googles *PageRank*-Algorithmus; das resultierende Programm *GeneRank* ordnet Gene u.a. nach dem Kriterium des Expressionsstatus von physiologisch funktionell mit ihm verknüpften Genen in eine Rangliste ein. Ein frei wählbarer Parameter steuert dabei das Verhältnis, in dem Expression und physiologische Verknüpfung gewichtet werden. Diesen Punkt kritisierten Daigle & Altman (2008) als willkürlich und stellten das Programm *M-BISON* vor, mit dem Anspruch, die Signifikanz der differentiellen Expression eines Gens unter Berücksichtigung experimenteller Störfaktoren und der vor dem Hintergrund des biologischen Wissens erwarteten Coregulationrelation mit anderen Genen darzustellen. Wang *et al.* (2010) stellten fest, dass für diese Problematik noch keine definitive Lösung gefunden wurde, da alle bisherigen Lösungen auf heuristischen Ansätzen beruhen. Sie führen die *Coordinative Component Analysis* (COCA) ein.

Hedegaard *et al.* (2009), die mehreren Gruppen zu Vergleichszwecken identische Genexpressionsdaten<sup>11</sup> zu Verfügung stellten, fanden, dass einerseits die biologische Interpretation solcher Experimente erheblich von den angewandten statistischen und analytischen Verfahren abhängt, andererseits methodenübergreifend über manche biologische Schlussfolgerungen ein Konsens erzielt werden konnte.

## 2.4 QTL und Kandidatengene der Eutergesundheit

Für die Mastitisempfindlichkeit wurden bereits eine Reihe von QTL beschrieben. Einen Überblick über die in der Onlineresource *AnimalQTLdb* (Hu *et al.*, 2007) bis dato verfügbaren Informationen zu denjenigen der 29 Autosomen (zuzüglich der Geschlechtschromosomen) des Rindergenoms, für die beim Milchrind für das Merkmal SCS bereits QTL bekannt sind, gibt Tabelle 32 (im Anhang).

Potentielle funktionelles Kandidatengene, welche für einzelne QTL diskutiert werden, sollen im Folgenden kurz vorgestellt werden. Im auf BTA2 von Bennewitz *et al.* (2003) gefundenen QTL liegt das Gen ***IL8RB***, das einen Rezeptor für das proinflammatorische Cytokin Interleukin 8 (IL-8) codiert. Die Identität von *IL8RB* als Träger der kausalen Mutation des QTL ist gegenwärtig in der Literatur umstritten (Leyva-Baca *et al.*, 2007, 2008; Goertz *et al.*, 2009).

---

<sup>11</sup>zur zeitpunktabhängigen Genexpression im Jejunum von Hühnern nach Infektionsversuchen, gemessen mit dem gesamtgenomischen *ARK-Genomics Gallus gallus 20 K v1.0*

Mehrere diskutierte Kandidatengene liegen auf BTA6. Zum einen **SPP1**, welches das Cytokin Osteopontin codiert, ein in T-Lymphocyten und Makrophagen exprimiertes Cytokin mit vielfältigen immunmodulatorischen Funktionen, u.a als Chemokin für Makrophagen und bei der Induktion einer Th1-Immunantwort (Wang & Denhardt, 2008). Es vermittelt Resistenz gegen eine Anzahl intrazellulärer Pathogene (Weber *et al.*, 2002). Alain *et al.* (2009) identifizierten in diesem Gen Polymorphismen mit signifikanter Assoziation zum Merkmal SCS im Kanadischen Holstein. In der telomeren Region von BTA6, in einem QTL für klinische Mastitis, ist das Gen **MUC7** lokalisiert, dessen Genprodukt, das Glykoprotein MG2, bei der Beseitigung von Bakterien auf mukösen Oberflächen (Grys *et al.*, 2005) eine Rolle spielt. Die Feinkartierungs- und SNP-Assoziationsergebnisse von Nilsen *et al.* (2009a) weisen auf dieses Gen als wahrscheinlichstes Kandidatengen dieses QTL hin.

Das antiinflammatorische Cytokin Interleukin 10 (IL-10) und sein Rezeptor IL-10R, deren Rolle im Verlauf der klinischen Mastitis gut untersucht ist (Bannerman *et al.*, 2004; Oviedo-Boyso *et al.*, 2007), wurden von Verschoor *et al.* (2009) auf Sequenzpolymorphismen untersucht, wobei sie nur im Gen für die  $\alpha$ -Kette von IL-10R, **IL10RA** SNPs mit signifikanter Assoziation mit dem Merkmal SCS fanden. **IL10RA** ist auf BTA15 in enger räumlicher Nähe zum von Ashwell *et al.* (2004) detektierten QTL für SCS lokalisiert.

Ein SCS-QTL auf BTA22 beinhaltet mehrere vielversprechende Kandidatengene. **LTF**, das Gen für Laktotransferrin, ein Protein mit eisenbindender, antimikrobieller und immunmodulatorischer Funktion (Gonzalez-Chavez *et al.*, 2009), wird seit langem diskutiert, ohne dass ein definitiver Zusammenhang mit Mastitisresistenz bisher nachgewiesen werden konnte (Pawlik *et al.*, 2009), aber vielversprechend erscheint (Arnould *et al.*, 2009). Auf der anderen Seite ziehen Patel *et al.* (2009) Laktotransferrin gar als Pathogenitätsfaktor für Euterinfektionen mit *S. uberis* in Betracht.

Ebenfalls in diesem QTL liegt ein Cluster von Chemokinrezeptorgenen. In einem dieser Gene, **CCR2**, beschreiben Leyva-Baca *et al.* (2007) einen SNP mit signifikanter Assoziation mit dem Merkmal SCS. Darüber hinaus beschreiben Griffin *et al.* (2005) im QTL einen Cluster von mindestens 11 Mitgliedern der Familie der *cathelicidin antimicrobial peptide* (**CAMP**)-Gene.

Aus der Neurobiologie ist das Gen **FEZL** (*forebrain embryonic zinc finger like*) bekannt, das einen in der neuronalen Entwicklung wichtigen Transkriptionsfaktor codiert. Jedoch wird seine Expression auch im Verlauf von Mastitiden heraufreguliert und führt, vermittelt über Semaphorin 5A (SEMA5A) seinerseits die Induktion von TNF- $\alpha$  und IL-8. Die Effizienz dieser Induktion und als Konsequenz letztlich die Resistenz gegen Mastitis wird reduziert durch eine Insertion von 3 Nukleotiden im **FEZL**-Gen, die zur Verlängerung einer Kette

von 12 Glycinen um einen weiteren Glycinrest führt. Der Effekt dieses Polymorphismus auf das Merkmal SCS wurde ursprünglich in fernöstlichen Herden beschrieben (Sugimoto *et al.*, 2006) und wurde kürzlich in Holstein-Populationen bestätigt (Sugimoto *et al.*, 2011).

Lokalisiert ist *FEZL* ebenfalls auf BTA22.

Auch der Haupthistokompatibilitätskomplex **MHC** (*major histocompatibility complex*, beim Rind: *bovine leukocyte antigen*, BoLa), scheint eine Rolle bei der Vermittlung von Resistenz gegen Mastitis zu spielen. Moleküle des MHC sind an der Antigenpräsentation während der Induktion einer adaptiven Immunreaktion beteiligt, sowie nach neueren Erkenntnissen auch an der Apoptoseinduktion.

Für verschiedene Allele des MHC-Klasse 2-Lokus **BoLA DRB3.2**, dessen Genprodukte für die Präsentation von Peptidantigenen extrazellulären Ursprungs (z.B. durch Phagozytose internalisierte) an CD4-positive T-Lymphocyten zuständig sind, konnte ein signifikanter Einfluss auf SCS und klinische Mastitis nachgewiesen werden (Kulberg *et al.*, 2007; Rupp *et al.*, 2007; Yoshida *et al.*, 2008).

Dies gilt ebenso für die Oberflächenrezeptoren **TLR2** und **TLR4** (Pant *et al.*, 2008; Sharma *et al.*, 2006). Sie gehören zur Familie der Toll-like-Rezeptoren, die als Komponenten des angeborenen Immunsystems konservierte Oberflächenstrukturen von Pathogenen (*pathogen associated molecular pattern*, PAMP) erkennen; mindestens zehn verschiedene TLRs wurden bisher beim Rind identifiziert (Menzies & Ingham, 2006). TLR2 ist ein Lipoproteinrezeptor<sup>12</sup>, während TLR4 spezifisch auf das Endotoxin gram-negativer Bakterien Lipopolysaccharid reagiert.

Ogorevc *et al.* (2009) haben in einer Metadatenbank die in verschiedenen Quellen verfügbaren Informationen bezüglich Kandidatengen und QTL für Mastitis (und Milchproduktion) zusammengetragen und aufbereitet. Tabelle 31 (im Anhang) stellt darauf basierend eine Zusammenfassung verschiedener Evidenztypen vor sowie in einer eingebetteten Abbildung eine detailliertere Darstellung für BTA18.

---

<sup>12</sup>Zahringer *et al.* (2008); diskutiert wird auch eine breitere und unspezifischere Palette von Liganden.

## Teil II

# Material und Methoden

## 1 Ein bioinformatischer heuristischer Ansatz zur Ermittlung einer Rangfolge potentieller funktioneller Kandidatengene

### 1.1 Ermittlung und Beschaffung (*Retrieval*) relevanter Publikationen für die positionellen Kandidatengene

In den folgenden Abschnitten wird der bioinformatische heuristische datenbankbasierte Ansatz beschrieben, mithilfe dessen für eine Liste positioneller Kandidatengene ein stichwortbasiertes Ranking potentieller funktioneller Kandidatengene erstellt wurde.

Hierbei kamen hauptsächlich in der Programmiersprache Perl geschriebene Softwareskripte zum Einsatz, wobei ein gemischter prozeduraler/objektorientierter Ansatz verfolgt wurde. Letzterem entsprechend wurden für die wesentlichen hier verwendeten Informationstypen jeweils eigene Klassen konstruiert. Jede Klasse definiert dabei einen Datencontainer beliebiger Komplexität inklusive Zugriffsmethoden für einzelne Daten. Tabelle 3 gibt eine nähere Beschreibung der drei definierten Klassen.

**Tabelle 3:** Der hier beschriebene Ansatz zur stichwortbasierten Datenbankrecherche wurde objektorientiert konzipiert. Diese Methode beruht auf der Definition von „Klassen“, welche Datenstrukturen und Methoden für den Zugriff auf die Daten definieren. Die titelgebenden „Objekte“ sind die konkreten Realisierungen dieser Klassen. Hier wurde für die drei wesentlichen Entitäten des Ansatzes jeweils eine Klasse definiert.

Name	Beschreibung
<i>Gene</i>	Ein Objekt dieser Klasse wurde für jedes positionelle Kandidatengen erzeugt. Es stellt Speicherplätze ( <i>Slots</i> ) und Zugriffsmethoden sowohl für Basisinformationen wie Gensymbol, EntrezGeneID etc. als auch für eine Liste der identifizierten Literaturverweise, ferner für eine Liste mit Objekten der Klasse <i>Article</i> (s.u.) bereit.
<i>Article</i>	Für jede heruntergeladene Artikelzusammenfassung ( <i>Abstract</i> ) wurde ein Objekt dieser Klasse erzeugt, welches Slots und Zugriffsmethoden für den gesamten <i>abstract</i> als auch für Teilinformationen wie die Artikelidentifikationsnummer (PMID) und die Liste der MeSH-Terme enthält.
<i>Theme</i>	Für die Kollektion hypothesenbasierter Stichworte wurde ein Objekt dieser Klasse erzeugt. Es verfügt über Slots und Zugriffsmethoden für je eine Liste von Suchbegriffen und eine Liste auszuschließender Begriffe.

Diese Klassendefinitionen wurden in einer Perl-Klassenbibliothek abgelegt, welche von den Perl-Arbeitsskripten beim Programmstart eingebunden wurde.

Das erste Skript, welches im Prozess aufgerufen wurde, war für die Ermittlung genrelevanter Literaturverweise und die Bereitstellung der entsprechenden Literatur in einem auswertbaren Format verantwortlich. Es las zunächst die Liste positioneller Kandidatengene in Tabellenform ein, erzeugte für jedes Kandidatengen ein Objekt der Klasse *Gene* und schrieb die Geninformationen aus der Tabelle (die mindestens die EntrezGeneID beinhaltete) in die jeweils dafür vorgesehenen Speicherplätze des Objektes.

Für die anschließende Ermittlung genrelevanter Literaturverweise wurde die textbasierte Datenbank Gene2PubMed eingesetzt. Ihre Struktur wird in Tabelle 4 beschrieben.

**Tabelle 4:** Die Datenbank Gene2PubMed, die vom NCBI unter der Webadresse `ftp://ftp.ncbi.nih.gov/gene/DATA/gene2pubmed.gz` bereitgestellt wird, bildet MEDLINE-indizierte Literatur auf Gene, für deren Beschreibung sie relevant sind, ab. Gene2PubMed wird als dreispaltige Tabelle mit einem Datensatz pro Zeile veröffentlicht; die Bedeutung der drei Felder pro Datensatz ist hier dargestellt.

Feldname	Beschreibung
tax_ID	Die taxonomische Identifikationsnummer der Spezies, z.B. 9913 für <i>Bos taurus</i> und 9606 für <i>Homo sapiens</i>
EntrezGeneID	Die unikale Identifikationsnummer des Gens innerhalb des Datenbestandes des NCBI; außerhalb, z.B. in Ensembl, als EntrezGeneID o.ä. geführt. Sowohl im heuristischen datenbankbasierten Ansatz als auch im kombinierten Datenbank/Transkriptom-Ansatz wurden die Gene durchgehend mit dieser ID angesprochen.
PubMedID	Die unikale Identifikationsnummer einer in PubMed gelisteten Publikation (PubMedID, PMID)

Sie wurde als *flat file database* lokal gespeichert und vom hier behandelten Skript in eine Datenstruktur eingelesen, welche jeder EntrezGeneID eine Liste von eindeutigen PubMed-Identifikationsnummern (PMIDs) zuwies<sup>13</sup>. Indem nun über die *Gene*-Objekte iteriert wurde, konnte durch den Abgleich der EntrezGeneIDs jedem positionellen Kandidatengen eine Liste relevanter PMIDs zugewiesen und im entsprechenden Slot des Objektes gespeichert werden.

Die über ihre PMID identifizierten wissenschaftlichen Publikationen bzw. deren für die automatisierte Auswertung relevanten Teile mussten für die weitere Analyse lokal gespeichert werden. Hier kamen die vom NCBI bereitgestellten *Entrez Programming Utilities* (EUtils)

<sup>13</sup>Dies wurde im Skripts als *Hash of Arrays* (HoA) realisiert

zum Einsatz. Diese bilden eine Kollektion von sieben<sup>14</sup> serverseitigen Programmen, die ein Interface für das vom NCBI zur Verwaltung ihres Datenbestandes eingesetzten System Entrez bereitstellen. Eine Übersicht über diese sieben Programme bietet Tabelle 5.

**Tabelle 5:** Die in den EUtils zusammengefassten Programme können durch eine URL der Form `http://eutils.ncbi.nlm.nih.gov/entrez/eutils/Name_des_Service.fcgi?` angesprochen werden. Dieser Basis-URL werden Anfragen in der Form *Parameter = Wert* angefügt; mehrere Parameter-Wert-Paare werden durch ein `&`-Symbol getrennt.

Name des Service	Parameter	Ausgabe
EInfo	Datenbankname	Zahl der Einträge, letzte Aktualisierung etc.
EGQuery	Suchbegriff	Zahl der Fundstellen
ESearch	Suchbegriff, Datenbank	UID(s) der Fundstelle(n)
ESummary	UIDs, Datenbank	Dokument-zusammenfassung
EPost	UIDs, Datenbank	Zugriffsschlüssel, Web-Umgebung
EFetch	UIDs, Datenbank	Datensatz
ELink	UIDs, Datenbank	Verlinkung

Für jeden dieser Services stellt NCBI folgende Basis-URL (*uniform resource locator*) bereit:

„`http://eutils.ncbi.nlm.nih.gov/entrez/eutils/Name_des_Service.fcgi?`“.

Durch Anhängen von Textbausteinen, die jeweils aus einem Parameter und dem gewünschten Wert dieses Parameters bestehen, an diese Basis-URL kann schrittweise eine Suchanfrage zusammengesetzt und auf geeignetem Wege an die Server des NCBI übertragen werden. In dieser Arbeit wurde für die Beschaffung der mittels der Datenbank Gene2PubMed ermittelten relevanten wissenschaftlichen Artikel der Service EFetch genutzt.

Zunächst wurde dabei die oben gezeigte Basis-URL mit den für die Aufgabenstellung notwendigen Parametern verknüpft. Jeder Parameter wurde dabei in der Form *Parameter=Wert* angegeben. Mehrere Parameter-Wert-Paare wurden dabei durch ein `&` getrennt. Die eingesetzten Parameter sind in Tabelle 6 aufgeführt.

Tabelle 6 zeigt die Konfigurationsparameter des Services EFetch, die für den hier verfolgten Zweck der Beschaffung genrelevanter Publikationen eingesetzt wurden.

<sup>14</sup>Ein achttes Programm, *ESpell*, wurde nach Abschluss der hier vorgestellten Arbeit in das NCBI-Angebot aufgenommen.

**Tabelle 6:** Für die positionellen Kandidatengene wurden die Identifikationsnummern (PMIDs) von relevanten, in MEDLINE indizierten Publikationen ermittelt und als Abfrage mittels des Services EFetch an die NCBI-Datenbanken übermittelt. Die in der Tabelle gezeigten Parameter wurden in der Form *Parameter = Wert* an die Basis-URL (siehe Text und Tab. 5) angefügt.

Parameter	Wert	Bedeutung
db	pubmed	Benutzung der Datenbank PubMed. Die Auswahlmöglichkeiten bei den anderen Parametern werden durch diese Wahl vorgegeben
ret_mode	text	weist EFetch an, seine Ergebnisse als einfachen Text auszugeben, Alternativen sind XML,HTML und asn.1
ret_type	medline	die angeforderten Publikationen werden im MEDLINE-Format ausgegeben

Es seien nun P1 und P2 zwei PMIDs von in der PubMed indizierten wissenschaftlichen Artikeln mit Relevanz für ein positionelles Kandidatengen. Abbildung 3 zeigt beispielhaft auf, wie in diesem Falle die vollständige Abfrage-URL (die in der Variable „\$search\_string“ enthalten ist) aussieht. Sie zeigt ebenfalls, wie mithilfe des Perl-Modules LWP diese Anfrage an die Server des NCBI übermittelt und das zurückgegebene Resultat in der Variable „\$result“ zwischengespeichert wurde.

```

my $search_string =
'http://eutils.ncbi.nlm.nih.gov/entrez/eutils/efetch.fcgi?db=pubmed&
ret_mode=text&ret_type=medline&id=P1+P2';
:
use LWP::UserAgent;
my $agent = LWP::UserAgent->new();
$agent->proxy('http', <Proxyadresse>);
my $returned = $agent->get($search_string);
my $result = $returned->content;

```

**Abbildung 3:** Die nach den im Text und in Tabelle 6 vorgestellten Prinzipien aus Textbausteinen skriptgesteuert zusammengesetzten Datenbankabfragen wurden mithilfe des Perl-Modules LWP an die Programmierschnittstelle EUtils des NCBI übermittelt und das zurückgegebene Ergebnis lokal gespeichert. Die Box zeigt dafür einen Beispielcode in Perl, in dem das Resultat der Abfrage „\$searchstring“ an den Service EFetch in der Variable „\$result“ gespeichert wird.

Die Variable „\$result“ aus dem Beispiel in Abbildung 3 enthielt nun die angeforderten, für das jeweilige positionelle Kandidatengen relevanten Artikel im MEDLINE-Format als fortlaufenden Text. Dieser wurde nun in die einzelnen Artikeldatensätze aufgetrennt und für

jeden Artikel ein Objekt der Klasse *Article* erzeugt. Eine Liste mit allen Artikelobjekten wurde nun im dafür vorgesehenen Speicherplatz des jeweiligen Genobjektes abgelegt.

## 1.2 Ermittlung des datenbankbasierten Rangkoeffizienten

### 1.2.1 MEDLINE-Format

Das MEDLINE-Format, in welchem die Artikel angefordert wurden, hat mehrere für diese Untersuchung nützliche Eigenschaften. Datensätze im MEDLINE-Format enthalten die gängigen Informationen wie Titel, Zusammenfassung (*abstract*), Autorenliste, Datum der Einreichung, Datum der Publikation, Zeitschrift etc. Hinzu kommt eine Liste von *Medical Subject Headings* (MeSH), die die Grundlage für die in dieser Untersuchung dargestellten Bewertungsmethode bildeten und in der Einleitung bereits vorgestellt wurden. Für die automatisierte Auswertung mittels Computerprogrammen wie in dieser Untersuchung durchgeführt, ist das MEDLINE-Format in besonderer Weise geeignet, indem jeder Information eine Identifikationsmarke (Label) vorangestellt ist, anhand derer Programme den Typ der jeweiligen Information ermitteln können. So wird z.B. der Titel der Publikation mit dem Label *TI* eingeleitet, die Zusammenfassung (*abstract*) mit *AB* und MeSH-Terme mit *MH*. Ein Beispiel für das MEDLINE-Format bietet die Box auf S. XX (im Anhang).

Das Perl-Modul *Data::Dumper* bietet die Möglichkeit, Perl-Datenstrukturen als physisches Image auf der Festplatte zu speichern und in späteren Programmläufen wiederherzustellen. Diese Funktionalität wurde an diesem Punkt des Programmablaufes genutzt, um die gesammelten Datensätze, d.h. Gen-Objekte mit den ihnen zugewiesenen Listen von Referenzpublikationen als wiederverwendbares Zwischenergebnis zu speichern.

In einem weiteren Perlskript erfolgte nun die Ermittlung des datenbankbasierten Score für die positionellen Kandidatengene. Die zuvor erzeugte Imagedatei mit den *Gene*-Objekten wurde eingelesen und die Objektliste rekonstruiert.



### 1.2.2 Reguläre Ausdrücke

Das Kandidatengenranking wurde durch einen Vergleich der MeSH-Indizierung der identifizierten relevanten wissenschaftlichen Publikationen mit einer hypothesenbasierten Stichwortliste erzielt. Dementsprechend wurden die gewählten Stichworte von existierenden MeSH-Termen abgeleitet und als einfache reguläre Ausdrücke, die jeweils potentiell auf eine Mehrzahl von MeSH-Termen passten, angelegt. Reguläre Ausdrücke (*regular expressions*) entstammen der Theoretischen Informatik und sind eine Art von formaler Sprache, mit der sich nahezu beliebige Textmuster exakt beschreiben lassen. In einem regulären Ausdruck kann ein Zeichen entweder für sich selbst stehen oder als Metasymbol eingesetzt werden. Letztere können ganze Zeichenklassen repräsentieren, andere Zeichen quantifizieren oder die Verankerung des regulären Ausdrucks an einer spezifischen Position im Suchtext<sup>15</sup> erzwingen.

Die in dieser Arbeit tatsächlich verwendeten Metasymbole sind mit ihren Bedeutungen in Tabelle 7 aufgeführt.

**Tabelle 7:** Die in dieser Arbeit für die Implementierung des heuristischen datenbankbasierten Ansatzes genutzten Metasymbole in Regulären Ausdrücken.

Symbol	Bedeutung
\b	markiert eine Wortgrenze
\s	Leerzeichen
\d	Ziffer
\/	literaler Schrägstrich
+	das vorherige Zeichen tritt ein- oder mehrfach auf

Die Implementierung der Stichworte als möglichst allgemein gehaltene reguläre Ausdrücke brachte mit sich, dass potentiell neben erwünschten, da in Übereinstimmung mit der Arbeitshypothese, auch unerwünschte MeSH-Terme als Treffer erkannt wurden. Aus diesem Grund wurde eine weitere Liste mit regulären Ausdrücken, die spezifisch auf solche unerwünschten Begriffe passten, angelegt, die später als Ausschlussliste Verwendung fand.

Beide Stichwortlisten wurden in einer Textdatei abgespeichert und vom für die Ermittlung des datenbankbasierten Score geschriebenen Skript eingelesen, das daraus ein Objekt der Klasse *Theme* mit Zugriffsmethoden für die erwünschten und nicht erwünschten Stichworte erzeugte.

<sup>15</sup>sogenannte Zusicherungen der Länge Null

### 1.2.3 Der datenbankbasierte Rangkoeffizient

Der in Abbildung 4 gezeigte Pseudocode in Kombination mit Formel 3 beschreiben, wie aus den vorliegenden Annotationen für jedes positionelle Kandidatengen ein datenbankbasierter Rangkoeffizient (DbRK) ermittelt wurde. Im Pseudocode wird die in Perl gebräuchliche Syntax für die Verwendung regulärer Ausdrücke verwendet. Die Form:

$$A \sim /B/i$$

bedeutet: Eine durch den regulären Ausdruck B definierte Zeichenkette befindet sich an beliebiger Stelle in der Zeichenkette A. Abweichungen in der Groß/Kleinschreibung werden dabei ignoriert (Parameter i).

Die Variable *Zaehler* gibt die Zahl der Übereinstimmungen zwischen der hypothesenbasierten Stichwortliste und der Gesamtheit der MeSH-Annotationen der für ein gegebenes Kandidatengen gefundenen Literatur an. Die Länge der Liste *Unikale\_Terme* ist ein Maß für den Umfang der für dieses Gen annotierten Literatur.

Der Datenbank-Rangkoeffizient (DbR-Koeffizient, DbRK) wird nun errechnet, indem beide Werte ins Verhältnis gesetzt werden:

$$DbRK = \frac{Zaehler}{n_{Unikale\_Terme}} \quad (3)$$

```

Einschlussliste = Theme->extrahiere_Einschlussliste();
Ausschlussliste = Theme->extrahiere_Ausschlussliste();

for Gen ( Liste_der_Genobjekte ) {
  Zaehler = 0;
  Unikale_Terme = erzeuge_leereListe();

  Liste_der_Artikelobjekte = Gen->extrahiere_Artikelobjekte();
  for Artikel ( Liste_der_Artikelobjekte ) {
    Liste_der_MeSH-Terme = Artikel->extrahiere_MeSH-Terme();
    for Term ( Liste_der_MeSH-Terme ) {
      if ( Term ist noch nicht Element von Unikale_Terme )
      {
        Term wird in Liste Unikale_Terme aufgenommen
      }

      if
      (
        Term =~ /Wortmuster/i für ein beliebiges Wortmuster
                               aus der Einschlussliste
        UND NICHT
        Term =~ /Wortmuster/i für ein beliebiges Wortmuster
                               aus der Ausschlussliste
      )
      {
        Zaehler = Zaehler + 1;
      }
    }
  }
}

```

**Abbildung 4:** Pseudocode (in reale Syntax der Programmiersprache Perl eingebettete Programmanweisungen in Klartext) der Prozedur zur Ermittlung des stichwortbasierten DbR-Koeffizienten eines positionellen Kandidatengens. Zentrales Element sind die durch das Schlüsselwort *for* eingeleiteten Schleifen (Iterationen) über die Liste der Objekte für jedes Gen (äußerste Schleife), die aus diesem extrahierte Liste repräsentativer Artikel (mittlere Schleife) und die aus diesen Artikeln extrahierte MeSH-Terme (innere Schleife). In der inneren Schleife erfolgt, abhängig vom Ergebnis des *pattern matching* der MeSH-Terme gegen die in einem Objekt vorgehaltenen hypothesenbasierten Stichworte gegebenenfalls die Inkrementierung des Zählers von Formel 3. Zu den Objekten und deren Klassen siehe Text, besonders Tab. 3.

### 1.3 Der kombinierte Datenbank-/Transkriptom-Ansatz: Integration der Datensätze

Die EntrezGeneIDs der positionellen Kandidatengene wurden auf die Annotation des bovinen *GeneChip*<sup>®</sup> (Hintermair, 2007) abgebildet und somit eine Liste boviner Probesets, die positionelle Kandidatengene repräsentieren, erstellt.

Für jedes Probeset wurde dann ermittelt, ob, und wenn ja, in welchen der fünf untersuchten Eutergeweben es mit einer Irrtumswahrscheinlichkeit von  $\leq 5\%$  zwischen den resistenten (Gruppe QQ) und anfälligen (Gruppe qq, zum Design der Studie Kap. II.4.3) differentiell exprimiert ist.

Für die Gewebe mit  $p \leq 0.05$  wurden die Faktoren der Expressionsunterschiede (*fold changes*,  $FC_{sig}$ ) bestimmt:

$$FC_{sig} = \begin{cases} \frac{Exp_{QQ}}{Exp_{qq}} & \text{für } Exp_{QQ} > Exp_{qq} \\ \frac{Exp_{qq}}{Exp_{QQ}} & \text{für } Exp_{QQ} < Exp_{qq} \end{cases} \quad (4)$$

Die hier ermittelten Werte wurden nun benutzt, um aufbauend auf dem heuristischen datenbankbasierten Ansatz einen kombinierten Rangkoeffizienten (Datenbank-/Transkriptom-Rangkoeffizient, DbTR-Koeffizient, DbTRK) für jedes positionelle Kandidatengen zu errechnen (Formel 5).

$$DbTRK = DbRK * \begin{cases} 1 & \text{für } n_{sig} = 0 \\ \sum FC_{sig} & \text{für } n_{sig} > 0 \end{cases} \quad (5)$$

Es sind:

**DbRK:** Der Datenbank-Rangkoeffizient aus Gleichung 3

**$FC_{sig}$ :** Der Faktor des Expressionsunterschiedes *fold change* zwischen der resistenten (QQ) und anfälligen (qq) Gruppe, für den eine Signifikanz bei 5% Irrtumswahrscheinlichkeit festgestellt wurde.

**$n_{sig}$ :** Die Anzahl der Gewebe, in denen für das positionelle Kandidatengen eine differentielle Expression ermittelt wurde.

## 2 Statistische Auswertung experimenteller Daten

### 2.1 Hardy-Weinberg-Gleichgewicht

Es seien A und a die Allele eines gegebenen SNP mit den beobachteten relativen Allelfrequenzen

$$p = \frac{2 * AA + Aa}{2 * (AA + Aa + aa)} \quad \text{und} \quad q = \frac{2 * aa + Aa}{2 * (AA + Aa + aa)} \quad (6)$$

Ein Hardy-Weinberg-Ungleichgewicht liegt vor, wenn durch einen  $\chi^2$ -*goodness-of-fit*-Test (bei  $df = 1$ ) die Nullhypothese, dass die Verteilung der beobachteten Genotypfrequenzen  $O_{AA}$ ,  $O_{Aa}$  und  $O_{aa}$  den erwarteten Genotypwahrscheinlichkeiten  $E_{AA} = p^2$ ,  $E_{Aa} = 2pq$  und  $E_{aa} = q^2$  folgt, zurückgewiesen wird.

### 2.2 Kopplungsungleichgewicht

Das Kopplungsungleichgewicht (LD, *linkage disequilibrium*) zwischen zwei Loci kann auf verschiedene Weise berechnet werden. In dieser Arbeit wird das Maß D' (Lewontin, 1964) verwendet, das in einem zweistufigen Verfahren ermittelt wird.

$$D = P_{AB} - p_A * p_B \quad (7)$$

$$D' = D * \begin{cases} \min(p_A * (1 - p_B), (1 - p_A) * p_B)^{-1} & \text{wenn } D > 0 \\ \max(p_A * p_B, (1 - p_A) * (1 - p_B))^{-1} & \text{wenn } D < 0 \end{cases} \quad (8)$$

$p_A$ : Die relative Frequenz des häufigeren Allels am Locus 1, A

$p_B$ : Die relative Frequenz des häufigeren Allels am Locus 2, B

$P_{AB}$ : Die relative Frequenz der Kombination AB

### 2.3 Euklidische Distanz

Die euklidische Distanz D zwischen zwei Datenreihen (Vektoren)  $x$  und  $y$  derselben Dimension  $n$  berechnet sich nach folgender Formel:

$$D = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (9)$$

## 2.4 SNP-Assoziationsuntersuchungen

Die statistische Auswertung der Genotypisierungsdaten erfolgte mithilfe der Software *ASReml* (Gilmour *et al.*, 2006). Diese implementiert besonders effiziente Algorithmen für die Anwendung gemischter linearer Modelle auf große Datensätze, wie sie in der Tiergenetik auftreten, mittels der *Residual Maximum Likelihood* (REML)-Methode (Patterson & Thompson, 1971).

Mittels Perl-Skripten wurden die für die Berechnung jeweils benötigten Phänotyp- und Genotypdaten aus einer MySQL-Datenbank abgerufen und im korrekten Format exportiert. Die Genotypen der Bullen in der jeweils ersten Zeile einer Datei definierten für den entsprechenden SNP den Basislevel (*intercept*) für die Berechnung von Genotypeffekten (s.u.).

Zusätzlich wurden die Stammbaumdatei (*pedigree*, in Form einer dreispaltigen Tabelle) und eine Steuerdatei, die u.a. die Spezifikation des jeweiligen gemischten linearen Modells enthielt, benötigt und bereitgestellt.

Als proprietäre Software lief *ASReml* auf einem separaten Linuxserver, der per *ssh* (*secure shell*) vom Arbeitsplatzrechner gesteuert wurde.

Für jede SNP-Phänotypkombination und jedes Modell wurde ein separater Programmaufbau durchgeführt. Die Steuerung der Programmaufrufe, Sicherung der Ergebnisdateien und spätere Extraktion der relevanten Werte aus den Ergebnisdateien erfolgte durch selbsterstellte Perlskripte.

### Modelle

Die Beziehungen zwischen einem gegebenen Phänotyp und den Genotypen der SNPs wurden mit gemischten linearen Modellen beschrieben, die in Abbildung 5 gezeigt sind. Die SNP-Genotypen wurden dabei als fixe Effekte modelliert. *ASReml* nutzt für die Varianzanalyse fixer Effekte eine  $\chi^2$ -verteilte *Wald*-Statistik, die unter den hier gegebenen Bedingungen einer F-Statistik entspricht. Die Denominator-Freiheitsgrade dieser F-Statistik wurden nach der Methode von Kenward & Roger (1997) bestimmt.

Singuläres Modell

$$y_i = \mu + G_i + S_i + \varepsilon_i \quad (10)$$

Additives Modell

$$y_i = \mu + G_{1i} + G_{2i} + S_i + \varepsilon_i \quad (11)$$

- $y_i$  Der Phänotyp des Bullen i, angegeben als DYD und gewichtet durch die EDC  
 $\mu$  Der gemeinsame Mittelwert (*grand mean*)  
 $G_i$  Der Genotyp des Bullen i am SNP<sup>a</sup>  
 $G_{1i}$  Der Genotyp des Bullen i am SNP 1<sup>b</sup>  
 $G_{2i}$  Der Genotyp des Bullen i am SNP 2<sup>c</sup>  
 $S_i$  Ein zufälliger polygener Effekt, determiniert durch den Stammbaum des Bullen i  
 $\varepsilon_i$  Der Restfehler des Bullen i

---


$$\text{DYD !WT EDC} \sim \begin{cases} \mu \text{ SNP !r animal} & \Leftarrow \text{singuläres Modell (nach Formel 10)} \\ \mu \text{ SNP1 SNP2 !r animal} & \Leftarrow \text{additives Modell (11)} \end{cases}$$

---

<sup>a</sup>nur in Modell 10
<sup>b</sup>Modell 11<sup>c</sup>Modell 11

**Abbildung 5:** Für die Analyse der Assoziation von SNPs in ausgewählten Kandidatengen mit Merkmalen der Eutergesundheit, insbesondere dem SCS in den ersten drei Laktationen, in einer DH-Halbgeschwisterpopulation. Es wurden jeweils zwei SNPs in einem Kandidatengen typisiert. Für jedes SNP-Paar wurden die Effekte einzeln und additiv berechnet.

## Covarianzstruktur

Die Covarianzstruktur der typisierten Population und damit der zufällige polygenetische Effekt wurde durch den Pedigree definiert (Kap. II.4.4). Dieser wurde von ASReml eingelesen, das aus diesen Daten eine inverse Verwandtschaftsmatrix errechnete. Dem Faktor Tieridentifikationsnummer (im Modell: *animal*) wurde das Label *!P* zugewiesen und damit die inverse Verwandtschaftsmatrix mit diesem Faktor assoziiert.

## Varianzdekomposition

Um die Signifikanz der einzelnen fixen Effekte im additiven Modell (mit mehreren fixen Effekten, nach Formel 11) zu bestimmen, wurde für jeden fixen Effekt eine *Wald*-Statistik nach der konditionalen Methode errechnet (vgl. Gilmour *et al.* (2006), S. 20ff.). Bei dieser wird für jeden fixen Effekt ein vollständiges Modell mit allen Effekten verglichen mit einem reduzierten Modell ohne den jeweiligen fixen Effekt. Die Signifikanz des fixen Effektes bestimmt sich dann aus der durch das vollständige Modell erzielten *sum of squares reduction*, also der zusätzlichen erklärten Varianz gegenüber dem reduzierten Modell. Die Reihenfolge der fixen Effekte spielt im Gegensatz zur inkrementellen Varianzdekomposition dabei keine Rolle.

Der additive genetische Effekt eines SNP auf ein gegebenes Merkmal wird aus den Allelsubstitutionseffekten der homozygoten Genotypen berechnet:

$$\text{Add.GE} = \frac{\Delta \text{Effekt}_{\text{Homozygot A}}, \text{Effekt}_{\text{Homozygot B}}}{2} \quad (12)$$



### 3 Verwendete Software

Der folgende Abschnitt gibt einen Überblick über die in dieser Arbeit verwendete Software.

#### 3.1 Perl

Perl ist eine freie, plattformunabhängige und interpretierte Skriptsprache, die vom Linguisten Larry Wall entworfen, 1987 in einer ersten Version veröffentlicht und seither von einer großen Unterstützergemeinde ständig weiterentwickelt wurde. Ihre Stärken liegen unter anderem in der effizienten Verarbeitung großer Datenmengen und der hervorragenden Integration Regulärer Ausdrücke. Mit Version 5 von Perl wurden Referenzen und damit die Möglichkeit der Realisierung komplexer Datenstrukturen eingeführt; sie sind neben *packages* (Namensräumen) auch eine der Säulen der objektorientierten Programmierung. Die im Rahmen dieser Arbeit eingesetzten Skripte, die von den genannten Vorzügen vielfältigen Gebrauch machen, basieren auf Version 5.8 von Perl.

Das *Comprehensive Perl Archive Network* (CPAN<sup>16</sup>) stellt eine sehr große Auswahl (zur Zeit mehr als 20000) an Zusatzpaketen (Module) bereit. Eine auf Anwendungen in der Bioinformatik spezialisierte und aufeinander abgestimmte Sammlung von Modulen ist BioPerl (Stajich *et al.*, 2002).

Tabelle 33 (im Anhang) führt einige Zusatzmodule auf, die in den Skripten, welche für die hier vorgestellte Arbeit entwickelt wurden, Verwendung fanden.

#### 3.2 R

Sehr viele der hier vorgestellten statistischen Berechnungen und Abbildungen wurden mit der statistischen Programmiersprache R, dem *project for statistical computing* (R Development Core Team, 2008), durchgeführt bzw. erzeugt. Dieses *Open Source* Projekt wurde von Ross Ihaka und Robert Gentleman an der Universität von Auckland auf der Basis des proprietären Systems S und mit Anleihen an die Sprache Scheme initiiert und wird vom *R Development Core Team* kontinuierlich weiterentwickelt. Es ist durch eine große Menge an Zusatzpaketen, deren Erstellung durch das objektorientierte Konzept von R begünstigt wird, für die verschiedensten Aufgaben erweiterbar; ein zentrales Paketarchiv steht mit dem Comprehensive R Archive Network (CRAN<sup>17</sup>) zur Verfügung. Auf bioinformatische Aufgaben spezialisierte Pakete werden durch das Projekt Bioconductor<sup>18</sup> (Gentleman *et al.*, 2004) zusammengefasst.

Der Vielseitigkeit und freien Verfügbarkeit steht eine kontraintuitive Bedienung über eine Textkonsole gegenüber; jedoch gibt es eine Reihe graphischer Oberflächen, beispielsweise *RStudio* und *Tinn-R*.

---

<sup>16</sup><http://www.cpan.org>

<sup>17</sup><http://www.cran.r-project.org/>

<sup>18</sup><http://www.bioconductor.org/>

### 3.3 Emboss

Die *European Molecular Biology Open Software Suite* (EMBOSS<sup>19</sup>, Rice *et al.*, 2000) ist eine frei verfügbare Sammlung von hauptsächlich kommandozeilengesteuerten Werkzeugen. Ein Großteil von ihnen ermöglicht die Analyse und Darstellung von Nukleotid- oder Proteinsequenzen unter den verschiedensten Aspekten, wobei jedes Einzelprogramm für eine spezifische Aufgabe konzipiert ist; gemeinsame Dateiformate ermöglichen den problemlosen Austausch von Daten.

Beispielsweise wurden die Abbildungen 12, 13 und 15 mit dem EMBOSS-Programm *lindna* erzeugt.

### 3.4 BLAST

Das *Basic Local Alignment Search Tool* (BLAST, Altschul *et al.*, 1997) implementiert Algorithmen zur Identifizierung lokaler Übereinstimmungen zwischen Nukleotid- bzw. Aminosäuresequenzen.

Neben der Nutzung der einzelnen in BLAST enthaltenen Programme über eine Weboberfläche (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) besteht die Möglichkeit der Einrichtung einer lokalen Installation.

---

<sup>19</sup><http://www.sanger.ac.uk/Software/EMBOSS>

## 4 Materialien und Tiere

Die experimentelle Identifizierung potentiell merkmalsrelevanter Sequenzpolymorphismen (SNPs) in den ausgewählten potentiellen funktionellen Kandidatengenomen erfolgte sequentiell in mehreren ausgewählten Populationen der Milchviehrasse Deutsche Holstein.

### 4.1 Bullen mit extrem differenter Verteilung des SCS-Schätzwertes MA-BLUP

Die vergleichende Poolsequenzierung als erste Stufe im Selektionsverfahren wurde an zwei Gruppen von Bullen aus dem DHV-Lückenfüllerprogramm (2002) durchgeführt, die vor Versuchsbeginn an den Mikrosatellitenmarkern *BM7109*, *ILSTS002*, *BMS2639*, *BM2078* und *TGLA227* typisiert worden sind. Für diese Bullen wurde nach der in Kühn *et al.* (2008) beschriebenen Methode der Zuchtwert MA-BLUP<sup>20</sup> (Fernando & Grossman, 1989) für das Merkmal Zellzahl berechnet, wobei der Mikrosatellitenmarker *TGLA227* zugrunde gelegt wurde (Stand 05/2005). Unter der Bedingung, dass jede Gruppe maximal vier Söhne eines gemeinsamen Bullenvaters enthalten durfte, wurden auf Grundlage dieser Berechnung je 15 Bullen mit extrem niedrigem bzw. extrem hohem MA-BLUP-Schätzwert in die entsprechende Testgruppe eingeteilt.

Neben dem MA-BLUP ist für diese Bullen auch der Relative Zuchtwert Zellzahl (RZS, Kap. I.1.4) bekannt.

Die Verteilung beider Zuchtwerte in den Testgruppen ist in Abbildung 6 graphisch dargestellt.

### 4.2 Bullen mit Heterozygosie am BTA18-SCS-QTL

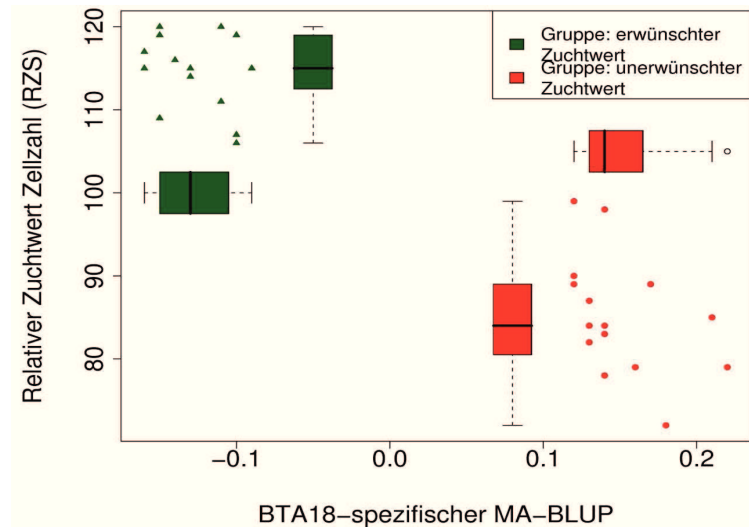
Weiterhin standen für die Verifikation potentiell merkmalsassoziierter SNPs vier ausgewählte DH-Bullen zur Verfügung, die vor Versuchsbeginn am BTA18-telomeren Mikrosatellitenmarker *TGLA227*, welcher als mit dem BTA18-SCS-QTL gekoppelt angenommen wurde, heterozygot typisiert worden sind. Diese Bullen sollen hier mit den Buchstaben  $\alpha$ ,  $\beta$ ,  $\gamma$  und  $\delta$  bezeichnet werden.

### 4.3 Auswahl der Stichprobe für die Transkriptomstudie des Verbundprojektes M.A.S.-Net

Das Verbundprojekt M.A.S.-Net innerhalb der Förderinitiative FUGATO des BMBF beinhaltete eine vergleichende Transkriptomstudie an zwei Gruppen von Färsen der Milchviehrasse Deutsche Holstein (DH), die sich durch eine stark differentielle genetische Prädisposition für die Resistenz gegen Euterinfektionen auszeichneten. Diese Stichprobe wurde unter den

---

<sup>20</sup>Marker Assisted Best Linear Unbiased Prediction



**Abbildung 6:** Scatterplot des gesamtgenomischen Zellzahlzuchtwerthes RZS gegen den den BTA18-SCS-QTL berücksichtigenden Zuchtwert MA-BLUP, sowie Boxplots, für 30 DH-Bullen in zwei Gruppen, die in der selektiven Poolsequenzierung, Kap. II.4.1, eingesetzt wurden. MA-BLUP und RZS sind negativ korreliert (siehe Text); die Bullengruppe mit den in der Zucht auf verbesserte Eutergesundheit erwünschten, wesentlich niedrigeren Werten für MA-BLUP weist deutlich höhere Werte für den RZS aus.

Töchtern von drei Bullen der DH-Gesamtpopulation rekrutiert, bei denen eine große Differenz zwischen den SCS-MA-BLUP-Zuchtwerten (s.o.) ihrer beiden Haplotypen an den Mikrosatellitenmarkern im Bereich des BTA18-SCS-QTL (Xu *et al.*, 2006) vermuten lässt, dass diese Haplotypen mit verschiedenen Allelen des QTL (Allel Q vermittelt einen niedrigen, Allel q einen hohen SCS), gekoppelt sind. Für eine Kohorte von Färsen<sup>21</sup> unter diesen Töchter wurden die wahrscheinlichsten paternal vererbten Haplotypen und damit die wahrscheinlichsten paternalen QTL-Allele bestimmt. In die Studie aufgenommen wurden 6 Färsen in der Gruppe **QQ** mit paternal vererbtem Q-Allel und 5 Färsen in der Gruppe **qq** mit paternal vererbtem q-Allel, die als weitere Bedingung von Müttern abstammen, deren Markergenotypen sowie eigene und paternale hohe bzw. niedrige Werte für den RZS mit hoher Wahrscheinlichkeit auf Homozygotie der Tiere für das Q- bzw. q-Allel am BTA18-SCS-QTL hinweisen.

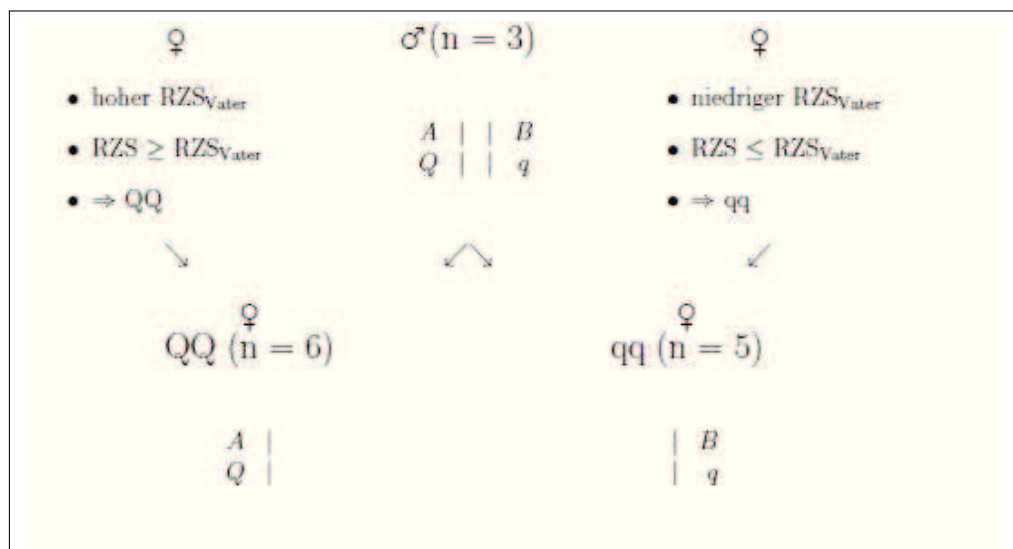
Tabelle 8 schlüsselt auf, mit wie vielen Töchtern die drei Bullen in den beiden Testgruppen jeweils vertreten waren. Darüber hinaus gibt sie die Mittelwerte und Standardabweichungen für den Parameter RZS in beiden Gruppen an.

Das Selektionsverfahren wird umfassend in Kühn *et al.* (2008) dargestellt (vgl. Abb. 7).

<sup>21</sup>geboren zwischen Februar und September 2003

**Tabelle 8:** Merkmale der in der FUGATO M.A.S.-Net-Transkriptomstudie untersuchten Färsengruppen QQ und qq. Die Bezeichnungen QQ bzw. qq stehen für Homozygotie von Allelen des BTA18-SCS-QTL, die mit entgegengesetzten Extremen der genetischen Prädisposition für Euterinfektionen assoziiert sind. Alle Färsen sind direkte Nachkommen von drei DH-Zuchtbullen, die mit einer unterschiedlichen Zahl von Töchtern in den Gruppen vertreten sind. Die Verteilung des Merkmales RZS, hier dargestellt durch die Parameter Mittelwert  $\mu$  und Standardabweichung  $\sigma$ , unterscheidet sich stark zwischen beiden Gruppen.

Gruppe	Töchter des Bullen			RZS $\mu \pm \sigma$
	A	B	C	
QQ (hohe Resistenz)	3	2	1	107,4 $\pm$ 6,17
qq (niedrige Resistenz)	3	1	1	94 $\pm$ 5,11



**Abbildung 7:** In der vergleichenden Transkriptomstudie des Verbundprojektes M.A.S.-Net wurden die DH-Färsengruppen QQ und qq, die entgegengesetzte Extreme der genetischen Prädisposition für Euterinfektionen repräsentieren, vergleichend untersucht. QQ enthält 6 Färsen, die von ihren Vätern das mit dem Markerhaplotyp A gekoppelte, züchterisch erwünschte, Allel Q des BTA18-SCS-QTL geerbt haben und qq deren paternale Halbschwester (n=5), die das mit dem Markerhaplotyp B gekoppelte, züchterisch unerwünschte, Allel q geerbt haben. Nur Töchter von Müttern mit höchstwahrscheinlich homozygotem QTL-Genotyp wurden berücksichtigt.

## 4.4 Die für die Assoziationsuntersuchungen genutzte DH-Halbgeschwisterpopulation

### Familienstruktur und Stammbaum

Für die Assoziationsuntersuchungen wurde die von Baes *et al.* (2009) beschriebene Population genutzt. Es handelte sich um sechs paternale Halbgeschwisterfamilien der Rasse Deutsche Holstein, die ihren Ursprung in den Halbgeschwisterdesigns (*granddaughter designs*) hatten, welche von Kühn *et al.* (2003) und Xu *et al.* (2006) für QTL-Kartierungen eingesetzt wurden. Die für die selektive Poolsequenzierung eingesetzten 30 Bullen (vgl. Kap. II.4.1) waren nicht Bestandteil dieser Population. Einen Überblick über die paternalen Verwandtschaftsverhältnisse gibt Tabelle 37 (im Anhang).

Der Stammbaum der genotypisierten Bullen umfasste insgesamt 7627 Datensätze aus jeweils einem Tier, dessen Vater und Mutter, alle über ihre Ohrmarke identifiziert, davon 5322 vollständige und 2305 unvollständige (d.h. eines oder beide Elterntiere sind nicht erfasst). Ausgehend von den typisierten Tieren sind bis zu fünf Vorgängergenerationen im Stammbaum vertreten; für mehr als 95% der 919 Bullen, für die verwertbare Typisierungsdaten gewonnen wurden, ist der Stammbaum bis zur dritten Parentalgeneration (Urgroßeltern), für mehr als 75% sogar bis zur vierten Parentalgeneration komplett.

### Phänotypdaten

Für die Bullen der DH-Halbgeschwisterpopulation (DH: Deutsche Holstein) stand ein vom VIT<sup>22</sup> bereitgestellter Phänotypendatensatz zur Verfügung. Dieser bestand jeweils aus Wertepaaren für die mittlere Töchterabweichung (*Daughter Yield Deviation*, DYD) und den effektiven Töchterbeitrag (*Effective Daughter Contribution*, EDC), für das Merkmal SCS in den ersten drei Laktationen sowie für 19 Konformations- und Milchleistungsphänotypen.

Die DYD eines Bullen für einen gegebenen Phänotyp ist definitionsgemäß der Durchschnitt der Leistung seiner Töchter, korrigiert für fixe und nichtgenetische zufällige Effekte der Töchter und genetische Effekte der angepaarten Kühe. Die hier verwendeten Werte wurden nach einem Testtagsmodell berechnet; zu den Details siehe Liu *et al.* (2004). Bei der EDC handelt es sich um ein Maß für den Umfang der Datenbasis, die für die Zuchtwertberechnung eines Bullen zur Verfügung steht. Der Wert kann die Zahl der Töchter des Bullen übersteigen, insbesondere, wenn mehrere korrelierte Merkmale betrachtet werden.

Neben den Parametern arithmetisches Mittel und Standardabweichung sind für die Phänotypen in Tabelle 36 (im Anhang) auch der Residuenstandardfehler des linearen Modells in Gleichung 13 angegeben.

---

<sup>22</sup>Vereinigte Informationssysteme Tierzucht, Verden, siehe Kap. I.1.4

$$DYD_i = \mu + Bullenwater_i + \varepsilon_i \quad (13)$$

Tabelle 40 (im Anhang) zeigt schematische Darstellungen einiger der berücksichtigten Konformationsphänotypen.

## 5 Labormethoden

### 5.1 Polymerasekettenreaktion

Für die Amplifikation genomischer DNA als Voraussetzung für die Sequenzierung wurde das Verfahren der Polymerasekettenreaktion (*polymerase chain reaction, PCR*) eingesetzt.

**Primerdesign** Als Ausgangspunkt für das Design genomischer Primer für die Amplifikation und Sequenzierung der ausgewählten Kandidatengene wurden die in Tabelle 19 gelisteten genomischen, auf der Version 3.1 des bovinen Genoms beruhenden Nukleotidsequenzen der Gene verwendet. Durchgeführt wurde das Primerdesign mithilfe der Onlinesoftware *Primer3* (Rozen & Skaletsky, 2000).

Tabelle 34 (im Anhang) zeigt die Nukleotidsequenzen der genomischen Primer für die Amplikons, welche die in dieser Arbeit gefundenen SNPs enthalten.

**Reaktionsansatz** Die in Tabelle 9 aufgeführten Reagenzien wurden in einem auf Eis gekühlten 200  $\mu$ l-Eppendorfgefäß gemischt. Für die folgende Amplifikation wurde das Verfahren der Touchdown-PCR eingesetzt, bei der in den ersten Zyklen der Reaktion eine höhere Temperatur für den Primeranlagerungsschritt (*Annealing*) eingestellt wird, als durch die Schmelztemperatur  $T_m$  der Primer vorgegeben, und diese schrittweise verringert wird. Dadurch wird eine höhere Spezifität der PCR erzielt (Korbie & Mattick, 2008).

Die PC-Reaktion erfolgte in einem Thermocycler nach dem in Tabelle 10 gezeigten Protokoll.

### 5.2 Agarosegelelektrophorese

Zu Kontroll- und Aufreinigungszwecken wurden PCR-Produkte auf 1-2 %ige Agarosegele mit TAE als Elektrophoresepuffer aufgetragen, wobei die Auftragsmenge jeweils 5-20  $\mu$ l betrug. Die Identifizierung der DNA unter UV-Licht wurde durch Zugabe von Ethidiumbromid (Endkonzentration 0,4  $\mu$ g/ml) ermöglicht, welches der Agarosesuspension vor dem Gießen des Gels zugesetzt wurde. Als Längenstandard diente *pBR322 AluI*.

**Tabelle 9:** Genomische DNA wurde sequenziert, um SNPs zu identifizieren und deren Allelverteilung zu bestimmen. Die Tabelle zeigt die Zusammensetzung der Reaktionsansätze, die für die Generierung der hierzu benötigten genomischen Amplikons mittels Polymerase-Kettenreaktion (PCR) eingesetzt wurden.

Komponente	Endkonzentration	Volumen ( $\mu\text{l}$ )
PCR-Puffer, 10fach	1fach	2,5
dATP	0,8 mM	0,5
dCTP	0,8 mM	0,5
dGTP	0,8 mM	0,5
dTTP	0,8 mM	0,5
5'-Primer	250 nM	2,5
3'-Primer	250 nM	2,5
dH <sub>2</sub> O		12
Taq-Polymerase	0,025 U/ $\mu\text{l}$	2,5
genom. DNA	1 ng/ $\mu\text{l}$	1
		25

**Tabelle 10:** Thermocyclerprogramm für die Amplifikation genomischer DNA mit der Methode der TouchDown-Polymerasekettenreaktion (TouchDown-PCR);  $-0,5\text{ }^\circ\text{C}/\text{Zyklus}$ .

Temperatur (in $^\circ\text{C}$ )	Dauer	Vorgang	Zyklen
94	30 sec	Denaturierung	
94	30 sec	Denaturierung	
$T_m + 5$ bis $T_m^a$	30 sec	Annealing	11
72	1 min	Elongation	
96	30 sec	Denaturierung	
$T_m$	30 sec	Annealing	34
72	1 min	Elongation	
72	9 min	finale Elongation	
4	$\infty$		

<sup>a</sup> $\Delta T = -0,5^\circ\text{C}/\text{Zyklus}$



### 5.3 Bestimmung der DNA-Konzentration

Die Bestimmung der Konzentration von DNA-Lösungen erfolgte an einem Photometer des Typs NanoDrop<sup>®</sup> ND1000 nach Vorgabe des Herstellers (Fa. ThermoFisher Scientific).

### 5.4 DNA-Sequenzierung nach Sanger

Die Identifizierung von DNA-Sequenzpolymorphismen erfolgte durch vergleichende Sequenzierung mithilfe der durch F. Sanger entwickelten Strangabbruchmethode (Sanger *et al.*, 1977).

Für die Sequenzierung der PCR-Produkte wurden 5 ng des PCR-Produktes pro 100 bp Länge benötigt (z.B. 25 ng DNA für ein PCR-Fragment von 500 bp). Ein dem entsprechendes Volumen der DNA-Lösung wurde mit ddH<sub>2</sub>O auf ein Volumen von 8  $\mu$ l ergänzt. Dazu wurden 0,5  $\mu$ l einer 10 mM-Lösung des 5'- bzw. 3'-PCR-Primers, mit dem das jeweilige Fragment initial amplifiziert wurde, und 1,5  $\mu$ l des BigDye-Terminator<sup>®</sup>-Kits (Applied Biosystems) gegeben. Amplifizierung und Einbau der Didesoxynukleotide fanden nach dem Protokoll in Tabelle 11 statt:

**Tabelle 11:** Thermocyclerprogramm für die Sequenzierung von DNA-Fragmenten nach der Sanger-Methode (Sanger *et al.*, 1977).

Temperatur (in °C)	Dauer	Schritt	Zyklen
96	1 min	Denaturierung	
96	10 sec	Denaturierung	
$T_m$	5 sec	Annealing	25
60	4 min	Elongation	
6	$\infty$		

Der Reaktionsansatz wurde anschließend zusammen mit 15  $\mu$ l dH<sub>2</sub>O auf einer Sephadex-Trennsäule (Centri-Sep<sup>™</sup>) gereinigt und 4  $\mu$ l des Produktes für die Sequenzierung eingesetzt. Diese erfolgte durch Kapillarelektrophorese in einem Gerät der Marke „AbiPrism 3130 Genetic Analyzer“ (Fa. Applied Biosystems). Die Software *Sequencing Analysis 5.3*, ebenfalls von Applied Biosystems, wurde für die Generierung der Sequenzchromatogramme eingesetzt; diese standen als Chromatogrammdateien (Dateiendung *.ab1* für die weitere Auswertung zur Verfügung).

Aus diesen Chromatogrammdateien wurden mithilfe des EMBOSS-Programms *abiview* die Chromatogramme als Bilddateien und die Sequenzinformationen als Textdateien im FASTA-Format extrahiert. Die Untersuchung auf SNPs wurde durch visuelle Inspektion der

Chromatogrammbilder und lokales Sequenzalignment mittels des Programms *bl2seq*, einem Bestandteil von BLAST (Altschul *et al.*, 1997), durchgeführt.

## 5.5 Pyrosequenzierung<sup>TM</sup>

Für die Genotypisierung der für die Assoziationsstudie selektierten SNPs mittels des Verfahrens der Pyrosequenzierung (Ronaghi *et al.*, 1996) wurden die in Tabelle 35 (im Anhang) gezeigten Primer eingesetzt. Die Durchführung der Experimente wurde im übrigen durch Mitarbeiter übernommen, denen an dieser Stelle Dank ausgesprochen werden soll.

## Teil III

# Ergebnisse

## 1 Identifizierung von potentiellen funktionellen positionellen Kandidatengenem im BTA18-SCS-QTL mittels eines heuristischen bioinformatischen Ansatzes

### 1.1 Identifizierung positioneller Kandidatengene des BTA18-SCS-QTL mittels der vergleichenden Genkarten

Das Verbundprojekt M.A.S.-Net innerhalb der BMBF-geförderten Forschungsinitiative FUGATO<sup>23</sup> realisierte eine Feinkartierung des BTA18-SCS-QTL auf Basis einer kombinierten LA/LD-Methode (Meuwissen & Goddard, 2000), mit der der QTL auf zwei Markerintervalle mit der höchsten Wahrscheinlichkeit für das Auftreten des kausalen Sequenzpolymorphismus eingegrenzt werden konnte (Baes *et al.*, 2009).

Für die begrenzenden Marker der Intervalle 5 (A) und 9 (B) einer intermediären Kartierung (Abb. 2) wurden in der bovinen Genkarte (Build 3.1) die jeweils proximalen Gensymbole gesucht, diese als flankierende Gene der QTL-syntenen humanen Genregion interpretiert und in die humane Genkarte (Build 36.3, HSA19) eingesetzt. Die so flankierten humanen Gensymbole wurden anschließend auf EntrezGeneIDs abgebildet; dies erfolgte mittels der in *R* durch die Bibliothek *biomaRt* (Durinck *et al.*, 2005) implementierten Schnittstelle zum Datenbankservice BioMart (Kasprzyk *et al.*, 2004; Smedley *et al.*, 2009) und der damit hergestellten Anbindung an die (humane) Ensembl-Genbank (Hubbard *et al.*, 2009).

Durch diese Vorgehensweise wurde für die QTL-Intervalle eine gemeinsame Liste von 254 über ihre unikale EntrezGeneID identifizierten positionellen humanen Kandidatengenem generiert. Tabelle 12 zeigt die begrenzenden Marker der bovinen Intervalle, die Grenzen der korrespondierenden humanen Region(en) und jeweils die Zahl der dort lokalisierten humanen Gene.

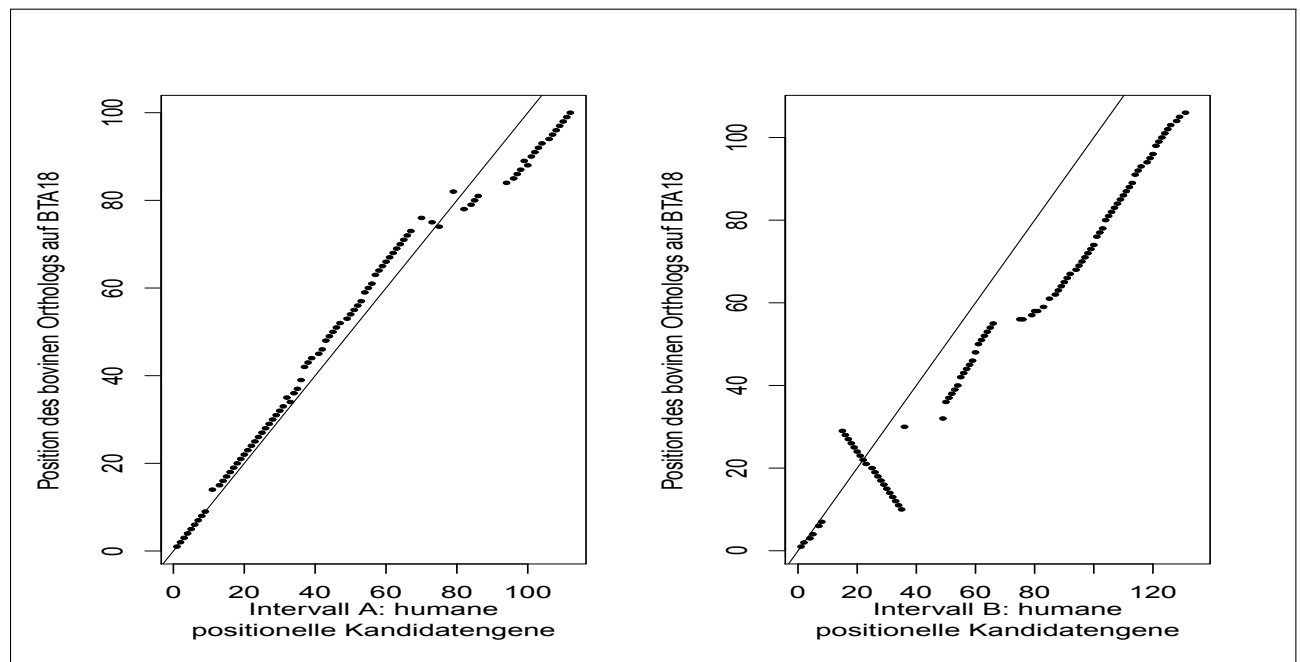
Wie in Abbildung 8 gezeigt, ist für die beiden Intervalle eine weitgehende, in Intervall A jedoch deutlich ausgeprägtere Konservierung der Gene und ihrer Positionen innerhalb der Genreihenfolge zwischen HSA19 und BTA18 gegeben. Bedeutsame Ausnahmen sind eine größere Inversion im Bereich von Intervall B sowie mehrere Instanzen aufeinanderfolgender Gene, die auf HSA19, nicht aber auf BTA18 vorhanden sind.

---

<sup>23</sup>Funktionelle Genomanalyse im tierischen Organismus, <http://www.fugato-forschung.de>

**Tabelle 12:** Für den BTA18-SCS-QTL wurden mittels einer LA/LD-Feinkartierung zwei Kandidatenintervalle, 5 (A) und 9(B), ermittelt. Die Tabelle gibt die Namen und Positionen der Marker, die diese Intervalle begrenzen, in der bovinen genetischen Karte an. Zusätzlich sind die Gensymbole derjenigen Gene, die in der intervall-syntenen humanen Region auf HSA19q besagten Markern per Sequenzvergleich räumlich zugeordnet werden können, nebst ihren genomischen Positionen angegeben. Die Union der beiden Listen der jeweils zwischen diesen Genen lokalisierten humanen Gene konstituiert die im weiteren untersuchte Liste positioneller Kandidatengene des BTA18-SCS-QTL.

Typ	Lage	Bezeichner	Position
Intervall A (113 EntrezGeneIDs)			
Marker	proximal	BMON117	71,1 cM
	distal	DIK4672	73,1 cM
Gen	proximal	<i>PEPD</i>	38,57 Mb
	distal	<i>CAPN12</i>	43,91 Mb
Intervall B (141 EntrezGeneIDs)			
Marker	proximal	BB710	77.8 cM
	distal	DIK3014	82.1 cM
Gen	proximal	<i>HNRPUL1</i>	46,46 Mb
	distal	<i>IGFL2</i>	51,34 Mb



**Abbildung 8:** In den Dotplots ist für die zu den QTL-Intervallen A (links) und B (rechts) syntenen Regionen von HSA19 auf der X-Achse für jedes Gen die Position in der Rangfolge der Gene und auf der Y-Achse die Position seines in Ensembl annotierten bovinen Homologs (sofern vorhanden) innerhalb der Genreihenfolge auf BTA18 aufgetragen. Berücksichtigt wurden nur Gene, die mit einer EntrezGeneID annotiert sind. In Intervall B ist eine größere Inversion evident. Die durchgezogenen Geraden symbolisieren eine ideale 1:1-Konservierung der Gene.

## 1.2 Die Anwendung des heuristischen Datenbankansatzes zur Identifizierung potentieller funktioneller Kandidatengene

Die Größenordnung von mehr als 250 positionellen Kandidatengen, die gemäß vorigem Abschnitt für den BTA18-SCS-QTL gefunden wurden, erwies sich für eine experimentelle Bearbeitung als zu umfangreich. Der in Kap. II.1.1 beschriebene heuristische stichwort- und datenbankbasierte Ansatz wurde angewendet, um zunächst auf Grundlage der verfügbaren einschlägigen wissenschaftlichen Referenzpublikationen einen genweisen Datenbankrangkoeffizienten (DbRK) zu ermitteln, auf Grundlage dessen die positionellen Kandidatengene in eine Rangliste potentieller funktioneller Kandidaten eingeordnet werden konnten.

### 1.2.1 Ermittlung und Beschaffung (*Retrieval*) relevanter Literatur

Mittels der (skriptbasierten) Auswertung der humanen Sektion der Datenbank Gene2PubMed wurden zunächst den positionellen Kandidatengen eine Liste von Referenzpublikationen zugeordnet (Kap. II.1.1). Die minimale Größe der einem Gen zugewiesenen Liste betrug 1, das Maximum 1871 Publikation(en). Tabelle 13 gibt Lageparameter der diesbezüglichen Verteilung an: über die gesamte Stichprobe aller 254 positionellen Kandidatengene sowie separiert nach dem Kriterium, ob für ein gegebenes Gen im weiteren Verlauf ein DbR-Koeffizient  $M > 0$  ermittelt werden konnte ( $n = 129$ ), oder nicht ( $n = 125$ ).

Die Diskrepanz zwischen Median und arithmetischem Mittel über die gesamte Stichprobe weist auf eine schiefe Verteilung, die wenige Genen mit sehr vielen annotierten Referenzpublikationen aufweist, hin. Es zeigte sich, dass dies fast ausschließlich auf eine starke Ungleichverteilung in der Gruppe der Gene mit ermittelbarem DbR-Koeffizienten zurückgeht, während für die Gegengruppe *a priori* relativ gleichmäßig, und (verifiziert durch einen verteilungsfreien Mann-Whitney-Wilcoxon-Rangsummentest) signifikant weniger Referenzpublikationen für eine Auswertung zur Verfügung standen (Tab. 13).

**Tabelle 13:** Die Anzahl der annotierten Referenzpublikationen, die für die insgesamt 254 über ihre EntrezGeneID identifizierten positionellen Kandidatengene in der Datenbank Gene2PubMed gefunden wurden, für alle positionellen Kandidatengene bzw. getrennt nach dem Kriterium, ob für die Gene im weiteren Verlauf ein Datenbank-Rangkoeffizient (DbRK) von  $M > 0$  errechnet werden konnte. Ein Mann-Whitney-Wilcoxon-Rangsummentest (U-Test) zeigt, dass sich dieses Merkmal in den beiden Gruppen hochsignifikant unterscheidet.

	Gene	Referenzpublikationen		U-Test
		arithm. Mittel	Median	
Gesamt	254	29,3	7	
mit DbRK	129	53	15	W=1963,5
ohne DbRK	125	4,8	4	p=1,6*10 <sup>-25</sup>

Mit dem im entsprechenden Methodenabschnitt dargelegten Verfahren, d.h. durch Nutzung der Programmierschnittstelle EUtils, welche den Zugriff auf die Entrez-Datenbanken erlaubt, wurden die selektierten Referenzpublikationen im MEDLINE-Format heruntergeladen.

Aus den heruntergeladenen MEDLINE-Datensätzen (für ein Beispiel dieses Formates siehe Abb. 21 im Anhang) wurden die Terme der MeSH-Annotation als Zeichenketten extrahiert.

Der quantitative Umfang der MeSH-Annotation variiert zwischen den ausgewerteten Referenzpublikationen. Für einige Gene ( $n = 7$ ), konnte kein MeSH-Term extrahiert werden; bei diesen Genen handelt es sich ausnahmslos um solche, für die lediglich eine einzelne Referenzpublikation in der Datenbank Gene2PubMed verzeichnet war.

In Formel 3 zur Berechnung des DbR-Koeffizienten wird die Gesamtzahl der in den Annotationen aller für ein gegebenes positionelles Kandidatengen identifizierten Referenzpublikationen vertretenen MeSH-Terme als Denominator eingesetzt, wobei hier jeder MeSH-Term nur ein einziges Mal gezählt wird.

Diese Zahl einzigartiger (unikaler) MeSH-Terme für ein gegebenes Kandidatengen ist in der hier untersuchten Menge der 247 Kandidatengene, für welche MeSH-Terme extrahiert wurden, sehr hoch korreliert (Produkt-Moment-Korrelation nach Pearson,  $r^2 \approx 0.97$ ) mit der Gesamtzahl der analysierten Referenzpublikationen, die mit ihnen annotiert wurden.

### 1.2.2 Die Berechnung des datenbankbasierten Rangkoeffizienten auf Grundlage der ermittelten Referenzliteratur

Die MeSH-Annotationen der für jedes positionelle Kandidatengen ermittelten Publikationen wurden mit einer Liste hypothesebasierter Stichworte nach der auf S. 38 gezeigten Prozedur abgeglichen, woraus nach Formel 3 (Kap. II.1.2.3) der DbR-Koeffizient des jeweiligen Gens ermittelt wurde. Die Stichworte repräsentieren gemäß der zugrundeliegenden Hypothese verschiedene Konzepte der angeborenen und adaptiven Immunabwehr. Zunächst mussten die gewählten Stichworte in reguläre Ausdrücke umgewandelt werden (Kap. II.1.2.2); diese sind in Tabelle 14 gezeigt. Unterschiede in der Groß- und Kleinschreibung zwischen den MeSH-Termen und der Stichwortliste wurden beim Vergleich ignoriert.

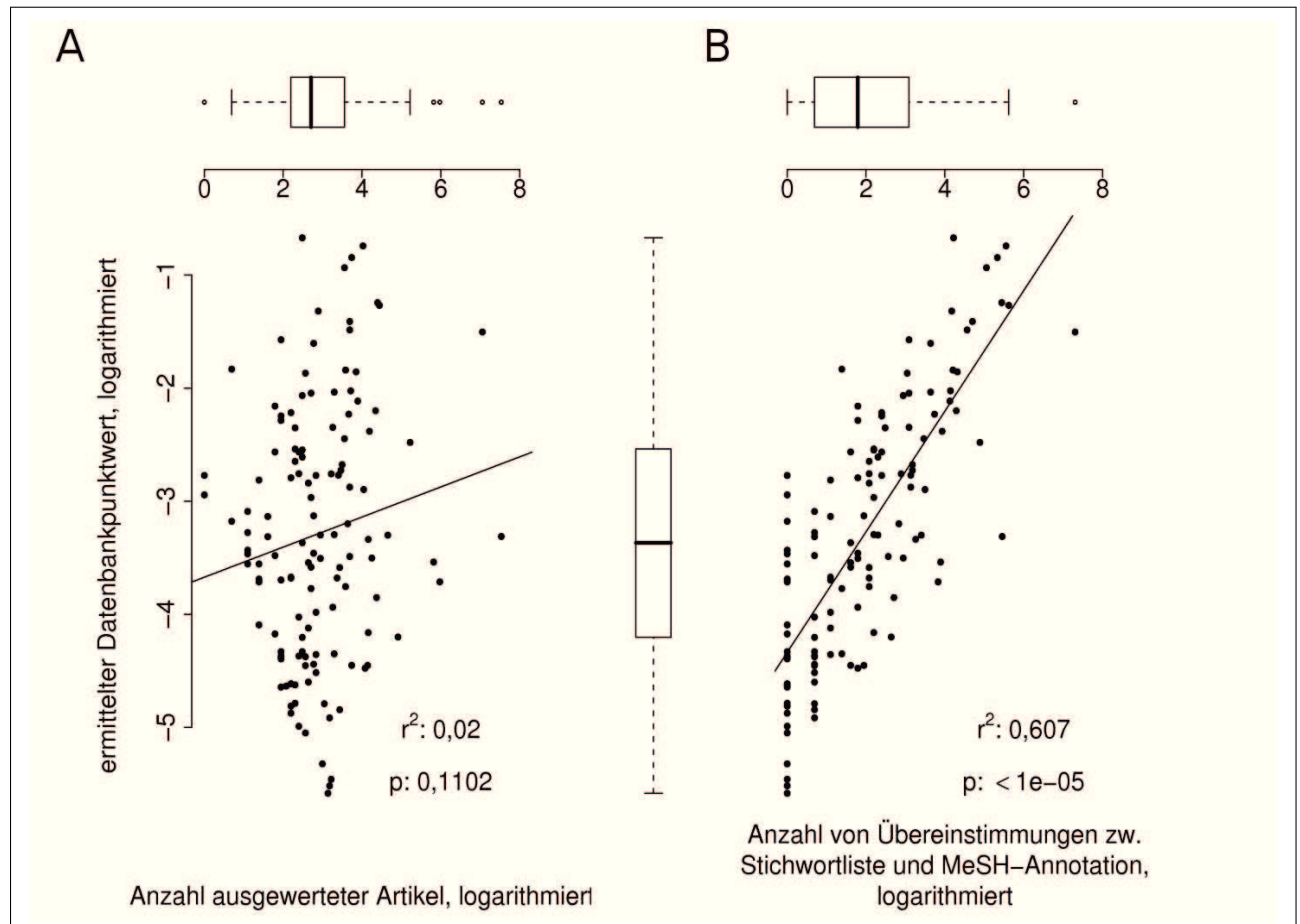
**Tabelle 14:** Liste der für die Beschreibung der immunologischen Hypothese verwendeten Stichworte in Form einfacher Regulärer Ausdrücke, die von MeSH-Termen abgeleitet sind. Die Bedeutungen der mit einem \ eingeleiteten Metasymbole sind Tabelle 7 zu entnehmen.

---

• \bB-Cell	• cd\d+	• immun	• Mast-cell
• \bcomplement\b	• Chemokin	• infection	• Mhc
• \bHla	• Cytokin	• Inflamm	• Neutrophil
• \bT-Cell	• Dendritic	• Interferon	• Nf-kappa\s b
• Anergy	• Eosinophil	• Interleukin	• Peptidoglycan
• Anti-bacterial	• Epitope	• Killer\scell	• Phagocyt
• Antigen	• Fc\	• Leukocyte	• Regulation-\immunology
• B-Lymphocyt	• Germinal\scenter	• Lymph	• stat\d
• Bone\s marrow	• Granulocyte	• Lysosom	• T-Lymphocyte
• Cd\	• Histocompatib	• Macrophage	• Toll-like

---

Es konnte keine Korrelation zwischen der Anzahl der ausgewerteten Artikel und dem berechneten DbR-Koeffizient für die 129 Gene, für die ein solcher Wert errechnet werden konnte, festgestellt werden (Abb. 9A). Dagegen ist die Anzahl der ermittelten Übereinstimmungen zwischen der Stichwortliste und den MeSH-Annotationen der ausgewerteten Publikationen mit dem DbR-Koeffizient korreliert; sie zeichnet für etwa 60% der Gesamtvarianz des DbR-Koeffizienten verantwortlich ( $r^2 = 0,6$ , Abb. 9B). Tabelle 15 zeigt die 15 nach dem DbR-Koeffizienten bestbewerteten positionellen Kandidatengene, also die nach diesem Kriterium wahrscheinlichsten potentiellen funktionellen Kandidatengene im BTA18-SCS-QTL.



**Abbildung 9:** Die Abhängigkeit des Datenbankrangkoeffizienten für 129 positionelle Kandidatengene, für welche ein solcher ermittelt werden konnte, von zwei Parametern: **A)** der Anzahl der für ein gegebenes Gen ausgewerteten Referenzpublikationen, bzw. **B)** der Gesamtzahl von MeSH-Annotationstermen in den ausgewerteten Referenzpublikationen, welche von (mindestens) einem Term der als Liste Regulärer Ausdrücke implementierten hypothesenbasierten Stichwortliste (Tab. 14) als Treffer (*regular expression match*) erkannt wurde. Der vertikale Boxplot, der die Verteilung des Datenbankrangkoeffizienten zeigt, gilt für die beiden Teilgraphiken A und B. Alle Variablen wurden logarithmiert. Der Rangkoeffizient weist mit dem ersteren Parameter nur eine schwache, mit letzterem, der als Numerator in der Formel für den Rangkoeffizienten auftritt, eine signifikante (Produkt-Moment)-Korrelation auf. Die Koeffizienten der Regressionsgeraden wurden mit der R-Funktion *lm()* berechnet.



**Tabelle 15:** Die wahrscheinlichsten potentiellen funktionellen Kandidatengene der Eutergesundheit im BTA18-SCS-QTL gemäß dem heuristischen datenbankbasierten Ansatz (254 Gene insgesamt, 129 mit Datenbank-Rangkoeffizient). **Gen, ID:** Gensymbol bzw. (humane) EntrezGeneID des positionellen Kandidatengens, **M:** Datenbank-Rangkoeffizient, **PMIDs:** Zahl der für dieses Gen ausgewerteten PubMedartikel, **Matches:** Zahl der Übereinstimmungen (*matches*) zwischen der hypothesenbasierten Stichwortliste und der MeSH-Annotation der ausgewerteten Artikel, **MeSH-Terme:** Anzahl der unikalen MeSH-Terme.

	<b>Gen</b>	<b>ID</b>	<b>M</b>	<b>PMIDs</b>	<b>Matches</b>	<b>MeSH-Terme</b>
1	HCST	10870	0,51	12	68	133
2	CD79A	973	0,48	56	258	542
3	CD22	933	0,43	42	207	482
4	TYROBP	7305	0,39	35	157	400
5	RELB	5971	0,29	81	231	802
6	CEACAM1	634	0,28	85	276	981
7	CEACAM3	1084	0,27	18	65	243
8	CEACAM6	4680	0,24	40	110	451
9	CEACAM5	1048	0,23	40	96	424
10	TGFB1	7040	0,22	1158	1489	6698
11	IRF2BP1	26145	0,21	7	22	106
12	CEACAM8	1088	0,2	16	38	189
13	CEACAM4	1089	0,16	2	4	25
14	PVRL2	5819	0,16	36	67	422
15	BCL3	602	0,16	47	75	480

### 1.3 Der kombinierte Ansatz: Integration von Datenbankansatz und Transkriptomstudie

Mithilfe des kombinierten Datenbank-/Transkriptom-Ansatzes (Kap. II.1.3) wurden die datenbankbasierten Rangkoeffizienten mit den Resultaten der M.A.S.-Net-Transkriptomstudie integriert, um eine verfeinerte und aussagekräftigere Rangliste potentieller funktioneller Kandidatengene für den BTA18-SCS-QTL zu erstellen.

#### 1.3.1 BTA18-SCS-QTL-spezifische Ergebnisse der M.A.S.-Net Transkriptomstudie

Das Konzept des Verbundprojektes M.A.S.-Net zur Entschlüsselung der dem BTA18-SCS-QTL im DH-Rind zugrundeliegenden kausalen Mutation beinhaltet eine vergleichende Transkriptomstudie von Färsen unter Berücksichtigung ihres Genotyps an diesem QTL. Das favorisierte QTL-Allel, d.h. das zu einem niedrigeren Wert für den SCS und damit vermutlich einhergehender erhöhter Resistenz gegen Euterinfektionen beitragende Allel, soll im folgenden mit **Q**, das negative Allel mit entgegengesetzter Wirkung mit **q** bezeichnet werden.

Untersucht wurden zwei Gruppen von Färsen, Töchter QTL-heterozygoter Bullen, welche jeweils entgegengesetzt homozygote Allelkombinationen für den QTL aufweisen, also **QQ** in der folglich als Hochresistenzgruppe bezeichneten Färsengruppe versus **qq** in der Niedrigresistenzgruppe; vgl. Kap. II.4.3.

Von diesen Tieren wurden im Rahmen von M.A.S.-Net Expressionsprofile für jeweils fünf Eutergewebe erstellt. Bei diesen Geweben handelte es sich um:

**LN:** Euterlymphknoten

**SK:** Strichkanal

**FR:** Fürstenbergsche Rosette

**Pa:** Parenchym

**DZ:** Drüsenzisterne

Einen für den BTA18-SCS-QTL spezifischen Überblick über dabei identifizierte differentiell exprimierte Probesets gibt Tabelle 29 (im Anhang).

Für die vorliegende Arbeit standen diese Daten als Paare von Werten für den *fold change*, d.h. den Quotienten der Mittelwerte beider Testgruppen (Kap. II.1.3) und der Irrtumswahrscheinlichkeit  $p$  eines zweiseitigen ungepaarten Welch-Test (T-Test für Gruppen mit ungleichen Varianzen) für jedes Probeset (vgl. Kap I.2.2.1) und jedes der fünf Eutergewebe zur Verfügung.

Mittels der von Hintermair (2007) erarbeiteten speziesübergreifenden Annotation des verwendeten bovinen *GeneChip*<sup>®</sup> wurden die Probesetnamen in humane EntrezGeneIDs konvertiert; diese wurden in den innerhalb dieser Arbeit genutzten Programmen und Skripten durchweg zur Gen-Identifizierung genutzt.

Der Expressionsdatensatz wurde nun durch Abgleich der EntrezGeneID auf diejenigen Probesets, die den 254 positionellen Kandidatengenegen der Region des BTA18-SCS-QTL zugeordnet wurden, eingegrenzt. Für 142 positionelle Kandidatengene (56%) konnten auf diesem Wege repräsentative Probesets identifiziert werden. Für einen Teil dieser Gene gilt, dass sie von mehr als einem Probeset repräsentiert werden; das Maximum liegt bei fünf Probesets pro Gen. Daraus resultiert eine Gesamtzahl von 212 Probesets, welche die QTL-Region abdecken.

Tabelle 29 (im Anhang) gibt für diejenigen dieser 212 Probesets, die in mindestens einem der untersuchten Gewebe eine signifikante ( $p \leq 0,05$ ) differentielle Expression aufweisen, einen Überblick über die *fold changes* zwischen den Färsengruppen QQ und qq über alle untersuchten Gewebe. Für die dort gezeigten Probesets gibt Tabelle 30 (im Anhang) eine Übersicht über die Expressionswerte in den einzelnen Tieren.

Abbildung 20 (im Anhang) zeigt die relativen und absoluten Frequenzen der 142 Gene der BTA18-SCS-QTL-Region in Abhängigkeit von der Anzahl der ihnen zugeordneten Probesets im Vergleich mit den entsprechenden Werten des gesamten Array. Ein  $\chi^2$ -*goodness-of-fit*-Test zeigt, dass die Frequenzverteilungen nicht signifikant voneinander abweichen ( $\chi^2 = 4,29$ , Freiheitsgrade = 4,  $p = 0,368$ ).

Die Anzahl differentiell exprimierter ( $p_{T-Test} < 0,05$ ) Probesets des BTA18-SCS-QTL in Abhängigkeit von ihrer Regulationsrichtung in den fünf analysierten Eutergeweben zeigt Tabelle 16.

### 1.3.2 Integration der QTL-spezifischen Transkriptomresultate in den heuristischen datenbankbasierten Ansatz

Die Ergebnisse der Transkriptomstudie wurden in den heuristischen datenbankbasierten Ansatz integriert. Zunächst ergab ein Abgleich der Genannotationen, vergleiche hierzu Tabelle 17, dass ein signifikanter Zusammenhang zwischen den Parametern „Ein Datenbank-Rangkoeffizient konnte berechnet werden“ und „Das Gen ist auf dem *GeneChip*<sup>®</sup> vertreten“ für die untersuchten positionellen Kandidatengene besteht ( $\chi^2 = 13,22$  (mit Yates-Korrektur), Freiheitsgrad = 1,  $p < 0,001$ ).

Für 87 der insgesamt 254 positionellen Kandidatengene trafen beide Kriterien zu.

**Tabelle 16:** Die Zahl differentiell exprimierter Probesets (mit  $p < 0,05$ ) unter den 212 Probesets, welche die positionellen Kandidatengene repräsentieren, in den fünf Eutergeweben. QQ bzw. qq: Färsen mit homozygoten Q-Allelen (für hohe Euterinfektionsresistenz) bzw. q-Allelen (für verminderte Euterinfektionsresistenz) am BTA18-SCS-QTL (siehe Text). Da jedes Gen durch mehrere Probesets auf dem bovinen *GeneChip*<sup>®</sup> vertreten sein kann (Abb. 20 im Anhang), ist gegebenenfalls in Klammern angegeben, wie viele Gene tatsächlich durch die differentiell exprimierten Probesets repräsentiert werden. Abk.: LN, Lymphknoten; FR, Fürstenbergsche Rosette; Pa, Parenchym; DZ, Drüsenzisterne; SK, Strichkanal

höher exprimiert in	Gewebe				
	LN	FR	Pa	DZ	SK
QQ	17	3	1	4	6
qq	1	21 (19)	4	13	8

**Tabelle 17:** Kontingenztabelle des Abgleichs der EntrezGeneID-Annotationen von 254 positionellen Kandidatengen mit der Annotation des verwendeten bovinen *GeneChip*<sup>®</sup>. DbRk: Datenbank-Rangkoeffizient

mit DbRK		mit Probeset(s)		$\chi^2$	p
		nein	ja		
	nein	70	55	13,22	0,00028
	ja	42	87		

Diesen 87 Genen konnten 135 Probesets zugeordnet werden. Aus Abbildung 20 (im Anhang) geht hervor, dass die Frequenzverteilung der Gene in Abhängigkeit von der Zahl der sie repräsentierenden Probesets derjenigen der Gesamtheit der positionellen Kandidatengene auf dem Array folgt (statistisch validiert durch einen  $\chi^2$ -goodness-of-fit-Test;  $\chi^2 = 1,2729$ , Freiheitsgrad = 4,  $p = 0,866$ ).

Jedem dieser 135 Probesets wurde der DbR-Koeffizient des zugehörigen positionellen Kandidatengens zugewiesen. Zusammen mit den 42 positionellen Kandidatengen mit errechnetem DbR-Koeffizienten, aber ohne Repräsentation auf dem Array, ergeben sich 177 Datensätze, für 129 positionelle Kandidatengene, die für die Ermittlung der wahrscheinlichsten potentiellen funktionellen Kandidatengene des BTA18-SCS-QTL nach dem kombinierten Datenbank/Transkriptom-Ansatz im folgenden in Betracht gezogen wurden.

Die Berechnung des kombinierten DbTR-Koeffizienten erfolgte nach den in Kap. II.1.3 dargestellten Formeln.

Für alle Datensätze von nicht auf dem Array vertretenen Genen, sowie denjenigen, die in keinem der fünf Gewebe eine signifikante (p-Wert eines ungepaarten zweiseitigen Welch-T-

Test  $\leq 0,05$ ) differentielle Expression zwischen den hochresistenten QQ- und den niedrigresistenten qq-Färsen zeigen, gilt dabei: kombinierter DbTR-Koeffizient = DbR-Koeffizient. Für 140 der 177 Datensätze (79%) ist dies hier der Fall.

37 Probesets zeigten eine signifikante differentielle Expression in mindestens einem Gewebe, daher gilt: kombinierter DbTR-Koeffizient  $>$  DbR-Koeffizient.

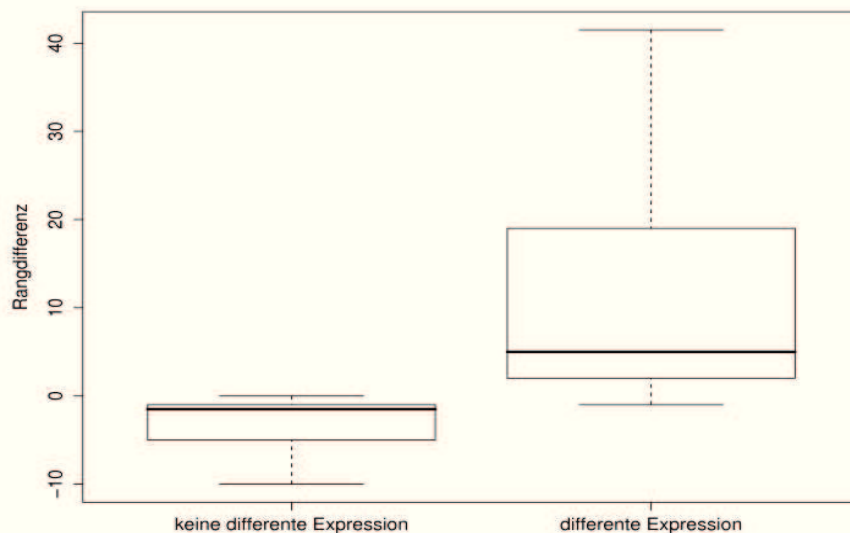
Die Sortierung der 177 Datensätze nach dem kombinierten DbTR-Koeffizienten ergibt eine Rangliste, die sich von derjenigen nach dem DbR-Koeffizienten unterscheidet.

Für jeden Datensatz wurde durch Vergleich der nach beiden Werten erhaltenen Ranglisten der Wert

$$\Delta Rang = Rang_{\text{kombinierter DbTR-Koeffizient}} - Rang_{\text{DbR-Koeffizient}}$$

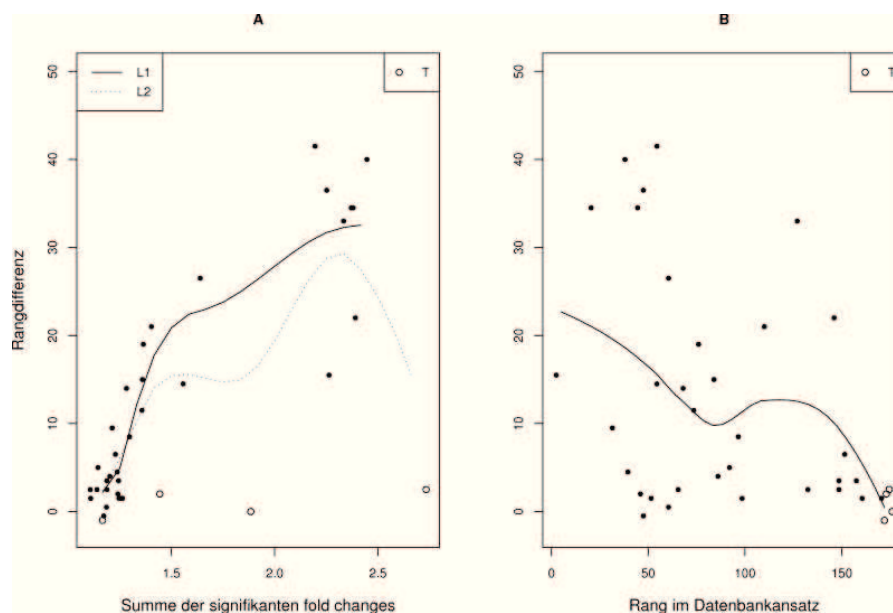
errechnet.

Gene ohne differentielle Expression erfahren im kombinierten Datenbank/Transkriptom-Ansatz zumeist einen Rangverlust (Median von  $\Delta Rang$ : -1,5), während solche mit einer differentiellen Expression ( $n = 37$ ) Ränge hinzugewinnen (Median: 5). Abbildung 10 zeigt die Boxplots der Verteilung von  $\Delta Rang$  über alle Datensätze ( $n = 177$ ) in Abhängigkeit von ihrem Expressionsstatus.



**Abbildung 10:** Für 177 Probesets, die 129 Gene repräsentieren, für welche im Datenbankansatz ein Rangkoeffizient  $\neq 0$  ermittelt werden konnte (siehe Text), wurde die Rangdifferenz  $\Delta Rang = Rang_{\text{kombinierter Ansatz}} - Rang_{\text{Datenbankansatz}}$  ermittelt. Dargestellt ist  $\Delta Rang$  in Abhängigkeit davon, ob das Probeset zwischen der resistenten QQ- und der empfindlichen qq-Gruppe in mindestens einem der untersuchten fünf Eutergewebe signifikant differentiell exprimiert ist.

Für die Datensätze der Probesets mit signifikant differentieller Expression in mindestens einem Gewebe wurde untersucht, in welcher Relation  $\Delta Rang$  zu der Summe der statistisch signifikanten *fold changes* des Probesets zwischen der QQ- und der qq-Gruppe steht (In der Formel 5 zur Berechnung des kombinierten DbTR-Koeffizienten wird diese Summe als Faktor eingesetzt). Diese Beziehung ist anhand eines Scatterplots in Abbildung 11A dargestellt. Zu beobachten ist, dass der Zugewinn an Rängen im kombinierten Datenbank/Transkriptom-Ansatz desto höher ausfällt, je höher der *fold change* ist. Limitiert ist dieser Zusammenhang bei Genen, denen im heuristischen datenbankbasierten Ansatz ein DbR-Koeffizient zugewiesen wurde, der zur Einstufung als eines der besten funktionellen Kandidatengene führte, da dies die Möglichkeiten für weiteren Rangzugewinn einschränkt. Ein Scatterplot der Rangzugewinne gegen den Rang des Gens im Datenbankansatz für diese 37 Datensätze, der einen solchen Zusammenhang andeutet, ist in Abbildung 11B zu sehen. An die in beiden Abbildungen gezeigten Daten wurden polynomiale lokale Regressionsfunktionen (*locally estimated scatterplot smoothing*, LOESS) der Form Rangdifferenz  $\sim$  *fold change* bzw. Rangdifferenz  $\sim$  DbR-Koeffizient angepasst und graphisch in der jeweiligen Abbildung dargestellt.



**Abbildung 11:** Die Abhängigkeit der Rangdifferenz eines Probesets, durch das ein positionelles Kandidatengen repräsentiert wird, von **A)** der Summe der Faktoren der Expressionsunterschiede (*fold change*) des Probesets in einem oder mehreren Geweben mit signifikant differentieller Expression, und **B)** dem vom Kandidatengen im datenbankbasierten Ansatz eingenommenen Rang. Dargestellt sind die Werte für diejenigen 37 Probesets, die sowohl einen DbR-Koeffizienten als auch eine signifikant differentielle Expression in mindestens einem Gewebe aufweisen.

T: Gene, welche im heuristischen datenbankbasierten Ansatz zu den fünf bestbewerteten Genen gehörten, L1: LOESS-Regression unter Ausschluss der Gene in T (nur in A), L2: LOESS-Regression unter Einschluss aller 37 Datensätze.

Auf Basis des kombinierten Rangkoeffizienten wurde eine Rangliste potentieller funktioneller Kandidatengene des BTA18-SCS-QTL erstellt, die aufgrund der Integration mehrerer Datenquellen von hoher Relevanz die informierte Auswahl von Kandidatengenen für die weitere experimentelle Untersuchung ermöglicht.

Die nach dem Datenbank/Transkriptom-Ansatz bestbewerteten Kandidatengene sind in Tabelle 18 dargestellt, zusammen mit wesentlichen Parametern der einzelnen Ansätze, aus denen sich die kombinierte Bewertung zusammensetzt. Dies sind im einzelnen der Datenbank-Rangkoeffizient der Gene als Maß für die Involviertheit der Gene in immunologischen Prozessen (vgl. Tab. 15) sowie die Magnituden (*fold change*) eventueller signifikanter Expressionsunterschiede zwischen den Färsengruppen mit genetisch determinierter differenter Resistenz gegen Euterinfektionen. In der Tabelle sind letztere weiter aufgeschlüsselt nach den Geweben, in denen sie auftreten. Es wird deutlich, dass eine signifikante differentielle Expression der hochpriorisierten Gene v.a. in den besonders immunrelevanten Eutergeweben Lymphknoten und Fürstenbergsche Rosette zu verzeichnen ist (daneben auch in der Drüsenzisterne); vgl. dazu auch Tabelle 16. Insbesondere zeigen die im Lymphknoten differentiell exprimierten Transkripte eine gleichgerichtete Regulation, nämlich eine Heraufregulation in der hochresistenten Färsengruppe.

Die Darstellung der Gene erfolgt hier auf der Ebene der einzelnen Probesets. Da jedes Gen durch mehrere Probesets auf dem *GeneChip*<sup>®</sup> repräsentiert sein kann (vgl. Tab. 20 im Anhang), kann es aufgrund der separaten Berechnung des kombinierten DbTR-Koeffizienten für jedes Probeset, abhängig vom Expressionsstatus der Probesets in den Geweben, auch mehrfach in der Tabelle auftreten. Dies betrifft konkret die Gene *CD22* und *CEACAM1*.

Einem anderen Gen, nämlich *HCST*, konnte kein Probeset zugewiesen werden; seine Präsenz unter den am höchsten priorisierten Genen ist demnach allein auf seine durch den Datenbankansatz ermittelte hohe Relevanz für immunologische Prozesse zurückzuführen.

Für die weitere experimentelle Bearbeitung wurden letztlich vier Gene als potentielle funktionelle Kandidatengene ausgewählt. Neben dem erwähnten *HCST* waren dies *CD22*, *CD79A* und *TYROBP*. Diese Gene wurden durch beide Ansätze hoch priorisiert, zwei von ihnen, *CD22* und *TYROBP* erfuhren im kombinierten Ansatz einen Rangzugewinn.

Mit der Auswahl dieser Gene sind beide Intervalle des feinkartierten BTA18-SCS-QTL im experimentellen Teil der Arbeit vertreten, da *CD22*, *HCST* und *TYROBP* im Intervall A (5) sowie *CD79A* im Intervall B (9) lokalisiert sind.

**Tabelle 18:** Die wahrscheinlichsten potentiellen funktionellen Kandidatengene der Eutergesundheit im BTA18-SCS-QTL gemäß dem kombinierten Datenbank/Transkriptom-Ansatz. Die Darstellung erfolgt hier auf der Ebene des einzelnen Probesets, d.h. ein Gen kann mehrmals in der Tabelle vertreten sein, wenn es durch mehrere Probesets repräsentiert wird (z.B. CD22, siehe Text und Abb. 20 im Anhang). **Gen, ID:** Gensymbol bzw. EntrezGeneID des dem Probeset zugeordneten Gens; **DbRK:** Datenbank-Rangkoeffizient (nach Formel 3); **DbTRK:** kombinierter Rangkoeffizient (nach Formel 5);  $\sum FC$ : Summe der signifikanten *fold changes*<sup>a</sup> (nach Formel 4);  $\delta$ : Hat das positionelle Kandidatengen im Vergleich zur Rangliste nach dem datenbankbasierten Rangkoeffizienten (vergleiche Tabelle 15) Ranglistenplätze gewonnen (+) oder verloren (-)? Die fünf folgenden Spalten schlüsseln den Wert für  $\sum FC$  nach den einzelnen untersuchten Geweben auf (vgl. Tab. 16);  $\uparrow$ : Das Probeset wird in der Gruppe mit QQ-Genotyp für den BTA18-SCS-QTL (d.h. mit höherer Mastitisresistenz) höher exprimiert; eine entgegengesetzte Regulation ist durch  $\downarrow$  gekennzeichnet. Abk. LN: Lymphknoten; DZ: Drüsenzisterne; FR: Fürstenbergsche Rosette; Pa: Parenchym; SK: Strichkanal

	ProbesetID	Gen	ID	DbRK	DbTRK	$\sum FC$	$\delta$	LN	DZ	FR	Pa	SK
1	Bt.4338.1.S1_at	CD22	933	0,43	1,17	2,73	+	$\uparrow$ 1,45	$\uparrow$ 1,27			
2	Bt.4436.1.S1_a.at	CD79A	973	0,48	0,9	1,88	0	$\uparrow$ 1,88				
3	Bt.8939.1.S1_at	TYROBP	7305	0,39	0,57	1,44	+			$\downarrow$ 1,44		
4	-	HCST	10870	0,51	0,51	1	-					
5	Bt.4338.2.S1_at	CD22	933	0,43	0,43	1	-					
6	Bt.10027.1.S2_at	CEACAM1	634	0,28	0,36	1,26	+		$\uparrow$ 1,26			
7	Bt.17669.1.S1_a.at	RELB	5971	0,29	0,34	1,17	-	$\uparrow$ 1,17				
8	Bt.10027.2.S1_a.at	CEACAM1	634	0,28	0,28	1	-					
9	-	CEACAM3	1084	0,27	0,27	1	0					
10	Bt.2466.1.S1_at	POLR2I	5438	0,11	0,26	2,39	+	$\uparrow$ 1,24		$\downarrow$ 1,15		
11	-	CEACAM6	4680	0,24	0,24	1	-					
12	-	CEACAM5	1048	0,23	0,23	1	-					
13	Bt.469.1.A1_at	TGFB1	7040	0,22	0,22	1	-					
14	Bt.19527.1.S1_at	IRF2BP1	26145	0,21	0,21	1	-					
15	-	CEACAM8	1088	0,2	0,2	1	-					

<sup>a</sup>d.h. Faktoren der Expressionsunterschiede



## 2 Identifizierung und vergleichende Sequenzierung von SNPs in den potentiellen funktionellen Kandidatengenen

Der kombinierte Datenbank/Transkriptom-Ansatz, dessen Ergebnisse auf den vorigen Seiten vorgestellt werden, identifizierte vier potentielle funktionelle Kandidatengene der Region des BTA18-SCS-QTL: *CD22*, *CD79A*, *HCST* und *TYROBP*.

Das vorliegende Kapitel beschreibt die Identifizierung potentiell mit dem Merkmal des Gehaltes somatischer Zellen in der Milch (SCS) assoziierter SNPs innerhalb dieser Gene. Diese erfolgte in einem mehrstufigen Verfahren.

Mithilfe der SNP-Referenzdatenbank des NCBI, *dbSNP*<sup>24</sup> wurden die Kandidatengene zunächst *in silico* auf bereits beschriebene SNPs untersucht. SNP-enhaltende Sequenzabschnitte wurden im folgenden priorisiert.

Die für die Sequenzierung abgeleiteten genomischen PCR- und Sequenzierprimer (Tab. 34 im Anhang), basieren auf der Sequenz des bovinen genomischen Contig mit der GenBank-Accessionnummer NC\_007316.2, welcher Teil der zu Beginn des Versuchszeitraumes aktuellen Version 3.1 der bovinen Genomsequenz ist (Tab. 19).

Durch vergleichende Sequenzierung der Kandidatengene in zwei Extremwertpools mit zusammen 30 DH-Referenzbullen wurden insgesamt 16 SNPs detektiert. Die Gene wurden dabei nicht vollständig sequenziert, vielmehr wurde das hier vorgestellte mehrstufige Verfahren für detektierte SNPs wiederholt, bis eine genügend große Stichprobe von potentiell merkmalsassoziierten SNPs für die Überprüfung auf eine tatsächliche Assoziation in einer großen Halbgeschwisterpopulation (Kap. III.3) zusammengestellt war.

Die gefundenen 16 SNPs werden zunächst in ihrem jeweiligen genomischen Kontext vorgestellt.

Darauf werden die beiden Stufen des Verfahrens zur Auffindung potentiell merkmalsassoziiierter SNPs mit ihren Ergebnissen präsentiert: Die selektive Poolsequenzierung ab S. 75 und der Konkordanztest ab S. 78.

### 2.1 Genomische Darstellung der SNPs

Die Reihenfolge, in der die potentiellen funktionellen Kandidatengene der Region des BTA18-SCS-QTL und die in ihnen detektierten SNPs im genomischen Kontext vorgestellt werden, folgt der Rangfolge der Gene nach dem kombinierten DbTR-Koeffizienten (Tab. 18). Dieses Schema wird auch in späteren Ausführungen innerhalb dieser Arbeit beibehalten.

Der Darstellung der genomischen Organisation der potentiellen funktionellen Kandidatengene und der SNP-Positionen liegt die Version 4.0 des *Bos taurus*-Genoms zugrunde.

---

<sup>24</sup>URL: <http://www.ncbi.nlm.nih.gov/projects/SNP/>

**Tabelle 19:** Die vier positionellen und potentiellen funktionellen Kandidatengene *CD22*, *CD79A*, *HCST* und *TYROBP* des BTA18-SCS-QTL wurden nach Maßgabe des kombinierten Datenbank/Transkriptom-Ansatzes für die Identifizierung potentiell merkmalsassoziiierter SNPs mittels vergleichender Genomsequenzierung ausgewählt. Angegeben sind die bovine EntrezGeneID (die IDs der humanen Homologen gibt Tab. 18), die genomische Ausdehnung des gesamten Transkriptes sowie die Start- und Endpositionen der genomischen Referenzsequenz, d.h. des Transkriptes plus flankierende Region, auf BTA18, die u.a. für die Ableitung der genomischen Primer genutzt wurde. Die Angaben basieren auf der Version 3.1 der bovinen Genomsequenz (Btau3.1), die zum Zeitpunkt des Versuchsbeginns aktuell war.

Gen	bovine EntrezGeneID	Ausdehnung in Kb	genomische Referenz (BTA18, Btau 3.1)	
			Startposition	Endposition
<i>CD22</i>	514582	14,5	45798633	45815152
<i>CD79A</i>	281674	3,6	48561414	48556751
<i>TYROBP</i>	282390	2	41595000	41601000
<i>HCST</i>	767892	2	41595000	41601000

Die genomischen (bovinen) Sequenzen inklusive 1 bis 5 Kb der flankierenden Regionen der Kandidatengene sowie die gesamte Sequenz des BTA18 wurden im Fastaformat sowie (nur für die Gensequenzen) im annotierten GenBank-Format lokal gespeichert.

Die genomischen Positionen der identifizierten SNPs relativ zum ersten Nukleotid von BTA18 bzw. zum Beginn des ersten annotierten Exon wurden durch BLAST (lokale Installation, Programm *bl2seq* (Kap. II.3.4) ermittelt, während die Start- und Endpositionen der Exons der Annotation entnommen wurden.

Die Namensgebung der SNPs folgt diesem Schema:

$$[\textit{Gensymbol}]_{[\textit{Position relativ zum ersten Exon}]}:[\textit{Allel A}]>[\textit{Allel B}]$$

Mit der Reihenfolge, in der die Allele im Namen angegeben werden, ist keine Aussage über Allelfrequenzen verbunden.

Die hier identifizierten SNPs wurden mit den in der *dbSNP* annotierten, bereits beschriebenen SNPs abgeglichen. Sofern sie dort bereits aufgenommen waren, sind sie mit ihrer Referenzidentifikationsnummer (*rsID*) angegeben, andernfalls (in drei Fällen) wurden sie als Neueintragung an die *dbSNP* übermittelt und sind mit der ihnen zugewiesenen *SubmissionID*, *ssID* angegeben.

### 2.1.1 SNPs im Gen *CD22*

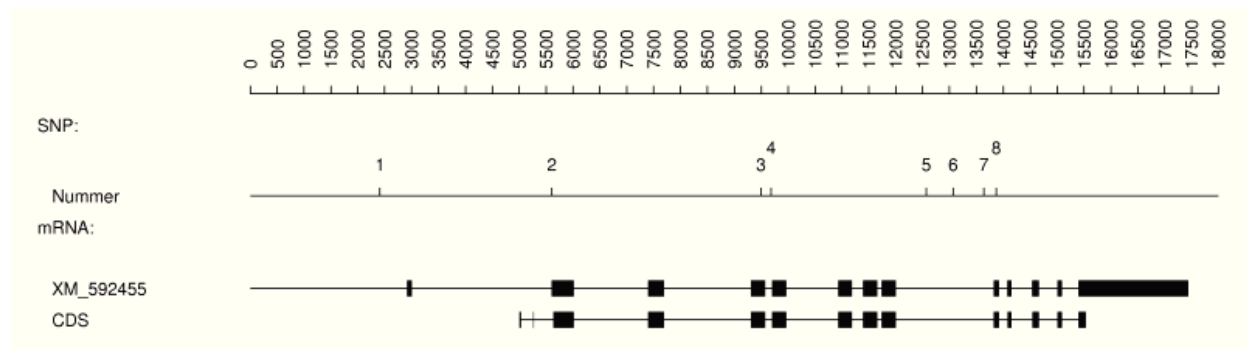
Die Sequenz mit der Accessionnummer XM\_592455 ist die ursprünglich in GenBank eingestellte Referenzsequenz (*RefSeq*) für die bovine mRNA des Gens *CD22* und wurde in der Zwischenzeit zurückgezogen; momentan (Stand 03.02.2010) gibt es in GenBank für dieses Gen keine bovine Referenz-mRNA. Die für XM\_592455 codierende genomische Sequenz umfaßt 14,5 Kb, diejenige für die in der aktuellen Annotation anstelle einer *RefSeq* angegebene *coding DNA sequence* (CDS) 10,9 Kb. Letztere Sequenz zuzüglich je 5 Kb der flankierenden Regionen wurden für die hier gezeigte Annotation verwendet.

Identifiziert wurden acht SNPs, davon fünf intronische. Bei einem von diesen, *ss159831343*, handelte es sich um eine Neueintragung in *dbSNP*. Zwei weitere sind in Exons gelegen (Tab. 20). Der erste SNP (mit der laufenden Nummer 1) liegt 509 bp stromaufwärts von Exon 1 der unter *accession*-Nummer XM\_592455 in GenBank abgelegten Nukleotidsequenz. Dieses Exon fehlt in der CDS, stattdessen ist dort ein alternatives erstes Exon etwa 2,1 Kb stromabwärts annotiert.

Wesentliche Parameter und die Positionen der im Gen *CD22* detektierten SNPs sind Tabelle 20 und der zugehörigen Abbildung 12 zu entnehmen.

**Tabelle 20:** Im Gen *CD22* aus dem QTL-Intervall 5 wurden acht SNPs, davon zwei exonische und fünf intronische gefunden, zusätzlich einer in der 5'-flankierenden Region. Die Position relativ zum Beginn des ersten Exon, welche Bestandteil des Namens ist (siehe Text) bzw. zum ersten Nukleotid des BTA18 wurde durch lokalen BLAST gegen den Build 4.0 des bovinen Genomes ermittelt. Die Angaben zu Start- und Endpunkt (erste bzw. letzte Zeile) beziehen sich auf Abb. 12.

Nr.	Name	dbSNP	Pos. im Gen	Nukleotid- position auf BTA18
Start der Nukleotidsequenz				45457502
1	CD22_-2595:C>T	rs41883763	5'-flankierende Region	45459907
2	CD22_611:A>C	rs41883767	Intron 1	45463112
3	CD22_4494:C>T	rs41883772	Exon 4	45466995
4	CD22_4680:A>G	rs41639753	Intron 4	45467181
5	CD22_7568:C>T	rs41883777	Intron 8	45470069
6	CD22_8067:A>G	ss159831343	Intron 8	45470568
7	CD22_8636:C>T	rs41883778	Intron 8	45471137
8	CD22_8867:A>G	rs41883779	Exon 9	45471368
Ende der Nukleotidsequenz				45478415



**Abbildung 12:** Die genomische Organisation des bovinen Kandidatengenes *CD22* und die Positionen der detektierten SNPs. Gezeigt werden die Exon/Intronstrukturen der unter der *accession*-Nummer XM\_592455 in GenBank abgelegten Nukleotidsequenz, welche formell nicht mehr gültig ist, hier zu Vergleichszwecken aber trotzdem gezeigt wird, sowie des aktuell annotierten CDS (*coding sequence*). Beide unterscheiden sich durch die Lage des ersten und die Ausdehnung des letzten Exons. Gezeigt sind hier nur Nukleotide 1-18000. Zur Lage von Start- und Endpunkt der Nukleotidsequenz auf BTA18 und der Identität der mit Nummern bezeichneten SNPs vgl. Tab. 20.

### 2.1.2 SNPs im Gen *CD79A*

Das Gen *CD79A* liegt in reverser Orientierung (d.h. auf dem Minusstrang) im Genom. Für die in Abbildung 13 gezeigte Annotation wurden 5,6 Kb genomische Sequenz des Minusstranges, die dem Gen plus jeweils 1 Kb der flankierenden Regionen entsprechen, genutzt (Tab. 19). Als mRNA-Referenz dient die Nukleotidsequenz mit der GenBank-*Accession*-Nummer NM\_174266.2. Gefunden wurden drei SNPs, davon zwei intronische und einer im 5'-flankierenden Bereich. Alle drei waren bereits in der *dbSNP* annotiert; Tabelle 21 zeigt die Namen und Positionen, letztere bezogen auf den Transkriptionsstart des Gens (in Transkriptionsrichtung) bzw. als absolute Basenposition auf der publizierten Nukleotidsequenz von BTA18.

Wesentliche Parameter und die Positionen der im Gen *CD79A* detektierten SNPs sind Tabelle 21 und der zugehörigen Abbildung 13 zu entnehmen.

#### Der genomische Kontext des SNP *CD79A*.-182:A>G

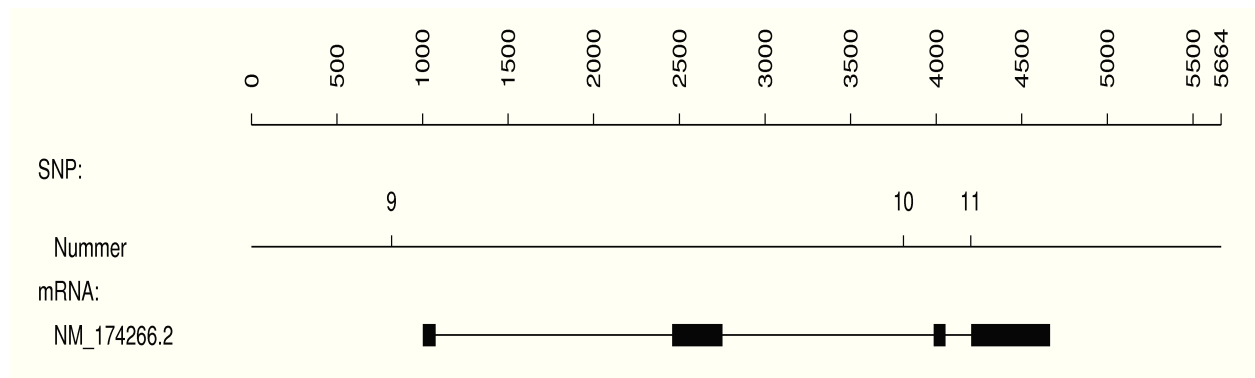
Im Fall von *CD79A*.-182:A>G (rs41886731) ist der unmittelbare genomische Kontext dieses SNP von besonderem Interesse, da es sich hierbei um die (erweiterte) Promotorregion des Gens *CD79A* handelt. Dieser wurde daher genauer untersucht. Die Nukleotidsequenz von Position -200 bis Position -150, relativ zum ersten Exon des Gens, die besagten SNP enthält, ist in Abbildung 14 dargestellt, annotiert mit einigen DNA-Elementen.

Der mit dem SNP verbundene Austausch von A gegen G bzw. von T gegen C (auf dem Gegenstrang, auf dem *CD79A* codiert ist) zerstört eine (imperfekte) palindromische

Sequenz und gleichzeitig ein in diesem enthaltenes putatives regulatorisches Element, ein  $\gamma$ -IRE (*interferone responsive element*), das durch die Konsensussequenz CWKKANNY (=C<sup>T</sup><sub>AGG</sub><sup>T</sup><sub>T<sup>T</sup><sub>TANN<sup>T</sup><sub>C</sub>) charakterisiert ist (Yang *et al.*, 1990). Die genannten Elemente befinden sich in unmittelbarer räumlicher Nähe zu einer Bindestelle für NF- $\kappa$ B.</sub></sub>

**Tabelle 21:** Im Gen *CD79A* aus dem QTL-Intervall 9 wurden drei SNPs, zwei intronische und ein Promotor-SNP gefunden. Die Position relativ zum Beginn des ersten Exon, welche Bestandteil des Namens ist (siehe Text) bzw. zum ersten Nukleotid des BTA18 wurde durch lokalen BLAST gegen den Build 4.0 des bovinen Genomes ermittelt. Die Angaben zu Start- und Endpunkt (erste bzw. letzte Zeile) beziehen sich auf Abb. 13.

Nr.	Name	dbSNP	Pos. im Gen	Pos. auf BTA18
Start der Nukleotidsequenz				51080791
9	CD79A_-182:A>G	rs41886731	5'-flankierende Region	51079974
10	CD79A_2808:C>T	rs41886735	Intron 2	51076984
11	CD79A_3201:A>G	rs41886736	Intron 3	51076591
Ende der Nukleotidsequenz				51075128



**Abbildung 13:** Die genomische Organisation des bovinen Kandidatengenes *CD79A* und die Positionen der detektierten SNPs. Das Gen befindet sich in reverser Orientierung auf dem Chromosom, gezeigt wird hier der Minusstrang. Die Exon/Intronstruktur folgt der GenBank-Annotation der Transkriptsequenz *NM\_174266.2*. Zur Lage von Start- und Endpunkt der Nukleotidsequenz auf BTA18 und der Identität der mit Nummern bezeichneten SNPs vgl. Tab. 21.



**Abbildung 14:** Der genomische Kontext, d.h. der Bereich von -200 bis -150 Nucleotiden relativ zum ersten Exon von *CD79A* bzw. 800 bis 850 der in Abbildung 13 dargestellten Region, des Promotor-SNP *CD79A*.-182:A>G (rs41886731). Siehe dazu auch Youn *et al.* (1997).

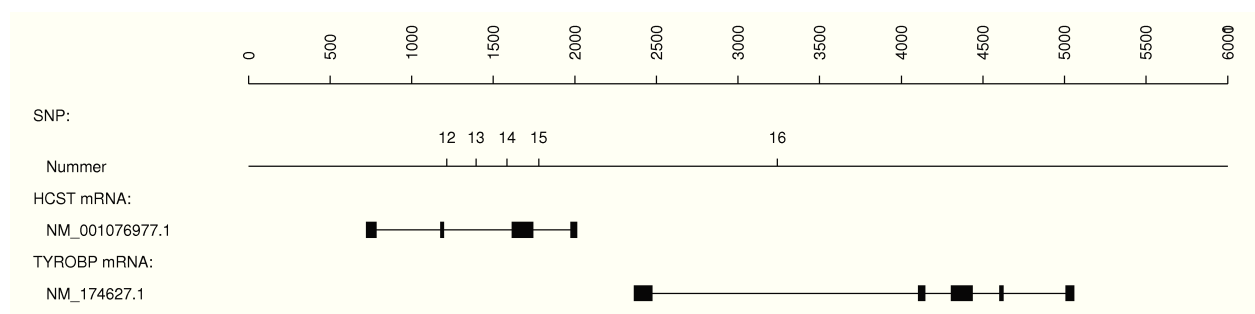
### 2.1.3 SNPs in den Genen *HCST* und *TYROBP*

Die Gene *HCST* (*DAP10*) und *TYROBP* (*DAP12*) liegen unmittelbar benachbart in einer Kopf-an-Kopf-Anordnung (*TYROBP* befindet sich dabei in reverser Orientierung) im Genom, darüberhinaus sind beide Gene sehr kurz; sie werden deshalb in einer gemeinsamen Graphik dargestellt. Die relative Position des im *TYROBP* gelegenen, zum Versuchszeitpunkt in *dbSNP* nicht gefundenen SNP *ss159831344* wurde ungeachtet dessen gleich den anderen Fällen in Relation zum Transkriptionsstart von *TYROBP* in Transkriptionsrichtung dieses Genes bestimmt. Zusätzlich zu diesem SNP konnten im Gen *HCST* weitere vier SNPs detektiert werden; alle SNPs sind intronisch, einer von ihnen, *ss159831345*, war zum Versuchszeitpunkt nicht in *dbSNP* enthalten.

Wesentliche Parameter und die Positionen der in den Genen *HCST* und *TYROBP* detektierten SNPs sind der Tabelle 22 und zugehörigen Abbildung 15 zu entnehmen.

**Tabelle 22:** Namen und Positionen der 5 in den Genen *HCST* und *TYROBP* erfassten SNPs. In den benachbarten Genen *HCST* und *TYROBP*, die im QTL-Intervall 5 lokalisiert sind, wurden fünf SNPs (vier intronische (*HCST*) bzw. ein intronischer (*TYROBP*)) gefunden. Die Position relativ zum Beginn des ersten Exon, welche Bestandteil des Namens ist (siehe Text) bzw. zum ersten Nukleotid des BTA18 wurde durch lokalen BLAST gegen den Build 4.0 des bovinen Genomes ermittelt. Die Angaben zu Start- und Endpunkt (erste bzw. letzte Zeile) beziehen sich auf Abb. 15.

Nr.	Name	dbSNP	Pos. im Gen	Pos. auf BTA18
Start der Nukleotidsequenz				46014000
12	HCST_497:A>C	rs41889075	Intron 2	46015214
13	HCST_675:A>G	rs41889076	Intron 2	46015392
14	HCST_866:C>T	rs41889077	Intron 2	46015583
15	HCST_1060:A>C	ss159831345	Intron 3	46015777
16	TYROBP_1820:A>G	ss159831344	Intron 4	46017239
Ende der Nukleotidsequenz				46020000



**Abbildung 15:** Genomische Organisation der Kandidatengene *HCST* und *TYROBP* sowie Positionen der identifizierten SNPs. Die Gene liegen in einer *head to head*-Tandemkonfiguration mit *TYROBP* auf dem Gegenstrang. Zur Lage von Start- und Endpunkt der Sequenz auf BTA18 und der Identität der mit Nummern bezeichneten SNPs vgl. Tab.22.

## 2.2 Selektive Poolgenotypisierung

Das erste Kriterium für die Beurteilung einer potentiellen Merkmalsassoziation eines SNP ergab sich aus der genaueren Auswertung der für die initiale Identifizierung der SNPs genutzten Daten; daher entspricht die Zahl der SNPs, für die dieses Kriterium bestimmt wurde der Gesamtzahl identifizierter SNPs. Die hierfür vergleichend sequenzierten beiden Pools von DH-Referenzbulln wurden selektiv nach dem Prinzip einer maximalen Differenz in der Verteilung des BTA18-SCS-QTL-spezifischen Zuchtwertes MA-BLUP für das Merkmal SCS zusammengestellt (Kühn *et al.*, 2008); Abbildung 6 gibt dazu einen Überblick. Der Pool der Bullen mit den züchterisch im Sinne der Verbesserung der Eutergesundheit erwünschten niedrigen Werten für den MA-BLUP wird im folgenden mit  $MA-BLUP_{min}$  bezeichnet.

In Anlehnung an Darvasi & Soller (1994) bzw. Lipkin *et al.* (1998), die die SDP-Methode, d.h. die Sequenzierung von Individuen aus den Schwänzen (*tails*) einer phänotypischen Verteilung (vgl. Kap. I.2.1.4), zur Identifizierung merkmalskausaler Regionen etablierten, wird als Arbeitshypothese angenommen, dass potentiell merkmalsassoziierte SNPs der Kandidatengene in den hier getesteten Extremwertpools durch eine deutliche Verschiebung in der Verteilung der Allele zwischen den Pools charakterisiert sind.

Die relativen Allelfrequenzen wurden aus den Sequenzchromatogrammen durch den Vergleich der Peakhöhen abgeschätzt. Abbildung 16 gibt zwei Beispiele dafür; je eines für einen SNP mit stark differenter Verteilung und einen SNP mit kaum differenter Verteilung beider SNP-Allele zwischen den Extremwertpools (die SNPs CD79A\_-182:A>G bzw. HCST\_1060:A>C).

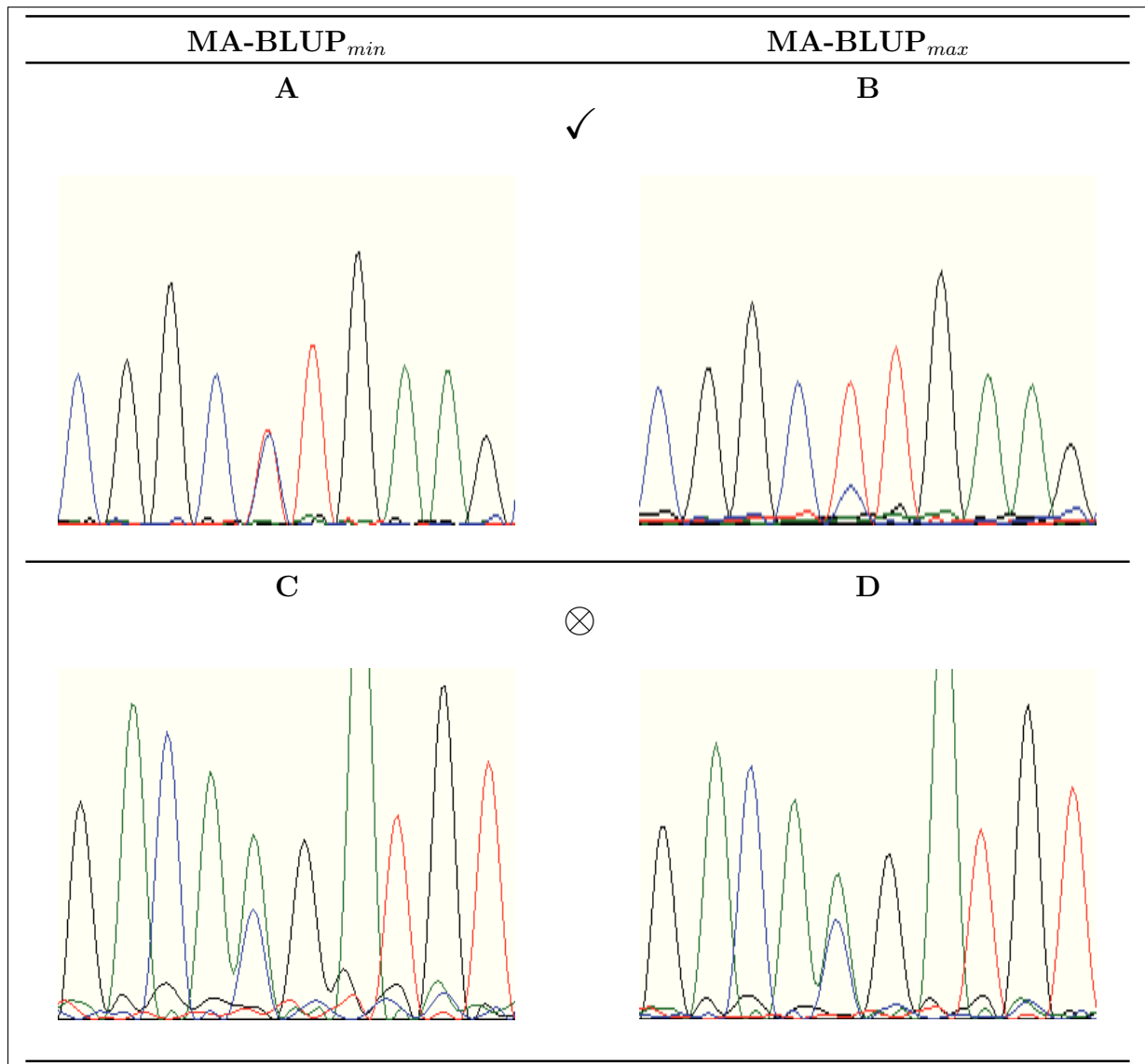
In Tabelle 23 sind die abgeschätzten relativen Frequenzen der beteiligten Allele für alle 16 SNPs in beiden Extremwertpools gezeigt. Die Verhältnisse werden dabei anhand eines Kontinuums von Symbolen, die im unteren Teil der Tabelle erläutert sind, dargestellt.

SNPs, bei denen in den gegensätzlichen Extremwertpools das jeweils gegensätzliche Allel stark dominant war, wurden in dieser Untersuchung nicht gefunden, jedoch in einigen Fällen eine schwächer ausgeprägte gegensätzliche Dominanz. Für mehrere SNPs, darunter HCST\_866:C>T, CD79A\_-182:A>G und CD79A\_2808:C>T, gilt, dass im Extremwertpool  $MA-BLUP_{min}$  beide Allele des SNP in annähernd der gleichen Frequenz auftreten, während im Extremwertpool  $MA-BLUP_{max}$  ein Allel sehr stark dominiert.



**Tabelle 23:** Die relativen Allelverhältnisse von 16 in den ausgewählten Kandidatengenomen identifizierten SNPs in zwei DNA-Pools von je 15 DH-Bullen, welche die entgegengesetzten Extreme der Verteilung von BTAT18-spezifischen Zuchtwerten (MA-BLUP) für das Merkmal SCS repräsentieren. MA-BLUP<sub>min</sub> umfasst Bullen mit favorisiertem Zuchtwert, d.h. diese Bullen vererben vermutlich eine verbesserte Eutergesundheit. Durch ein Häkchen in der Spalte **AS** sind die SNPs markiert, die in der Gesamtwürdigung dieses Ansatzes und des Konkordanztest (Tab. 25) für die Assoziationsstudie in einer DH-Halbgeschwisterpopulation ausgewählt wurden. Die verwendeten Symbole sind im unteren Teil der Tabelle erklärt.

SNP	Allele	MA-BLUP <sub>max</sub>	MA-BLUP <sub>min</sub>	AS
CD22_-2595:C>T	C,T	$C \gg T$	$C > T$	
CD22_611:A>C	A,C	$A \gg C$	$A > C$	
CD22_4494:C>T	C,T	$C \gg T$	$C = T$	
CD22_4680:A>G	A,G	$G \gg A$	$G \gg \gg A$	
CD22_7568:C>T	C,T	$C \gg T$	$C < T$	✓
CD22_8067:A>G	A,G	$G \gg A$	$G \gg \gg A$	✓
CD22_8636:C>T	C,T	$T > C$	$T = C$	
CD22_8867:A>G	A,G	$A > G$	$A = G$	
CD79A_-182:A>G	A,G	$A \gg \gg G$	$A = G$	✓
CD79A_2808:C>T	C,T	$C \gg \gg T$	$C > T$	✓
CD79A_3201:A>G	A,G	$G \gg \gg A$	$G \gg \gg A$	
HCST_497:A>C	A,C	$C > A$	$C \gg A$	
HCST_675:A>G	A,G	$A \gg G$	$A < G$	✓
HCST_866:C>T	C,T	$C \gg \gg T$	$C = T$	✓
HCST_1060:A>C	A,C	$A > C$	$A > C$	
TYROBP_1820:A>G	A,G	$G = A$	$G > A$	
Erklärung der Symbole				
>>>	stark überrepräsentiert			
>>	deutlich überrepräsentiert			
>	geringfügig überrepräsentiert			
=	gleichverteilt			
<	geringfügig unterrepräsentiert			
<<	deutlich unterrepräsentiert			
<<<	stark unterrepräsentiert			



**Abbildung 16:** Als Resultate der vergleichenden Sequenzierung von potentiellen funktionellen Kandidatengen in zwei Extremwertpools ( $MA-BLUP_{min}$  bzw.  $MA-BLUP_{max}$ ) wurden Sequenzchromatogramme wie die hier ausschnittsweise gezeigten erhalten und für die Abschätzung der relativen Allelfrequenzen der in den Genen identifizierten SNPs in den Pools ausgewertet. Das obere Panel (**A** und **B**) vergleicht beispielhaft die Chromatogramme eines SNP (CD79A\_-182:A>G; beachte, dass hier der Komplementärstrang gezeigt ist) mit sehr differenter Frequenzverteilung der SNP-Allele in den Pools; dieser SNP wurde in den weiteren Untersuchungen mit hoher Priorität untersucht. Das untere Panel (**C** und **D**) stellt dagegen die Chromatogramme eines SNP (HCST\_1060:A>C) mit nahezu gleicher Allelverteilung in den Pools gegenüber; dieser SNP spielt im Folgenden nur eine untergeordnete Rolle.

Farbcodierung der Nukleotide: ● A; ● C; ● G; ● T

### 2.3 Konkordanztest

Das zweite Selektionskriterium für potentiell merkmalsassoziierte SNPs in den untersuchten potentiellen funktionellen Kandidatengen der Region des BTA18-SCS-QTL war ein Test auf Konkordanz. Geprüft wurde die Hypothese, dass die SNPs in Bullen, welche heterozygote Allele des BTA18-SCS-QTL tragen, ebenfalls heterozygote Genotypen aufweisen (d.h. konkordant mit dem QTL sind). Hierfür standen vier QTL-heterozygote Testbullen ( $\alpha$ ,  $\beta$ ,  $\gamma$  und  $\delta$ ) zur Verfügung.

Nicht alle der im Rahmen der selektiven Poolsequenzierung identifizierten SNPs wurden ebenfalls in den Testbullen genotypisiert. Primär wurden solche SNPs mit differenter Allelverteilung in den Extrempools ausgewählt. Hinzu kamen SNPs, die in für die primären SNPs generierten Amplikons lokalisiert waren, so dass schließlich für 14 der 16 insgesamt identifizierten SNPs Genotypisierungsergebnisse in den Testbullen vorlagen.

Für keinen dieser 14 SNPs konnte vollständige Konkordanz, d.h. Heterozygotie in allen vier Testbullen, nachgewiesen werden. Zwei SNPs im Gen *CD79A* sind in drei der vier Bullen heterozygot, wobei es sich in beiden Fällen um die Bullen  $\alpha$ ,  $\beta$  und  $\gamma$  handelt. Jedes der vier potentiellen funktionellen Kandidatengene enthält mindestens einen SNP, der in mindestens einem der Testbullen heterozygot ist. Bei sieben SNPs wurde festgestellt, dass in den Testbullen beide entgegengesetzt homozygote SNP-Genotypen auftreten.

Tabelle 24 summarisiert die Sequenzierungsergebnisse für jedes potentiell funktionelle Kandidatengen und Tabelle 25 zeigt die einzelnen Genotypen der SNPs in den Testbullen.

**Tabelle 24:** Konkordanztest: 14 der 16 insgesamt identifizierten SNPs wurden auf ihren Allelstatus in vier Testbullen, die heterozygote Allele des BTA18-SCS-QTL tragen, getestet: Die Tabelle gibt für  $n = 1 - 4$  an, wie viele der in jedem Kandidatengen gefundenen SNPs in  $n$  Bullen einen heterozygoten Genotyp aufweist, d.h. konkordant mit dem QTL ist. Die einzelnen Genotypisierungsergebnisse sind Tabelle 25 zu entnehmen.

	heterozygote SNPs in $n$ Bullen				
	$n = 4$	3	2	1	0
<i>CD22</i>	0	0	0	2	5
<i>CD79A</i>	0	2	0	1	0
<i>HCST</i>	0	0	1	2	0
<i>TYROBP</i>	0	0	0	1	0
gesamt	0	2	1	6	5

**Tabelle 25:** Genotypisierungsergebnisse für 14 der insgesamt 16 identifizierten SNPs (die nicht genotypisierten SNPs sind mit einem \* gekennzeichnet) in vier Testbullen mit heterozygoten Allelen des BTA18-SCS-QTL. Die zu prüfende These lautete, dass potentiell mit der Eutergesundheit assoziierte SNPs in möglichst vielen der Testbullen ebenfalls heterozygot (d.h. konkordant mit dem QTL) sind. Heterozygote Genotypen sind daher fettgedruckt hervorgehoben. Durch ein Häkchen in der Spalte **AS** sind die SNPs markiert, die in der Gesamtwürdigung dieses Ansatzes und der selektiven Poolsequenzierung (Tab. 23) für die Assoziationsstudie in einer DH-Halbgeschwisterpopulation ausgewählt wurden.

SNP	Testbulle				AS
	$\alpha$	$\beta$	$\gamma$	$\delta$	
CD22_-2595:C>T*	-	-	-	-	
CD22_611:A>C	AA	AA	CC	CC	
CD22_4494:C>T	CC	CC	TT	TT	
CD22_4680:A>G	GG	<b>AG</b>	GG	GG	
CD22_7568:C>T	CC	CC	-	TT	✓
CD22_8067:A>G	GG	<b>AG</b>	GG	GG	✓
CD22_8636:C>T	TT	TT	CC	CC	
CD22_8867:A>G	AA	AA	GG	GG	
CD79A_-182:A>G	<b>AG</b>	<b>AG</b>	<b>AG</b>	AA	✓
CD79A_2808:C>T	<b>CT</b>	<b>CT</b>	<b>CT</b>	CC	✓
CD79A_3201:A>G	GG	<b>AG</b>	GG	GG	
HCST_497:A>C	AA	<b>AC</b>	CC	CC	
HCST_675:A>G	AA	AA	GG	<b>AG</b>	✓
HCST_866:C>T	CC	CC	<b>CT</b>	<b>CT</b>	✓
HCST_1060:A>C*	-	-	-	-	
TYROBP_1820:A>G	GG	GG	<b>AG</b>	GG	

## 2.4 Fazit des mehrstufigen Auswahlverfahrens zur Bestimmung potentiell merkmalsassoziiierter SNPs

Für jeden der 16 SNPs, die in den vier positionellen und potentiell funktionellen Kandidatengenomen des BTA18-SCS-QTL identifiziert wurden, lagen Datensätze zur relativen Allelverteilung in Extremwertpools vor; der Allelstatus in für den BTA18-SCS-QTL heterozygot typisierten Bullen wurde für 14 der 16 SNPs erfasst. In beiden Experimenten stachen zwei SNPs im Gen *CD79A* besonders hervor, ohne dass sie die Kriterien für die Postulierung einer Merkmalsassoziation gänzlich erfüllten; d.h. keine Dominanz gegensätzlicher Allele in den gegensätzlichen Extremwertpools, stattdessen nahezu Fixierung eines Allels in einem Pool versus Gleichverteilung im anderen. Drei der vier Testbullen mit heterozygoten Allelen des BTA18-SCS-QTL wiesen ebenfalls Heterozygotie an diesen SNPs auf; damit liegt keine vollständige Konkordanz vor. In der Gesamtschau beider Experimente und unter Berücksichtigung von noch zu diskutierenden Erwägungen wurden diese Befunde als Indizien für eine potentielle Merkmalsassoziation und damit als Rechtfertigung einer Typisierung im DH-Halbgeschwisterpanel angesehen. Dies gilt ebenso für je zwei SNPs in den Genen *CD22* und *HCST*, wobei hier die Indizienlage für eine potentielle Merkmalsassoziation mit jeweils maximal zwei QTL-heterozygoten Bullen mit heterozygoten QTL-Allelen bzw. teils geringer ausgeprägten Allelverschiebungen zwischen den Extremwertpools etwas weniger positiv als für *CD79A* ausfällt. Hier kam zusätzlich die Zielstellung zum Tragen, beide untersuchten QTL-Intervalle (5 und 9) in die angestrebte Assoziationsuntersuchung einzubeziehen; *CD79A* befindet sich in Intervall 9, beide andere Gene in Intervall 5.

Im Ganzen wurden demnach sechs (= 37,5% von 16 identifizierten) SNPs für die Typisierung im DH-Halbgeschwisterpanel selektiert: • CD22\_7568:C>T, • CD22\_8067:A>G, • HCST\_675:A>G, • HCST\_866:C>T, • CD79A\_-182:A>G und • CD79A\_2808:C>T.

Diese SNPs sind in den Tabellen 23 und 25 durch Häkchen markiert.

### 3 Merkmalsassoziationen

#### 3.1 Genotypisierung

Für die Assoziationsanalysen an den sechs ausgewählten SNPs stand eine DH-Halbgeschwisterpopulation mit insgesamt 1054 Bullen von 6 Bullenvätern, die in Kap. II.4.4 beschrieben ist, zur Verfügung. Diese Tiere, nicht aber die Bullenväter, wurden mit der Methode der Pyrosequenzierung (Kap. I.2.1.4 bzw., für die verwendeten Primer, Kap. II.5.5) an sechs der insgesamt 16 gefundenen SNPs genotypisiert. Letztlich lagen Genotypdaten für 919 Tiere vor (entspricht 87%).

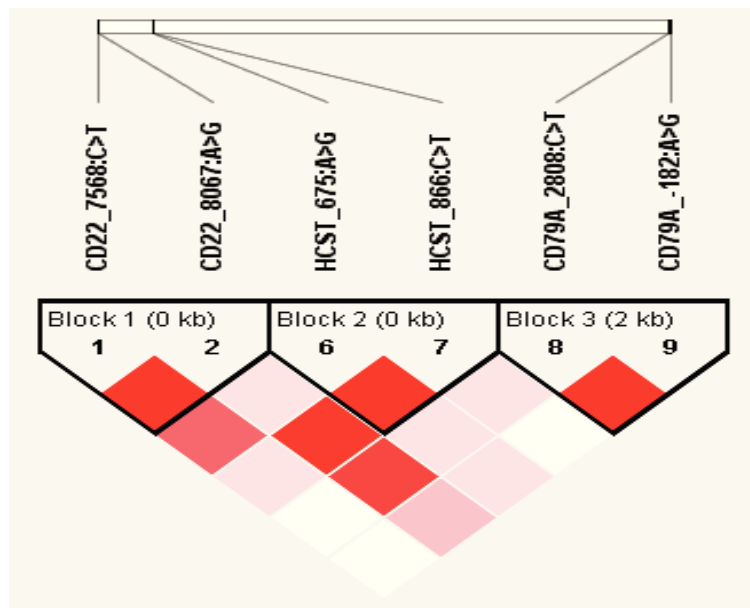
Tabelle 26 gibt einen Überblick über die Genotypisierungsergebnisse mit den beobachteten Allel- und Genotypfrequenzen an den SNPs. Die relativen Frequenzen der unterrepräsentierten Allele (*minor allele frequency, MAF*) betrugen global zwischen 0,18 und 0,36. Zwischen den einzelnen Halbgeschwisterfamilien traten dabei große Unterschiede in den MAF auf; mit zwei Ausnahmen waren aber die Allele mit global niedrigerer Frequenz auch in den Familien unterrepräsentiert.

**Tabelle 26:** Zusammenfassung der Genotypisierung von je zwei SNPs in drei potentiellen funktionellen Kandidatengen (*CD22*, *CD79A*, *HCST*) der Region des BTA18-SCS-QTL in einer DH-Halbgeschwisterpopulation ( $n = 919$ , 6 Bullenväter). Die Typisierungen erfolgten mit der Methode der Pyrosequenzierung (Kap. I.2.1.4, zu den verwendeten Primern und Laborprotokollen Kap. II.5.5). Angegeben sind für jeden SNPs das höherfrequente (**A**) und geringerfrequente (**a**) Allel, die ihnen zugehörigen beobachteten relativen Allelfrequenzen **p** und **q** sowie die beobachteten Genotypfrequenzen. Weiterhin sind die Teststatistik und der p-Wert (bei  $df = 1$ ) eines  $\chi^2$ -goodness-of-fit-Test der Nullhypothese, dass die Genotypfrequenzen den durch das Hardy-Weinberg-Gesetz (HWG, Kap. II.2.1) für die beobachteten Allelfrequenzen geforderten folgen, angegeben. Für die relativen Allelfrequenzen sind gerundete Werte angegeben, die Berechnung der  $\chi^2$ -Statistik erfolgte mit ungerundeten Werten.

SNP	Allele		rel. Allelfreq.		Genotypfreq.			HWG	
	A	a	p	q	AA	Aa	aa	$\chi^2$	p
CD22_7568:C>T	C	T	0,77	0,23	528	348	39	3,84	0,05
CD22_8067:A>G	G	A	0,82	0,18	614	271	30	0,0004	0,99
CD79A_2808:C>T	G	A	0,67	0,33	393	430	88	3,64	0,056
CD79A_-182:A>G	T	C	0,64	0,36	366	440	105	2,5	0,11
HCST_675:A>G	A	G	0,68	0,32	412	420	85	2,25	0,13
HCST_866:C>T	C	T	0,77	0,23	531	344	42	2,15	0,14

Mit dem Programm HaploView (Barrett *et al.*, 2005) wurden die typisierten SNPs auf Kopplungsungleichgewichte (*Linkage Disequilibrium*, *LD*) getestet und Haplotypblöcke bestimmt. Als Maß für LD werden die Parameter *D* (Formel 7) und *D'* (Formel 8) genutzt. Die gegebenen Verwandtschaftsstrukturen wurden hierbei ignoriert und die Bullen als unverwandt angenommen.

Starkes LD konnte jeweils zwischen den paarweise in einem Gen typisierten SNPs nachgewiesen werden (Abb. 17). Keine der in HaploView verfügbaren Algorithmen detektierte gegenüberliegende Haplotypblöcke.



**Abbildung 17:** Plot der paarweisen Kopplungsungleichgewichte (LD, *linkage disequilibrium*) zwischen den sechs in der DH-Halbgeschwisterpopulation typisierten SNPs (vgl. Tab. 26). LD wurde unter Nichtberücksichtigung der Verwandtschaftsstrukturen auf Basis von Lewontins *D'* (Lewontin (1964), Formeln 7 und 8) ermittelt (im Plot farblich kodiert; je intensiver der Rotton, desto höher das LD) und Haplotypblöcke mit der Konfidenzintervall-Methode von Gabriel *et al.* (2002) bestimmt. Beachte, dass *CD79A* auf dem *antisense*-Strang kodiert ist; daher liegt der SNP *CD79A\_-182:A>G* telomer. Für die Berechnungen und die Erzeugung des Plots wurde die Software HaploView (Barrett *et al.*, 2005) verwendet.

Die *CD22*-SNPs fallen auf, da sie trotz geringer physikalischer und genetischer Distanz einen sehr großen Unterschied in der Übereinstimmung der erwarteten mit den beobachteten Genotypfrequenzen (Hardy-Weinberg-Gleichgewicht, HWG) zeigen (Tab. 26); die anderen SNPs weichen in dieser Beziehung weniger voneinander ab.

Dies ist zumindest teilweise darauf zurückzuführen, dass nicht alle Bullenväter für alle SNPs heterozygot waren. Dies kann zumindest für diejenigen Väter deduziert werden, unter deren Söhnen Homozygote eines Allels fehlten; reale Typisierungen der Väter an den SNPs lagen jedoch nicht vor.

In Ermangelung parentaler Genotypen konnte die Berechnung des HWG nicht auf die maternal vererbten Allele, wie z.B. von Flori *et al.* (2009) praktiziert, beschränkt werden.

Die für die Nutzung in HaploView nötige Konvertierung der Genotypisierungsdaten in das LINKAGE-Format wurde mittels eines Perlskriptes mit Anbindung an die für die Verwaltung der Daten eingesetzte MySQL-Datenbank realisiert.

## 3.2 Modellierung

In Tabelle 36 (im Anhang) werden für die zur Verfügung stehende DH-Halbgeschwisterpopulation Phänotypen für Eutergesundheit, hauptsächlich repräsentiert durch das Merkmal SCS in drei aufeinanderfolgenden Laktationsperioden (vgl. Kap. I.1.4) sowie Konformations- und Milchleistungsphänotypen (einige von letzteren sind in der abgebildeten Tabelle 40 im Anhang näher beschrieben) zusammen mit populationspezifischen Kenngrößen vorgestellt.

Mithilfe gemischter linearer Modelle wurden Assoziationen zwischen den in der Halbgeschwisterpopulation typisierten SNPs (siehe voriges Kapitel) und den mittleren Töchterabweichungen (DYD, *Daughter Yield Deviation*), die durch die EDC (*Effective Daughter Contribution*) gewichtet wurden (Kap. I.1.4) berechnet.

Hierbei wurden die drei untersuchten Gene jeweils separat betrachtet und auf jedes Gen zwei verschiedene gemischte lineare Modelle angewendet, welche neben einem zufälligen polygenen Effekt die je zwei SNPs im Gen in unterschiedlichen Faktorkombinationen als fixe Effekte berücksichtigten (Kap. II.2.4, speziell die Box in Abb. 5). Die Berechnungen wurden mit der Software ASReml (Gilmour *et al.*, 2006) durchgeführt.

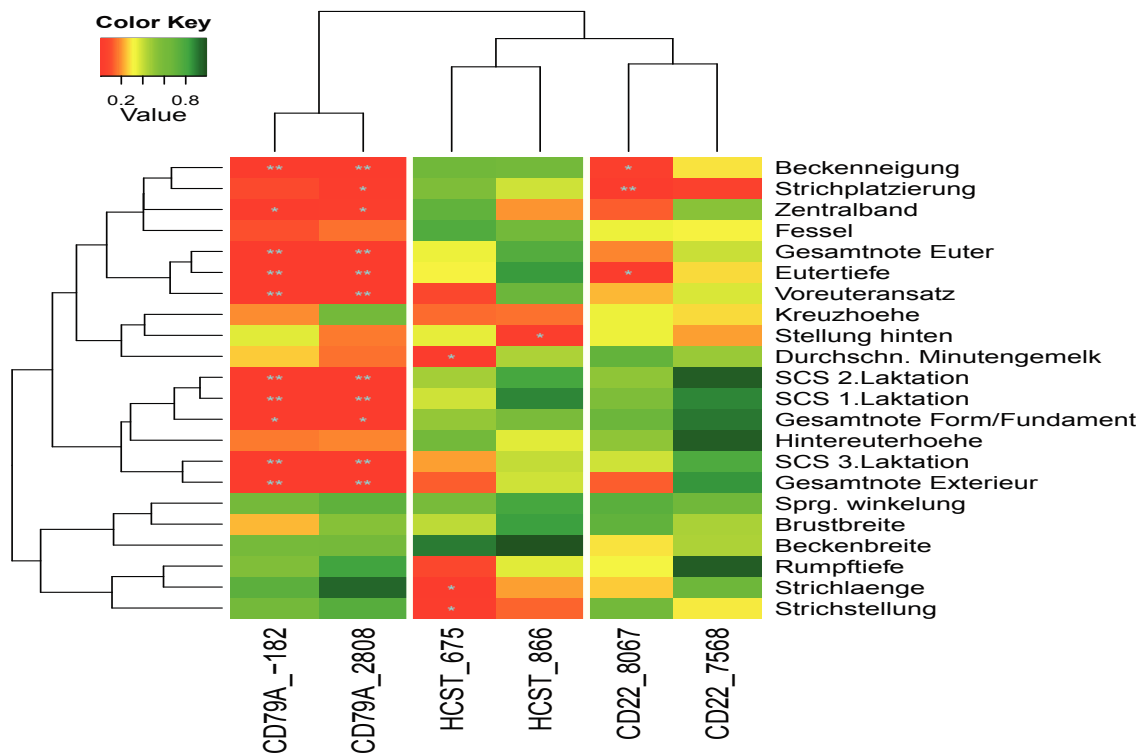
Zunächst wurden die SNPs einzeln als fixe Effekte in das gemischte lineare Modell eingesetzt (Singuläres Modell, Formel 10).

Für zwei SNP-Phänotyp-Paare, CD79A\_-182:A>G und SCS in der 2. Laktation bzw. HCST\_866:C>T und SCS in der 1. Laktation, konnte durch die Software der Denominator-Freiheitsgrad nicht korrekt ermittelt werden und wurde daher manuell bestimmt.

Abbildung 18 gibt einen Überblick über die Ergebnisse anhand einer Cluster-*Heatmap* der Matrix der Irrtumswahrscheinlichkeiten  $p$  der Merkmalsassoziationen. Diese sind darin farbkodiert, wobei intensive Rottöne einen niedrigen  $p$ -Wert anzeigen. Für die exakten  $p$ -Werte sei auf Tab. 38 im Anhang verwiesen.

Das Spaltendendrogramm ordnet die genweisen SNP-Paare jeweils separaten Clustern zu; die Distanz der SNPs in *CD22* und *HCST* untereinander ist geringer als zu *CD79A*-SNP. Damit folgt das Muster der SNP-Assoziationen der genomischen Organisation der SNPs und Gene.





**Abbildung 18:** Cluster-*Heatmap* der Resultate der SNP-Assoziationsberechnungen nach dem Einzel-SNP-Modell (gemischtes lineares Modell mit dem SNP als fixen Effekt nach Formel 10). Dargestellt sind die Irrtumswahrscheinlichkeiten  $p$  für eine Assoziation jedes von sechs SNPs mit der gewichteten mittleren Töchterabweichung (DYD, *daughter yield deviation*) von 22 Phänotypen für Eutergesundheit, Konformation und Milchleistung. Zur Interpretation des Farbspektrums siehe die Legende oben links. Hochsignifikante ( $p \leq 0,01$ ) bzw. signifikante ( $p \leq 0,05$ ) Assoziationen sind mit \*\* bzw. \* speziell gekennzeichnet. Die der Darstellung zugrundeliegenden genauen Werte sind der Tabelle 38 im Anhang zu entnehmen. Die Dendrogramme für Spalten (SNPs) und Zeilen (Phänotypen) wurden durch agglomeratives hierarchisches Clustering auf Basis der paarweisen euklidischen Distanzen (Formel 9) ermittelt. Die Heatmap wurde mit der Funktion *heatmap.2* aus dem R-Paket *gplots* erstellt.

SNP-Merkmalspaare mit hochsignifikanten ( $p \leq 0,01$ ) bzw. signifikanten ( $p \leq 0,05$ ) Assoziationen sind in der Abbildung neben der Farbkodierung durch Asterisken speziell gekennzeichnet. Diese werden in den folgenden Abschnitten detaillierter untersucht. Insbesondere werden in den Tabellen 27 und 28 neben verschiedenen Modellparametern die absoluten Abweichungen (Effektgrößen) der jeweiligen homozygoten Genotypen vom heterozygoten Referenzgenotyp, deren Standardfehler und die nach Formel 12 berechneten additiven genetischen Effekte der SNPs (in Relation zum Residuenstandardfehler des Phänotyps nach Formel 13) angegeben.

Für die *CD79A*-SNPs werden die signifikanten Effekte zusätzlich graphisch in Relation zu den jeweiligen Residuenstandardfehlern der Phänotypen (vgl. Formel 13 und Tabelle 36 im Anhang) in Abb. 19 dargestellt.

Das gegebenenfalls zusätzlich berechnete Modell (Formel 11) enthielt beide SNPs per Gen als fixe Effekte. Die Varianzdekomposition erfolgte in diesen Fällen mittels der konditionalen Methode (Typ III, vgl. Kap. II.2.4).

Die F-Statistiken, deren Freiheitsgrade und die daraus resultierenden Irrtumswahrscheinlichkeiten, die für die SNPs in den Tabellen 27 und 28 angegeben werden, sind die Werte, die für den entsprechenden SNP verbleiben, nachdem der Effekt des jeweils anderen SNP im selben Gen nach oben dargestelltem Prinzip eliminiert wurde.

Im folgenden wird jeweils das SNP-Allel, welches in der hier behandelten Halbgeschwisterpopulation die höhere Frequenz aufweist, als Wildtypallel bezeichnet.

### 3.2.1 CD22

Keiner der beiden SNPs CD22\_7568:C>T und CD22\_8067:A>G zeigte eine Assoziation mit dem Merkmal SCS in einer der drei untersuchten Laktationsperioden. Von vier mit der Euterbeschaffenheit in Verbindung stehenden Phänotypen, mit denen CD22\_8067:A>G eine signifikante Assoziation aufweist, wurde nur für einen, die Strichplatzierung, eine tendenzielle Assoziation ( $p = 0,056$ ) mit CD22\_7568:C>T gefunden. Für diesen Phänotyp zeigen die jeweiligen homozygoten Wildtypvarianten gegensätzliche Allelsubstitutionseffekte gegenüber dem heterozygoten Referenzgenotyp, nämlich bei CD22\_7568:C>T, mit C als Wildtypallel, in numerisch negative, bei CD22\_8067:A>G, mit G als Wildtypallel dagegen in numerisch positive Richtung. Für die homozygoten Nichtwildtypallele gilt entsprechend die umgekehrte Aussage. Der additive genetische Effekt des SNP CD22\_8067:A>G auf das Merkmal Strichplatzierung ist mit einer Größe von 0,31 Standardabweichungen (genauer: des Residuenstandardfehlers des linearen Modelles aus Formel 13) der größte in dieser Untersuchung beobachtete.

**Tabelle 27:** Die SNPs CD22\_7568:C>T und CD22\_8067:A>G bzw. HCST\_675:A>G und HCST\_866:C>T wurden jeweils einzeln und additiv (Formeln 10 und 11) als fixe Effekte in gemischten linearen Modellen auf ihre Assoziation mit verschiedenen Phänotypen in einer Halbgeschwisterpopulation der Milchviehrasse Deutsche Holstein untersucht; zu einer Übersicht über die Signifikanzen siehe Abb. 18 und Tab. 38 (im Anhang). Nur für im Einzel-SNP-Modell signifikante SNP-Phänotyp-Assoziationen gibt diese Tabelle die errechneten **F**-Statistiken (die im additiven Modell durch konditionale Varianzdekomposition bestimmt wurden, daher  $F_{con}$ ), die Denominator-Freiheitsgrade **DF** und die Signifikanzen der **F**-Werte bei 2 und **DF** Freiheitsgraden. Zusätzlich sind für das Einzel-SNP-Modell die Schätzwerte der homozygoten Genotypen relativ zum heterozygoten Referenzgenotyp (**Estimate**) sowie ihre Standardfehler (**SE**) und schließlich der sich daraus ergebende additive genetische Effekt (= der Hälfte der Differenz der Homozygoten-Effekte) angegeben.

Phänotyp	SNP	Einzel			Additiv			Homozygot WT		Homozygot NWT		Add. GE
		DF	F	p	DF	$F_{con}$	p	Estimate	SE	Estimate	SE	
Strichplatzierung	7568:C>T	756,4	2,92	0,056	757,4	1,28	0,279	-59,15	25,18	26,73	53,75	0,13
	8067:A>G	808,8	7,58	0,001	801,4	5,92	0,003	69,12	24,22	-125,23	59,92	0,31
Beckenneigung	7568:C>T	781,1	1,22	0,295	777,6	1,37	0,256	-20,21	25,71	-75,88	55,19	0,08
	8067:A>G	824,4	3,26	0,041	821,1	3,40	0,036	43,14	24,79	-93,41	61,39	0,21
Eutertiefe	7568:C>T	785,7	1,25	0,289	787,5	2,48	0,086	38,72	25,84	-25,39	55,82	0,1
	8067:A>G	841,2	3,69	0,027	836,2	4,93	0,008	31,79	24,97	-132,12	61,94	0,25
	HCST...	DF	F	p	DF	$F_{con}$	p	Estimate	SE	Estimate	SE	
DMG	675:A>G	543,0	4,03	0,020	543,3	3,64	0,028	-0,02	0,01	-0,04	0,02	0,08
	866:C>T	544,4	0,80	0,454	543,3	0,43	0,653	-0,01	0,01	-0,02	0,03	0,03
Stellung hinten	675:A>G	333,8	1,03	0,358	326	0,15	0,863	38,83	27,18	-3,21	49,39	0,04
	866:C>T	332,5	3,28	0,040	323,3	2,38	0,096	63,00	26,42	89,38	67,15	0,03
Strichstellung	675:A>G	225,8	3,67	0,028	231,9	1,97	0,144	-19,00	36,74	-184,17	68,48	0,16
	866:C>T	229,3	1,92	0,150	232,1	0,30	0,744	-19,72	37,05	-192,90	98,98	0,17
Strichlänge	675:A>G	895,1	4,01	0,020	867,4	2,78	0,065	8,72	24,70	116,78	41,23	0,17
	866:C>T	893,9	1,53	0,218	864,9	0,30	0,740	10,10	24,44	94,53	54,15	0,13

### 3.2.2 CD79A

Für beide SNPs in diesem Gen, die in der DH-Halbgeschwisterpopulation typisiert wurden, konnten in der Einzel-SNP-Analyse signifikante Assoziationen mit der mittleren Töchterabweichung (DYD) des Merkmals SCS in den ersten drei Laktationen festgestellt werden.

Der Promotor-SNP CD79A\_-182:A>G zeigt in allen Laktationen einen numerisch positiven Allelsubstitutionseffekt der homozygoten Wildtypvariante im Vergleich mit der heterozygoten Variante, der aber mit jeder Laktation kleiner wird. Im Gegensatz dazu ist der Effekt der homozygoten Nichtwildtypvariante numerisch stets negativ und steigt von der ersten bis zur dritten Laktation.

Ein ähnliches Muster zeigt der Intron-SNP CD79A\_2808:C>T. Die ermittelten F-Statistiken für diesen SNP sind jedoch um einiges größer als für CD79A\_-182:A>G. Die Differenz zwischen den Allelsubstitutionseffekten beider SNPs ist für den homozygoten Nichtwildtyp-Genotyp größer als für den homozygoten Wildtyp.

Das Maximum der ermittelten additiven genetischen Effekte für das Merkmal SCS wurde für den SNP CD79A\_2808:C>T in der dritten Laktation mit einem Wert von  $\approx 0,21$  Einheiten des Residuenstandardfehlers ermittelt.

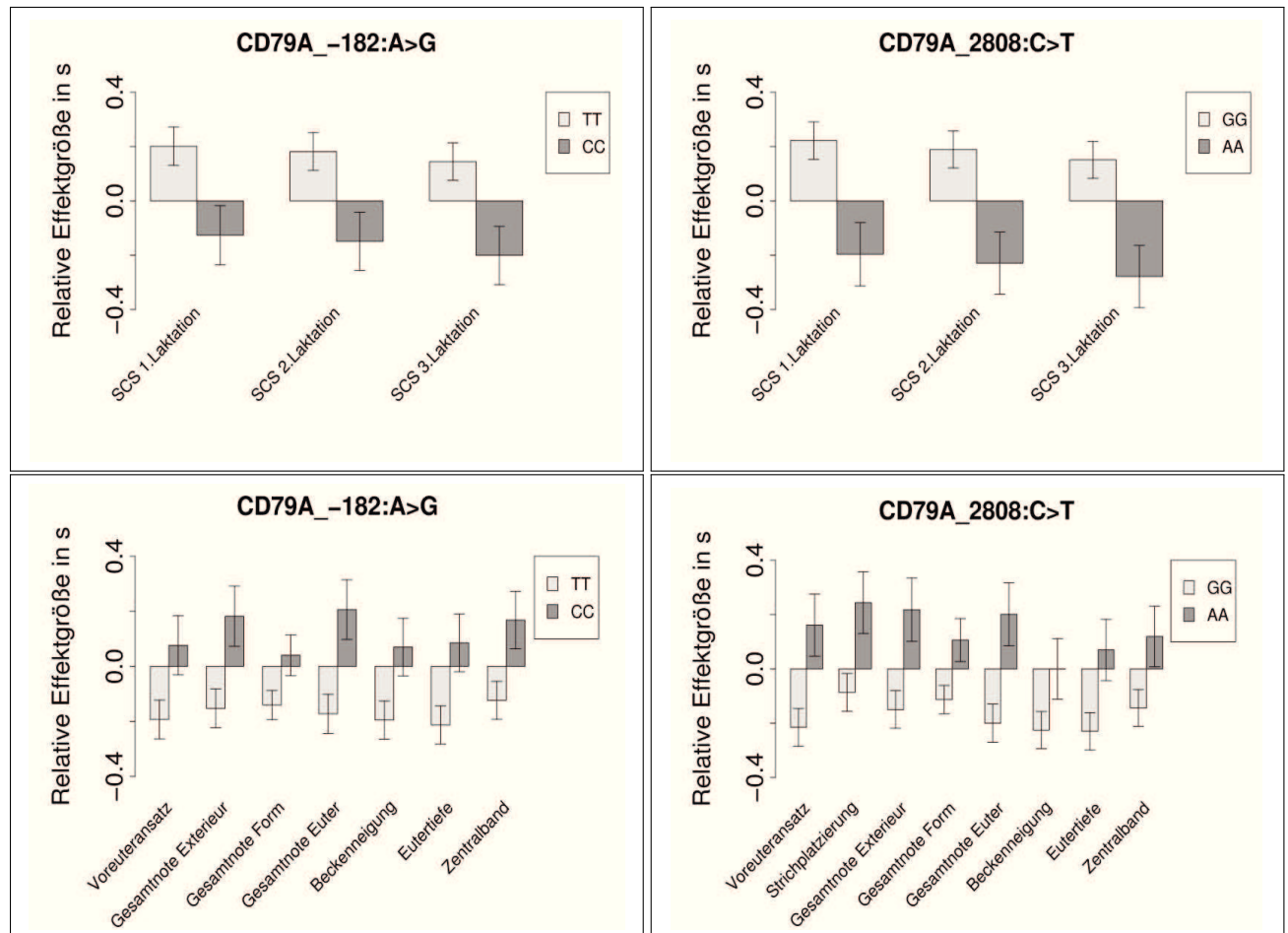
Für das Merkmal SCS in der 2. Laktation konnte der Denominatorfreiheitsgrad des F-Testes durch ASReml numerisch nicht korrekt ermittelt werden und wurde deshalb geschätzt; auf dieser Schätzung beruht die Irrtumswahrscheinlichkeit.

Weiteren Aufschluss über die Beziehung der beiden SNPs zum Merkmal SCS gibt die Analyse des additiven Modells (Formel 11). Hier zeigte sich ein Gegensatz in den statistischen Aussagen bezüglich der ersten und den beiden folgenden Laktationsperioden. Das additive Modell erklärt signifikant mehr der phänotypischen Varianz des Merkmals SCS in der ersten Laktation als ein reduziertes Modell, das nur den Effekt des SNP CD79A\_-182:A>G beinhaltet. Dies kann aus dem Wert für  $F_{con}$  des SNP CD79A\_2808:C>T geschlossen werden, der mit 3,14 höher liegt als der Schwellenwert von 3,005 für eine 5%ige Irrtumswahrscheinlichkeit einer F-Verteilung bei den gegebenen Freiheitsgraden (2 und 904). Dagegen kann die Nullhypothese, dass jeder der beiden SNPs alleine die phänotypische Varianz des Merkmals SCS in den beiden folgenden Laktationen erklärt, nicht zurückgewiesen werden; die Werte für  $F_{con}$  liegen hier für CD79A\_2808:C>T (2. Laktation: 2,57, 3. Laktation: 2,22) weit höher als für CD79A\_-182:A>G (0,64 und 0,38), jedoch unterhalb des Schwellenwertes.

Der SNP CD79A\_2808:C>T weist ebenfalls signifikante Assoziationen mit den mittleren Töchterabweichungen für acht Konformationsphänotypen auf, darunter mehrere Parameter der Euterform und Gesamtbewertungen für Euter, Exterieur und Form/Fundament. Abgesehen vom Merkmal Strichplatzierung ( $p = 0.091$ ) wurden signifikante Assoziationen mit diesen

Merkmalen auch für den SNP CD79A\_-182:A>G berechnet. In allen Fällen hatte der homozygote Nichtwildtyp-Genotyp einen numerisch positiven, der homozygote Wildtyp-Genotyp dagegen einen numerisch negativen Effekt auf den Phänotyp,

Für sieben dieser Phänotypen gilt, dass ein reduziertes Modell, welches jeweils nur einen der SNPs erhält, nicht signifikant weniger der phänotypischen Varianz erklärt als das additive Modell, das die Effekte beider SNPs enthält. Die Ausnahme bildet das Merkmal Voreuteransatz.



**Abbildung 19:** Relative Schätzwertgrößen für die Effekte der homozygoten Genotypen der SNPs CD79A\_-182:A>G (links) und CD79A\_2808:C>T (rechts) auf die mittlere Töchterabweichung des Merkmals SCS der ersten drei Laktationen (oben) und mehrere Konformationsphänotypen (unten) in einer Halbgeschwisterpopulation der Milchviehrasse Deutsche Holstein, in Relation zum Residuenstandardfehler des Modells in Gleichung 13. Die zugrundeliegenden absoluten Schätzwerte sind Tab. 28 und die Residuenstandardabweichungen Tab. 36 (im Anhang) zu entnehmen.

**Tabelle 28:** Die SNPs CD79A\_-182:A>G und CD79A\_2808:C>T wurden jeweils einzeln und additiv (siehe Formeln 10 und 11) als fixe Effekte in gemischten linearen Modellen auf ihre Assoziation mit der mittleren Töchterabweichung (DYD) der Merkmale Somatische Zellzahl in der Milch (SCS) der Laktationen eins bis drei sowie verschiedenen Konformations- und Milchleistungsphänotypen in einer Halbgeschwisterpopulation der Milchviehrasse Deutsche Holstein untersucht. Die Tabelle zeigt für beide Modelle die errechneten **F**-Statistiken (die im additiven Modell durch konditionale Varianzdekomposition bestimmt wurden, daher  $F_{con}$ ), die Denominator-Freiheitsgrade **DF** und die Signifikanzen der **F**-Werte bei 2 und **DF** Freiheitsgraden. Zusätzlich sind für das Einzel-SNP-Modell die absoluten Allelsubstitutionseffekte der homozygoten Genotypen relativ zum heterozygoten Referenzgenotyp (**Estimate**) sowie ihre Standardfehler (**SE**) und schließlich der sich daraus ergebende additive genetische Effekt (= der Hälfte der Differenz der Homozygoten-Effekte) angegeben. Aufgeführt sind nur die elf Phänotypen, darunter der SCS in allen drei untersuchten Laktationen, für die signifikante Assoziationen mit mind. einem der SNPs gefunden wurden.

Phänotyp	SNP	Einzel			Additiv			Homozygot WT		Homozygot NWT		Add. GE
		CD79A...	DF	F	p	DF	$F_{con}$	p	Estimate	SE	Estimate	
SCS 1. Laktation	-182:A>G	906,0	5,82	0,004	904,0	0,81	0,440	0,07	0,03	-0,05	0,04	0,16
	2808:C>T	906,0	8,19	0,001	904,0	3,14	0,040	0,08	0,03	-0,07	0,04	0,21
SCS 2. Laktation	-182:A>G			0,005	904,0	0,64	0,530	0,07	0,03	-0,06	0,04	0,17
	2808:C>T				904,0	2,57	0,079	0,08	0,03	-0,09	0,05	0,21
SCS 3. Laktation	-182:A>G	906,0	5,13	0,007	904,0	0,38	0,690	0,06	0,03	-0,09	0,05	0,17
	2808:C>T	906,0	7,00	0,001	904,0	2,22	0,111	0,07	0,03	-0,12	0,05	0,22
Voreuteransatz	-182:A>G	844,1	4,73	0,010	820,6	1,58	0,210	-58,35	21,32	23,06	32,46	0,14
	2808:C>T	845,1	7,20	0,001	820,0	4,02	0,021	-65,02	20,94	48,69	34,65	0,19
Strichplatzierung	-182:A>G	791,9	2,43	0,091	761,4	1,07	0,344	-15,30	22,63	63,6	34,01	0,12
	2808:C>T	790,1	3,89	0,022	760,6	2,52	0,083	-27,62	22,24	77,7	36,31	0,17
Gesamtnote Ext	-182:A>G	903,0	4,84	0,009	901,0	0,10	0,906	-123,65	57,19	147,75	88,57	0,17
	2808:C>T	903,0	5,24	0,006	901,0	0,49	0,614	-121,33	56,28	176,56	94,7	0,18
Gesamtnote Form	-182:A>G	291,8	4,21	0,017	305,8	1,82	0,166	-16,82	6,35	4,81	8,89	0,09
	2808:C>T	287,2	4,08	0,019	301,9	1,70	0,186	-13,58	6,28	12,73	9,52	0,11
Gesamtnote Euter	-182:A>G	845,1	6,21	0,003	822,5	0,21	0,815	-11,85	4,89	14,15	7,43	0,19
	2808:C>T	845,1	7,10	0,001	821,9	1,11	0,338	-13,73	4,81	13,8	7,95	0,2
Beckenneigung	-182:A>G	814,5	4,92	0,009	787,1	0,72	0,491	-64,24	22,84	23,02	34,45	0,13
	2808:C>T	815,6	5,79	0,004	786,4	1,58	0,208	-74,35	22,46	-0,12	36,83	0,11
Eutertiefe	-182:A>G	824,6	5,95	0,003	798,2	0,05	0,953	-70,4	23,01	28,3	34,86	0,15
	2808:C>T	826,1	6,67	0,002	797,5	0,76	0,471	-76,04	22,63	23,03	37,28	0,15
Zentralband	-182:A>G	829,1	3,83	0,024	804,6	0,80	0,454	-32,72	18,22	44,56	27,62	0,15
	2808:C>T	831,6	3,53	0,032	803,9	0,50	0,611	-38,07	17,95	31,54	29,59	0,13

### 3.2.3 HCST

Keiner der beiden im Gen *HCST* typisierten SNPs zeigte eine signifikante Assoziation mit dem Merkmal SCS in einer der ersten drei Laktationsperioden.

Für den SNP HCST\_675:A>G wurden signifikante Einzeleffekte auf drei Phänotypen, davon zwei mit direktem Bezug zur Zitze, nämlich Strichstellung und Strichlänge, gefunden; wobei jeweils nur der homozygote (nach der hier benutzten Definition) Nichtwildgenotyp GG einen deutlich von Null verschiedenen Effekt zeigt. Dieser ist in den beiden Phänotypen gegenläufig: negativ für Strichstellung und positiv für Strichlänge, jeweils mit im Mittel  $\approx 0,4$  Einheiten des Residuenstandardfehlers des Modells 13.

Dagegen konnte für den SNP HCST\_866:C>T lediglich ein Effekt auf den einzelnen Phänotyp „Stellung hinten“ gefunden werden, der formal signifikant, jedoch sehr klein ist.

## Teil IV

# Diskussion

Für die Identifizierung merkmalsbeeinflussender Gene bzw. deren Sequenzpolymorphismen im Nutztier haben Ron & Weller (2007) die Devise *...winning by points rather than knockout* ausgegeben.

Sie soll zum Ausdruck bringen, dass in Ermangelung spezialisierter experimenteller Populationen wie transgener Inzuchtstämme (beispielsweise *-sic!*- Knockout-Mutanten) und deren Kreuzungen, die bei Modellorganismen potentiell verfügbar sind, der Nutztierforscher meist auf die im regulären Zuchtbetrieb vorhandenen Populationsstrukturen angewiesen ist. Jede von diesen ist inhärent weniger aussagekräftig, um dies zu kompensieren, müssen für ein vorgeschlagenes potentielles kausales Kandidatengen bzw. Sequenzpolymorphismus unabhängige Indizien zusammengetragen werden.

In dieser Arbeit wird ein Ansatz vorgestellt, der in obigem Sinne eine hypothesengestützte Priorisierung positioneller Kandidatengene auf Basis von funktioneller Annotation quantitativ mit Ergebnissen einer ebenfalls hypothesenbasierten hochdimensionalen Transkriptomstudie integriert. Für die mittels dieses bioinformatischen Protokolles ermittelten potentiell funktionellen Kandidatengene eines QTL für den Gehalt somatischer Zellen in der Milch (SCS-QTL) beim Deutschen Holsteinrind (DH) wurden daraufhin in unabhängigen Studienpopulationen dieser Rasse, teilweise erfolgreich, Indizien für die Assoziation von Einzelnucleotidpolymorphismen (SNPs) in diesen Genen mit Merkmalen der Eutergesundheit erbracht.

Der erwähnte SCS-QTL ist auf der distalen Hälfte des bovinen Chromosoms 18 (BTA18, bovine Autosomen sind akrozentrisch) lokalisiert und gehört zu den am häufigsten und in verschiedenen Milchviehrassen dokumentierten merkmalsbeeinflussenden Regionen für SCS und damit für die Eutergesundheit (z.B. Ashwell *et al.*, 1997; Schrooten *et al.*, 2000; Schulman *et al.*, 2004; Xu *et al.*, 2006).

Die Region dieses QTL ist im Vorfeld der hier vorgestellten Untersuchungen, ebenfalls im Rahmen des Verbundprojektes M.A.S.-Net, durch Anwendung einer kombinierten LA/LD-Analyse feinkartiert worden (Kap. I.2.1.3 und Baes *et al.*, 2009, vgl. auch Meuwissen *et al.*, 2002).

Ein QTL für Milchproduktion auf BTA6 konnte durch eine analoge Methodik auf einen Bereich von 7,5 cM eingegrenzt werden (Olsen *et al.*, 2004). Dieser Genombereich enthält  $\approx 30$  Gene<sup>25</sup>; die Autoren verweisen auf die Notwendigkeit einer noch genaueren Feinkartierung,

---

<sup>25</sup>gemäß der aktuellen Ensembl-Annotation



um die Identifizierung der kausalen Mutation zu ermöglichen<sup>26</sup>.

Im Gegensatz dazu umfassen die ca. 6,3 cM der beiden hier untersuchten Intervalle mit lokalen Maxima des LRT mehr als 250 Gene. Diese sehr hohe Zahl positioneller Kandidatengene des QTL erforderte die Anwendung bioinformatischer Methoden zur Priorisierung experimentell zu bearbeitender Gene.

## 1 Der bioinformatische Ansatz

Im Zusammenhang mit dem zu diesem Zweck hier entwickelten heuristischen Ansatz sollen im folgenden einige Punkte speziell betrachtet werden.

Zunächst beruht er auf der Nutzung der humanen Genkarte und des für humane Gene vorliegenden Korpus wissenschaftlicher Literatur anstelle der entsprechenden originalen bovinen Daten. Dies war motiviert durch die nicht in großem Umfang und/oder erst spät verfügbare Genom- und Literaturinformation, bovine Gene betreffend. Beispielsweise liegt in der in dieser Arbeit verwendeten Version der Literaturdatenbank Gene2PubMed der Median der pro humanem Gen annotierten wissenschaftlichen Publikationen bei 5 mit einem oberen Quartil von 13 und einem Maximum von 2716, während die entsprechenden Werte für bovine Gene bei lediglich 2, 2 und 112 liegen.

Durch Brunner *et al.* (2003) sowie in anderen Publikationen (vgl. Kap. I.2.1.2) wurde eine Synteniebeziehung zwischen dem distalen Bereich von BTA18 und dem langen Arm des humanen Autosoms 19 (HSA19q) gefunden. Daher kann es als valide Arbeitsgrundlage betrachtet werden, die Gene, welche den QTL-Intervall-begrenzenden Markern im bovinen Genom am nächsten liegen, in die humane Genkarte von HSA19q einzusetzen und die resultierende, durch diese Gene flankierte Liste positioneller Kandidatengene als Annäherung an die Verhältnisse beim Rind anzusehen. Für beide auf diese Weise etablierten humanen Genlisten wird in Abbildung 8 gezeigt, dass die große Majorität der dort annotierten Gene, mit der Einschränkung, dass nur Gene mit einer EntrezGeneID berücksichtigt wurden, tatsächlich über ein bovines Ortholog auf BTA18 verfügen. Hierbei erscheint die Reihenfolge der Gene in beiden Spezies hochgradig konserviert, mit den Ausnahmen einer ca. 20 Gene umfassenden (in Intervall B) sowie einiger kleinerer Inversionen. In einigen Fällen konnte für mehrere aufeinanderfolgende Gene kein bovines Ortholog auf BTA18 identifiziert werden (z.B. an den Flanken der größeren Inversion); diesen liegen meist Cluster paraloger Gene zugrunde, von denen HSA19q sehr viele aufweist (u.a. Grimwood *et al.* (2004) und die Literaturübersicht), und die teilweise erst nach der Trennung der evolutionären Linien, die zu Primaten bzw. Wiederkäuern führten, evolvierten.

---

<sup>26</sup>Dies wurde durch Olsen *et al.* (2005) realisiert.

Ein weiterer wichtiger Punkt betrifft die Auswahl der Literatur, die den Annotationskorpus für jedes positionelle Kandidatengen bereitstellt. Prinzipiell steht dafür die Gesamtheit der in MEDLINE indizierten Publikationen zur Verfügung. Für eine gegebene Publikation einen qualitativen Zusammenhang mit einem Gen, für dessen Beschreibung die Publikation relevant ist, herzustellen, ist jedoch mit rein computergestützten Verfahren schwierig und erfordert z.B. Methoden des maschinellen Lernens, wie sie etwa in von Nakazato *et al.* (2009) angewendet wurden. Stellvertretend für die auftretenden Probleme sei hier auf die *Gene Symbol Disambiguation*, *GSD*, hingewiesen, also die Entscheidung, welche der möglicherweise vorhandenen synonymen<sup>27</sup> bzw. homonymen<sup>28</sup> Bedeutungen eines gegebenen Gensymbols tatsächlich gemeint ist (Chen *et al.*, 2005; Xu *et al.*, 2007).

Ein alternativer Ansatz, der dieses und andere Probleme umgeht, besteht in der Beschränkung auf manuell, durch Experten kuratierte Datenbanken von Gen-Literaturbeziehungen wie die Gene2PubMed-Datenbank, welche in dieser Arbeit und von anderen, beispielsweise Chen *et al.* (2007), genutzt wurde. Der Median der Zahl von in Gene2PubMed identifizierten Referenzpublikationen für die Stichprobe der positionellen Kandidaten lag etwas höher (7 vs. 5) im Vergleich zur Gesamtheit der in dieser Datenbank annotierten humanen Gene. In der Gruppe der Gene mit annotiertem bovinen Ortholog (s.o.) wurden hochsignifikant mehr Publikationen pro Gen gefunden als in der Komplementärgruppe (Daten nicht gezeigt); für die *de facto* letztlich zu berücksichtigenden positionellen Kandidatengene erhöht dies den Umfang des verfügbaren Annotationskorpus.

Die für jedes positionelle Kandidatengen durch skriptgesteuerte Auswertung der Datenbank Gene2PubMed gefundenen Referenzpublikationen wurden unter Nutzung einer NCBI-seitigen Programmierschnittstelle wiederum skriptbasiert heruntergeladen, in einem Format (MEDLINE), das neben den Zusammenfassungen (*abstracts*) u.a. die MeSH-Annotationen, nicht aber den Volltext, enthielt. Die MeSH-Terme wurden für den skriptgesteuerten Abgleich mit den Termen der hypothesenbasierten Stichwortliste genutzt.

In Übereinstimmung mit der Arbeitshypothese, die die Resistenz gegenüber Mastitis als Funktion immunologischer Parameter beschreibt, ohne diese näher zu charakterisieren, wurden die 40 Stichworte, die in Tabelle 14 gezeigt sind, so gewählt, dass sie ein möglichst breites Spektrum immunologischer Prozesse abdecken. Es sind sowohl Begriffe, die der adaptiven als auch der angeborenen, der zellulären als auch der humoralen Immunität zuzuordnen sind, enthalten. Die Skala reicht dabei von sehr allgemein gehaltenen Begriffen wie *immun* oder *Inflamm* bis zu hochspezifischen wie *Germinal center*, wobei der Grad der Allgemeinheit potentiell mit der Anzahl thematisch verwandter MeSH-Terme, die von diesem Stichwort

---

<sup>27</sup>verschiedene Symbole beschreiben dasselbe Gen

<sup>28</sup>ein Symbol dient der Identifizierung verschiedener Gene

erfasst werden und damit potentiell dem Anteil dieses Stichwortes am Nenner der Formel 3, korreliert.

Allgemeinere Stichworte sind demnach im Wesentlichen für die grobe Strukturierung der aus der Anwendung von Formel 3 auf die positionellen Kandidatengene resultierenden Rangliste potentieller funktioneller Kandidatengene verantwortlich, während die spezifischeren Stichworte eher an der Feinsortierung beteiligt sind.

An dieser Stelle muss darauf hingewiesen werden, dass die Zahl der erfassbaren MeSH-Terme gerade bei den spezifischeren Stichworten nicht nur durch ihre Spezifität, sondern auch dadurch eingeschränkt wurde, dass das Prinzip, nach dem ein Stichwort auf ein minimales, nicht sinnentstellendes Fragment reduziert werden sollte, nicht in jedem Fall stringent eingehalten wurde (wie im Falle von *Anti-bacterial*, zu dem thematisch eng verwandte, aber von diesem Stichwort nicht erfasste MeSH-Terme existieren).

Wie Tabelle 13 zeigt, konnte für 129 der 254 identifizierten humanen positionellen Kandidatengen für Eutergesundheit des BTA18-SCS-QTL ein datenbankbasierter Rangkoeffizient (DbR-Koeffizient)  $\neq 0$  errechnet werden, d.h. die Analyse der manuell kuratierten Datenbank für Referenzliteratur Gene2PubMed führte zu einer Annotationsquote durch die hypothesenbasierten Stichwortterme von 50,8%. Dieser Wert ist niedriger, aber nicht unmittelbar vergleichbar mit der MeSH-Annotationsquote von 60,1%, welche für eine genomweite Studie durch Nakazato *et al.* (2008) angegeben wird, da jene sich auf die Gesamtheit aller MeSH-Terme, nicht auf eine hypothesenbasierte Teilmenge, bezieht.

Andererseits muss hier aufgrund der bekannten Identität des langen Armes von HSA19 (HSAq) als immunologischer *HotSpot* (Hannan *et al.*, 2010) auch eine nach oben gerichtete Verzerrung der erwarteten Annotationsquote durch Terme immunrelevanter Mechanismen, wie sie gemäß der in dieser Arbeit verfolgten immunologischen Hypothese der Eutergesundheit zum Einsatz kamen, einkalkuliert werden. Verkompliziert wird dies dadurch, dass es sich, wie in der Literaturübersicht gezeigt, bei einem nicht geringen Teil der betreffenden positionellen Kandidatengene auf HSA19q um Mitglieder größerer Gencluster mit immunologischer Bedeutung handelt.

Zudem ist die Wahrscheinlichkeit für ein positionelles Kandidatengen, dem ein hypothesenbasierter DbR-Koeffizient zugewiesen wurde, durch ein oder mehrere Probesets in der Annotation des für die vergleichende Transkriptomstudie verwendeten bovinen *Gene-Chip*<sup>®</sup> (vgl. Hintermair, 2007) vertreten zu sein, statistisch signifikant höher als für ein Kandidatengen ohne einen solchen Rangkoeffizienten, wie Tabelle 17 zeigt; die Ursache kann in der tatsächlichen Nichtberücksichtigung dieser Gene im Design des Microarray oder in unzulänglichen bzw. mehrdeutigen Gendatenbankinformationen, die eine zuverlässige Annotation durch die von Hintermair (2007) eingesetzten Algorithmen nicht zuließen, liegen.

Aus diesen Erwägungen scheint es gerechtfertigt, diese positionellen Kandidatengene mit einem DbR-Koeffizienten = 0 in den weiteren Untersuchungen nicht weiter zu berücksichtigen.

Die Variable *Zaehler* aus dem auf S. 38 dargestellten Pseudocode, d.h. die Zahl von Übereinstimmungen zwischen den MeSH-Annotationen und der hypothesenbasierten Stichwortliste, als DbR-Koeffizient einzusetzen, hätte zu einer unerwünschten Abhängigkeit des DbR-Koeffizienten eines positionellen Kandidatengens von der Anzahl der für dieses Kandidatengen analysierten Publikationen führen können.

Um diese Abhängigkeit aufzulösen, wurde als Korrekturfaktor die Gesamtzahl einzigartiger (unikaler) MeSH-Terme in den Annotationen der für jedes positionelle Kandidatengen gefundenen Publikationen in die Formel 3 zur Berechnung des DbR-Koeffizienten (S. 37) eingeführt. Aufgrund seiner sehr hohen und hochsignifikanten Korrelation mit der Gesamtzahl Artikel, mit denen sie annotiert wurden, ist er ebenso wie diese ein Maß für den Umfang der für ein Gen vorliegenden Literatur, bietet aber durch den größeren Wertebereich eine potentiell schärfere Trennung.

Die durch den vorgestellten Ansatz top-priorisierten Gene haben wohldefinierte Funktionen in immunologischen Mechanismen und stehen damit im Einklang mit der zugrundeliegenden Arbeitshypothese. Herausgehoben werden soll an dieser Stelle das Gen *HCST*, das trotz seiner in Relation zu anderen hochpriorisierten Genen recht niedrigen Zahl von Referenzpublikationen korrekt eingeordnet wurde. Dieses Gen wurde in einer Arbeit zum (humanen) Abdominalen Aortenaneurysma, für das eine Kandidatengenregion auf HSA19q mit der in dieser Arbeit diskutierten überlappt, im Zuge einer Suche nach immunologischen Kandidatengenen ebenfalls hochpriorisiert (Lillvis *et al.*, 2011). Allerdings erfolgte dies dort erst durch kombinierte Betrachtung mehrerer separater Annotationsdatenquellen und nicht automatisiert.

Für die Umsetzung des Datenbankansatzes wurde die interpretierte<sup>29</sup> Skriptsprache Perl eingesetzt, die aufgrund ihrer Vielseitigkeit und Flexibilität gerade in der Bioinformatik breite Anwendung findet. Sie ermöglicht sowohl prozedurale als auch objektorientierte Programmierung, ist also für kleinste bis sehr umfangreiche Projekte gleichermaßen geeignet. Besonders hervorzuheben sind die hochentwickelten Methoden zur Verarbeitung von Zeichenketten, die in dieser Arbeit angewendet wurden. Die in Perl verfügbare Implementierung und Erweiterung regulärer Ausdrücke hat als Softwarebibliothek *Perl Compatible Regular Expressions*, PCRE, Eingang in andere Softwareprojekte gefunden (<http://www.pcre.org>).

---

<sup>29</sup>d.h., es werden keine ausführbaren Programme erzeugt oder benötigt, sondern die Quelltexte werden bei jedem Programmstart neu in Maschinensprache übersetzt. Dies ermöglicht eine höhere Flexibilität, bedeutet aber zumindest potentiell eine Laufzeitverlängerung.

Der Datenbankansatz wurde nach dem Prinzip der objektorientierten Programmierung entwickelt. Hierzu wurden zunächst die drei grundlegenden Komponenten des Ansatzes identifiziert: *Gene*, *Publikationen* und *Stichworte*. Diese stellten jeweils unterschiedliche Anforderungen an die Verarbeitung der zugehörigen Daten. Folglich wurde für jede dieser Komponenten eine Klasse definiert, welche diesen Anforderungen gerecht werdende Datenstrukturen bereitstellte. Indem im Hauptprogramm der Zugriff auf die Daten ausschließlich über in den Klassen definierte Zugriffsmethoden erfolgte, wurde ein Grundsatz der objektorientierten Programmierung realisiert, nämlich die strikte Trennung der Speicherung von Daten vom Zugriff auf diese. Dies hat mehrere Vorteile. Zum Einen senkt die Auslagerung der Datenstruktur- und Verarbeitungsdefinitionen in einen separaten Bereich Umfang und Komplexität des Hauptprogramms und damit potentiell die Fehlerquote. Darüber hinaus haben notwendige oder wünschenswerte Änderungen an den Datenstrukturen keine Auswirkungen auf das Hauptprogramm, da beide Programmteile nur über das Methodeninterface kommunizieren. Mit der hier praktizierten modularen Schreibweise steigen Wartbarkeit des Codes und Wiederverwendbarkeit einzelner Codefragmente.

Dem kombinierten Ansatz, also der quantitativen Integration der Ergebnisse einer hypotesenbasierten Transkriptomstudie, lagen einige Prämissen zugrunde. Zunächst handelte es sich bei einer möglichen differentiellen Expression eines positionellen Kandidatengens zwischen den untersuchten Stichproben weder um eine notwendige noch um eine hinreichende Bedingung für die Berücksichtigung als funktionelles Kandidatengens. Gene, für die mit den Mitteln des vorgestellten Datenbankansatzes keine immunrelevante Funktion annotiert werden konnte, wurden mangels klarer Relevanz für die Arbeitshypothese nicht weiter berücksichtigt. Umgekehrt muss für hypotesenrelevante Gene nicht zwangsweise eine starke differentielle Expression gefordert werden. Dies ergibt sich aus dem besonderen Design der zu integrierenden Studie. Die Besonderheit besteht hier nicht so sehr darin, dass eine Hochdurchsatztranskriptomstudie für die Dissektion einer genetischen Resistenz beim Nutztier angewendet wurde. Solche Untersuchungen sind bereits vielfach beschrieben worden, vgl. Kap. I.2.2.2. Vielmehr hebt sie sich von diesen ab, indem hier die Einteilung der Tiere in die genetischen Resistenzgruppen (QQ,qq) nicht aufgrund phänotypischer Beobachtungen (wie z.B. in Araujo *et al.* (2009) und Nino-Soto *et al.* (2008) geschehen) oder Zugehörigkeit zu verschiedenen Rassen (Glass & Jensen, 2007; Jensen *et al.*, 2008) vorgenommen wurde, sondern aufgrund des Besitzes komplementärer Allele des BTA18-SCS-QTL; die hier verfolgte Strategie einer maximalen genetischen Dissimilarität auf Basis von QTL-Information erhöht die Aussagekraft von *genetical genomics*-Studien deutlich (Jin *et al.*, 2004; de Koning *et al.*, 2005; Cardoso *et al.*, 2008).

In dieser, wie auch in anderen Transkriptomstudien zur Resistenzdissektion, fehlt ein externer Reiz, etwa eine Pathogenstimulation bzw. eine darauf erfolgende Immunreaktion. Die

*a priori* erwarteten, nur auf die unterschiedliche genetische Ausstattung zurückgehenden Expressionsunterschiede sind daher gering; folglich werden in solchen Studien vergleichsweise niedrige Werte für den minimal für ein differentiell exprimiertes Gen geforderten Expressionsunterschied (*fold change*, FC) zwischen den Resistenzgruppen angesetzt, so z.B. 1,1 in Keane *et al.* (2006) oder 1,2 in Araujo *et al.* (2009). In dieser Arbeit wurde auf einen FC-Schwellenwert verzichtet und nur auf eine moderate statistische Signifikanz (Irrtumswahrscheinlichkeit  $\alpha \leq 0,05$ ) des *fold change* abgestellt.

In Szenarien, in denen eine Vielzahl von Nullhypothesen simultan getestet werden, wie es bei den Transkriptomdaten hier der Fall ist, wird gewöhnlich eine Korrektur der Irrtumswahrscheinlichkeiten  $p$  (durch Kontrolle der *family-wise error rate* bzw. *false discovery rate*) durchgeführt. Mit dem Ziel, die Anzahl „überlebender“, für den integrierten Ansatz verfügbarer Datensätze zu erhöhen, wurde hier auf eine solche Korrektur verzichtet. Von Lu *et al.* (2008) wurde eine vergleichbare Logik auf einen integrativen *transcriptomics*-Ansatz bei der Maus angewendet.

Der für die Bewertung des kombinierten Ansatzes entscheidende Parameter ist nicht der absolute kombinierte Datenbank/Transkriptom-Rangkoeffizient der Gene, sondern deren Rangdifferenz im Vergleich zum heuristischen datenbankbasierten Ansatz und in Abhängigkeit von ihrem Expressionsstatus. Die Abbildungen 10 und 11 nutzen daher den Parameter Rangdifferenz, um die Rolle des Faktors Genexpression im kombinierten Ansatz zu untersuchen. Eine wünschenswerte Eigenschaft der Berechnungsmethode des kombinierten DbTR-Koeffizienten und im Boxplot in Abbildung 10 gezeigt, ist, dass keines der 177 Probesets, das keine differentielle Expression in mindestens einem der untersuchten fünf Gewebe aufweist, einen Rangzugewinn verzeichnete. Es kam jedoch zu Rangverlusten in dieser Gruppe, die nur durch eine Verdrängung durch differentiell regulierte Gene, die Ränge hinzugewannen, erklärt werden kann. Da es hier keine den nichtregulierten Genen intrinsische Ursachen gibt, sind die Rangverluste in der weit überwiegenden Zahl erwartungsgemäß minimal.

Das komplexe Auswahlverfahren der in die Transkriptomstudie aufgenommenen Tiere bedingte allerdings einen relativ geringen Stichprobenumfang ( $n = 5$  bzw.  $6$ ). Hier liegt Potential, durch Anwendung fortschrittlicher univariater Methoden wie *Limma* (*linear models for microarray data*, Smyth, 2004) die statistischen Ergebnisse besser abzusichern. Bei *Limma* geschieht dies etwa durch empirische Bayes-Verfahren.

## 2 Die ausgewählten Kandidatengene

In der Literaturübersicht wurde herausgearbeitet, dass es deutliche Hinweise auf eine Beteiligung adaptiver Immunmechanismen an der Ausprägung einer Resistenz gegen Euterinfektionen gibt, insbesondere unter Umständen, die zu einer Beeinträchtigung nichtadaptiver Mechanismen führen. Tatsächlich wurden Gene, die diesem Bereich zuzuordnen sind, bereits als potentielle funktionelle Kandidatengene für die bovine Mastitisresistenz in Betracht gezogen, bzw. es wurden merkmalsassoziierte Polymorphismen in diesen Genen kartiert. Beispiel sind z.B. der Rezeptor des vornehmlich von B-Lymphocyten sekretierten Interleukin-10 (Verschoor *et al.*, 2009) und die J-Kette, die der Multimerisierung von Immunglobulinen der Klassen A und M dient (Koshland, 1985; Nilsen *et al.*, 2009a). Daneben sind Allele des bovinen Haupthistokompatibilitätskomplex Typ II (*bovine leukocyte antigen*, BoLA bzw. speziesübergreifend *major histocompatibility complex*, MHC), namentlich BoLA-DRB3.2, mit der Resistenz gegen Mastitis assoziiert (Kulberg *et al.*, 2007; Rupp *et al.*, 2007).

Von den sowohl im stichwortbasierten Datenbankansatz als auch im auf diesem aufbauenden kombinierten Ansatz wurden zwei positionelle Kandidatengene top-priorisiert und für die experimentelle Bearbeitung ausgewählt, die der adaptiven Immunität zuzurechnen sind, nämlich *CD22* und *CD79A*. Bei den Produkten beider Gene handelt es sich um Membranrezeptoren auf B-Lymphocyten.

*CD79A* (syn. Ig- $\alpha$ , Mb-1) bildet mit *CD79B* (syn. Ig- $\beta$ , lokalisiert auf BTA19 bzw. HSA17) Heterodimere aus, die über ihre cytoplasmatischen Domänen für die Transduktion von Signalen des B-Zell-Rezeptorkomplexes (BCR) verantwortlich sind.

Orthologe Vertreter wurden in vielen Spezies mit adaptiver Immunität beschrieben; auch im Rind (*Bos taurus*) wurde es kloniert (Youn *et al.*, 1996) und kartiert (Youn *et al.*, 1997). *CD79A/B*-vermittelten Signalen des BCR kommt eine wesentliche Rolle im Prozess der Reifung von B-Lymphocyten in Knochenmark und Peripherie (Torres *et al.*, 1996; Koyama *et al.*, 1997; Pike & Ratcliffe, 2005; Fuentes-Panana *et al.*, 2006; Storch *et al.*, 2007) und bei der Antigenerkennung reifer B-Lymphocyten in der Peripherie (Williams *et al.*, 1994; de Franco, 1997) zu. Das *CD79A*-Molekül besitzt eine cytoplasmatische ITAM-Domäne (*immunoreceptor tyrosine-based activation motif*) mit zwei Tyrosinresten (Reth, 1989; Cambier, 1995; Reth & Wienands, 1997; Bannish *et al.*, 2001), die im Zuge der BCR-Aktivierung phosphoryliert werden und damit Bindestellen für Phosphotyrosinkinasen mit tyrosinbindenden SH2-Domänen aus der Src bzw. Syk-Familie bilden (z.B. Cambier, 1995; Kurosaki *et al.*, 1995). Ein konservierter Tyrosinrest außerhalb des ITAM ermöglicht im phosphorylierten Zustand die Bindung des Adaptormoleküls BLNK (*B-cell linker protein*), welches anschließend die Bildung des Signalosomes initiiert (Siemasko *et al.*, 2002; Patterson *et al.*, 2006). Für die erfolgreiche Heterodimerisierung von *CD79A/B* sind neben den cytoplasmatischen Domänen auch die extrazellulären- und Transmembrandomänen beider Moleküle entschei-

dend (Dylke *et al.*, 2007).

Neben der Funktion als Corezeptor des BCR fungieren CD79A/B-Heterodimere auch als Corezeptor des MHC-II (Lang *et al.*, 2001; Siemasko *et al.*, 1999).

CD22 aus der Familie der SIGLECs (*sialyl acid-binding Ig-like lectins*) nimmt wichtige Funktionen in Homöostase und Erhaltung von B-Lymphocyten wahr (Tedder *et al.*, 2005; Crocker *et al.*, 2007), hauptsächlich durch Dämpfung der Signaltransduktion des BCR.

CD22 verfügt im cytoplasmatischen Teil des Transmembranproteins über mehrere tyrosinbasierte Motive, darunter 3 inhibitorische ITIMs (*immunoreceptor tyrosine based inhibitory motif*), die infolge der Aktivierung des BCR phosphoryliert werden und daraufhin die Tyrosinphosphatase SHP-1 (*SH2-domain-containing protein tyrosine phosphatase 1*) in die Nähe des B-Zell-Rezeptorkomplexes rekrutieren. Daraufhin wird dieser durch SHP-1 dephosphoryliert und seine Signaltransduktion dadurch gedrosselt.

Durch Rekrutierung stimulativer Effektoren, z.B. GRB2, SHC, PI3K und PLC $\gamma$ 2, kann CD22 jedoch auch alternative Mechanismen der Regulation von B-Zellen aktivieren (Tedder *et al.*, 2005).

CD22-defiziente Mäuse (-/-) zeigen eine Hyperaktivierung des adaptiven Immunsystems (Crocker *et al.*, 2007), wird dagegen der Ligand ausgeschaltet, beispielsweise durch Mutation des Enzyms Sialyltransferase (ST6GAL1), resultiert eine Immundefizienz (Grewal *et al.*, 2006).

Hauptsächlich dem Bereich der angeborenen Immunität gehören dagegen die beiden verbleibenden durch den kombinierten Ansatz identifizierten und experimentell untersuchten potentiellen funktionellen Kandidatengene *HCST* (*DAP10*) und *TYROBP* (*DAP12*) an. Beide sind im Genom in einer *head to head*-Tandemkonfiguration benachbart und auch funktionell eng verwandt. Es handelt sich bei ihren Genprodukten um Transmembranproteine, die u.a. auf Natürlichen Killerzellen (NK), Monocyten und Makrophagen exprimiert werden. Sie wirken dort als Corezeptoren für Immunrezeptoren und vermitteln deren Signale ins Cytoplasma. NK-Zellen werden durch TYROBP-gekoppelte NKG2-Rezeptoren (syn. KLR, *killer cell lectin like receptor*), die verschiedene Liganden, darunter nichtklassische MHC-1-Moleküle, erkennen, aktiviert (Takaki *et al.*, 2006). Makrophagen und Monocyten können durch Signale eines TREM (*triggering receptor expressed on myeloid cells*)-1/TYROBP-Komplexes zur Produktion proinflammatorischer Chemokine und Cytokine angeregt werden; dieser Effekt ist synergistisch mit gleichzeitiger Stimulation von Toll-like-Rezeptoren (Tessarz & Cerwenka, 2008). Durch einen anderen TYROBP-gekoppelten Rezeptor auf Monocyten und Makrophagen, TREM-2, werden inhibitorische Signale vermittelt (Hamerman *et al.*, 2006).

Hauptpartner, und beim Mensch dessen einziger Corezeptor, von HCST ist der Rezeptor NKG2D, mit dem es auf der Oberfläche hauptsächlich von NK-Zellen, aber z.B. auch  $\gamma\delta$ -T-Zellen, exprimiert wird. Es ist dort, ähnlich wie TYROBP, an der Erkennung aberranter,



nichtklassischer MHC-I-Moleküle, die z.B. in pathogen-infizierten Zellen, exprimiert werden, beteiligt. Derartige Muster treten aber auch bei der Tumorgenese auf, und HCST ist mit der Immunität gegen Tumore in Verbindung gebracht worden (Hyka-Nouspikel & Phillips, 2006; Hyka-Nouspikel *et al.*, 2007). An der TREM2/TYROBP-abhängigen Aktivierung von PI3K scheint HCST ebenfalls beteiligt (Peng *et al.*, 2010).

### 3 Identifizierung potentiell merkmalsassoziierter SNPs

Das in dieser Arbeit primär interessierende phänotypische Merkmal ist der Gehalt somatischer Zellen in der Milch (SCS), welcher in der Rinderzucht als Indikator für die Eutergesundheit genutzt wird (vgl. Kap. I. 1.4).

Durch Genotypisierung in drei unabhängigen und verschieden strukturierten experimentellen Populationen und Kombination dieser Ergebnisse in einem konsistenten Gesamtbild im Sinne von Ron & Weller (2007) konnten innerhalb der ausgewählten potentiellen funktionellen Kandidatengene nur für SNPs im Gen *CD79A* Hinweise für eine Assoziation mit dem Merkmal SCS erbracht werden. Diese SNPs, CD79A\_-182:A>G (im folgenden als Promotor-SNP bezeichnet) und CD79A\_2808:C>T (Intron-SNP) werden daher im Fokus der folgenden Ausführungen stehen.

Unter den ebenfalls untersuchten Konformationsphänotypen sind die verhältnismäßig meisten mit den beiden SNPs im Gen *CD79A* assoziiert (Einzel-SNP-Modell), wobei es sich in der Mehrzahl um direkt das Euter beschreibende Merkmale handelt. Mit einer Ausnahme, der Strichplatzierung, sind die Sets der signifikanten ( $p < 0,05$ ) Assoziationen beider SNPs in diesem Gen identisch ( $p$  für die Strichplatzierung beim Promotor-SNP  $< 0,1$ ). Für eines der Paare linearer Merkmale, die von Nemcova *et al.* (2011) mit besonders hohen genetischen und phänotypischen positiven Korrelationen angegeben werden und die in der hier untersuchten Halbgeschwisterpopulation hochsignifikant phänotypisch korreliert sind ( $r = 0,57$ ), nämlich Voreuteransatz und Eutertiefe, weisen die homozygoten Genotypen des in der Halbgeschwisterpopulation überrepräsentierten Allels konsistent negative Effekte auf.

In einer genomischen Assoziationsstudie fanden Cole *et al.* (2011) keine gemeinsame Effekte von SNPs auf Voreuteransatz und Eutertiefe (jedoch überlappende Genregionen), wohl aber auf Voreuteransatz und den SCS; diese SNPs sind allerdings nicht auf BTA18 lokalisiert.

Wie die in Kapitel III.3.2.2 dargestellten Resultate zeigen, weisen von den SNPs in *CD79A* beide in der Einzel-SNP-Analyse einen statistisch signifikanten ( $p < 0,01$ ) Effekt auf das Merkmal SCS in allen drei in der Untersuchung berücksichtigten Laktationsperioden auf. Dabei sind es in beiden Fällen die in der Halbgeschwisterpopulation überrepräsentierten Allele (die im weiteren als Wildtypallele bezeichnet werden), die in ihren homozygoten Va-

rianten einen numerisch positiven Effekt für das Merkmal zeigen, d.h. mit einem erhöhten SCS assoziiert sind.

Aufgrund der vielfach beschriebenen negativen Beziehung zwischen einer hohen Milchleistung und der Eutergesundheit (z.B. Rauw *et al.*, 1998) ist es plausibel, dass als Ergebnis einer Zuchtpolitik beim Deutschen Holstein, die lange einen hohen Milchertrag priorisierte (siehe die Literaturübersicht) die Frequenz von Allelen, die einen niedrigeren SCS und damit eine verbesserte Eutergesundheit vermitteln, sinkt.

Hier kommen entweder pleiotrope Effekte eines einzelnen Gens auf diese Merkmale zum Tragen, wobei es jedoch schwierig ist, einen Effekt von *CD79A* auf die Milchleistung aus seiner bekannten Funktion heraus zu postulieren.

Ein Alternativmodell sieht ein enges Kopplungsungleichgewicht zwischen einem potentiellen funktionellen Kandidatengen der Eutergesundheit, beispielsweise *CD79A*, und einem für Milchleistung vor, wobei negative Allele des ersteren in einem „Hitchhiking“-Effekt (Nilsen *et al.*, 2009a) zusammen mit erwünschten, d.h. die Milchleistung fördernden Allelen des letzteren selektiert werden.

Dieser Mechanismus wird beispielsweise für den telomeren Bereich von BTA6, in dem QTL für klinische Mastitis und Milchleistungsparametern kartiert wurden (Klungland *et al.*, 2001; Nilsen *et al.*, 2009b), diskutiert. Nilsen *et al.* (2009a) beschreiben u.a. Haplotypen, welche simultan mit einer Reduktion klinischer Mastitiden und einem höheren Proteingehalt der Milch assoziiert sind. Diese Haplotypen umspannen einen Gencluster für Caseine, die zusammen etwa 80% des Proteingehaltes von Kuhmilch ausmachen (Farrell *et al.*, 2004), sowie die beiden immunologisch interessanten (und möglicherweise mit den Caseinen evolutionär verwandten (Rijnkels, 2002)) Gene *Muc7*, das das antimikrobielle Glykoprotein MG2 codiert, und *IGJ* für die *joint*-Region von Immunglobulinen; merkmalsassoziierte SNPs wurden in beiden Gruppen von Genen detektiert.

Als Indiz für die Beteiligung von Polymorphismen des *CD79A* an der Ausprägung einer genetischen Resistenz können Ergebnisse von Gautier *et al.* (2009) herangezogen werden. Diese Autoren analysierten mithilfe des *BovineSNP50 chip assay* (Matukumalli *et al.*, 2009) Spuren adaptiver genetischer Selektion in westafrikanischen Rinderrassen als Folge von Migrationsbewegungen. Als einen infolgedessen wichtigen Anpassungsprozess geben die Autoren die Entwicklung von Pathogenresistenzen an. Identifiziert wurden 53 genomische Regionen unter positiver Selektion, darunter eine 4,4 Mb große Region auf BTA18. Der SNP, der innerhalb dieser Region den Spitzenwert (*peak*) aufweist, liegt im Gen *CD79A*, dieses wird daher von den Autoren als eines der wahrscheinlichsten Kandidatengene für die Adaption an differente Umweltbedingungen im gegebenen Kontext betrachtet. Dieser Befund darf jedoch hinsichtlich der Rolle von *CD79A* bei der Resistenzausbildung gegen Euterinfektionen aufgrund des differenten Kontextes nicht überbewertet werden.

Es gibt hierzu jedoch auch Gegenbeispiele, so etwa durch Verschoor *et al.* (2009). Dort hatte das unterrepräsentierte Allel eines signifikant mit dem Zuchtwert für SCS in kanadischen Holsteinbullen assoziierten SNP im Gen für eine Kette des IL10-Rezeptors einen numerisch positiven Effekt mit der Folge der Vererbung einer reduzierten Eutergesundheit.

Unterstützung für den angenommenen Status von *CD79A* als potentielles funktionelles Kandidatengen der Eutergesundheit erwächst aus den Resultaten der beiden Experimente zur vergleichenden Sequenzierung, die in den Kapiteln zur selektiven Poolsequenzierung (III.2.2) und Sequenzierung QTL-heterozygoter Bullen (III.2.3), insbesondere die Tabellen 23 und 25, gezeigt werden. Gemäß der Konzeption dieser Arbeit dienten diese Experimente als Selektionsstufen, in denen unter den insgesamt 16 in den vier ausgewählten Kandidatengen detektierten SNPs diejenigen sechs, die in die Assoziationsstudie übernommen wurden, ausgewählt wurden.

Sie können aber auch retrospektiv herangezogen werden, um die Plausibilität der in der Assoziationsstudie erzielten Ergebnisse zu prüfen. Unter einem anderen Aspekt können letztere zu einer kritischen Reflektion über die Validität der gewählten Methoden vergleichender Sequenzierung, speziell der untersuchten Populationen, für die Detektion potentiell merkmalsassoziierter Genvarianten innerhalb der untersuchten Intervalle des feinkartierten BTA18-SCS-QTL beitragen.

Dabei ist natürlich ein gewisser Selektionsbias zu berücksichtigen, da nach der Konzeption dieser Arbeit idealerweise nur die in den Selektionsstufen hinsichtlich einer Merkmalsassoziation vielversprechendsten SNPs in die umfangreiche und aufwendige Assoziationsstudie aufgenommen werden sollten. Relativiert wird der Bias dadurch, dass gemäß der Priorisierung einer Assoziationsstestung von zwei SNPs pro untersuchtem Kandidatengen auch SNPs zur Assoziationsstudie zugelassen wurden, deren Resultate in der selektiven Poolsequenzierung und dem Konkordanztest verglichen mit denen des Promotor-SNP und des Intron-SNP von *CD79A* eine Merkmalsassoziation in etwas geringerem Maß wahrscheinlich erscheinen ließen.

Sowohl für den Promotor-SNP, den Intron-SNP (beide in *CD79A*) als auch für mehrere SNPs in den anderen potentiellen funktionellen Kandidatengen wurden in den DNA-Pools mit extrem gegensätzlichen Zuchtwerten für SCS, d.h. der ersten Selektionsstufe, gravierende Allelverschiebungen festgestellt, gleichwohl erwies sich die daraus abgeleitete Vermutung einer Assoziation der SNPs mit diesem Merkmal nur für die ersten beiden als zutreffend.

Die Zusammenstellung der in Kap. III.2.2 analysierten Extremwertpools erfolgte nach dem durch Darvasi & Soller (1994) vorgeschlagenen und theoretisch begründeten Verfahren der selektiven Poolsequenzierung, SDP (vgl. I.2.1.4). Wie in Abschnitt II.4.1, speziell Abbildung

6 gezeigt, weisen die Pools stark divergierende Werte für den gesamtgenomischen Relativen Zuchtwert Zellzahl (RZS) auf, vererben also eine unterschiedliche genetische Disposition für die Resistenz gegen Euterinfektionen. Das untere Quantil der Bullenstichprobe mit hohem RZS (112,5) und das obere Quantil der Bullenstichprobe mit niedrigem RZS (89) liegen jeweils  $\approx$  eine Standardabweichung oberhalb bzw. unterhalb des Mittelwertes; da der RZS definitionsgemäß normalverteilt ist, bedeutet das, dass je drei Viertel jeder Bullenstichprobe im jeweiligen extremsten Sechstel der RZS-Verteilung liegen. Die Normalverteilung des untersuchten Merkmals ist im Übrigen gemäß Sen *et al.* (2009) ein wichtiges Kriterium für die Anwendbarkeit der SDP.

Analog zusammengesetzte Pools wurden von Sharma *et al.* (2006) erfolgreich bei der Identifizierung SCS-assoziierter SNPs im bovinen *TLR4*-Gen eingesetzt.

Der RZS war jedoch nicht das tatsächliche Selektionskriterium, vielmehr kamen hierfür BLUP-Schätzwerte für den SCS, die für den Effekt verschiedener Allele des BTA18-SCS-QTL korrigiert wurden (*marker assisted best linear unbiased predictor*, MA-BLUP, Fernando & Grossman, 1989; Kühn *et al.*, 2008) zur Anwendung. Von den extrem differierenden MA-BLUP-Schätzwerten (vergleiche wieder Abbildung 6) wird angenommen, dass sie divergierende Allele des BTA18-SCS-QTL reflektieren, so dass die begründete Erwartung besteht, dass sich der kausale SNP bzw. ein mit diesem gekoppelter durch eine Allelverschiebung zwischen den Pools manifestiert.

Die Annahme einer Kopplung zwischen einem der hier identifizierten SNPs und dem QTL, der der MA-BLUP-Selektion zugrunde liegt, ist jedoch problematisch, da für die Zuweisung der Allele des letzteren der Mikrosatellitenmarker TGLA227 genutzt wurde. Dieser liegt extrem telomer und damit mehr als 10 Mb vom am weitesten telomer gelegenen SNP in dieser Arbeit entfernt.

Beim Hausrind erstrecken sich Kopplungsungleichgewichte (LD), u.a. aufgrund einer geringen effektiven Populationsgröße und gerichteter Selektion, über größere Bereiche als beim Mensch (Nsengimana & Baret, 2004; McVean, 2007). Kim & Kirkpatrick (2009) geben Ausdehnungen von LD-Blocks von bis zu 1 Mb, und in einer Studie von Kim *et al.* (2009) in derselben Population trat ein *long range linkage disequilibrium* über 5 Megabasen hinweg auf.

Khatkar *et al.* (2006) und Qanbari *et al.* (2010) zeigen ausgedehntes LD in genomischen Bereichen, die starker Selektion in Milchviehassen unterliegen, u.a. auf BTA6 (vgl. Hayes *et al.*, 2009). Letztere Autoren konnten zwar genauso wenig wie Flori *et al.* (2009) für die erweiterte Region von BTA18, in der die verschiedenen Instanzen des hier behandelten SCS-QTL

lokalisiert wurden, Spuren genomischer Selektion nachweisen; diese ist dennoch angesichts der mehrfachen und unabhängigen Kartierung eines SCS-QTL durchaus plausibel.

Auch wenn die Annahme einer Kopplung zwischen SNPs in den untersuchten Intervallen und dem der Selektion zugrundeliegenden QTL also möglicherweise nicht aufrechterhalten werden kann, wobei dies noch am ehesten für das mehrere Mb telomer gelegene Intervall 9 der Fall ist, in dem bezeichnenderweise die beiden tatsächlich mit dem SCS assoziierten *CD79A*-SNPs liegen, bleibt aufgrund der extremen RZS-Verteilung in den Pools (s.o.) die aus einer Allelverschiebung abgeleitete Hypothese einer Merkmalsassoziation valide.

In diesem Falle wäre der den Intervallen zugrundeliegende QTL nicht mit dem QTL der MA-BLUP-Selektion identisch. Tatsächlich gibt es Indizien für die Präsenz mehrerer QTL für Eutergesundheit/SCS im distalen Bereich von BTA18 (z.B. Brand *et al.*, 2009); Cole *et al.* (2011) fanden in einer genomweiten Assoziationsstudie ebenfalls mehrere SCS-assoziierte SNPs in großer Distanz voneinander in diesem Bereich.

Der in Kapitel II.2.3 vorgestellte Konkordanztest hat sich als wichtiges Kriterium bei der Beurteilung potentiell merkmalsassoziierter SNPs etabliert (Ron & Weller, 2007). Hierbei wird überprüft, ob ein SNP in Testbullen, die heterozygote Allele des mit dem SNP als gekoppelt gedachten QTL tragen, ebenfalls heterozygot ist. Das Fehlen dieses Nachweises im Fall des von Schnabel *et al.* (2005a) postulierten Kandidatengens *SPP1* für einen Milchleistungs-QTL auf BTA6 trug zur Zurückweisung dieses Gens zugunsten von *ABCG2* (Cohen-Zinder *et al.*, 2005) durch Seroussi (2009) bei.

Die Interpretation der hier vorgestellten Ergebnisse wird allerdings dadurch erschwert, dass die QTL-Heterozygotie der Testbullen ebenfalls vom Genotyp am TGLA227 abgeleitet wurde. Im Gegensatz zur vergleichenden Sequenzierung in den Extremwertpools steht hier kein Surrogatmerkmal zur Verfügung, mit dem die Validität des Ansatzes aufrechterhalten werden kann. Es ist jedoch bemerkenswert, dass drei der vier Testbullen sowohl für den Promotor-SNP als auch den Intron-SNP in *CD79A* heterozygot sind; keiner der anderen SNPs ist in mehr als zwei Bullen heterozygot. Der Konkordanztest erweist sich somit in diesem Fall in der Voraussage einer dann tatsächlich festgestellten Merkmalsassoziation der vergleichenden Poolsequenzierung als überlegen.

Einen analogen Fall präsentierten Blott *et al.* (2003) für einen höchstwahrscheinlich kausalen SNP für einen Milchleistungs-QTL auf BTA20, der im Gen für den Wachstumshormonrezeptor (*GHR*) lokalisiert ist, und der in zwei Bullen mit heterozygoten QTL-Allelen als homozygot klassifiziert wurde.

Von den beiden *CD79A*-SNPs, die mit dem Merkmal SCS in allen Laktationsperioden signifikant assoziiert waren, zeigte nur der Intron-SNP (*CD79A\_2808:C>T*) bei einer Irrtums-

wahrscheinlichkeit von 5,6% eine annähernd signifikante Abweichung vom Hardy-Weinberg-Gleichgewicht (HWG, Tab. 26). Dies kann ein Indiz für eine Selektion an diesem SNP sein: In einer Arbeit mit dem Ziel, ein Panel von SNPs zu entwickeln, welche die eindeutige Identifizierung von Rindern mittels eines genetischen Profils ermöglichen, wurden nur solche SNPs berücksichtigt, die sich im Hardy-Weinberg-Gleichgewicht befinden, da bei diesen davon ausgegangen wird, dass sie keinem Selektionsprozess unterliegen (Allen *et al.*, 2010). Auch Visscher & Le Hellard (2003) weisen darauf hin, dass bei krankheitsassoziierten Loci regelmäßig kein HWG zu erwarten ist.

Dieses Kriterium sollte nicht absolut gesetzt werden, es ist jedoch bemerkenswert, dass an beiden SNPs etwa 10% weniger Homozygote des mutierten Allels auftreten als durch die ohnehin niedrigere Frequenz dieses Allels unter Annahme eines HWG vorhergesagt (88 (105) beobachtete gegenüber 101 (116) angenommenen beim Intron-SNP bzw. Promotor-SNP, vgl. Tabelle 26).

Im Einzel-SNP-Modell zeigte der Intron-SNP verglichen mit dem Promotor-SNP sowohl geringere Irrtumswahrscheinlichkeiten für eine Assoziation mit dem Merkmal SCS, als auch die (absolut) größeren Effekte auf dieses Merkmal, hauptsächlich für die homozygoten Nicht-wildtypallele, und zwar konsistent über alle untersuchten Laktationen hinweg.

Das zweite der berechneten Modelle, in dem beide SNPs gemeinsam als additive fixe Effekte berücksichtigt wurden, brachte Ergebnisse, die es möglich erscheinen lassen, dass sich die Beziehung beider SNPs im Hinblick auf das Merkmal SCS laktationsabhängig ändert. Die Genotypenverteilung des Intron-SNP erklärt signifikant mehr ( $p < 0,05$ ) der phänotypischen Varianz des SCS in der ersten Laktation als die des Promotor-SNP. In der zweiten und dritten Laktation ist dies nicht der Fall, wobei hier ein Trend zu einer höheren Irrtumswahrscheinlichkeit zu beobachten ist (2. Laktation:  $p < 0,1$ , 3. Laktation:  $p > 0,1$ ). Dies korrespondiert mit einer Verstärkung der numerisch negativen (also mit einer verbesserten Eutergesundheit gekoppelten) Effekte der Nichtwildtypallele von der ersten bis zur dritten Laktation, die für den Promotor-SNP ausgeprägter ist.

Sofern es sich hierbei nicht um ein statistisches Artefakt handelt, erscheint es nicht abwegig, für Polymorphismen in einem Gen, das der adaptiven Immunantwort zuzuordnen ist, wie im Fall von *CD79A*, zumal für einen Promotor-SNP, einen longitudinalen Effekt auf die Eutergesundheit zu vermuten. So konnten etwa Ohtsuka *et al.* (2009) zeigen, dass sich mit zunehmendem Alter einer Kuh, das mit einer höheren Laktationszahl koinzidiert, die Zahl der peripheren B-Lymphocyten in der Vorkalbperiode vermindert; gleichzeitig steigt die Aktivität von Immunzellen des Euters (Ohtsuka *et al.*, 2010). Für einen SNP im 3'-UTR des bovinen Osteopontingens (*SPP1*), das als u.a. T-Zell-Regulator auch in adaptiven Immun-

reaktionen eine Rolle spielt, weisen Alain *et al.* (2009) ebenfalls eine laktationsabhängige Änderung der Stärke einer Assoziation mit dem Merkmal SCS aus.

Unter funktionellen Gesichtspunkten ist von den beiden mit dem Merkmal SCS assoziierten *CD79A*-SNPs der Promotor-SNP möglicherweise der interessantere. Er liegt innerhalb einer 11 Nukleotide langen Sequenz, die sich von den Positionen -175 bis -182 Nukleotide (relativ zum ersten Exon) erstreckt, und ein durch ein Nukleotid unterbrochenes Palindrom aus zwei Pentameren bildet (Abb. 14). Palindrome sind vielversprechende Kandidaten für Bindestellen von Transkriptionsfaktoren (TF). Dies ist u.a. darauf zurückzuführen, dass die Präsenz zweier identischer DNA-Elemente in einer *tail-to-tail*-Konfiguration, die ein Palindrom kennzeichnet, die simultane Bindung von homodimeren TF ermöglicht. Gut untersucht ist dies z.B. für einen TF der Ets-Familie (Baillat *et al.*, 2002; Babayeva *et al.*, 2010), aber auch für *Runx1* (Bowers *et al.*, 2010), den Tumorsuppressor *p53* (Kitayner *et al.*, 2006) und den *HNRNPK*-Promotor (Mikula *et al.*, 2010).

Innerhalb des Palindroms und über das separierende Nukleotid hinweg befindet sich ein  $\gamma$ -IRE (*gamma interferon response element*). Dieses Element wurde von Yang *et al.* (1990) in den Promotoren humaner MHC-II Gene detektiert und seither speziesübergreifend in den Promotoren verschiedener Gene beschrieben. Eine Auswahl bietet Tabelle 39 (im Anhang). Die Konsensussequenz dieses oktameren Elementes lautet  $C_{AGG}^{TTT}ANN_C^T$ ; nach der IUPAC-Nomenklatur entspricht dies dem Schema CWKKANNY.

In der weit überwiegenden Zahl der bis dato beschriebenen  $\gamma$ -IR-Sequenzen ist das Dinukleotid KK an Position 3/4 homogen; dies ist jedoch nicht obligatorisch und beim  $\gamma$ -IRE des *CD79A*-Promotors auch nicht der Fall.

Über die konkrete Funktion dieses Elementes herrscht noch keine Klarheit; ebenso für das bindende Protein bzw. den bindenden Proteinkomplex (Franklin *et al.*, 2009).

Durch den SNP *CD79A*\_-182:A>G (T>C auf dem Gegenstrang), wird das T an Position 2 des Elementes, durch ein C ersetzt, das nicht dem Muster der Konsensussequenz entspricht; fakultativ wäre hier ein A. Die putative Bindestelle des TF wird dadurch disruptiert, mit möglichen Folgen für die Expression von *CD79A*. Tatsächlich waren Expressionsunterschiede des Gens in den Färsengruppen der M.A.S.-Net-Transkriptomstudie mit einer signifikanten Heraufregulation des Gens in hochresistenten Färsen einer der ursprünglichen Faktoren in seiner Selektion durch den kombinierten bioinformatischen Ansatz als potentielles funktionelles Kandidatengen der Eutergesundheit.

Der ebenfalls als potentiell merkmalsassoziiert in Betracht gezogene der beiden intronischen SNPs befindet sich 2990 Nukleotide stromabwärts<sup>30</sup> in Intron 2 des Gens *CD79A* und

<sup>30</sup>bezogen auf den 5'-Strang stromaufwärts, da *CD79A* auf dem Gegenstrang codiert ist

mit dem Promotor-SNP in engem Kopplungsungleichgewicht (vgl. auch Abb. 17). Für diesen SNP lässt sich nicht ohne weiteres eine Funktion postulieren. Er führt nicht zur Bildung einer aberranten Spleißstelle; doch können auch intronische SNPs außerhalb von Spleißkonsensussequenzen an aberranten Spleißvorgängen beteiligt sein (Toscano *et al.*, 2006).

Für die SNP im Gen *CD79A* wurden zwar, ganz im Sinne von Ron & Weller (2007), experimentelle Indizien für eine Merkmalsassoziation beigebracht, während der vorgestellte bioinformatische Ansatz dies für das Gen selbst leistete. Der konkrete Nachweis steht jedoch noch aus. Denkbar wären hier beispielsweise Untersuchungen zu DNA-Protein-Bindungen, etwa durch Chromatin-Immunopräzipitation (ChIP) oder *Electrophoretic Mobility Shift Assay* (EMSA), mit denen die Bindung eines Kandidaten-TF an regulatorische Elemente des *CD79A* gezeigt oder widerlegt werden kann.

Wenn der Nachweis, dass es sich bei einem der hier untersuchten SNP um das kausale Nukleotid des QTL, also des QTN, handelt, nicht geführt werden kann, besteht die Möglichkeit, dass die tatsächliche kausale Mutation mit einem von ihnen in engem Kopplungsungleichgewicht (LD) steht. In diesem Fall können Haplotypen, die den SNP, das QTN sowie gegebenenfalls weitere im LD stehende Nukleotide umfassen, abgeleitet und für genomische Zuchtprogramme eingesetzt werden. Die hier vorgestellten SNP können dabei möglicherweise als *Tag-SNP* (de Bakker *et al.*, 2005) zur Charakterisierung des gesamten Haplotyps genutzt werden.

Die experimentelle Untersuchung dieser letztlich sehr interessanten SNPs wurde ermöglicht durch die vorherige Selektion ihrer „Wirts“gene aus einer *a priori* sehr umfangreichen Liste positioneller Kandidatengene des BTA18-SCS-QTL. Die vorgestellte Methode, welche einen heuristischen bioinformatischen, auf der automatisierten Auswertung kuratierter Referenzliteratur basierenden Ansatz zur hypothesengeleiteten Genpriorisierung quantitativ mit den Resultaten einer Hochdurchsatz-Transkriptomstudie integriert, erwies sich hierbei als sinnvoll und zielführend.

Mit der immer schneller voranschreitenden Verbreitung von *-omics*-Verfahren auch in den Nutztier- bzw. Nutzpflanzenwissenschaften wird der Bedarf an bioinformatischen Werkzeugen zur Filterung der anfallenden Datenmengen ebenfalls wachsen. Vor diesem Hintergrund erscheint eine Weiterentwicklung und Anwendung des vorgestellten Ansatzes vielversprechend.



## Teil V

# Zusammenfassung

Die Eutergesundheit beim Milchvieh der Rasse Deutsche Holstein (DH) ist eng mit dem Gehalt somatischer Zellen in der Milch (ausgedrückt durch den Parameter SCS), einem quantitativen polygenen Merkmal, verknüpft. Ein gut bestätigter QTL für dieses Merkmal auf Chromosom 18 (BTA18-SCS-QTL), der im Vorfeld dieser Arbeit durch Feinkartierung hochaufgelöst wurde, und der eine syntene Region auf dem langen Arm des humanen Chromosom 19 besitzt, weist dennoch eine sehr hohe Zahl verbleibender positioneller Kandidatengene auf. Ziel der vorliegenden Arbeit war es, erstens innerhalb dieser Region vielversprechende funktionelle Kandidatengene, die gemäß einer Arbeitshypothese einen Bezug zu immunologischen Prozessen aufweisen, zu identifizieren und zweitens in einer Auswahl dieser Kandidatengene Einzelbasenaustausche (SNPs) zu bestimmen und auf ihre potentielle Assoziation mit dem Merkmal SCS zu untersuchen.

Die im ersten Schwerpunkt erforderliche Priorisierung von Kandidatengen wurde erreicht, indem zunächst für jedes positionelle Kandidatengen eine Liste kuratierter wissenschaftlicher Referenzartikel innerhalb der Literaturlatenbank PubMed ermittelt wurde; dies erfolgte durch skriptgesteuerte Auswertung der vom NCBI zur Verfügung gestellten Datenbank *Gene2Pubmed*. Über eine Programmierschnittstelle wurden die identifizierten Artikel im zusammenfassenden MEDLINE-Format bezogen. Ein hypothesenbasierter Datenbank-Rangkoeffizient wurde für jedes positionelle Kandidatengen berechnet, indem das Ausmaß der Übereinstimmung zwischen den MeSH-Annotationen seiner Referenzliteratur und einer die Arbeitshypothese reflektierenden Stichwortliste, korrigiert um die Gesamtmenge der für ein Gen ermittelten MeSH-Annotationen, untersucht wurde.

Dieser Ansatz wurde holistisch erweitert, indem Expressionsdaten (genauer: *fold changes*), die im Rahmen der M.A.S.-Net-Transkriptomstudie (deren Design spezifisch auf die Detektion der dem BTA18-SCS-QTL zugrundeliegenden Variation zugeschnitten war) an fünf Eutergeweben erhoben wurden, auf die positionellen Kandidatengene der QTL-Region abgebildet und nach Anwendung eines Signifikanzfilters in diesen integriert wurden.

Der erweiterte, integrierte Ansatz resultierte in der Erstellung einer informierten Rangliste potentieller funktioneller Kandidatengene des BTA18-SCS-QTL, aus der die vier hochpriorisierten und tatsächlich in hohem Maße immunrelevanten Kandidatengene *CD22*, *CD79A*, *HCST* und *TYROBP* für die experimentelle Bearbeitung im zweiten Schwerpunkt ausgewählt wurden.

Insgesamt 16 SNPs innerhalb dieser Gene wurden identifiziert und in einem mehrstufigen reduktionistischen Protokoll auf ihre potentielle Assoziation mit dem Merkmal SCS getestet. Dies erfolgte zunächst durch vergleichende Sequenzierung in zwei DNA-Pools von DH-Bullen, die gegensätzliche Extreme der genetisch bedingten Resistenz gegen Euterinfektionen repräsentierten. Dem schloss sich für SNPs, die dort eine für eine Merkmalsassoziation erwartete, zwischen den Pools stark differente Allelverteilung zeigten, ein Test der Hypothese ihrer Heterozygotie in vier DH-Testbullen mit heterozygoten Allelen des BTA18-SCS-QTL an.

Nach gemeinsamer Betrachtung beider Ansätze qualifizierten sich sechs SNPs, je zwei davon in den Genen *CD22*, *HCST* und *CD79A* für eine Assoziationsstudie in einer großen Halbgeschwisterpopulation der Rasse DH. Hierfür lagen neben dem Merkmal SCS in den ersten drei Laktationsperioden Daten zu mehreren Konformations- und Milchleistungsphänotypen vor.

Die Berechnungen ergaben signifikante Effekte auf das Merkmal SCS in allen Laktationsperioden sowie mehrere Konformationsphänotypen, hauptsächlich das Euter betreffend, für die beiden SNPs des Gens *CD79A*. Einer dieser SNPs ist im *CD79A*-Promotor innerhalb putativer regulatorischer Elemente lokalisiert. *CD79A* empfiehlt sich für eine weitergehende Untersuchung im Hinblick auf seine Rolle in der bovinen Eutergesundheit. Der vorgestellte bioinformatische Ansatz wurde damit erfolgreich für die Identifizierung potentieller funktionaler Kandidatengene eines quantitativen Merkmales beim Nutztier eingesetzt.

## Teil VI

## Appendix

## Tabellen

**Tabelle 29:** Resultate der M.A.S.-Net Transkriptomstudie. Gezeigt sind die mittleren relativen Expressionsunterschiede (*fold change*) zwischen den Färsengruppen QQ (hohe Resistenz) und qq (niedrige Resistenz) über die fünf untersuchten Eutergewebe (Lymphknoten **LN**, Drüsenzisterne **DZ**, Fürstenbergsche Rosette **FR**, Parenchym **Pa** und Strichkanal **SK**) für Probesets, die folgende Bedingungen erfüllen: Das Probeset ist für ein positionelles Kandidatengen der hier untersuchten Intervalle des BTA18-SCS-QTL annotiert (zutreffend für insgesamt 212 Probesets) und es ist in mindestens einem der Eutergewebe signifikant ( $p \leq 0,05$ ) differentiell exprimiert. Die Gewebe, für die letzteres zutrifft, sind jeweils mit einem \* markiert.  $\uparrow$ : Das Probeset wird in der Gruppe mit QQ-Genotyp für den BTA18-SCS-QTL höher exprimiert; eine entgegengesetzte Regulation ist durch  $\downarrow$  gekennzeichnet. Die Präprozessierung der DNA-Microarrays führte (gewebespezifisch) zum Ausschluss einer Reihe von nichtexprimierten Probesets von der weiteren Analyse; für diese liegen demgemäß keine Daten vor, sie sind hier durch NA gekennzeichnet.

Probeset	Symbol	LN		DZ		FR		PA		SK	
Bt.18570.1.S1_at	ACTN4	NA		NA		1,12	$\uparrow$	NA		1,4	* $\downarrow$
Bt.21775.1.S1_at	ARHGEF1	1,36	* $\uparrow$	NA		1,06	$\uparrow$	NA		1,06	$\downarrow$
Bt.20087.1.S1_at	ATP5SL	1,27	* $\uparrow$	1,02	$\downarrow$	1,2	* $\downarrow$	1,05	$\downarrow$	1,05	$\uparrow$
Bt.1960.2.S1_a.at	B9D2	1,01	$\downarrow$	NA		1,22	* $\downarrow$	NA		NA	
Bt.16930.1.A1_at	BCL3	1,08	$\uparrow$	NA		1,19	$\uparrow$	NA		1,19	* $\uparrow$
Bt.28412.1.A1_at	C19orf15	1,31	* $\uparrow$	1,12	$\downarrow$	1,23	$\uparrow$	NA		1,12	$\uparrow$
Bt.22161.1.S1_at	C19orf61	1,16	* $\uparrow$	NA		1,03	$\downarrow$	NA		1,02	$\downarrow$
Bt.28795.1.A1_at	CAPNS1	1,15	$\uparrow$	1,08	$\downarrow$	1,11	* $\downarrow$	1,02	$\downarrow$	1,17	$\downarrow$
Bt.13207.1.S1_at	CCDC97	1,23	* $\uparrow$	1,07	$\downarrow$	1,02	$\uparrow$	NA		1,03	$\uparrow$
Bt.4338.1.S1_at	CD22	1,46	* $\uparrow$	1,28	* $\uparrow$	1,07	$\uparrow$	1,08	$\downarrow$	1,01	$\downarrow$
Bt.4436.1.S1_a.at	CD79A	1,88	* $\uparrow$	NA		NA		NA		NA	
Bt.10027.1.S2_at	CEACAM1	1,16	$\uparrow$	1,26	* $\uparrow$	1,06	$\uparrow$	1,04	$\downarrow$	1,04	$\uparrow$
Bt.21019.1.S1_at	CIC	1,3	* $\uparrow$	1,18	$\downarrow$	1,1	$\uparrow$	1,03	$\uparrow$	1,08	* $\uparrow$
Bt.75.1.S1_at	COX6B1	1,08	$\uparrow$	1	$\downarrow$	1,21	* $\downarrow$	1,02	$\uparrow$	1,03	$\uparrow$
Bt.4589.1.S1_at	DEDD2	1,2	$\uparrow$	1,01	$\downarrow$	1,02	$\uparrow$	NA		1,28	* $\uparrow$
Bt.20340.2.S1_at	DMPK	1,19	* $\uparrow$	1,35	$\downarrow$	1,03	$\uparrow$	1,01	$\uparrow$	1,31	$\downarrow$
Bt.20340.3.S1_at	DMPK	NA		1,16	$\downarrow$	1	$\downarrow$	NA		1,64	* $\downarrow$
Bt.13633.1.A1_at	EIF3K	1,56	* $\uparrow$	NA		1,12	$\uparrow$	NA		1,1	$\downarrow$
Bt.23512.1.S1_at	EIF3K	1,01	$\uparrow$	1,12	$\downarrow$	1,11	* $\downarrow$	1,03	$\uparrow$	1,09	* $\downarrow$
Bt.6079.1.S1_a.at	ERCC1	1,1	$\uparrow$	1,09	$\downarrow$	1,17	* $\downarrow$	1,06	$\downarrow$	1,03	$\uparrow$
Bt.6079.2.S1_a.at	ERCC1	1,04	$\uparrow$	1,04	$\downarrow$	1,17	* $\downarrow$	1,05	$\downarrow$	1,08	* $\downarrow$
Bt.16945.1.S1_at	ETHE1	1,14	$\uparrow$	1,17	* $\downarrow$	1,27	* $\downarrow$	1,13	$\downarrow$	1,09	$\uparrow$
Bt.7019.1.S1_at	ETV2	1,3	* $\uparrow$	1	$\uparrow$	1,19	* $\downarrow$	1,03	$\downarrow$	1,06	$\downarrow$

Fortsetzung auf der nächsten Seite

Probeset	Symbol	LN	DZ	FR	PA	SK
Bt.2831.1.S1_at	EXOSC5	1,08 ↓	1,13 ↓	1,09 ↓	1,23 * ↓	1,02 ↑
Bt.29542.1.A1_at	FFAR2	1,07 ↑	1,02 ↑	1,02 ↓	1,11 * ↑	1,05 ↓
Bt.24863.1.S1_at	FOSB	NA	1,3 * ↑	1,1 ↑	1,14 ↓	1,05 ↑
Bt.20096.2.S1_a.at	FXYD5	1,2 ↑	1,21 ↓	1,22 ↓	1,1 ↓	1,36 * ↓
Bt.20510.1.S1_at	FXYD7	1,33 * ↑	1,4 ↓	1,04 ↓	NA	1,25 ↓
Bt.7919.1.A1_at	FXYD7	1,42 * ↑	1,17 ↑	1,06 ↑	NA	1,02 ↑
Bt.13505.1.S1_at	GPI	1,01 ↓	1,19 * ↓	1,12 ↓	1,07 ↓	1,01 ↓
Bt.26582.2.S1_at	GPI	1,15 ↓	NA	1,24 * ↓	NA	NA
Bt.7478.1.S1_at	HSPB6	1,32 ↑	1,07 ↓	1,24 * ↑	1,11 ↓	1,09 ↓
Bt.12204.1.S1_at	KIAA0355	1,16 ↑	1,13 ↓	1,23 * ↑	1,06 ↑	1,24 * ↑
Bt.14202.1.A1_at	LGI4	1,21 * ↑	1,12 ↓	1,08 ↑	NA	1,11 ↓
Bt.14202.2.S1_at	LGI4	1,11 ↓	1,17 * ↓	1,09 ↑	1,1 ↑	1,1 ↓
Bt.4486.1.S1_a.at	LIN37	1,55 * ↑	1,07 ↓	1,13 ↓	1,06 ↓	1,04 ↓
Bt.13143.1.A1_at	LIPE	1 ↑	1,21 * ↑	NA	1,04 ↓	NA
Bt.24735.1.A1_at	LSM14A	1,05 ↓	1,29 * ↓	1,1 ↓	1,16 * ↓	1,02 ↓
Bt.9656.1.S1_at	MARK4	NA	NA	1,2 * ↑	NA	1,17 * ↑
Bt.4040.1.S1_at	PAFAH1B3	1,11 ↑	1,17 * ↓	1,26 ↓	1,1 ↓	1,13 ↑
Bt.13950.1.A1_at	PDCD2L	1,01 ↓	1,03 ↑	1,17 * ↓	1,11 ↓	1,01 ↑
Bt.3082.1.S1_at	PEPD	1,36 * ↓	1,07 ↓	1,01 ↓	1,04 ↓	1,03 ↓
Bt.2466.1.S1_at	POLR2I	1,24 * ↑	1,16 ↓	1,15 * ↓	1,08 ↓	1,02 ↓
Bt.9048.2.S1_a.at	PSENNEN	1,07 ↑	1,14 * ↓	1,12 * ↓	1,08 ↓	1,04 ↓
Bt.4943.2.S1_at	PVRL2	NA	1,26 * ↓	1,04 ↑	1,14 ↓	1,01 ↓
Bt.17669.1.S1_a.at	RELB	1,17 * ↑	NA	1,04 ↓	NA	NA
Bt.1973.1.S1_at	SNRPD2	1,09 ↑	1,2 * ↓	1,13 * ↓	1,04 ↓	1,04 ↓
Bt.15844.1.S1_at	TBCB	1,03 ↓	1,34 * ↓	1,06 ↓	1,13 ↓	1,07 ↓
Bt.4533.1.S1_at	TMEM147	1,19 ↑	1,02 ↓	1,14 * ↓	1,04 ↓	1,04 ↑
Bt.8939.1.S1_at	TYROBP	1,01 ↑	1,07 ↓	1,44 * ↓	1,08 ↑	1,21 ↓
Bt.5384.1.S1_at	U2AF1L4	1,19 ↑	1,24 ↓	1,2 ↓	1,1 ↓	1,24 * ↓
Bt.4294.1.S1_at	USF2	1,21 ↑	1,14 ↑	1,2 * ↓	1,02 ↓	1 ↓
Bt.29377.1.S1_at	VASP	1 ↑	1,09 ↓	1,14 * ↓	1,05 ↓	1,11 ↓
Bt.27918.1.S1_at	WTIP	1,05 ↑	1,27 ↓	1,01 ↓	NA	1,24 * ↓
Bt.3059.1.S1_at	YIF1B	1,07 ↓	1,04 ↓	1,26 * ↓	1,01 ↓	1,18 ↓
Bt.28639.1.S1_at	ZNF146	1,23 ↓	1,25 * ↓	1,11 ↓	1,09 ↓	1,06 ↓
Bt.25310.1.A1_at	ZNF181	1,21 ↓	1,21 ↓	1,1 ↓	1,11 * ↓	1,04 ↓
Bt.22265.1.S1_at	ZNF225	1 ↑	1,09 ↓	1,24 ↑	1,22 * ↓	1,61 ↑
Bt.5423.1.S1_at	ZNF233	1,15 ↑	1,01 ↓	1,11 * ↓	1 ↓	1,16 * ↓
Bt.18686.1.A1_at	ZNF235	1,05 ↓	1,23 * ↓	1,16 * ↓	1,08 ↓	1,04 ↓
Bt.27569.1.A1_at	ZNF565	1 ↑	1,15 * ↓	1,16 ↓	1,11 ↑	1,12 ↓
Bt.19400.1.A1_at	ZNF567	1,07 ↓	1,3 * ↓	1,32 ↓	1,06 ↓	1,14 ↓
Bt.2661.1.S1_at	ZNF574	1,11 ↑	NA	1,11 ↑	NA	1,11 * ↑

**Tabelle 30:** Resultate der M.A.S.-Net Transkriptomstudie. Gezeigt sind für jedes der (für diese Arbeit relevanten) untersuchten Tiere der Gruppen QQ (hohe Resistenz) bzw. qq (geringe Resistenz) die Expressionswerte für Probesets, die folgende Bedingungen erfüllen: Das Probeset ist für ein positionelles Kandidatengen der hier untersuchten Intervalle des BTA18-SCS-QTL annotiert (zutreffend für insgesamt 212 Probesets) und es ist in mindestens einem der Eutergewebe signifikant ( $p \leq 0,05$ ) differentiell exprimiert. Expressionswerte wurden durch Anwendung des Algorithmus PLIER (*Probe Logarithmic Intensity Error*) für die Prozessierung der Signalintensitäten ermittelt und werden als log2-transformierte Werte präsentiert. Beachte, dass das Tier mit der laufenden Nummer 6 komplett aus der Studie ausgenommen wurde, es ist hier nur der Vollständigkeit halber aufgeführt. Einzelne Microarrays wurden wegen mangelnder Qualität ebenfalls ausgenommen, die entsprechenden Werte sind durch NA ersetzt. LN: Lymphknoten; FR: Fürstenbergsche Rosette; DZ: Drüsenzisterne; Pa: Parenchym; SK: Strichkanal

Probeset	Gewebe	QQ						qq				
		1	2	3	4	5	6	7	8	9	10	11
Bt.4338.1.S1.at	LN	10,656	10,263	10,277	10,782	10,507	NA	10,15	NA	9,701	10,115	9,824
Bt.4436.1.S1.a.at	LN	11,229	10,134	10,888	11,241	11,333	NA	10,052	NA	10,155	10,339	9,617
Bt.8939.1.S1.at	LN	10,331	10,547	10,29	10,429	10,302	NA	10,329	NA	10,481	10,43	10,205
Bt.10027.1.S2.at	LN	9,416	9,376	9,199	9,602	9,102	NA	9,268	NA	8,654	9,266	9,305
Bt.17669.1.S1.a.at	LN	6,992	6,842	7,118	7,116	7,214	NA	6,799	NA	7,016	6,637	6,773
Bt.4943.2.S1.at	LN	5,417	5,223	5,68	5,682	5,566	NA	5,137	NA	5,631	5,762	5,855
Bt.16930.1.A1.at	LN	8,75	8,357	8,637	9,006	8,769	NA	8,641	NA	8,571	8,598	8,519
Bt.1973.1.S1.at	LN	11,839	11,912	11,826	11,709	11,883	NA	11,781	NA	11,768	11,748	11,553
Bt.2831.1.S1.at	LN	7,476	7,455	7,611	7,377	7,268	NA	7,423	NA	7,999	7,44	7,34
Bt.13505.1.S1.at	LN	9,532	9,652	10,46	9,641	10,083	NA	9,519	NA	10,15	9,981	9,913
Bt.26582.2.S1.at	LN	3,656	5,046	4,201	3,252	4,1	NA	4,803	NA	4,482	4,283	4,372
Bt.2466.1.S1.at	LN	9,803	9,745	9,675	9,649	9,883	NA	9,497	NA	9,457	9,526	9,264
Bt.29542.1.A1.at	LN	7,05	6,815	6,634	7,08	7,045	NA	6,468	NA	6,751	6,969	7,07
Bt.18570.1.S1.at	LN	6,496	6,126	6,515	6,145	6,318	NA	6,004	NA	6,757	6,694	6,075
Bt.20340.2.S1.at	LN	5,193	5,268	5,302	4,849	5,358	NA	4,472	NA	4,943	4,835	5,067
Bt.20340.3.S1.at	LN	6,204	5,971	6,125	6,44	6,382	NA	6,004	NA	5,731	5,775	5,831
Bt.24863.1.S1.at	LN	5,981	5,929	5,933	6,457	6,291	NA	5,937	NA	6,095	6,382	6,933
Bt.9656.1.S1.at	LN	5,77	6,125	6,003	5,989	6,429	NA	6,163	NA	5,698	5,693	5,948
Bt.28795.1.A1.at	LN	7,29	6,807	7,214	7,1	7,314	NA	6,624	NA	6,98	7,062	7,004
Bt.3082.1.S1.at	LN	10,737	10,796	10,724	10,847	10,385	NA	10,931	NA	10,876	11,249	11,54
Bt.20096.2.S1.a.at	LN	9,812	9,649	10,149	9,604	10,352	NA	9,718	NA	9,989	9,416	9,475
Bt.29377.1.S1.at	LN	10,152	10,5	10,675	10,399	10,459	NA	10,336	NA	10,537	10,553	10,295
Bt.4294.1.S1.at	LN	9,296	8,882	9,385	9,507	9,596	NA	9,035	NA	9,104	9,296	8,747
Bt.4589.1.S1.at	LN	6,508	6,028	6,782	6,395	6,953	NA	5,851	NA	6,562	6,372	6,108
Bt.16945.1.S1.at	LN	8,845	8,606	9,018	8,73	9,194	NA	8,442	NA	8,804	8,9	8,585
Bt.13633.1.A1.at	LN	7,565	7,663	8,272	7,564	7,91	NA	7,549	NA	6,43	6,973	7,405
Bt.23512.1.S1.at	LN	11,497	11,698	11,818	11,378	11,715	NA	11,505	NA	11,711	11,689	11,543
Bt.21775.1.S1.at	LN	9,55	9,152	9,733	9,524	9,757	NA	9,292	NA	9,242	9,134	8,69
Bt.4533.1.S1.at	LN	10,175	9,673	10,126	10,064	10,396	NA	9,666	NA	9,861	10,065	9,753
Bt.28639.1.S1.at	LN	9,111	8,625	8,436	9,427	8,302	NA	9,761	NA	8,559	8,639	9,374
Bt.7478.1.S1.at	LN	5,435	6,101	6,173	6,488	6,564	NA	5,654	NA	5,741	5,844	5,37
Bt.6079.1.S1.a.at	LN	8,163	8,164	8,618	8,173	8,549	NA	8,018	NA	8,263	8,396	8,105
Bt.6079.2.S1.a.at	LN	7,977	7,224	8,259	7,892	8,207	NA	7,502	NA	8,132	8,118	7,646
Bt.5384.1.S1.at	LN	10,258	9,834	10,267	9,863	10,314	NA	9,77	NA	10,129	9,832	9,662
Bt.13143.1.A1.at	LN	5,449	4,859	4,798	4,631	5,192	NA	5,324	NA	5,002	4,914	4,657
Bt.21019.1.S1.at	LN	9,312	9,295	9,519	9,571	9,786	NA	9,201	NA	9,143	9,16	8,937
Bt.9048.2.S1.a.at	LN	10,303	10,256	10,518	10,525	10,521	NA	10,278	NA	10,321	10,438	10,272
Bt.27918.1.S1.at	LN	6,075	6,269	6,496	6,093	5,908	NA	6,023	NA	6,195	6,287	5,804
Bt.27569.1.A1.at	LN	6,184	6,908	6,352	6,668	6,197	NA	6,876	NA	6,46	6,149	6,367
Bt.20510.1.S1.at	LN	7,315	7,641	7,778	7,921	8,022	NA	7,4	NA	7,325	7,309	7,13
Bt.14202.1.A1.at	LN	5,937	6,029	6,382	6,002	6,059	NA	5,91	NA	5,654	5,606	5,72
Bt.14202.2.S1.at	LN	5,049	4,929	5,281	5,052	5,428	NA	5,224	NA	5,557	5,584	5,096
Bt.24735.1.A1.at	LN	7,735	8,1	7,674	7,892	7,637	NA	8,456	NA	7,509	7,828	7,754
Bt.28412.1.A1.at	LN	6,088	5,941	6,274	6,33	6,117	NA	6,106	NA	5,618	5,151	5,731
Bt.3059.1.S1.at	LN	7,083	6,625	7,31	6,991	7,297	NA	6,741	NA	7,407	7,318	7,215
Bt.12204.1.S1.at	LN	8,034	8,378	7,785	8,442	7,796	NA	7,9	NA	7,651	8	7,907
Bt.75.1.S1.at	LN	11,677	11,751	11,781	11,627	11,833	NA	11,668	NA	11,634	11,692	11,522
Bt.25310.1.A1.at	LN	5,523	5,285	5,38	5,796	5,123	NA	6,536	NA	5,382	5,176	5,991
Bt.13950.1.A1.at	LN	6,626	7,158	7,11	6,672	6,903	NA	6,936	NA	7,014	6,923	6,812
Bt.4486.1.S1.a.at	LN	8,296	7,626	8,491	8,363	8,929	NA	7,914	NA	7,633	7,745	7,373

Fortsetzung auf der nächsten Seite

Probeset	Gewebe	QQ						qq				
		1	2	3	4	5	6	7	8	9	10	11
Bt.15844.1.S1.at	LN	9,989	9,949	10,25	9,84	10,183	NA	9,934	NA	10,2	10,155	10,039
Bt.2661.1.S1.at	LN	7,165	7,399	7,362	7,222	7,706	NA	7,312	NA	7,196	7,294	7,023
Bt.1960.2.S1.a.at	LN	5,87	4,568	5,751	5,213	5,754	NA	5,043	NA	5,728	5,764	5,332
Bt.7919.1.A1.at	LN	6,996	7,023	7,629	7,595	7,743	NA	6,891	NA	7,116	6,942	6,347
Bt.5423.1.S1.at	LN	8,341	8,472	8,131	8,506	8,3	NA	7,898	NA	8,283	8,459	7,892
Bt.13207.1.S1.at	LN	8,515	8,873	8,714	8,411	8,592	NA	8,349	NA	8,454	8,419	8,005
Bt.22161.1.S1.at	LN	7,303	7,026	7,226	7,236	7,547	NA	6,984	NA	7,137	7,004	6,973
Bt.22265.1.S1.at	LN	9,044	10,07	9,681	9,903	9,429	NA	9,412	NA	10,073	9,924	9,079
Bt.7019.1.S1.at	LN	8,02	8,229	8,393	7,95	8,382	NA	7,871	NA	8,103	7,789	7,378
Bt.18686.1.A1.at	LN	5,453	6,402	5,975	5,965	6,039	NA	6,223	NA	5,872	6,177	5,954
Bt.19400.1.A1.at	LN	6,494	6,299	6,259	7,14	6,518	NA	7,186	NA	6,27	6,177	6,977
Bt.20087.1.S1.at	LN	6,854	6,123	6,505	6,596	6,545	NA	5,794	NA	6,281	6,244	6,143
Bt.4040.1.S1.at	LN	8,445	8,533	8,549	8,635	8,781	NA	8,343	NA	8,55	8,756	8,071
Bt.4338.1.S1.at	FR	6,628	6,465	6,874	6,842	6,57	NA	6,434	6,7	6,506	6,543	6,619
Bt.4436.1.S1.a.at	FR	4,897	5,115	4,782	4,815	4,82	NA	4,258	4,725	4,794	5,012	4,988
Bt.8939.1.S1.at	FR	7,732	7,644	8,068	7,789	8,138	NA	8,083	8,014	8,82	8,709	8,539
Bt.10027.1.S2.at	FR	7,41	7,452	7,231	7,333	7,273	NA	7,155	7,301	7,497	7,308	7,007
Bt.17669.1.S1.a.at	FR	6,31	6,051	6,194	5,853	6,162	NA	6,061	6,276	6,416	6,068	6,085
Bt.4943.2.S1.at	FR	7,18	7,113	7,153	6,873	7,149	NA	6,679	6,946	7,369	7,007	7,175
Bt.16930.1.A1.at	FR	8,652	9,099	9,167	8,225	8,757	NA	8,351	8,446	8,723	8,376	8,7
Bt.1973.1.S1.at	FR	11,154	11,207	11,403	11,067	11,268	NA	11,312	11,328	11,447	11,453	11,46
Bt.2831.1.S1.at	FR	6,375	6,614	6,485	6,532	6,813	NA	6,373	6,455	7,061	6,794	6,83
Bt.13505.1.S1.at	FR	10,355	10,767	10,517	10,594	10,745	NA	10,639	10,778	10,78	10,554	11,065
Bt.26582.2.S1.at	FR	3,531	3,826	3,492	4,176	4,439	NA	4,602	4,377	4,228	4,532	4,79
Bt.2466.1.S1.at	FR	9,106	9,289	9,325	9,144	9,387	NA	9,283	9,327	9,601	9,585	9,501
Bt.29542.1.A1.at	FR	6,062	6,117	6,073	5,865	5,79	NA	5,862	5,706	5,753	6,344	6,364
Bt.18570.1.S1.at	FR	7,84	7,476	7,683	7,635	7,813	NA	7,679	7,781	7,485	6,93	7,641
Bt.20340.2.S1.at	FR	11,355	11,155	11,417	11,58	11,564	NA	11,644	11,567	11,117	11,166	11,356
Bt.20340.3.S1.at	FR	8,93	8,59	8,957	8,986	9,19	NA	9,365	9,376	8,787	8,31	8,819
Bt.24863.1.S1.at	FR	7,414	7,118	7,663	7,283	7,049	NA	6,962	7,379	7,16	7,051	7,251
Bt.9656.1.S1.at	FR	6,617	6,562	6,8	6,836	6,726	NA	6,465	6,573	5,98	6,424	6,527
Bt.28795.1.A1.at	FR	7,393	7,467	7,437	7,283	7,332	NA	7,454	7,471	7,572	7,602	7,618
Bt.3082.1.S1.at	FR	10,47	10,259	10,555	10,475	10,475	NA	10,656	10,689	10,408	10,218	10,338
Bt.20096.2.S1.a.at	FR	8,937	8,942	9,121	8,777	9,424	NA	9,399	9,485	9,339	8,881	9,54
Bt.29377.1.S1.at	FR	9,703	9,72	9,871	9,915	10,041	NA	10,069	10,033	10,019	9,97	10,148
Bt.4294.1.S1.at	FR	8,408	8,645	8,628	8,662	8,73	NA	8,973	8,907	8,857	8,742	8,971
Bt.4589.1.S1.at	FR	5,412	5,977	5,537	6,115	5,645	NA	5,567	5,491	5,838	5,985	5,685
Bt.16945.1.S1.at	FR	8,822	9,214	9,127	9,046	9,104	NA	9,155	9,306	9,732	9,326	9,575
Bt.13633.1.A1.at	FR	6,927	6,806	6,746	6,589	6,39	NA	6,168	6,514	6,558	6,493	6,749
Bt.23512.1.S1.at	FR	11,203	11,343	11,351	11,213	11,253	NA	11,389	11,314	11,423	11,514	11,45
Bt.21775.1.S1.at	FR	7,641	7,535	7,618	7,528	7,673	NA	7,495	7,358	7,71	7,466	7,537
Bt.4533.1.S1.at	FR	9,951	10,322	10,261	10,101	10,227	NA	10,338	10,253	10,511	10,415	10,301
Bt.28639.1.S1.at	FR	9,382	9,617	9,43	9,647	9,555	NA	9,764	9,659	9,466	9,79	9,732
Bt.7478.1.S1.at	FR	11,384	11,295	11,241	11,275	11,261	NA	11,176	11,38	10,816	10,629	10,885
Bt.6079.1.S1.a.at	FR	8,888	8,97	8,911	8,732	9,083	NA	9,257	9,163	9,143	8,982	9,221
Bt.6079.2.S1.a.at	FR	8,812	8,788	8,938	8,651	9,035	NA	9,091	9,112	9,116	8,98	9,096
Bt.5384.1.S1.at	FR	9,218	8,766	9,315	8,909	8,88	NA	9,464	9,45	9,282	9,033	9,193
Bt.13143.1.A1.at	FR	5,604	5,288	5,358	5,366	5,154	NA	4,891	5,093	4,644	5,187	5,334
Bt.21019.1.S1.at	FR	9,426	9,372	9,307	9,516	9,558	NA	9,285	9,498	9,324	9,038	9,346
Bt.9048.2.S1.a.at	FR	10,295	10,527	10,606	10,626	10,656	NA	10,607	10,678	10,755	10,808	10,713
Bt.27918.1.S1.at	FR	8,042	7,912	7,78	8,119	8,252	NA	8,053	8,08	7,876	7,968	8,191
Bt.27569.1.A1.at	FR	6,108	6,489	6,066	6,41	6,308	NA	6,779	6,703	6,085	6,595	6,498
Bt.20510.1.S1.at	FR	10,878	11,073	11,211	11,187	11,107	NA	11,357	11,37	10,812	11,252	10,976
Bt.14202.1.A1.at	FR	8,689	8,689	8,545	8,755	8,824	NA	8,964	9,051	8,167	8,3	8,419
Bt.14202.2.S1.at	FR	7,224	7,018	7,331	7,4	7,47	NA	7,456	7,601	6,747	6,868	7,044
Bt.24735.1.A1.at	FR	9,204	9,284	9,404	9,195	9,365	NA	9,839	9,457	9,164	9,487	9,207
Bt.28412.1.A1.at	FR	6,258	5,85	5,945	5,971	6,076	NA	5,709	5,732	5,896	4,84	5,897
Bt.3059.1.S1.at	FR	6,934	6,847	6,829	6,776	6,983	NA	6,924	7,253	7,481	7,11	7,465
Bt.12204.1.S1.at	FR	8,632	8,949	8,72	8,751	8,741	NA	8,539	8,644	8,392	8,499	8,136
Bt.75.1.S1.at	FR	11,102	11,377	11,339	11,108	11,305	NA	11,443	11,423	11,567	11,603	11,575
Bt.25310.1.A1.at	FR	6,537	6,943	6,539	7,007	7,034	NA	7,043	6,91	6,98	6,93	6,958
Bt.13950.1.A1.at	FR	5,723	6,126	5,862	6,275	6,256	NA	6,224	6,204	6,308	6,463	6,437
Bt.4486.1.S1.a.at	FR	7,133	7,48	7,404	7,568	7,747	NA	7,434	7,841	7,632	7,754	7,662
Bt.15844.1.S1.at	FR	10,617	10,606	10,655	10,631	10,864	NA	10,845	10,709	10,816	10,638	10,801
Bt.2661.1.S1.at	FR	7,796	7,602	7,706	7,626	7,732	NA	7,597	7,724	7,512	7,315	7,534
Bt.1960.2.S1.a.at	FR	5,291	5,66	5,426	5,713	5,869	NA	6,303	5,879	5,744	5,796	6,097
Bt.7919.1.A1.at	FR	6,587	6,596	6,782	6,953	6,822	NA	6,343	6,651	6,63	6,734	6,839
Bt.5423.1.S1.at	FR	7,848	7,96	8,016	7,88	8,059	NA	8,192	8,093	8,096	8,072	8,082
Bt.13207.1.S1.at	FR	8,346	8,205	8,123	7,991	8,334	NA	8,198	8,148	8,376	8,021	8,111
Bt.22161.1.S1.at	FR	7,366	6,935	7,227	7,085	7,237	NA	6,866	7,118	7,693	6,895	7,51
Bt.22265.1.S1.at	FR	10,571	10,121	9,894	9,818	10,358	NA	9,628	9,55	9,72	10,255	10,035
Bt.7019.1.S1.at	FR	7,653	7,433	7,682	7,76	8,062	NA	8,095	8,027	7,897	7,914	8,038

Fortsetzung auf der nächsten Seite

Probeset	Gewebe	QQ						qq				
		1	2	3	4	5	6	7	8	9	10	11
Bt.18686.1.A1.at	FR	6,381	6,607	6,367	6,675	6,711	NA	6,947	6,877	6,685	6,886	6,594
Bt.19400.1.A1.at	FR	5,855	6,401	5,839	6,963	6,535	NA	6,877	6,768	6,364	7,049	6,93
Bt.20087.1.S1.at	FR	6,763	6,849	6,826	6,85	6,967	NA	7,265	7,304	6,989	7	7,173
Bt.4040.1.S1.at	FR	8,143	8,522	8,079	8,04	8,57	NA	8,319	8,315	9,047	8,641	8,761
Bt.4338.1.S1.at	DZ	6,767	6,718	6,526	6,769	6,532	NA	6,524	6,139	5,985	6,228	6,328
Bt.4436.1.S1.a.at	DZ	5,339	5,396	5,086	5,412	6,09	NA	5,496	5,351	5,229	5,086	5,601
Bt.8939.1.S1.at	DZ	6,975	7,012	6,898	6,802	6,47	NA	7,121	6,85	7,541	7,073	6,054
Bt.10027.1.S2.at	DZ	7,334	7,783	7,427	7,781	7,562	NA	7,299	6,977	7,581	7,268	6,938
Bt.17669.1.S1.a.at	DZ	4,926	4,939	5,403	5,064	5,085	NA	5,526	5,255	5,335	4,895	4,434
Bt.4943.2.S1.at	DZ	5,742	5,729	5,912	5,682	5,676	NA	6,272	6,125	6,084	6,289	6,072
Bt.16930.1.S1.at	DZ	7,321	7,363	7,635	7,484	7,101	NA	7,768	7,663	7,363	7,555	7,042
Bt.1973.1.S1.at	DZ	10,067	10,28	10,369	9,979	9,894	NA	10,424	10,442	10,456	10,481	10,131
Bt.2831.1.S1.at	DZ	7,633	7,249	7,112	6,893	7,1	NA	7,173	7,081	7,564	7,464	7,653
Bt.13505.1.S1.at	DZ	8,721	8,942	9,128	8,722	8,747	NA	9,007	9,023	9,399	9,116	8,995
Bt.26582.2.S1.at	DZ	3,812	3,662	4,246	3,834	3,911	NA	3,38	3,349	3,894	3,796	4,275
Bt.2466.1.S1.at	DZ	8,476	8,438	8,358	7,931	8,322	NA	8,429	8,533	8,694	8,42	8,557
Bt.29542.1.A1.at	DZ	5,834	5,965	5,66	5,854	5,942	NA	6,09	6,107	5,505	5,759	5,629
Bt.18570.1.S1.at	DZ	6,362	6,269	6,732	6,452	6,574	NA	7,059	6,742	6,095	6,463	6,353
Bt.20340.2.S1.at	DZ	6,819	7,657	8,657	7,894	7,666	NA	8,5	8,426	8,136	8,625	7,356
Bt.20340.3.S1.at	DZ	5,128	5,628	6,902	5,978	5,53	NA	5,95	6,325	5,805	6,733	5,789
Bt.24863.1.S1.at	DZ	7,197	6,984	7,09	7,117	7,656	NA	6,667	6,581	6,595	6,829	7,252
Bt.9656.1.S1.at	DZ	6,102	6,315	5,935	6,171	5,848	NA	5,871	6,051	5,575	5,675	5,462
Bt.28795.1.A1.at	DZ	6,131	5,975	6,299	6,154	5,861	NA	6,561	6,281	6,019	6,367	5,866
Bt.3082.1.S1.at	DZ	8,764	8,885	8,927	8,8	8,701	NA	8,772	9,027	9,032	9,118	8,597
Bt.20096.2.S1.a.at	DZ	6,889	7,199	7,24	6,98	6,897	NA	7,589	7,483	7,587	7,254	6,833
Bt.29377.1.S1.at	DZ	8,187	8,615	9,086	8,675	8,218	NA	8,807	8,827	8,667	8,811	8,314
Bt.4294.1.S1.at	DZ	7,952	8,039	8,13	8,046	7,784	NA	7,801	7,937	7,957	7,645	7,619
Bt.4589.1.S1.at	DZ	5,811	5,009	5,519	4,997	5,229	NA	5,274	5,155	5,244	5,282	5,779
Bt.16945.1.S1.at	DZ	7,962	8,04	8,081	7,931	7,719	NA	8,245	8,242	8,11	8,213	8,153
Bt.13633.1.A1.at	DZ	6,696	7,115	6,579	6,669	7,111	NA	6,913	6,734	8,373	6,38	7,061
Bt.23512.1.S1.at	DZ	10,375	10,431	10,549	10,234	10,222	NA	10,551	10,748	10,55	10,592	10,177
Bt.21775.1.S1.at	DZ	6,891	6,88	7,181	7,22	6,855	NA	7,184	7,165	7,215	7,116	6,938
Bt.4533.1.S1.at	DZ	9,356	9,279	9,337	9,308	9,231	NA	9,361	9,309	9,358	9,421	9,234
Bt.28639.1.S1.at	DZ	8,561	8,722	8,758	8,535	8,074	NA	8,922	8,983	8,956	8,723	8,719
Bt.7478.1.S1.at	DZ	7,859	8,853	9,273	8,797	8,142	NA	8,838	9,177	9,285	8,828	7,311
Bt.6079.1.S1.a.at	DZ	7,898	7,775	7,74	7,624	7,564	NA	7,808	7,859	7,869	7,863	7,848
Bt.6079.2.S1.a.at	DZ	8,007	7,958	7,946	7,81	7,857	NA	7,941	7,913	7,729	8,016	8,309
Bt.5384.1.S1.at	DZ	7,787	7,288	7,724	7,086	6,964	NA	7,857	7,933	7,778	7,532	7,47
Bt.13143.1.A1.at	DZ	5,758	5,727	5,638	5,891	5,821	NA	5,54	5,545	4,869	5,454	5,451
Bt.21019.1.S1.at	DZ	7,725	8,05	8,3	8,164	7,675	NA	8,395	8,384	8,424	8,144	7,83
Bt.9048.2.S1.a.at	DZ	9,401	9,67	9,631	9,525	9,48	NA	9,65	9,791	9,789	9,774	9,663
Bt.27918.1.S1.at	DZ	5,855	6,498	6,689	6,463	5,896	NA	6,945	6,665	6,712	6,999	6,081
Bt.27569.1.A1.at	DZ	5,343	5,579	5,49	5,705	5,405	NA	5,46	5,902	5,879	5,945	5,671
Bt.20510.1.S1.at	DZ	6,653	7,485	8,098	7,534	6,956	NA	8,14	8,337	7,793	8,335	6,717
Bt.14202.1.A1.at	DZ	5,448	5,021	5,447	5,64	5,148	NA	5,601	5,259	5,664	5,655	5,695
Bt.14202.2.S1.at	DZ	5,043	4,976	5,356	5,404	4,639	NA	5,384	5,249	5,526	5,446	5,519
Bt.24735.1.A1.at	DZ	7,262	7,568	7,926	7,553	7,635	NA	8,068	8,057	7,991	8,082	7,735
Bt.28412.1.A1.at	DZ	4,499	5,549	5,553	5,475	4,721	NA	5,027	5,466	5,926	5,193	5,428
Bt.3059.1.S1.at	DZ	6,761	6,187	6,669	6,231	6,465	NA	6,605	6,45	6,314	6,406	6,852
Bt.12204.1.S1.at	DZ	7,077	7,703	7,797	7,66	7,413	NA	7,501	7,841	8,071	7,935	7,243
Bt.75.1.S1.at	DZ	11,226	11,086	11,151	11,006	10,886	NA	10,9	11,049	11,202	11,051	11,172
Bt.25310.1.A1.at	DZ	5,823	6,271	5,98	6,092	5,289	NA	6,004	6,194	6,613	6,255	6,122
Bt.13950.1.A1.at	DZ	6,217	6,045	5,991	5,79	5,721	NA	6,082	5,551	5,853	5,916	6,092
Bt.4486.1.S1.a.at	DZ	6,583	6,716	6,635	6,245	6,202	NA	6,712	6,812	7,157	6,074	6,209
Bt.15844.1.S1.at	DZ	8,305	8,608	8,705	8,216	8,026	NA	8,993	8,922	8,857	8,956	8,34
Bt.2661.1.S1.at	DZ	6,859	7,065	6,981	7,021	6,565	NA	6,831	7,038	7,005	6,981	6,838
Bt.1960.2.S1.a.at	DZ	4,168	4,397	4,241	4,185	4,608	NA	4,158	4,316	4,439	4,612	4,85
Bt.7919.1.A1.at	DZ	7,024	7,121	6,721	7,025	7,28	NA	6,813	7,046	6,506	6,831	6,681
Bt.5423.1.S1.at	DZ	8,19	8,162	7,961	8,021	8,109	NA	8,069	8,145	8,148	8,146	7,978
Bt.13207.1.S1.at	DZ	7,4	7,227	7,241	7,052	6,984	NA	7,209	7,349	7,69	7,228	6,93
Bt.22161.1.S1.at	DZ	8,058	7,884	7,652	8,007	7,787	NA	7,915	7,35	7,944	7,633	8,114
Bt.22265.1.S1.at	DZ	8,543	10,631	10,622	10,022	10,254	NA	10,148	9,848	10,537	10,31	9,893
Bt.7019.1.S1.at	DZ	9,082	9,215	8,959	8,941	9,104	NA	8,716	8,903	9,555	9,085	9,029
Bt.18686.1.A1.at	DZ	5,591	5,974	6,045	5,727	5,608	NA	6,011	6,218	6,46	6,241	5,891
Bt.19400.1.A1.at	DZ	4,981	5,969	6,034	5,868	5,512	NA	5,936	6,153	6,411	6,276	6,044
Bt.20087.1.S1.at	DZ	6,547	6,234	6,447	6,19	6,219	NA	6,535	6,293	6,401	6,238	6,35
Bt.4040.1.S1.at	DZ	8,347	8,338	8,147	8,078	8,136	NA	8,417	8,433	8,396	8,34	8,655
Bt.4338.1.S1.at	Pa	6,967	NA	6,77	6,851	6,558	NA	6,931	7,334	6,857	6,923	6,52
Bt.4436.1.S1.a.at	Pa	5,524	NA	5,608	5,933	5,783	NA	5,722	5,851	5,485	5,496	5,533
Bt.8939.1.S1.at	Pa	6,556	NA	5,873	5,939	5,769	NA	5,723	5,809	6,018	5,965	5,966
Bt.10027.1.S2.at	Pa	8,075	NA	8,152	8,035	8,356	NA	8,126	7,727	8,42	8,323	8,467
Bt.17669.1.S1.a.at	Pa	4,652	NA	4,606	4,713	4,452	NA	4,607	5,008	4,541	4,49	4,389

Fortsetzung auf der nächsten Seite

Probeset	Gewebe	QQ						qq				
		1	2	3	4	5	6	7	8	9	10	11
Bt.4943.2.S1.at	Pa	5,196	NA	4,861	4,908	4,43	NA	5,099	4,407	5,097	5,532	5,511
Bt.16930.1.A1.at	Pa	7,055	NA	7,02	6,856	7,02	NA	7,119	6,921	6,979	7,059	7,024
Bt.1973.1.S1.at	Pa	9,828	NA	9,725	9,537	9,612	NA	9,803	9,69	9,81	9,496	9,831
Bt.2831.1.S1.at	Pa	6,782	NA	6,9	6,748	6,717	NA	7,223	7,043	7,273	6,903	7,184
Bt.13505.1.S1.at	Pa	8,094	NA	8,455	8,459	8,28	NA	8,218	8,444	8,281	8,481	8,693
Bt.26582.2.S1.at	Pa	3,405	NA	2,684	3,804	3,946	NA	3,622	3,22	3,972	3,635	3,792
Bt.2466.1.S1.at	Pa	8,349	NA	8,175	7,814	8,034	NA	8,226	8,009	8,334	7,972	8,486
Bt.29542.1.A1.at	Pa	5,971	NA	6,049	5,854	5,99	NA	5,727	5,91	5,847	5,703	5,701
Bt.18570.1.S1.at	Pa	5,896	NA	5,969	6,213	6,447	NA	5,961	6,313	5,922	6,138	6,233
Bt.20340.2.S1.at	Pa	6,945	NA	7,644	7,628	7,308	NA	7,371	7,474	7,231	7,71	7,025
Bt.20340.3.S1.at	Pa	5,876	NA	5,652	5,771	5,737	NA	5,555	5,959	5,748	5,719	5,209
Bt.24863.1.S1.at	Pa	6,35	NA	7,003	7,229	7,078	NA	7,177	6,866	7,19	7,232	7,169
Bt.9656.1.S1.at	Pa	5,9	NA	5,733	5,584	5,761	NA	5,97	5,461	5,988	5,668	5,484
Bt.28795.1.A1.at	Pa	5,575	NA	5,11	5,687	5,318	NA	5,565	5,633	5,429	5,703	4,992
Bt.3082.1.S1.at	Pa	8,465	NA	7,959	8,208	8,087	NA	8,268	8,199	8,247	8,227	8,27
Bt.20096.2.S1.a.at	Pa	5,627	NA	5,859	5,606	5,979	NA	5,857	6,254	6,141	5,642	5,85
Bt.29377.1.S1.at	Pa	7,547	NA	7,86	7,39	7,765	NA	7,667	7,702	7,537	7,847	7,867
Bt.4294.1.S1.at	Pa	7,681	NA	7,554	7,445	7,307	NA	7,585	7,447	7,437	7,546	7,664
Bt.4589.1.S1.at	Pa	5,026	NA	5,508	4,784	4,837	NA	4,833	4,861	4,846	4,334	5,398
Bt.16945.1.S1.at	Pa	7,03	NA	6,955	7,487	7,128	NA	7,384	7,346	7,228	7,154	7,62
Bt.13633.1.A1.at	Pa	6,927	NA	6,845	7,157	6,957	NA	6,965	6,843	7,052	6,775	6,979
Bt.23512.1.S1.at	Pa	10,231	NA	9,937	9,996	9,91	NA	10,081	10,215	9,859	9,794	9,957
Bt.21775.1.S1.at	Pa	6,223	NA	6,332	5,884	6,05	NA	6,322	6,251	6,369	6,335	6,405
Bt.4533.1.S1.at	Pa	9,014	NA	9,039	8,998	8,867	NA	9,098	8,935	9,068	8,985	9,079
Bt.28639.1.S1.at	Pa	8,068	NA	7,726	7,575	7,401	NA	7,829	7,758	7,69	7,693	8,171
Bt.7478.1.S1.at	Pa	5,915	NA	6,037	6,362	6,035	NA	5,963	6,174	6,465	6,485	6,23
Bt.6079.1.S1.a.at	Pa	7,305	NA	7,303	7,227	7,294	NA	7,447	7,364	7,3	7,286	7,465
Bt.6079.2.S1.a.at	Pa	7,911	NA	7,7	8,028	7,837	NA	8,042	7,84	7,87	7,806	8,136
Bt.5384.1.S1.at	Pa	7,14	NA	7,018	6,305	6,468	NA	6,957	7,023	6,935	6,564	7,032
Bt.13143.1.A1.at	Pa	5,408	NA	5,678	5,499	6,007	NA	5,497	5,85	5,699	5,841	5,719
Bt.21019.1.S1.at	Pa	7,731	NA	7,29	6,95	7,126	NA	7,358	7,056	7,342	7,046	7,356
Bt.9048.2.S1.a.at	Pa	9,278	NA	9,357	9,278	9,356	NA	9,541	9,468	9,545	9,19	9,389
Bt.27918.1.S1.at	Pa	4,593	NA	4,411	4,683	5,077	NA	4,784	4,787	4,873	4,682	4,965
Bt.27569.1.A1.at	Pa	5,662	NA	5,036	5,378	5,281	NA	5,011	5,164	4,742	5,156	5,528
Bt.20510.1.S1.at	Pa	5,065	NA	5,448	4,972	4,849	NA	5,158	5,326	5,242	5,037	4,974
Bt.14202.1.A1.at	Pa	4,393	NA	5,277	5,459	4,672	NA	5,005	5,544	5,406	5,562	5,284
Bt.14202.2.S1.at	Pa	4,866	NA	5,12	5,159	4,924	NA	4,637	4,889	4,545	5,157	4,758
Bt.24735.1.A1.at	Pa	7,055	NA	7,055	6,949	6,89	NA	7,476	7,157	7,131	7,175	7,211
Bt.28412.1.A1.at	Pa	3,816	NA	4,209	3,904	4,068	NA	4,454	4,27	4,686	4,45	4,386
Bt.3059.1.S1.at	Pa	6,336	NA	6,678	6,466	6,383	NA	6,223	6,481	6,371	6,669	6,654
Bt.12204.1.S1.at	Pa	6,196	NA	6,541	6,238	5,932	NA	6,054	5,736	6,216	6,152	6,433
Bt.75.1.S1.at	Pa	11,026	NA	11,33	11,123	10,856	NA	11,173	10,854	11,275	10,824	11,116
Bt.25310.1.A1.at	Pa	5,646	NA	5,565	5,613	5,446	NA	5,855	5,748	5,791	5,68	5,776
Bt.13950.1.A1.at	Pa	5,55	NA	5,634	5,327	5,671	NA	5,598	5,832	5,822	5,567	5,921
Bt.4486.1.S1.a.at	Pa	6,083	NA	5,839	5,918	5,224	NA	5,85	5,779	5,964	5,781	6,038
Bt.15844.1.S1.at	Pa	7,738	NA	7,411	7,205	7,323	NA	7,599	7,479	7,79	7,335	7,841
Bt.2661.1.S1.at	Pa	6,815	NA	6,886	6,906	6,872	NA	6,927	7,001	7,182	6,957	7,017
Bt.1960.2.S1.a.at	Pa	4,11	NA	4,573	4,57	4,389	NA	4,511	4,116	4,515	4,455	4,604
Bt.7919.1.A1.at	Pa	6,785	NA	6,924	6,679	6,991	NA	6,794	6,642	6,284	6,877	6,832
Bt.5423.1.S1.at	Pa	8,163	NA	8,275	8,041	8,228	NA	8,266	8,316	8,247	8,098	7,977
Bt.13207.1.S1.at	Pa	7,074	NA	6,57	6,337	6,334	NA	6,45	6,536	6,715	6,473	6,765
Bt.22161.1.S1.at	Pa	9,143	NA	8,658	8,981	9,123	NA	9,163	8,534	9,053	8,587	8,756
Bt.22265.1.S1.at	Pa	7,23	NA	7,303	7,396	7,834	NA	7,779	7,866	7,654	7,828	7,671
Bt.7019.1.S1.at	Pa	9,357	NA	9,503	9,415	9,423	NA	9,415	9,483	9,617	9,367	9,463
Bt.18686.1.A1.at	Pa	4,627	NA	4,933	4,774	5,04	NA	5,337	4,86	4,944	4,755	5,202
Bt.19400.1.A1.at	Pa	5,185	NA	4,444	5,056	3,272	NA	4,792	4,858	4,571	4,065	5,192
Bt.20087.1.S1.at	Pa	6,09	NA	6,409	6,338	5,94	NA	6,195	6,469	5,916	6,272	6,555
Bt.4040.1.S1.at	Pa	8,254	NA	7,765	7,604	7,834	NA	8,201	7,951	8,064	7,447	8,399
Bt.4338.1.S1.at	SK	6,249	6,446	6,352	6,321	6,526	NA	6,41	6,345	6,346	6,501	6,325
Bt.4436.1.S1.a.at	SK	4,446	5,651	4,911	4,638	4,959	NA	4,931	4,335	4,46	4,867	4,771
Bt.8939.1.S1.at	SK	7,133	6,846	7,592	7,773	7,655	NA	7,352	7,6	7,541	8,117	7,867
Bt.10027.1.S2.at	SK	6,772	6,425	6,859	6,659	6,94	NA	6,711	6,876	6,741	6,518	6,516
Bt.17669.1.S1.a.at	SK	4,68	4,947	5,345	5,081	5,233	NA	5,192	5,299	4,945	5,112	5,054
Bt.4943.2.S1.at	SK	6,364	5,972	6,217	6,048	6,426	NA	5,976	6,005	6,502	6,384	6,236
Bt.16930.1.A1.at	SK	8,049	8,242	8,043	8,019	8,034	NA	7,524	7,659	7,835	8,139	7,902
Bt.1973.1.S1.at	SK	11,329	11,232	11,421	11,234	11,218	NA	11,279	11,39	11,434	11,279	11,373
Bt.2831.1.S1.at	SK	7,386	7,038	7,204	7,313	6,861	NA	6,74	6,635	7,535	7,355	7,369
Bt.13505.1.S1.at	SK	10,957	11,041	10,889	10,876	10,737	NA	10,768	10,806	11,046	10,76	11,189
Bt.26582.2.S1.at	SK	3,588	3,03	3,763	3,787	4,154	NA	4,08	3,018	4,179	4,391	4,457
Bt.2466.1.S1.at	SK	9,762	9,672	9,921	9,69	9,675	NA	9,585	9,663	9,92	9,777	9,949
Bt.29542.1.A1.at	SK	5,582	5,511	5,282	5,65	5,578	NA	5,596	5,758	5,66	5,552	5,538
Bt.18570.1.S1.at	SK	7,1	6,994	6,535	7,096	7,312	NA	7,213	7,144	7,622	7,635	8,124

Fortsetzung auf der nächsten Seite



Probeset	Gewebe	QQ						qq				
		1	2	3	4	5	6	7	8	9	10	11
Bt.20340.2.S1.at	SK	9,693	9,35	9,644	9,875	10,248	NA	10,625	10,664	9,802	9,848	9,849
Bt.20340.3.S1.at	SK	7,192	6,989	6,431	7,53	7,699	NA	8,296	8,119	7,721	7,739	7,875
Bt.24863.1.S1.at	SK	6,721	6,29	6,499	6,361	6,669	NA	6,861	6,474	6,303	6,165	6,328
Bt.9656.1.S1.at	SK	6,83	6,888	6,849	6,508	6,673	NA	6,661	6,536	6,215	6,622	6,39
Bt.28795.1.A1.at	SK	6,721	6,175	6,348	6,369	6,675	NA	6,844	6,953	6,603	6,733	6,47
Bt.3082.1.S1.at	SK	9,677	9,497	9,797	9,754	9,933	NA	10,047	10,123	9,315	9,619	9,804
Bt.20096.2.S1.a.at	SK	8,095	7,654	7,804	7,837	8,303	NA	8,425	8,729	8,165	8,121	8,597
Bt.29377.1.S1.at	SK	9,486	9,341	9,412	9,465	9,469	NA	9,443	9,611	9,57	9,464	9,883
Bt.4294.1.S1.at	SK	8,322	8,243	8,494	8,189	8,28	NA	8,553	8,392	8,171	8,166	8,264
Bt.4589.1.S1.at	SK	6,934	6,96	6,943	6,924	6,343	NA	6,499	6,021	6,68	6,501	6,316
Bt.16945.1.S1.at	SK	9,842	9,977	10,232	9,683	9,774	NA	9,821	9,741	9,817	9,59	9,91
Bt.13633.1.A1.at	SK	6,322	5,238	6,157	6,174	6,441	NA	6,403	6,245	6,288	6,218	6,104
Bt.23512.1.S1.at	SK	11,363	11,346	11,368	11,382	11,392	NA	11,41	11,488	11,539	11,48	11,563
Bt.21775.1.S1.at	SK	7,617	7,124	7,362	7,528	7,378	NA	7,287	7,295	7,707	7,361	7,806
Bt.4533.1.S1.at	SK	10,841	10,771	10,932	10,586	10,671	NA	10,693	10,623	10,798	10,652	10,722
Bt.28639.1.S1.at	SK	8,929	8,926	9,027	9,299	9,122	NA	9,037	9,206	8,865	9,241	9,413
Bt.7478.1.S1.at	SK	9,646	9,799	9,803	10,002	10,489	NA	10,504	10,764	9,596	9,707	9,82
Bt.6079.1.S1.a.at	SK	9,015	8,931	9,014	8,855	8,847	NA	8,858	9,002	8,862	8,83	8,916
Bt.6079.2.S1.a.at	SK	8,867	8,837	8,833	8,851	8,891	NA	8,954	8,931	8,96	8,946	9,069
Bt.5384.1.S1.at	SK	8,5	8,193	8,855	8,496	8,521	NA	8,813	8,961	8,951	8,629	8,815
Bt.13143.1.A1.at	SK	3,991	3,534	3,78	3,576	3,482	NA	4,392	3,846	3,542	3,689	3,754
Bt.21019.1.S1.at	SK	8,928	8,881	8,945	8,961	8,845	NA	8,779	8,763	8,769	8,865	8,802
Bt.9048.2.S1.a.at	SK	10,401	10,526	10,614	10,601	10,573	NA	10,546	10,634	10,598	10,527	10,696
Bt.27918.1.S1.at	SK	6,555	6,68	6,483	6,906	7,078	NA	7,015	7,279	7,006	7,057	7,089
Bt.27569.1.A1.at	SK	5,71	6,044	5,957	6,248	5,784	NA	5,954	6,033	5,738	6,489	6,532
Bt.20510.1.S1.at	SK	9,208	9,411	9,486	9,888	10,091	NA	10,242	10,53	9,585	9,907	9,471
Bt.14202.1.A1.at	SK	7,719	7,342	7,468	7,762	7,994	NA	7,929	8,274	7,652	7,76	7,509
Bt.14202.2.S1.at	SK	6,405	5,949	6,304	6,291	6,559	NA	6,669	6,712	6,333	6,454	6,158
Bt.24735.1.A1.at	SK	8,719	8,737	8,874	8,929	9,044	NA	9,034	8,963	8,678	9,108	8,673
Bt.28412.1.A1.at	SK	5,64	5,082	5,131	4,722	5,472	NA	4,996	5,22	5,349	4,405	4,894
Bt.3059.1.S1.at	SK	7,127	6,993	6,984	6,744	6,84	NA	7,009	7,023	7,237	7,159	7,558
Bt.12204.1.S1.at	SK	8,538	8,64	8,733	8,576	8,569	NA	8,3	8,279	8,342	8,453	8,084
Bt.75.1.S1.at	SK	11,583	11,53	11,619	11,436	11,446	NA	11,503	11,456	11,529	11,427	11,463
Bt.25310.1.A1.at	SK	7,205	6,996	7,133	6,559	6,773	NA	6,835	6,884	7,002	7,09	7,215
Bt.13950.1.A1.at	SK	6,343	6,399	6,38	6,142	5,964	NA	6,227	6,07	6,344	6,027	6,466
Bt.4486.1.S1.a.at	SK	7,128	7,096	7,322	7,268	7,386	NA	7,059	7,542	6,934	7,341	7,6
Bt.15844.1.S1.at	SK	10,545	10,552	10,667	10,509	10,615	NA	10,57	10,537	10,736	10,705	10,816
Bt.2661.1.S1.at	SK	7,766	7,666	7,692	7,533	7,52	NA	7,4	7,456	7,476	7,393	7,601
Bt.1960.2.S1.a.at	SK	4,703	4,71	4,253	4,722	4,796	NA	4,891	5,023	5,125	5,263	5,321
Bt.7919.1.A1.at	SK	6,162	6,595	6,325	6,284	6,473	NA	6,331	6,381	6,506	6,244	6,191
Bt.5423.1.S1.at	SK	7,2	7,123	7,282	7,083	7,23	NA	7,509	7,458	7,478	7,321	7,355
Bt.13207.1.S1.at	SK	8,444	8,135	8,278	8,033	8,15	NA	8,066	7,913	8,34	8,187	8,317
Bt.22161.1.S1.at	SK	6,323	6,541	6,529	6,717	6,826	NA	6,497	6,693	6,613	6,539	6,759
Bt.22265.1.S1.at	SK	10,726	10,336	9,28	10,733	9,192	NA	10,606	8,74	8,952	9,11	9,372
Bt.7019.1.S1.at	SK	7,101	7,185	7,415	7,535	7,398	NA	7,374	7,62	7,492	7,367	7,259
Bt.18686.1.A1.at	SK	5,962	5,723	5,467	5,799	5,868	NA	5,903	6,013	5,451	5,894	5,891
Bt.19400.1.A1.at	SK	6,255	6,347	6,096	7,091	6,146	NA	6,517	6,548	6,368	6,735	6,881
Bt.20087.1.S1.at	SK	6,961	6,88	7,055	6,752	6,942	NA	6,883	6,981	6,73	6,678	6,932
Bt.4040.1.S1.at	SK	9,174	9,072	8,993	8,866	8,702	NA	8,625	8,685	8,915	8,591	9,067

**Tabelle 31:** Angelehnt an Ron & Weller (2007), aber mit erweitertem Kriterienkatalog stellten Ogorevc *et al.* (2009) durch Analyse von Datenbanken eine Metadatenbank zusammen, die den seinerzeitigen (Dezember 2008) Stand der im Zusammenhang mit Mastitis untersuchten Kandidatenloci abbildet. Links ist die gesamtgenomische Perspektive in einer Tabelle dargestellt, während die Abbildung rechts die Verhältnisse für BTA18 detaillierter zeigt; die dort verwendete Farbcodierung entspricht der der Tabelle. Nur für vier der acht einbezogenen Kategorien, darunter drei auf der Ebene von Genen sowie QTL-Kartierungen wurden auf BTA18 Kandidaten identifiziert. Die ausgewerteten Transkriptomstudien (●), mit der Ausnahme von Pfaffl *et al.* (2003), untersuchten infizierte vs. nichtinfizierte Individuen; dieses Design ist wegen der auftretenden Confounding-Effekte suboptimal für die Dissektion genetischer Resistenz (vgl. Kap. I.2.2.2).

Kriterium		gesamt
Knock-Outs und Transgene	●	143
SCS-QTL bzw. QTL für klinische Mastitis	●	71
SNP-Assoziationen	●	10
AFLP Marker Assoziationen		27
Transkriptomstudien	●	107
genetisch variable Milchproteine		9
in den Milchdrüsen exprimierte miRNAs		32
Epigenetische Faktoren		1
<b>Total</b>		<b>391</b>

**BTA18**

- NOD2
- CDH3
- CDH1
- HP
- USF2
- CEBPB
- RELB
- KCNN4
- BAX
- C5AR1

cM

centromer (0 cM)

BMS2213 (24,5 cM)

INRA121 (30,15 cM)

BM8151 (40,21 cM)

BM7109 (47 cM)

ILSTS002 (54,8 cM)

BMS078 (76,8 cM)

TGLA227 (84,1 cM)

0  
5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60  
65  
70

**Tabelle 32:** Merkmalsbeeinflussende Regionen (QTL) für die Merkmale SCC und SCS in verschiedenen Milchviehrassen sind in der *Animal Quantitative Trait Loci Database* (AnimalQTLdb, Hu *et al.*, 2007) zum Abruf verfügbar. Mit Stand vom 11.03.2010 wurden 29 einschlägige Publikationen gefunden, von denen SNP-Assoziationsstudien und Feinkartierungen von QTL nicht in die Tabelle aufgenommen wurden.

Milchviehrasse	Chromosom(en) (BTA)	Referenz
Israelisches Holstein	7	Ron <i>et al.</i> (2004)
Dänisches Holstein	5, 6, 8, 13, 22, 23, 24	Lund <i>et al.</i> (2008)
Finnish Ayrshire, Schwedisches Rot-Weiß, Dänisches Rotvieh	9, 11, 14, 18	Lund <i>et al.</i> (2007)
US Holstein	18	Muncie <i>et al.</i> (2006)
Israelisches Holstein	7	Weller <i>et al.</i> (2008)
Finnish Ayrshire	3, 29, 1, 11, 14, 18, 21, 23, 24, 27	Schulman <i>et al.</i> (2004)
Französisches Holstein, Normandie, Montbeliarde	9, 10, 15, 21, 23	Boichard <i>et al.</i> (2003)
Holstein	5, 7, 15, 20, 23, 26, 29	Ashwell <i>et al.</i> (2004)
<i>multipel</i>	1, 3, 4, 5, 8, 9, 10, 13, 14, 15, 18, 21, 22, 23, 26	Rupp & Boichard (2003), Metaanalyse
Holstein	4, 11, 14, 26	Zhang <i>et al.</i> (1998)
US Holstein	12	Ashwell <i>et al.</i> (1998)
US Holstein	5, 7, 21, 22, 23	Heyen <i>et al.</i> (1999)
US Holstein	14	Rodriguez-Zas <i>et al.</i> (2002)
Holstein	2, 18, 19	Bennewitz <i>et al.</i> (2003)
Holstein	1, 2, 3, 4, 5, 6, 11, 14, 16, 19, 23, 25, 27, 29	Daetwyler <i>et al.</i> (2008)
Italienisches Holstein	4, 26	Longeri <i>et al.</i> (2006)
US Holstein	8, 10, 11, 21	Schnabel <i>et al.</i> (2005b)
Schwedisches Rot-Weiß, Schwedisches Holstein	5, 9, 11, 23	Holmberg & Andersson-Eklund (2004)
Norwegisches Rotvieh (NRF)	3, 4, 6, 8	Klungland <i>et al.</i> (2001)
Deutsche Holstein	7, 10, 18, 27	Kühn <i>et al.</i> (2003)
Holstein Friesian	3, 18	Schrooten <i>et al.</i> (2000)
Kanadisches Holstein	1, 13, 18, 23, 26, 28, 29	Kolbehdari <i>et al.</i> (2009)

**Tabelle 33:** Die für die Erstellung dieser Arbeit entwickelten Perlskripte nutzten teilweise die durch externe Module bereitgestellten zusätzlichen Funktionen. Einige dieser Module entstammen der bioinformatisch orientierten Projekt BioPerl (Stajich *et al.*, 2002), andere dem CPAN (*Comprehensive Perl Archive Network*).

Paket	Modul	Autor(en)	Einsatzzweck
DBI		Tim Bunce	Anbindung an Datenbanken, z.B. MySQL
SeqIO		Ewan Birney, Lincoln Stein	BioPerl; Repräsentation annotierter DNA-Sequenzen
SearchIO	blast	Jason Stajich	BioPerl; Textparser für gängige bioinformatische Programme
LWP	UserAgent	Gisle Aas, Martin Koster	Zugriff auf Onlineressourcen via URL
Data	Dumper	Gurusamy Sarathy	physikalische Speicherung von Datenstrukturen

**Tabelle 34:** Die für die Generierung und anschließende Sequenzierung der Amplikons, welche die identifizierten SNPs innerhalb der potentiellen funktionellen Kandidatengene enthielten, generierten genomischen Primer in 5'-3'-Orientierung.  $T_m$ : Schmelztemperatur, GC (%): Anteil von G und C an der Sequenz in %.

Gen	Typ	Sequenz (in 5'-3'-Orientierung)	$T_m$ (°C)	GC (%)
CD22	forward	ATCTGGGCCTCCTGAATTG	60,02	52,63
	reverse	TGTCACACTGGCAGGAGAAG	60,02	55
	reverse	GATGTTTCCCCACCTTCCTC	60,69	55,00
	forward	ACAGTCCATAGGGTCGCAAG	60,13	55,00
	forward	CAAGGAGTGGCTCCAGACTC	59,99	60
	reverse	TTTCTGCTTGCTGTCCCTTC	60,52	50
	reverse	TTTCTGCTTGCTGTCCCTTC	60,52	50
	forward	CAAGGAGTGGCTCCAGACTC	59,99	60
	reverse	CGCCTTAGCACCAGCTTTAC	60,04	55
	forward	TTGCTCCCTGAGAACTCTGC	60,68	55
	reverse	GATTCAGGCACCCGAAGAG	60,75	57,89
	forward	GATTAGTGATTGGAGCCAGGAC	59,97	50
	reverse	CCTCTTTCTGCCCTCCTACC	59,51	55
	forward	AGCACAAGTTCACACGCAAG	60,1	50
	reverse	CCTCTTTCTGCCCTCCTACC	59,51	55
	forward	AGCACAAGTTCACACGCAAG	60,1	50
HCST	forward	ATAAACTTCTGGGCCGCTTC	60,58	50
	reverse	GCAGTGACACAACAGCATCC	60,33	55
	reverse	AGCCCTGTCCCTACTCTCAC	58,34	60
	forward	CTTCAGGTCCCTCCTCCATTC	60,06	52,38
	reverse	AGCCCTGTCCCTACTCTCAC	58,34	60
	forward	CTTCAGGTCCCTCCTCCATTC	60,06	52,38
	forward	ACCCTTCAGGTTCCCTGTTC	60,35	55
	reverse	ATTTGGTCCATACCCTGGTG	59,53	50
TYROBP	forward	CGCCAGCATCTAGCCTAAAC	60,00	55,00
	reverse	TGTACGGAGGGGAAATTCAG	59,93	50,00
CD79A	reverse	CTGAGCATCTCCGATGTGG	60,37	57,89
	forward	AATGCAGGAGACGTTGGTTC	60,12	50,00
	reverse	CTGAGCATCTCCGATGTGG	60,37	57,89
	forward	AATGCAGGAGACGTTGGTTC	60,12	50,00
	reverse	CTCAGCAATCTCCCAGTTCC	59,80	55,00
	forward	TCTCCTTTCCAACCCCAAG	60,03	52,63

**Tabelle 35:** Für die Typisierung an einer Halbgeschwisterpopulation wurden sechs SNPs ausgewählt (vgl. Tabelle 26). Die Nukleotidsequenzen der für die Pyrosequenzierung der SNPs generierten PCR- und Pyrosequenzierungsprimer sind in dieser Tabelle jeweils in 5'-3'-Orientierung angegeben. Die beiden SNPs im Gen HCST lagen im selben Amplikon, so daß hier nur ein PCR-Primerpaar zum Einsatz kam. Die Abk. *Sequenz.* bezeichnet den tatsächlich für die Sequenzierreaktion benutzten Primer; mit den mittels *PCR* gekennzeichneten Primern wurden die Fragmente amplifiziert. Es bezeichnen weiterhin: fwd.: *forward*, rev.: *reverse*,  $T_m$ : Schmelztemperatur, GC (%): Anteil von G und C an der Sequenz in %.

SNP	Zweck	Typ	Sequenz	$T_m$ (°C)	GC (%)
CD22_7568:C>T	PCR	fwd.	TCTCCTTTCTTGCTCCCTGA	60,1	50
	PCR	rev.	TGCAGGCAGATTCTTTACCA	59,4	45
	Sequenz.	fwd.	TGTGCACAATAGGCC		
CD22_8067:A>G	PCR	fwd.	TCTGGAAACATAGGGGGAAA	59,4	45
	PCR	rev.	GAGACCTAGGGAACGTGGTG	59,6	60
	Sequenz.	fwd.	ACCTAAGTGTGTCATTTCG		
CD79A_2808:C>T	PCR	fwd.	CCGAGCGACTAAAACAACAA	59	45
	PCR	rev.	TTCTGCCATCGTTTCTGTTG	59,8	45
	Sequenz.	fwd.	TCTATTTGTGGGGAAGA		
CD79A_-182:A>G	PCR	fwd.	TCTCTTCCCCCACCCTAA	63,8	61
	PCR	rev.	CCCCACTGCTGGGCTCTT	63,7	66,7
	Sequenz.	fwd.	CTCTCCACTCACGGC		
HCST_675:A>G	PCR	fwd.	CTCTGGGTCCCCACTTCAC	60,5	63,2
	PCR	rev.	GCCAGGAGTGGCAGAGAG	59,6	66,7
	Sequenz.	fwd.	ATGAGCCAACCCCCC		
HCST_866:C>T	PCR	fwd.	GCCAGGAGTGGCAGAGAG	59,6	66,7
	PCR	rev.	GCCAGGAGTGGCAGAGAG	59,6	66,7
	Sequenz.	fwd.	GTCTGGAACCCCCCT		

**Tabelle 36:** Phänotypen der genotypisierten Bullenpopulation wurden vom VIT zur Verfügung gestellt;  $M, s$ : nichtstratifizierter Mittelwert bzw. Standardabweichung des Phänotyps;  $s_r$ : Residuenstandardfehler (= Quadratwurzel der Residuenvarianz) des linearen Modells aus Gleichung 13.

<b>Englisch</b>	<b>Phänotyp Deutsch</b>	<b>M</b>	<b>s</b>	<b><math>s_r</math></b>
SCS 1 lakt	<i>Somatic Cell Score</i> , 1. Laktation	-0,1	0,39	0,37
SCS 2 lakt	<i>Somatic Cell Score</i> , 2. Laktation	-0,14	0,42	0,4
SCS 3 lakt	<i>Somatic Cell Score</i> , 3. Laktation	-0,14	0,46	0,43
udder support	Zentralband	136,6	285,42	264,81
udder depth	Eutertiefe	94,73	333,51	330,57
teat length	Strichlänge	-37,61	347,81	331,59
stature	Kreuzhöhe	32,67	100,6	96,02
rump width	Beckenbreite	-36,86	358,4	328,39
rump angle	Beckenneigung (Winkel zwischen Hüftknochen und	28,45	351,9	329,43
rear udder height	Hintereuterhöhe	97,9	291,4	278,6
rear teat placement	Strichstellung	155,73	506,28	501,36
rear leg set	Sprg. winklung	-16,51	289,28	282,87
rear leg rear view	Stellung hinten	-33,74	523,89	521,07
overall udder score	Gesamtnote Euter	27,63	71,58	68,63
overall feet leg		24,5	119,63	119,85
overall conformation	Gesamtnote Exterieur	9439,74	894,52	810,67
front teat placement	Strichplatzierung	45,37	346,57	318,44
fore udder	Voreuteransatz	20,82	315,98	302,38
foot angle	Fessel	26,11	245,67	239,03
chest width	Brustbreite	-72,22	298,77	291,82
body depth	Rumpftiefe	7,43	333,92	315,14
dmg	Durchschnittliches Minutengemelk	0,03	0,14	0,14

**Tabelle 37:** Die Verteilung der für die Assoziationsstudie an sechs SNPs typisierten DH-Bullen auf die sechs Halbgeschwisterfamilien und paternale Verwandtschaftsverhältnisse. Die Zuordnung der Identifikationsbuchstaben der Bullen ist arbiträr und hat über diese Tabelle hinaus keine Bedeutung.

Vater	typisiert	Generation			
		-1	-2	-3	-4
A	66	♂			
B	91	♂	♂	♂	♂
C	113	♂			
D	260				
E	59	♂	♂	♂	♂
F	346	♂	♂		



**Tabelle 38:** Übersicht über die Ergebnisse der Assoziation von sechs SNPs mit dem SCS in den ersten drei Laktationsperioden sowie mehreren Konformations- und Milchleistungsphänotypen. Gezeigt werden die Irrtumswahrscheinlichkeiten  $p$  für den fixen Effekt der SNPs im Einzel-SNP-Modell (Modell 10 auf S. 42) auf die gewichtete mittlere Töchterabweichung (DYD, *daughter yield deviation*) des Phänotyps. Statistisch signifikante Phänotyp-SNP-Kombinationen ( $p < 0,05$ , Ausnahme: CD22.:7568:C>T-Strichplatzierung mit  $p = 0,056$ ) sind farblich hervorgehoben.


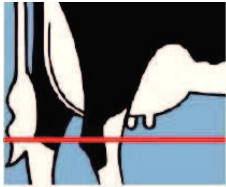





\*: Der Denominator-Freiheitsgrad (*denDF*), der für die exakte Berechnung des p-Wertes der verwendeten Wald-Statistik nötig ist, konnte von der Software nicht einwandfrei ermittelt werden. In diesem Fall wurde *denDF* geschätzt; dies erlaubt eine hinreichend genaue Ermittlung des p-Wertes.

Merkmal	CD22_		CD79A_		HCST_	
	7568:C>T	8067:A>G	2808:C>T	-182:A>G	675:A>G	866:C>T
SCS	0,8920	0,5790	0,0010	0,0040	0,3930	0,8958*
1.Laktation						
SCS	0,9730	0,5110	0,0010	0,0039*	0,4740	0,8120
2.Laktation						
SCS	0,7930	0,3940	0,0010	0,0070	0,2120	0,4170
3.Laktation						
Gesamtnote	0,4050	0,1880	0,0010	0,0030	0,3460	0,7710
Euter						
Strichplatzierung	0,0560	0,0010	0,0220	0,0910	0,5780	0,3990
Gesamtnote	0,8700	0,1350	0,0060	0,0090	0,1380	0,3940
Exterieur						
Sprg. winklung	0,6630	0,7510	0,7330	0,6220	0,6030	0,8160
Hintereuterhöhe	0,9770	0,5180	0,1900	0,1760	0,6410	0,3610
Gesamtnote	0,9240	0,6810	0,0190	0,0170	0,5050	0,5940
Form/ Fundament						
Zentralband	0,5350	0,1370	0,0320	0,0240	0,7280	0,2100
Strichlänge	0,6710	0,2670	0,9570	0,7420	0,0200	0,2180
Rumpftiefe	0,9790	0,3280	0,8340	0,5610	0,0880	0,3680
Voreuteransatz	0,3750	0,2460	0,0010	0,0100	0,0900	0,6870
Eutertiefe	0,2890	0,0270	0,0020	0,0030	0,3180	0,8580
Kreuzhöhe	0,2870	0,3480	0,6330	0,2000	0,1520	0,1670
Beckenneigung	0,2950	0,0410	0,0040	0,0090	0,6640	0,6370
Beckenbreite	0,4560	0,2920	0,6260	0,6170	0,9170	0,9910
Stellung hinten	0,2110	0,3480	0,1710	0,3610	0,3580	0,0400
Strichstellung	0,3070	0,6440	0,7670	0,6350	0,0280	0,1500
Fessel	0,3190	0,3460	0,1630	0,1190	0,7830	0,6480
Brustbreite	0,4620	0,7260	0,5430	0,2410	0,4300	0,8440
Durchschn.	0,4950	0,7180	0,1660	0,2640	0,0200	0,4540
Minutengemelk						

**Tabelle 39:** Ein  $\gamma$ -IR-Element befindet sich im Promotor des bovinen *CD79A*-Genes und wird durch den in dieser Arbeit untersuchten SNP *CD79A*.-182:A>G beeinflusst. Die Tabelle gibt einige Arbeiten, in denen solche Element, mit der Konsensussequenz CWKKANNY ( $=C_{AGG}^{TTT}ANN_C^T$ ) in Promotoren beschrieben und teilweise funktionell untersucht wurden. Bei der durch Jeyaseelan *et al.* (2000) untersuchten Speikobra handelt es sich um *Naja sputatrix*.

Gen	Produkt	Spezies	Referenz
<i>SLPI</i>	<i>secretory leukoprotease inhibitor</i>	Maus	Kikuchi <i>et al.</i> (1998)
<i>LIF</i>	<i>leukaemia inhibitory factor</i>	Rind	Piedrahita <i>et al.</i> (1997)
<i>NRAMP1</i>	<i>natural resistance associated macrophage protein</i>	Hund	Altet <i>et al.</i> (2002)
<i>DPA</i>	MHC-II-Komplex	Mensch	Yang <i>et al.</i> (1990)
<i>NOS2</i>	induzierbare Nitritoxidase (iNOS)	Ratte	Sanchez <i>et al.</i> (2003)
<i>PLA<sub>2</sub></i>	Phospholipase A <sub>2</sub>	Speikobra	Jeyaseelan <i>et al.</i> (2000)
<i>PLA<sub>2</sub></i>	Phospholipase A <sub>2</sub>	Mensch	Wu <i>et al.</i> (1994)
<i>PCI</i>	Protein C Inhibitor	Mensch	Hayashi <i>et al.</i> (1998)
<i>MT-I</i>	Metallothionein I	Maus	Ghoshal <i>et al.</i> (2001)
<i>S100A8</i>	Untereinheit von Calprotectin	Mensch	Ren <i>et al.</i> (2009)

**Tabelle 40:** Verschiedene Parameter der Beschaffenheit von Euter und Fundament werden für die Ermittlung des Zuchtwertes eines Bullen an dessen Töchtern beurteilt. Für die hier gezeigten Parameter wurden vom VIT Verden Phänotypdaten (DYD) und Gewichtungen (EDC) für die in der Assoziationsstudie untersuchte DH-Halbgeschwisterpopulation zur Verfügung gestellt; vgl. Tab. 36. Die Abbildungen und Angaben sind einer Broschüre des VIT Verden, nach Informationen des DHV, entnommen (VIT, 2011).

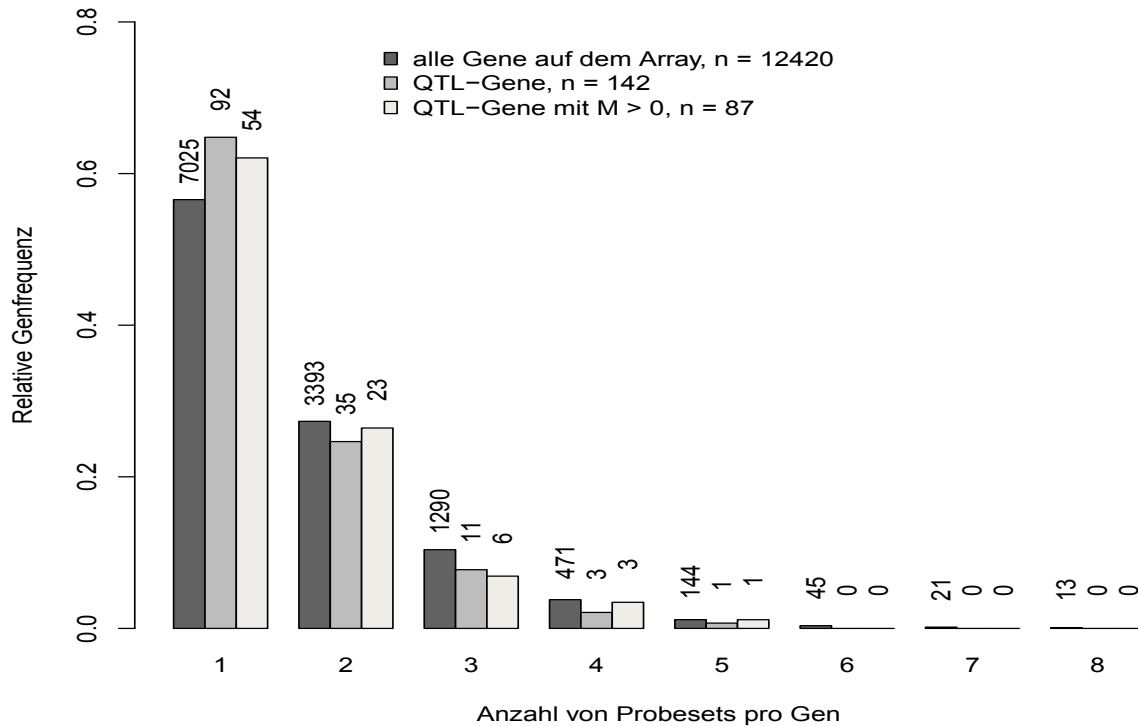
Merkmals	Beschreibung	Abbildung	positiv	negativ
Zentralband	Länge des Euterspaltes nach oben, sowie seine Ausprägung		stark	schwach
Eutertiefe	Abstand zwischen der gedachten waagerechten Linie durch die Hacke und den Euterboden		hoch	tief
Strichlänge	Länge der Vorderstriche		mittel	sehr kurz, sehr lang
Kreuzhöhe	Körperhöhe in der Mitte des Kreuzbeines			< 142 cm, > 152 cm
Beckenbreite	Abstand der Mittelpunkte der Sitzbeinhöcker		breit	schmal
Beckenneigung	Neigung der gedachten Linien zwischen Sitzbein und Hüftbeinen		leicht geneigt	stark ansteigend, abfallend
Hinter-euterhöhe	Abstand zwischen unterem Scheidenrand und Beginn des Euterdrüsengewebes		hoch	niedrig

Fortsetzung auf der nächsten Seite

Tabelle 40, fortgesetzt

Merkmal	Beschreibung	Abbildung	positiv	negativ
Strichstellung	Ansatz der Striche unter den Eutervierteln		vorne: mittig bis innen; hinten: mittig	vorne: außen; hinten: außen, sehr weit innen
Sprunggelenkwinkel	Winkel des Hinterbeines in Höhe des Sprunggelenkes		mittel	stark gewinkelt, steil
Hinterbeinstellung	Stellung der hinteren Klauen		leicht nach außen bis parallel	sehr nach außen
Voreuteransatz	Winkel des Euters mit der Bauchdecke		fest	schwach
Fessel ( <i>foot angle</i> )	Winkel des vorderen Teils der Klaue mit dem Boden		hoch	flach
Brustbreite	Abstand zwischen den Buggelenken		stark	schwach, sehr stark
Körpertiefe	Flankentiefe in Höhe der letzten Rippe		tief	aufgezogen, sehr tief

## Abbildungen



**Abbildung 20:** Die Frequenzverteilung von Genen, denen gemäß Hintermair (2007) mehrere Probesets auf dem bovinen *GeneChip*<sup>®</sup> zugeordnet werden, in Abhängigkeit von der Anzahl der Probesets pro Gen. Von 254 positionellen Kandidatengenens des BTA18-SCS-QTL wurden 142 Gene mit insgesamt 212 Probesets, von denen maximal fünf dasselbe Gen repräsentieren, auf dem Array gefunden (mittlere Balken).  $\approx 60\%$  der Gene sind durch ein singuläres Probeset vertreten. Die relativen Frequenzen gleichen den für den gesamten *GeneChip*<sup>®</sup> ermittelten (linke Balken); hier treten auch Gene auf, die von mehr als 8 Probesets (Maximum: 51) repräsentiert werden, diese sind der Übersichtlichkeit halber nicht dargestellt. In der Untergruppe von 87 QTL-Genen, für die ein DbR-Koeffizient  $M > 0$  durch den datenbankbasierten Ansatz errechnet wurde (rechte Balken), zeigt sich eine sehr ähnliche Frequenzverteilung. Über den Balken für die relativen Frequenzen sind die absoluten Genfrequenzen als Zahlenwerte dargestellt.

```

PMID- 19359050
OWN - NLM
STAT- MEDLINE
DA - 20090504
DCOM- 20090707
LR - 20111117
IS - 0165-6147 (Print)
IS - 0165-6147 (Linking)
VI - 30
IP - 5
DP - 2009 May
TI - Siglecs as targets for therapy in immune-cell-mediated disease.
PG - 240-8
AB - The sialic-acid-binding immunoglobulin-like lectins (siglecs) comprise a family
...
AD - Department of Chemical Physiology, The Scripps Research Institute, La Jolla CA
92037, USA.
FAU - O'Reilly, Mary K
AU - O'Reilly MK
FAU - Paulson, James C
AU - Paulson JC
LA - eng
GR - AI050143/AI/NIAID NIH HHS/United States
GR - GM060938/GM/NIGMS NIH HHS/United States
GR - R01 AI050143-07/AI/NIAID NIH HHS/United States
GR - R01 GM060938-10/GM/NIGMS NIH HHS/United States
PT - Journal Article
PT - Research Support, N.I.H., Extramural
PT - Research Support, Non-U.S. Gov't
PT - Review
DEP - 20090407
PL - England
TA - Trends Pharmacol Sci
JT - Trends in pharmacological sciences
JID - 7906158
RN - 0 (Aminoglycosides)
RN - 0 (Antibodies)
RN - 0 (Antibodies, Monoclonal)
RN - 0 (Antibodies, Monoclonal, Humanized)
RN - 0 (Antigens, CD)
RN - 0 (Antigens, CD22)
RN - 0 (Antigens, Differentiation, Myelomonocytic)
RN - 0 (CD33 antigen)
RN - 0 (Lectins)
RN - 0 (Polysaccharides)
RN - 0 (gemtuzumab)
RN - 0 (sialic acid binding Ig-like lectin)
SB - IM
MH - Aminoglycosides/therapeutic use
MH - Animals
MH - Antibodies/therapeutic use
MH - Antibodies, Monoclonal/therapeutic use
MH - Antibodies, Monoclonal, Humanized
MH - Antigens, CD/immunology
MH - Antigens, CD22/immunology
MH - Antigens, Differentiation, Myelomonocytic/immunology
MH - Clinical Trials as Topic
MH - Drug Delivery Systems/*methods
MH - Humans
MH - Immune System Diseases/*drug therapy
MH - *Lectins/chemistry/metabolism/therapeutic use
MH - Leukemia/drug therapy
MH - Lymphocyte Depletion
MH - Lymphoma, B-Cell/drug therapy
MH - Models, Biological
MH - Polysaccharides/metabolism/therapeutic use
RF - 104
PMC - PMC2830709
MID - NIHMS177195
OID - NLM: NIHMS177195
OID - NLM: PMC2830709
EDAT- 2009/04/11 09:00
MHDA- 2009/07/08 09:00
CRDT- 2009/04/11 09:00
PHST- 2009/01/13 [received]
PHST- 2009/02/18 [revised]
PHST- 2009/02/19 [accepted]
PHST- 2009/04/07 [aheadofprint]
AID - S0165-6147(09)00052-2 [pii]
AID - 10.1016/j.tips.2009.02.005 [doi]
PST - ppublsh
SO - Trends Pharmacol Sci. 2009 May;30(5):240-8. Epub 2009 Apr 7.

```

**Abbildung 21:** Beispiel für das MEDLINE-Textformat von in der PubMed indizierten Publikationen. Wiedergegeben ist der komplette Eintrag eines mit MeSH-Termen annotierten Artikels; jedoch wurde der *Abstract* aus Platzersparnisgründen gekürzt. Beachte das Label, welches jede Zeile einleitet, den Eintragstyp dieser Zeile definiert und dieses Format für die maschinelle Auswertung prädestiniert. MeSH-Terme werden durch das Label MH eingeleitet.

## Teil VII

## Literaturverzeichnis

## Literatur

- Abdallah, J. M., Goffinet, B., Cierco-Ayrolles, C. & Perez-Enciso, M. (2003). Linkage disequilibrium fine mapping of quantitative trait loci: a simulation study. *Genet Sel Evol*, **35** (5), 513–532.
- Abecasis, G. R., Cardon, L. R. & Cookson, W. O. (2000). A general test of association for quantitative traits in nuclear families. *American Journal of Human Genetics*, **66** (1), 279–292.
- Alain, K., Karrow, N. A., Thibault, C., St-Pierre, J., Lessard, M. & Bissonnette, N. (2009). Osteopontin: an early innate immune marker of *Escherichia coli* mastitis harbors genetic polymorphisms with possible links with resistance to mastitis. *BMC Genomics*, **10**, 444.
- Allen, A. R., Taylor, M., McKeown, B., Curry, A. I., Lavery, J. F., Mitchell, A., Hartshorne, D., Fries, R. & Skuce, R. A. (2010). Compilation of a panel of informative single nucleotide polymorphisms for bovine identification in the Northern Irish cattle population. *BMC Genetics*, **11**, 5.
- Altet, L., Francino, O., Solano-Gallego, L., Renier, C. & Sanchez, A. (2002). Mapping and sequencing of the canine NRAMP1 gene and identification of mutations in leishmaniasis-susceptible dogs. *Infection and Immunity*, **70** (6), 2763–2771.
- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, **25** (17), 3389–3402.
- Andronicos, N., Hunt, P. & Windon, R. (2010). Expression of genes in gastrointestinal and lymphatic tissues during parasite infection in sheep genetically resistant or susceptible to *Trichostrongylus colubriformis* and *Haemonchus contortus*. *International Journal for Parasitology*, **40** (4), 417–429.
- Araujo, R. N., Padilha, T., Zarlenga, D., Sonstegard, T., Connor, E. E., van Tassel, C., Lima, W. S., Nascimento, E. & Gasbarre, L. C. (2009). Use of a candidate gene array to delineate gene expression patterns in cattle selected for resistance or susceptibility to intestinal nematodes. *Veterinary Parasitology*, **162** (1-2), 106–115.
- Arcade, A., Labourdette, A., Falque, M., Mangin, B., Chardon, F., Charcosset, A. & Joets, J. (2004). BioMercator: integrating genetic maps and QTL towards discovery of candidate genes. *Bioinformatics*, **20** (14), 2324–2346.
- Arias, J. A., Keehan, M., Fisher, P., Coppieters, W. & Spelman, R. (2009). A high density linkage map of the bovine genome. *BMC Genetics*, **10**, 18.
- Arnould, V. M., Soyeyurt, H., Gengler, N., Colinet, F. G., Georges, M. V., Bertozzi, C., Portetelle, D. & Renaville, R. (2009). Genetic analysis of lactoferrin content in bovine milk. *Journal of Dairy Science*, **92** (5), 2151–2158.
- Asai, K., Komine, Y., Kozutsumi, T., Yamaguchi, T., Komine, K. & Kumagai, K. (2000). Predominant subpopulations of T lymphocytes in the mammary gland secretions during lactation and intraepithelial T lymphocytes in the intestine of dairy cows. *Veterinary Immunology and Immunopathology*, **73** (3-4), 233–240.
- Ashwell, M. S., Da, Y., van Tassell, C. P., van Raden, P. M., Miller, R. H. & Rexroad, Jr,

- C. E. (1998). Detection of putative loci affecting milk production and composition, health, and type traits in a United States Holstein population. *Journal of Dairy Science*, **81** (12), 3309–3314.
- Ashwell, M. S., Heyen, D. W., Sonstegard, T. S., van Tassell, C. P., Da, Y., van Raden, P. M., Ron, M., Weller, J. I. & Lewin, H. A. (2004). Detection of quantitative trait loci affecting milk production, health, and reproductive traits in Holstein cattle. *Journal of Dairy Science*, **87** (2), 468–475.
- Ashwell, M. S., Rexroad, C. E., Miller, R. H., van Raden, P. M. & Da, Y. (1997). Detection of loci affecting milk production and health traits in an elite us holstein population using microsatellite markers. *Animal Genetics*, **28** (3), 216–222.
- Babayeva, N. D., Wilder, P. J., Shiina, M., Mino, K., Desler, M., Ogata, K., Rizzino, A. & Tahirov, T. H. (2010). Structural basis of ets1 cooperative binding to palindromic sequences on stromelysin-1 promoter dna. *Cell Cycle*, **9** (15), 3054–3062.
- Baes, C., Brand, B., Mayer, M., Kühn, C., Liu, Z., Reinhardt, F. & Reinsch, N. (2009). Refined positioning of a quantitative trait locus affecting somatic cell score on chromosome 18 in the German Holstein using linkage disequilibrium. *Journal of Dairy Science*, **92** (8), 4046–4054.
- Baillat, D., Begue, A., Stehelin, D. & Aumercier, M. (2002). Ets-1 transcription factor binds cooperatively to the palindromic head to head ets-binding sites of the stromelysin-1 promoter by counteracting autoinhibition. *Journal of Biological Chemistry*, **277** (33), 29386–29398.
- Bajdik, C. D., Kuo, B., Rusaw, S., Jones, S. & Brooks-Wilson, A. (2005). CGMIM: automated text-mining of Online Mendelian Inheritance in Man (OMIM) to identify genetically-associated cancers and candidate genes. *BMC Bioinformatics*, **6**, 78.
- Band, M. R., Larson, J. H., Rebeiz, M., Green, C. A., Heyen, D. W., Donovan, J., Windish, R., Steining, C., Mahyuddin, P., Womack, J. E. & Lewin, H. A. (2000). An ordered comparative map of the cattle and human genomes. *Genome Research*, **10** (9), 1359–1368.
- Bannerman, D. D., Paape, M. J., Lee, J. W., Zhao, X., Hope, J. C. & Rainard, P. (2004). Escherichia coli and Staphylococcus aureus elicit differential innate immune responses following intramammary infection. *Clinical and Diagnostic Laboratory Immunology*, **11** (3), 463–472.
- Bannish, G., Fuentes-Panana, E. M., Cambier, J. C., Pear, W. S. & Monroe, J. G. (2001). Ligand-independent signaling functions for the B lymphocyte antigen receptor and their role in positive selection during B lymphopoiesis. *Journal of Experimental Medicine*, **194** (11), 1583–1596.
- Barrett, J. C., Fry, B., Maller, J. & Daly, M. J. (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, **21** (2), 263–265.
- Bayani, J., Marrano, P., Graham, C., Zheng, Y., Li, L., Katsaros, D., Lassus, H., Butzow, R., Squire, J. A. & Diamandis, E. P. (2011). Genomic instability and copy-number heterogeneity of chromosome 19q, including the kallikrein locus, in ovarian carcinomas. *Mol Oncol*, **5** (1), 48–60.
- Beaudeau, F., Fourichon, C., Seegers, H. & Bareille, N. (2002). Risk of clinical mastitis in dairy herds with a high proportion of low individual milk somatic-cell counts. *Preventive Veterinary Medicine*, **53** (1-2), 43–54.
- Beilharz, R. G., Luxford, B. G. & Wilkinson, J. L. (1993). Quantitative genetics and evolution - is our understanding of genetics sufficient to explain evolution. *Journal of Animal*



- Breeding and Genetics*, **110** (3), 161 – 170.
- Bennewitz, J., Reinsch, N., Grohs, C., Leveziel, H., Malafosse, A., Thomsen, H., Xu, N., Looft, C., Kühn, C., Brockmann, G. A., Schwerin, M., Weimann, C., Hiendleder, S., Erhardt, G., Medjugorac, I., Russ, I., Forster, M., Brenig, B., Reinhardt, F., Reents, R., Averdunk, G., Blumel, J., Boichard, D. & Kalm, E. (2003). Combined analysis of data from two granddaughter designs: A simple strategy for QTL confirmation and increasing experimental power in dairy cattle. *Genet Sel Evol*, **35** (3), 319–338.
- Bentwich, I., Avniel, A., Karov, Y., Aharonov, R., Gilad, S., Barad, O., Barzilai, A., Einat, P., Einav, U., Meiri, E., Sharon, E., Spector, Y. & Bentwich, Z. (2005). Identification of hundreds of conserved and nonconserved human microRNAs. *Nature Genetics*, **37** (7), 766–770.
- Bink, M. C. & van Eeuwijk, F. A. (2009). A Bayesian QTL linkage analysis of the common dataset from the 12th QTLMAS workshop. *BMC Proceedings*, **3 Suppl 1**, S4.
- Blott, S., Kim, J. J., Moisisio, S., Schmidt-Kuntzel, A., Cornet, A., Berzi, P., Cambisano, N., Ford, C., Grisart, B., Johnson, D., Karim, L., Simon, P., Snell, R., Spelman, R., Wong, J., Vilkki, J., Georges, M., Farnir, F. & Coppieters, W. (2003). Molecular dissection of a quantitative trait locus: a phenylalanine-to-tyrosine substitution in the transmembrane domain of the bovine growth hormone receptor is associated with a major effect on milk yield and composition. *Genetics*, **163** (1), 253–266.
- Boichard, D., Grohs, C., Bourgeois, F., Cerqueira, F., Faugeras, R., Neau, A., Rupp, R., Amigues, Y., Boscher, M. Y. & Leveziel, H. (2003). Detection of genes influencing economic traits in three French dairy cattle breeds. *Genet Sel Evol*, **35** (1), 77–101.
- Bowers, S. R., Calero-Nieto, F. J., Valeaux, S., Fernandez-Fuentes, N. & Cockerill, P. N. (2010). Runx1 binds as a dimeric complex to overlapping runx1 sites within a palindromic element in the human gm-csf enhancer. *Nucleic Acids Research*, **38** (18), 6124–6134.
- Boysen, P. & Storset, A. K. (2009). Bovine natural killer cells. *Veterinary Immunology and Immunopathology*, **130** (3-4), 163–177.
- Brand, B., Baes, C., Mayer, M., Reinsch, N. & Kühn, C. (2009). Identification of a two-marker-haplotype on *Bos taurus* autosome 18 associated with somatic cell score in German Holstein cattle. *BMC Genetics*, **10**, 50.
- Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C., Aach, J., Ansorge, W., Ball, C. A., Causton, H. C., Gaasterland, T., Glenisson, P., Holstege, F. C., Kim, I. F., Markowitz, V., Matese, J. C., Parkinson, H., Robinson, A., Sarkans, U., Schulze-Kremer, S., Stewart, J., Taylor, R., Vilo, J. & Vingron, M. (2001). Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nature Genetics*, **29** (4), 365–371.
- Brink, M. (2003). *A contribution to the detailed mapping of QTL regions for udder health in cattle*. PhD thesis,.
- Bruckmaier, R. M. (2005). Gene expression of factors related to the immune reaction in response to intramammary *Escherichia coli* lipopolysaccharide challenge. *Journal of Dairy Research*, **72 Spec No**, 120–124.
- Brunner, R. M., Sanftleben, H., Goldammer, T., Kühn, C., Weikard, R., Kata, S. R., Womack, J. E. & Schwerin, M. (2003). The telomeric region of BTA18 containing a potential QTL region for health in cattle exhibits high similarity to the HSA19q region in humans. *Genomics*, **81** (3), 270–278.
- Buret, A. G. (2010). Immuno-modulation and anti-inflammatory benefits of antibiotics: the example of tilmicosin. *Canadian Journal of Veterinary Research*, **74** (1), 1–10.

- Burgess, D. L., Gefrides, L. A., Foreman, P. J. & Noebels, J. L. (2001). A cluster of three novel Ca<sup>2+</sup> channel gamma subunit genes on chromosome 19q13.4: evolution and expression profile of the gamma subunit gene family. *Genomics*, **71** (3), 339–350.
- Burton, J. L., Burnside, E. B., Kennedy, B. W., Wilkie, B. N. & Burton, J. H. (1989). Antibody responses to human erythrocytes and ovalbumin as marker traits of disease resistance in dairy calves. *Journal of Dairy Science*, **72** (5), 1252–1265.
- Burton, J. L., Madsen, S. A., Yao, J., Sipkovsky, S. S. & Coussens, P. M. (2001). An immunogenomics approach to understanding periparturient immunosuppression and mastitis susceptibility in dairy cows. *Acta Veterinaria Scandinavica*, **42** (3), 407–424.
- Cambier, J. C. (1995). Antigen and Fc receptor signaling. The awesome power of the immunoreceptor tyrosine-based activation motif (ITAM). *Journal of Immunology*, **155** (7), 3281–3285.
- Capuco, A. V., Bright, S. A., Pankey, J. W., Wood, D. L., Miller, R. H. & Bitman, J. (1992). Increased susceptibility to intramammary infection following removal of teat canal keratin. *Journal of Dairy Science*, **75** (8), 2126–2130.
- Cardoso, F. F., Rosa, G. J., Steibel, J. P., Ernst, C. W., Bates, R. O. & Tempelman, R. J. (2008). Selective transcriptional profiling and data analysis strategies for expression quantitative trait loci mapping in outbred F2 populations. *Genetics*, **180** (3), 1679–1690.
- Carlen, E., Strandberg, E. & Roth, A. (2004). Genetic parameters for clinical mastitis, somatic cell score, and production in the first three lactations of Swedish holstein cows. *Journal of Dairy Science*, **87** (9), 3062–3070.
- Carragher, D. M., Rangel-Moreno, J. & Randall, T. D. (2008). Ectopic lymphoid tissues and local immunity. *Seminars in Immunology*, **20** (1), 26–42.
- Chen, J., Xu, H., Aronow, B. J. & Jegga, A. G. (2007). Improved human disease candidate gene prioritization using mouse phenotype. *BMC Bioinformatics*, **8**, 392.
- Chen, L., Liu, H. & Friedman, C. (2005). Gene name ambiguity of eukaryotic nomenclatures. *Bioinformatics*, **21** (2), 248–256.
- Clements, J. A., Willemsen, N. M., Myers, S. A. & Dong, Y. (2004). The tissue kallikrein family of serine proteases: functional roles in human disease and potential as clinical biomarkers. *Critical Reviews in Clinical Laboratory Sciences*, **41** (3), 265–312.
- Cleveland, M. A. & Deeb, N. (2009). Evaluation of a genome-wide approach to multiple marker association considering different marker densities. *BMC Proceedings*, **3 Suppl 1**, S5.
- Cohen-Zinder, M., Seroussi, E., Larkin, D. M., Loor, J. J., Everts-van der Wind, A., Lee, J. H., Drackley, J. K., Band, M. R., Hernandez, A. G., Shani, M., Lewin, H. A., Weller, J. I. & Ron, M. (2005). Identification of a missense mutation in the bovine ABCG2 gene with a major effect on the QTL on chromosome 6 affecting milk yield and composition in Holstein cattle. *Genome Research*, **15** (7), 936–944.
- Cole, J. B., van Raden, P. M., O'Connell, J. R., van Tassell, C. P., Sonstegard, T. S., Schnabel, R. D., Taylor, J. F. & Wiggans, G. R. (2009). Distribution and location of genetic effects for dairy traits. *Journal of Dairy Science*, **92** (6), 2931–2946.
- Cole, J. B., Wiggans, G. R., Ma, L., Sonstegard, T. S., Lawlor, Jr, T. J., Crooker, B. A., van Tassell, C. P., Yang, J., Wang, S., Matukumalli, L. K. & Da, Y. (2011). Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary u.s. holstein cows. *BMC Genomics*, **12**, 408.
- Coletti, M. H. & Bleich, H. L. (2001). Medical subject headings used to search the biomedical literature. *Journal of the American Medical Informatics Association*, **8** (4), 317–323.

- Collins, R. A., Parsons, K. R. & Bland, A. P. (1986). Antibody-containing cells and specialised epithelial cells in the bovine teat. *Research in Veterinary Science*, **41** (1), 50–55.
- Concha, C., Holmberg, O. & Morein, B. (1978). Proportion of B- and T-lymphocytes in normal bovine milk. *Journal of Dairy Research*, **45** (2), 287–290.
- Connor, E. E., Ashwell, M. S., Schnabel, R. & Williams, J. L. (2006). Comparative mapping of bovine chromosome 27 with human chromosome 8 near a dairy form QTL in cattle. *Cytogenetic and Genome Research*, **112** (1-2), 98–102.
- Connor, E. E., Meyer, M. J., Li, R. W., van Amburgh, M. E., Boisclair, Y. R. & Capuco, A. V. (2007). Regulation of gene expression in the bovine mammary gland by ovarian steroids. *Journal of Dairy Science*, **90 Suppl 1**, E55–65.
- Corl, C. M., Gandy, J. C. & Sordillo, L. M. (2008). Platelet activating factor production and proinflammatory gene expression in endotoxin-challenged bovine mammary endothelial cells. *Journal of Dairy Science*, **91** (8), 3067–3078.
- Costa, E. O., Ribeiro, A. R., Watanabe, E. T. & Melville, P. A. (1998). Infectious bovine mastitis caused by environmental organisms. *Zentralblatt für Veterinärmedizin. Reihe B*, **45** (2), 65–71.
- Cox, D. R., Burmeister, M., Price, E. R., Kim, S. & Myers, R. M. (1990). Radiation hybrid mapping: a somatic cell genetic method for constructing high-resolution maps of mammalian chromosomes. *Science*, **250** (4978), 245–250.
- Crocker, P. R., Paulson, J. C. & Varki, A. (2007). Siglecs and their roles in the immune system. *Nature Reviews Immunology*, **7** (4), 255–266.
- Crooks, L., Sahana, G., de Koning, D. J., Lund, M. S. & Carlborg, O. (2009). Comparison of analyses of the QTLMAS XII common dataset. II: genome-wide association and fine mapping. *BMC Proceedings*, **3 Suppl 1**, S2.
- Daetwyler, H. D., Schenkel, F. S., Sargolzaei, M. & Robinson, J. A. (2008). A genome scan to detect quantitative trait loci for economically important traits in Holstein cattle using two methods and a dense single nucleotide polymorphism map. *Journal of Dairy Science*, **91** (8), 3225–3236.
- Daigle, Jr, B. J. & Altman, R. B. (2008). M-BISON: microarray-based integration of data sources using networks. *BMC Bioinformatics*, **9**, 214.
- Darvasi, A. & Soller, M. (1994). Selective DNA pooling for determination of linkage between a molecular marker and a quantitative trait locus. *Genetics*, **138** (4), 1365–1373.
- Darvasi, A., Weinreb, A., Minke, V., Weller, J. I. & Soller, M. (1993). Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. *Genetics*, **134** (3), 943–951.
- de Bakker, P. I. W., Yelensky, R., Pe'er, I., Gabriel, S. B., Daly, M. J. & Altshuler, D. (2005). Efficiency and power in genetic association studies. *Nature Genetics*, **37** (11), 1217–1223.
- de Franco, A. L. (1997). The complexity of signaling pathways activated by the BCR. *Current Opinion in Immunology*, **9** (3), 296–308.
- de Koning, D. J., Carlborg, O. & Haley, C. S. (2005). The genetic dissection of immune response using gene-expression studies and genome mapping. *Veterinary Immunology and Immunopathology*, **105** (3-4), 343–352.
- de Risi, J., Penland, L., Brown, P. O., Bittner, M. L., Meltzer, P. S., Ray, M., Chen, Y., Su, Y. A. & Trent, J. M. (1996). Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nature Genetics*, **14** (4), 457–460.
- Diez-Tascon, C., Keane, O. M., Wilson, T., Zadissa, A., Hyndman, D. L., Baird, D. B.,

- McEwan, J. C. & Crawford, A. M. (2005). Microarray analysis of selection lines from outbred populations to identify genes involved with nematode parasite resistance in sheep. *Physiological Genomics*, **21** (1), 59–69.
- Diggle, M. A. & Clarke, S. C. (2004). Pyrosequencing: sequence typing at the speed of light. *Molecular Biotechnology*, **28** (2), 129–137.
- Djebbari, A., Karamycheva, S., Howe, E. & Quackenbush, J. (2005). MeSHer: identifying biological concepts in microarray assays based on PubMed references and MeSH terms. *Bioinformatics*, **21** (15), 3324–3326.
- Dodd, F. H. (1983). Mastitis—progress on control. *Journal of Dairy Science*, **66** (8), 1773–1780.
- Donaldson, L., Vuocolo, T., Gray, C., Strandberg, Y., Reverter, A., McWilliam, S., Wang, Y., Byrne, K. & Tellam, R. (2005). Construction and validation of a Bovine Innate Immune Microarray. *BMC Genomics*, **6**, 135.
- Duhamel, G. E., Bernoco, D., Davis, W. C. & Osburn, B. I. (1987). Distribution of T and B lymphocytes in mammary dry secretions, colostrum and blood of adult dairy cattle. *Veterinary Immunology and Immunopathology*, **14** (2), 101–122.
- Durink, S., Moreau, Y., Kasprzyk, A., Davis, S., de Moor, B., Brazma, A. & Huber, W. (2005). BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics*, **21** (16), 3439–3440.
- Düvel, A. (2010). *Untersuchungen zur Prostaglandin E2- beeinflussten Differenzierung boviner Monozyten in Makrophagen*. PhD thesis,.
- DVG (2002). Leitlinien zur Bekämpfung der Mastitis des Rindes als Bestandsproblem. ISBN: 3-936815-53-4.
- Dylke, J., Lopes, J., Dang-Lawson, M., Machtaler, S. & Matsuuchi, L. (2007). Role of the extracellular and transmembrane domain of Ig-alpha/beta in assembly of the B cell antigen receptor (BCR). *Immunology Letters*, **112** (1), 47–57.
- Elsik, C. G., Tellam, R. L., Worley, K. C., Gibbs, R. A., Muzny, D. M., Weinstock, G. M., Adelson, D. L., Eichler, E. E., Elnitski, L., Guigo, R., Hamernik, D. L., Kappes, S. M., Lewin, H. A., Lynn, D. J., Nicholas, F. W., Raymond, A., Rijnkels, M., Skow, L. C., Zdobnov, E. M., Schook, L., Womack, J., Alioto, T., Antonarakis, S. E., Astashyn, A., Chapple, C. E., Chen, H. C., Chrast, J., Camara, F., Ermolaeva, O., Henrichsen, C. N., Hlavina, W., Kapustin, Y., Kiryutin, B., Kitts, P., Kokocinski, F., Landrum, M., Maglott, D., Pruitt, K., Sapojnikov, V., Searle, S. M., Solovyev, V., Souvorov, A., Ucla, C., Wyss, C., Anzola, J. M., Gerlach, D., Elhaik, E., Graur, D., Reese, J. T., Edgar, R. C., McEwan, J. C., Payne, G. M., Raison, J. M., Junier, T., Kriventseva, E. V., Eyraes, E., Plass, M., Donthu, R., Larkin, D. M., Reecy, J., Yang, M. Q., Chen, L., Cheng, Z., Chitko-McKown, C. G., Liu, G. E., Matukumalli, L. K., Song, J., Zhu, B., Bradley, D. G., Brinkman, F. S., Lau, L. P., Whiteside, M. D., Walker, A., Wheeler, T. T., Casey, T., German, J. B., Lemay, D. G., Maqbool, N. J., Molenaar, A. J., Seo, S., Stothard, P., Baldwin, C. L., Baxter, R., Brinkmeyer-Langford, C. L., Brown, W. C., Childers, C. P., Connelley, T., Ellis, S. A., Fritz, K., Glass, E. J., Herzig, C. T., Iivanainen, A., Lahmers, K. K., Bennett, A. K., Dickens, C. M., Gilbert, J. G., Hagen, D. E., Salih, H., Aerts, J., Caetano, A. R., Dalrymple, B., Garcia, J. F., Gill, C. A., Hiendleder, S. G., Memili, E., Spurlock, D., Williams, J. L., Alexander, L., Brownstein, M. J., Guan, L., Holt, R. A., Jones, S. J., Marra, M. A., Moore, R., Moore, S. S., Roberts, A., Taniguchi, M., Waterman, R. C., Chacko, J., Chandrabose, M. M., Cree, A., Dao, M. D., Dinh, H. H., Gabisi, R. A., Hines, S., Hume, J., Jhangiani, S. N., Joshi, V., Kovar, C. L., Lewis, L. R., Liu, Y. S., Lopez,

- J., Morgan, M. B., Nguyen, N. B., Okwuonu, G. O., Ruiz, S. J., Santibanez, J., Wright, R. A., Buhay, C., Ding, Y., Dugan-Rocha, S., Herdandez, J., Holder, M., Sabo, A., Egan, A., Goodell, J., Wilczek-Boney, K., Fowler, G. R., Hitchens, M. E., Lozado, R. J., Moen, C., Steffen, D., Warren, J. T., Zhang, J., Chiu, R., Schein, J. E., Durbin, K. J., Havlak, P., Jiang, H., Liu, Y., Qin, X., Ren, Y., Shen, Y., Song, H., Bell, S. N., Davis, C., Johnson, A. J., Lee, S., Nazareth, L. V., Patel, B. M., Pu, L. L., Vattathil, S., Williams, Jr, R. L., Curry, S., Hamilton, C., Sodergren, E., Wheeler, D. A., Barris, W., Bennett, G. L., Eggen, A., Green, R. D., Harhay, G. P., Hobbs, M., Jann, O., Keele, J. W., Kent, M. P., Lien, S., McKay, S. D., McWilliam, S., Ratnakumar, A., Schnabel, R. D., Smith, T., Snelling, W. M., Sonstegard, T. S., Stone, R. T., Sugimoto, Y., Takasuga, A., Taylor, J. F., van Tassell, C. P., Macneil, M. D., Abatepaulo, A. R., Abbey, C. A., Ahola, V., Almeida, I. G., Amadio, A. F., Anatriello, E., Bahadue, S. M., Biase, F. H., Boldt, C. R., Carroll, J. A., Carvalho, W. A., Cervelatti, E. P., Chacko, E., Chapin, J. E., Cheng, Y., Choi, J., Colley, A. J., de Campos, T. A., de Donato, M., Santos, I. K., de Oliveira, C. J., Deobald, H., Devinoy, E., Donohue, K. E., Dovic, P., Eberlein, A., Fitzsimmons, C. J., Franzin, A. M., Garcia, G. R., Genini, S., Gladney, C. J., Grant, J. R., Greaser, M. L., Green, J. A., Hadsell, D. L., Hakimov, H. A., Halgren, R., Harrow, J. L., Hart, E. A., Hastings, N., Hernandez, M., Hu, Z. L., Ingham, A., Iso-Touru, T., Jamis, C., Jensen, K., Kapetis, D., Kerr, T., Khalil, S. S., Khatib, H., Kolbehdari, D., Kumar, C. G., Kumar, D., Leach, R., Lee, J. C., Li, C., Logan, K. M., Malinverni, R., Marques, E., Martin, W. F., Martins, N. F., Maruyama, S. R., Mazza, R., McLean, K. L., Medrano, J. F., Moreno, B. T., More, D. D., Muntean, C. T., Nandakumar, H. P., Nogueira, M. F., Olsaker, I., Pant, S. D., Panzitta, F., Pastor, R. C., Poli, M. A., Poslusny, N., Rachagani, S., Ranganathan, S., Razpet, A., Riggs, P. K., Rincon, G., Rodriguez-Osorio, N., Rodriguez-Zas, S. L., Romero, N. E., Rosenwald, A., Sando, L., Schmutz, S. M., Shen, L., Sherman, L., Southey, B. R., Lutzow, Y. S., Sweedler, J. V., Tammen, I., Telugu, B. P., Urbanski, J. M., Utsunomiya, Y. T., Verschoor, C. P., Waardenberg, A. J., Wang, Z., Ward, R., Weikard, R., Welsh, Jr, T. H., White, S. N., Wilming, L. G., Wunderlich, K. R., Yang, J. & Zhao, F. Q. (2009). The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science*, **324** (5926), 522–528.
- Emanuelson, U., Danell, B. & Philipsson, J. (1988). Genetic-parameters for clinical mastitis, somatic-cell counts, and milk-production estimated by multiple-trait restricted maximum-likelihood. *Journal of Dairy Science*, **71** (2), 467–476.
- Everts-van der Wind, A., Kata, S. R., Band, M. R., Rebeiz, M., Larkin, D. M., Everts, R. E., Green, C. A., Liu, L., Natarajan, S., Goldammer, T., Lee, J. H., McKay, S., Womack, J. E. & Lewin, H. A. (2004). A 1463 gene cattle-human comparative map with anchor points defined by human genome sequence coordinates. *Genome Research*, **14** (7), 1424–1437.
- Everts-van der Wind, A., Larkin, D. M., Green, C. A., Elliott, J. S., Olmstead, C. A., Chiu, R., Schein, J. E., Marra, M. A., Womack, J. E. & Lewin, H. A. (2005). A high-resolution whole-genome cattle-human comparative map reveals details of mammalian chromosome evolution. *Proceedings of the National Academy of Sciences of the United States of America*, **102** (51), 18526–18531.
- Ewens, W. J., Li, M. & Spielman, R. S. (2008). A review of family-based tests for linkage disequilibrium between a quantitative trait and a genetic marker. *PLoS Genetics*, **4** (9), e1000180.
- Fan, X., Lobenhofer, E. K., Chen, M., Shi, W., Huang, J., Luo, J., Zhang, J., Walker, S. J., Chu, T.-M., Li, L., Wolfinger, R., Bao, W., Paules, R. S., Bushel, P. R., Li, J., Shi, T., Nikolskaya, T., Nikolsky, Y., Hong, H., Deng, Y., Cheng, Y., Fang, H., Shi, L. & Tong, W.

- (2010). Consistency of predictive signature genes and classifiers generated using different microarray platforms. *Pharmacogenomics Journal*, **10** (4), 247–257.
- Fare, T. L., Coffey, E. M., Dai, H., He, Y. D., Kessler, D. A., Kilian, K. A., Koch, J. E., LeProust, E., Marton, M. J., Meyer, M. R., Stoughton, R. B., Tokiwa, G. Y. & Wang, Y. (2003). Effects of atmospheric ozone on microarray data quality. *Analytical Chemistry*, **75** (17), 4672–4675.
- Farnir, F., Grisart, B., Coppieters, W., Riquet, J., Berzi, P., Cambisano, N., Karim, L., Mni, M., Moisisio, S., Simon, P., Wagenaar, D., Vilkki, J. & Georges, M. (2002). Simultaneous mining of linkage and linkage disequilibrium to fine map quantitative trait loci in outbred half-sib pedigrees: revisiting the location of a quantitative trait locus with major effect on milk production on bovine chromosome 14. *Genetics*, **161** (1), 275–287.
- Farrell, Jr, H. M., Jimenez-Flores, R., Bleck, G. T., Brown, E. M., Butler, J. E., Creamer, L. K., Hicks, C. L., Hollar, C. M., Ng-Kwai-Hang, K. F. & Swaisgood, H. E. (2004). Nomenclature of the proteins of cows' milk—sixth revision. *Journal of Dairy Science*, **87** (6), 1641–1674.
- Fernando, R. & Grossman, M. (1989). Marker assisted selection using best linear unbiased prediction. *Genetics Selection Evolution*, **21** (4), 467–477.
- Flori, L., Fritz, S., Jaffrezic, F., Boussaha, M., Gut, I., Heath, S., Foulley, J. L. & Gautier, M. (2009). The genome response to artificial selection: a case study in dairy cattle. *PLoS ONE*, **4** (8), e6595.
- Fragkou, I. A., Dagleish, M. P., Papaioannou, N., Cripps, P. J., Boscós, C. M., Ververidis, H. N., Orfanou, D. C., Solomakos, N., Finlayson, J., Govaris, A., Kyriazakis, I. & Fthenakis, G. C. (2010). The induction of lymphoid follicle-like structures in the ovine teat duct following experimental infection with *Mannheimia haemolytica*. *Veterinary Journal*, **184** (2), 194–200.
- Frandsen, R. D. (1986). *Anatomy and Physiology of Farm Animals*. 4th edition, Lea & Febiger, Philadelphia.
- Franklin, B. S., Parroche, P., Ataide, M. A., Lauw, F., Ropert, C., de Oliveira, R. B., Pereira, D., Tada, M. S., Nogueira, P., da Silva, L. H., Bjorkbacka, H., Golenbock, D. T. & Gazzinelli, R. T. (2009). Malaria primes the innate immune response due to interferon-gamma induced enhancement of toll-like receptor expression and function. *Proceedings of the National Academy of Sciences of the United States of America*, **106** (14), 5789–5794.
- Fuentes-Panana, E. M., Bannish, G., Karnell, F. G., Trembl, J. F. & Monroe, J. G. (2006). Analysis of the individual contributions of Igalpha (CD79a)- and Igbeta (CD79b)-mediated tonic signaling for bone marrow B cell development and peripheral B cell maturation. *Journal of Immunology*, **177** (11), 7913–7922.
- Fusco, J. C., Tong, W. & Shi, L. (2007). QA/QC issues to aid regulatory acceptance of microarray gene expression data. *Environmental and Molecular Mutagenesis*, **48** (5), 349–353.
- Gabriel, S. B., Schaffner, S. F., Nguyen, H., Moore, J. M., Roy, J., Blumenstiel, B., Higgins, J., de Felice, M., Lochner, A., Faggart, M., Liu-Cordero, S. N., Rotimi, C., Adeyemo, A., Cooper, R., Ward, R., Lander, E. S., Daly, M. J. & Altshuler, D. (2002). The structure of haplotype blocks in the human genome. *Science*, **296** (5576), 2225–2229.
- Gallais, A., Moreau, L. & Charcosset, A. (2007). Detection of marker-qtL associations by studying change in marker frequencies with selection. *Theoretical and Applied Genetics*, **114** (4), 669–681.
- Gao, Q. & Womack, J. E. (1997). Comparative mapping of anchor loci from HSA19 to cattle

- chromosomes 7 and 18. *Journal of Heredity*, **88** (6), 524–527.
- Gautier, M., Flori, L., Riebler, A., Jaffrezic, F., Laloe, D., Gut, I., Moazami-Goudarzi, K. & Foulley, J. L. (2009). A whole genome Bayesian scan for adaptive genetic divergence in West African cattle. *BMC Genomics*, **10**, 550.
- Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., Hornik, K., Hothorn, T., Huber, W., Iacus, S., Irizarry, R., Leisch, F., Li, C., Maechler, M., Rossini, A. J., Sawitzki, G., Smith, C., Smyth, G., Tierney, L., Yang, J. Y. & Zhang, J. (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology*, **5** (10), R80.
- Georges, M., Nielsen, D., Mackinnon, M., Mishra, A., Okimoto, R., Pasquino, A. T., Sargeant, L. S., Sorensen, A., Steele, M. R., Zhao, X. & et al. (1995). Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. *Genetics*, **139** (2), 907–920.
- Ghoshal, K., Majumder, S., Zhu, Q., Hunzeker, J., Datta, J., Shah, M., Sheridan, J. F. & Jacob, S. T. (2001). Influenza virus infection induces metallothionein gene expression in the mouse liver and lung by overlapping but distinct molecular mechanisms. *Molecular and Cellular Biology*, **21** (24), 8301–8317.
- Gilmour, A., Gogel, B., Cullis, B. & Thompson, R. (2006). *ASReml User Guide Release 2.0*. VSN International Ltd Hemel Hempstead, HP1 1ES, UK.
- Glass, E. J. & Jensen, K. (2007). Resistance and susceptibility to a protozoan parasite of cattle—gene expression differences in macrophages from different breeds of cattle. *Veterinary Immunology and Immunopathology*, **120** (1-2), 20–30.
- Goddard, M. & Beilharz, R. (1977). Natural selection and animal breeding. In *Proceedings of the 3rd International Congress SABRAO, Canberra, Australia* vol. 4, pp. 19–21.
- Goeman, J. J., van de Geer, S. A., de Kort, F. & van Houwelingen, H. C. (2004). A global test for groups of genes: testing association with a clinical outcome. *Bioinformatics*, **20** (1), 93–99.
- Goertz, I., Baes, C., Weimann, C., Reinsch, N. & Erhardt, G. (2009). Association between single nucleotide polymorphisms in the CXCR1 gene and somatic cell score in Holstein dairy cattle. *Journal of Dairy Science*, **92** (8), 4018 – 4022.
- Goldammer, T., Kata, S. R., Brunner, R. M., Dorroch, U., Sanftleben, H., Schwerin, M. & Womack, J. E. (2002). A comparative radiation hybrid map of bovine chromosome 18 and homologous chromosomes in human and mice. *Proceedings of the National Academy of Sciences of the United States of America*, **99** (4), 2106–2111.
- Goldammer, T., Kata, S. R., Brunner, R. M., Kühn, C., Weikard, R., Womack, J. E. & Schwerin, M. (2004). High-resolution comparative mapping between human chromosomes 4 and 8 and bovine chromosome 27 provides genes and segments serving as positional candidates for udder health in cattle. *Genomics*, **84** (4), 696–706.
- Gonzalez-Chavez, S. A., Arevalo-Gallegos, S. & Rascon-Cruz, Q. (2009). Lactoferrin: structure, function and applications. *International Journal of Antimicrobial Agents*, **33** (4), 301.e1–8.
- Grapes, L., Firat, M. Z., Dekkers, J. C., Rothschild, M. F. & Fernando, R. L. (2006). Optimal haplotype structure for linkage disequilibrium-based fine mapping of quantitative trait loci using identity by descent. *Genetics*, **172** (3), 1955–1965.
- Grewal, P. K., Botton, M., Ramirez, K., Collins, B. E., Saito, A., Green, R. S., Ohtsubo, K., Chui, D. & Marth, J. D. (2006). ST6Gal-I restrains CD22-dependent antigen receptor endocytosis and Shp-1 recruitment in normal and pathogenic immune signaling. *Molecular*

- and Cellular Biology*, **26** (13), 4970–4981.
- Griffin, K. B., Michal, J. J., Fox, L. K., Gaskins, C. T. & Jiang, Z. (2005). Fine mapping of the bovine chromosome 22q24 region that harbours antimicrobial genes and a QTL for somatic cell score. *Animal Genetics*, **36** (5), 448–450.
- Grimwood, J., Gordon, L. A., Olsen, A., Terry, A., Schmutz, J., Lamerdin, J., Hellsten, U., Goodstein, D., Couronne, O., Tran-Gyamfi, M., Aerts, A., Altherr, M., Ashworth, L., Bajorek, E., Black, S., Branscomb, E., Caenepeel, S., Carrano, A., Caoile, C., Chan, Y. M., Christensen, M., Cleland, C. A., Copeland, A., Dalin, E., Dehal, P., Denys, M., Detter, J. C., Escobar, J., Flowers, D., Fotopulos, D., Garcia, C., Georgescu, A. M., Glavina, T., Gomez, M., Gonzales, E., Groza, M., Hammon, N., Hawkins, T., Haydu, L., Ho, I., Huang, W., Israni, S., Jett, J., Kadner, K., Kimball, H., Kobayashi, A., Larionov, V., Leem, S. H., Lopez, F., Lou, Y., Lowry, S., Malfatti, S., Martinez, D., McCready, P., Medina, C., Morgan, J., Nelson, K., Nolan, M., Ovcharenko, I., Pitluck, S., Pollard, M., Popkie, A. P., Predki, P., Quan, G., Ramirez, L., Rash, S., Retterer, J., Rodriguez, A., Rogers, S., Salamov, A., Salazar, A., She, X., Smith, D., Slezak, T., Solovyev, V., Thayer, N., Tice, H., Tsai, M., Ustaszewska, A., Vo, N., Wagner, M., Wheeler, J., Wu, K., Xie, G., Yang, J., Dubchak, I., Furey, T. S., De Jong, P., Dickson, M., Gordon, D., Eichler, E. E., Pennacchio, L. A., Richardson, P., Stubbs, L., Rokhsar, D. S., Myers, R. M., Rubin, E. M. & Lucas, S. M. (2004). The DNA sequence and biology of human chromosome 19. *Nature*, **428** (6982), 529–535.
- Grisart, B., Coppieters, W., Farnir, F., Karim, L., Ford, C., Berzi, P., Cambisano, N., Mni, M., Reid, S., Simon, P., Spelman, R., Georges, M. & Snell, R. (2002). Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. *Genome Research*, **12** (2), 222–231.
- Grohn, Y. T., Eicker, S. W., Ducrocq, V. & Hertl, J. A. (1998). Effect of diseases on the culling of Holstein dairy cows in New York State. *Journal of Dairy Science*, **81** (4), 966–978.
- Grys, T. E., Siegel, M. B., Lathem, W. W. & Welch, R. A. (2005). The StcE protease contributes to intimate adherence of enterohemorrhagic *Escherichia coli* O157:H7 to host cells. *Infection and Immunity*, **73** (3), 1295–1303.
- Guda, K., Cui, H., Garg, S., Dong, M., Nambiar, P. R., Achenie, L. E. & Rosenberg, D. W. (2003). Multistage gene expression profiling in a differentially susceptible mouse colon cancer model. *Cancer Letters*, **191** (1), 17–25.
- Günther, J., Koczan, D., Yang, W., Nürnberg, G., Repsilber, D., Schuberth, H. J., Park, Z., Maqbool, N., Molenaar, A. & Seyfert, H. M. (2009). Assessment of the immune capacity of mammary epithelial cells: comparison with mammary tissue after challenge with *Escherichia coli*. *Veterinary Research*, **40** (4), 31.
- Guo, L., Lobenhofer, E. K., Wang, C., Shippy, R., Harris, S. C., Zhang, L., Mei, N., Chen, T., Herman, D., Goodsaid, F. M., Hurban, P., Phillips, K. L., Xu, J., Deng, X., Sun, Y. A., Tong, W., Dragan, Y. P. & Shi, L. (2006). Rat toxicogenomic study reveals analytical consistency across microarray platforms. *Nature Biotechnology*, **24** (9), 1162–1169.
- Haile-Mariam, M., Goddard, M. E. & Bowman, P. J. (2001). Estimates of genetic parameters for daily somatic cell count of Australian dairy cattle. *Journal of Dairy Science*, **84** (5), 1255–1264.
- Halasa, T., Huijps, K., Osteras, O. & Hogeveen, H. (2007). Economic effects of bovine mastitis and mastitis management: a review. *Veterinary Quarterly*, **29** (1), 18–31.
- Hamerman, J. A., Jarjoura, J. R., Humphrey, M. B., Nakamura, M. C., Seaman, W. E. &



- Lanier, L. L. (2006). Cutting edge: inhibition of TLR and FcR responses in macrophages by triggering receptor expressed on myeloid cells (TREM)-2 and DAP12. *Journal of Immunology*, **177** (4), 2051–2055.
- Hamosh, A., Scott, A. F., Amberger, J., Bocchini, C., Valle, D. & McKusick, V. A. (2002). Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Research*, **30** (1), 52–55.
- Hannan, F. M., Nesbit, M. A., Turner, J. J., Stacey, J. M., Cianferotti, L., Christie, P. T., Conigrave, A. D., Whyte, M. P. & Thakker, R. V. (2010). Comparison of human chromosome 19q13 and syntenic region on mouse chromosome 7 reveals absence, in man, of 11.6 Mb containing four mouse calcium-sensing receptor-related sequences: relevance to familial benign hypocalciuric hypercalcaemia type 3. *European Journal of Human Genetics*, **18** (4), 442–447.
- Hardiman, G. (2004). Microarray platforms—comparisons and contrasts. *Pharmacogenomics*, **5** (5), 487–502.
- Harp, J. A., Runnels, P. L. & Pesch, B. A. (1988). Lymphocyte recirculation in cattle: patterns of localization by mammary and mesenteric lymph node lymphocytes. *Veterinary Immunology and Immunopathology*, **20** (1), 31–39.
- Harp, J. A., Waters, T. E. & Goff, J. P. (2004). Lymphocyte subsets and adhesion molecule expression in milk and blood of periparturient dairy cattle. *Veterinary Immunology and Immunopathology*, **102** (1-2), 9–17.
- Hastbacka, J., de la Chapelle, A., Kaitila, I., Sistonen, P., Weaver, A. & Lander, E. (1992). Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland. *Nature Genetics*, **2** (3), 204–211.
- Hayashi, T., Usui, M., Nishioka, J., Zhang, Z. X. & Suzuki, K. (1998). Regulation of the human protein C inhibitor gene expression in HepG2 cells: role of Sp1 and AP2. *Biochemical Journal*, **332** ( Pt 2), 573–582.
- Hayes, B. & Goddard, M. E. (2001). The distribution of the effects of genes affecting quantitative traits in livestock. *Genet Sel Evol*, **33** (3), 209–229.
- Hayes, B. J., Chamberlain, A. J., Maceachern, S., Savin, K., McPartlan, H., MacLeod, I., Sethuraman, L. & Goddard, M. E. (2009). A genome map of divergent artificial selection between *Bos taurus* dairy cattle and *Bos taurus* beef cattle. *Animal Genetics*, **40** (2), 176–184.
- Hayes, H., Elduque, C., Gautier, M., Schibler, L., Cribiu, E. & Eggen, A. (2003). Mapping of 195 genes in cattle and updated comparative map with man, mouse, rat and pig. *Cytogenetic and Genome Research*, **102** (1-4), 16–24.
- Hedegaard, J., Arce, C., Biciato, S., Bonnet, A., Buitenhuis, B., Collado-Romero, M., Conley, L. N., Sancristobal, M., Ferrari, F., Garrido, J. J., Groenen, M. A., Hornshoj, H., Hulsegge, I., Jiang, L., Jimenez-Marin, A., Kommadath, A., Lagarrigue, S., Leunissen, J. A., Liaubet, L., Neerincx, P. B., Nie, H., van der Poel, J., Prickett, D., Ramirez-Boo, M., Rebel, J. M., Robert-Granie, C., Skarman, A., Smits, M. A., Sorensen, P., Tossier-Klopp, G. & Watson, M. (2009). Methods for interpreting lists of affected genes obtained in a DNA microarray experiment. *BMC Proceedings*, **3 Suppl 4**, S5.
- Hein, W. R. & Mackay, C. R. (1991). Prominence of gamma delta T cells in the ruminant immune system. *Immunology Today*, **12** (1), 30–34.
- Heuven, H. C., Bastiaansen, J. W. & van den Berg, S. M. (2009). Efficient detection of QTL with large effects in a simulated pig-type pedigree using selective genotyping. *BMC Proceedings*, **3 Suppl 1**, S8.

- Heyen, D. W., Weller, J. I., Ron, M., Band, M., Beever, J. E., Feldmesser, E., Da, Y., Wiggans, G. R., van Raden, P. M. & Lewin, H. A. (1999). A genome scan for QTL influencing milk production and health traits in dairy cattle. *Physiological Genomics*, **1** (3), 165–175.
- Hibbitt, K., Craven, N. & Batten, E. (1996). Bovine Medicine: Diseases and Husbandry of Cattle. chapter Anatomy, physiology and immunology of the udder, pp. 273–278. Blackwell.
- Hillerton, J. E., West, J. G. & Shearn, M. F. (1992). The cost of summer mastitis. *Veterinary Record*, **131** (14), 315–317.
- Hintermair, V. (2007). Comparative Analysis of Bos Taurus and Homo Sapiens DNA Microarrays. Master thesis LMU München.
- Hocquette, J. F., Lehnert, S., Barendse, W., Cassar-Malek, I. & Picard, B. (2007). Recent advances in cattle functional genomics and their application to beef quality. *Animal*, **1** (1), 159–173.
- Hoffman, S. M., Nelson, D. R. & Keeney, D. S. (2001). Organization, structure and evolution of the CYP2 gene cluster on human chromosome 19. *Pharmacogenetics*, **11** (8), 687–698.
- Holmberg, M. & Andersson-Eklund, L. (2004). Quantitative trait loci affecting health traits in Swedish dairy cattle. *Journal of Dairy Science*, **87** (8), 2653–2659.
- Houwen, R. H., Baharloo, S., Blankenship, K., Raeymaekers, P., Juyn, J., Sandkuijl, L. A. & Freimer, N. B. (1994). Genome screening by searching for shared segments: mapping a gene for benign recurrent intrahepatic cholestasis. *Nature Genetics*, **8** (4), 380–386.
- Hovey, R. C., McFadden, T. B. & Akers, R. M. (1999). Regulation of mammary gland growth and morphogenesis by the mammary fat pad: a species comparison. *Journal of Mammary Gland Biology and Neoplasia*, **4** (1), 53–68.
- Hsu, K. C., Chida, S., Geraghty, D. E. & Dupont, B. (2002). The killer cell immunoglobulin-like receptor (KIR) genomic region: gene-order, haplotypes and allelic polymorphism. *Immunological Reviews*, **190**, 40–52.
- Hu, Z. L., Fritz, E. R. & Reecy, J. M. (2007). AnimalQTLdb: a livestock QTL database tool set for positional QTL information mining and beyond. *Nucleic Acids Research*, **35** (Database issue), 604–609.
- Hubbard, T. J., Aken, B. L., Ayling, S., Ballester, B., Beal, K., Bragin, E., Brent, S., Chen, Y., Clapham, P., Clarke, L., Coates, G., Fairley, S., Fitzgerald, S., Fernandez-Banet, J., Gordon, L., Graf, S., Haider, S., Hammond, M., Holland, R., Howe, K., Jenkinson, A., Johnson, N., Kahari, A., Keefe, D., Keenan, S., Kinsella, R., Kokocinski, F., Kulesha, E., Lawson, D., Longden, I., Megy, K., Meidl, P., Overduin, B., Parker, A., Pritchard, B., Rios, D., Schuster, M., Slater, G., Smedley, D., Spooner, W., Spudich, G., Trevanion, S., Vilella, A., Vogel, J., White, S., Wilder, S., Zadissa, A., Birney, E., Cunningham, F., Curwen, V., Durbin, R., Fernandez-Suarez, X. M., Herrero, J., Kasprzyk, A., Proctor, G., Smith, J., Searle, S. & Flicek, P. (2009). Ensembl 2009. *Nucleic Acids Research*, **37** (Database issue), 690–697.
- Hummel, M., Meister, R. & Mansmann, U. (2008). GlobalANCOVA: exploration and assessment of gene group effects. *Bioinformatics*, **24** (1), 78–85.
- Hyka-Nouspikel, N., Lucian, L., Murphy, E., McClanahan, T. & Phillips, J. H. (2007). DAP10 deficiency breaks the immune tolerance against transplantable syngeneic melanoma. *Journal of Immunology*, **179** (6), 3763–3771.
- Hyka-Nouspikel, N. & Phillips, J. H. (2006). Physiological roles of murine DAP10 adapter protein in tumor immunity and autoimmunity. *Immunological Reviews*, **214**, 106–117.

- Ingvartsen, K., Dewhurst, R. & Friggens, N. (2003). On the relationship between lactational performance and health: is it yield or metabolic imbalance that cause production diseases in dairy cattle? A position paper. *Livestock Production Science*, **83** (2-3), 277–308.
- Itoh, T., Watanabe, T., Ihara, N., Mariani, P., Beattie, C. W., Sugimoto, Y. & Takasuga, A. (2005). A comprehensive radiation hybrid map of the bovine genome comprising 5593 loci. *Genomics*, **85** (4), 413–424.
- Jansen, R. C. & Nap, J. P. (2001). Genetical genomics: the added value from segregation. *Trends in Genetics*, **17** (7), 388–391.
- Jaster, E. H. (2005). Evaluation of quality, quantity, and timing of colostrum feeding on immunoglobulin G1 absorption in Jersey calves. *Journal of Dairy Science*, **88** (1), 296–302.
- Jeffery, I. B., Higgins, D. G. & Culhane, A. C. (2006). Comparison and evaluation of methods for generating differentially expressed gene lists from microarray data. *BMC Bioinformatics*, **7**, 359.
- Jensen, K., Paxton, E., Waddington, D., Talbot, R., Darghouth, M. A. & Glass, E. J. (2008). Differences in the transcriptional responses induced by *Theileria annulata* infection in bovine monocytes derived from resistant and susceptible cattle breeds. *International Journal for Parasitology*, **38** (3-4), 313–325.
- Jensen, K., Talbot, R., Paxton, E., Waddington, D. & Glass, E. J. (2006). Development and validation of a bovine macrophage specific cDNA microarray. *BMC Genomics*, **7**.
- Jeyaseelan, K., Armugam, A., Donghui, M. & Tan, N. H. (2000). Structure and phylogeny of the venom group I phospholipase A(2) gene. *Molecular Biology and Evolution*, **17** (7), 1010–1021.
- Jin, C., Lan, H., Attie, A. D., Churchill, G. A., Bulutuglo, D. & Yandell, B. S. (2004). Selective phenotyping for increased efficiency in genetic mapping studies. *Genetics*, **168** (4), 2285–2293.
- Johnston, D., Nelson, S., Schulman, J., Savage, A. & Powell, T. (1998). Redefining a thesaurus: Term-centric no more. *Journal of the American Medical Informatics Association*, **Suppl**, 1025.
- Kadarmideen, H. & Simm, G. (2002). Selection responses expected from index selection including disease resistance, fertility and longevity in dairy cattle. *Proceedings of the 7th World Congress on Genetics Applied to Livestock Production, Montpellier, France, August, 2002. Session 1, - (-), -*.
- Kadarmideen, H. N. & Pryce, J. E. (2001). Genetic and economic relationships between somatic cell count and clinical mastitis and their use in selection for mastitis resistance in dairy cattle. *Animal Science*, **73**, 19 – 28.
- Kasprzyk, A., Keefe, D., Smedley, D., London, D., Spooner, W., Melsopp, C., Hammond, M., Rocca-Serra, P., Cox, T. & Birney, E. (2004). Ensmart: a generic system for fast and flexible access to biological data. *Genome Research*, **14** (1), 160–169.
- Keane, O. M., Zadissa, A., Wilson, T., Hyndman, D. L., Greer, G. J., Baird, D. B., McCulloch, A. F., Crawford, A. M. & McEwan, J. C. (2006). Gene expression profiling of naive sheep genetically resistant and susceptible to gastrointestinal nematodes. *BMC Genomics*, **7**, 42.
- Kelly, A. L., Tiernan, D., O'Sullivan, C. & Joyce, P. (2000). Correlation between bovine milk somatic cell count and polymorphonuclear leukocyte level for samples of bulk milk and milk from individual cows. *Journal of Dairy Science*, **83** (2), 300–304.
- Kenward, M. G. & Roger, J. H. (1997). Small sample inference for fixed effects from restricted

- maximum likelihood. *Biometrics*, **53** (3), 983–997.
- Khatkar, M. S., Thomson, P. C., Tammen, I., Cavanagh, J. A., Nicholas, F. W. & Raadsma, H. W. (2006). Linkage disequilibrium on chromosome 6 in Australian Holstein-Friesian cattle. *Genet Sel Evol*, **38** (5), 463–477.
- Khatkar, M. S., Thomson, P. C., Tammen, I. & Raadsma, H. W. (2004). Quantitative trait loci mapping in dairy cattle: review and meta-analysis. *Genet Sel Evol*, **36** (2), 163–190.
- Kikuchi, T., Abe, T., Hoshi, S., Matsubara, N., Tominaga, Y., Satoh, K. & Nukiwa, T. (1998). Structure of the murine secretory leukoprotease inhibitor (Slpi) gene and chromosomal localization of the human and murine SLPI genes. *American Journal of Respiratory Cell and Molecular Biology*, **19** (6), 875–880.
- Kim, E. S. & Kirkpatrick, B. W. (2009). Linkage disequilibrium in the North American Holstein population. *Animal Genetics*, **40** (3), 279–288.
- Kim, E. S., Shi, X., Cobanoglu, O., Weigel, K., Berger, P. J. & Kirkpatrick, B. W. (2009). Refined mapping of twinning-rate quantitative trait loci on bovine chromosome 5 and analysis of insulin-like growth factor-1 as a positional candidate gene. *Journal of Animal Science*, **87** (3), 835–843.
- Kitayner, M., Rozenberg, H., Kessler, N., Rabinovich, D., Shaulov, L., Haran, T. E. & Shakked, Z. (2006). Structural basis of dna recognition by p53 tetramers. *Molecular Cell*, **22** (6), 741–753.
- Klebanov, L. & Yakovlev, A. (2007). How high is the level of technical noise in microarray data? *Biology Direct*, **2**, 9.
- Klos, K., Shimmin, L., Ballantyne, C., Boerwinkle, E., Clark, A., Coresh, J., Hanis, C., Liu, K., Sayre, S. & Hixson, J. (2008). APOE/C1/C4/C2 hepatic control region polymorphism influences plasma apoE and LDL cholesterol levels. *Human Molecular Genetics*, **17** (13), 2039–2046.
- Klungland, H., Sabry, A., Heringstad, B., Olsen, H. G., Gomez-Raya, L., Vage, D. I., Ol-saker, I., Odegard, J., Klemetsdal, G., Schulman, N., Vilkki, J., Ruane, J., Aasland, M., Ronningen, K. & Lien, S. (2001). Quantitative trait loci affecting clinical mastitis and somatic cell count in dairy cattle. *Mammalian Genome*, **12** (11), 837–842.
- Knott, S. A., Elsen, J. M. & Haley, C. S. (1996). Methods for multiple-marker mapping of quantitative trait loci in half-sib populations. *Theoretical and Applied Genetics*, **93** (1-2), 71–80.
- Kolbehdari, D., Wang, Z., Grant, J. R., Murdoch, B., Prasad, A., Xiu, Z., Marques, E., Stothard, P. & Moore, S. S. (2009). A whole genome scan to map QTL for milk production traits and somatic cell score in Canadian Holstein bulls. *Journal of Animal Breeding and Genetics*, **126** (3), 216–227.
- Korbie, D. J. & Mattick, J. S. (2008). Touchdown PCR for increased specificity and sensitivity in PCR amplification. *Nature Protocols*, **3** (9), 1452–1456.
- Korol, A., Frenkel, Z., Cohen, L., Lipkin, E. & Soller, M. (2007). Fractioned DNA pooling: a new cost-effective strategy for fine mapping of quantitative trait loci. *Genetics*, **176** (4), 2611–2623.
- Koshland, M. E. (1985). The coming of age of the immunoglobulin J chain. *Annual Review of Immunology*, **3**, 425–453.
- Kossaibati, M. A. & Esslemont, R. J. (1997). The costs of production diseases in dairy herds in England. *Veterinary Journal*, **154** (1), 41–51.
- Koyama, M., Ishihara, K., Karasuyama, H., Cordell, J. L., Iwamoto, A. & Nakamura, T.

- (1997). CD79 alpha/CD79 beta heterodimers are expressed on pro-B cell surfaces without associated mu heavy chain. *International Immunology*, **9** (11), 1767–1772.
- Krensky, A. M. (2000). Granulysin: a novel antimicrobial peptide of cytolytic T lymphocytes and natural killer cells. *Biochemical Pharmacology*, **59** (4), 317–320.
- Kühn, C., Bennewitz, J., Reinsch, N., Xu, N., Thomsen, H., Looft, C., Brockmann, G. A., Schwerin, M., Weimann, C., Hiendleder, S., Erhardt, G., Medjugorac, I., Forster, M., Brenig, B., Reinhardt, F., Reents, R., Russ, I., Averdunk, G., Blumel, J. & Kalm, E. (2003). Quantitative trait loci mapping of functional traits in the German Holstein cattle population. *Journal of Dairy Science*, **86** (1), 360–368.
- Kühn, C., Reinhardt, F. & Schwerin, M. (2008). Marker assisted selection of heifers improved milk somatic cell count compared to selection on conventional pedigree breeding values. *Archiv für Tierzucht*, **51** (1), 23 – 32.
- Kulberg, S., Heringstad, B., Guttersrud, O. A. & Olsaker, I. (2007). Study on the association of BoLA-DRB3.2 alleles with clinical mastitis in Norwegian Red cows. *Journal of Animal Breeding and Genetics*, **124** (4), 201–207.
- Kurosaki, T., Johnson, S. A., Pao, L., Sada, K., Yamamura, H. & Cambier, J. C. (1995). Role of the Syk autophosphorylation site and SH2 domains in B cell antigen receptor signaling. *Journal of Experimental Medicine*, **182** (6), 1815–1823.
- Lam, A. C., Powell, J., Wei, W. H., de Koning, D. J. & Haley, C. S. (2009). A combined strategy for quantitative trait loci detection by genome-wide association. *BMC Proceedings*, **3 Suppl 1**, S6.
- Lang, P., Stolpa, J. C., Freiberg, B. A., Crawford, F., Kappler, J., Kupfer, A. & Cambier, J. C. (2001). TCR-induced transmembrane signaling by peptide/MHC class II via associated Ig-alpha/beta dimers. *Science*, **291** (5508), 1537–1540.
- Larkin, D. M., Everts-van der Wind, A., Rebeiz, M., Schweitzer, P. A., Bachman, S., Green, C., Wright, C. L., Campos, E. J., Benson, L. D., Edwards, J., Liu, L., Osoegawa, K., Womack, J. E., de Jong, P. J. & Lewin, H. A. (2003). A cattle-human comparative map built with cattle BAC-ends and human genome sequence. *Genome Research*, **13** (8), 1966–1972.
- Ledur, M. C., Navarro, N. & Perez-Enciso, M. (2009). Data modeling as a main source of discrepancies in single and multiple marker association methods. *BMC Proceedings*, **3 Suppl 1**, S9.
- Lehnert, S., Wang, Y. & Byrne, K. (2004). Development and application of a bovine cDNA microarray for expression profiling of muscle and adipose tissue. *Australian Journal of Experimental Agriculture*, **44** (11), 1127–1133.
- Leitner, G., Krifucks, O., Jacoby, S., Lavi, Y. & Silanikove, N. (2008). Concentrations of ganglioside type M1 and immunoglobulin G in colostrum are inversely related to bacterial infection at early lactation in cows. *Journal of Dairy Science*, **91** (9), 3337–3342.
- Leung, Y. F. & Cavalieri, D. (2003). Fundamentals of cDNA microarray data analysis. *Trends in Genetics*, **19** (11), 649–659.
- Lewontin, R. C. (1964). The Interaction of Selection and Linkage. I. General Considerations; Heterotic Models. *Genetics*, **49** (1), 49–67.
- Leyva-Baca, I., Schenkel, F., Martin, J. & Karrow, N. A. (2008). Polymorphisms in the 5' upstream region of the CXCR1 chemokine receptor gene, and their association with somatic cell score in Holstein cattle in Canada. *Journal of Dairy Science*, **91** (1), 407–417.
- Leyva-Baca, I., Schenkel, F., Sharma, B. S., Jansen, G. B. & Karrow, N. A. (2007). Identi-

- fication of single nucleotide polymorphisms in the bovine CCL2, IL8, CCR2 and IL8RA genes and their association with health and production in Canadian Holsteins. *Animal Genetics*, **38** (3), 198–202.
- Liebich, H., Reese, S. & Budras, K. (2004). Funktionelle Histologie der Haussäugetiere. chapter Allgemeine Körperdecke, pp. 312–330. Schattauer 4. edition.
- Lillvis, J. H., Kyo, Y., Tromp, G., Lenk, G. M., Li, M., Lu, Q., Igo, Jr, R. P., Sakalihan, N., Ferrell, R. E., Schworer, C. M., Gatalica, Z., Land, S. & Kuivaniemi, H. (2011). Analysis of positional candidate genes in the AAA1 susceptibility locus for abdominal aortic aneurysms on chromosome 19. *BMC Medical Genetics*, **12** (1), 14.
- Lindmark-Mansson, H., Braenning, C., Alden, G. & Paulsson, M. (2006). Relationship between somatic cell count, individual leukocyte populations and milk components in bovine udder quarter milk. *International Dairy Journal*, **16** (7), 717 – 727.
- Lipkin, E., Mosig, M. O., Darvasi, A., Ezra, E., Shalom, A., Friedmann, A. & Soller, M. (1998). Quantitative trait locus mapping in dairy cattle by means of selective milk DNA pooling using dinucleotide microsatellite markers: analysis of milk protein percentage. *Genetics*, **149** (3), 1557–1567.
- Liu, G., Loraine, A. E., Shigeta, R., Cline, M., Cheng, J., Valmeekam, V., Sun, S., Kulp, D. & Siani-Rose, M. A. (2003). NetAffx: Affymetrix probesets and annotations. *Nucleic Acids Research*, **31** (1), 82–86.
- Liu, H. C., Cheng, H. H., Tirunagaru, V., Sofer, L. & Burnside, J. (2001). A strategy to identify positional candidate genes conferring Marek's disease resistance by integrating DNA microarrays and genetic mapping. *Animal Genetics*, **32** (6), 351–359.
- Liu, Z., Reinhardt, F., Bunger, A. & Reents, R. (2004). Derivation and calculation of approximate reliabilities and daughter yield-deviations of a random regression test-day model for genetic evaluation of dairy cattle. *Journal of Dairy Science*, **87** (6), 1896–1907.
- Longeri, M., Polli, M., Strillacci, M. G., Samore, A. B. & Zanotti, M. (2006). Short communication: quantitative trait loci affecting the somatic cell score on chromosomes 4 and 26 in Italian Holstein cattle. *Journal of Dairy Science*, **89** (8), 3175–3177.
- Lu, L., Wei, L., Peirce, J. L., Wang, X., Zhou, J., Homayouni, R., Williams, R. W. & Airey, D. C. (2008). Using gene expression databases for classical trait QTL candidate gene discovery in the BXD recombinant inbred genetic reference population: mouse forebrain weight. *BMC Genomics*, **9**, 444.
- Lund, M. S., Guldbandsen, B., Buitenhuis, A. J., Thomsen, B. & Bendixen, C. (2008). Detection of quantitative trait loci in Danish Holstein cattle affecting clinical mastitis, somatic cell score, udder conformation traits, and assessment of associated effects on milk yield. *Journal of Dairy Science*, **91** (10), 4028–4036.
- Lund, M. S., Sahana, G., Andersson-Eklund, L., Hastings, N., Fernandez, A., Schulman, N., Thomsen, B., Viitala, S., Williams, J. L., Sabry, A., Viinalass, H. & Vilkki, J. (2007). Joint analysis of quantitative trait loci for clinical mastitis and somatic cell score on five chromosomes in three Nordic dairy cattle breeds. *Journal of Dairy Science*, **90** (11), 5282–5290.
- Lunemann, A., Lunemann, J. D. & Munz, C. (2009). Regulatory NK-cell functions in inflammation and autoimmunity. *Molecular Medicine*, **15** (9-10), 352–358.
- Mackay, T. F. (2001). The genetic architecture of quantitative traits. *Annual Review of Genetics*, **35**, 303–339.
- Mallard, B. A., Dekkers, J. C., Ireland, M. J., Leslie, K. E., Sharif, S., Vankampen, C. L., Wagter, L. & Wilkie, B. N. (1998). Alteration in immune responsiveness during the

- peripartum period and its ramification on dairy cow and calf health. *Journal of Dairy Science*, **81** (2), 585–595.
- Maouche, S., Poirier, O., Godefroy, T., Olaso, R., Gut, I., Collet, J. P., Montalescot, G. & Cambien, F. (2008). Performance comparison of two microarray platforms to assess differential gene expression in human monocyte and macrophage cells. *BMC Genomics*, **9**, 302.
- Mariasegaram, M., Robinson, N. A. & Goddard, M. E. (2007). Empirical evaluation of selective DNA pooling to map QTL in dairy cattle using a half-sib design by comparison to individual genotyping and interval mapping. *Genet Sel Evol*, **39** (3), 267–283.
- Mark, T., Fikse, W., Emanuelson, U. & Philipsson, J. (2002). International genetic evaluations of Holstein sires for milk somatic cell and clinical mastitis. *Journal of Dairy Science*, **85** (9), 2384–2392.
- Mattapallil, M. J., Augello, A., Cheadle, C., Teichberg, D., Becker, K. G., Chan, C. C., Mattapallil, J. J., Pennesi, G. & Caspi, R. R. (2008). Differentially expressed genes in MHC-compatible rat strains that are susceptible or resistant to experimental autoimmune uveitis. *Investigative Ophthalmology and Visual Science*, **49** (5), 1957–1970.
- Matukumalli, L. K., Lawley, C. T., Schnabel, R. D., Taylor, J. F., Allan, M. F., Heaton, M. P., O'Connell, J., Moore, S. S., Smith, T. P., Sonstegard, T. S. & van Tassell, C. P. (2009). Development and characterization of a high density SNP genotyping assay for cattle. *PLoS ONE*, **4** (4), e5350.
- Mavrogianni, V. S., Cripps, P. J., Brooks, H., Taitzoglou, I. A. & Fthenakis, G. C. (2007). Presence of subepithelial lymphoid nodules in the teat of ewes. *Anatomia Histologia Embryologia*, **36** (3), 168–171.
- Mavrogianni, V. S., Fthenakis, G. C., Brooks, H., Papaioannou, N., Cripps, P. J., Taitzoglou, I., Brellou, G. & Saratsis, P. (2005). The effects of inoculation of *Mannheimia haemolytica* into the teat of lactating ewes. *Veterinary Research*, **36** (1), 13–25.
- McGall, G., Labadie, J., Brock, P., Wallraff, G., Nguyen, T. & Hinsberg, W. (1996). Light-directed synthesis of high-density oligonucleotide arrays using semiconductor photoresists. *Proceedings of the National Academy of Sciences of the United States of America*, **93** (24), 13555–13560.
- McInerney, J., Howe, K. & Schepers, J. (1990). A framework for the economical analysis of disease in farm livestock. Report of a research project (Ref CSA 873).
- McVean, G. (2007). The structure of linkage disequilibrium around a selective sweep. *Genetics*, **175** (3), 1395–1406.
- Menzies, F. D., Bryson, D. G., McCallion, T. & Matthews, D. I. (1995). A study of mortality among suckler and dairy cows in Northern Ireland in 1992. *Veterinary Record*, **137** (21), 531–536.
- Menzies, M. & Ingham, A. (2006). Identification and expression of Toll-like receptors 1-10 in selected bovine and ovine tissues. *Veterinary Immunology and Immunopathology*, **109** (1-2), 23–30.
- Meuwissen, T. H. & Goddard, M. E. (2000). Fine mapping of quantitative trait loci using linkage disequilibria with closely linked marker loci. *Genetics*, **155** (1), 421–430.
- Meuwissen, T. H., Karlsen, A., Lien, S., Olsaker, I. & Goddard, M. E. (2002). Fine mapping of a quantitative trait locus for twinning rate using combined linkage and linkage disequilibrium mapping. *Genetics*, **161** (1), 373–379.
- Miklos, G. L. & Maleszka, R. (2004). Microarray reality checks in the context of a complex disease. *Nature Biotechnology*, **22** (5), 615–621.

- Mikula, M., Gaj, P., Dzwonek, K., Rubel, T., Karczmarski, J., Paziewska, A., Dzwonek, A., Bragoszewski, P., Dadlez, M. & Ostrowski, J. (2010). Comprehensive analysis of the palindromic motif tctcgcgaga: a regulatory element of the hnrnpk promoter. *DNA Research*, **17** (4), 245–260.
- Miller, R. H., Norman, H. D., Wright, J. R. & Cole, J. B. (2009). Impact of genetic merit for milk somatic cell score of sires and maternal grandsires on herd life of their daughters. *Journal of Dairy Science*, **92** (5), 2224–2228.
- Mömke, S., Kuiper, H., Spotter, A., Drogemuller, C. & Distl, O. (2005). A refined radiation hybrid map of the telomeric region of bovine chromosome 18q25-q26 compared with human chromosome 19q13. *Animal Genetics*, **36** (2), 141–145.
- Morris, D. G., Waters, S. M., McCarthy, S. D., Patton, J., Earley, B., Fitzpatrick, R., Murphy, J. J., Diskin, M. G., Kenny, D. A., Brass, A. & Wathes, D. C. (2009). Pleiotropic effects of negative energy balance in the postpartum dairy cow on splenic gene expression: repercussions for innate and adaptive immunity. *Physiological Genomics*, **39** (1), 28–37.
- Morrison, J. L., Breitling, R., Higham, D. J. & Gilbert, D. R. (2005). GeneRank: using search engine technology for the analysis of microarray experiments. *BMC Bioinformatics*, **6**, 233.
- Mrode, R. & Swanson, G. (2003). Estimation of genetic parameters for somatic cell count in the first three lactations using random regression. *Livestock Production Science*, **79** (2-3), 239–247.
- Muncie, S. A., Cassady, J. P. & Ashwell, M. S. (2006). Refinement of quantitative trait loci on bovine chromosome 18 affecting health and reproduction in US Holsteins. *Animal Genetics*, **37** (3), 273–275.
- Murphy, W. J., Larkin, D. M., Everts-van der Wind, A., Bourque, G., Tesler, G., Auvil, L., Beever, J. E., Chowdhary, B. P., Galibert, F., Gatzke, L., Hitte, C., Meyers, S. N., Milan, D., Ostrander, E. A., Pape, G., Parker, H. G., Raudsepp, T., Rogatcheva, M. B., Schook, L. B., Skow, L. C., Welge, M., Womack, J. E., O'Brien, S. J., Pevzner, P. A. & Lewin, H. A. (2005). Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps. *Science*, **309** (5734), 613–617.
- Myllys, V., Honkanen-Buzalski, T., Huovinen, P., Sandholm, M. & Nurmi, E. (1994). Association of changes in the bacterial ecology of bovine mastitis with changes in the use of milking machines and antibacterial drugs. *Acta Veterinaria Scandinavica*, **35** (4), 363–369.
- Nakazato, T., Bono, H., Matsuda, H. & Takagi, T. (2009). Gendoo: functional profiling of gene and disease features using MeSH vocabulary. *Nucleic Acids Research*, **37** (Web Server issue), 166–169.
- Nakazato, T., Takinaka, T., Mizuguchi, H., Matsuda, H., Bono, H. & Asogawa, M. (2008). BioCompass: a novel functional inference tool that utilizes MeSH hierarchy to analyze groups of genes. *In Silico Biology*, **8** (1), 53–61.
- Natarajan, J. & Ganapathy, J. (2007). Functional gene clustering via gene annotation sentences, MeSH and GO keywords from biomedical literature. *Bioinformation*, **2** (5), 185–193.
- Nelson, S. J., Schopen, M., Savage, A. G., Schulman, J. L. & Arluk, N. (2004). The MeSH translation maintenance system: structure, interface design, and implementation. *Studies in Health Technology and Informatics*, **107** (Pt 1), 67–69.
- Nemcova, E., Stipkova, M. & Zavadilova, L. (2011). Genetic parameters for linear type traits in czech holstein cattle. *Czech Journal of Animal Science*, **56** (4), 157–162.



- Nickerson, S. C., Pankey, J. W. & Boddie, N. T. (1984). Distribution, location, and ultrastructure of plasma cells in the uninfected, lactating bovine mammary gland. *Journal of Dairy Research*, **51** (2), 209–217.
- Nilsen, H., Olsen, H. G., Hayes, B., Nome, T., Sehested, E., Svendsen, M., Meuwissen, T. H. & Lien, S. (2009a). Characterization of a QTL region affecting clinical mastitis and protein yield on BTA6. *Animal Genetics*, **40** (5), 701–712.
- Nilsen, H., Olsen, H. G., Hayes, B., Sehested, E., Svendsen, M., Nome, T., Meuwissen, T. & Lien, S. (2009b). Casein haplotypes and their association with milk production traits in Norwegian Red cattle. *Genet Sel Evol*, **41**, 24.
- Nino-Soto, M. I., Heriazon, A., Quinton, M., Miglior, F., Thompson, K. & Mallard, B. A. (2008). Differential gene expression of high and low immune responder Canadian Holstein dairy cows. *Dev Biol (Basel)*, **132**, 315–320.
- Nordfors, L., Jansson, M., Sandberg, G., Lavebratt, C., Sengul, S., Schalling, M. & Arner, P. (2002). Large-scale genotyping of single nucleotide polymorphisms by Pyrosequencing trade mark and validation against the 5' nuclease (Taqman((R))) assay. *Human Mutation*, **19** (4), 395–401.
- Nsengimana, J. & Baret, P. V. (2004). Linkage disequilibrium and the genetic distance in livestock populations: the impact of inbreeding. *Genet Sel Evol*, **36** (3), 281–296.
- Ogorevc, J., Kunej, T., Razpet, A. & Dovc, P. (2009). Database of cattle candidate genes and genetic markers for milk production and mastitis. *Animal Genetics*, **40**, 832–851.
- Ohtsuka, H., Terasawa, S., Watanabe, C., Kohiruimaki, M., Mukai, M., Ando, T., Petrovski, K. R. & Morris, S. (2010). Effect of parity on lymphocytes in peripheral blood and colostrum of healthy Holstein dairy cows. *Canadian Journal of Veterinary Research*, **74** (2), 130–135.
- Ohtsuka, H., Uematsu, M., Saruyama, Y., Ono, M., Kohiruimaki, M., Ando, T. & Kawamura, S. (2009). Age-related alterations in peripheral leukocyte population of healthy Holstein dairy cows during the pre-calving period. *Journal of Veterinary Medical Science*, **71** (8), 1121–1124.
- Olsen, H. G., Lien, S., Gautier, M., Nilsen, H., Roseth, A., Berg, P. R., Sundsaasen, K. K., Svendsen, M. & Meuwissen, T. H. (2005). Mapping of a milk production quantitative trait locus to a 420-kb region on bovine chromosome 6. *Genetics*, **169** (1), 275–283.
- Olsen, H. G., Lien, S., Svendsen, M., Nilsen, H., Roseth, A., Aasland Opsal, M. & Meuwissen, T. H. (2004). Fine mapping of milk production QTL on BTA6 by combined linkage and linkage disequilibrium analysis. *Journal of Dairy Science*, **87** (3), 690–698.
- Oltenacu, P. A. & Algers, B. (2005). Selection for increased production and the welfare of dairy cows: are new breeding goals needed? *Ambio*, **34** (4-5), 311–315.
- Oltenacu, P. A. & Broom, D. M. (2010). The impact of genetic selection for increased milk yield on the welfare of dairy cows. *Animal Welfare*, **19** (Suppl. S), 39–49.
- Oltenacu, P. A. & Ekesbo, I. (1994). Epidemiological study of clinical mastitis in dairy cattle. *Veterinary Research*, **25** (2-3), 208–212.
- Osteras, O., Solbu, H., Refsdal, A. O., Roalkvam, T., Filseth, O. & Minsaas, A. (2007). Results and evaluation of thirty years of health recordings in the norwegian dairy cattle population. *Journal of Dairy Science*, **90** (9), 4483–4497.
- Oviedo-Boyso, J., Valdez-Alarcon, J. J., Cajero-Juarez, M., Ochoa-Zarzosa, A., Lopez-Meza, J. E., Bravo-Patino, A. & Baizabal-Aguirre, V. M. (2007). Innate immune response of bovine mammary gland to pathogenic bacteria responsible for mastitis. *Journal of Infection*, **54** (4), 399–409.

- Paape, M. J., Bannerman, D. D., Zhao, X. & Lee, J. W. (2003). The bovine neutrophil: Structure and function in blood and milk. *Veterinary Research*, **34** (5), 597–627.
- Pant, S. D., Schenkel, F. S., Leyva-Baca, I., Sharma, B. S. & Karrow, N. A. (2008). Identification of polymorphisms in bovine TLR2 and CARD15, associations between CARD15 polymorphisms and milk somatic cell score in Canadian Holsteins, and functional relevance of SNP c.3020A>T. *Dev Biol (Basel)*, **132**, 247–253.
- Pareek, R., Wellnitz, O., van Dorp, R., Burton, J. & Kerr, D. (2005). Immunorelevant gene expression in LPS-challenged bovine mammary epithelial cells. *Journal of Applied Genetics*, **46** (2), 171–177.
- Park, Y. H., Fox, L. K., Hamilton, M. J. & Davis, W. C. (1992). Bovine mononuclear leukocyte subpopulations in peripheral blood and mammary gland secretions during lactation. *Journal of Dairy Science*, **75** (4), 998–1006.
- Patel, D., Almeida, R. A., Dunlap, J. R. & Oliver, S. P. (2009). Bovine lactoferrin serves as a molecular bridge for internalization of *Streptococcus uberis* into bovine mammary epithelial cells. *Veterinary Microbiology*, **137** (3-4), 297–301.
- Patterson, H. & Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika*, **58** (3), 545–554.
- Patterson, H. C., Kraus, M., Kim, Y. M., Ploegh, H. & Rajewsky, K. (2006). The B cell receptor promotes B cell activation and proliferation through a non-ITAM tyrosine in the Igalpha cytoplasmic domain. *Immunity*, **25** (1), 55–65.
- Pawlik, A., Sender, G. & Korwin-Kossakowska, A. (2009). Bovine lactoferrin gene polymorphism and expression in relation to mastitis resistance - a review. *Animal Science Papers and Reports*, **27** (4), 263–271.
- Peng, Q., Malhotra, S., Torchia, J. A., Kerr, W. G., Coggeshall, K. M. & Humphrey, M. B. (2010). TREM2- and DAP12-dependent activation of PI3K requires DAP10 and is inhibited by SHIP1. *Sci Signal*, **3** (122), ra38.
- Pfaffl, M. W., Wittmann, S. L., Meyer, H. H. & Bruckmaier, R. M. (2003). Gene expression of immunologically important factors in blood cells, milk cells, and mammary tissue of cows. *Journal of Dairy Science*, **86** (2), 538–545.
- Philipsson, J. & Lindhe, B. (2003). Experiences of including reproduction and health traits in Scandinavian dairy cattle breeding programmes. *Livestock Production Science*, **83** (2-3), 99 – 112.
- Philipsson, J., Ral, G. & Berglund, B. (1995). Somatic-cell count as a selection criterion for mastitis resistance in dairy-cattle. *Livestock Production Science*, **41** (3), 195–200.
- Piedrahita, J. A., Weaks, R., Petrescu, A., Shrode, T. W., Derr, J. N. & Womack, J. E. (1997). Genetic characterization of the bovine leukaemia inhibitory factor (LIF) gene: isolation and sequencing, chromosome assignment and microsatellite analysis. *Animal Genetics*, **28** (1), 14–20.
- Piepers, S., De Meulemeester, L., de Kruif, A., Opsomer, G., Barkema, H. W. & De Vliegher, S. (2007). Prevalence and distribution of mastitis pathogens in subclinically infected dairy cows in Flanders, Belgium. *Journal of Dairy Research*, **74** (4), 478–483.
- Piepers, S., Opsomer, G., Meyer, E., Demeyere, K., Barkema, H. W., de Kruif, A. & De Vliegher, S. (2009). Heifer and quarter characteristics associated with periparturient blood and milk neutrophil apoptosis in healthy heifers and in heifers with subclinical mastitis. *Journal of Dairy Science*, **92** (9), 4330–4339.
- Pike, K. A. & Ratcliffe, M. J. (2005). Dual requirement for the Ig alpha immunoreceptor tyrosine-based activation motif (ITAM) and a conserved non-Ig alpha ITAM tyrosine in

- supporting Ig alpha beta-mediated B cell development. *Journal of Immunology*, **174** (4), 2012–2020.
- Poso, J. & Mantysaari, E. A. (1996). Relationships between clinical mastitis, somatic cell score, and production for the first three lactations of Finnish Ayrshire. *Journal of Dairy Science*, **79** (7), 1284 – 1291.
- Qanbari, S., Pimentel, E. C., Tetens, J., Thaller, G., Lichtner, P., Sharifi, A. R. & Simianer, H. (2010). The pattern of linkage disequilibrium in German Holstein cattle. *Animal Genetics*, **41** (4), 346–356.
- R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing Vienna, Austria. ISBN 3-900051-07-0.
- Rainard, P. & Riollot, C. (2006). Innate immunity of the bovine mammary gland. *Veterinary Research*, **37** (3), 369–400.
- Rauw, W. (2008). *Ressource Allocation Theory applied to farm animals*. CAB International, Wallingford.
- Rauw, W. M., Kanis, E., Noordhuizen-Stassen, E. N. & Grommers, F. J. (1998). Undesirable side effects of selection for high production efficiency in farm animals: a review. *Livestock Production Science*, **56** (1), 15 – 33.
- Reents, R., Dekkers, J. C. & Schaeffer, L. R. (1995). Genetic evaluation for somatic cell score with a test day model for multiple lactations. *Journal of Dairy Science*, **78** (12), 2858–2870.
- Ren, X. Y., Xu, L., Meng, H. X., Zhao, H. S., Lu, R. F., Chen, Z. B., Feng, X. H., Shi, D., Zhang, L. & Tian, Y. (2009). Family-based association analysis of S100A8 genetic polymorphisms with aggressive periodontitis. *Journal of Periodontal Research*, **44** (2), 184–192.
- Reth, M. (1989). Antigen receptor tail clue. *Nature*, **338** (6214), 383–384.
- Reth, M. & Wienands, J. (1997). Initiation and processing of signals from the B cell antigen receptor. *Annual Review of Immunology*, **15**, 453–479.
- Rice, P., Longden, I. & Bleasby, A. (2000). EMBOSS: the European Molecular Biology Open Software Suite. *Trends in Genetics*, **16** (6), 276–277.
- Rijnkels, M. (2002). Multispecies comparison of the casein gene loci and evolution of casein gene family. *Journal of Mammary Gland Biology and Neoplasia*, **7** (3), 327–345.
- Rinaldi, M., Li, R. W., Bannerman, D. D., Daniels, K. M., Evock-Clover, C., Silva, M. V., Paape, M. J., van Ryssen, B., Burvenich, C. & Capuco, A. V. (2010). A sentinel function for teat tissues in dairy cows: dominant innate immune response elements define early response to *E. coli* mastitis. *Functional & Integrative Genomics*, **10** (1), 21–38.
- Riquet, J., Coppieters, W., Cambisano, N., Arranz, J. J., Berzi, P., Davis, S. K., Grisart, B., Farnir, F., Karim, L., Mni, M., Simon, P., Taylor, J. F., Vanmanshoven, P., Wagenaar, D., Womack, J. E. & Georges, M. (1999). Fine-mapping of quantitative trait loci by identity by descent in outbred populations: application to milk production in dairy cattle. *Proceedings of the National Academy of Sciences of the United States of America*, **96** (16), 9252–9257.
- Rivas, A. L., Quimby, F. W., Blue, J. & Coksaygan, O. (2001). Longitudinal evaluation of bovine mammary gland health status by somatic cell counting, flow cytometry, and cytology. *Journal of Veterinary Diagnostic Investigation*, **13** (5), 399–407.
- Rodriguez-Zas, S. L., Southey, B. R., Heyen, D. W. & Lewin, H. A. (2002). Detection of quantitative trait loci influencing dairy traits using a model for longitudinal data. *Journal*

- of *Dairy Science*, **85** (10), 2681–2691.
- Ron, M., Feldmesser, E., Golik, M., Tager-Cohen, I., Kliger, D., Reiss, V., Domochofsky, R., Alus, O., Seroussi, E., Ezra, E. & Weller, J. I. (2004). A complete genome scan of the Israeli Holstein population for quantitative trait loci by a daughter design. *Journal of Dairy Science*, **87** (2), 476–490.
- Ron, M., Israeli, G., Seroussi, E., Weller, J. I., Gregg, J. P., Shani, M. & Medrano, J. F. (2007). Combining mouse mammary gland gene expression and comparative mapping for the identification of candidate genes for QTL of milk production traits in cattle. *BMC Genomics*, **8**, 183.
- Ron, M. & Weller, J. I. (2007). From QTL to QTN identification in livestock - winning by points rather than knock-out: a review. *Animal Genetics*, **38** (5), 429–439.
- Ronaghi, M., Karamohamed, S., Pettersson, B., Uhlen, M. & Nyren, P. (1996). Real-time DNA sequencing using detection of pyrophosphate release. *Analytical Biochemistry*, **242** (1), 84–89.
- Roosen, S., Exner, K., Paul, S., Schroder, J. M., Kalm, E. & Looft, C. (2004). Bovine beta-defensins: identification and characterization of novel bovine beta-defensin genes and their expression in mammary gland tissue. *Mammalian Genome*, **15** (10), 834–842.
- Rozen, S. & Skaletsky, H. (2000). Primer3 on the WWW for general users and for biologist programmers. *Methods in Molecular Biology*, **132**, 365–386.
- Rupp, R. & Boichard, D. (2003). Genetics of resistance to mastitis in dairy cattle. *Veterinary Research*, **34** (5), 671–688.
- Rupp, R., Hernandez, A. & Mallard, B. A. (2007). Association of bovine leukocyte antigen (BoLA) DRB3.2 with immune response, mastitis, and production and type traits in Canadian Holsteins. *Journal of Dairy Science*, **90** (2), 1029–1038.
- Sahana, G., Lund, M. S., Andersson-Eklund, L., Hastings, N., Fernandez, A., Iso-Touru, T., Thomsen, B., Viitala, S., Sorensen, P., Williams, J. L. & Vilkki, J. (2008). Fine-mapping QTL for mastitis resistance on BTA9 in three Nordic red cattle breeds. *Animal Genetics*, **39** (4), 354–362.
- Sanchez, A. C., Davis, R. L. & Syapin, P. J. (2003). Identification of cis-regulatory regions necessary for robust *Nos2* promoter activity in glial cells: indirect role for NF-kappaB. *Journal of Neurochemistry*, **86** (6), 1379–1390.
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. 1977. *Biotechnology*, **24**, 104–108.
- Schena, M., Shalon, D., Davis, R. W. & Brown, P. O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, **270** (5235), 467–470.
- Schibler, L., Roig, A., Mahe, M. F., Laurent, P., Hayes, H., Rodolphe, F. & Cribiu, E. P. (2006). High-resolution comparative mapping among man, cattle and mouse suggests a role for repeat sequences in mammalian genome evolution. *BMC Genomics*, **7**, 194.
- Schnabel, R. D., Kim, J.-J., Ashwell, M. S., Sonstegard, T. S., van Tassell, C. P., Connor, E. E. & Taylor, J. F. (2005a). Fine-mapping milk production quantitative trait loci on *bta6*: analysis of the bovine osteopontin gene. *Proceedings of the National Academy of Sciences of the United States of America*, **102** (19), 6896–6901.
- Schnabel, R. D., Sonstegard, T. S., Taylor, J. F. & Ashwell, M. S. (2005b). Whole-genome scan to detect QTL for milk production, conformation, fertility and functional traits in two US Holstein families. *Animal Genetics*, **36** (5), 408–416.

- Schrooten, C., Bovenhuis, H., Coppeters, W. & van Arendonk, J. A. (2000). Whole genome scan to detect quantitative trait loci for conformation and functional traits in dairy cattle. *Journal of Dairy Science*, **83** (4), 795–806.
- Schulman, N. F., Sahana, G., Iso-Touru, T., Lund, M. S., Andersson-Eklund, L., Viitala, S. M., Varv, S., Viinalass, H. & Vilkki, J. H. (2009). Fine mapping of quantitative trait loci for mastitis resistance on bovine chromosome 11. *Animal Genetics*, **40** (4), 509–515.
- Schulman, N. F., Viitala, S. M., de Koning, D. J., Virta, J., Maki-Tanila, A. & Vilkki, J. H. (2004). Quantitative trait Loci for health traits in Finnish Ayrshire cattle. *Journal of Dairy Science*, **87** (2), 443–449.
- Schwerin, M., Czernek-Schafer, D., Goldammer, T., Kata, S. R., Womack, J. E., Pareek, R., Pareek, C., Walawski, K. & Brunner, R. M. (2003). Application of disease-associated differentially expressed genes—mining for functional candidate genes for mastitis resistance in cattle. *Genet Sel Evol*, **35 Suppl 1**, S19–34.
- Sen, S., Johannes, F. & Broman, K. W. (2009). Selective genotyping and phenotyping strategies in a complex trait context. *Genetics*, **181** (4), 1613–1626.
- Seroussi, E. (2009). The concordance test emerges as a powerful tool for identifying quantitative trait nucleotides: lessons from BTA6 milk yield QTL. *Animal Genetics*, **40** (2), 230–234.
- Sewalem, A., Miglior, F., Kistemaker, G. J. & van Doormaal, B. J. (2006). Analysis of the relationship between somatic cell score and functional longevity in canadian dairy cattle. *Journal of Dairy Science*, **89** (9), 3609–3614.
- Sharma, B. S., Leyva, I., Schenkel, F. & Karrow, N. A. (2006). Association of toll-like receptor 4 polymorphisms with somatic cell score and lactation persistency in Holstein bulls. *Journal of Dairy Science*, **89** (9), 3626–3635.
- Sheldrake, R. F., Husband, A. J. & Watson, D. L. (1988). Origin of antibody-containing cells in the ovine mammary gland following intraperitoneal and intramammary immunisation. *Research in Veterinary Science*, **45** (2), 156–159.
- Shi, L., Reid, L. H., Jones, W. D., Shippy, R., Warrington, J. A., Baker, S. C., Collins, P. J., de Longueville, F., Kawasaki, E. S., Lee, K. Y., Luo, Y., Sun, Y. A., Willey, J. C., Setterquist, R. A., Fischer, G. M., Tong, W., Dragan, Y. P., Dix, D. J., Frueh, F. W., Goodsaid, F. M., Herman, D., Jensen, R. V., Johnson, C. D., Lobenhofer, E. K., Puri, R. K., Schrf, U., Thierry-Mieg, J., Wang, C., Wilson, M., Wolber, P. K., Zhang, L., Amur, S., Bao, W., Barbacioru, C. C., Lucas, A. B., Bertholet, V., Boysen, C., Bromley, B., Brown, D., Brunner, A., Canales, R., Cao, X. M., Cebula, T. A., Chen, J. J., Cheng, J., Chu, T. M., Chudin, E., Corson, J., Corton, J. C., Croner, L. J., Davies, C., Davison, T. S., Delenstarr, G., Deng, X., Dorris, D., Eklund, A. C., Fan, X. H., Fang, H., Fulmer-Smentek, S., Fuscoe, J. C., Gallagher, K., Ge, W., Guo, L., Guo, X., Hager, J., Haje, P. K., Han, J., Han, T., Harbottle, H. C., Harris, S. C., Hatchwell, E., Hauser, C. A., Hester, S., Hong, H., Hurban, P., Jackson, S. A., Ji, H., Knight, C. R., Kuo, W. P., LeClerc, J. E., Levy, S., Li, Q. Z., Liu, C., Liu, Y., Lombardi, M. J., Ma, Y., Magnuson, S. R., Maqsodi, B., McDaniel, T., Mei, N., Myklebost, O., Ning, B., Novoradovskaya, N., Orr, M. S., Osborn, T. W., Papallo, A., Patterson, T. A., Perkins, R. G., Peters, E. H., Peterson, R., Philips, K. L., Pine, P. S., Pusztai, L., Qian, F., Ren, H., Rosen, M., Rosenzweig, B. A., Samaha, R. R., Schena, M., Schroth, G. P., Shchegrova, S., Smith, D. D., Staedtler, F., Su, Z., Sun, H., Szallasi, Z., Tezak, Z., Thierry-Mieg, D., Thompson, K. L., Tikhonova, I., Turpaz, Y., Vallanat, B., Van, C., Walker, S. J., Wang, S. J., Wang, Y., Wolfinger, R., Wong, A., Wu, J., Xiao, C., Xie, Q., Xu, J., Yang, W., Zhang, L., Zhong, S., Zong, Y. &

- Slikker, Jr, W. (2006). The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nature Biotechnology*, **24** (9), 1151–1161.
- Shook, G. E. (1993). Genetic improvement of mastitis through selection on somatic cell count. *Veterinary Clinics of North America. Food Animal Practice*, **9** (3), 563–581.
- Shpigel, N. Y., Winkler, M., Ziv, G. & Saran, A. (1998). Clinical, bacteriological and epidemiological aspects of clinical mastitis in Israeli dairy herds. *Preventive Veterinary Medicine*, **35** (1), 1–9.
- Siemasko, K., Eisfelder, B. J., Stebbins, C., Kabak, S., Sant, A. J., Song, W. & Clark, M. R. (1999). Ig alpha and Ig beta are required for efficient trafficking to late endosomes and to enhance antigen presentation. *Journal of Immunology*, **162** (11), 6518–6525.
- Siemasko, K., Skaggs, B. J., Kabak, S., Williamson, E., Brown, B. K., Song, W. & Clark, M. R. (2002). Receptor-facilitated antigen presentation requires the recruitment of B cell linker protein to Igalpha. *Journal of Immunology*, **168** (5), 2127–2138.
- Simianer, H. & Pimentel, E. C. (2009). Robust QTL fine mapping by applying a quantitative transmission disequilibrium test to the Mendelian sampling term. *Journal of Animal Breeding and Genetics*, **126** (6), 432–442.
- Skelding, A., Schenkel, F. S., Sharma, B. S., Verschoor, C., Pant, S., Biggar, G., Boermans, H. & Karrow, N. (2010). Identification of single nucleotide polymorphisms in the bovine interleukin-12 and interleukin-23 receptor genes and their associations with health and production traits in Holstein cows. *Journal of Dairy Science*, **93** (10), 4860–4871.
- Smedley, D., Haider, S., Ballester, B., Holland, R., London, D., Thorisson, G. & Kasprzyk, A. (2009). Biomart–biological queries made easy. *BMC Genomics*, **10**, 22.
- Smith, K. & Hogan, J. (1995). Epidemiology of mastitis. In *Proceedings of the third IDF International Mastitis Seminar, Tel-Aviv, Israel* pp. 3–13,.
- Smyth, G. K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology*, **3**, Article3.
- Snelling, W. M., Casas, E., Stone, R. T., Keele, J. W., Harhay, G. P., Bennett, G. L. & Smith, T. P. (2005). Linkage mapping bovine EST-based SNP. *BMC Genomics*, **6** (1), 74.
- Snelling, W. M., Chiu, R., Schein, J. E., Hobbs, M., Abbey, C. A., Adelson, D. L., Aerts, J., Bennett, G. L., Bosdet, I. E., Boussaha, M., Brauning, R., Caetano, A. R., Costa, M. M., Crawford, A. M., Dalrymple, B. P., Eggen, A., Everts-van der Wind, A., Floriot, S., Gautier, M., Gill, C. A., Green, R. D., Holt, R., Jann, O., Jones, S. J., Kappes, S. M., Keele, J. W., de Jong, P. J., Larkin, D. M., Lewin, H. A., McEwan, J. C., McKay, S., Marra, M. A., Mathewson, C. A., Matukumalli, L. K., Moore, S. S., Murdoch, B., Nicholas, F. W., Osoegawa, K., Roy, A., Salih, H., Schibler, L., Schnabel, R. D., Silveri, L., Skow, L. C., Smith, T. P., Sonstegard, T. S., Taylor, J. F., Tellam, R., Van Tassell, C. P., Williams, J. L., Womack, J. E., Wye, N. H., Yang, G. & Zhao, S. (2007). A physical map of the bovine genome. *Genome Biology*, **8** (8), R165.
- Solinas-Toldo, S., Lengauer, C. & Fries, R. (1995). Comparative genome map of human and cattle. *Genomics*, **27** (3), 489–496.
- Sordillo, L. (2005). Factors affecting mammary gland immunity and mastitis susceptibility. *Livestock Production Science*, **98** (1-2), 89–99.
- Sordillo, L. M., Kendall, J. T., Corl, C. M. & Cross, T. H. (2005). Molecular characterization of a saposin-like protein family member isolated from bovine lymphocytes. *Journal of*

- Dairy Science*, **88** (4), 1378–1390.
- Sordillo, L. M., O’Boyle, N., Gandy, J. C., Corl, C. M. & Hamilton, E. (2007). Shifts in thioredoxin reductase activity and oxidant status in mononuclear cells obtained from transition dairy cattle. *Journal of Dairy Science*, **90** (3), 1186–1192.
- Spielman, R. S., McGinnis, R. E. & Ewens, W. J. (1993). Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *American Journal of Human Genetics*, **52** (3), 506–516.
- Stajich, J. E., Block, D., Boulez, K., Brenner, S. E., Chervitz, S. A., Dagdigian, C., Fuellen, G., Gilbert, J. G., Korf, I., Lapp, H., Lehvaslaiho, H., Matsalla, C., Mungall, C. J., Osborne, B. I., Pocock, M. R., Schattner, P., Senger, M., Stein, L. D., Stupka, E., Wilkinson, M. D. & Birney, E. (2002). The Bioperl toolkit: Perl modules for the life sciences. *Genome Research*, **12** (10), 1611–1618.
- Stevenson, M., Guo, Y., Gaizauskas, R. & Martinez, D. (2008). Disambiguation of biomedical text using diverse sources of information. *BMC Bioinformatics*, **9 Suppl 11**, 7.
- Storch, B., Meixlsperger, S. & Jumaa, H. (2007). The Ig-alpha ITAM is required for efficient differentiation but not proliferation of pre-B cells. *European Journal of Immunology*, **37** (1), 252–260.
- Storset, A. K., Slettedal, I. O., Williams, J. L., Law, A. & Dissen, E. (2003). Natural killer cell receptors in cattle: a bovine killer cell immunoglobulin-like receptor multigene family contains members with divergent signaling motifs. *European Journal of Immunology*, **33** (4), 980–990.
- Suchyta, S. P., Sipkovsky, S., Kruska, R., Jeffers, A., McNulty, A., Coussens, M. J., Tempelman, R. J., Halgren, R. G., Saama, P. M., Bauman, D. E., Boisclair, Y. R., Burton, J. L., Collier, R. J., de Peters, E. J., Ferris, T. A., Lucy, M. C., McGuire, M. A., Medrano, J. F., Overton, T. R., Smith, T. P., Smith, G. W., Sonstegard, T. S., Spain, J. N., Spiers, D. E., Yao, J. & Coussens, P. M. (2003). Development and testing of a high-density cDNA microarray resource for cattle. *Physiological Genomics*, **15** (2), 158–164.
- Sugimoto, M., Fujikawa, A., Womack, J. E. & Sugimoto, Y. (2006). Evidence that bovine forebrain embryonic zinc finger-like gene influences immune response associated with mastitis resistance. *Proceedings of the National Academy of Sciences of the United States of America*, **103** (17), 6454–6459.
- Sugimoto, M., Itoh, T., Gotoh, Y., Kawahara, T., Moriya, H., Uchimura, Y. & Sugimoto, Y. (2011). Short communication: Enhanced clinical mastitis resistance in Holsteins with a FEZL p.Gly105(12.13) polymorphism. *Journal of Dairy Science*, **94** (4), 2103–2107.
- Suriyasathaporn, W., Schukken, Y. H., Nielen, M. & Brand, A. (2000). Low somatic cell count: a risk factor for subsequent clinical mastitis in a dairy herd. *Journal of Dairy Science*, **83** (6), 1248–1255.
- Tadepally, H. D., Burger, G. & Aubry, M. (2008). Evolution of C2H2-zinc finger genes and subfamilies in mammals: species-specific duplication and loss of clusters, genes and effector domains. *BMC Evolutionary Biology*, **8**, 176.
- Takaki, R., Watson, S. R. & Lanier, L. L. (2006). DAP12: an adapter protein with dual functionality. *Immunological Reviews*, **214**, 118–129.
- Tarres, J., Guillaume, F. & Fritz, S. (2009). A strategy for QTL fine-mapping using a dense SNP map. *BMC Proceedings*, **3 Suppl 1**, S3.
- Taylor, B. C., Dellinger, J. D., Cullor, J. S. & Stott, J. L. (1994). Bovine milk lymphocytes display the phenotype of memory T cells and are predominantly CD8+. *Cellular Immunology*, **156** (1), 245–253.

- Tedder, T. F., Poe, J. C. & Haas, K. M. (2005). CD22: a multifunctional receptor that regulates B lymphocyte survival and signal transduction. *Advances in Immunology*, **88**, 1–50.
- Terwilliger, J. D. (1995). A powerful likelihood method for the analysis of linkage disequilibrium between trait loci and one or more polymorphic marker loci. *American Journal of Human Genetics*, **56** (3), 777–787.
- Tessarz, A. S. & Cerwenka, A. (2008). The TREM-1/DAP12 pathway. *Immunology Letters*, **116** (2), 111–116.
- Torres, R. M., Flaswinkel, H., Reth, M. & Rajewsky, K. (1996). Aberrant B cell development and immune response in mice with a compromised BCR complex. *Science*, **272** (5269), 1804–1808.
- Toscano, C., Klein, K., Bliedernicht, J., Schaeffeler, E., Saussele, T., Raimundo, S., Eichelbaum, M., Schwab, M. & Zanger, U. M. (2006). Impaired expression of CYP2D6 in intermediate metabolizers carrying the \*41 allele caused by the intronic SNP 2988G<sub>A</sub>: evidence for modulation of splicing events. *Pharmacogenetics and Genomics*, **16** (10), 755–766.
- Uleberg, E. & Meuwissen, T. H. (2007). Fine mapping of multiple QTL using combined linkage and linkage disequilibrium mapping—a comparison of single QTL and multi QTL methods. *Genet Sel Evol*, **39** (3), 285–299.
- Uribe, H. A., Kennedy, B. W., Martin, S. W. & Kelton, D. F. (1995). Genetic-parameters for common health disorders of holstein cows. *Journal of Dairy Science*, **78** (2), 421–430.
- Vaccarelli, G., Miccoli, M. C., Antonacci, R., Pesole, G. & Ciccarese, S. (2008). Genomic organization and recombinational unit duplication-driven evolution of ovine and bovine T cell receptor gamma loci. *BMC Genomics*, **9**, 81.
- van Knegsel, A. T., de Vries Reilingh, G., Meulenberg, S., van den Brand, H., Dijkstra, J., Kemp, B. & Parmentier, H. K. (2007). Natural antibodies related to energy balance in early lactation dairy cows. *Journal of Dairy Science*, **90** (12), 5490–5498.
- Verschoor, C. P., Pant, S. D., Schenkel, F. S., Sharma, B. S. & Karrow, N. A. (2009). SNPs in the bovine IL-10 receptor are associated with somatic cell score in Canadian dairy bulls. *Mammalian Genome*, **20** (7), 447–454.
- Verschoor, C. P., Pant, S. D., You, Q., Kelton, D. F. & Karrow, N. A. (2010). Gene expression profiling of pbmcs from holstein and jersey cows sub-clinically infected with mycobacterium avium ssp. paratuberculosis. *Veterinary Immunology and Immunopathology*, **137** (1-2), 1–11.
- Veyrieras, J. B., Goffinet, B. & Charcosset, A. (2007). MetaQTL: a package of new computational methods for the meta-analysis of QTL mapping experiments. *BMC Bioinformatics*, **8**, 49.
- Visscher, P. M. & Le Hellard, S. (2003). Simple method to analyze SNP-based association studies using DNA pools. *Genetic Epidemiology*, **24** (4), 291–296.
- VIT (2011). Beschreibung der Zuchtwertschätzung für Milchleistungsmerkmale, Zellzahl, Exterieurmerkmale, Nutzungsdauer und Zuchtleistungsmerkmale. Broschüre des VIT Verden.
- Wagter, L. C., Mallard, B. A., Wilkie, B. N., Leslie, K. E., Boettcher, P. J. & Dekkers, J. C. (2000). A quantitative approach to classifying Holstein cows based on antibody responsiveness and its relationship to peripartum mastitis occurrence. *Journal of Dairy Science*, **83** (3), 488–498.
- Wang, C., Xuan, J., Li, H., Wang, Y., Zhan, M., Hoffman, E. P. & Clarke, R. (2010).



- Knowledge-guided gene ranking by coordinative component analysis. *BMC Bioinformatics*, **11** (1), 162.
- Wang, D. & Nettleton, D. (2008). Combining classical trait and microarray data to dissect transcriptional regulation: a case study. *Theoretical and Applied Genetics*, **116** (5), 683–690.
- Wang, K. X. & Denhardt, D. T. (2008). Osteopontin: Role in immune regulation and stress responses. *Cytokine & Growth Factor Reviews*, **19** (5-6), 333–345.
- Watkins, C., McKellar, A., Jensen, K., George, A., Jones, D., Sharp, M. J., Stevenson, K. & Hopkins, J. (2008). Development and validation of an oligonucleotide microarray for immuno-inflammatory genes of ruminants. *Veterinary Research Communications*, **32** (8), 647–657.
- Wayne, M. L. & McIntyre, L. M. (2002). Combining mapping and arraying: An approach to candidate gene identification. *Proceedings of the National Academy of Sciences of the United States of America*, **99** (23), 14903–14906.
- Weber, G., Zawaideh, S., Hikita, S., Kumar, V., Cantor, H. & Ashkar, S. (2002). Phosphorylation-dependent interaction of osteopontin with its receptors regulates macrophage migration and activation. *Journal of Leukocyte Biology*, **72** (4), 752–761.
- Weikard, R., Goldammer, T., Laurent, P., Womack, J. E. & Kühn, C. (2006). A gene-based high-resolution comparative radiation hybrid map as a framework for genome sequence assembly of a bovine chromosome 6 region associated with QTL for growth, body composition, and milk performance traits. *BMC Genomics*, **7**, 53.
- Weller, J. I., Golik, M., Reikhav, S., Domochofsky, R., Seroussi, E. & Ron, M. (2008). Detection and analysis of quantitative trait loci affecting production and secondary traits on chromosome 7 in Israeli Holsteins. *Journal of Dairy Science*, **91** (2), 802–813.
- Weller, J. I., Kashi, Y. & Soller, M. (1990). Power of daughter and granddaughter designs for determining linkage between marker loci and quantitative trait loci in dairy cattle. *Journal of Dairy Science*, **73** (9), 2525–2537.
- Wendt, K., Bostedt, H., Mielke, H. & Fuchs, H.-W. (1994). *Euter und Gesäugekrankheiten*. Gustav Fischer Verlag, Jena.
- Wessner, B., Gryadunov-Masutti, L., Tschan, H., Bachl, N. & Roth, E. (2010). Is there a role for microRNAs in exercise immunology? A synopsis of current literature and future developments. *Exercise Immunology Review*, **16**, 22–39.
- Williams, G. T., Peaker, C. J., Patel, K. J. & Neuberger, M. S. (1994). The alpha/beta sheath and its cytoplasmic tyrosines are required for signaling by the B-cell antigen receptor but not for capping or for serine/threonine-kinase recruitment. *Proceedings of the National Academy of Sciences of the United States of America*, **91** (2), 474–478.
- Wolfova, M., Stipkova, M. & Wolf, J. (2006). Incidence and economics of clinical mastitis in five Holstein herds in the Czech Republic. *Preventive Veterinary Medicine*, **77** (1-2), 48–64.
- Wu, T., Ikezono, T., Angus, C. W. & Shelhamer, J. H. (1994). Characterization of the promoter for the human 85 kDa cytosolic phospholipase A2 gene. *Nucleic Acids Research*, **22** (23), 5093–5098.
- Xu, H., Fan, J. W., Hripcsak, G., Mendonca, E. A., Markatou, M. & Friedman, C. (2007). Gene symbol disambiguation using knowledge-based profiles. *Bioinformatics*, **23** (8), 1015–1022.
- Xu, N. Y., Paul, S., Bennewitz, J., Reinsch, N., Thaller, G., Reinhard, F., Kühn, C., Schwerin, M., Erhardt, G., Weimann, C., Thomsen, H., Mishra, S. & Kalm, E. (2006). Confirmation

- of quantitative trait loci for somatic cell score on bovine chromosome 18 in the German Holstein. *Archiv für Tierzucht*, **49** (2), 111 – 119.
- Yang, Z., Sugawara, M., Ponath, P. D., Wessendorf, L., Banerji, J., Li, Y. & Strominger, J. L. (1990). Interferon gamma response region in the promoter of the human DPA gene. *Proceedings of the National Academy of Sciences of the United States of America*, **87** (23), 9226–9230.
- Yauk, C. L. & Berndt, M. L. (2007). Review of the literature examining the correlation among DNA microarray technologies. *Environmental and Molecular Mutagenesis*, **48** (5), 380–394.
- Yauk, C. L., Rowan-Carroll, A., Stead, J. D. & Williams, A. (2010). Cross-platform analysis of global microrna expression technologies. *BMC Genomics*, **11**, 330.
- Yoshida, T., Mukoyama, H., Furuta, H., Holmes, C. W., Kosugiyama, M. & Tomogane, H. (2008). Allelic frequency of PCR-RFLP type of the BoLA-DRB3 in Japanese Holstein herds and the relation to mastitis. *Animal Science Journal*, **79** (4), 409–416.
- Youn, H. Y., Cho, K. W., Nishimura, Y., Goitsuka, R., Watari, T., Tsujimoto, H. & Hasegawa, A. (1997). Genomic structure of the bovine mb-1 gene encoding the Ig-alpha subunit of the B cell antigen receptor complex. *Veterinary Immunology and Immunopathology*, **56** (3-4), 247–257.
- Youn, H. Y., Goitsuka, R., Kato, H., Mason, D. Y., Watari, T., Tsujimoto, H. & Hasegawa, A. (1996). Molecular cloning of bovine mb-1 cDNA. *Veterinary Immunology and Immunopathology*, **52** (3), 191–200.
- Zahringer, U., Lindner, B., Inamura, S., Heine, H. & Alexander, C. (2008). TLR2 - promiscuous or specific? A critical re-evaluation of a receptor expressing apparent broad specificity. *Immunobiology*, **213** (3-4), 205–224.
- Zecchinon, L., Fett, T. & Desmecht, D. (2005). How *Mannheimia haemolytica* defeats host defence through a kiss of death mechanism. *Veterinary Research*, **36** (2), 133–156.
- Zhang, Q., Boichard, D., Hoeschele, I., Ernst, C., Eggen, A., Murkve, B., Pfister-Genskow, M., Witte, L. A., Grignola, F. E., Uimari, P., Thaller, G. & Bishop, M. D. (1998). Mapping quantitative trait loci for milk production and health of dairy cattle in a large outbred pedigree. *Genetics*, **149** (4), 1959–1973.
- Zhou, Z., Poe, A. C., Limor, J., Grady, K. K., Goldman, I., McCollum, A. M., Escalante, A. A., Barnwell, J. W. & Udhayakumar, V. (2006). Pyrosequencing, a high-throughput method for detecting single nucleotide polymorphisms in the dihydrofolate reductase and dihydropteroate synthetase genes of *Plasmodium falciparum*. *Journal of Clinical Microbiology*, **44** (11), 3900–3910.

## Danksagung

Diese Arbeit wäre ohne die Hilfe einer Reihe von Personen nicht möglich gewesen. Bedanken möchte ich mich besonders bei ...

... **Prof. Dr. Manfred Schwerin** für die Vergabe des Themas, die Begleitung der Arbeit, die hilfreichen Kommentare, Denkanstöße und Diskussionen, sein Gutachten und nicht zuletzt seine Geduld.

... **Frau Prof. Renate Horn** für die Betreuung am Fachbereich sowie viele Verbesserungsvorschläge, konstruktive Kritik und die Begutachtung meiner Arbeit.

... meinem externem Gutachter **Prof. Georg Erhardt**, auch für die persönliche Teilnahme am Kolloquium.

... **Dr. Siriluck Ponsuksili** für die reibungslose Labororganisation, viele Tipps und speziell die Konzeption der Pyrosequenzierungsexperimente.

... den technischen Mitarbeiterinnen **Annette Jugert, Joana Bittner, Nicole Gentz** und **Kerstin Jahnke** für ihre großartige Arbeit im Labor und ganz besonders für die praktische Durchführung der Pyrosequenzierungsexperimente.

... **Dr. Anja Hartmann** für die Analyse der M.A.S.-Net-Transkriptomdaten, deren Bereitstellung und ihre angenehme Art als Büro- und Laborkollegin.

... **Dr. Christina Walz, Dr. Katrin Komolka, Albrecht Rief** und **Isabell Oestreich**, meinen sehr netten und hilfsbereiten Kollegen in der Arbeitsgruppe Funktionelle Genomanalyse.

... **Dr. Christa Kühn** für die Bereitstellung von Probenmaterial und Daten im Rahmen des Projektes M.A.S.-Net.

... **Dr. Christine Baes** und **Dr. Bodo Brand** für die Arbeit an der Feinkartierung des BTA18-SCS-QTL.

... **Dr. Manfred Mayer** für die Unterstützung bei der statistischen Auswertung.

... ebenfalls allen anderen Kooperationspartnern im Projekt M.A.S.-Net.

... **Dr. Sylvio** und **Peggy Tetzlaff**, deren Unterstützung während fast der gesamten Entstehungszeit der Arbeit gar nicht hoch genug geschätzt werden kann.

... **Dr. Dirk Repsilber**, der mich mit der Statistiksprache R vertraut machte und auf sympathische Weise stets mit Rat und Tat behilflich war.

... **Dr. Mark Möller**, der die Arbeit korrekturgelesen und auf viele Verbesserungsmöglichkeiten hingewiesen hat.

Ein ganz besonderes Dankeschön geht an meine Mentorin **Frau Prof. Edda Siegl**, die im Hintergrund stets für mein wissenschaftliches Fortkommen gewirkt hat. Ihr habe ich sehr viel zu verdanken.

Und natürlich möchte ich mich bei meiner Familie für die dauerhafte, liebevolle Unterstützung bedanken.