

*Untersuchung der differentiellen Genexpression mit DNA-Microarrays bei  
Akuter Lymphoblastischer Leukämie im Vergleich mit normalen B-Lymphozyten*

**Dissertation**

zur Erlangung des akademischen Grades

doctor rerum naturalium (Dr. rer. nat.)

vorgelegt der

Mathematisch-Naturwissenschaftlich-Technischen Fakultät  
(mathematisch-naturwissenschaftlicher Bereich)  
der Martin-Luther-Universität Halle-Wittenberg

von Herrn Dipl.-Biol. Uwe Eduard Hattenhorst  
geboren am 01.07.1966 in Wuppertal

Gutachter:

1. Prof. Dr. rer. nat. Gunter Reuter
2. Prof. Dr. med. Stefan Burdach
3. Prof. Dr. rer. nat. Friedemann Horn

Halle (Saale), 20. Oktober 2003, Tag der Verteidigung

**urn:nbn:de:gbv:3-000005644**

[<http://nbn-resolving.de/urn/resolver.pl?urn=nbn%3Ade%3Agbv%3A3-000005644>]

*La vraie générosité envers l'avenir  
consiste à tout donner au présent.*

*Albert Camus*

<b>1 Einleitung</b>	<b>1</b>
1.1 <i>Die akute lymphoblastische Leukämie</i>	1
1.2 <i>Leukämogenese, die Entstehung von Leukämien</i>	3
1.3 <i>Hämatopoese, die Entstehung der zellulären Bestandteile des Blutes</i>	8
1.4 <i>Funktionelle Genomik</i>	13
1.5 <i>Differentielle Genexpression und funktionelle Genomik der cALL</i>	15
<b>2 Material und Methoden</b>	<b>17</b>
2.1 <i>Chemikalien und Enzyme</i>	17
2.2 <i>Untersuchungsmaterial</i>	18
2.2.1 <i>Übersicht über die verwendeten Proben</i>	18
2.2.2 <i>Patientencharakteristika</i>	19
2.3 <i>Aufarbeitung des Untersuchungsmaterials</i>	20
2.3.1 <i>Isolation von Lymphozyten über Dichtegradientenzentrifugation</i>	20
2.3.2 <i>Gewebedisaggregation und Herstellung von Einzelzellsuspensionen</i>	21
2.4 <i>Kryokonservierung humaner Zellen</i>	22
2.5 <i>Zellzahlbestimmung</i>	22
2.6 <i>Separation von B-Lymphozyten</i>	22
2.7 <i>Durchflusszytometrische Untersuchung des Probenmaterials</i>	23
2.8 <i>Präparation der Nukleinsäuren für die Analyse der Genexpression</i>	24
2.8.1 <i>Präparation von Gesamt-RNA mittels Phenol-Chloroform-Extraktion</i>	24
2.8.2 <i>Aufreinigung der präparierten RNA</i>	24
2.8.3 <i>Optionale Ethanol-Präzipitation von RNA</i>	25
2.8.4 <i>Erst-Strang cDNA-Synthese</i>	25
2.8.5 <i>Zweit-Strang-Synthese</i>	25
2.8.6 <i>Phenol-Chloroform-Extraktion der DNA</i>	26
2.8.7 <i>in vitro-Transkription</i>	26
2.8.8 <i>Aufreinigung der IVT-Produkte</i>	27
2.8.9 <i>Fragmentierung der cDNA</i>	27
2.9 <i>Qualitätskontrolle der präparierten Nukleinsäuren</i>	27
2.9.1 <i>Native Gelelektrophorese</i>	27
2.9.2 <i>Photometrische Konzentrationsbestimmung</i>	28
2.10 <i>Affymetrix Gene Chip Instrument System</i>	28
2.10.1 <i>High-density Oligonucleotide Arrays</i>	28
2.10.2 <i>Eigenschaften der verwendeten Arrays</i>	30
2.10.3 <i>Hybridisierung</i>	31
2.10.4 <i>Waschen und Färben der Microarrays</i>	31
2.10.5 <i>Microarray Scan</i>	32

2.10.6 Waschen und Färben mit Antikörperverstärkung	32
2.10.7 Microarray Scan nach Antikörperverstärkung	32
<b>2.11 Computer-gestützte Analyse der Genexpression</b>	<b>32</b>
2.11.1 Normalisierung	34
2.11.1.1 Normalisierung der EOS-H Genechips	34
2.11.1.2 Lineare Regression der EOS-H Genechips	35
2.11.1.3 Normalisierung der Affymetrix HG-U133A Genechips	36
2.11.2 Methoden zur Auswertung der Genexpressionsdaten	37
2.11.2.1 Identifikation differentiell exprimierter Transkripte	37
2.11.2.2 Identifikation von Klassen und Clustern	39
<b>3 Ergebnisse</b>	<b>43</b>
3.1 <i>Eigenschaften der verwendeten Proben</i>	43
3.1.1 Die Leukämieproben	43
3.1.2 Die Kontrollproben	43
3.2 <i>Qualitätskontrolle der Genechips und Ausschluss von Proben</i>	46
3.2.1 EOS-H Genechip	46
3.2.2 HG-U133A Genechip	47
3.3 <i>Expressionsprofile typischer differenzierungsrelevanter Marker</i>	50
3.4 <i>Identifikation differentiell exprimierter Gene</i>	52
3.4.1 EOS-H Genechip	53
3.4.2 HG-U133A Genechip	54
3.4.3 Klassifikation der Proben mit differentiell exprimierten Markergenen	55
3.5 <i>Genexpressionsprofile ermittelt mit unterschiedlichen Genechips</i>	58
3.5.1 Gegenüberstellung der Ergebnisse von EOS-H und HG-U133A	58
3.5.2 Analyse der auf beiden Genechips untersuchten Proben	60
3.6 <i>Funktionelle Gruppen differentiell exprimierter Gene</i>	62
3.6.1 Clustern der differentiell exprimierten Gene	62
3.6.2 Einteilung der Gene in funktionelle Gruppen	64
3.6.2.1 Entwicklung von B-Lymphozyten	66
3.6.2.2 Immunsystem und Inflammation	70
3.6.2.3 Replikation von DNA, Transkription, Proteinsynthese	74
3.6.2.4 Signaltransduktion und Regulation von Genexpression	76
3.6.2.5 Regulation von Differenzierung und Entwicklung	80
3.6.2.6 Regulation von Zellzyklus und Zellproliferation	83
3.6.2.7 Allgemeiner Metabolismus	84
3.6.2.8 Gene mit bekannten Zusammenhängen zu Krebserkrankungen	86
3.6.2.9 Gene, die sich keiner der Gruppen zuordnen lassen	90
3.6.3 Gene, die mit beiden Chiptypen detektiert wurden	91

3.7	<i>Untergruppen der Leukämien</i>	92
3.8	<i>cALL mit t(12;21)(p13;q22) TEL/AML1-Translokation</i>	97
3.8.1	<i>Identifikation differentiell exprimierter Gene</i>	97
3.8.2	<i>Klassifikation der Proben mit differentiell exprimierten Genen</i>	100
3.8.3	<i>Differentiell exprimierte Gene der TEL/AML1-cALL</i>	102
3.8.4	<i>Unterschiede der Leukämien mit und ohne TEL/AML1-Translokation</i>	105
<b>4</b>	<b>Diskussion</b>	<b>108</b>
4.1	<i>Leukämische Blasten der cALL und gesunde B-Lymphozyten</i>	108
4.2	<i>EOS-H und HG-U133A. Zwei unterschiedliche Genechiptypen</i>	112
4.3	<i>Genexpressionsprofil der cALL im Kindesalter</i>	117
4.4	<i>cALL mit TEL/AML1-Translokation</i>	123
4.5	<i>Vergleich der Ergebnisse mit anderen Genexpressionsstudien</i>	127
4.6	<i>Gene mit besonderem Einfluss auf die Pathologie der cALL</i>	132
4.7	<i>Leukämische Blasten sind Zellen mit hoher Proliferationsrate</i>	136
4.8	<i>Überleben und programmierter Zelltod leukämischer Zellen</i>	140
4.9	<i>Die cALL ist eine Erkrankung von B-Vorläuferzellen</i>	142
4.10	<i>TGF<math>\beta</math>/BMP und WNT, kooperative Signalwege leukämischer Zellen</i>	148
4.11	<i>Proliferation oder Differenzierung?</i>	154
<b>5</b>	<b>Zusammenfassung</b>	<b>159</b>
<b>6</b>	<b>Anhang</b>	<b>161</b>
6.1	<i>Tabellen</i>	161
6.2	<i>Abkürzungsverzeichnis</i>	166
6.3	<i>Anmerkungen zur verwendeten Nomenklatur</i>	168
6.4	<i>Literaturverzeichnis</i>	169
6.5	<i>Publikationsliste</i>	184

# 1 Einleitung

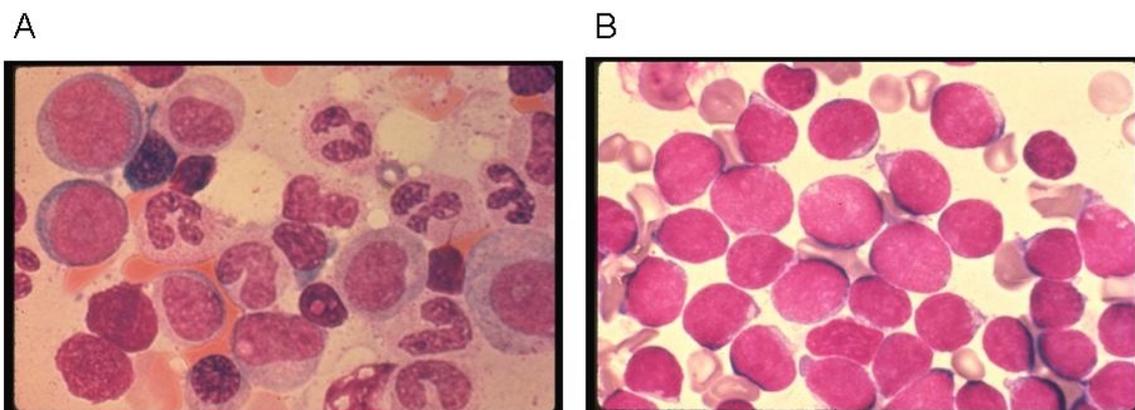
## 1.1 Die akute lymphoblastische Leukämie

Leukämien (abgeleitet von griechisch *leukós* ›weiß‹ und *haĩma* ›Blut‹) sind maligne Erkrankungen der weißen Reihe des hämatopoetischen Systems. Aufgrund des klinischen Verlaufs können akute und chronische Krankheitsformen unterschieden werden. Je nach Abstammung der entarteten Zelle unterteilt man die Leukämien weiter in myeloische und lymphatische Leukämien. Die maligne Transformation hämatopoetischer Zellen ist das gemeinsame Merkmal aller Leukämien. Darüberhinaus lassen sich verschiedene Formen der Leukämie unterscheiden. Bei den lymphatischen Leukämien ist die normale Lymphopoese gestört, das heißt die Entwicklung und Reifung der Lymphozyten des Immunsystems sind krankhaft verändert. Myeloische Leukämien zeichnen sich durch Entartung der Myelopoese aus.

Die akute lymphoblastische Leukämie (ALL) umfasst eine heterogene Gruppe maligner Erkrankungen, die sich nicht nur in immunologisch relevanten Geweben wie dem Knochenmark, dem peripheren Blut oder den lymphatischen Geweben, sondern auch in allen anderen Geweben manifestieren können. Die ALL lässt sich in Abhängigkeit von der Zellreihe, in der die maligne Transformation stattgefunden hat, in T-Zell- und B-Zell-Leukämien weiter unterteilen. Knapp 15% der ALL gehen auf T-Zellblasten und 85% auf B-Zellblasten zurück (Burdach, in press; Kersey, 1997; Schrappe, 2000b). Nach der deutschen Studie „ALL-BFM 90“ (Schrappe, 2000a) der „Deutschen Gesellschaft für pädiatrische Onkologie und Hämatologie (GPOH)“ sind 64,4% der ALL vom Typ common ALL (cALL), 4,9% pro-B ALL, 17,0% prä-B ALL und 13,5% T ALL. Non-Hodgkin-Lymphome sind als lokalisierte Formen der ALL zu betrachten, bei der es zur Ausbildung solider Tumoren kommt. Definitionsgemäß sind hier neben dem lokalen Tumor nicht mehr als 25% Blasten im Knochenmark zu finden. Non-Hodgkin-Lymphome machen ca. 7 % der malignen Erkrankung des Kindes- und Jugendalters aus (Burdach, in press).

Die akute lymphoblastische Leukämie ist die häufigste Krebserkrankung im Kindesalter. Für den Beobachtungszeitraum von 1980 bis 1999 berichtet das Deutsche Kinderkrebsregister (Kinderkrebsregister, 1999) von jährlich rund 1800 neuerfassten Krebserkrankungen bei Kindern unter 15 Jahren. Der Altersmedian für die Erstdiagnose der ALL bei Kindern liegt bei vier Jahren und acht Monaten; das Geschlechterverhältnis liegt bei 1,3:1,0 (Jungen/Mädchen).

Die primäre B-Zell-ALL entsteht im Knochenmark, das normalerweise aus einer Vielzahl verschiedener Zelltypen besteht (vgl. Abbildung 1-1-A), insbesondere den unterschiedlichen Vorläuferzellen der verschiedenen hämatopoetischen Reihen. Die T-Zell-ALL entsteht im Thymus oder in primären lymphatischen Organen. Bei den B-lymphozytären Leukämien wird das Knochenmark von leukämischen Blasten diffus infiltriert und verdrängt die vorhandenen gesunden Zellen. Lichtmikroskopisch auffallend ist die Monotonie des leukämischen Knochenmarks im Ausstrich (vgl. Abbildung 1-1-B), in dem fast nur Leukämiezellen zu erkennen sind. Mit zunehmender Infiltration durch die Blasten ist die normale Hämatopoese gestört, mit der Folge zunehmender Knochenmarksinsuffizienz. Die Geschwindigkeit des natürlichen Krankheitsverlaufs hängt wesentlich von der Proliferationsrate der Blasten ab. Eine weitere den Krankheitsverlauf beeinflussende Eigenschaft der Blasten ist ihre Fähigkeit, das Knochenmark zu verlassen und andere Gewebe zu infiltrieren. Die streuenden leukämischen Zellen können alle Organe und Gewebe des Körpers besiedeln. Sie bevorzugen in der Hauptsache Leber, Milz und Lymphknoten.



**Abbildung 1-1 Knochenmarksausstrich**

Das gesunde Knochenmark auf der linken Seite (A) zeigt eine vielfältige Zellzusammensetzung. Beim leukämischen Knochenmark der cALL auf der rechten Seite (B) wird deutlich, wie die Blasten nahezu alle anderen Zellen bereits verdrängt haben.

Atlas of Hematology, Nagayo University School of Medicine, <http://pathy.med.nagoya-u.ac.jp>

Der Verlauf der Krankheit ist schleichend. So kann der Anamnesezeitraum vom Auftreten klinischer Symptome bis zur endgültigen Diagnosestellung von vier Wochen bis zu einigen Monaten dauern. Die ersten auftretenden klinischen Symptome (Fieber, Abgeschlagenheit, Blässe, Appetitlosigkeit oder Knochenschmerzen) sind eher unspezifisch und lassen an eine Virusinfektion denken. Die proliferierenden Blasten im Knochenmark verdrängen die anderen dort angesiedelten Zellen, und mit fortschreitendem Krankheitsverlauf nimmt die Knochenmarksinsuffizienz zu. Es kommt zur Anämie, zu Hautblutungen aufgrund von Gerinnungsstörungen wegen fehlender

Thrombozyten (Thrombozytopenie) und zu schweren lokalen und generalisierten Infektionen wegen fehlender neutrophiler Granulozyten (Neutropenie). Eine ausgeprägte Vergrößerung von Milz und Leber (Splénomegalie und Hepatomegalie) ist bei 2/3 der betroffenen Kinder zu beobachten, die Lymphknoten sind bei 1/2 der betroffenen Kinder vergrößert.

Für die Behandlung der ALL gibt es Therapieprotokolle, die unter anderem von der deutschen Studie „ALL-BFM“ der GPOH erarbeitet wurden. Die Intensivierung der Therapie und die gleichzeitige Kombination mehrerer Medikamente haben dazu geführt, dass die Heilungschancen der Kinder erheblich verbessert werden konnten. Hinter dem Einsatz verschiedener Medikamente steht die Idee, durch die verschiedenen Wirkmechanismen und Angriffsmodalitäten der Medikamente die Entwicklung chemotherapieresistenter Zellen zu verhindern. Auf eine intensive Induktionstherapie zur Erzielung einer Remission, also dem Rückgang der Krankheitserscheinungen, folgen weitere Therapieblöcke, die schließlich in eine Dauertherapie für einen Zeitraum von zwei bis drei Jahren übergehen. Bei einem Rezidiv oder Wiederaufleben der ALL erfolgt eine erneute Behandlung nach einem gesonderten Rezidivprotokoll. Bei schlechtem Ansprechen auf die Chemotherapie oder einem besonders frühen Rezidiv wird eine allogene Stammzelltransplantation durchgeführt.

Durch die Therapieintensivierung hat sich die Prognose der ALL allgemein in den vergangenen Jahrzehnten erheblich verbessert. Die Drei-Jahres-Überlebenschance erreicht bei den Kindern derzeit 87%. Die Fünf-Jahres-Überlebenschance liegt bei 84%. Für erwachsene Patienten mit gleicher Diagnose liegt diese lediglich bei 24-40% (Kinderkrebsregister, 1999). Eine individuelle Prognose ist von vielen Faktoren abhängig. Es gibt große Bemühungen nationaler und internationaler Studiengruppen, Risikogruppen zu definieren und darauf angepasste Therapieprotokolle zu erstellen. Entscheidenden Einfluss auf die individuelle Prognose haben das Alter des Patienten zum Zeitpunkt der Diagnose, die initiale Zahl der Blasten im Knochenmark, der Nachweis bestimmter genetischer oder chromosomaler Veränderungen, das Ansprechen auf die Therapie und der Zeitraum zwischen Remission und einem Rezidiv.

## *1.2 Leukämogenese, die Entstehung von Leukämien*

Leukämien entstehen aufgrund einer Fehlregulation der klonalen Expansion unreifer lymphoider (ALL) oder myeloischer (akute myeloische Leukämie, AML) Vorläuferzellen, die in einem bestimmten Stadium der Differenzierung arretiert werden, sich aber weiterhin teilen. Die leukämischen Lymphoblasten besitzen keine spezifischen

morphologischen oder zytochemischen Eigenschaften. Somit ist die Diagnose stark von der Immunphänotypisierung abhängig. Hierbei werden mit Hilfe fluoreszenzmarkierter Antikörper bestimmte Antigene der Zelloberfläche nachgewiesen. Die differenzierungsrelevanten Antigene werden systematisch mit einer CD-Nomenklatur (Cluster of Differentiation) beschrieben. Durch Bestimmung des Expressionsmusters differenzierungsrelevanter Antigene können leukämische Zellen entsprechend der normalen Reifungsreihe der B- und T-Lymphozyten klassifiziert werden. Weil solche Zelldifferenzierungsantigene, mit Ausnahme weniger zytoplasmatischer Antigene, keine Spezifität für einzelne Differenzierungsreihen besitzen, wird zur Klassifikation der Zellen ein Panel von Antikörpern verwendet. Die verwendeten Panels lassen in 99% der Fälle eine Diagnose zu (Pui, 1995; Pui, 1998).

Zur Diagnose der ALL werden verschiedene Parameter bestimmt. Hierzu gibt es Empfehlungen der internationalen French-American-British (FAB) Cooperative Group und der deutschen Studiengruppe ALL-BFM (Schrappe, 2000b). Eine ALL wird diagnostiziert, wenn mehr als 25% Lymphoblasten im Knochenmark enthalten sind oder Lymphoblasten im peripheren Blut nachzuweisen sind. Für die immunphänotypische Diagnose und Klassifikation der cALL muss folgendes Expressionsmuster erfüllt sein (+ Antigen nachweisbar, +/- Antigen kann, muss aber nicht nachweisbar sein, - Antigen nicht nachweisbar): CD19+, CD10+, CD20+/-, CD22+, CD24+/-, CD34+/-, CD79a+, TdT+/-, HLA-DR+, zytoplasmatisches IgM+. Ungefähr 1/4 der pädiatrischen und 1/3 der adulten ALL-Patienten zeigen eine Koexpression von Antigenen, die normalerweise auf Zellen der myeloischen Reifungsreihe zu finden sind (Lucio, 2001). Von einer ALL mit myeloischer Koexpression wird gesprochen, wenn mindestens 20% der leukämischen Zellen die Antigene CD13, CD33 und CD67 exprimieren.

Für Leukämien sind Hyperploidien beschrieben worden. Folglich wird bei den Patienten regelmäßig der DNA-Index bestimmt. Hierzu wird der DNA-Gehalt leukämischer Blasten mit dem normaler diploider Lymphozyten verglichen. Ein Index  $\geq 1,6$  wird als hyperploid bezeichnet und ist bei 24,7% der Kinder nachzuweisen (Schrappe, 2000a). Die Inzidenz für die Hyperploidie nimmt beim Heranwachsenden zu. In seltenen Fällen (1%) wird eine Hypoploidie beschrieben (Pui, 1998; van den Berg, 2000). Bei 1,2% der Patienten ist die ALL mit einem Down-Syndrom (Trisomie 21) assoziiert (Pui, 2000).

Mehr als 200 chromosomale Veränderungen im Zusammenhang mit Leukämien sind bereits beschrieben worden, und mindestens 50% der Patienten zeigen definierte molekulargenetische Abnormalitäten. Die durch chromosomale Translokationen eingeführten genetischen Veränderungen tragen mit großer Wahrscheinlichkeit zur Entstehung von Leukämien bei. Durch Translokationen können neue Fusionsgene

entstehen, die hybride Proteine mit veränderter Funktion und Reaktivität kodieren. Ein weiterer pathogener Mechanismus ist die durch die Translokation verursachte Juxtaposition regulativer Elemente, wodurch die Transkription eines Onkogens aktiviert und/oder fehlreguliert werden kann.

**Tabelle 1-1 Chromosomale Aberrationen der ALL**

1: Frequenz für alle drei MLL-Translokationen (MLL-AFX1, MLL-AF4, MLL-ENL) gemeinsam. 2: Frequenz für alle drei MYC-Translokationen (MYC-IgH, MYC-IgK, MYC-IgL) gemeinsam.

Chromosomale Veränderungen	Translokation	Frequenz (%)	
		Kinder	Erwachsene
t(12;21) (p13;q22)	TEL-AML1 (ETV6-CBFA2)	22	2
t(9;22) (q34;q11)	BCR-ABL (Ph+)	4	25
t(1;19) (q23;p13)	E2A-PBX1	5	3
11q23	MLL-AFX1	6 <sup>1</sup>	7 <sup>1</sup>
8q24	MYC-IgH	2 <sup>2</sup>	4 <sup>2</sup>
14q11	TCR $\alpha\delta$	4	6
7q35	TCR $\beta$	3	2
Hyperploidie >50		25	6
Hypoploidie <45		1	4
andere		28	41

Die häufigsten chromosomalen Abweichungen mit ihren Inzidenzen sind in Tabelle 1-1 zusammengestellt (Pui, 1998; van den Berg, 2000). Bei den Neugeborenen-Leukämien machen 11q23 Translokationen, bei denen das MLL-Gen (Mixed Lineage Leukemia, auch Myeloid/Lymphoid Leukemia) betroffen ist, mit 60-70% den größten Anteil aus (Biondi, 2000; Felix, 1999). Bei den älteren Kindern ist mit 20-22% die TEL/AML1-Translokation (Harbott, 1997) und bei den Erwachsenen mit 25% die BCR/ABL-Translokation (Philadelphia-Chromosom) nachzuweisen. Die meisten pädiatrischen Leukämiepatienten tragen Translokationen, bei denen Gene für Transkriptionsfaktoren involviert sind. Translokationen im Bereich von Genen, die Antigenrezeptoren kodieren, sind eher selten. Die US-amerikanische Studiengruppe Pediatric Oncology Group (POG) berichtet weiterhin, dass Aneuploidien und Translokationen bei den akuten B-Zell Leukämien häufiger anzutreffen sind als bei T-Zell Formen (Pullen, 1999). Zur Bestimmung der häufig beobachteten Chromosomentranslokationen werden die lymphoiden Blasten im Rahmen der Diagnose ebenfalls zytogenetisch untersucht.

Die Ursachen für die Leukämieentstehung sind nicht geklärt. Es gibt bis heute verschiedene Erklärungsansätze. Die akuten Leukämien sind als Krankheitsgruppe heterogen. Selbst die cALL kann als eine Gruppe ähnlicher Krankheiten betrachtet werden, wenn man die unterschiedlichen dokumentierten chromosomalen Veränderungen berücksichtigt. Daher ist es schwierig, ein allgemeingültiges Modell der Pathogenese der cALL bzw. aller akuten Leukämien zu formulieren. Dennoch gibt es

viele Befunde, die auf vergleichbare Entstehungsmechanismen von Leukämien hinweisen.

Die Ursachen für die Entstehung der Translokationen sind noch spekulativ. Hier werden Exposition von Mutter und Fötus mit petrochemischen Produkten, Lösungsmitteln, Pestiziden oder Alkohol diskutiert. Auch Topoisomerase 2 Inhibitoren oder Substanzen, die dazu metabolisiert werden können (Antibiotika, Flavonoide, Benzolmetabolite), stehen im Verdacht, chromosomale Translokationen hervorzurufen (Greaves, 1999). Für die Entstehung der Translokationen im Fötus gibt es verschiedene Erklärungsansätze. Die Exposition von Mutter und Ungeborenem mit Topoisomerasehemmern (Anthracycline, Epipodophyllotoxine) könnte einer der Gründe sein (Greaves, 1997). Die Topoisomerasen Typ 1 und Typ 2 sind Enzyme, die in doppelsträngigen DNA-Molekülen Strangbrüche erzeugen und wieder verschließen können. Sie unterstützen mit dieser Funktion bei der normalen Replikation im Zellzyklus das Entwinden der DNA-Stränge. Anthracyclin, der Wirkstoff im Krebsmedikament Doxorubicin, hingegen bewirkt durch die Stabilisierung des intermediären, geöffneten Doppelstrangs einen zur Topoisomerase 2 antagonistischen und in der Folge zytotoxischen Effekt. In hohen Konzentrationen löst der Komplex aus DNA, Topoisomerase und Topoisomerasehemmer Apoptose aus, in geringeren Konzentrationen kann es jedoch zur illegitimen Rekombination zwischen Chromosomenregionen kommen. Hier wird eine direkte Beteiligung bei der Entstehung chromosomaler Translokationen, zum Beispiel an 11q23, und damit an der Entstehung der Leukämie diskutiert (Greaves, 1997).

Einige äußere Faktoren, die entscheidenden Einfluss auf die Entwicklung der Krankheit haben können, sind identifiziert worden. Als wichtigste Risikofaktoren sind zu nennen: ionisierende Strahlen und bestimmte chemische Verbindungen, insbesondere organische Lösungsmittel wie Benzol. Allerdings ist lediglich der Einfluss ionisierender Strahlen auf die Entstehung akuter Leukämien näher untersucht worden (Wiemels, 1999a). Epidemiologische Daten zeigen einen signifikanten Zusammenhang zwischen Alkohol- oder Marihuanakonsum der schwangeren Mutter und einer ALL des Kindes. Ein Zusammenhang zwischen viraler Infektion und Leukämieentstehung wurde ebenfalls untersucht. Die Bedeutung einer Infektion mit dem *Epstein-Barr Virus* (EBV) für die Entstehung von Lymphomen - das sind solide Tumoren aus reifen B-Lymphozyten - ist seit langem bekannt. Als auslösender Mikroorganismus von akuten Leukämien wurde bisher ein Virus identifiziert. In der Folge einer Infektion mit dem *Human T Cell Leukemia Virus Type-1* (HTLV-1) entwickeln einige Patienten eine T-Zell Leukämie. Im Zusammenhang mit den akuten Leukämien der B-Lymphozytenreihe sind bislang keine Viren identifiziert worden.

Für therapieinduzierte oder Sekundärleukämien sind die auslösenden Faktoren zum Teil bekannt. Durch Radio- oder Chemotherapie, die zur Behandlung anderer maligner oder nicht maligner Erkrankungen eingesetzt werden, können akute Leukämien induziert werden. Die verwendeten Therapeutika sind genotoxisch und können DNA-Strangbrüche verursachen. Das bei der Behandlung von Hodgkin-Lymphomen eingesetzte Etoposid steht im Verdacht, bei den Patienten eine akute myeloische Leukämie als Sekundärleukämie auszulösen.

Leukämien und andere Krebserkrankungen gehen mit vielen verschiedenen chromosomalen und genetischen Veränderungen einher. Aufgrund der Tatsache, dass diese Veränderungen bei der ALL auf den leukämischen Klon beschränkt und in anderen Zellen der Patienten nicht nachzuweisen sind, liegt es nahe zu vermuten, diese Veränderungen seien erworben und nicht vererbt. Lediglich 5% der akuten Leukämien (myeloische und lymphatische) weisen vererbte genetische Veränderungen auf, die eine Prädisposition für die Entstehung einer Leukämie darstellen. Betroffen sind hier Gene, die einen direkten Einfluss auf DNA-Stabilität und DNA-Reparatur haben. Es scheint keine vererbten Mutationen in Tumorsuppressorgenen zu geben, die für die Entstehung der Leukämie verantwortlich sein könnten. Ein solcher Mechanismus der Krebsentstehung ist zum Beispiel für das Retinoblastom beschrieben worden. Dieser seltene solide Tumor des Kindesalters nimmt seinen Ursprung in der Netzhaut und ist eine Modellerkrankung für die erbliche Prädisposition für Krebserkrankungen. Das Retinoblastomprotein RB funktioniert als Tumorsuppressor und kontrolliert die Progression des Zellzyklus'. Mit Vererbung einer inaktivierenden Mutation im entsprechenden Gen kann es zur Tumorentstehung kommen (Brantley, 2001).

Die mit den Leukämien einhergehenden Translokationen entstehen vermutlich intrauterin. Fehlende Hinweise auf einen Vererbungsmechanismus und die Untersuchungen an Zwillings- und Drillingsgeburten deuten auf ein pränatales Geschehen hin. Es wurden monozygote Zwillingspaare untersucht, bei denen jeweils ein Zwilling im Alter von zwei Monaten bis 14 Jahren an einer ALL erkrankt war. Das Blut der Zwillinge auf Guthrie-Testkarten wurde retrospektiv untersucht. Die auf den Karten drei bis vier Tage nach der Geburt asservierten Blutstropfen wurden auf vorhandene Translokationen der Leukozyten hin analysiert. In allen Fällen konnten dieselben Translokationen bei beiden Zwillingkindern nachgewiesen werden. Dies legt die Vermutung nahe, dass es in einer Zelle zur Chromosomentranslokation gekommen ist, diese präleukämische Zelle klonal *in utero* expandiert und auf den zweiten Fötus übertragen wurde. Die Konkordanz einer Leukämieerkrankung bei Zwillingen liegt jedoch nur bei 5%, woraus geschlossen wird, dass die Translokation allein nicht leukämieauslösend ist und ein weiteres Ereignis stattfinden muss (Wiemels, 1999a).

Weiterhin wird ein Fall von Drillingskindern berichtet, bei denen die beiden monozygoten Zwillinge dieselbe erworbene, chromosomale TEL/AML1-Translokation trugen und beide an einer ALL erkrankten. Das dritte Kind zeigte keine chromosomalen Veränderungen und erkrankte auch nicht an der Leukämie. Häufig finden in dem nicht durch Translokation umgelagerten TEL-Allel sekundäre Deletionen statt. Die beiden erkrankten Zwillinge hatten bei gleichem Fusionstyp der TEL/AML1-Translokation unterschiedliche TEL-Deletionen. Daraus wurde geschlossen, dass die Deletionen als sekundäre, unabhängige, postnatale Ereignisse zu bewerten sind, und dass für die Entstehung einer ALL mehrere Schritte notwendig sind (Maia, 2001). In einer weiteren Arbeit konnte nachgewiesen werden, dass die Frequenz der TEL/AML1-Translokation in den dort untersuchten Nabelschnurblutproben 100fach höher liegt als die Wahrscheinlichkeit, an einer ALL zu erkranken (Mori, 2002). Die Translokation als singuläres Ereignis reicht also nicht für die Entwicklung der ALL aus.

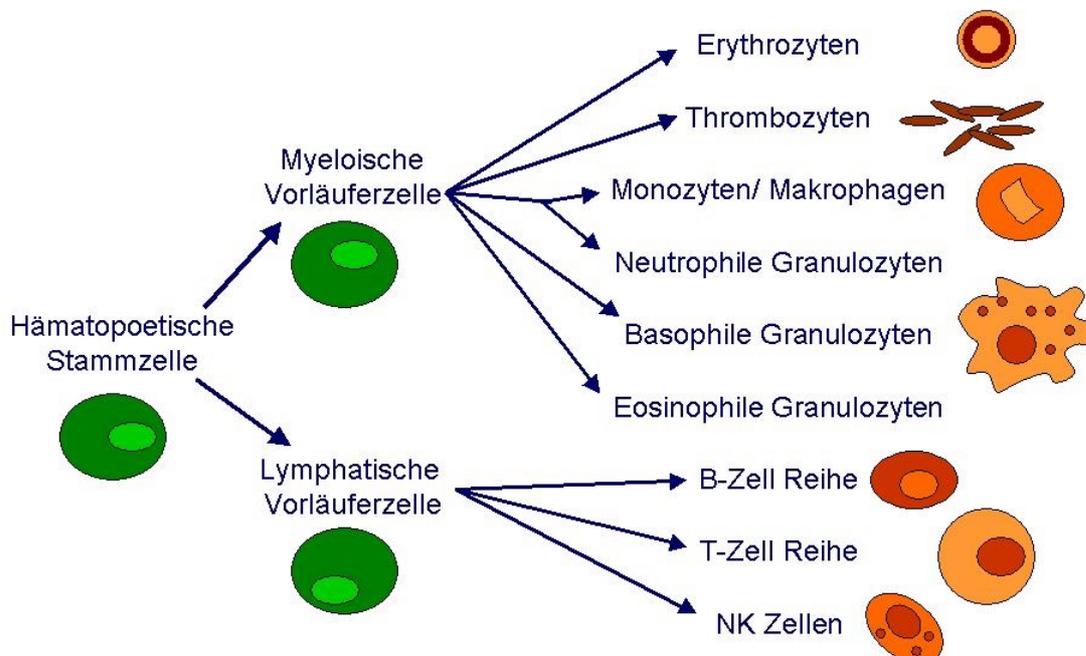
Aufgrund der oben genannten Erkenntnisse wird ein „Multi-Step-Model“ der ALL-Pathogenese mit sequentieller Kombination pränataler und postnataler Ereignisse diskutiert. Nach diesem steht am Anfang ein Ereignis, das, ausgelöst durch Exposition von Mutter und Kind mit einem zytotoxischen Agens, zur chromosomalen Translokation führt. Vermutlich sind dann weitere Umwelteinflüsse für die Krebsentstehung verantwortlich, die die Evolution des leukämischen Klon provozierten. Ein solches zweites Ereignis könnte eine abnormale, verspätete oder inadäquate Immunantwort auf eine gewöhnliche Infektion sein, die einen proliferativen Stress auf die B-Vorläuferzellen ausübt und die präleukämischen Zellen zur unkontrollierten Zellteilung anregt (Greaves, 1999; Greaves, 1997).

### *1.3 Hämatopoese, die Entstehung der zellulären Bestandteile des Blutes*

Die zellulären Bestandteile des Blutes - hierzu gehören die roten Blutkörperchen oder Erythrozyten, die den Sauerstoff transportieren, die Plättchen oder Thrombozyten, die eine zentrale Rolle bei der Blutgerinnung einnehmen, und die weißen Blutkörperchen oder Leukozyten, die Zellen des Immunsystems - entspringen alle derselben hämatopoetischen Stammzelle, die sich im Knochenmark befindet. Diese hämatopoetischen Stammzellen zeichnen sich durch die Fähigkeit aus, sich ständig selbst erneuern zu können. Die hämatopoetische Stammzelle wird als pluripotent bezeichnet, da sich aus ihr sämtliche unterschiedlichen Zellen des Blutes und des Immunsystems entwickeln können.

Das derzeit gültige Modell der Hämatopoese, die beim Menschen vollständig im Knochenmark stattfindet, geht von einer pluripotenten Stammzelle aus, die sich selbst von einer übergeordneten, omnipotenten Stammzelle ableitet. Aus der pluripotenten

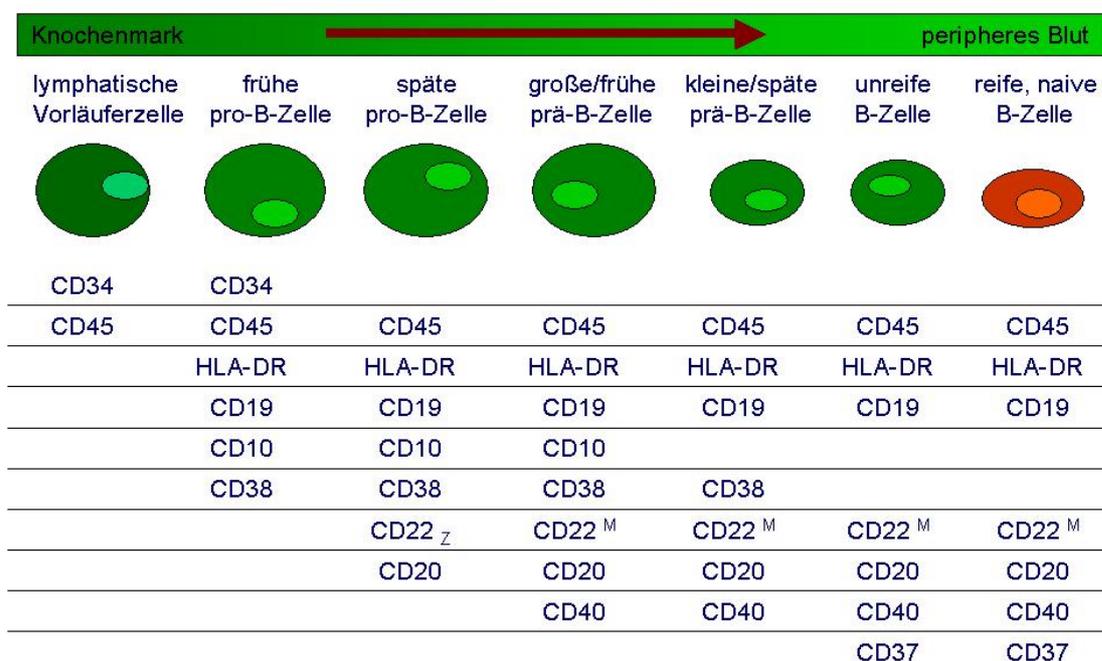
hämatopoetischen Stammzelle (Abbildung 1-2) entwickeln sich zwei verschiedene Typen von Vorläuferzellen, die myeloischen und die lymphoiden Progenitorzellen. Die Progenitorzellen besitzen nur noch eine eingeschränkte Fähigkeit zur Selbsterneuerung und differenzieren nur in bestimmte Zelltypen. Der myeloische Progenitor ist der Vorläufer der Monozyten, die weiter in Makrophagen, Dendritische Zellen und in die unterschiedlichen Granulozyten (Eosinophile, Basophile und Neutrophile) differenzieren. Wahrscheinlich können die myeloischen Progenitoren auch in Retikulozyten, die Vorstufe der Erythrozyten, und in Megakaryozyten, die Vorläufer der Thrombozyten, differenzieren. Aus dem lymphoiden Progenitor entwickeln sich die unterschiedlichen Zellen des adaptiven Immunsystems wie die B- und T-Lymphozyten und die Natürlichen Killerzellen (NK). Die Abbildung 1-2 gibt eine vereinfachte Übersicht der Hämatopoese nach dem heutigen Kenntnisstand wieder. Insbesondere die Entwicklung der Natürlichen Killerzellen sowie die Abstammung der Eosinophilen und der Basophilen Granulozyten sind noch Gegenstand wissenschaftlicher Diskussionen. Bei der Entwicklung vom undifferenzierten Progenitor zur reifen Zelle des Blutes oder des Immunsystems mit festgelegter Funktion durchlaufen die Zellen verschiedene Reifungsstadien mit spezifischen morphologischen und funktionalen Charakteristika, die sich zum Beispiel in stadienspezifischen Expressionsmustern von Oberflächenantigenen feststellen lassen.



**Abbildung 1-2 Hämatopoese**

Aus der pluripotenten hämatopoetischen Stammzelle entwickeln sich über die myeloischen und lymphoiden Progenitorzellen alle Zellen des Blut- und Immunsystems.

Lymphozyten sind die einzigen Zellen des Körpers, die Antigene spezifisch erkennen und unterscheiden können. Sie sind damit verantwortlich für die beiden kennzeichnenden Eigenschaften des adaptiven Immunsystems: Spezifität und Erinnerung. Lymphozyten lassen sich gemäß ihrer Funktion und anhand der Expression charakteristischer Oberflächenmoleküle in Untergruppen einteilen. Morphologisch lassen sie sich nicht unterscheiden. T-Lymphozyten sind die Effektorzellen der spezifischen zellulären Immunantwort. Sie besitzen über die von ihnen sezernierten Zytokine immunmodulatorische Eigenschaften und bewirken die zellvermittelte Zytotoxizität des Immunsystems. B-Lymphozyten produzieren die für die spezifische Immunantwort notwendigen Antikörper und bilden somit die humorale Immunität.

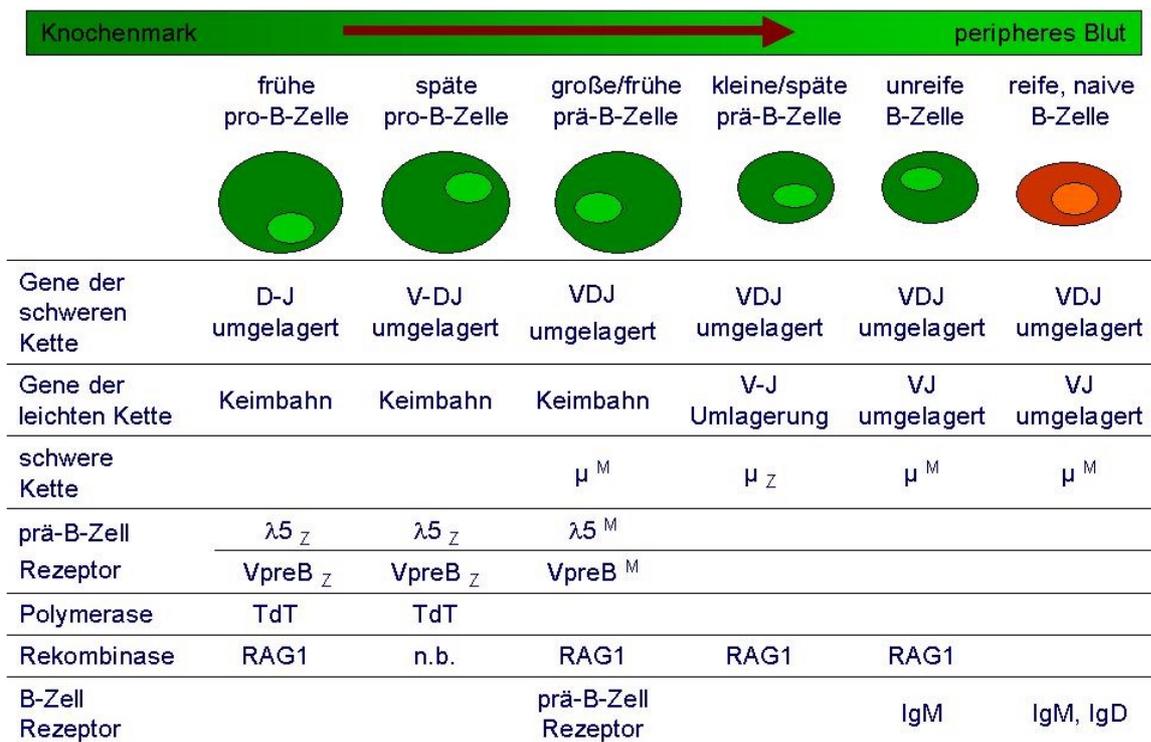


**Abbildung 1-3 Reifungsreihe der B-Lymphozyten**

In Abhängigkeit vom Reifungsstadium exprimieren die B-Vorläuferzellen charakteristische, differenzierungsrelevante Antigene an der Zelloberfläche die sich durchflusszytometrisch nachweisen lassen. Z: Zytoplasma, M: Zellmembran. (Janeway, 2001; Melchers, 1998).

Während der fetalen Entwicklung beginnt die Entwicklung der B-Lymphozyten in Inseln hämatopoetischer Zellen der fetalen Leber. Die Funktion der Blutbildung wird mit zunehmender Entwicklung des Fötus vom Knochenmark übernommen. Nur 2% der Lymphozyten des Körpers befinden sich im Blut. Die restlichen 98% befinden sich in den primären (Knochenmark, Thymus) oder sekundären (Milz, Lymphknoten) lymphatischen Organen. Ausgehend vom lymphoiden Progenitor läuft die Reifungsreihe der B-Lymphozyten beim Menschen vollständig im Knochenmark als primärem lymphatischem Organ ab. Der Progenitor proliferiert und differenziert sich zunächst über die Zwischenstadien der frühen pro-B-Zelle und der späten pro-B-Zelle zur großen prä-

B-Zelle und entwickelt sich dann über die kleine prä-B-Zelle und unreife B-Zelle zur reifen, naiven B-Zelle, die schließlich in das periphere Blut übertritt. Hier zirkulieren die Zellen, besiedeln die sekundären lymphatischen Organe und entwickeln sich im Falle einer Immunantwort zu antikörperproduzierenden Plasmazellen oder zu Gedächtniszellen. Die aufeinanderfolgenden Stadien der B-Zell-Reifung sind unter anderem durch Expression und Umlagerung der Immunglobulingene, unterschiedliche Abhängigkeit von Wachstumsfaktoren sowie durch die Expression verschiedener Oberflächenantigene und intrazellulärer Proteine gekennzeichnet (Janeway, 2001). Die Oberflächenantigene lassen sich durchflusszytometrisch mit Hilfe von Antikörpern nachweisen. Die Abbildung 1-3 zeigt die stadienspezifische Expression der wichtigsten Oberflächenantigene, die zur Identifizierung des Entwicklungsstadiums von B-Lymphozyten herangezogen werden können und die bei der diagnostischen Identifizierung leukämischer Zellen verwendet werden.



**Abbildung 1-4 Entwicklung des B-Zellrezeptors**

Die Gene der variablen Regionen der leichten und schweren Antikörperketten werden während der B-Zellreifung sukzessive umgelagert. Erst mit dem Erscheinen des prä-B-Zellrezeptors an der Oberfläche der prä-B-Zellen beginnt die Umlagerung der variablen Segmente der leichten Ketten. Z: im Zytoplasma nachweisbar, M: auf der Zellmembran nachweisbar, nb: nicht bestimmt. (Janeway, 2001; Melchers, 1998).

Die Entwicklung der B-Lymphozyten über die verschiedenen Stadien der Reifungsreihe ist eng mit der Entstehung der Diversität der B-Zell Rezeptoren und den damit notwendigen Umlagerungen der Immunglobulingene verbunden. Der B-Zell

Rezeptor der reifen B-Lymphozyten setzt sich aus jeweils zwei identischen schweren Ketten und zwei leichten Ketten zusammen. Jedes dieser Immunglobulinmoleküle verfügt über konstante Regionen und über variable Regionen, die Antigene spezifisch binden können. Weiterhin gehören zum B-Zell Rezeptorkomplex jeweils zwei akzessorische Moleküle zur Signaltransduktion,  $Ig\alpha$  (CD79a) und  $Ig\beta$  (CD79b). Zum Erreichen der außerordentlichen Diversität an B-Zell Rezeptoren werden die entsprechenden Immunglobulingene während der B-Zellreifung umgelagert. Dies geschieht jeweils in bestimmten Stufen der verschiedenen B-Vorläuferzellen (Abbildung 1-4) (Melchers, 1998).

Bis zum Stadium der großen prä-B-Zelle sind die V-, D- und J-Gensegmente der schweren Kette umgelagert und das fertige Immunglobulin ist an der Zelloberfläche nachweisbar. Die beiden Gensegmente der leichten Ketten, V und J, bleiben bis dahin im Zustand der ursprünglichen Keimbahn-DNA. An der Oberfläche der großen prä-B-Zelle kommt es zur Expression des intermediären prä-B-Zellrezeptors. An die Stelle der noch nicht gebildeten leichten Ketten tritt die sogenannte Surrogate Light Chain (SLC), die sich aus zwei Bestandteilen zusammensetzt. Diese beiden Proteine  $V_{preB}$  und  $\lambda 5$  sind charakteristisch für die große prä-B-Zelle. Der Ligand dieses intermediären Rezeptorkomplexes ist bislang nicht bekannt. Erst mit dem Übergang zur kleinen prä-B-Zelle werden auch die Gensegmente der leichten Kette umgelagert und der prä-B-Zellrezeptor verschwindet vollständig von der Oberfläche. Übereinstimmend mit diesem Muster der Gensegmentumlagerungen der variablen Regionen sind in der Zelle die dafür erforderlichen Rekombinasen (zum Beispiel RAG1) stadienspezifisch nachweisbar. Der Polymerase TdT, die charakteristisch in B-Vorläuferzellen exprimiert ist, wird ebenfalls eine Rolle bei der Gensegmentumlagerung zugeschrieben und wird als ein Marker zum Nachweis der cALL herangezogen. Mit Beendigung der Gensegmentumlagerungen und Präsentation des fertigen B-Zellrezeptors auf der Oberfläche tritt die unreife B-Zelle schliesslich als reife B-Zelle in das periphere Blut über (Janeway, 2001).

Der Anteil der B-Lymphozyten im Knochenmark ändert sich mit dem Alter. In den ersten vier Lebensjahren beträgt der Anteil der CD19+ B-Zellen mehr als 65% aller Lymphozyten. Mehr als 80% dieser Zellen machen die unreifen CD19/CD10 doppelpositiven B-Vorläuferzellen aus. Mit zunehmendem Alter geht die Zahl der B-Lymphozyten im Knochenmark zugunsten der T-Lymphozyten zurück, bis sich ein ungefähres Gleichgewicht einstellt (33,6% B-Zellen und 34,8% T-Zellen). Unabhängig davon sinkt der Anteil der unreifen CD10+ Zellen, und der Anteil reifer CD20+ B-Zellen steigt an (Rego, 1998).

#### 1.4 Funktionelle Genomik

Der Begriff der Genomik beschreibt ein Forschungsfeld, das sich mit der Entschlüsselung, das heißt Sequenzierung, vollständiger Genome von Organismen beschäftigt. Mit dem Jahr 2001 ist das Genom des Menschen komplett sequenziert worden. Es deutet sich an, dass es weniger als 40.000 Gene enthält. Dennoch ist man heute weit davon entfernt, mit der vollständigen Sequenzierung des Genoms alle Gene des Menschen, deren Position und Organisation im Genom und vor allem deren Funktionen bestimmt zu haben. Als logische Folge hat sich daraus die funktionelle Genomik als weiteres Forschungsgebiet entwickelt. Funktionelle Genomik wird definiert als die Untersuchung aller zu einem Zeitpunkt in einer Zelle exprimierten Gene und die Veränderung dieses Expressionsmusters während Wachstum, Krankheit oder unter wechselnden Umweltbedingungen (Freeman, 2000).

Mit verschiedenen Ansätzen versucht man nun, die Funktionen der Gene und der kodierten Proteine umfassend zu verstehen (Fields, 1999). Die Verwendung von DNA-Microarrays zur Analyse der Expressionsmuster möglichst vieler Gene in einem Experiment ist einer dieser Ansätze. Die funktionelle Genomik versucht, einige fundamentale Fragen zu beantworten: Wie sieht das Zusammenspiel der verschiedenen exprimierten Proteine in den komplexen Stoffwechselfvorgängen auf der Ebene von Zellen, Organen oder des gesamten Organismus aus? Welche Gene werden zu welchen Zeitpunkten und unter welchen Bedingungen exprimiert und an welchen Orten? Und letztlich, wie resultiert diese differentielle Genexpression in der Ausprägung unterschiedlicher Zelltypen oder Organe?

Es gibt verschiedene Methoden, die Genexpression auf mRNA-Ebene zu untersuchen. Diese Methoden sind aber nicht umfassend. Hierzu gehören unter anderem Northern-Blots, RT-PCR, cDNA-Sequenzierung oder cDNA-Fragment Fingerprinting. Eine andere Methode, die bereits zur simultanen Untersuchung der Genexpressionsmuster verschiedener Zellen angewandt wurde, ist die „Serial Analysis of Gene Expression“ (SAGE) (Zhang, 1997). Eine neuere Methode zur Genexpressionsanalyse stellt die Sequenzierung mit Microbeads dar, das „Massively Parallel Signature Sequencing“ (MPSS) (Brenner, 2000). Die beiden zuletzt genannten Methoden haben den Vorteil, dass eine vorhergehende Kenntnis der Gensequenzen nicht nötig ist, wodurch sich die differentielle Genexpression unbekannter Gene aufdecken lässt. Wegen der fehlenden Automatisierungsmöglichkeit und des Aufwandes, große Mengen an Proben einzeln untersuchen zu müssen, sind die Anwendungsmöglichkeiten jedoch limitiert. In den letzten Jahren sind neue Methoden zur vergleichenden Genexpressionsanalyse entwickelt worden. DNA-Microarrays sind eine davon. Das grundlegende Prinzip dieser Microarrays ist die Fixierung von Sonden

(Oligonukleotide oder cDNA) auf festen Oberflächen. Nach Fluoreszenzmarkierung der zu untersuchenden Probe (RNA oder cDNA) wird diese auf den Microarray hybridisiert und die Fluoreszenzintensität an jeder Position auf dem Microarray gemessen. Die einzelnen gemessenen Fluoreszenzintensitäten repräsentieren die Menge der verschiedenen Transkripte.

Die Anwendung von Microarrays geht heute bereits über die funktionelle Genomik hinaus. Neu ist zum Beispiel ein Genechip zur Detektion und Identifikation chromosomaler Translokationen maligner hämatologischer Erkrankungen. Mit Hilfe dieses Arrays ließen sich sowohl der Typ der Translokation als auch die unterschiedlichen Spleißvarianten identifizieren (Nasedkina, 2002). Diese Methode kann die arbeitsintensive und anfällige Multiplex-PCR ersetzen. Eine ungewöhnliche und gänzlich neue Anwendung von Microarrays verbindet die Erkenntnisse der vollständigen Sequenzierung des menschlichen Genoms direkt mit den Fragestellungen der funktionellen Genomik. Ziel dieser Studie (Kapranov, 2002) war die empirisch ermittelte Kartierung derjenigen Abschnitte der Chromosomen 21 und 22, die tatsächlich in der Zelle transkribiert werden. Zu diesem Zweck wurden Genechips hergestellt, die praktisch die beiden Chromosomen komplett abfragen können. Durch die Untersuchung von elf humanen Zelllinien konnten die Autoren feststellen, dass nicht nur die ca. 770 beschriebenen oder mit Computeralgorithmen vorausgesagten Gene transkribiert werden. Vielmehr wird ein erheblich größerer Anteil genomischer Sequenzen in mRNA transkribiert und in das Zytosol transportiert. Dieses „versteckte Transkriptom“ nichtkodierender RNA übernimmt wahrscheinlich eine Vielzahl verschiedener zellulärer Funktionen: beim RNA-Spleißen, bei der Ausbildung ribosomaler oder nukleärer Strukturen, bei der Addition telomerischer Sequenzen oder bei der chromosomalen Inaktivierung von Genen.

Die wichtigste Anwendung für Microarrays liegt derzeit in der funktionellen Genomik. Die Sammlung der Gene, die von genomischer DNA exprimiert oder transkribiert werden, bezeichnet man als das Expressionsprofil oder das Transkriptom der Zelle. Das Transkriptom ist eine der Hauptdeterminanten zellulärer Phänomene oder Funktionen. Die Transkription genomischer DNA in mRNA ist der erste Schritt zur Proteinsynthese. Unterschiede in der Genexpression zeigen sich in morphologischen und phänotypischen Zellunterschieden, spiegeln aber auch zelluläre Reaktionen auf veränderte Umweltbedingungen und Stimuli wider. Im Gegensatz zum Genom ist das Transkriptom hoch dynamisch und verändert sich als Antwort auf Veränderungen sehr rasch. Bei normalen zellulären Ereignissen wie der DNA Replikation oder der Zellteilung ist das Transkriptom ständigen Veränderungen unterworfen (Lockhart, 2000). Wenngleich die Regulation des Proteinvorkommens einer Zelle sicherlich nicht allein

über die Regulation der mRNA gesteuert wird, so sind doch praktisch alle Unterschiede zwischen Zelltypen oder Zellzuständen (z.B. aktivierte Immunzellen, pathologische Zellen) mit Unterschieden in den mRNA Mengen zahlreicher Gene korreliert (DeRisi, 1997).

### *1.5 Differentielle Genexpression und funktionelle Genomik der cALL*

Die vorliegende Arbeit zur funktionellen Genomik der cALL untersucht im direkten Vergleich die Genexpressionsmuster der pathologisch veränderten Leukämiezellen mit gesunden Zellen eines entsprechenden Differenzierungsstadiums, also gesunden B-Vorläuferzellen. Im Vordergrund steht dabei die Frage nach den Veränderungen im Expressionsmuster, die an der Ausprägung des pathologischen Phänotyps beteiligt sind. Im Zusammenhang der Erforschung von Krankheiten, Diagnose- und Therapiemöglichkeiten kann die Genexpressionsanalyse helfen, Ursachen und Konsequenzen pathologischer Veränderungen zu bestimmen. Aus den Daten lassen sich kausale Zusammenhänge zwischen Genexpression und der Biologie der pathologischen Zellen herstellen. Auf molekularer Ebene kann zum Beispiel betrachtet werden, mit welchen Mechanismen die Krebszellen von außen eintreffende Apoptosesignale unterdrücken und ihr eigenes Überleben sichern, welche Signaltransduktionswege in den Zellen vorrangig genutzt werden, welche Genprodukte die große Proliferationsrate der Krebszellen steuern oder welches Differenzierungsstadium die Lymphoblasten exakt besitzen. Neben diesen grundlegenden Fragen zum besseren Verständnis der Vorgänge in maligne entarteten Zellen bieten die Genexpressionsdaten medizinische Anwendungsmöglichkeiten.

Die globale Genexpressionsanalyse kann so zum Beispiel Subtypen der Leukämie definieren, deren Heterogenität auf verschiedene chromosomale Veränderungen oder auf unterschiedliche Differenzierungsstadien der erkrankten Zelltypen zurückzuführen ist. Für diese Subtypen lassen sich dann maßgeschneiderte Therapien entwickeln, die auf die zugrundeliegenden Veränderungen spezifisch eingehen können. Die klinische Anwendung der Genexpressionsdaten kann einer objektiveren Diagnose, Prognose oder Beurteilung des Behandlungserfolges dienen. Es kann auf molekularer Ebene prinzipiell untersucht werden, wie neue Wirkstoffe die Zelle beeinflussen, welche differentiell exprimierten Genprodukte selbst möglicherweise therapeutisch eingesetzt werden können oder sich als Zielstrukturen therapeutischer Intervention eignen (Lockhart, 2000; Sgroi, 1999). So ließen sich zum Beispiel in einer Studie 95 Gene detektieren, die geeignet sind, in Leukämien mit BCR/ABL-Translokation die Sensitivität gegenüber einem Chemotherapeutikum vorherzusagen (Hofmann, 2002). Für Lymphome ist es bereits gelungen, Patientengruppen zu definieren, deren Genexpressionsmuster

mit dem Krankheitsverlauf korreliert (Alizadeh, 2000). Aus Genexpressionsprofilen lassen sich somit Informationen gewinnen, die eine Aussage zum Krankheitsverlauf zulassen.

Bislang veröffentlichte Arbeiten beschäftigen sich vorrangig mit dem Vergleich verschiedener Krankheitsentitäten untereinander, so zum Beispiel mit dem Vergleich der ALL mit AML (Golub, 1999; Yeoh, 2002). Für die hier zu untersuchenden leukämischen Zellen der cALL soll zum ersten Mal das Genexpressionsprofil im Vergleich zu gesunden Zellen mit einem entsprechenden Differenzierungsstadium analysiert werden. Das vorrangige Ziel dabei ist die Aufklärung krankheitsspezifischer Eigenschaften auf molekularer Ebene, die einen grundlegenden Einblick in kausale Zusammenhänge zwischen Genexpression und der Biologie beziehungsweise der Pathologie der Krebszellen zulassen.

## 2 Material und Methoden

### 2.1 Chemikalien und Enzyme

Soweit hier nicht anders gekennzeichnet, wurden alle im Text erwähnten Chemikalien für die Versuche von Sigma Aldrich Chemicals (Steinheim) oder von Merck Eurolab (Darmstadt) bezogen.

2 <sup>nd</sup> -Strand-Buffer	Invitrogen, Karlsruhe
anti-CD10-FITC Antikörper	Becton Dickinson Bioscience, Heidelberg
anti-CD10-PE-Antikörper	Becton Dickinson Bioscience, Heidelberg
anti-CD19-FITC-Antikörper	Becton Dickinson Bioscience, Heidelberg
anti-CD19-PE-Antikörper	Becton Dickinson Bioscience, Heidelberg
anti-CD3-FITC-Antikörper	Becton Dickinson Bioscience, Heidelberg
anti-CD45-PerCP-Antikörper	Becton Dickinson Bioscience, Heidelberg
anti-FITC MB-Antikörper	Miltenyi Biotech, Bergisch Gladbach
Antifoam	Sigma Aldrich Chemicals, Steinheim
B-Cell Isolation Kit	Miltenyi Biotech, Bergisch Gladbach
Biotin-CTP	Roche Diagnostics, Mannheim
Biotin-UTP	Roche Diagnostics, Mannheim
Biotin-anti-Streptavidin-AK	Vector Laboratories, Burlingame, USA
BSA	Invitrogen, Karlsruhe
Chloroform	Carl Roth Chemie, Karlsruhe
Control Oligonucleotide B2	Affymetrix, Santa Clara, USA
DEPC-Wasser	Ambion, Austin, USA
DMSO	Sigma Aldrich Chemicals, Steinheim
DNA-Ligase ( <i>E. coli</i> )	Invitrogen, Karlsruhe
DNA Polymerase I ( <i>E. coli</i> )	Invitrogen, Karlsruhe
dNTP-Mix	Invitrogen, Karlsruhe
Ethanol absolut	Sigma Aldrich Chemicals, Steinheim
Erythrozytenlysepuffer	Universitätsapotheke Halle
Eucaryotic Hybridization Controls	Affymetrix, Santa Clara, USA
Ficoll-Paque	Amersham Pharmacia Biotech, Freiburg
Genechip EOS-H	EOS Biotech, South San Francisco, USA
Genechip HG-U133A	Affymetrix, Santa Clara, USA
Glykogen	Invitrogen, Karlsruhe
Hering Sperm DNA	Promega, Mannheim
Humanes Serum Albumin	Bayer, Wuppertal
Isotypkontrolle g1-FITC/g2-PE	Becton Dickinson Bioscience, Heidelberg
MEGAscript T7 Kit	Ambion, Austin, USA
PBS	Biochrom, Berlin
Phase Lock Gel Tubes	Eppendorf, Hamburg
Phenol:Chloroform:Isoamylalkohol	Sigma Aldrich Chemicals, Steinheim
Phycoerythrin Streptavidin	Molecular Probes, Leiden, Niederlande
RnaseH	Invitrogen, Karlsruhe
RNeasy Mini Kit	Qiagen, Hilden
SSPE (20×)	BioWhittaker, Verviers, Belgien
Superscript Choice System	Invitrogen, Karlsruhe
T7-(dT)24-Primer	Invitrogen, Karlsruhe
TRIzol	Invitrogen, Karlsruhe
Trypanblau-Lösung	Invitrogen, Karlsruhe
Tween 20	Pierce Chemicals, Rockford, USA
Ziegen-IgG	Sigma Aldrich Chemicals, Steinheim

## 2.2 Untersuchungsmaterial

Das Untersuchungsmaterial der Studie war Knochenmark (KM) oder peripheres Blut (PB) pädiatrischer Patienten, die an einer Akuten Lymphoblastischen Leukämie (common Acute Lymphoblastic Leukemia, cALL) erkrankt waren. Es wurden ausschließlich Proben verwendet, die zum Zeitpunkt der Erstmanifestation der Krankheit asserviert und kryokonserviert worden waren. Es wurden keine Proben nach Therapiebeginn oder Proben einer Rezidivkrankung in der Untersuchung eingesetzt. Die Leukämieproben stammen von Patienten der Universitätsklinik und Poliklinik für Kinder- und Jugendmedizin der Martin-Luther-Universität Halle-Wittenberg, der Universitätsklinik und Poliklinik für Kinder und Jugendliche der Universität Leipzig (Prof. Dr. D. Körholz) oder von Patienten der Studiengruppe ALL-BFM der GPOH (Prof. Dr. M. Schrappe).

Die Kontrollzellen der Untersuchung stammen aus verschiedenen Quellen freiwilliger Spender. Es wurde Knochenmark von drei gesunden Erwachsenen untersucht, das eigens zu diesem Zweck entnommen worden war. Das entnommene Knochenmark wurde komplett kryokonserviert und später analysiert. Sortierte B-Lymphozyten wurden aus zwei unterschiedlichen Quellen gewonnen. Zum einen wurden sie aus Gaumenmandeln isoliert, die pädiatrischen Patienten operativ entfernt worden waren. Zum Zeitpunkt der Tonsillektomie lagen keine Anzeichen einer entzündlichen Erkrankung vor. Zum anderen wurden B-Lymphozyten aus Nabelschnurrestblut (cord blood, CB) gewonnen, das nach der Geburt der Kinder in der Universitätsklinik für Geburtshilfe und Reproduktionsmedizin der Martin-Luther-Universität aus der abgetrennten Nabelschnur entnommen werden konnte. Die Proben wurden unmittelbar nach Entnahme aufgearbeitet und die B-Lymphozyten isoliert. Die isolierten Zellen wurden entweder sofort weiterverarbeitet oder bis zur Extraktion der Ribonukleinsäuren in flüssigem Stickstoff gelagert.

Die Entnahme des Patienten- und des gesunden Spendermaterials sowie die damit durchgeführten Untersuchungen sind von der Ethik-Kommission der Martin-Luther-Universität mit Datum vom 25.05.2000 genehmigt worden.

### 2.2.1 Übersicht über die verwendeten Proben

Insgesamt wurde von 56 pädiatrischen Patienten mit der Diagnose cALL Knochenmark aufbereitet. Von 36 dieser Proben konnte ausreichend RNA in genügender Qualität präpariert werden, die anschließend untersucht wurde. Eine Übersicht aller verwendeten Leukämieproben und deren Eigenschaften liefert die Tabelle 6-1. Zehn

der 35 Leukämieproben wurden auf zwei unterschiedlichen Arraytypen (EOS-H und HG-U133A) doppelt untersucht. Eine Patientenprobe wurde nachträglich untersucht und diente zur Kontrolle der Ergebnisse (vgl. 3.4.3), die mit dem Affymetrix Genechip HG-U133A erzielt wurden.

Von drei gesunden, erwachsenen Spendern wurde Knochenmark gesammelt und die RNA für die Genexpressionsanalyse aufbereitet. Eine vorherige Lymphozytenpräparation oder eine andere Zellselektion wurde an diesen Proben nicht durchgeführt. Zur Ergänzung dieser Kontrollgruppe wurde eine vierte Knochenmarkprobe verwendet, die von unserem Kooperationspartner EOS Biotech analysiert worden ist. Die RNA dieser Probe besteht aus einem Pool von 69 erwachsenen Individuen (Alter 22 bis 59 Jahre). Diese Daten sind uns ebenso wie die Expressionsdaten von zwei Proben peripheren Blutes von Erwachsenen freundlicherweise von unserem Kooperationspartner zur Verfügung gestellt worden.

Insgesamt wurden 60 Proben Nabelschnurrestblut gesammelt. Das Blut wurde direkt nach der Geburt aus den Venen der Nabelschnur von gesunden, reifgeborenen Kindern im Kreißsaal entnommen. Von diesen Proben konnten 15 Proben soweit aufbereitet werden, dass sie der Genexpressionsanalyse unterzogen werden konnten. Die verbleibenden Proben lieferten entweder insgesamt zu wenig Zellen, die sortierten Zellen enthielten zu wenig B-Zellen, oder die Menge oder Qualität der präparierten RNA war nicht ausreichend. Das aus der Nabelschnurvene entnommene Blutvolumen der verwendeten Proben reichte von 42ml bis 90ml (Durchschnitt  $63.62 \pm 12.17$  ml) und enthielt nach der Lymphozytenaufreinigung (vgl. 2.3.1) zwischen  $6,9$  und  $20 \times 10^7$  Zellen (Durchschnitt  $11,56 \pm 3,82 \times 10^7$  Zellen). Die Tabelle 6-2 im Anhang fasst die Eigenschaften der in die Analyse eingeflossenen Kontrollproben zusammen. Acht der 15 Proben wurden auf den beiden unterschiedlichen Arraytypen doppelt untersucht.

Als weitere Kontrollgruppe wurden Tonsillen von zwei pädiatrischen Patienten präpariert. Zum Zeitpunkt der chirurgischen Entfernung der Tonsillen lagen keine Anzeichen einer Entzündung vor. Angaben zu diesen Proben finden sich ebenfalls in der Tabelle 6-2 im Anhang. Diese Tabelle gibt eine detaillierte Übersicht aller aus verschiedenen Quellen stammenden sortierten B-Zellen (vgl. hierzu 2.6, Separation von B-Lymphozyten).

### 2.2.2 Patientencharakteristika

Sämtliche Patientenproben stammen von Kindern mit einer akuten lymphoblastischen Leukämie vom Typ cALL. Die Diagnosen wurden jeweils in den behandelnden Krankenhäusern gestellt. Die Patientendaten wurden über die ALL-BFM Studienzentrale der GPOH übermittelt. Die durchflusszytometrische Untersuchung des

Patientenmaterials aus Halle zur Diagnosestellung wurde von Frau Dr. rer. nat. O. Diwan im Interdisziplinären hämatologisch-onkologischen Labor der Martin-Luther-Universität durchgeführt. Das durchschnittliche Alter der Patienten bei Diagnosestellung war sechs Jahre und sechs Monate (Minimum zwei Jahre ein Monat, Maximum 15 Jahre acht Monate). Ausgenommen davon ist die Probe ALL37UEH. Hierbei handelt es sich um die einzige Säuglingsleukämie (Alter zum Zeitpunkt der Diagnose war sechs Monate). Dies ist die einzige Probe, die nicht dem Knochenmark, sondern dem peripheren Blut entnommen wurde. Die Geschlechterverteilung liegt bei 20 weiblichen zu 15 männlichen Patienten. Die Prozentzahl der leukämischen Blasten im Knochenmark der Patienten lag zwischen 40 und 99%, wobei in 32 der 35 Proben über 90% Blasten nachgewiesen werden konnten. Für sieben Patienten konnte eine  $t(12;21)(p13;q22)$  TEL/AML1-Translokation und für einen Patienten eine  $t(9;22)(q34;q11)$  BCR/ABL-Translokation nachgewiesen werden. Andere Translokationen kamen im untersuchten Patientengut nicht vor. Bei allen Patienten ließ sich durchflusszytometrisch die Expression von CD10 an der Zelloberfläche und zytoplasmatisch die Expression der Terminalen Deoxynucleotidyltransferase (TdT), zwei wichtigen Markern der cALL, nachweisen. Aufgrund klinischer Parameter waren elf Patienten einer Standardrisikogruppe, 20 Patienten einer mittleren Risikogruppe und vier Patienten einer Hochrisikogruppe laut ALL-BFM Studienprotokoll zugeordnet worden.

Bei der Probe ALL58GAT handelt es sich um eine nachträglich hinzugefügte Probe, die nicht in die eigentliche Analyse der Genexpression einbezogen wurde. Es handelt sich hierbei um die einzige Probe eines B-Zell-Lymphoms im Rezidivstadium. Die Lymphoblasten wurden aus dem Liquor des Patienten isoliert und exprimieren auf ihrer Oberfläche die Antigene CD10, CD19, CD20 und teilweise CD34. Die leukämischen Blasten der cALL zeigen ein entsprechendes Expressionsmuster, sind aber im Knochenmark nachzuweisen. Der Patient ALL58GAT hatte im Knochenmark keine Lymphoblasten. Diese Probe konnte später als Kontrolle der ermittelten Ergebnisse dienen (vgl. 3.4.3). Eine Übersicht aller verwendeten Leukämieproben und den klinischen Eigenschaften liefert die Tabelle 6-1 im Anhang.

### *2.3 Aufarbeitung des Untersuchungsmaterials*

#### *2.3.1 Isolation von Lymphozyten über Dichtegradientenzentrifugation*

Aus dem humanen Untersuchungsmaterial PB, KM, Zellsuspension oder CB wurden die Lymphozyten mittels einer Dichtegradientenzentrifugation isoliert. Hierzu wurde Ficoll-Paque verwendet. Die Probe wurde im Verhältnis 1:1 mit PBS verdünnt. In ein Zentrifugenröhrchen wurde, je nach Menge der aufzuarbeitenden Probe, 3ml (10ml)

Ficoll-Paque vorgelegt und mit 4ml (15ml) verdünnter Probe vorsichtig überschichtet, ohne die beiden Phasen miteinander zu vermischen. Die Proben wurden für 30min bei Raumtemperatur (RT) mit einer Beschleunigung von 400×g zentrifugiert. Nach Beendigung wurde die Zentrifuge ohne zu bremsen auslaufen gelassen. Die oberste Schicht, sie enthält Plasma und Thrombozyten, wurde abgenommen und verworfen. Die Interphase, sie enthält die Lymphozyten, wurde abgenommen und weiterbearbeitet. Die unterste Schicht, sie enthält in der Hauptsache Granulozyten und pelletierte Erythrozyten, wurde verworfen. Die Lymphozytensuspension wurde mit mindestens drei Volumenteilen PBS aufgefüllt, vorsichtig durchmischt und für 10min bei 300×g zentrifugiert. Der Überstand wurde verworfen, das Zellpellet in Erythrozytenlysepuffer (Ammoniumchlorid 8,29g, Kaliumhydrogencarbonat 1,00g, EDTA-Natriumsalz 0,0372g, ad 1000ml H<sub>2</sub>O, pH 7,3) resuspendiert und für 15min auf Eis inkubiert, um verbliebene Erythrozyten zu lysieren. Die Suspension wurde mit PBS aufgefüllt, für 10min bei 300×g zentrifugiert, der Überstand verworfen, das Zellpellet in PBS resuspendiert und erneut für 10min bei 300×g zentrifugiert. Der Überstand wurde verworfen und das Zellpellet in 10ml PBS resuspendiert. Die Zellzahl wurde mit Hilfe einer Neubauer-Zählkammer im Lichtmikroskop bestimmt (vgl. 2.5).

### 2.3.2 Gewebedisaggregation und Herstellung von Einzelzellsuspensionen

Um aus den entnommenen Gaumenmandeln B-Lymphozyten isolieren zu können, wurden durch mechanische Disaggregation des Gewebes direkt nach der Entnahme Einzelzellsuspensionen hergestellt. Hierfür wurde das automatisierte Medimachine System (DAKO Diagnostics, Hamburg) verwendet. Dieses System besteht aus dem eigentlichen Gerät, den sterilen Einweg-Disaggregationskammern Medicons (Porengröße 50µm zur Vereinzelnung von Zellen bis ca. 230µm Größe) und den sterilen Filtereinheiten Filcons (Porengröße ebenfalls 50µm). Die Zerkleinerung des Gewebes erfolgt mechanisch, in dem es durch das Metallsieb des Medicons gedrückt wird.

Die Gaumenmandeln wurden von störendem Gewebe befreit, in Stücke von ca. 10-20mm<sup>3</sup> geschnitten und in PBS mit 2mM EDTA aufbewahrt. Das Medicon wurde mit 2ml Medium gespült, danach mit 2ml der Gewebesuspension befüllt und das Gerät für 15sec laufen gelassen. Die Suspension wurde entnommen, über ein Filcon filtriert und gesammelt. Dieser Vorgang wurde bis zur völligen Aufarbeitung der Gaumenmandel wiederholt.

Nach einer Zentrifugation der Zellsuspension für 10min bei 250×g bei RT wurde die Zellsuspension einer Dichtegradientenzentrifugation unterzogen (vgl. 2.3.1).

#### 2.4 Kryokonservierung humaner Zellen

Humane Zellen lassen sich in flüssigem Stickstoff bei  $-195^{\circ}\text{C}$  dauerhaft lagern.  $1 \times 10^6 - 1 \times 10^7$  Zellen wurden in 1ml Einfriermedium (Humanes Serum Albumin (Stammlösung 5%) + 10% DMSO (Endkonzentration)) in Kryogefäße überführt. In einem speziellen Container, der mit Isopropanol gefüllt ist, wurden die befüllten Kryogefäße in einer Kühltruhe langsam mit einer Abkühlungsrate von  $1^{\circ}\text{C}/\text{min}$  auf  $-70^{\circ}\text{C}$  abgekühlt, bevor sie im flüssigen Stickstoff dauerhaft eingelagert werden konnten.

#### 2.5 Zellzahlbestimmung

10 $\mu\text{l}$  einer Zellsuspension wurden mit 90 $\mu\text{l}$  einer Trypanblau-Lösung vermischt, für drei Minuten inkubiert und die Zahl lebender Zellen mit Hilfe einer Neubauer-Zählkammer im Lichtmikroskop bestimmt.

#### 2.6 Separation von B-Lymphozyten

Aus Lymphozytensuspensionen, die aus Nabelschnurrestblut oder aus Gaumenmandeln stammten, wurden B-Lymphozyten unter Verwendung des B-Cell Isolation Kits (Miltenyi Biotech, Bergisch Gladbach) isoliert. Das Trennprinzip beruht auf einer Antikörpermarkierung der Zellen, wobei paramagnetische Partikel (Microbeads, MB) an die Antikörper gebunden sind (Miltenyi, 1990). Hierfür gibt es direktmarkierte Antikörper oder eine Kombination aus Hapten-konjugierten Antikörpern in Verbindung mit markierten anti-Hapten Zweitantikörpern. Bei der Passage durch ein starkes Magnetfeld werden die so markierten Zellen zurückgehalten und nur unmarkierte Zellen können passieren. Dieses als Magnetic Activating Cell Sorting (MACS) bezeichnete Verfahren ermöglicht die Isolation einzelner antikörperspezifischer Zellen. Das verwendete B-Cell Isolation Kit enthält Antikörper gegen die Antigene CD2, CD4, CD11b, CD16, CD36 und IgE-Antikörper, wodurch alle Nicht-B-Zellen depletiert werden. Die B-Zellen bleiben unmarkiert und können gesammelt werden (Bauer, 1999; Kim, 1999).

Die Zellkonzentration wurde mit MACS-Puffer (PBS, 2mM EDTA, 0,5% HSA, entgast) auf  $10^7/60\mu\text{l}$  Zellen eingestellt. Pro  $10^7$  Zellen wurden 20 $\mu\text{l}$  des FcR Blockierungs-Reagenzes und 20 $\mu\text{l}$  des Gemisches Hapten-konjugierter Antikörper zugesetzt, gemischt und für 15min bei  $12^{\circ}\text{C}$  inkubiert. Die Zellen wurden mit dem 20fachen Volumen des Inkubationsansatzes mit MACS-Puffer gewaschen und für 10min bei  $300 \times g$  bei  $12^{\circ}\text{C}$  zentrifugiert. Dieser Waschschrift wurde einmal wiederholt.

Die Zellen wurden in 80 $\mu\text{l}$  MACS-Puffer pro  $10^7$  Zellen resuspendiert und mit 20 $\mu\text{l}$  MACS anti-Hapten Microbeads pro  $10^7$  Zellen magnetisch markiert. Die Inkubation wurde für 15min bei  $12^{\circ}\text{C}$  durchgeführt. Die Zellen wurden mit dem 20fachen Volumen

des Inkubationsansatzes mit MACS-Puffer gewaschen und für 10min bei  $300\times g$  bei  $12^{\circ}\text{C}$  zentrifugiert, anschließend in  $500\mu\text{l}$  MACS-Puffer pro  $10^8$  Zellen resuspendiert und auf die vorbereiteten Trennsäulen gegeben.

Die unmarkierten Zellen wurden als B-Zellfraktion in insgesamt 12ml MACS-Puffer gesammelt. Die markierten Zellen wurden nach Entfernen des Magnetfeldes als Nicht-B-Zellfraktion ebenfalls in 12ml MACS-Puffer gesammelt. Zur Qualitätskontrolle wurden anschließend eine Zellzählung (vgl. 2.5) und eine FACS-Analyse (vgl. 2.7) durchgeführt. Von den so isolierten B-Zellen wurde Gesamt-RNA nach 2.8 präpariert. Eine Übersicht zu allen sortierten B-Zellen liefert die Tabelle 6-2 im Anhang.

Die Proben CB\*1NP, CB\*2NP, TS01NP und TS02NP wurden mit einer anderen Strategie sortiert. Die ursprüngliche Idee war die direkte Markierung der Zielzellen und damit eine Positivselektion aus dem Zellgemisch. Dazu wurden die Zellen zuerst mit einem anti-CD10-FITC Antikörper und dann mit einem magnetisch markierten anti-FITC MB-Antikörper inkubiert und im Magnetfeld separiert. Zur Entfernung der verbliebenen CD10-positiven Granulozyten wurden die Zellen mit einem magnetisch markierten anti-CD19 Antikörper inkubiert, der nur B-Lymphozyten markiert, und ein zweites Mal im Magnetfeld separiert. Als Positivfraktion erhält man CD10/CD19 doppelpositive B-Zellen. Da die Ergebnisse der hier getesteten Sortierstrategie insgesamt nicht befriedigend waren, insbesondere war die Zellausbeute zu gering, wurden in der Folge alle weiteren Kontrollen nach der oben geschilderten Methode der Depletion der Nicht-B-Zellen sortiert. Man muss davon ausgehen, dass die in den Proben CB\*1NP und CB\*2NP enthaltenen Zellpopulationen nicht mit denen aus den übrigen Nabelschnurblutkontrollen gewonnenen identisch sind. Daher werden diese Probe bei der Analyse der Daten als eigenständige Kontrollgruppen betrachtet.

### *2.7 Durchflusszytometrische Untersuchung des Probenmaterials*

Die aus den verschiedenen Quellen stammenden sortierten B-Lymphozyten-Präparationen wurden vor und nach der Separation auf ihre zelluläre Zusammensetzung hin untersucht. Für die durchflusszytometrische Untersuchung (Fluorescence Activated Cell Analysis, FACS) von Oberflächenmolekülen der Zellen wurden diese mit spezifischen, monoklonalen Maus-gegen-Mensch Antikörpern markiert. Es wurden nur direktmarkierte Antikörper (AK) verwendet.

Pro Färbung wurden  $100\mu\text{l}$  Vollblut, KM oder  $1,0\text{-}1,5\times 10^6$  isolierte Zellen in  $100\mu\text{l}$  PBS mit  $10\mu\text{l}$  Antikörperlösung für 15min bei RT im Dunkeln inkubiert, anschließend mit  $2\text{ml}$  PBS gewaschen und zentrifugiert. Nach einem weiteren Waschschrift wurden die Zellen in  $200\mu\text{l}$  PBS aufgenommen und im Durchflusszytometer analysiert.

Die verwendeten Proben wurden auf ihren Gehalt an B-Lymphozyten hin untersucht. Hierzu wurde ein standardisierter Färbeansatz mit den folgenden Antikörpern verwendet:

- |                    |                                |                 |
|--------------------|--------------------------------|-----------------|
| 1: Isotypkontrolle | $\gamma$ 1-FITC/ $\gamma$ 2-PE |                 |
| 2: anti-CD19-FITC  | anti-CD10-PE                   |                 |
| 3: anti-CD3-FITC   | anti-CD19-PE                   | anti-CD45-PerCP |

Es wurden 10.000 Zellen je Färbeansatz analysiert. Mit Hilfe eines FACScan (Becton Dickinson Bioscience, Heidelberg) wurden die Proben durchflusszytometrisch untersucht. Zur Auswertung der gemessenen Daten standen die „FACScan“- und die „LYSIS II“-Software zur Verfügung.

## 2.8 Präparation der Nukleinsäuren für die Analyse der Genexpression

### 2.8.1 Präparation von Gesamt-RNA mittels Phenol-Chloroform-Extraktion

Für die Analyse wurde ausschließlich Gesamt-RNA ohne Aufreinigung von poly(A)<sup>+</sup>RNA präpariert. Die Selektion der mRNA zur Analyse erfolgte unter Verwendung eines T7-(dT)<sub>24</sub>-Primers in der Erst-Strang-Synthese (vgl. 2.8.4). Die Zellen wurden in 1ml TRIzol, einem gebrauchsfertigen Phenol-Isothiozyanat-Gemisch, pro 50-100mg Gewebe oder 5-10×10<sup>6</sup> Zellen durch mehrmaliges Auf- und Abziehen der Suspension durch eine Kanüle homogenisiert und die Suspension anschließend bei 12.000×g für 10min bei 4°C in 1,5ml Gefäßen zentrifugiert. Der Überstand wurde ohne Zellklumpen in ein neues Reaktionsgefäß überführt und für 5min bei RT inkubiert. Nach Zugabe von 0,2ml Chloroform pro ml eingesetztem TRIzol wurde die Probe für 15sec kräftig geschüttelt und bei RT für 3min inkubiert, anschließend bei 12.000×g für 15min bei 4°C zentrifugiert. Die obere, klare, wässrige Phase wurde in ein neues Reaktionsgefäß überführt. Es wurden 0,5ml Isopropanol pro ml eingesetztem TRIzol zur wässrigen Phase hinzugegeben, leicht gemischt und bei RT für 10min inkubiert, anschließend bei 12.000×g für 10min bei 4°C zentrifugiert. Der Überstand wurde verworfen, das Pellet wurde in 1ml kaltem 75-80% Ethanol pro ml eingesetztem TRIzol gewaschen und anschließend bei 7.500×g für 5min bei 4°C zentrifugiert. Der Überstand wurde verworfen, das Pellet getrocknet und in 50µl DEPC-Wasser resuspendiert. Die präparierte RNA wurde einer Qualitätskontrolle unterzogen (vgl. 2.9).

### 2.8.2 Aufreinigung der präparierten RNA

Zur Aufreinigung der nach 2.8.1 präparierten Gesamt-RNA wurde das RNeasy Mini Kit verwendet. Das Volumen der Probe wurde mit Rnase-freiem Wasser auf 100µl eingestellt und die Probe nach Vorschrift des Herstellers aufgearbeitet. Auf den Zusatz

von  $\beta$ -Mercaptoethanol wurde verzichtet, da die mit TRIzol gereinigte RNA bereits kaum noch Protein enthält. In Abweichung vom Herstellerprotokoll wurde die RNA in zwei aufeinanderfolgenden Schritten mit jeweils 25 $\mu$ l DEPC-Wasser eluiert. Das Endvolumen der aufgereinigten RNA in wässriger Lösung betrug somit 50 $\mu$ l. Die präparierte RNA wurde einer Qualitätskontrolle unterzogen (vgl. 2.9).

### 2.8.3 Optionale Ethanol-Präzipitation von RNA

Für die nachfolgende cDNA-Synthese wurden 15 $\mu$ g RNA in maximal 10 $\mu$ l H<sub>2</sub>O verwendet. War die RNA-Konzentration nach der Aufreinigung (vgl. 2.8.2) zu gering, musste eine Ethanol-Präzipitation durchgeführt werden. Es ist nicht angeraten, die RNA bei höherer Temperatur eindampfen zu lassen.

Die RNA-Lösung wurde mit DEPC-Wasser auf ein Volumen von 150 $\mu$ l eingestellt, die gleiche Menge 5M NH<sub>4</sub>OAc hinzugegeben. Es wurden 2,5 Volumenteile Ethanol und 0,5 $\mu$ l Glykogen hinzugegeben und gemischt. Nach Inkubation für 10min bei  $-20^{\circ}\text{C}$  wurde die Probe für 30min bei 12.000 $\times$ g und  $16^{\circ}\text{C}$  zentrifugiert. Das Pellet wurde anschließend mit 500 $\mu$ l kaltem 80% Ethanol gewaschen und für 5min bei 12.000 $\times$ g bei  $16^{\circ}\text{C}$  zentrifugiert. Nach Lufttrocknung des Pellets wurden die RNA in Rnase-freiem Wasser resuspendiert und die Konzentration bestimmt (vgl. 2.9.2).

### 2.8.4 Erst-Strang cDNA-Synthese

Das Volumen der RNA wurde auf 15 $\mu$ g/10 $\mu$ l eingestellt. Für die Primer-Hybridisierung wurde 1 $\mu$ l T7-(dT)<sub>24</sub>-Primer (100pmol/ $\mu$ l, Sequenz 5'-GGCCAGTGAATTGTAATACG-ACTCACTATAGGGAGGCGG-(dT)<sub>24</sub>-3') zu 10 $\mu$ l RNA-Lösung gegeben, der Ansatz für 10min bei  $70^{\circ}\text{C}$  inkubiert, anschließend 5sec zentrifugiert und für 2min auf Eis gestellt. Für die Erst-Strang-Synthese, in der an die vorhandene RNA ein komplementärer DNA-Strang synthetisiert wird, wurde das Superscript Choice System verwendet. Zu den 11 $\mu$ l der mit dem Primer hybridisierten RNA wurden 4 $\mu$ l 1<sup>st</sup>-Strand-Buffer (5 $\times$  konzentriert), 2 $\mu$ l 0,1M DTT (Endkonzentration 0,01M) und 1 $\mu$ l 10mM dNTP (Endkonzentration 0,5mM) gegeben und der Ansatz für 2min bei  $37^{\circ}\text{C}$  inkubiert. Nach Zugabe von 2 $\mu$ l der Reversen Transkriptase Superscript RT II wurde der Ansatz für 60min bei  $37^{\circ}\text{C}$  inkubiert.

### 2.8.5 Zweit-Strang-Synthese

In der Zweit-Strang-Synthese wird aus dem cDNA-Strang ein DNA-Doppelstrang gebildet, wobei der ursprüngliche RNA-Strang abgebaut wird. Der Syntheseansatz wurde wie folgt hergestellt: zu 30 $\mu$ l 2<sup>nd</sup>-Strand-Buffer (5 $\times$  konzentriert) wurden 3 $\mu$ l dNTP (10mM, Endkonzentration 0,2mM), 1 $\mu$ l *E. coli* DNA-Ligase (10u/ $\mu$ l), 4 $\mu$ l *E. coli* DNA

Polymerase I (10u/μl), 1μl RnaseH (2u/μl) und 95μl H<sub>2</sub>O gegeben. Zu diesen 134μl kommt der komplette Ansatz aus der Erst-Strang-Synthese. Der Ansatz wurde für 2h bei 16°C inkubiert und die Reaktion anschließend durch Zugabe von 10μl 0,5M EDTA gestoppt.

#### 2.8.6 Phenol-Chloroform-Extraktion der DNA

Über eine Phenol-Chloroform-Extraktion wurde die doppelsträngige DNA aus dem Syntheseansatz aufgereinigt. Dazu wurden Phase Lock Gel Tubes für 1min bei maximaler Drehzahl zentrifugiert. Diese Phase Lock Gel Tubes erleichtern die Trennung der organischen von der wässrigen Phase durch eine Gelschicht, die sich zwischen die beiden Phasen legt. Der komplette Ansatz der Synthese wurde in die so vorbereiteten Reaktionsgefäße überführt. Nach Zugabe von 166μl Phenol:Chloroform:Isoamylalkohol (Verhältnis 25:24:1) wurde der Ansatz kurz geschüttelt und für 2min bei 12.000×g zentrifugiert. Die obere Phase wurde in ein neues Reaktionsgefäß überführt und die gleiche Menge 5M NH<sub>4</sub>OAc zugegeben. Nach Zugabe von 2,5 Volumenteilen Ethanol und von 0,5μl Glycogen wurde der Ansatz kurz gemischt und für 10min bei -20°C inkubiert. Der Ansatz wurde für 30min bei 12.000×g bei 16°C zentrifugiert, das Pellet anschließend mit 500μl kaltem 80% Ethanol gewaschen und erneut für 5min bei 12.000×g bei 16°C zentrifugiert. Der letzte Waschschrift wurde einmal wiederholt, bevor das Pellet getrocknet und die DNA in 3μl Rnase-freiem Wasser resuspendiert wurde.

#### 2.8.7 *in vitro*-Transkription

Bei der nun folgenden *in vitro*-Transkription (IVT) wird die doppelsträngige DNA wieder in RNA umgeschrieben. Dabei werden neben unmarkierten auch mit Biotin gekoppelte Nukleotide verwendet, die später zum Markieren der RNA-Moleküle dienen. Bei den vorangegangenen Syntheseschritten wurden die mRNA Moleküle im Verhältnis 1:1 umgeschrieben. Die IVT Reaktion stellt eine 30-100fache, dabei lineare Amplifikation dar (Lipshutz, 1999). Diese Linearität ist eine Voraussetzung für die genaue Bestimmung der relativen Expressionslevel der betrachteten Transkripte. Die gesamte Methode zur Amplifikation heterogener Populationen von antisense- bzw. komplementärer RNA (cRNA) unter Verwendung von T7 RNA Polymerase wurde erstmalig von Van Gelder *et al.* (Van Gelder, 1990) beschrieben.

Für die *in vitro*-Transkription wurde das MEGAscript T7 Kit verwendet. Der zur Transkription erforderliche Ansatz setzt sich wie folgt zusammen: jeweils 2μl ATP (75mM, Endkonzentration 7,5mM) und GTP (75mM, Endkonzentration 7,5mM), jeweils 1,5μl CTP (75mM, Endkonzentration 5,6mM) und UTP (75mM, Endkonzentration

5,6mM), jeweils 3,75µl Biotin-CTP (10mM, Endkonzentration 1,9mM) und Biotin-UTP (10mM, Endkonzentration 1,9mM), jeweils 2µl Puffer-Mix (10× konzentriert) und Enzym-Mix (10× konzentriert). Zum Ansatz wurden 1,5µl der frisch hergestellten und gereinigten DNA aus 2.8.6 gegeben. Inkubiert wurde für 6h bei 37°C.

#### 2.8.8 Aufreinigung der IVT-Produkte

Nach Abschluss der *in vitro*-Transkription wurde die synthetisierte RNA mit dem RNeasy Mini Kit gereinigt (vgl. 2.8.2) und einer Qualitätskontrolle unterzogen (vgl. 2.9.1 und 2.9.2).

#### 2.8.9 Fragmentierung der cDNA

Bevor die markierte RNA auf die Microarrays hybridisiert werden kann, muss sie fragmentiert werden. Dabei entstehen RNA-Fragmente mit einer Größe von 35 bis 200 Basenpaaren. Diese kleinen Fragmente ermöglichen eine bessere Hybridisierung auf den Genechips als große Fragmente.

Für die Fragmentierung wurden 15µg der cRNA in einem Volumen von maximal 16µl benötigt. Wurde diese Konzentration nicht erreicht, mussten die IVT-Produkte gefällt werden. Dazu wurde das entsprechende Volumen für 15µg entnommen und nach dem Protokoll 2.8.3 präzipitiert. Die cRNA wurde anschließend in 16µl Rnase-freiem Wasser resuspendiert.

Zu 8µl der cRNA wurden 2µl eines Fragmentierungspuffers (5× Fragmentierungspuffer: 200mM Tris-Acetat, pH 8,1, 500mM KOAc, 150mM MgOAc, sterilfiltriert) gegeben und der Ansatz für 35min bei 95°C inkubiert. Um Kondenswasser im Reaktionsgefäß zu sammeln wurde die Probe für 10sec zentrifugiert und bis zur anschließenden Hybridisierung auf Eis aufbewahrt.

### 2.9 Qualitätskontrolle der präparierten Nukleinsäuren

#### 2.9.1 Native Gelelektrophorese

Die Analyse der RNA erfolgte über die Auftrennung in neutralen Agarosegelen (1,0% Agarose in 1×TAE-Puffer mit Ethidiumbromid zur Anfärbung der Nukleinsäuren). Als Laufpuffer wurde 1×TAE (40mM Tris/HCl, 20mM Natriumacetat, 2mM EDTA, pH 7,8) verwendet. Vor der Messung wurde die RNA für 5min bei 65°C erhitzt und sofort für 2min auf Eis gestellt. Es wurde 1µg RNA pro Probe aufgetragen.

### 2.9.2 Photometrische Konzentrationsbestimmung

Mittels photometrischer Bestimmung der Absorption bei 260nm und 280nm wurden die Konzentration und die Reinheit der Nukleinsäuren bestimmt. Berechnung der Konzentration für RNA (gemessen bei 260nm, Schichtdicke 1cm): 1 OD = 40µg/ml. Vor der Messung wurde die RNA für 5min bei 65°C erhitzt und sofort für 2min auf Eis gestellt.

Durch die Bestimmung des Quotienten der bei 260nm und 280nm gemessenen Absorption ( $A_{260}:A_{280}$ ) erhält man eine Information über die Reinheit der Nukleinsäurepräparation. Liegt der Quotient für RNA im Bereich 1,9 – 2,0, handelt es sich um eine reine Präparation. Für die Aufreinigung mit TRIzol ist ein Wert zwischen 1,5 – 1,7 akzeptabel. Phenol absorbiert ebenfalls bei 260nm, wodurch eine höhere RNA-Konzentration vorgetäuscht werden kann. Es ist unbedingt darauf zu achten, bei der Präparation kein TRIzol mitzuschleppen. Bei niedrigeren Quotienten kann man eine Verunreinigung durch Proteine und/oder Phenol erkennen. Eine genaue Bestimmung der Nukleinsäuremenge ist dann nicht möglich.

## 2.10 Affymetrix Gene Chip Instrument System

### 2.10.1 High-density Oligonucleotide Arrays

DNA-Hybridisierungsarrays bieten die Möglichkeit, Genexpressionsanalysen im genomischen Maßstab durchzuführen, in dem Sinne, dass in einem Versuchsansatz Tausende Gene parallel untersucht werden können. Als Gen wird in diesem Zusammenhang die DNA-Sequenz definiert, die in der Zelle in mRNA transkribiert wird und ein Protein kodiert. Intron- und Promotorsequenzen sowie andere funktionelle Elemente auf DNA-Ebene, die nach heutiger Definition Bestandteile eines Gens sind, werden bei dieser funktionellen Genexpressionsanalyse nicht betrachtet. Weiterhin wird mit Hybridisierungsarrays die Expression von ESTs (Expressed Sequence Tags) untersucht, also von Produkten der Transkription, deren Sequenz bekannt, aber deren Zuordnung zu einem Gen oder Protein nicht definiert ist (Schuler, 1997; Schuler, 1996).

Hybridisierungsarrays zeichnen sich durch eine Konstruktion aus, bei der auf einer festen Matrix die genspezifischen Sequenzen (Sonde, engl. probe) immobilisiert werden und mit markierter RNA (Probe, engl. sample, target) hybridisieren können. In diesem Sinn kann man Hybridisierungsarrays als eine Umkehrung des Northern-Blots betrachten (Freeman, 2000), bei dem die Probe immobilisiert wird und mit den freien, markierten Sonden hybridisieren kann. Hintergrund für die Genexpressionsanalyse mit Hybridisierungsarrays ist die Theorie, dass, je stärker die Expression eines Gens ist, desto größer ist die Menge markierter RNA und folglich desto größer das zu

detektierende Signal ist. Hierbei können jedoch nur relative Transkriptionsniveaus und keine absoluten Mengen oder gar die Kopienzahl bestimmt und miteinander verglichen werden. Eine Voraussetzung der Genexpressionsanalyse ist die Linearität aller experimentellen Schritte. Nur so lassen sich reproduzierbare Aussagen über den Expressionslevel hoch oder niedrig exprimierter Gene machen und miteinander vergleichen. Bei den Syntheseschritten zur Erst- und Zweit-Strang Synthese handelt es sich um 1:1 Umschreibungen. Die IVT ist eine lineare Amplifikation. Ebenso besteht zwischen der Menge an fluoreszenzmarkierter Probe und der detektierten Signalintensität ein linearer Zusammenhang (Freeman, 2000).

Die in dieser Untersuchung verwendeten high-density Oligonucleotide Arrays (Lipshutz, 1999; Lockhart, 1996) (synonym werden die Ausdrücke Genechip oder Microarray verwendet) unterscheiden sich in einigen Eigenschaften grundlegend von anderen Hybridisierungsarrays. Im Gegensatz zu anderen Arrays werden bei diesem System die Sonden *in situ* auf dem Chip synthetisiert. Hierzu wird eine kombinierte Methode aus Nukleotidsynthese und Photolithographie verwendet. Die Photolithographie ist der Halbleiterindustrie entlehnt und erlaubt die gezielte schrittweise Synthese der Oligonukleotide, Base für Base, an definierten Positionen auf dem Array (Fodor, 1993; Fodor, 1991; McGall, 1996). Durch diese Technik ist man in der Lage, bis zu 300.000 unterschiedliche Sonden auf einer Fläche von nur 1,28cm × 1,28cm zu synthetisieren. Dabei werden pro Sonde auf einer Fläche von 20µm × 20µm  $1 \times 10^6$  bis  $1 \times 10^7$  Kopien eines Oligonukleotids erzeugt (Schadt, 2000).

Die Synthese von Oligonukleotiden mit mehr als 25 Basen Länge ist problematisch. Da diese Sondenlänge jedoch keine hochspezifische Hybridisierung der Ziel-RNA zulässt, werden für jedes zu detektierende Gen bis zu 40 verschiedene Sonden, in ihrer Gesamtheit als Sondensatz (probe set) bezeichnet, verwendet. Die Sonden sind unter anderem so ausgewählt, dass sie sich, wenn möglich, nicht überlappen. Die Hälfte dieser Sonden stellen die genspezifischen Sonden (perfect match probe, PM). Zu jeder genspezifischen Sonde gibt es genau eine Sonde, bei der an Position 13 des Oligonukleotids eine homomere Base ausgetauscht wurde (mismatch probe, MM). Die Idee der Mismatch-Sonden ist, Hintergrundsignale (zum Beispiel durch Eigenschaften des Scanners verursachtes Rauschen der Signale) oder unspezifische Interaktionen (zum Beispiel Kreuzhybridisierung oder unspezifische Hybridisierungen) zu quantifizieren und mit dem PM-Signal zu verrechnen (Lockhart, 1996; Wodicka, 1997). Oligonukleotid-Arrays haben den Vorteil, dass die Sonden sequenzspezifisch aus dem gesamten Genom gezielt ausgewählt werden können, wodurch sich Kreuzhybridisierungen minimieren lassen, ein Problem, welches bei den cDNA Hybridisierungsarrays mit 1 bis 2kb langen Sonden besteht (Golub, 2001).

Das „Gene Chip Instrument System“ der Firma Affymetrix besteht aus vier Gerätekomponenten. Eine zentrale Computerplattform dient der Steuerung der übrigen Geräte sowie der Analyse der erhobenen Messdaten. Im Hybridisierungssofen werden die mit der Probe beladenen Genechips unter Temperierung und Rotation hybridisiert. Die sogenannte Fluidics Station ist ein automatisiertes System, in dem vier Genechips gleichzeitig nach einem definierten Programm gewaschen und mit einem Fluoreszenzfarbstoff markiert werden. Der konfokale Laserscanner „Gene Array Scanner“ (Hewlett Packard, Palo Alto, CA, USA) letztlich ist mit einem Argonlaser ausgestattet. Er regt die auf dem Genechip gebundenen Fluoreszenzfarbstoffe der markierten und hybridisierten Nukleinsäuren an und misst das emittierte Licht bei einer Wellenlänge von 570nm. Gescannt werden die Genechips mit einer Auflösung von 3µm (Pixelgröße).

#### 2.10.2 Eigenschaften der verwendeten Arrays

Für die Analyse der Genexpression wurden zwei verschiedene Genechips verwendet, die humane Gensequenzen abfragen: EOS-H und Affymetrix HG-U133A. Bei den verwendeten EOS-H Genechips handelt es sich um Spezialanfertigungen. Im Gegensatz zu den unterschiedlichen Genechips der Firma Affymetrix (Santa Clara, California, USA), befinden sich hier auf einem Chip 35.356 Sondensätze für 5770 Gene und 29.586 EST-Cluster. Jeder Sondensatz besteht aus 20 verschiedenen Sonden. Weiterhin enthalten die hier verwendeten Genechips keine MM-Sonden. Durch Auswahl besonders zuverlässiger Sonden und durch Verwendung eigener Algorithmen zur Analyse konnte auf diese verzichtet werden (Glynn, 2000b). Jeder einzelne Sondensatz trägt eine eindeutige Kennzeichnung, mit deren Hilfe die Ursprungssequenzen in öffentlich zugänglichen Datenbanken identifiziert werden können (<http://www.ncbi.nlm.gov>, hier zum Beispiel GenBank). Diese Microarrays wurden auf der Basis einer wissenschaftlichen Kooperation für dieses Projekt zur Verfügung gestellt.

Der Affymetrix Genechip HG-U133A detektiert mit seinen 22.308 Sondensätzen 13.220 bekannte Gene und 5.413 EST-Cluster. Jeder Sondensatz setzt sich aus je elf PM- und MM-Sonden zusammen. Die Sequenzen, auf denen die Sonden beruhen, entstammen der „UniGene Database (Build 133, 20.04.2001)“ und wurden zusätzlich mit Daten aus zwei weiteren, öffentlich zugänglichen Datenbanken ergänzt („Washington University EST Trace Repository“ und „University of California, Santa Cruz Golden Path Human Genome Database“).

### 2.10.3 Hybridisierung

Für die Herstellung der Hybridisierungslösung wurden zuerst eine 12× MES Lösung (1,22M MES (2-[N-Morpholino]-Ethansulfonsäure), 0,89M MES-Natriumsalz, in Wasser, pH 6,5, sterilfiltriert) und ein 2× Hybridisierungspuffer (200mM MES, 2M [Na<sup>+</sup>], 40mM EDTA, 0,02% Tween20) hergestellt. Beide Lösungen wurden bei 4°C gelagert.

Zur Herstellung der Hybridisierungslösung wurden zu 15µg fragmentierter RNA (vgl. 2.8.9) (Endkonzentration 0,05µg/µl) 15µl Eucaryotic Hybridization Controls (Spike Controls, Endkonzentration 100pM), 5µl Control Oligonucleotide B2 (3nM, Endkonzentration 50pM), 3µl Hering Sperm DNA (3nM, Endkonzentration 0,1mg/ml), 3µl acetyliertes BSA (50mg/ml, Endkonzentration 0,5mg/ml), 150µl des 2× Hybridisierungspuffers und 104µl Wasser gegeben und gemischt. Die Hybridisierungslösung wurde für 5min bei 99°C und anschließend für 5min bei 45°C erhitzt. Die Hybridisierungslösung kann bei -80°C eingelagert werden. Vor der Inkubation auf den Arrays wurde die fragmentierte RNA enthaltende Hybridisierungslösung für 2min bei maximaler Drehzahl zentrifugiert, um unlösliche Partikel zu sedimentieren.

Die Arrays wurden auf RT erwärmt, mit 200µl des 1× Hybridisierungspuffers befüllt und für 10min bei 45°C unter Rotation im Hybridisierungsofen präinkubiert. Nach Entnahme des Puffers wurden die Arrays mit 200µl der Hybridisierungslösung befüllt und im Hybridisierungsofen für 16h bei 45°C inkubiert.

### 2.10.4 Waschen und Färben der Microarrays

Nach der Inkubation wurde die Hybridisierungslösung entnommen. Diese kann bei -80°C eingelagert werden und gegebenenfalls erneut verwendet werden. Die Arrays wurden mit 200µl des Puffers Non-Stringent-Wash A befüllt (6× SSPE (20× SSPE Stock aus 3 M NaCl, 0,2 M NaH<sub>2</sub>PO<sub>4</sub>, 0,02 M EDTA), 0,01% Tween20, in Wasser, sterilfiltriert, 0,005% Antifoam) und bis zur weiteren Verarbeitung bei RT im Dunkeln aufbewahrt. Des weiteren wurde der Puffer Stringent-Wash B hergestellt (100mM MES, 0,1M [Na<sup>+</sup>], 0,01% Tween20, sterilfiltriert).

Für die Herstellung der SAPE-Lösung (Streptavidin-Phycoerythrin) wurde zuerst ein 2× Färbe-Puffer hergestellt (200mM MES, 2M [Na<sup>+</sup>], 0,1% Tween20, sterilfiltriert, 0,01% Antifoam). Zu 400µl des 2× Färbe-Puffers wurden 360µl H<sub>2</sub>O, 32µl acetyliertes BSA (50mg/ml) und 8µ Streptavidin-Phycoerythrin (1mg/ml) gegeben. Die SAPE-Lösung muss lichtgeschützt verwahrt werden.

In der Fluidics-Station des Systems wurden die hybridisierten Arrays mit der SAPE-Lösung für 30min bei 25°C inkubiert und automatisch mit den beiden Lösungen Non-Stringent-Wash A und Stringent-Wash B gewaschen (Waschprogramm EukGE-WS1).

#### 2.10.5 Microarray Scan

Die fertigen Arrays wurden bei 570nm (Anregungswellenlänge 488nm) bei einer Auflösung von 3µm mit einer Wiederholung gescannt. Die gemessenen Lichtintensitäten aus beiden Bildern werden automatisch pro Pixel gemittelt und das endgültige Bild des Arrays gespeichert.

#### 2.10.6 Waschen und Färben mit Antikörperverstärkung

Die bereits mit SAPE gefärbten und gescannten Arrays können einem zweiten Färbeschritt unterzogen werden, bei dem es durch die Verwendung biotinylierter anti-Streptavidin-Antikörper und erneuter Färbung mit SAPE zu einer Verstärkung der Fluoreszenzsignale kommt. Für die Antikörper-Lösung wurden zu 400µl des 2× Färbepuffers (355,2µl H<sub>2</sub>O, 32µl acetyliertes BSA (50mg/ml), 8µl Ziegen-IgG 10mg/ml und 4,8µl biotinylierter anti-Streptavidin-Antikörper (0,5mg/ml)) gegeben.

In der Fluidics-Station des Systems wurden die hybridisierten und bereits einmal gefärbten Arrays erneut mit der SAPE-Lösung für 10min bei 25°C, dann mit der Antikörper-Lösung für 10min bei 25°C und wieder mit der SAPE-Lösung für 10min bei 25°C inkubiert und automatisch mit den beiden Lösungen Non-Stringent-Wash A und Stringent-Wash B gewaschen (Waschprogramm EukGE-WS2). Für die EOS-H Genechips entfällt der erste der drei Inkubationsschritte.

#### 2.10.7 Microrray Scan nach Antikörperverstärkung

Die fertigen Arrays wurden erneut bei 570nm (Anregungswellenlänge 488nm), bei einer Auflösung von 3µm mit einer Wiederholung gescannt. Die gemessenen Lichtintensitäten aus beiden Scans werden pro Pixel gemittelt und das endgültige Bild des Arrays gespeichert.

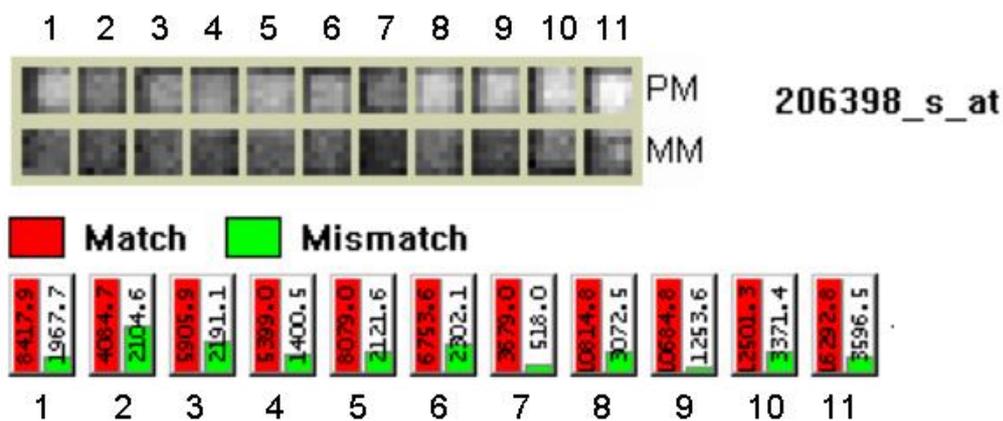
### 2.11 Computer-gestützte Analyse der Genexpression

Der konfokale Laserscanner generiert ein Bild, indem er jedes einzelne Sondenfeld mit seinem Laserlicht anregt, die resultierende Photonenemission der fluoreszenzmarkierten und hybridisierten cRNA detektiert und das Lichtsignal in einen 16-Bit Intensitätswert übersetzt. Die Bilder der gescannten Genechips stellen die Rohdaten eines jeden Experimentes dar.

Die fundamentalen Analyseschritte für jeden einzelnen Microarray sind dann die Aufteilung des Bildes in ein Raster, die Korrektur des Hintergrundes, die Skalierung für

die nachfolgenden Chip-zu-Chip Vergleiche und die statistische Berechnung des singulären Signals für jedes Gen, das als relatives Maß der Expressionsstärke steht. Für die HG-U133A Genechips erstellt die Affymetrix Software für jedes Bild eine Liste aller auf dem Array vorhandenen SONDENSÄTZE beziehungsweise Gene und macht neben dem Signalwert (Abbildung 2-1) eine Angabe darüber, ob das Gen exprimiert (P, present) oder nicht exprimiert (A, absent) ist. Das dargestellte Beispiel zeigt den SONDENSATZ für CD19 der Patientenprobe ALL54TEY. Im oberen Teil der Abbildung sind Ausschnitte des gescannten Gesamtbildes zusammengestellt. Jedes einzelne Feld zeigt die gemessenen Lichtsignale eines SONDENFELDES, die auf dem Genechip nicht in dieser Anordnung zu finden sind. Im unteren Teil der Abbildung sind die gemessenen Lichtintensitäten in Zahlenwerte übersetzt und die zueinandergehörenden SONDENPAARE (PM und MM) gegenübergestellt worden. Aus diesen Messwerten wird schließlich für jedes Gen ein Einzelwert, das Signal - in einer nicht mehr gebräuchlichen Nomenklatur als „Average Difference“ bezeichnet - berechnet. Dieses Signal macht eine relative Angabe über die ursprüngliche Menge eines Transkriptes und repräsentiert somit die Expressionsstärke eines Gens.

Aufgrund des unterschiedlichen Designs (vgl.2.10.2) können die EOS-H Genechips nicht mit der Affymetrix Software ausgewertet werden. Die Originalbilder wurden deshalb von unserem Kooperationspartner primär ausgewertet. Wir erhielten zur weiteren Analyse der Ergebnisse die vollständigen Listen mit den Signalwerten für alle Gene und ESTs zurück.



**Abbildung 2-1 Exemplarische Darstellung eines einzelnen SONDENSATZES für CD19**

Im oberen Teil der Abbildung sind die Ausschnitte eines gescannten Genechips dargestellt, die die verschiedenen SONDEN eines SATZES zur Detektion von CD19 tragen. Jeder PM-Sonde wird die entsprechende MM-Sonde gegenübergestellt. Der untere Teil der Abbildung stellt die aus den gemessenen Lichtintensitäten ermittelten Messwerte vom PM und MM paarweise gegenüber. Aus allen Messwerten wird schließlich ein einzelner Wert (Signal) berechnet und in der Datentabelle angegeben, der die Expressionsstärke von CD19 in dieser Probe repräsentiert. (Probe ALL54TEY)

### 2.11.1 Normalisierung

Nichtbiologische Faktoren können zur Variabilität der Daten in vielen biologischen Experimenten führen. Diese spielen auch bei der Untersuchung der differentiellen Genexpression mit vielen einzelnen experimentellen Schritten eine Rolle. Bereits geringe Variationen in Menge und Qualität der hybridisierten Proben, die Menge an verwendeten Färbereagenzien oder weitere experimentelle Bedingungen können insgesamt, trotz aller Standardisierungen der Methoden und Protokolle, die Hybridisierungsintensitäten beeinflussen. Um verlässlich verschiedene Genechip-Experimente miteinander vergleichen zu können, müssen die Unterschiede nichtbiologischen Ursprungs minimiert und die Datensätze vor der vergleichenden Analyse normalisiert werden. Die der Normalisierung zugrundeliegende Annahme ist, dass das gesamte Signal eines Genechips (nicht notwendigerweise die Verteilung der individuellen Fluoreszenzintensitäten von Sondensätzen), über verschiedene Hybridisierungsexperimente betrachtet, vergleichbar, das heißt ungefähr gleich sei. Diese Annahme ist vertretbar, da mit jeder Hybridisierungslösung dieselbe RNA-Menge auf einen Chip gebracht wird. Weiterhin wird vorausgesetzt, dass die betrachteten Variablen, also die Fluoreszenzintensitäten aller Sondensätze auf einem Chip, regelmäßig verteilt sind. Aufgrund des unterschiedlichen Designs der EOS-H und der HG-U133A Genechips lassen sich die Daten der unterschiedlichen Genechips jedoch nicht nach demselben Prinzip normalisieren.

#### 2.11.1.1 Normalisierung der EOS-H Genechips

Als Basisverteilung zur Normalisierung der Sondensatz-Intensitäten wurde die Gamma-Verteilung ausgewählt. Die Gamma-Verteilung  $\Gamma_{k,\lambda}(x)$  wird in der Statistik häufig angewendet und stellt ein statistisches Modell für Variablen dar, die nur positive Werte annehmen können (zum Beispiel Altersverteilungen oder Intensitätswerte). Die Gamma-Verteilung wird durch die beiden Größen Formparameter und Skalenparameter charakterisiert. Der Formparameter  $k$  beschreibt die Form der Verteilung, zum Beispiel als Exponential- oder Normalfunktion. Über den Skalenparameter  $\lambda$  wird die Varianz beschrieben. Durch geeignete Wahl beider Parameter können die meisten empirisch ermittelten Verteilungen inklusive ihrer Mittelwerte  $M=k/\lambda$  und Varianzen  $V=k/\lambda^2$  gut repräsentiert werden. Des Weiteren wird der kleinste positive Wert, den eine Gamma-Variable im Experiment angenommen hat, als Offset-Parameter betrachtet.

Die im folgenden kurz beschriebene Prozedur wird auf jeden einzelnen Genechip angewendet. Dabei fließen die Daten aus beiden Scans, vor und nach der Antikörperverstärkung, in die Normalisierungsprozedur ein. Die Normalisierung ist in dem Sinne chipimmanent, als dass für sie keine Daten eines anderen Genechips

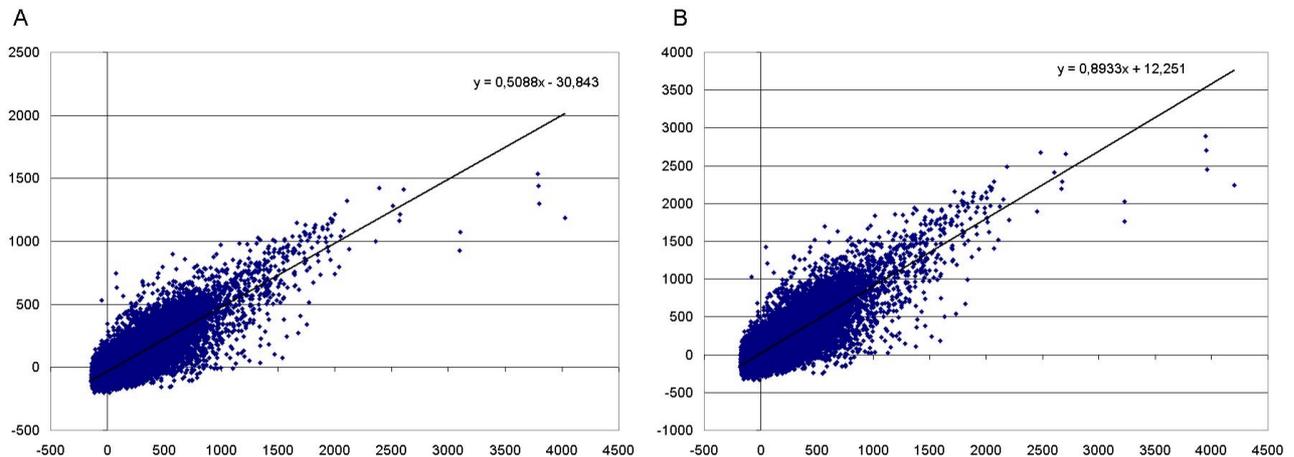
herangezogen werden. Bei der Normalisierung wird die Form der Verteilung aller Signale, bestimmt durch den Formparameter, beibehalten. Ursprung und Skalierung der Verteilung werden durch Einstellen von Offset-Parameter und Skalenparameter variiert. Aus allen gemessenen Intensitätswerten wird eine Gamma-Verteilung berechnet, die die Verteilung der Messwerte optimal beschreibt. Mittels einer Goodness-of-fit Statistik wird diese errechnete Gamma-Verteilung überprüft. Die ursprüngliche Gamma-Funktion wird nun so modifiziert, dass der Mittelwert der Verteilung den Wert 300 annimmt (dieser Wert wurde als Mittelwert festgelegt). Der die Form der Verteilung beschreibende Parameter bleibt dabei unverändert. Als nächstes wird aus allen Messwerten eine empirische Verteilungsfunktion berechnet. Letztlich wird dann jedem Wert auf der empirischen Verteilungsfunktion ein neuer Wert auf der angepassten Verteilungsfunktion – das ist die modifizierte Gamma-Verteilung – zugeordnet, der den normalisierten Intensitätswert darstellt, mit dem alle weiteren Analysen durchgeführt werden (persönliche Mitteilung, EOS Biotech).

#### 2.11.1.2 Lineare Regression der EOS-H Genechips

Nach der Normalisierung der EOS-H Genechip Daten wurden diese einer linearen Regression unterzogen, um die Ergebnisse besser miteinander vergleichen zu können. Trägt man jeden Messwert  $x$  von Chip  $j$  gegen jeden Messwert  $y$  von Chip  $i$  paarweise für jedes Gen gegeneinander auf, erhält man in der grafischen Darstellung eine Punktwolke aller Messwerte. Durch diese Punktwolke lässt sich eine Regressionsgerade legen, die durch die Formel:  $y = ax + b$  beschrieben wird. Für zwei identische Datensätze ergibt sich:  $y = x$ . Aus der Abbildung 2-2 wird deutlich, wie unterschiedlich zwei beliebige Proben zueinander sein können. Die Annahme ist jedoch, dass sich die miteinander zu vergleichenden Proben, zumindest innerhalb der Gruppen „Leukämie“ oder „CB-Kontrollen“, nur in einem geringen Prozentsatz differentiell exprimierter Gene unterscheiden. Die feststellbaren großen Unterschiede, im dargestellten Beispiel liegen die Maxima bei ungefähr 1.500 und 4.000, sind jedoch wahrscheinlich experimentell bedingt und verzerren die Auswertung.

Für die lineare Regression wurde als Basisdatensatz die Probe ALL25HCS als durchschnittliche Probe gewählt. Mit Hilfe dieses Datensatzes wurden für alle weiteren Proben Regressionsgeraden bestimmt und die Messwerte schließlich so verändert, dass sich alle Regressionsgeraden einander angleichen. Jedem Wert  $D_{ji}$ , mit  $j$  = Signal für Gen  $J$  der zu korrigierenden Probe und  $i$  = Signal des entsprechenden Gens des Basisdatensatzes, wird ein neuer Wert  $D_{ji}'$  zugeordnet, für den gilt:  $D_{ji}' = (D_{ji} - b)/a$ . Damit wird, bildlich gesprochen, die Lage der Messwertpaare des neu skalierten Datensatzes im Koordinatenkreuz verschoben. Die relative Lage der einzelnen

Messwertpaare zueinander wird dabei jedoch nicht verändert. Die Expressionsunterschiede eines Gens in verschiedenen Proben, die sich in den Messwerten niederschlagen, bleiben erhalten. Die Abbildung 2-2-B zeigt dasselbe Probenpaar wie oben nach durchgeführter linearer Regression. In der Gegenüberstellung der beiden Grafiken vor und nach der durchgeführten linearen Regression wird der Effekt deutlich.



**Abbildung 2-2 Punktdiagramm zweier Leukämieproben vor linearer Regression**

A: In diesem Beispiel wurden die Signalwerte für jeden Sondensatz von zwei Proben exemplarisch gegeneinander aufgetragen und die sich ergebende Regressionsgerade berechnet. Die angegebenen Intensitäten der Signale sind dimensionslose Zahlen. B: Dieselben Proben, jedoch nach erfolgter linearer Regression. Man sieht den Effekt der linearen Regression daran, dass sich die Form der Punktwolken nicht verändert haben. Verändert hat sich ihre Lage im zweidimensionalen Achsenkreuz.

### 2.11.1.3 Normalisierung der Affymetrix HG-U133A Genechips

Die Skalierung ist eine mathematische Methode um Unterschiede zu minimieren, die durch die Probenbearbeitung, die Hybridisierungsbedingungen, die Färbung oder durch unterschiedliche Herstellungsladungen der Genechips zustande kommen können. Die Skalierung ermöglicht den direkten Vergleich verschiedener Genechips miteinander, sofern sie vom selben Chiptyp sind.

Für die Skalierung werden alle vorhandenen Sondensätze berücksichtigt. Alle Einzelsignale eines Arrays werden zu einem bereinigten Mittelwert (Trimmed Mean Signal) zusammengefasst. Für diesen bereinigten Mittelwert werden vor der Berechnung die extremen Minima und Maxima entfernt. Der bereinigte Mittelwert wird dann auf eine vom Anwender zu definierende Gesamtintensität (Target Signal) korrigiert. Der hierfür erforderliche Skalierungsfaktor (Scale Factor, SF) ermittelt sich demnach aus:  $\text{Target Signal} = \text{Scale Factor} \times \text{Trimmed Mean Signal}$ . Die Idee für diese Art der Skalierung ist die Annahme, alle Genechips hätten in etwa die gleiche Gesamthelligkeit beziehungsweise eine gleiche mittlere Helligkeit.

Alle Arrays in diesem Experiment wurden auf dieselbe mittlere Targetintensität (Wert 500, Empfehlung Affymetrix) unter Berücksichtigung aller vorhandenen Sondensätze normalisiert. Die notwendigen Skalierungsfaktoren hierfür erreichten dabei Werte zwischen 2,3 und 42. Die Unterschiede in den Skalierungsfaktoren von Arrays, die direkt miteinander verglichen werden sollen, sollten allerdings nicht mehr als den Faktor 5 betragen. Bei zu großen Skalierungsfaktoren besteht die Gefahr, dass der lineare Zusammenhang zwischen auf dem Array hybridisierter RNA und dem gemessenen Fluoreszenzsignal verloren geht. Besonders schwache Signale würden unverhältnismäßig stark hervorgehoben und verfälschten somit das Ergebnis.

Im Gegensatz zu den EOS-H Genechips, bei denen die Daten vor und nach Antikörperverstärkung der Signale zur Normalisierung zusammengefasst wurden, müssen bei den HG-U133A diese beiden Datensätze unabhängig voneinander ausgewertet werden, wenngleich beiden Sätzen dieselbe Normalisierungsprozedur zugrunde liegt.

#### 2.11.2 Methoden zur Auswertung der Genexpressionsdaten

Die Ergebnisse der mit Affymetrix HG-U133A Genechip erhobenen Daten wurden mit den Programmen „Microarray Analysis Suite“ (MAS, Version 5.0), „Micro Database“ (MDB, Version 3.0) und „Data Mining Tool“ (DMT, Version 3.0) ausgewertet. Aufgrund des speziellen Designs der verwendeten EOS-H Genechips konnte auf die von Affymetrix angebotenen Computerprogramme zur Auswertung nicht zurückgegriffen werden. Die Analyse dieser Arrays wurde mit den Statistikfunktionen von „Microsoft Excel“ durchgeführt. Die Affymetrix HG-U133A Genechips wurden nach der Skalierung dann ebenfalls mit diesem Programm statistisch ausgewertet, um für beide Genechiptypen vergleichbare Methoden der Auswertung anwenden zu können.

##### 2.11.2.1 Identifikation differentiell exprimierter Transkripte

Aus den erhobenen Messdaten sollen diejenigen Transkripte identifiziert werden, die einen eindeutigen Unterschied zwischen den Leukämieproben und den Kontrollen zeigen. Diese Identifikation differentiell exprimierter Gene zwischen zwei definierten Gruppen von Proben wird als „Gene Selection“ oder „Class Comparison“ bezeichnet (Simon, 2003). Die Definition der Gruppen geschieht unabhängig von den Genexpressionsprofilen. Die Frage nach der richtigen Auswertemethode ist derzeit ein zentraler Diskussionspunkt der Microarray-gestützten Genexpressionsanalyse. Das generelle Problem der statistischen Analyse der Microarraydaten ist dabei, dass die Anzahl  $n$  der untersuchten Proben (Gewebe, Zelllinien, Patientenproben) klein ist. Dem gegenüber steht die sehr große Anzahl  $p$  der in einem Ansatz untersuchten Parameter

(Gene). In der Literatur wird vom „large p, small n“-Problem gesprochen. Es werden zur Zeit viele Artikel mit Ergebnissen aus Genexpressionsanalysen veröffentlicht. Die Zahl der Arbeiten, die sich gezielt mit den Möglichkeiten zur Auswertung befassen, ist allerdings vergleichsweise gering (Ben-Dor, 2000; Ben-Dor, 2001; Brazma, 2000; Thomas, 2001). Ein Konsens über angemessene Normalisierungs- oder Analysemethoden besteht nicht, wenngleich es Bestrebungen in diese Richtung gibt (Knight, 2002).

Häufig wird zur Identifikation differentiell exprimierter Gene der statistische  $t$ -Test für zwei unverbundene Proben angewandt (Arfin, 2000; Garber, 2001; Hedenfalk, 2001; Notterman, 2001; Tanaka, 2000; Virtaneva, 2001). Dieser Test geht von der Annahme aus, dass die Mittelwerte zweier Datensätze gleich sind. Mit  $t$ -Tests kann dann überprüft werden, ob die Mittelwerte zweier Stichproben zufällig unterschiedlich sind. Zu den Voraussetzungen des  $t$ -Test gehört die Normalverteilung der beiden Grundgesamtheiten, in diesem Falle der Genexpressionsdaten der beiden Gruppen „Leukämie“ und „Kontrolle“. Das Programm „Microsoft Excel“ bietet einen zweiseitigen  $t$ -Test für zwei unverbundene Stichproben mit unterschiedlichen Varianzen der Grundgesamtheiten an, der für die Auswertung herangezogen wurde. Die Zuverlässigkeit dieses Testverfahrens wurde bereits gezeigt (Thomas, 2001).

Es wurde mit einer Irrtumswahrscheinlichkeit  $\alpha$  von 1% (oder  $p=0,01$ ) getestet. Da sehr viele Gene gleichzeitig betrachtet werden und für jedes einzelne Gen der  $t$ -Test gemacht wird, erhöht sich die Wahrscheinlichkeit für zufällige Unterschiede in den Mittelwerten. Um dies zu berücksichtigen, sollte eine Bonferroni-Korrektur eingeführt werden. Auf diesem Weg lässt sich die Rate falsch-positiver Ergebnisse verringern (Thomas, 2001). Für die Bonferroni-Korrektur würde das gewünschte Signifikanzniveau durch die Gesamtzahl der durchgeführten Tests, also die Anzahl der auf den Genechips vorhandenen Sondensätze, geteilt werden. Das würde für die verwendeten Genechips bedeuten, dass nur Expressionsunterschiede mit  $p < 2,9 \times 10^{-7}$  beziehungsweise  $p < 4,9 \times 10^{-7}$  als signifikant betrachtet werden könnten. Die Größenordnung der Expressionsunterschiede und die erforderliche Messgenauigkeit sowie die zu fordernde Exaktheit aller durchgeführten Präparations- und Syntheseschritte lassen es kaum möglich erscheinen, ein solch hohes Signifikanzniveau experimentell zu erreichen. Die strikte Anwendung der Bonferroni-Korrektur würde viele biologisch relevante Unterschiede verwerfen (Claverie, 1999). Auf die Einführung der Bonferroni-Korrektur wurde daher verzichtet.

Die oben genannten Veröffentlichungen, die den  $t$ -Test zur Identifikation differentiell exprimierter Gene herangezogen haben, arbeiten mit unterschiedlichen Signifikanzniveaus von 5% bis 0,1%. Für die vorliegende Arbeit wurde das hohe

Signifikanzniveau 1% verwendet. Weiterhin gibt es keinen Konsens, vor allem auch keine Untersuchungen darüber, ab welcher Größe ein auf Transkriptionsebene gemessener Expressionsunterschied wirklich biologisch bedeutsam ist. Die Faktoren (Fold Change, FC), die von den unterschiedlichen Autoren als minimaler Unterschied zugelassen werden, reichen von 1,5 (Chen, 2001b) bis 4 (Notterman, 2001). In dieser Arbeit wurden FC-Werte  $>2$  als Expressionsunterschied akzeptiert.

#### 2.11.2.2 Identifikation von Klassen und Clustern

Die Identifikation differentiell exprimierter Gene beim Vergleich der Expressionsprofile zweier Untersuchungsgruppen kann als „Class Comparison“ bezeichnet werden (vgl. 2.11.2.1). Die zu vergleichenden Gruppen werden zu Beginn der Analyse definiert, ohne Kenntniss der Genexpressionsprofile. Die Klassifikation von Proben entsprechend eines Expressionsprofils, als „Class Prediction“ bezeichnet, ist der „Class Comparison“ in dem Sinne ähnlich, als das hier die Klassen vordefiniert werden. Ziel ist die Identifikation von Prediktoren, einer Anzahl von Schlüsselgenen, die eine Vorhersage ermöglichen, welcher Gruppe eine Probe zuzuordnen sei. Die „Class Discovery“, das eigentlich unüberwachte Clustern, definiert zu Beginn der Analyse keine Gruppen oder Klassen. Unter Clustern versteht man die Einteilung von Proben in Gruppen aufgrund von Ähnlichkeiten in verschiedenen Merkmalen, in diesem Falle ähnlicher Expressionsmuster, ohne dabei die Klassenzuordnung der Proben zu beachten. Das Einsortieren in zu bildende Gruppen oder Cluster geschieht nur unter Berücksichtigung der gemessenen, damit intrinsischen Probeneigenschaften. Aus dieser Eigenschaft rührt der Name des „unsupervised learning“ oder unüberwachten Clusters.

Eine Vielzahl an Clustermethoden ist auf die mit Microarrays ermittelten Expressionsprofile angewendet worden, darunter das „Hierarchische Clustern“ (Alon, 1999; Eisen, 1998), „Self-organizing Maps (SOM)“ (Tamayo, 1999), „Relevance Networks“ (Butte, 2000) oder das „*k*-Means Clustering“ (Sherlock, 2000). Die dahinterstehenden mathematischen Methoden sind nicht neu, werden aber zunehmend für die Analyse genomischer Daten benutzt. Jede dieser Methoden scheint dazu geeignet, die vorherrschenden Strukturen in einem Datensatz aufzuspüren (Golub, 2001). Prinzipiell eignen sich die unterschiedlichen Methoden zum Clustern und zum Klassifizieren.

Das von M. Eisen erstmalig auf Genexpressionsdaten angewandte hierarchische Clustern beruht mathematisch auf einer „Nearest Neighborhood Analysis“ auf Grundlage der Pearson-Korrelation und erkennt Muster in Genexpressionsprofilen. Das entsprechende Programm „Cluster and TreeView“ ist auf der Internetseite M. Eisens frei verfügbar und wurde bereits mehrfach für Analysen verwendet: bei der Untersuchung der Reaktion humaner Fibroblasten auf Serum (Iyer, 1999), bei der Untersuchung von

Adenokarzinomen (Notterman, 2001), bei der Analyse der Wirkung von Umwelteinflüssen auf die Bäckerhefe *Saccharomyces cerevisiae* (Gasch, 2000), bei der Identifikation von Subklassen des Brustkrebses (Sorlie, 2001) sowie bei Lungen- und Nierenkrebs-Studien (Chen, 2001a; Takahashi, 2001). Der Vorteil dieser Methode ist die Darstellung der Ergebnisse in hierarchisch organisierten Bäumen, die die Verwandtschaft oder Ähnlichkeit von Gruppen oder Clustern anschaulich machen. Einander ähnliche Proben sind durch kurze Äste miteinander verbunden. Je länger die Äste werden und je mehr Verzweigungen zwischen zwei Proben liegen, desto unähnlicher sind sie sich. Hierarchische Clustermethoden sind gut dazu geeignet, die Daten in zwei Dimensionen zu clustern. Erstens können die Proben zu Gruppen sortiert werden. Hierbei wird danach geschaut, welche Proben, über den jeweiligen Microarray betrachtet, Ähnlichkeiten aufweisen. Zweitens können Gene geclustert werden, indem nach Genen gesucht wird, die, über die verschiedenen Proben betrachtet, ähnliche Expressionsmuster besitzen.

Die Ähnlichkeit zwischen zwei Proben, also in diesem Fall zwischen zwei Datenreihen von Messwerten, wird mit Hilfe des Pearson Korrelationskoeffizienten bestimmt. Von den verschiedenen mathematischen Möglichkeiten, den Pearson Korrelationskoeffizienten zu berechnen, bietet das Programm vier an: nicht zentrierte Korrelation, zentrierte Korrelation, absolute nicht zentrierte Korrelation, absolute zentrierte Korrelation sowie zwei weitere Methoden zur Berechnung von Korrelationen: Spearman Rang Korrelation und Kendall's  $\tau$ -Test. Von allen möglichen Merkmalskombinationen werden die Korrelationskoeffizienten berechnet. Zum Clustern wird dann die Entfernung zwischen allen Datenreihen in einer Distanzmatrix ermittelt, wobei Distanz nichts anderes bedeutet als „Distanz = 1,0 – Korrelation“. Das Clustern beginnt mit dem Ermitteln der beiden Proben mit der geringsten Distanz, also mit dem größten Korrelationskoeffizienten. Diese werden über zwei Äste in einem Knoten miteinander verbunden. Die Länge der Äste repräsentiert die Distanz zwischen den beiden Proben. Diese beiden Proben werden aus der Distanzmatrix entfernt und als ein einzelnes Pseudo-Item wieder hinzugefügt. Der nächste Zyklus beginnt wieder mit der Berechnung der Korrelationskoeffizienten und der Distanzmatrix der nun veränderten Liste. Wieder wird das Probenpaar mit der geringsten Distanz ermittelt, dieses kann aus zwei einzelnen Proben bestehen oder aus einer Probe und einem Pseudo-Item. Über einen Knoten wird das Paar miteinander verbunden und als neues Pseudo-Item behandelt. Dieser Prozess wird so lange wiederholt, bis eine einzelne Probe übrigbleibt. Dabei werden immer mehr Proben zu Pseudo-Items zusammengefasst, die aus unterschiedlich vielen Proben bestehen können. Zum Zusammenfügen der Proben, also zum eigentlichen hierarchischen Clustern bietet das Programm drei Optionen an:

Average Linkage Clustering, Complete Linkage Clustering und Single Linkage Clustering. Die Unterschiede in den Algorithmen bestehen in der Berechnung der Distanzen und der Erstellung der Distanzmatrix.

Weitere Clustermethoden sind die „Self-organizing Maps“ (Tamayo, 1999) oder das „*k*-Means Clustern“. Diese Methoden eignen sich gut, komplexe Datensätze zu betrachten und profunde Muster der Genexpression aufzuspüren. Ein Anwendungsbeispiel liefert die Arbeit von R. Hoffmann *et al.* (Hoffmann, 2002), in der die Genexpression der murinen B-Zellreifungsreihe untersucht wurde. Die Genexpressionsprofile der verschiedenen Reifungsstadien von der prä-B-Zelle bis zur reifen B-Zelle wurden ermittelt und die Daten mit SOM geclustert. Auf diesem Wege konnten Gruppen von Genen identifiziert werden, die im zeitlichen Ablauf der Reifungsreihe einer parallelen Genregulation unterlagen. In der Anwendung haben *k*-Means Clustermethoden oder SOM jedoch den Nachteil, dass vom Nutzer selbst die Anzahl der zu generierenden Cluster festgelegt werden muss. Dies birgt die Schwierigkeit, dass eine biologisch angemessene Zahl an Gruppen im voraus kaum bekannt ist. Für die Betrachtung von nur zwei zellulären Zuständen, in diesem Fall „kranke Zelle“ und „gesunde Zelle“, eignen sich SOM als Clustermethode für Gene nicht sehr gut. Ein weiterer Nachteil der SOM ist, dass es keinen Hinweis auf die Verbindung oder den Grad der Ähnlichkeit zwischen einzelnen Clustern gibt.

Die „Principal Component Analysis“ (PCA) als weitere unüberwachte Methode erlaubt die Analyse und die Darstellung multidimensionaler Datensätze, die bei Genexpressionsanalysen aus  $m$  Beobachtungen (Anzahl der Proben) und  $n$  Variablen (Anzahl der untersuchten Gene auf einem Genechip) bestehen. Das Ziel der PCA ist dabei die Reduktion der Dimensionalität einer komplexen Datenmatrix durch die Identifikation neuer Variablen, den Principal Components, in denen Informationen konzentriert werden (Sturn, 2002). Die PCA erlaubt die Darstellung grundlegender Muster der Genexpression. Geringe Unterschiede in den Daten werden schwächer bewertet. Die Darstellung der Principal Components geschieht als Punkte in einem dreidimensionalen Koordinatensystem. Punkte, die in diesem Koordinatensystem nah beieinanderliegen, haben ähnliche Eigenschaften, beruhen daher auf ähnlichen Expressionsdaten. Punkte mit größerer Distanz zueinander sind sich nicht ähnlich und beruhen auf unterschiedlichen Expressionsdaten. Wird die PCA auf die untersuchten Proben angewandt, lassen sich Ähnlichkeiten in der Genexpression so leicht darstellen. Die Möglichkeit der PCA-Anwendung auf Genexpressionsdaten hat unter anderem N. Holter gezeigt (Holter, 2000). Bei seiner Analyse der Expressionsmuster akuter Leukämien konnte S. Armstrong mittels einer PCA zeigen, dass Patienten mit einer MLL-Translokation eine eigene Krankheitsentität bilden (Armstrong, 2001).

Alle unüberwachten Clustermethoden beinhalten zwei Schwierigkeiten, die bei der Interpretation der erhaltenen Ergebnisse berücksichtigt werden müssen. Erstens: die robuste Identifikation von Strukturen im Datensatz beinhaltet nicht notwendigerweise eine biologische Aussagekraft der Ergebnisse. So können Proben entsprechend der Datenqualität oder unterschiedlicher Präparationsmethoden geclustert werden. Diese Clusterergebnisse sind reproduzierbar und statistisch signifikant, hätten jedoch keine biologische Bedeutung. Die Ergebnisse sollten folglich mit anderen, experimentellen Methoden verifiziert werden. Zweitens: unüberwachte Clustermethoden eignen sich dazu, bislang unbekannte biologische Strukturen aufzudecken, wie zum Beispiel Untergruppen einer Krebserkrankung. Im Gegensatz dazu eignet sich diese Methode eher weniger dazu, Korrelationen zwischen der Expression einzelner Gene und beobachteten Unterschieden der Proben herzustellen, zum Beispiel bei der Differenzierung zwischen chemotherapieresistenten und chemotherapieempfindlichen Tumoren (Golub, 2001).

In der vorliegenden Arbeit wurde zur Bestimmung funktioneller Gruppen differentiell exprimierter Gene das hierarchische Clustern von M. Eisen verwendet (Eisen, 1998). Zur Untersuchung möglicher Untergruppen in den leukämischen Patientenproben wurde ebenfalls das hierarchische Clustern herangezogen, aber auch ein *k*-Means Clusterverfahren, welches freundlicherweise von Frau Dipl.-Inform. A. Thum, Institut für Informatik (Prof. Stefan Posch), Martin-Luther-Universität, erarbeitet worden ist. Für die PCA wurde das Programm „Genesis“ von A. Sturn verwendet (Sturn, 2002).

## 3 Ergebnisse

### 3.1 *Eigenschaften der verwendeten Proben*

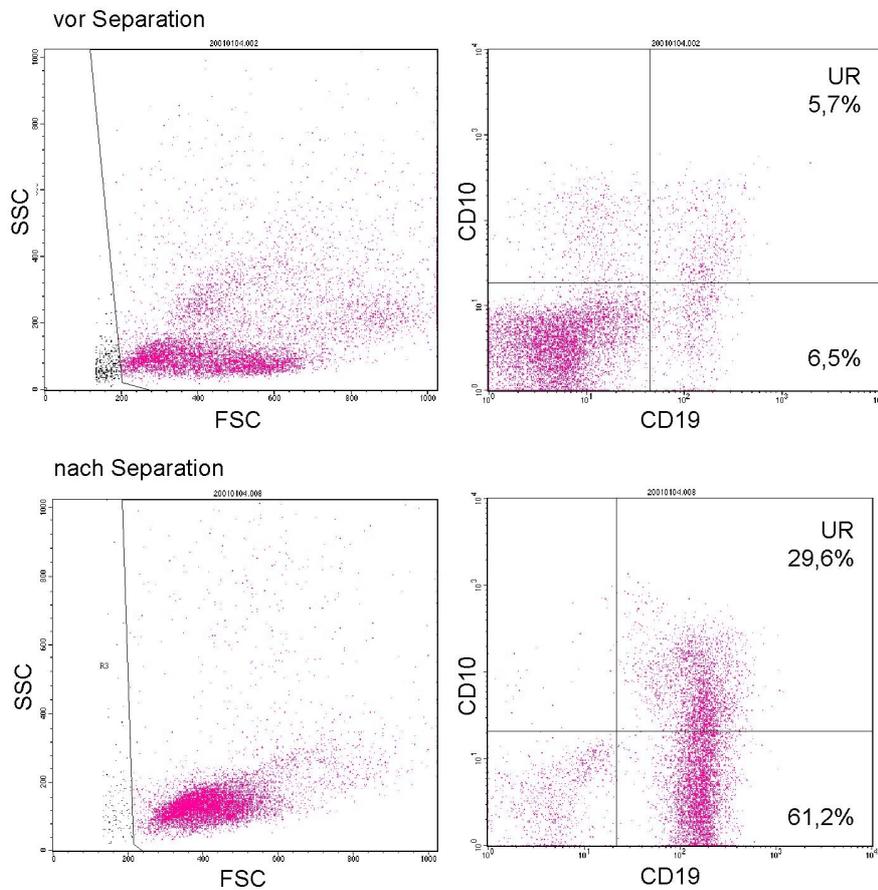
#### 3.1.1 Die Leukämieproben

Aus dem Knochenmark pädiatrischer Patienten mit einer cALL wurden zur Untersuchung der Genexpression lediglich die Lymphozyten mittels einer Dichtegradientenzentrifugation isoliert (vgl. 2.3.1). Dabei wurden zwischen  $1,0$  und  $10,0 \times 10^7$  Zellen isoliert. Die Prozentzahl der leukämischen Blasten im Knochenmark der Patienten lag zwischen 40% und 99%, wobei in 32 der 35 Proben über 90% Blasten nachgewiesen werden konnten. Aus diesem Grund wurde darauf verzichtet, die leukämischen Blasten mit zusätzlichen Methoden weiter anzureichern. Aus den Proben konnten zwischen  $1,8$  und  $87,4 \mu\text{g}$  Gesamt-RNA isoliert werden (vgl. 2.8.1 und 2.8.2). Ein direkter Zusammenhang zwischen der Anzahl der eingesetzten Zellen und der Menge daraus isolierter RNA lässt sich nicht erkennen. Aus der isolierten RNA wurden letztlich zwischen  $4,4$  und  $106,0 \mu\text{g}$  cRNA als Produkt der IVT gewonnen (vgl. 2.8.7). Aus der nachträglich hinzugenommenen Probe ALL58GAT wurden aus  $5 \times 10^6$  Zellen  $2,5 \mu\text{g}$  RNA gewonnen. Die Tabelle 6-3 im Anhang fasst die Eigenschaften aller verwendeten Proben zusammen.

#### 3.1.2 Die Kontrollproben

Das Nabelschnurblut gesunder reifgeborener Kinder wurde als Quelle für sortierte B-Lymphozyten als Kontrollgruppe herangezogen. Zum Zeitpunkt der Geburt gehören zum fetalen Blutkreislauf das Neugeborene selbst, die Nabelschnur und die fetale Seite der Plazenta. Nabelschnurblut ist leicht verfügbar, hat eine große Zelldichte und enthält mehr unreife Lymphozyten als peripheres Blut von Kindern oder Erwachsenen (Hoffkes, 1996; Rubinstein, 1995; Wu, 1999). Die Angaben in der Literatur über die Zusammensetzung des Nabelschnurblutes schwanken und sind mitunter aufgrund unterschiedlichen Studiendesigns schwer miteinander zu vergleichen. Für die frühen B-Zellen, die die Oberflächenantigene CD10 und CD19 exprimieren, werden Anteile von 0,5-1,5% aller Lymphozyten beschrieben. Die CD19 exprimierenden B-Zellen machen 12-14% der Lymphozyten aus (Calado, 1999; Comans-Bitter, 1997; O'Gorman, 1998). Unterschiede im Anteil der B-Lymphozyten im Nabelschnurblut ergeben sich aus der natürlichen Heterogenität menschlichen Untersuchungsmaterials. Mit großer Wahrscheinlichkeit haben Faktoren wie Gestationsalter, Ernährungszustand oder die

Art der Geburt einen Einfluss auf die Zusammensetzung des Blutes. Für diese Studie wurde nur das Nabelschnurblut von gesunden, reifgeborenen Kindern verwendet.



**Abbildung 3-1 Durchflusszytometrische Kontrolle der Zellseparation**

Durchflusszytometrische Kontrolle der Zellseparation am Beispiel der Nabelschnurblutprobe CB36NP, bei der B-Lymphozyten von ursprünglich 12,2% auf 90,8% angereichert werden konnten. Der Gehalt an B-Vorläuferzellen stieg dabei von 5,7% auf 29,6% an. Die Prozentzahlen geben den Anteil an lebenden Zellen in den Proben an. SSC = Seitwärts-Scatter, FSC = Vorwärts-Scatter.

Vor der Zellseparation wurden über die Dichtegradientenzentrifugation (vgl. 2.3.1) bereits die meisten Thrombozyten, Granulozyten und Erythrozyten aus den Proben entfernt. Durch die sich anschließende Zellseparation reduzierte sich die Zellzahl in den Proben durchschnittlich von  $11,6 \pm 3,8 \times 10^7$  Leukozyten auf  $1,1 \pm 0,4 \times 10^7$  Leukozyten. Dabei wurde der Anteil CD19 exprimierender B-Lymphozyten von vorher  $10,8 \pm 5,7\%$  auf  $79,3 \pm 10,1\%$  angereichert. Durch die Separation wurden die CD19/CD10 doppelt-positiven frühen B-Lymphozyten von  $4,0 \pm 2,2\%$  auf  $25,2 \pm 6,8\%$  angereichert. Das verwendete System zur Zellseparation (vgl. 2.6) enthält Antikörper, die T-Zellen und NK-Zellen (anti-CD2- und anti-CD4-Antikörper), Granulozyten und Monozyten (anti-CD11b-Antikörper), Monozyten und Thrombozyten (anti-CD16- und anti-CD36-Antikörper), Mastzellen und Basophile Granulozyten (IgE-Antikörper) binden und über

die magnetische Markierung aus der Zellsuspension entfernen. Damit bleiben lediglich die nicht markierten B-Lymphozyten zur weiteren Bearbeitung zurück. Der Anteil der Nicht-B-Lymphozyten (CD45+/CD19-) an den in der Probe vorhandenen Leukozyten reduzierte sich von ursprünglich  $88,9\pm 6,4\%$  auf  $20,6\pm 9,9\%$ . Als Beispiel zeigt die Abbildung 3-1 die durchflusszytometrische Kontrolle der Zellseparation der Nabelschnurrestblutprobe CB36NP, bei der B-Lymphozyten von ursprünglich 12,2% auf 90,8% angereichert werden konnten. Der Gehalt an B-Vorläuferzellen stieg dabei von 5,7% auf 29,6% an.

Die Kontamination der zur Genexpressionsanalyse herangezogenen Zellproben mit Nicht-B-Lymphozyten (CD45+/CD19-) geht wahrscheinlich auf Promyelozyten und Myelozyten im Nabelschnurblut zurück. Diese myeloischen Vorläuferzellen machen im Nabelschnurblut einen größeren Anteil aus als in vergleichbaren Quellen (peripheres Blut, fetales oder adultes Knochenmark) (Wu, 1999). Das verwendete System zur Anreicherung der B-Lymphozyten aus Nabelschnurblut ist eigentlich dafür ausgelegt, die B-Lymphozyten aus peripherem Blut Erwachsener zu isolieren, in dem keine myeloischen Vorläuferzellen nachzuweisen sind. Folglich enthält das System keine Antikörper zur spezifischen Markierung dieser Zellen. Diese Zellen werden also durch die verwendeten Antikörper zur Depletion nicht markiert und verbleiben als kontaminierende Zellen in der Probe.

Zu Beginn des Projektes wurden unterschiedliche Separationsstrategien zur Herstellung möglichst reiner CD19/CD10 doppelt-positiver Zellproben getestet (vgl. 2.6). Zum einen wurde versucht, die Zielzellen direkt mit fluoreszenzmarkierten Antikörpern gegen CD19 und CD10 zu markieren und durchflusszytometrisch zu sortieren, zum anderen wurde versucht, die Zellen in zwei aufeinanderfolgenden Schritten mit paramagnetisch markierten Antikörpern gegen CD19 und CD10 im Magnetfeld zu sortieren (vgl. 2.6). Beide Methoden einer Positivselektion der frühen B-Zellen lieferten nur unbefriedigende Ergebnisse. Die Reinheit der erhaltenen Proben war etwas größer als bei der schließlich verwendeten Negativselektion, aber die Ausbeute an Zellen war letztlich für die Genexpressionsanalyse zu gering. Aus diesem Grund wurde die unter 2.6 beschriebene Methode der Zellseparation verwendet. Lediglich die Kontrollproben CB\*1NP, CB\*2NP und die aus Tonsillen präparierten Kontrollproben TS01NP und TS02NP wurden auf dem Weg der Positivselektion mit markierten Antikörpern gegen CD19 und CD10 hergestellt. Für die TS Kontrollproben reduzierte sich die Zellzahl der Proben durch die Separation durchschnittlich von  $39,2\pm 17,2\times 10^7$  Leukozyten auf  $0,4\pm 0,02\times 10^7$  Leukozyten. Dabei wurden die CD19 exprimierenden B-Lymphozyten von vorher  $57,8\pm 5,0\%$  auf  $92,8\pm 1,4\%$  angereichert.

Durch die Separation wurden die Nicht-B-Zellen von  $39,6 \pm 2,4\%$  auf  $7,3 \pm 1,4\%$  abgereichert.

Aus den Proben konnten zwischen 8,6 und 106,1  $\mu\text{g}$  Gesamt-RNA isoliert werden (vgl. 2.8.1 und 2.8.2). Ein direkter Zusammenhang zwischen der Anzahl der eingesetzten Zellen und der Menge daraus isolierter RNA lässt sich auch hier nicht erkennen. Aus der isolierten RNA wurden letztlich zwischen 4,8 und 84,5  $\mu\text{g}$  cRNA als Produkt der *in vitro*-Transkription gewonnen (vgl. 2.8.7). Die Tabelle 6-4 im Anhang fasst die Eigenschaften aller verwendeten Kontrollproben zusammen.

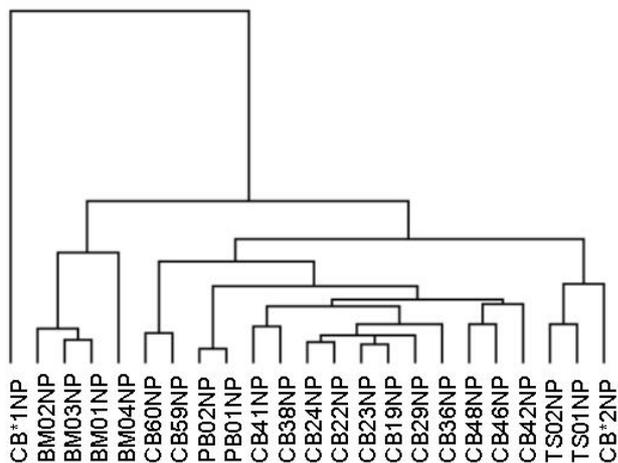
### 3.2 Qualitätskontrolle der Genechips und Ausschluss von Proben

#### 3.2.1 EOS-H Genechip

Auf dem EOS-H Genechip wurden insgesamt 16 cALL Proben (ALL01-ALL37) und 13 CB-Kontrollen sortierter B-Lymphozyten (CB19-CB60) sowie die zusätzlichen Kontrollen CB\*1 und CB\*2, die tonsillären B-Lymphozyten TS01 und TS02, die gesunden Knochenmarksproben BM01-BM04 und die beiden Proben peripheren Blutes PB01 und PB02 untersucht. Der Prozentsatz der als exprimiert detektierten Gene auf einem Genechip kann als ein Qualitätsmerkmal für die Hybridisierung herangezogen werden. Durchschnittlich wurden mit EOS-H 21,0% der auf den Genechips vorhandenen Gene detektiert. Dies entspricht für EOS-H Genechips einem mittleren Wert (EOS, persönliche Mitteilung). Lediglich die Probe ALL18YEM zeigte mit 1,5% detektierter Gene einen deutlich unterdurchschnittlichen Wert. Hier muss davon ausgegangen werden, dass entweder bei der Markierung der cRNA oder bei der Hybridisierung ein Arbeitsschritt nicht funktioniert hat. Die Probe ALL18YEM wurde daher von der weiteren Auswertung der Experimente ausgeschlossen. Die Tabelle 6-5 im Anhang gibt eine Zusammenfassung der Menge der detektierten Gene auf den Genechips.

Bei den Proben CB19NP bis CB60NP, CB1\*NP, CB2\*NP, TS01NP und TS\*02NP handelt es sich um Kontrollproben, deren RNA aus gesunden B-Lymphozyten gewonnen wurde. Jedoch stammen die Proben aus unterschiedlichen Quellen (Nabelschnurblut oder Tonsillen), und die B-Lymphozyten wurden mit unterschiedlichen Methoden präpariert (vgl. 2.6). Zur Beantwortung der Frage, ob sich diese unterschiedlichen Proben als einheitliche Kontrollgruppe zur Untersuchung der differentiellen Genexpression im Vergleich zu den cALL Proben eignen, wurden die Proben geclustert (vgl. 2.11.2.2). Das Clusterergebnis macht deutlich, dass es sich hier nicht um eine einheitliche Probengruppe handelt. Die Abbildung 3-2 zeigt das Ergebnis. Die beiden Proben CB1\*NP und CB2\*NP clustern nicht zusammen mit den anderen

ebenfalls aus Nabelschnurblut isolierten B-Lymphozyten. Es wird auch deutlich, dass diese beiden Proben selbst keine eigene Gruppe bilden. Die beiden aus Gaumenmandeln isolierten Proben hingegen bilden selbst eine Gruppe, clustern jedoch nicht mit den übrigen Proben sortierter B-Lymphozyten. Aufgrund dieser Befunde wird als Kontrollgruppe zur Identifikation differentiell exprimierter Gene in der cALL die Gruppe der Proben isolierter B-Lymphozyten aus Nabelschnurrestblut CB19NP bis CB60NP definiert. Die Abbildung macht deutlich, dass die verbleibenden Kontrollproben BM01NP bis BM04NP, PB01NP und PB02NP eigenständige Gruppen bilden. Die Ähnlichkeit zwischen peripherem Blut und Nabelschnurrestblut zeigt sich an der Tatsache, dass die PB-Proben als Gruppe zusammenbleiben, jedoch innerhalb der CB-Proben clustern. Die Kontrollproben aus gesundem Knochenmark hingegen werden eindeutig von den anderen Proben getrennt und bilden ein eigenes Cluster.



**Abbildung 3-2 Hierarchisches Clustern der Kontrollen auf dem EOS-H Genechip**

Die mit EOS-H untersuchten Kontrollen clustern in zwei Gruppen: Knochenmark (BM) und B-Lymphozyten. Die Proben CB\*1 und CB\*2 lassen sich nicht sinnvoll zuordnen. Die aus Tonsillen gewonnenen B-Lymphozyten (TS) verhalten sich anders als die aus Nabelschnurblut isolierten B-Zellen (CB) und stehen isoliert.

### 3.2.2 HG-U133A Genechip

Auf dem Affymetrix HG-U133A wurden 29 cALL-Proben (ALL25-ALL57) und acht CB-Kontrollen sortierter B-Lymphozyten (CB36-CB60) untersucht, dazu die nachträglich untersuchte Probe ALL58GAT. Es wurden 28,4% der Gene auf dem Genechip vor und 31,8% nach der Antikörperverstärkung detektiert (vgl. 2.10.3). Dies macht den gewünschten Effekt der Signalverstärkung sichtbar.

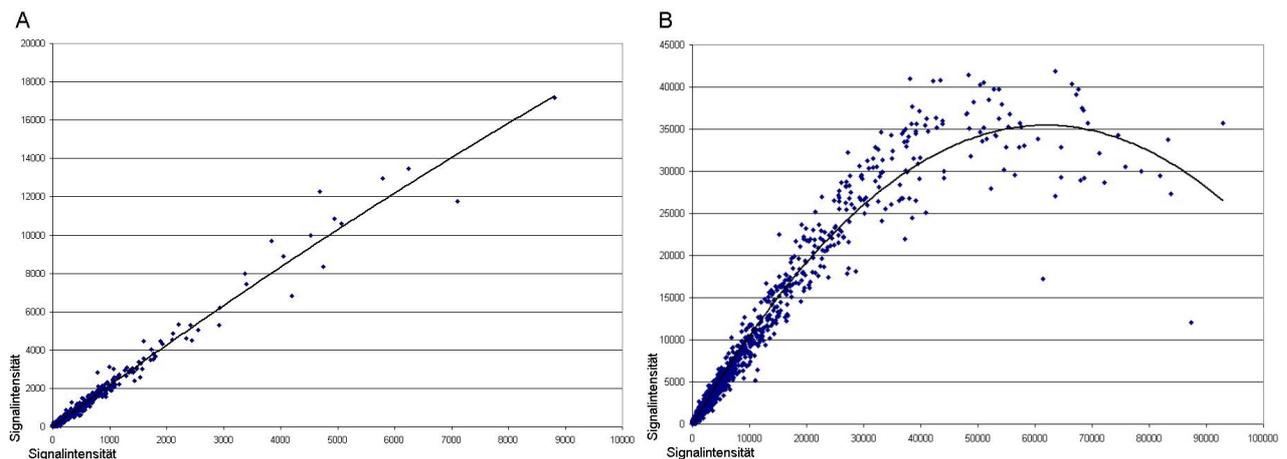
Zur Qualitätskontrolle der Experimente befinden sich auf den Genechips Sondensätze für vier konstitutiv exprimierte Gene: ISGF3 (Interferon-Stimulated Transcription Factor 3), 18S rRNA Gen, GAPDH (Glyceraldehyde-3-Phosphate Dehydrogenase) und ACTB (Beta-Actin). Diese Sonden detektieren Transkripte von

Genen, die üblicherweise in der Zelle konstitutiv exprimiert sind. Diese Transkripte dienen hier als interne Kontrolle zur Überprüfung der RNA Qualität der eingesetzten Proben. Die entsprechenden SONDENSÄTZE sind jeweils spezifisch für das 3'-, das 5'-Ende und die Mitte des Transkriptes gestaltet (die experimentellen SONDENSÄTZE zeigen keine Spezifität für bestimmte Bereiche der Transkripte). Unterschiede im Verhältnis der Signalintensitäten von 3'- zu 5'-Ende der konstitutiv exprimierten Haushalts-Gene, die größer als der Faktor 3 sind, deuten auf eine Degradation der Probe hin. Die Kontrollsonden der konstitutiv exprimierten Gene informieren somit über die Qualität der Reversen Transkription (vgl. 2.8.4) und der IVT (vgl. 2.8.7) der Probenbearbeitung. Eigentümlicherweise gab es regelmäßig bei ISGF3 einen Hinweis auf Degradation, auch wenn die übrigen Kontrolltranskripte in Ordnung erschienen.

Zur Hybridisierungslösung werden Eucaryotic Hybridization Controls (Spike Controls) hinzugegeben (vgl. 2.10.3). Hierbei handelt es sich um vier synthetisch hergestellte, prokaryotische Transkripte, die der Probe in definierten Mengen zugesetzt werden. BioB, BioC und BioDn sind bakterielle Gene aus *Escherichia coli* und CreX ist eine Rekombinase aus dem Bakteriophagen P1 (Endkonzentration in der Hybridisierungslösung: BioB 1,5 pM, BioC 5 pM, BioD 25 pM und CreX 100 pM). Diese Transkripte repräsentieren prokaryotische Gene, die im Menschen nicht existieren, deren SONDENSÄTZE als Kontrolle aber auf den Genechips vorhanden sind. Diese exogenen Transkripte mit ihren definiert eingesetzten Konzentrationen ermöglichen eine Qualitätskontrolle der Arbeitsschritte Hybridisierung, Waschen und Färben (vgl. 2.10.3 und 2.10.4). BioB wird in der geringsten Konzentration eingesetzt und stellt die untere Nachweisgrenze des Systems dar. BioB konnte auf allen Genechips detektiert werden. Unter Berücksichtigung aller dieser Qualitätsmerkmale konnten sämtliche HG-U133A Genechips weiter ausgewertet werden.

Betrachtet man die gemessenen Signalintensitäten aller Gene eines Genechips im Detail, so fällt auf, dass nach Antikörperverstärkung einige der Signale den Sättigungsbereich des Scanners erreichen. Diese Signale können vom Scanner nicht mehr korrekt detektiert werden. Da in diesem Sättigungsbereich die Messwerte nicht mehr der Menge an hybridisiertem Transkript und auch nicht der initial isolierten RNA entsprechen, werden die Ergebnisse an dieser Stelle verfälscht. In der Abbildung 3-3-A sind alle Signale eines Genechips vor und nach der Antikörperverstärkung gegeneinander aufgetragen worden (diese Daten wurden freundlicherweise von Frau Dr. Gesine Hansen zur Verfügung gestellt und sind nicht Bestandteil dieser Arbeit). Man sieht deutlich den linearen Zusammenhang zwischen den Messwerten und die Verstärkung der Signale, die ungefähr den Faktor 2 ausmacht. Die Abbildung 3-3-B zeigt als Beispiel die Leukämieprobe ALL42ORK der hier untersuchten Proben. Es wird

deutlich, dass kein linearer Zusammenhang mehr zwischen den Signalen vor und nach der Antikörperverstärkung besteht, was durch das eingefügte Regressionspolynom angezeigt wird. Als Effekt der Sättigung werden den Signalen nach Antikörperverstärkung Werte zugeordnet, die teilweise kleiner sind als die ursprünglichen Werte vor der Antikörperverstärkung. Aus diesem Grund musste auf die Auswertung der Ergebnisse aller Genechips nach Antikörperverstärkung verzichtet werden.



**Abbildung 3-3 Antikörperverstärkung der Signale**

A: Die Verstärkung der Signale mit fluoreszenzmarkierten Antikörpern macht ungefähr den Faktor zwei aus. Die Messwerte liegen auch nach der Antikörperverstärkung noch alle im Detektionsbereich des Scanners. Die Abhängigkeit zwischen den Signalen vor und nach der Antikörperverstärkung ist annähernd linear. Auf der Abszisse sind die Signalwerte eines Genechips vor, auf der Ordinate die entsprechenden Werte desselben Genechips nach Antikörperverstärkung aufgetragen. Die Intensitäten der Signale sind dimensionslose Zahlen. B: Die Verstärkung der gemessenen Signale mit fluoreszenzmarkierten Antikörpern lässt sich in diesem Beispiel nicht bewerten. Die Messwerte liegen nach der Antikörperverstärkung außerhalb des Detektionsbereichs des Scanners. Die lineare Abhängigkeit besteht nicht mehr und es ergeben sich nicht interpretierbare Signalwerte.

Um einen Vergleich der Genechips untereinander zu ermöglichen, wurden diese normalisiert und auf eine Targetintensität von 500 skaliert (vgl. 2.11.1.3). Die hierzu erforderlichen Skalierungsfaktoren lagen zwischen 2,3 und 42,0. Laut Angaben von Affymetrix ist ein Vergleich verschiedener Genechips jedoch nur dann korrekt möglich, wenn sich die Skalierungsfaktoren um höchstens den Faktor 5 unterscheiden. Bei zu großen Skalierungsfaktoren besteht die Gefahr, dass der lineare Zusammenhang zwischen der auf dem Genechip hybridisierten RNA und dem gemessenen Fluoreszenzsignal verloren geht. Besonders schwache Signale würden unverhältnismäßig stark hervorgehoben und verfälschten somit das Ergebnis. Für sechs der Genechips sind bei der Skalierung auf die gewählte Targetintensität Faktoren erforderlich gewesen, die größer als 11,5 ( $2,3 \times 5$ ) waren. Aus diesem Grund mussten zwei CB-Kontrollen (CB36NP und CB38NP) sowie vier cALL-Proben (ALL28DEM,

ALL29NIR, ALL30MOH und ALL32HCS) von der weiteren Auswertung ausgeschlossen werden. Die Tabelle 6-5 im Anhang gibt eine Zusammenfassung der Qualitätsmerkmale der analysierten Genechips und zeigt, dass die Genechips, die mit den größten Skalierungsfaktoren normalisiert wurden, die geringste Menge detektierter Gene aufweisen.

Die nachträglich untersuchte Probe ALL58GAT zeigte eine Detektion von 43,8% der Gene. Der Skalierungsfaktor lag mit 24,8 sehr hoch. Allerdings ist zwischen der Untersuchung der eigentlichen Proben der Studie und dieser zusätzlichen Probe die Sensitivität des Scanners aus wartungstechnischen Gründen verändert worden. Die Vergleichbarkeit zweier Experimente, die mit unterschiedlichen Scannersensitivitäten gemessen wurden, ist eingeschränkt. Es ist zu bemerken, dass die Probe ALL58GAT als Rezidiv eines Lymphoms nicht in die Analyse der Genexpression eingeflossen ist, sondern lediglich als Modellfall zur Überprüfung der erlangten Ergebnisse dient.

Eine gesonderte Definition der Kontrollgruppe entfällt hier, da auf HG-U133A nur cALL-Proben und sortierte B-Lymphozyten aus Nabelschnurrestblut untersucht wurden.

### 3.3 *Expressionsprofile typischer differenzierungsrelevanter Marker*

Die Leukozyten tragen auf ihrer Oberfläche charakteristische Antigene, die sich mit fluoreszenzmarkierten Antikörpern in der Durchflusszytometrie nachweisen lassen (vgl. 2.7). Einige dieser Antigene werden in Kombination mit anderen Antigenen exklusiv auf den Zellen bestimmter Entwicklungslinien exprimiert und können zur Identifikation herangezogen werden. Ebenso werden diese differenzierungsrelevanten Oberflächenantigene (Lineage Marker) zur Diagnose leukämischer Zellen herangezogen (vgl. 1.2). Genauso wie in der Durchflusszytometrie, wo das exprimierte Protein nachgewiesen wird, ist die Genexpressionsanalyse, hier wird die mRNA-Ebene betrachtet, dazu in der Lage, die Expression der Lineage Marker aufzuzeigen. Die folgenden Beispiele zeigen Ergebnisse der mit HG-U133A untersuchten Proben. Mit den EOS-H Genechips werden vergleichbare Ergebnisse erzielt.

Die beiden wichtigsten Marker zur Identifikation der cALL sind die Metallopeptidase CALLA (CD10) und die Terminale Deoxynukleotidyltransferase (TdT). Aus den Patientendaten ist zu ersehen, dass beide Marker bei allen Patienten nachgewiesen werden konnten. In der Genexpressionsanalyse zeigt sich ein entsprechendes Bild (Abbildung 3-4). Im oberen Teil der Abbildung ist im Blockdiagramm die Signalintensität beider Moleküle für alle untersuchten Proben dargestellt. Im Unteren Teil der Abbildung steht „P“ (present) für „exprimiert“, „M“ (marginal) für „geringfügig exprimiert“ und „A“ (absent) für „nicht exprimiert“. Diese Bezeichnungen werden vom Affymetrix Programm MAS automatisch ausgegeben, wobei die unterschiedlichen Intensitäten der PM- und





die jeweiligen Sondensätze, die ein Gen oder EST detektieren, lediglich aus PM-Sonden. Hingegen gibt es auf dem Affymetrix HG-U133A Genechip zu jedem PM-Sondensatz auch einen Satz an MM-Sonden. Das bedeutet, dass bei der Auswahl der Sondensequenzen für die Herstellung beider Genechiptypen unterschiedliche Kriterien angewandt wurden, was sich in unterschiedlichen Hybridisierungseigenschaften niederschlägt. Es ist folglich nicht möglich, ein Signal für ein bestimmtes Gen auf dem einen Genechiptyp mit dem Signal für das gleiche Gen auf dem anderen Genechiptyp in direkte Beziehung zueinander zu setzen. Es lassen sich demnach nur die Genechips eines Typs miteinander vergleichen und auswerten. Erst die Ergebnisse aus diesen beiden getrennten Analysen lassen sich dann inhaltlich miteinander vergleichen.

Auf den beiden Genechiptypen befinden sich Sondensätze für bekannte Gene, hypothetische Proteine und EST Cluster. Die Sonden sind jeweils so ausgewählt, dass nur die ein Protein kodierenden Sequenzen abgefragt werden. Entsprechend der Untersuchung des Transkriptoms, also der Genexpression auf mRNA-Ebene, werden andere chromosomale Sequenzen, die nach heutiger Auffassung ebenso zu einem Gen gehören (zum Beispiel Promotoren), nicht untersucht. Strenggenommen untersucht die funktionelle Genomik weder die Gene noch die Proteine, sondern das Intermediat des exprimierten Gens, die mRNA-Transkripte. Zur Vereinfachung sollen in der Folge alle von den Sondensätzen abgefragten Zielsequenzen als „Transkript“ oder „Gen“ mit jeweils synonymem Bedeutung bezeichnet werden. Die Genbezeichnungen folgen, soweit das möglich ist, den systematisierten Bezeichnungen der OMIM-Datenbank (vgl. 3.6.2). Gebräuchliche Synonyme werden darüber hinaus auch genannt. Zur verwendeten Nomenklatur siehe Kapitel 6.3.

#### 3.4.1 EOS-H Genechip

Die Identifikation differentiell exprimierter Gene geschah mit statistischen Methoden. Durch die Normalisierung der Daten (vgl. 2.11.1.1) erhielten die besonders schwachen Signale negative Vorzeichen. Nach Rücksprache mit dem Kooperationspartner EOS, wurde allen Werten, die kleiner als 50 sind, der Wert 50 zugeordnet. In diesem Bereich befindet man sich an der unteren Nachweisgrenze des Systems. Dieser Wert ist dann der Grenzwert, ab dem ein Gen auf EOS-H als exprimiert (present, P) verstanden wird. Signale, die kleiner oder gleich 50 sind, werden als nicht exprimiert (absent, A) gewertet.

Gene, die in der cALL Gruppe oder der Kontrollgruppe gesunder B-Lymphozyten aus Nabelschnurrestblut als differentiell exprimiert benannt werden sollen, müssen die folgenden Kriterien erfüllen: Die Mittelwerte für ein Gen aus beiden Gruppen sollen sich mindestens um den Faktor 2 unterscheiden (Fold Change, FC>2).

Expressionsunterschiede, die kleiner als zweifach sind, haben wahrscheinlich keine biologische Bedeutung für den Phänotyp einer Zelle, wenngleich sich Gene finden lassen, deren  $FC < 2$  ist, dieser dennoch statistisch signifikant ist mit  $p < 0,01$  (Glynn, 2000b). Weiterhin sollte das Ergebnis des  $t$ -Tests bei  $p < 0,01$  liegen. Insgesamt folgen diese Suchkriterien einer bereits veröffentlichten Arbeit, bei der mit EOS-H Genechips Unterschiede in der Genexpression zwischen ruhenden und aktivierten B-Lymphozyten untersucht worden sind. Der dort verwendete Grenzwert für den FC von 1,7 wurde hier mit 2,0 allerdings etwas strenger gefasst.

Mit diesen Kriterien werden 99 Sondensätze (maximaler FC 12,6) identifiziert, die in den cALL-Proben stärker exprimiert sind als in den Kontrollen. Diese 99 Sondensätze detektieren 95 unterschiedliche Gene und ESTs. Weiterhin werden so 263 Sondensätze (maximaler FC 18,7) identifiziert, die 256 unterschiedliche Gene und ESTs detektieren, welche in den cALL-Proben schwächer exprimiert sind als in den Kontrollen.

#### 3.4.2 HG-U133A Genechip

Bei der Normalisierung der Daten der HG-U133A Genechips mit der Affymetrix MAS Software treten keine negativen Werte auf. Weiterhin liefert dieses Programm bereits eine Kennzeichnung dafür, ob ein Gen als exprimiert (P) oder als nicht exprimiert (A) betrachtet wird. Mittels des Wilcoxon-Rangsummentests wird dabei überprüft, ob sich die PM- und die MM-Werte eines Sondensatzes signifikant unterscheiden. Über einen Grenzwert  $\alpha$  für den p-Wert wird dann festgelegt, ob das Gen exprimiert ist oder nicht. Auf die Hilfskonstruktion, alle Werte auf den Wert 50 zu setzen, wenn sie kleiner als 50 waren, wie bei den EOS-H Microarrays geschehen, konnte somit verzichtet werden.

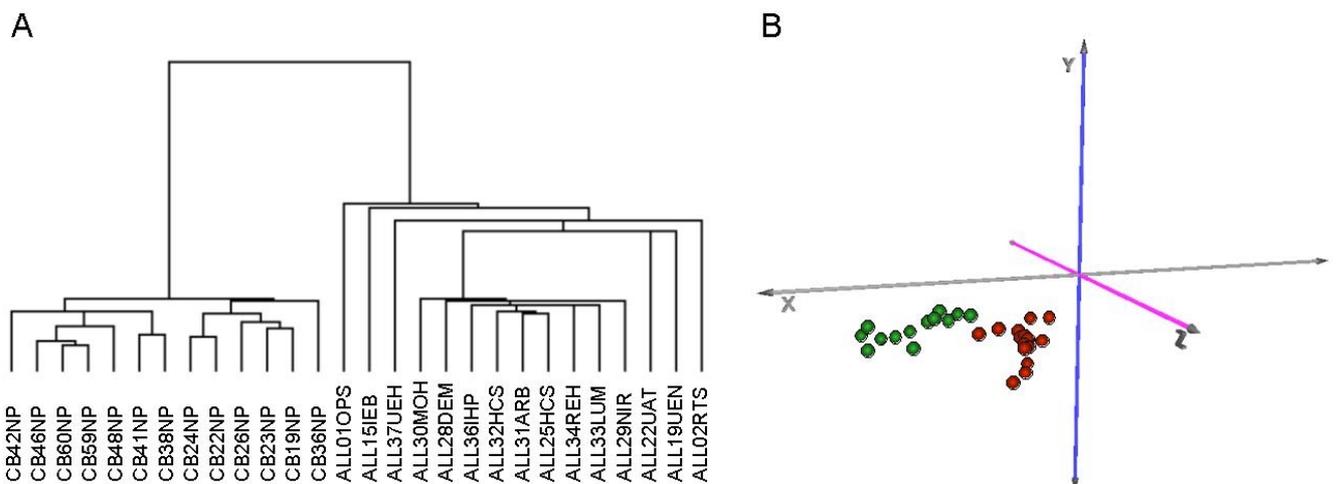
Gene, die in der cALL-Gruppe oder der Kontrollgruppe gesunder B-Lymphozyten aus Nabelschnurrestblut als differentiell exprimiert benannt werden sollen, müssen die folgenden Kriterien erfüllen: Die Mittelwerte für ein Gen aus beiden Gruppen sollten sich mindestens um den Faktor 2 unterscheiden ( $FC > 2$ ). Weiterhin sollte das Ergebnis des  $t$ -Tests bei  $p < 0,01$  liegen.

Mit den beiden oben aufgeführten Kriterien erhält man 575 Gene, die in den Leukämieproben stärker exprimiert sind (maximaler FC 257,0), und 1.002 Gene, die dort schwächer exprimiert sind (maximaler FC 129,5). Nicht zuletzt vor dem Hintergrund der Fragestellung nach einer Suche von differentiell exprimierten Genen, die sich als Zielstrukturen für Therapie oder Diagnose der cALL eignen, kann ein weiter einschränkendes Kriterium eingeführt werden. Ein differentiell exprimiertes Gen muss in wenigstens der Hälfte der Proben einer Gruppe (z.B. Leukämie) exprimiert sein und darf gleichzeitig in höchstens einer Probe der anderen Gruppe (z.B. Kontrolle)

exprimiert sein. (Wird dieses Kriterium enger gefasst - keine Expression in sämtlichen Proben der einen Gruppe bei gleichzeitiger Expression in allen Proben der anderen Gruppe – findet man keine differentiell exprimierte Gene. Dies spiegelt die individuellen Eigenschaften der Proben und die natürliche Varianz biologischer Untersuchungsmaterialien wider.) Auf diese Weise werden im HG-U133A Datensatz 132 Sondensätze, die 126 verschiedene Gene und ESTs detektieren, identifiziert, die ausschließlich in den cALL-Proben exprimiert sind. Weiterhin werden 138 Sondensätze für 135 Gene und ESTs identifiziert, die in den cALL-Proben nicht nachzuweisen sind. Diese 270 Sondensätze repräsentieren in stärkerem Maße die Gene, die exklusiv in der Gruppe der Leukämien exprimiert oder nicht exprimiert werden. Das heißt allerdings nicht, dass diese Transkripte die größten FC-Werte oder die größten Signifikanzen für die Unterschiede aufwiesen. Dieser eingeschränkte Datensatz lässt vielleicht weniger Aussagen zur Pathobiologie der cALL zu, weil geringe aber biologisch Relevante Genexpressionsunterschiede unberücksichtigt bleiben. Dafür finden sich hier vermutlich die besseren Zielstrukturen zum Einsatz in neuen therapeutischen Strategien.

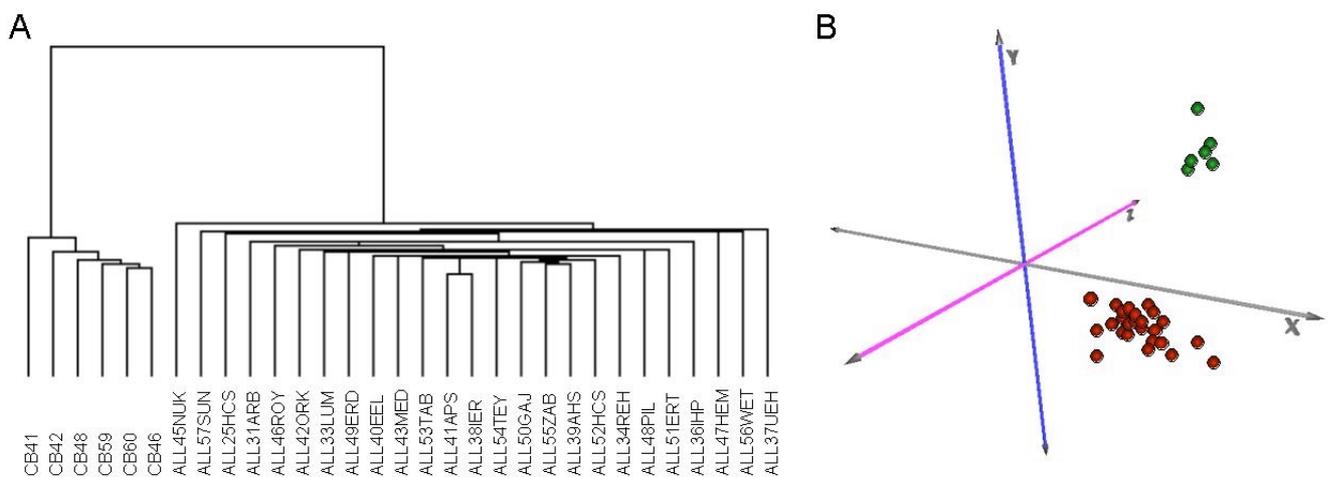
#### 3.4.3 Klassifikation der Proben mit differentiell exprimierten Markergenen

Mit den als differentiell exprimiert identifizierten Genen hat man einen Datensatz von Markergenen an der Hand, mit denen sich die Leukämien von gesunden Kontrollen unterscheiden lassen. Das Muster dieser herauf- und herunterregulierten Gene ist spezifisch für die cALL und lässt eine Zuordnung von Proben zur Leukämie- oder zur Kontrollgruppe zu. Zur Überprüfung dieser Frage wurden die Proben mit den beiden reduzierten Datensätzen differentiell exprimierter Gene klassifiziert. Dazu wird das Programm von M. Eisen verwendet (vgl. 2.11.2.2). Strenggenommen sind das hierarchische Clustern und die PCA keine Methoden zum Klassifizieren. Die mit diesen Programmen erzielten Ergebnisse deuten aber eindrücklich an, dass ein Klassifizieren der Proben mit den ermittelten Daten korrekt möglich ist. Beim Klassifizieren werden die Proben entsprechend ihren Merkmalsausprägungen, das sind hier die differentiell exprimierten Gene, sortiert. In diesem speziellen Fall werden als Ergebnis des Klassifizierens die beiden Klassen „cALL“ oder „CB-Kontrolle“ erwartet. Abbildung 3-6 und Abbildung 3-7 zeigen die beiden Ergebnisse des Klassifizierens und der PCA, jeweils mit den Daten der EOS-H und der HG-U133A Genechips. Das Resultat ist eindeutig. Mit Hilfe der 362 Gene von EOS-H und der 270 Gene von HG-U133A lassen sich jeweils alle Proben korrekt den Klassen der Leukämien oder der gesunden Kontrollen zuordnen. Insbesondere in Abbildung 3-7-B werden die Unterschiede zwischen den leukämischen und den gesunden Proben deutlich.



**Abbildung 3-6 EOS-H: Klassifikation und PCA mit ausgewählten Genen**

A. Beim Klassifizieren der Proben mit ausgewählten Genen, die einen Unterschied zwischen cALL-Proben und Kontrollen beschreiben und die mit EOS-H ermittelt wurden, trennen sich die Kontrollen und die cALL-Proben in zwei eindeutige Gruppen im Dendrogramm. B: Auch die PCA verdeutlicht, dass sich mit den ausgewählten Genen die Proben in die beiden Gruppen Leukämie (rot) und Kontrolle (grün) aufteilen lassen.

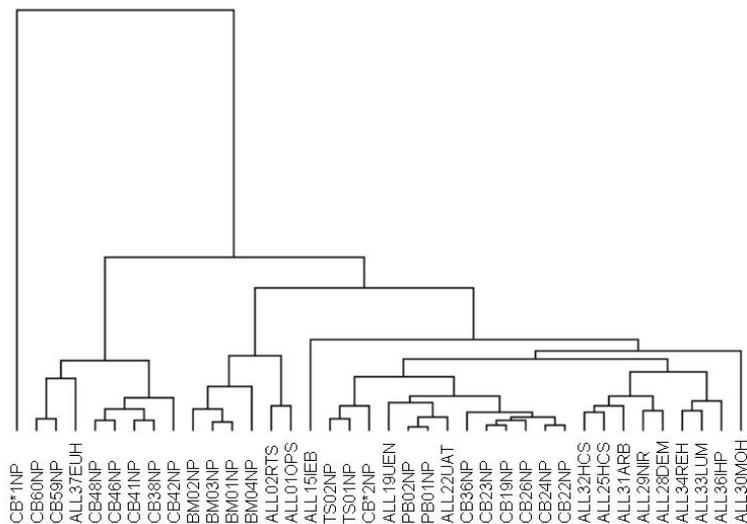


**Abbildung 3-7 HG-U133A: Klassifikation und PCA mit ausgewählten Genen**

A. Beim Klassifizieren der Proben mit ausgewählten Genen, die einen Unterschied zwischen cALL-Proben und Kontrollen beschreiben und die mit HG-U133A ermittelt wurden, trennen sich die Kontrollen und die cALL-Proben in zwei eindeutige Gruppen im Dendrogramm. B: Auch die PCA verdeutlicht, dass sich mit den ausgewählten Genen die Proben in die zwei Gruppen Leukämie (rot) und Kontrolle (grün) aufteilen lassen.

Versucht man hingegen, mit den Informationen aller Gene, die sich auf den Genechips befinden, zu clustern, so lassen sich die Gruppen nicht mehr eindeutig voneinander trennen. Die Abbildung 3-8 zeigt beispielhaft ein Experiment unter Verwendung aller SONDENSÄTZE des EOS-H Genechips. Die verschiedenen Leukämieproben werden nicht zusammen in eine Gruppe, sondern teilweise zwischen

die Kontrollen sortiert. Auch lassen sich die Kontrollen aus Knochenmark nicht mehr von den übrigen Kontrollen eindeutig absondern (vgl. Abbildung 3-2).

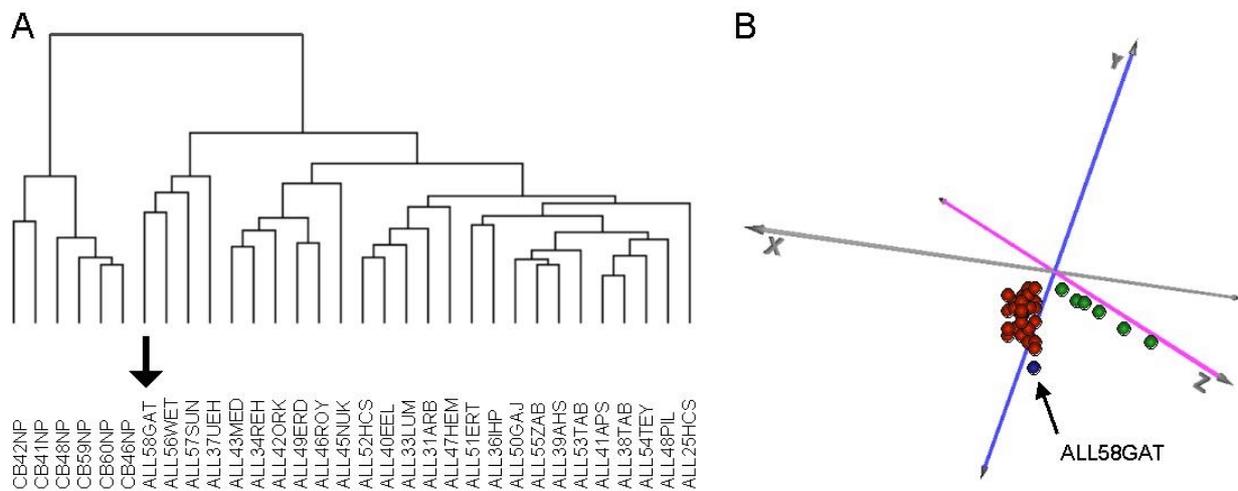


**Abbildung 3-8 Clustern mit allen Genen**

Am Beispiel des EOS-H Genechips wird gezeigt, dass bei Verwendung sämtlicher SONDENSÄTZE zum Clustern, sich die Proben nicht mehr eindeutigen Gruppen zuordnen lassen. Einzelne Leukämieproben werden zu den Kontrollen sortiert.

Die Probe ALL58GAT wurde nachträglich analysiert. Bei dieser Probe handelt es sich um die einzige Probe eines Patienten mit einem Rezidiv eines B-Zell-Lymphoms. Die Lymphoblasten wurden aus dem Liquor isoliert und auf HG-U133A untersucht. Die Daten sind nicht in die Analyse der differentiellen Genexpression zwischen cALL Blasten und gesunden B-Lymphozyten eingeflossen, da es sich definitionsgemäß nicht um eine Leukämie handelt. Aufgrund der verwandtschaftlichen Nähe von cALL und B-Zell-Lymphom kann diese Probe als eine Form der Kontrolle für die Spezifität des ermittelten Genexpressionsmusters der cALL angesehen werden. Verwendet man die mit HG-U133A ermittelten 270 differentiell exprimierte Gene zum Klassifizieren der Leukämieproben und der Kontrollen, unter Hinzunahme der Probe ALL58GAT, so werden, wie bereits gezeigt, Leukämien und Kontrollen eindeutig in zwei getrennte Gruppen eingeteilt, was die Abbildung 3-9-A verdeutlicht. Die neue Probe ALL58GAT wird dabei entsprechend der Verwandtschaft zur Leukämie korrekt zu den übrigen leukämischen Proben sortiert. Dies liefert einen Hinweis darauf, dass die ausgewählten Gene dazu geeignet sind, vorher unbekannte Proben richtig zu klassifizieren und zwischen Leukämie und gesunder Probe zu unterscheiden. Wichtiger noch, zeigt das korrekte Clustern der Lymphomprobe auch auf, wie nah verwandt diese beiden Krankheiten sind. Dass es sich bei dieser Probe allerdings nicht um eine Leukämie handelt, wird aus der PCA ersichtlich (Abbildung 3-9-B). Diese Probe steht im

dreidimensionalem Koordinatensystem in unmittelbarer Nähe zu den leukämischen Proben, ist aber ebenso deutlich von diesen abgegrenzt.



**Abbildung 3-9 Klassifikation und PCA mit der Lymphomprobe ALL58GAT**

A: Mit den differentiell exprimierten Genen, die zwischen cALL und gesunden B-Lymphozyten unterscheiden, lässt sich auch eine vorher unbekannte Probe (Pfeil), in diesem Fall eine Lymphom-Probe, „korrekt“ den Leukämien zuordnen. B: In der PCA bilden die Leukämien (rot) und die Kontrollen (grün) zwei eindeutig unterscheidbare Gruppen. Die Lymphomprobe (blau) steht in unmittelbarer Nähe zu den Leukämien, ist aber ebenso eindeutig von diesen abgegrenzt.

### 3.5 Genexpressionsprofile ermittelt mit unterschiedlichen Genechips

#### 3.5.1 Gegenüberstellung der Ergebnisse von EOS-H und HG-U133A

Das Genexpressionsprofil der Leukämiegruppe unterscheidet sich eindeutig vom Profil der Kontrollgruppe sortierter gesunder B-Lymphozyten aus Nabelschnurblut. Mit Hilfe der beiden unterschiedlichen Genechips konnten unter Verwendung derselben Kriterien 362 Gene auf EOS-H beziehungsweise 1577 Gene auf HG-U133A identifiziert werden, in denen sich die kranken von den gesunden Zellen unterscheiden. Im folgenden werden diese beiden ermittelten Datensätze differentiell exprimierter Gene miteinander verglichen, da, wie bereits oben erwähnt, eine gemeinsame Auswertung beider mit unterschiedlichen Genechips ermittelten Datensätze aufgrund des unterschiedlichen Chipdesigns nicht möglich ist.

Auf den Genechips wird jeder einzelne SONDENSATZ, der ein Gen oder EST detektieren soll, durch eine eindeutige Identifikationsnummer (UNIQUID) gekennzeichnet. Des Weiteren wird jedes Gen durch eine Beschreibung und durch Sequenzbezeichnungen (Accession-Number) gekennzeichnet. Diese Bezeichnungen entstammen den Datenbanken, aus denen die Sequenzen entnommen wurden und die die Grundlage für die Gestaltung des jeweiligen SONDENSATZES sind. Wie das folgende

Beispiel für CAP18 zeigt, sind die Angaben beider Genechips nicht immer direkt in Beziehung zueinander zu setzen:

**Tabelle 3-1 Unterschiedliche Bezeichnungen für dasselbe Gen auf EOS-H und HG-U133A**

Genechip	UNIQUID	Beschreibung	Accession-No.
EOS-H	EOS32481	cathelicidin antimicrobial peptide	AA251812
HG-U133A	210244_at	antimicrobial LPS-binding protein CAP18	NM_004345

Um feststellen zu können, welche Sondensätze auf den jeweiligen Genechips tatsächlich dieselben Gene detektieren, war es in vielen Fällen notwendig, die Ursprungssequenzen direkt miteinander zu vergleichen. Hierzu wurden die Datenbanken und Programme (zum Beispiel BLASTN, zum Vergleichen zweier Nukleotidsequenzen) der US-amerikanischen Internetseite <http://www.ncbi.nlm.gov> des National Center for Biotechnology Information (NCBI) verwendet. In manchen Fällen stimmten bereits die Beschreibungen oder die Accession-Number überein, was eine schnelle und eindeutige Zuordnung zuließ.

Von den 362 differentiell exprimierten Genen, die mit EOS-H identifiziert worden waren, wurden 159 Sondensätze auch mit HG-U133A identifiziert. Für 91 Sondensätze gab es auf HG-U133A keine entsprechenden Sondensätze. Diese können exklusiv nur mit EOS-H detektiert werden. Für 112 Sondensätze hingegen wird mit HG-U133A ein anderes Ergebnis erzielt. Diese Gene zeigen mit den Affymetrix Genechips keine differentielle Expression. Es handelt sich hierbei aber nicht um direkt widersprüchliche Ergebnisse in dem Sinne, dass ein Gen als heraufreguliert im einen Fall und im anderen Fall als herunterreguliert detektiert wurde. Für den Vergleich in umgekehrter Richtung zeigt sich, dass von den 1577 mit HG-U133A als differentiell exprimiert detektierten Sondensätzen 195 auch mit EOS-H detektiert wurden. Für 288 Sondensätze gab es auf EOS-H keine Entsprechungen und 1094 Sondensätze schließlich wurden auf HG-U133A als differentiell exprimiert detektiert, auf EOS-H jedoch nicht. Korrigiert man diese Zahlen um die Sondensätze, die exklusiv auf nur einem der beiden Chiptypen vorhanden sind, lassen sich die Ergebnisse gegenüberstellen. Für die mit EOS-H ermittelten Ergebnisse sind 58,7% identisch mit denen der HG-U133A Genechips. Für den umgekehrten Vergleich sind dies 15,1% identische Ergebnisse. Die Tabelle 3-2 fasst die Ergebnisse zusammen.

Die mit den beiden Genechiptypen gemessenen Signale lassen sich wegen der unterschiedlichen Hybridisierungseigenschaften der verschiedenen Sonden nicht direkt miteinander vergleichen. Ebenso ist eine direkte Gegenüberstellung der berechneten FC-Werte zwischen den beiden Genechiptypen aus demselben Grund nicht zulässig.

Die Ursachen für die Unterschiede in den Ergebnissen liegen in den unterschiedlichen Eigenschaften der Microarrays und in der unterschiedlichen Zusammensetzung der Untersuchungsgruppen. Bei all den hier festgestellten Unterschieden in den Ergebnissen, die mit den unterschiedlichen Chiptypen zustande gekommen sind, ist zu unterstreichen, dass es keine direkt widersprüchlichen Ergebnisse gibt. Man findet unter den differentiell exprimierten Genen der cALL keine, die auf dem einen Chiptyp als heraufreguliert und auf dem anderen als herunterreguliert bezeichnet wurden.

**Tabelle 3-2 Gegenüberstellung der mit beiden Genechiptypen ermittelten Ergebnisse**

Es wurden die Ergebnisse aller untersuchten Proben miteinander verglichen.

	EOS-H	HG-U133A
Anzahl detekierter Sondensätze differentiell exprimierter Gene	362	1577
Anzahl mit beiden Genechips detektierter Sondensätze	159	195
Anzahl mit nur einem Genechip detektierter Sondensätze	112	1094
Anzahl nur auf einem der beiden Genechips vorhandener Sondensätze	91	288
Bereinigte Anzahl detektierter Sondensätze*	271	1289
Prozent der mit beiden Genechips detektierten Sondensätze**	<b>58,7</b>	<b>15,1</b>
Prozent der mit nur einem Genechip detektierten Sondensätze**	<b>41,3</b>	<b>84,9</b>

\* nach Abzug der nur auf einem Genechip vorhandenen Sondensätze

\*\* Bezogen auf die bereinigte Anzahl detektierter Sondensätze

### 3.5.2 Analyse der auf beiden Genechips untersuchten Proben

Die unterschiedlichen Typen von Genechips, EOS-H und HG-U133A, liefern Ergebnisse, die nur teilweise identisch sind. Ein Grund hierfür liegt möglicherweise in den zum Teil verschiedenen Proben, die mit den entsprechenden Genechips untersucht worden sind. Bei der Größe der jeweils untersuchten Gruppen (16 cALL-Proben und 15 Nabelschnurblutkontrollen auf EOS-H, 29 cALL-Proben und acht Kontrollen auf HG-U133A) ist es denkbar, dass sich probenindividuelle Eigenschaften in den Ergebnissen bemerkbar machen. Ein anderer Grund können Genechip-spezifische Effekte sein. Um dies zu überprüfen wurden in zwei parallelen Ansätzen nur die Proben ausgewertet, die auf beiden Genechiptypen gemessen worden sind. Diese Untersuchung soll nicht der Identifikation von Genen zur inhaltlichen Auswertung dienen, sondern lediglich dem Vergleich der Ergebnisse, die mit verschiedenen Microarrays ermittelt wurden.

Zur Untersuchungsgruppe gehören: ALL25HCS, ALL31ARB, ALL33LUM, ALL34REH, ALL36IHP und ALL37UEH für die Leukämien, CB41NP, CB42NP, CB46NP, CB48NP, CB59NP und CB60NP für die Kontrollen. Dabei ist zu berücksichtigen, dass mehr als diese zwölf Proben auf beiden Genechiptypen

gemessen worden sind, aber wegen der Qualitätskontrolle von der Analyse ausgeschlossen wurden (vgl. 3.2). Beide Datensätze wurden unabhängig voneinander, aber mit demselben Kriterium untersucht, und die Ergebnisse miteinander verglichen. Gene, die als differentiell exprimiert bezeichnet werden sollen, müssen die folgenden Kriterien erfüllen:  $t$ -Test  $p < 0,01$  und Expressionsunterschied  $FC > 2$ . Auf EOS-H wurden somit 113 Gene identifiziert (FC 2,0–16,0), die in den Leukämien stärker als in den Kontrollen exprimiert sind, und 531 Gene (FC 2,0–35,2), die in den Leukämien schwächer exprimiert sind. Auf demselben Weg konnten auf HG-U133A 618 in den Leukämien relativ überexprimierte Gene (FC 2,0–67,3) und 800 relativ unterexprimierte Gene (FC 2,0–269,6) ausgemacht werden. Für den Vergleich beider Ergebnisse wurden die Gene nach absteigenden FC-Werten sortiert, getrennt nach über- und unterexprimierten Genen, und jeweils die 150 Gene mit dem größten Unterschied in der Expression weiter betrachtet. Das sind für EOS-H 113 + 150 Gene und für HG-U133A 150 + 150 Gene.

### **Tabelle 3-3 Gegenüberstellung der mit beiden Genechiptypen ermittelten Ergebnisse**

Es wurden nur die Ergebnisse derjenigen Proben miteinander verglichen, die mit beiden Genechiptypen untersucht wurden. Die Untersuchungsgruppen definieren sich wie folgt: ALL25HCS, ALL31ARB, ALL33LUM, ALL34REH, ALL36IHP und ALL37UEH für die Leukämien, CB41NP, CB42NP, CB46NP, CB48NP, CB59NP und CB60NP für die Kontrollen.

	EOS-H	HG-U133A
Anzahl detekierter Sondensätze differentiell exprimierter Gene	263	300
Anzahl mit beiden Genechips detektierter Sondensätze	47	52
Anzahl mit nur einem Genechip detektierter Sondensätze	185	183
Anzahl nur auf einem der beiden Genechips vorhandener Sondensätze	31	65
Bereinigte Anzahl detektierter Sondensätze*	232	235
Prozent der mit beiden Genechips detektierten Sondensätze**	<b>20,3</b>	<b>22,1</b>
Prozent der mit nur einem Genechip detektierten Sondensätze**	<b>79,7</b>	<b>77,9</b>

\* nach Abzug der nur auf einem Genechip vorhandenen Sondensätze

\*\* Bezogen auf die bereinigte Anzahl detektierter Sondensätze

Beim direkten Vergleich der mit den verschiedenen Chiptypen identifizierten Gene wird deutlich, dass die Unterschiede in den beiden Ergebnissen auch hier groß sind. Nur 47 Sondensätze auf EOS-H und 52 Sondensätze auf HG-U133A liefern auf beiden Genechips identische Ergebnisse. Für 31 Sondensätze auf EOS-H beziehungsweise 65 Sondensätze auf HG-U133A befinden sich keine entsprechenden Sondensätze zur Detektion auf dem jeweils anderen Genechip. Allerdings werden auf EOS-H 185 Sondensätze als differentiell exprimiert identifiziert, die auf HG-U133A keinen Unterschied in der Expressionsstärke zwischen Leukämie und Kontrollen zeigten. In der umgekehrten Betrachtung sind das 183 Sondensätze. Nach Abzug der

Sondensätze, die nur auf einem der beiden Genechiptypen vorhanden sind, sieht man, dass 20-22% der Ergebnisse (20,3% für EOS-H und 22,1% für HG-U133A) für beide Genechiptypen identisch sind. Die verbleibenden 78-80% stellen die Sondensätze dar, die mit auf nur einem der beiden Genechips als differentiell exprimiert detektiert wurden (vgl. Tabelle 3-3).

### 3.6 Funktionelle Gruppen differentiell exprimierter Gene

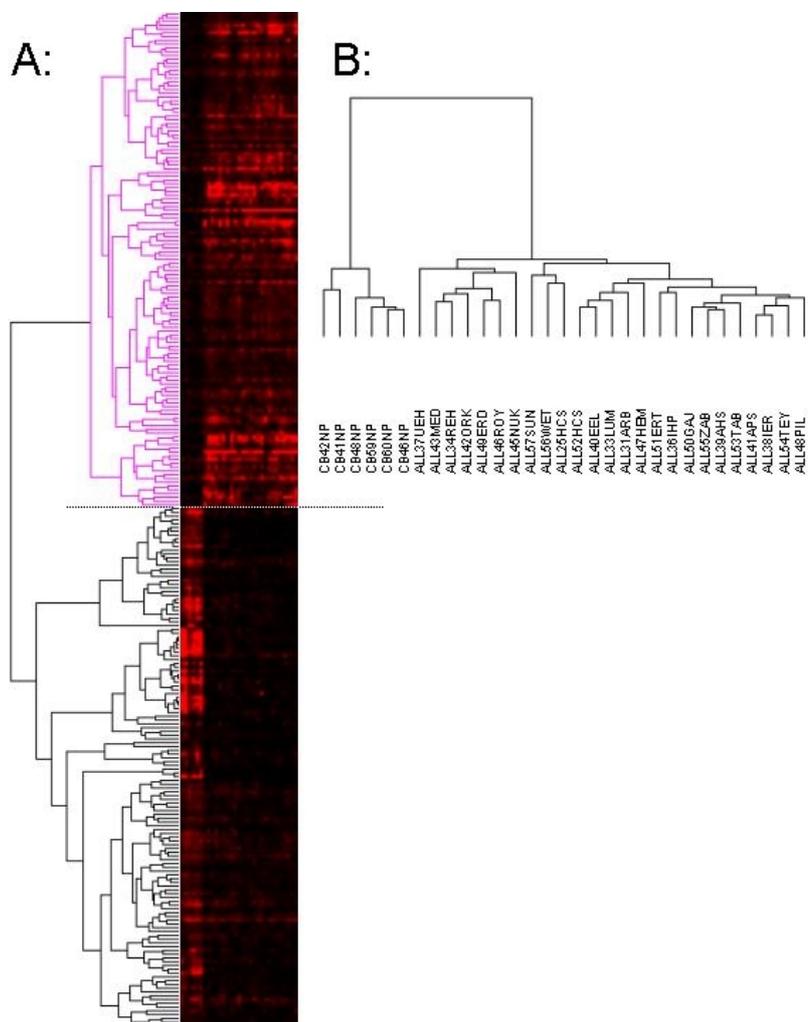
Die differentiell exprimierten Gene sind nicht nur in der Lage, zwischen krank und gesund zu unterscheiden (vgl. 3.4.3), sie sind auch dazu geeignet, Aussagen über die Pathologie der cALL zu machen. Gene, die in einem funktionellen oder regulatorischen Zusammenhang zueinander stehen, können vergleichbare Expressionsmuster und Expressionsstärken haben. Umgekehrt sollte es demnach möglich sein, aus gleichen Expressionsmustern verschiedener Gene funktionelle oder regulatorische Zusammenhänge abzuleiten, die bislang nicht bekannt waren. In einer bereits oben erwähnten Arbeit wurde (Hoffmann, 2002) mit Hilfe von Clusteranalysen gezeigt, welche Gene im Verlauf der Reifungsreihe von der murinen prä-B-Zelle bis zur reifen B-Zelle ähnliche Expressionsmuster aufweisen. Es ist also naheliegend zu untersuchen, in welchem funktionellen Zusammenhang die differentiell exprimierten Gene stehen, die den Unterschied zwischen leukämischen Blasten und gesunden B-Lymphozyten ausmachen.

#### 3.6.1 Clustern der differentiell exprimierten Gene

Unter Verwendung von EOS-H Genechips wurden 362 differentiell exprimierte Gene identifiziert. Die Untersuchung mit den HG-U133A Genechips lieferte 1575, mit den stringenten Abfragekriterien 270 Gene. In zwei getrennten Untersuchungsansätzen wurden diese Gene mit Hilfe des hierarchischen Clusters geclustert, um funktionelle Gruppen innerhalb der Gene zu identifizieren. Das Programm sucht die Gene, die, über alle Proben betrachtet, ein ähnliches Expressionsmuster aufweisen, und ordnet diese in ein Cluster ein. Die Abbildung 3-10 zeigt als Beispiel das Ergebnis der mit HG-U133A identifizierten 270 Gene. Zwei Cluster sind ganz offensichtlich und leicht erkennbar: unten die in den Kontrollen stärker exprimierten Gene und oben die in den Leukämien stärker exprimierten Gene. Die Identifikation weiterer eindeutiger Cluster gestaltet sich indes schwieriger.

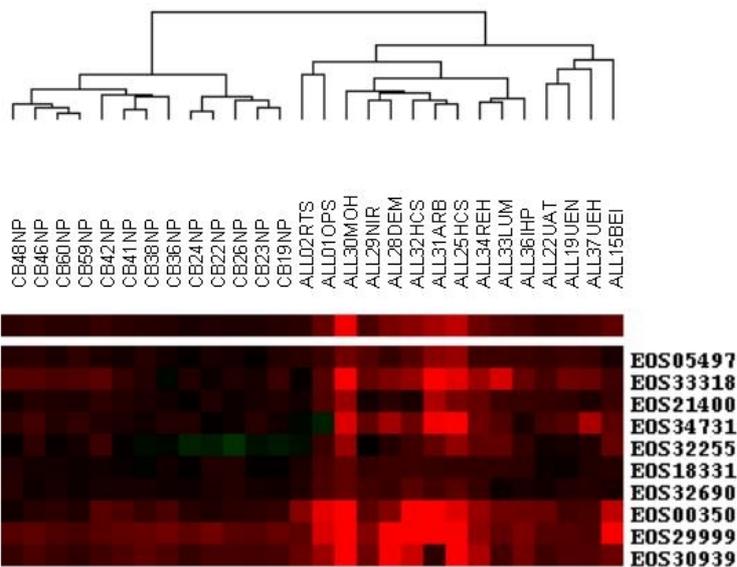
Beim verwendeten hierarchischen Clustern muß vom Anwender festgelegt werden, an welchen Stellen des Baums Cluster voneinander abgegrenzt werden können. Nach Festlegung von zehn in der Hierarchie des Clusterbaums übergeordneten und 23

untergeordneten Genclustern für EOS-H und von sechs übergeordneten und 15 untergeordneten Genclustern für HG-U133A müssen die einzelnen Mitglieder der jeweiligen Cluster näher betrachtet werden. Die Abbildung 3-11 zeigt als Beispiel ein Cluster von Genen des EOS-H Datensatzes. Neben drei ESTs finden sich in diesem Cluster unter anderem Gene für den Insulinrezeptor INSR, für Typ V Collagen, für die Cytochrom Reduktase Cyb5, für das Enzym  $\beta$ -Galaktose-Sulfotransferase, für ein nukleäres Ribonukleoprotein und für ein Protein, das in die p53-vermittelte Apoptose involviert ist. Offensichtlich lässt sich bei den Mitgliedern dieses Clusters kein funktioneller Zusammenhang feststellen. Bei der detaillierten Auswertung der übrigen Gencluster beider Datensätze lassen sich ebenfalls keine Aussagen über Verbindungen der Genprodukte zueinander machen.



**Abbildung 3-10 Gencluster**

A: Mit stringenten Kriterien wurden mit HG-U133A 270 Gene identifiziert, die eine eindeutige Unterscheidung zwischen cALL-Proben und Kontrollen zulassen. Diese differentiell exprimierten Gene clustern deutlich in zwei Gruppen. Oberhalb der punktierten Linie sind die in den Leukämien stärker exprimierten Gene, darunter die schwächer exprimierten Gene. B: Zur Ergänzung ist angegeben, wie die verschiedenen leukämischen Proben und Kontrollen mit dieser Genauswahl korrekt clustern.



**Abbildung 3-11 Einzelnes Gencluster**

Vergrößerung eines einzelnen Genclusters, ermittelt mit EOS-H Genechips. Die einzelnen Proben lassen sich mit den Genen korrekt klassifizieren. Die Gene in diesem Beispielcluster geben aber keine Hinweise auf eine funktionelle Zusammengehörigkeit.

### 3.6.2 Einteilung der Gene in funktionelle Gruppen

Die Ergebnisse des Clusters der Gene beider Genechips, wie unter Punkt 3.6.1 beschrieben, sind gleichermaßen wenig aufschlussreich für das Verständnis der cALL und lassen keine funktionellen Zusammenhänge von Genen innerhalb der Cluster erkennen. Inhaltlich, nach den jeweiligen bekannten Funktionen der durch sie kodierten Proteine, lassen sich die differentiell exprimierten Gene dennoch zu Gruppen zusammenfassen und erlauben so eine Analyse ihrer Rolle für die Pathologie oder den Phänotyp der cALL. Weiterhin können verfügbare Informationen über bereits bekannte Zusammenhänge einzelner Gene mit anderen Krebserkrankungen zum Verständnis ihrer Rolle bei der cALL beitragen ebenso wie Angaben über bestimmte Gewebe, in denen ihre Transkripte bereits nachgewiesen worden sind.

Neben veröffentlichten Publikationen stammen die meisten Informationen von der NetAffx Internetseite (<http://affymetrix.com>) oder aus der OMIM Datenbank (<http://www.ncbi.nlm.gov>). Die Seite NetAffx des Genechipherstellers Affymetrix liefert alle öffentlich verfügbaren Informationen zu den abgefragten Sequenzen, Nucleotiden und Proteinen (Liu, 2003). In der OMIM Datenbank (Online Mendelian Inheritance in Man) werden ebenfalls Informationen über Gene beziehungsweise Proteine und ihre zelluläre Funktion, ihre Gewebeverteilung, ihre Vererbungsmechanismen, ihre Mutationen und Polymorphismen, ihre biochemischen Eigenschaften und vor allem ihre

Bedeutung für Krankheiten gesammelt. Sofern in den folgenden Abschnitten Informationen der NetAffx oder der OMIM Datenbank verwendet wurden, sind diese als solche nicht weiter ausdrücklich mit der entsprechenden Quellenangabe gekennzeichnet worden.

Die in OMIM gespeicherten Daten geben bisweilen Aufschluss darüber, in welchen Geweben oder Zellen bestimmte Transkripte nachzuweisen sind. In der Regel handelt es sich hierbei um Ergebnisse von Northern-Blot Analysen. In einer großangelegten Studie (Su, 2002) wurden insgesamt 46 adulte und fetale humane Gewebe, Organe und normale und pathologisch veränderte Zelllinien auf ihr Genexpressionsprofil hin untersucht. Zu den untersuchten Organen zählen unter anderem Bauchspeicheldrüse, Herz, Hoden, Leber, Lunge, Niere, Plazenta, Schilddrüse, verschiedene Gewebe des Zentralnervensystems, Blut und Knochenmark. Unter den betrachteten pathologischen Zelllinien befinden sich verschiedene Burkitt Lymphome, Melanome sowie eine chronisch myeloische Leukämie (K562) und eine akute lymphatische Leukämie (MOLT-4, T-Zell Leukämie). Für die Studie wurden die Affymetrix Genechips U95A verwendet. Zum einen wurden in diesem Ansatz für jedes Gen die Orte seiner Expression bestimmt, was bereits einer Kartierung des menschlichen Transkriptoms nahe kommt. Zum anderen haben die Autoren gewebespezifische Transkripte identifizieren können, die exklusiv in bestimmten Geweben exprimiert sind. Die öffentlich zugängliche GNF-Expressionsdatenbank (<http://expression.gnf.org>) des „Genomics Institute of the Novartis Research Foundation“ (GNF) ermöglicht nun eine umfassende Recherche danach, welche Transkripte in welchen Geweben nachzuweisen sind. Bei der Charakterisierung der in den cALL-Proben differentiell exprimierten Gene wurde neben OMIM und NetAffx auf diese Datenbank zurückgegriffen.

Bei Betrachtung aller differentiell exprimierten Gene lassen sich sieben funktionelle Gruppen festlegen. Die Definition dieser Gruppen geschah im Hinblick auf mögliche Zusammenhänge mit der Pathologie der cALL beziehungsweise in Hinblick auf zelluläre Eigenschaften von Leukämie- und Krebszellen. Die Schwerpunkte liegen dabei auf Wachstum, Entwicklung und Differenzierung von Zellen sowie auf der Immunologie. Das Ziel dieser Einteilung ist es, die Bedeutung einzelner Gene für die Pathologie der Krankheit besser zu verstehen oder funktionelle Zusammenhänge zwischen Genprodukten zu verdeutlichen. Es kann durchaus sinnvoll sein, bei anderen Fragestellungen andere Gruppeneinteilungen vorzunehmen. Die Zuordnung der Gene zu einer bestimmten Gruppe ist nicht immer zwingend die einzig mögliche. Entscheidend für die Zuordnung sollte schließlich die Bedeutung des jeweiligen Gens für die cALL sein. So kodiert zum Beispiel E2F5 einen Transkriptionsfaktor. Seine Bedeutung liegt jedoch in der Regulation des Zellzyklus'. Da es sich bei der cALL um

eine Erkrankung stark proliferierender Zellen handelt, bei denen die Kontrolle des Zellzyklus' gestört sein dürfte, ist die Zuordnung von E2F5 in die Gruppe „Regulation von Zellzyklus und Zellproliferation“ für das Verständnis der cALL aufschlussreicher als eine Zuordnung in die Gruppe „Regulation der Genexpression“.

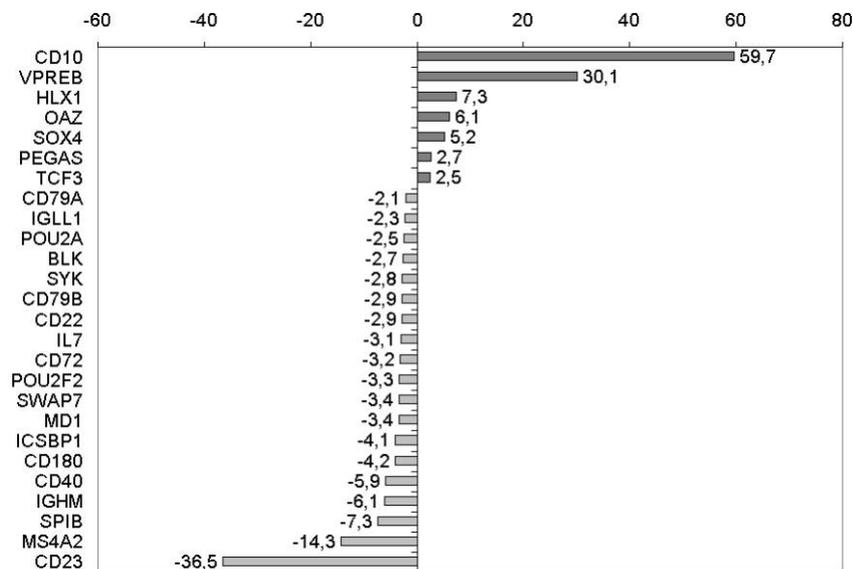
Im einzelnen lassen sich die Gene in die folgenden neun Gruppen sinnvoll einteilen: 1. Entwicklung von B-Lymphozyten, 2. Immunsystem und Inflammation, 3. Replikation von DNA, Transkription, Proteinsynthese, 4. Signaltransduktion und Regulation von Genexpression, 5. Regulation von Differenzierung und Entwicklung, 6. Regulation von Zellzyklus und Zellproliferation und 7. allgemeiner Metabolismus. Eine weitere Gruppe bildet keine funktionelle Einheit, sondern umfasst Gene, die bereits in Zusammenhang mit Leukämien oder anderen Krebserkrankungen beschrieben worden sind (hierbei kann es zu Überschneidungen mit den Gruppen 1–7 kommen). Eine letzte Gruppe fasst unter anderem die Gene zusammen, für die keine Informationen über die Funktion des von ihnen kodierten Proteins verfügbar sind. Hierunter fallen vor allem ESTs, hypothetische Proteine (KIAA1020, FLJ10246) und Sequenzen isolierter Proteine unbekannter Funktion, wie zum Beispiel ANXA4, das aus der Plazenta isoliert wurde. Den größten Anteil machen in der Tat die Gene und ESTs aus, zu deren Funktion es derzeit keine weiteren Informationen gibt.

In den folgenden Abschnitten sollen die Ergebnisse der Untersuchungen beider Genechiptypen inhaltlich gemeinsam betrachtet werden. Sie stehen gleichberechtigt nebeneinander. Das heißt im einzelnen wird bei den Genen nicht unterschieden, mit welchem der Arrays sie detektiert wurden. Die zusammenfassende Darstellung der Gene in den einzelnen funktionellen Gruppen ist nicht vollständig. Ein Teil der differentiell exprimierte Gene wird hier nicht näher erläutert, weil die verfügbaren Informationen vielleicht zu allgemein oder zu ungenau sind oder weil sich ein vernünftiger Bezug zur Fragestellung der Arbeit nicht herstellen lässt.

#### 3.6.2.1 Entwicklung von B-Lymphozyten

In dieser Gruppe kommt es zu Überschneidungen mit anderen Gruppen, da es sich hier strenggenommen nicht um eine funktionell einheitliche Gruppe handelt. Die Gene dieser Gruppe können, entsprechend ihrer Funktion, doppelt erwähnt sein. Im Hinblick auf die Pathogenese der cALL und den Zusammenhang mit einer gestörten Differenzierung der B-Lymphozyten erscheint es jedoch sinnvoll, diese Gene hier gemeinschaftlich zu betrachten. Mindestens sieben unterschiedliche Transkripte, die in den Zellen der cALL heraufreguliert sind haben eine direkte Verbindung zur B-Zellreifung. Es sind hingegen 18 Gene herunterreguliert. Ein Teil der herunterregulierten Gene ist auch in Wachstum und Aktivierung reifer B-Zellen involviert

(vgl. Abbildung 3-12). Weiterhin sind herunterreguliert 55 Transkripte, die verschiedene Bestandteile von Immunglobulinen kodieren. Ein Teil davon kodiert Segmente der variablen Immunglobulinketten, die während der B-Zellreifung umgelagert werden.



**Abbildung 3-12 Übersicht differentiell exprimierter Gene der B-Zellreifung**

Sechs Gene mit Funktionen in der B-Zellreifung sind in den Zellen der cALL mit durchschnittlichen FC-Werten von 2,5 bis 59,7 überexprimiert (positiver FC-Wert, dunkler Balken). Die übrigen dargestellten Gene zeigen einen Zusammenhang mit der Differenzierung von B-Zellen, sind aber in der cALL schwächer exprimiert als in den Kontrollen (negativer FC-Wert, heller Balken).

An erster Stelle der heraufregulierten Gene ist CD10 zu nennen, weil es ein charakteristisches Oberflächenantigen unreifer B-Lymphozyten und damit ein wichtiger Marker zur Diagnose der cALL ist. Diese als Neprysilin bezeichnete Metallopeptidase ist während der B-Zellreifung transient exprimiert und an der Aktivierung und Deaktivierung von Peptiden beteiligt, in dem sie endständige Aminosäuren dieser Peptide abspaltet (Bene, 1997). CD10 ist allerdings auch auf anderen gesunden Zellen nachzuweisen, die nicht lymphatischen Ursprungs sind.

VpreB ist ebenfalls ein Zelloberflächenmolekül und wird im gesunden Menschen nur in den B-Vorläuferzellen exprimiert. Es ist weder auf reifen B-Lymphozyten noch auf anderen Zellen des Blutes nachzuweisen und in den leukämischen Zellen stark exprimiert. Zusammen mit dem Protein Lambda-5 ( $\lambda 5$ , IGLL1) bildet VpreB die Surrogate Light Chain (SLC) des prä-B-Zellrezeptors unreifer B-Zellen, die im weiteren Verlauf der B-Zellreifung durch die leichte Kette des B-Zellrezeptors ersetzt wird. Der zweite Bestandteil dieser SLC, IGLL1, ist zwar in allen leukämischen Proben anwesend, im Vergleich mit den Kontrollen aber deutlich schwächer exprimiert. Mit dem dritten Bestandteil der B-Zellrezeptoren, der schweren Kette IGHM, verhält es sich

entsprechend. Auf dem HG-U133A Genechip befinden sich fünf Sondensätze, die Bereiche von IGHM detektieren. Für alle fünf Sondensätze zeigte sich eine deutlich schwächere Expression in den leukämischen Proben. Als weitere Bestandteile des prä-B-Zell- und des B-Zellrezeptors gelten die Oberflächenantigene CD79a (Ig $\alpha$ ) und CD79b (Ig $\beta$ ). Diese beiden Rezeptoren sind in den leukämischen Zellen ebenfalls nur schwach exprimiert. Es wird deutlich, wie die einzelnen Bestandteile des prä-B-Zellrezeptors in den leukämischen Zellen unterschiedlich exprimiert werden.

TdT ist die Terminale Deoxynukleotidyltransferase, die die Insertion von Nukleotiden spezifisch an den Verbindungsstellen der Immunglobulingensegmente V(H)-D und D-J(H) während der Umlagerung dieser Gene im Verlauf der B-Zellreifung katalysiert. TdT ist nur in unreifen B-Zellen, in Thymozyten und in leukämischen Zellen nachzuweisen. Die charakteristische Überexpression von TdT in den Leukämieblasten, die auch mit den Genechips gezeigt werden konnte, wird bereits zur cALL-Diagnose herangezogen.

SOX4, PEGASUS, HLX1 und OAZ sind Gene für vier heraufregulierte Transkriptionsfaktoren der cALL. SOX4 kodiert einen lymphozytenspezifischen Aktivator der Transkription. Mäuse, denen SOX4 fehlt, zeigen Defekte in der B-Zellreifung, aber nicht in der Reifung von Erythrozyten, Myelozyten oder T-Lymphozyten (Melchers, 1998). Das Zinkfingerprotein PEGASUS gehört der Ikaros-Familie von Transkriptionsfaktoren an, die vorrangig in Lymphozyten exprimiert werden, und die eine essentielle Rolle bei der Reifung von Lymphozyten übernehmen. PEGASUS ist erst kürzlich beschrieben worden und scheint die Expression seiner Zielgene zu unterdrücken (Perdomo, 2000). Von HLX1 ist bekannt, dass es in hämatopoetischen Progenitorzellen exprimiert wird und dass es die Differenzierung der Progenitorzellen zu reifen hämatopoetischen Zellen reguliert. OAZ (OLF1/EBF-Assoziiertes Zinkfinger Protein) wird als Aktivator der Transkription beschrieben, dessen Expression in vielen Geweben nachgewiesen wurde. OAZ selbst wird als Antwort auf BMP2 oder BMP4 durch Mitglieder der MADH-Familie gesteuert. Es gibt weitere Hinweise auf die Beteiligung eines OAZ-Homologen an der Ausbildung der Posterior/Anterior-Achse in der Entwicklung von *Xenopus laevis*.

Mindestens vier Gene für Transkriptionsfaktoren, POU2F2, POU2AF1, ICSBP1 und SPIB, und für die Tyrosinkinase SYK, die alle an der Differenzierung der B-Zellen beteiligt sind, sind in den cALL-Zellen herunterreguliert. POU2F2 (OCT2, OTF2) ist ein wichtiger, aber nicht essentieller Faktor der B-Zellreifung und der Transkription von Immunglobulinen (Schubart, 2001) nach Stimulation der reifen B-Lymphozyten im peripheren Blut. POU2AF1 (OBF1) ist ein OCT2-Bindungsfaktor, der die Aktivität von OCT2 modulieren kann. OBF1-defiziente Mäuse sind überlebensfähig und haben eine normale B-Zellentwicklung, zeigen jedoch geringe Mengen an IgG im Blut und eine

geringere Zellzahl an B-Lymphozyten. SPIB zeigt eine Verwandtschaft zu den Transkriptionsfaktoren SPI1 und Pu.1, hat aber ein eigenes Transkriptionsmuster. Es wird nur in Lymphozyten, vorrangig in B-Zellen, exprimiert, und steigt während der B-Zellreifung in der Expression an, während es in T-Lymphozyten mit fortschreitender Differenzierung abnimmt. Es verfügt über eine ETS-Domäne und ist für die Entwicklung von B-Lymphozyten essentiell (Su, 1996). Von der Tyrosinkinase SYK ist bekannt, dass Mäuse, denen diese Kinase fehlt, einen Defekt in der B-Zellreifung aufweisen.

Weiterhin gehören zu den herunterregulierten Genen einige, die mit der Aktivierung reifer B-Zellen in Verbindung gebracht werden. Hierzu zählen der Burkitt Lymphom Rezeptor BLR1, der in den reifen Zellen für die Wanderung der Lymphozyten zu Milz oder Lymphknoten verantwortlich ist, oder CD180, das eine Rolle in der Aktivierung reifer B-Zellen spielt. Bei einer Genexpressionsanalyse, in der gesunde B-Lymphozyten mit Zellen der chronischen lymphoblastischen Leukämie untersucht wurden, konnte gezeigt werden, dass BLR1 auch in der CLL überexprimiert ist (Aalto, 2001). CD23, ein niedrigaffiner IgE-Rezeptor, und CD83, ein monozytenbindendes Zelladhäsionsmolekül, werden auch mit der Aktivierung ruhender B-Zellen in Verbindung gebracht. Dieses Oberflächenmolekül ist auf proliferierenden, aber nicht auf zirkulierenden B-Lymphozyten des Blutes nachweisbar, sowie auf dendritischen Zellen. CD83 scheint weiterhin Funktionen in unreifen B-Zellstadien einzunehmen. Bei einem Vergleich der Genexpressionsprofile von hämatopoetischen Stammzellen mit prä-B-Zellen zeigte sich CD83 als charakteristisch für die unreifen B-Lymphozyten (Muschen, 2002). In den hier untersuchten Leukämien ist die Expression verringert. CD22 ist ein Zelladhäsionsmolekül und zeigt mit drei unabhängigen Sondensätzen eine verringerte Expression in den leukämischen Blasten. Es ist in pro-B-Zellen als zytoplasmatisches Molekül nachzuweisen, bevor es in prä-B-Zellen an der Zelloberfläche erscheint. In den prä-B-Zellen ist CD22 an der Regulation der Signalweiterleitung über den B-Zellrezeptor beteiligt. Dieses Molekül gehört zu den Markern, die zur Diagnose der cALL herangezogen werden und in der durchflusszytometrischen Analyse nachzuweisen sein müssen. Dennoch scheint die Expression in den Leukämiezellen geringer zu sein. Bei vier Patienten ist CD22 auf den Genechips überhaupt nicht nachzuweisen. Das auf Ebene der Genexpression in den Blasten nicht nachzuweisende Zytokin IL7 schließlich ist für die Differenzierung der B-Vorläuferzellen bedeutend. Zusammen mit dem Wachstumsfaktor HGF als Kofaktor vermittelt IL7 ein Differenzierungssignal. HGF selbst zeigt keine veränderte Expression.

In die Gruppe der Gene mit Relevanz bei der B-Zellreifung gehören 55 Loci, deren Sonden nur auf HG-U133A vertreten sind und die in den leukämischen Proben nicht exprimiert sind, in den Kontrollen aber sehr wohl nachgewiesen werden können.

Hierbei handelt es sich um Segmente, die konstante Ketten von Immunglobulinen kodieren, und um verschiedene Segmente der variablen Regionen von Immunglobulinen, die bei der Ausbildung des Immunsystems umgelagert werden und somit die Diversität der Immunglobuline bedingen. Die Umlagerung dieser Gene auf chromosomaler Ebene ist ein fest regulierter Bestandteil der B-Zellreifungsreihe (vgl. 1.3 und Abbildung 1-4). Die fehlende Expression dieser Segmente deutet an, dass die Bildung der Immunglobuline unterbrochen ist und dass die Zellen in einem Stadium der B-Zellreifung arretiert sind und sich eventuell nicht weiter differenzieren können.

Von besonderem Interesse für die Reifung von B-Lymphozyten sind die gemeinschaftlich heraufregulierten Gene, die in Zusammenhang mit den Signaltransduktionswegen von TGF $\beta$  (Transforming Growth Factor) und BMP (Bone Morphogenetic Protein) stehen. Diese Transkripte werden im Abschnitt über die Signaltransduktion (3.6.2.4) detailliert betrachtet.

### 3.6.2.2 Immunsystem und Inflammation

Bemerkenswerterweise stehen einer Vielzahl herunterregulierter Gene mit einer immunologischen oder inflammatorischen Funktion nur sehr wenige heraufregulierte Gene gegenüber. Zunächst seien die im Vergleich relativ überexprimierten Gene genannt. Überexprimiert ist VAMP3, ein Mitglied der Synaptobrevin Genfamilie. VAMP3 ist in vielen Geweben nachweisbar und steuert unter anderem die Funktion von Thrombozyten. Der Koagulationsfaktor F13A1 ist ebenfalls in die Blutgerinnung involviert und praktisch nur in gesundem Vollblut und in der Plazenta nachweisbar. Der Makrophagen-Mannoserezeptor MRC1 ist in der Lage, Mannosereste an der Oberfläche potentiell pathogener Viren, Bakterien und Pilze zu binden, und hat vermutlich eine Funktion in der Bakterienabwehr. Das Expressionsmuster dieses Rezeptors ist nicht weiter beschrieben. Von PAG wird vermutet, dass es die Initiierung einer Immunantwort steuert und Lymphozyten in einem ruhenden Status halten kann. PAG ist nahezu ubiquitär exprimiert und zeigt die höchste Expression in Zellen des Immunsystems, in der Lunge, in Herz und Plazenta. Im Vergleich mit den Leukozyten der Kontrollen ist die Expression von PAG in den leukämischen Blasten weiter erhöht. Die Major Histocompatibility Complex (MHC) Antigene der Klasse II werden im Menschen auf B-Zellen, Makrophagen, aktivierten T-Zellen und auf weiteren antigenpräsentierenden Zellen (Langerhans Zellen oder Dendritische Zellen) nachgewiesen. Der Transkriptionsfaktor NF- $\kappa$ B ist für die Expression der MHC Klasse II Gene essentiell. Er setzt sich aus den beiden Untereinheiten NFYA und NFYB zusammen. In den leukämischen Zellen ist NFYA relativ überexprimiert, wohingegen das NFYB Transkript überhaupt nicht detektiert werden konnte. CD99 (MIC2) ist in den

leukämischen Blasten ebenfalls relativ überexprimiert. Von diesem Molekül mit möglichen Funktionen in Zelladhäsion und Signaltransduktion ist bekannt, dass es in prä-B-ALL hochreguliert sein kann.

Es sind auffallend sehr viel mehr immunologisch relevante Gene in den Blasten schwächer exprimiert als in den gesunden Kontrollzellen. Die von einigen dieser herunterregulierten Gene kodierten Proteine übernehmen Funktionen bei der Blutgerinnung, so zum Beispiel der Koagulationsfaktor F7, das thrombozyteninhibierende Protein VASP oder PGD2, das ebenfalls die Aggregation von Thrombozyten verhindern kann. Hier zeigt sich also, dass verschiedene Faktoren, die an der Blutgerinnung oder an der Steuerung der Thrombozytenfunktion beteiligt sind, in den Leukämiezellen herunter- oder auch heraufreguliert (F13A1) sein können.

Andere herunterregulierte Genprodukte sind direkt an entzündlichen Vorgängen beteiligt, wie zum Beispiel Lipocalin LCN2, das in fast allen Geweben, inklusive Leukozyten, nachzuweisen ist und das ein Modulator von Entzündungsreaktionen ist. ALOX5AP kodiert einen nicht näher charakterisierten Faktor zur Leukotriensynthese. Dieses wiederum ist in die verschiedenen entzündlichen Vorgänge bei Krankheiten wie Asthma, Arthritis oder Psoriasis involviert. Mit ANX1 ist ein antiinflammatorisch wirkendes Gen herunterreguliert. An Entzündungsstellen kommt es durch die Aktivität von antimikrobiell wirkenden neutrophilen Elastasen zu Gewebeschädigungen. Um diese Schädigungen einzugrenzen, wird ein entsprechender Elastase-Inhibitor ELANH2 gebildet, der in den Blasten ebenfalls fehlt.

An antimikrobiellen Immunreaktionen beteiligt sind unter anderem die herunterregulierten Transkripte für Lysozym LYZ, das hauptsächlich in Lungenmakrophagen und in der Milz nachzuweisende Cathepsin CTSS und PGLYRP, das an bakterielle Peptidoglykane bindet. Einige der Genprodukte in dieser Gruppe wirken regulatorisch oder modulatorisch auf Immunreaktionen. Das Adhäsionsmolekül ICAM3 ist essentiell für die Initiierung einer Immunantwort. Der Rezeptor SPAP1 bindet den Fc-Teil von Immunglobulinen und kann so Immunreaktionen modulieren. Eine ähnliche Funktion haben die IgE-Rezeptoren MS4A2 und MS4A3, die in die Familie der CD20-ähnlichen Rezeptoren gehören. MS4A2 ist an der Manifestation von allergischem Asthma bronchiale beteiligt. Lactotransferrin LTF ist ein eisenbindendes Molekül, welches den Eisenmetabolismus, die Hämatopoese und Immunreaktionen modulieren kann. Es ist in der Lage, das pathogene Potential von *Haemophilus influenzae* zu unterdrücken, indem es selektiv bestimmte Proteasen hemmt. Für die Aktivierung von B-Lymphozyten ist LYN erforderlich, das neben vielen adulten Geweben in den meisten fetalen Geweben nachzuweisen ist und das Homologie zum V-YES-1 Onkogen zeigt. LYN ist ein Marker der reifen B-Lymphozyten.

Der Burkitt Lymphom Rezeptor BLR1 und das Lymphozyten Adhäsionsmolekül SELL sind zwei Oberflächenrezeptoren, die die Wanderung und Lokalisierung von Lymphozyten in lymphatische Gewebe steuern. Beide sind in den Zellen der cALL herunterreguliert. SELL ist nicht nur entsprechend seiner Funktion auf Lymphozyten, sondern auch auf Neutrophilen Granulozyten und auf Monozyten nachzuweisen. Im vorliegenden Datensatz wird deutlich, dass SELL in praktisch allen Kontrollen unterschiedlicher Herkunft exprimiert wird, jedoch bis auf zwei Ausnahmen nicht in den Leukämien. Die Oberflächenantigene CD32, CD39 und CD55 sind in verschiedenen funktionellen Zusammenhängen an der Immunabwehr fremder Antigene beteiligt. HLA-DOB kodiert ein Oberflächenmolekül unbekannter Funktion und gehört in die Gruppe der MHC Klasse II Gene. Am Expressionsmuster für HLA-DOB ist auffällig, dass es im vorliegenden Datensatz ausschließlich in den B-Lymphozyten aus Nabelschnurrestblut exprimiert wird und nicht in den Leukämiezellen oder einer der anderen Kontrollen, die mit EOS-H untersucht worden sind. HLA-DQA1 und HLA-DQB1 sind zwei weitere herunterregulierte Antigene. Sie gehören zur HLA-DR Familie, der auf B-Zellen exprimierten HLA-Moleküle.

Lymphotoxin (LTB) und LY117 sind zwei immunmodulatorische Moleküle. Interessanterweise sind die genomischen Sequenzen der beiden Gene auf Chromosom 6p21 direkt benachbart. Der Panleukozytenmarker CD45 ist auf allen Leukozyten nachweisbar ebenso auf allen hier untersuchten leukämischen Blasten. Dennoch ist die Expression deutlich niedriger. Die Funktion von CD45 liegt zum einen in der Steuerung der Differenzierung von Leukozyten, zum anderen spielt CD45 eine Schlüsselrolle in der Aktivierung von Lymphozyten bei der Immunantwort. Zu diesem Zweck interagiert CD45 mit PTPRCAP, welches ebenfalls herunterreguliert ist. Die Bindung eines Antigens an den B-Zellrezeptor reifer B-Lymphozyten aktiviert intrazelluläre Proteinkinasen. Diese Aktivierung der B-Zellen mündet in den „class switch“ der schweren Ig-Ketten. Die konstante  $\mu$ -Region des IgM wird ersetzt durch konstante Regionen anderer Ig-Klassen,  $\gamma$  (IgG),  $\alpha$  (IgA) oder  $\epsilon$  (IgE). Swap70 wurde ursprünglich in der Maus isoliert und ist Bestandteil des Proteinkomplexes, der die Rekombination zwischen zwei Switch-Regionen unterstützt.

Eine Reihe von Genen mit immunologischen Funktionen, die alle in den leukämischen Proben nicht exprimiert und nur in bestimmten Kontrollen vorhanden sind, lassen sich gesondert zusammenfassen. Es handelt sich hierbei um Gene, die typisch für die frühen Zellstadien innerhalb der Granulozytenreifungsreihe sind. Neutrophile Granulozyten sind ein zentrales Element des Immunsystems der Neugeborenen in der Abwehr bakterieller Infektionen. Sie enthalten zytoplasmatische Granula mit antibiotisch wirksamen Proteinen und Peptiden (Levy, 1999). Die

Vorläuferzellen reifer Granulozyten sind später noch im Knochenmark nachzuweisen, wo sie, wie alle Leukozyten, ihre Reifung und Differenzierung durchmachen. Bei den hier identifizierten Transkripten handelt es sich um solche, aus denen insbesondere diese antimikrobiellen Proteine azurophiler Granula der Neutrophilen Granulozyten gebildet werden (Borregaard, 1997). Unter anderem sind dies Azurocidin AZU1, Myeloperoxidase MPO, Elastase ELA2, Defensin DEFA4, neutrophile Kollagenase MMP8 und die drei Calgranuline S100A8, S100A9 und S100A12. Betrachtet man die Expression dieser Gene in den mit EOS-H untersuchten Proben und Kontrollen, wird das besondere Muster deutlich. Die Transkripte sind lediglich in den Nabelschnurrestblut-Proben und in den Knochenmarksproben vorhanden. In beiden Quellen sind die frühen Stadien der Granulozyten nachzuweisen (Carr, 2000). In den Tonsillen, einem typischen sekundären lymphatischen Organ, und im peripheren Blut sind diese Zellen nicht anzutreffen. Dementsprechend konnte mit Hilfe der Genechips in diesen Proben keine Expression gezeigt werden. Bei den im Nabelschnurrestblut somit nachgewiesenen Granulozyten handelt es sich wahrscheinlich um eine zelluläre Kontamination, die daher rührt, dass das zur Zellseparation verwandte System (vgl. 2.6) zur Selektion von B-Lymphozyten aus adultem peripheren Blut gedacht ist. Die Auswahl der dort vorhandenen Antikörper ist nicht in der Lage, unreife Granulozyten zu depletieren. Sie verbleiben unmarkiert in der Fraktion der separierten B-Lymphozyten. Weiterhin fehlen in den Leukämieproben drei Ribonukleasen, Rnase2, Rnase3 und Rnase6, welche in den Granula der Eosinophilen Granulozyten vorkommen. Ihre ribonukleolytischen Eigenschaften werden besonders bei der Abwehr von Helminthen eingesetzt. Möglicherweise ist die Detektion dieser Ribonukleasen auch auf verbliebene Nicht-B-Zellen in den Kontrollproben zurückzuführen, wenngleich speziell Rnase6 fast ubiquitär nachgewiesen werden konnte.

Einige wenige der detektierten Gene, die in den Kontrollen überexprimiert sind, kodieren Proteine, die direkt mit T-Zellen in Verbindung stehen: CD1c und CD1d als Bestandteile des T-Zellrezeptors, CD103 für die Gewebeverteilung von T-Zellen, CD229 für die Adhäsion von T-Lymphozyten mit akzessorischen Zellen, SLAP als negativer Regulator von T-Zellreaktionen und SCYC2, ein Chemokinligand, der auf aktivierten T-Zellen nachzuweisen ist. Es ist möglich, dass die Detektion dieser Gene auf Reste ebenfalls kontaminierender T-Lymphozyten zurückzuführen ist.

In die Gruppe der in den Leukämien herunterregulierten Gene mit immunologischer Bedeutung gehören 55 Loci, deren Sonden nur auf HG-U133A vertreten sind. Hierbei handelt es sich um einzelne Segmente der variablen Regionen von Immunglobulinen, die bei der Ausbildung des Immunsystems umgelagert werden. Da die Umlagerung

dieser Gensegmente eine zentrale Rolle bei der Reifung der Lymphozyten besitzt, sind sie bereits in Abschnitt 3.6.2.1 besprochen worden.

### 3.6.2.3 Replikation von DNA, Transkription, Proteinsynthese

Zu den hier vertretenen Genen zählen diejenigen, welche Proteine kodieren, die an zellulären Prozessen der Replikation, Transkription, Translation oder an der Reparatur beschädigter chromosomaler DNA beteiligt sind. Es sind hier keine direkt regulatorischen Proteine der Genexpression, zum Beispiel Transkriptionsfaktoren, vertreten (vgl. dafür 3.6.2.4: Signaltransduktion und Regulation der Genexpression).

Die folgenden Gene sind in den cALL-Proben stärker exprimiert als in den Kontrollen. An der Replikation von DNA direkt beteiligt sind unter anderem TOP1, TOP2A, CHAF1 und SMARCA4. Die Topoisomerase TOP2A katalysiert die Entwindung doppelsträngiger superspiralisierter DNA Moleküle, wohingegen das Substrat für TOP1 einzelsträngige DNA ist. CHAF1A ist Bestandteil des aus drei Untereinheiten bestehenden Proteins CHAF, welches replizierte DNA *in vitro* mit Histonoktameren assembliert. H1F2, H1FX, H2AFO, H2AFY, H2BFL und H3FB kodieren unterschiedliche, heraufregulierte Histonproteine und bilden die Nukleosomenstrukturen chromosomaler DNA aus. SMARCA4 schließlich ist ein chromatinassoziiertes Protein mit unklarer Funktion.

Die Proteinkinase TTK ist in der cALL ebenfalls heraufreguliert und ist an der Duplikation der Centrosomen beteiligt und lässt sich besonders in den sich schnell teilenden Zellen des Thymus und der Hoden nachweisen. In der GNF-Expressionsdatenbank (vgl. 3.6.2) wird TTK als ein hodenspezifisches Transkript beschrieben. Eine Mutation im homologen Gen der Bäckerhefe führt zu Defekten in der Ausbildung der Spindelpolkkörperchen und zu Defekten der Zellzykluskontrolle. NEK2 (NIMA) ist ein Protein, das für den Eintritt der Zelle in die Mitose benötigt wird und während der Zellteilung mit den Zentrosomen assoziiert bleibt. MCM4 stabilisiert die DNA während der Mitose. Das zur SMC-Familie (Structural Maintenance of Chromosomes) gehörende CAPG ist mitbeteiligt an der Kondensation der Chromosomen während der Mitose. DEEPEST ist ein Molekül, welches an den mitotischen Spindelapparat assoziiert ist. Das ROK1 Genprodukt ist eine ATP-verbrauchende RNA-Helikase und wird zur RNA-Prozessierung benötigt, vor allem aber zum Fortschreiten des Zellzyklus'. KIN17 kodiert ein Zinkfingerprotein, welches bei der Reparatur geschädigter DNA eine Rolle spielt. Es ist ubiquitär exprimiert und zeigt besonders in Muskeln, im Herzen und im Hoden eine hohe Expression.

PCNA wird nur im Kern proliferierender Zellen nachgewiesen und hat daher eine vermutete Funktion während der Zellteilung. POLE und POLQ schließlich sind

Untereinheiten von DNA-Polymerasen, die in die Replikation und die Reparatur von DNA involviert sind. Die Überexpression dieser an sich ubiquitär exprimierten Gene stimmt gut mit der Tatsache überein, dass es sich bei den leukämischen Zellen um schnell proliferierende Zellen handelt. Laut GNF-Expressionsdatenbank sind alle diese Gene am stärksten in den Geweben exprimiert, die sich durch hohe Zellteilungsraten auszeichnen, zum Beispiel im Hoden und den in dieser Studie untersuchten Zelllinien maligner Erkrankungen. POLE und TOP2A sind darüber hinaus bereits in Rhabdomyosarkomen, einem Tumor der quergestreiften Muskulatur, mit Hilfe von Genechips als dort überexprimiert beschrieben worden (Khan, 1998).

Während bei den heraufregulierten Genen diejenigen vorherrschen, die an der Replikation beteiligt sind, stehen die nun folgenden herunterregulierten Gene eher im Zusammenhang mit Transkription und Translation, wenngleich vereinzelt Gene herunterreguliert sein können (POLD4), die im Zusammenhang mit der Replikation gesehen werden müssen. Sechs Gene sind herunterreguliert, die Strukturproteine der Ribosomen sind und die damit an der Proteinbiosynthese beteiligt sind. Diese ribosomalen Proteine sind RPL5, RPL8, RPL12, RPL13, RPL29 und RPL36. SLU7 kodiert einen Spleißfaktor. ADARB1 ist ein Protein der RNA-Editierung und RNGTT ist für die Prozessierung (Capping) des 5'-Endes neu entstehender prä-mRNA verantwortlich. ELL2 kodiert einen RNA-Polymerase Elongationsfaktor, der die katalytische Aktivität der RNA-Polymerase II steigert. Das PCF11 homologe Protein spaltet prä-mRNA. MGC15749 ist an der Metabolisierung von mRNA beteiligt und GARS an der Biosynthese von tRNA-Molekülen. EIF4G1 ist für die Rekrutierung von mRNA zu den Ribosomen verantwortlich. Die beiden Proteine EIF2S3 und EEF1G schließlich übernehmen Funktionen beim Transport von tRNA Molekülen zu den Ribosomen. Letzteres Gen ist mit einer erhöhten Expression in Pankreastumoren beschrieben worden. DICE steht für „Deleted In Cancer“ und wird als Tumorsuppressorprotein betrachtet, welches vermutlich mit DNA-Reparatur oder RNA-Spleißen zusammenhängt. Die Regulation der Transkription von Genen wird unter anderem über den Acetylierungsstatus von Histonen gesteuert, die die entsprechenden chromosomalen Abschnitte der Gene bedecken. Für die Acetylierung und Deacetylierung von Histonen gibt es eine Reihe von Proteinen, die aus verschiedenen Untereinheiten bestehen können, von denen zwei herunterreguliert sind: HDAC9 und ADA3. Das ebenfalls herunterregulierte MBD4 ist an der Reparatur von DNA beteiligt.

In den Leukämiezellen sind vorrangig Gene heraufreguliert, die an der Replikation von DNA beteiligt sind und damit auf proliferierende Zellen verweisen. Hingegen kodieren die wenigen herunterregulierten Gene Proteine der Proteinbiosynthese und

deuten damit an, dass in den leukämischen Blasten zumindest keine generell erhöhte Genexpression zu verzeichnen ist.

#### 3.6.2.4 Signaltransduktion und Regulation von Genexpression

Signaltransduktion und Regulation der Genexpression sind hier weit gefasste Begriffe. An diesen Prozessen sind die verschiedensten Proteine beteiligt, angefangen von extrazellulär wirkenden Zytokinen und ihren entsprechenden Rezeptoren, über Proteine im Zytoplasma, die eingegangene Signale zum Beispiel über G-Proteine vermitteln, bis zu Transkriptionsfaktoren, die die Expression einzelner Gene zu steuern vermögen. Nicht nur die Genexpression ist direkt abhängig von der Signaltransduktion. Viele andere Prozesse, zum Beispiel Stoffwechsel oder Immunantworten, sind bestimmt durch die Weiterleitung von außen eintreffender Signale. Nicht nur aus den genannten Gründen ist diese Gruppe heterogener als die anderen Gruppen. In menschlichen Zellen gibt es eine Vielzahl unterschiedlicher Signaltransduktionswege, die verschachtelt sind und ineinander greifen.

Häufig werden von außen an die Zelle kommende Signale über Rezeptoren in das Zytoplasma weitergeleitet, die direkt mit G-Proteinen interagieren. Dieser vielfältig anzutreffende Mechanismus erfordert unterschiedliche G-Protein-assoziierte Rezeptoren oder G-Protein-gekoppelte zytoplasmatische Proteine, zum Beispiel Kinasen. Einige Transkripte solcher Proteine können in den leukämischen Zellen nicht nachgewiesen werden. Hierzu zählen die G-Protein-gekoppelten Rezeptoren GPR18, P2RY10, die Kinasen GPR6 und PKIKG oder die GTPasen-aktivierenden Proteine RGS14 und RAP1GDS1.

Die meisten der mit Signaltransduktion in Zusammenhang stehenden detektierten Transkripte zeigen funktionelle Verbindungen zu der RAS-Superfamilie GTP-bindender Proteine oder sind verwandte Proteine dieser Familie. Zu den in den leukämischen Blasten heraufregulierten Transkripten zählen unter anderem das durch Dexamethason induzierte RAS-verwandte RASD1 oder ein dem RAS-verwandten Protein RAB15 der Ratte homologes Protein. NET1, GAB2 und RGL sind drei Proteine, die in RAS/RAL Signalwege involviert sind. Ebenfalls an diesen Signalwegen beteiligt sind die nun folgenden, herunterregulierten Transkripte RALGDS, REPS2 und ARHGEF3. Über RAL Proteine ist REPS2 an der Aktivierung oder Inaktivierung von RALGDS beteiligt, welches selbst mit der aktiven Form von RAS interagiert. Die bereits erwähnten ARL4, RAB22 und RAB32 sind RAS-verwandte GTP-bindende Proteine mit geringer Expression in den leukämischen Blasten. Bei SOS1 handelt es sich vermutlich um einen positiven Regulator von RAS, wohingegen RSU1 ein RAS-Suppressor ist.

Sieben Transkriptionsfaktoren, die in die Gruppe der ETS-Transkriptionsfaktoren gehören, sind in den leukämischen Blasten differentiell exprimiert. ETS2, ELF2 und ETV5 sind in den Blasten stärker nachzuweisen als in den Kontrollen. Die Aktivität von ETS2 wird durch Signale verstärkt, die über die RAS/RAF-Kinase Kaskade vermittelt werden. Es wird besonders stark in CD34+ Stammzellen exprimiert. ETV5 zeigt eine Expression in praktisch allen untersuchten Geweben, mit Ausnahme der Lymphozyten, wo es im gesunden Menschen normalerweise nicht nachzuweisen ist. Weiterhin konnte für ETV5 in den Proben von Patienten mit einer B-CLL (B-Zell chronisch lymphozytäre Leukämie) eine deutlich erhöhte Expression nachgewiesen werden (Korz, 2002). ETS1 und SPIB sind hingegen drei Mitglieder dieser Familie von Transkriptionsfaktoren, die in den Blasten abwesend sind. SPIB ist ein für die Entwicklung von B-Lymphozyten und myeloischen Zellen essentieller Transkriptionsfaktor. Ebenso sind die beiden Transkriptionsfaktoren EWS und FLI1 in den Leukämien schwächer exprimiert. Diese beiden Transkriptionsfaktoren bilden in Tumoren der Ewing-Familie Fusionstranskripte, die vermutlich zur Entartung der betroffenen Zellen führen.

Ein wichtiger Effektor in Wingless-Signalwegen (WNT) ist das  $\beta$ -Catenin CTNNB1, von dem vermutet wird, dass es an der Metastasierung von Tumoren beteiligt ist. Catenine zeigen Homologie zu Plakoglobin und zu Armadillo-Proteinen aus *D. melanogaster*. Das  $\beta$ -Catenin ist ein multifunktionelles Protein mit Funktionen in der Zelladhäsion und der Koaktivierung von T-Zell Transkriptionsfaktoren (TCF) und Lymphozyten Enhancer Faktoren (LEF). Der Transkriptionsfaktor LEF1, der im gesunden Menschen in B- und T-Lymphozyten exprimiert ist, assoziiert mit CTNNB1 und wird somit aktiviert (Hovanes, 2001). CTNNB1 interagiert mit Mitgliedern der TCF-Familie von Transkriptionsfaktoren. So wird zum Beispiel die WNT-abhängige Aktivierung von TCF7L2 über CTNNB1 vermittelt. Sowohl das  $\beta$ -Catenin als auch die beiden Transkriptionsfaktoren LEF1 und TCF7L2 sind in den leukämischen Blasten heraufreguliert. In Zellen des Dickdarmkrebses konnte gezeigt werden, dass die Überexpression des Komplexes LEF1/CTNNB2 zu einer Fehlregulation der Transkription führt (Hovanes, 2001). Weiterhin gibt es Hinweise, dass  $\beta$ -Catenin eine wichtige Rolle in Proliferation, Adhäsion und zum Überleben leukämischer Blasten besitzt (de Lau, 2001). DVL3 ist dem *dsh* aus *D. Melanogaster* homolog und ebenfalls hoch exprimiert. Dieses Gen kodiert ein zytoplasmatisches Phosphoprotein, welches die Zellproliferation in Entwicklungsprozessen reguliert, und wichtiger Bestandteil von WNT-Signalwegen ist. DVL3 wird unter anderem in B-Zellen, Nieren und Blase exprimiert und in Tumorgewebe aus der Brust oder dem Mastdarm. Ein negativer

Regulator dieses Signaltransduktionsweges ist NOTCH2, das in den Zellen der cALL herunterreguliert ist.

In JAK/STAT (Janus Kinase/ Signal Transducer and Activator of Transcription) vermittelte Signaltransduktionswege sind mindestens sechs der identifizierten Transkripte involviert. PIASX kodiert einen Inhibitor des aktivierten STATX und CSRP2 für einen nicht näher charakterisierten Faktor des JAK/STAT Signalwegs. Beide zeigen eine erhöhte Expression in den Leukämiezellen. Die Tyrosin-Phosphatase PTP $\epsilon$  (PTPRE) ist ein Inhibitor im JAK/STAT Signalweg. Sie zeigt die höchste Expression in den Zellen des Blutes. In den Leukämieproben ist die Expression der hier detektierten Rezeptorisoform nochmals deutlich erhöht. Protein-Tyrosin-Phosphatasen (PTPasen) stehen in einem regulativen Gleichgewicht mit Proteinkinasen (PTK), kontrollieren die Tyrosinphosphorylierung und steuern so die Funktion ihrer verschiedenen Zielproteine, zum Beispiel in der Zellproliferation. Ein Ungleichgewicht zwischen PTK und PTPase ist möglicherweise an der Onkogenese beteiligt (Tanuma, 2000). STAT2 (SOCS2) ist ein Inhibitor in JAK/STAT Signalwegen. So ist es in der Lage, die durch IGF vermittelte Aktivierung von STAT3 zu inhibieren.

Deutlich herunterreguliert in den leukämischen Zellen ist indes STAT3. Als Antwort auf Interleukin-5, Interleukin-6, Epidermaler Wachstumsfaktor EGF, Leukämie-inhibierender Faktor LIF, Onkostatin M oder Interleukin-11 wird der Transkriptionsfaktor STAT3 sehr schnell in der Zelle aktiviert. In vielen humanen Krebserkrankungen sowie in transformierten Zelllinien ist STAT3 dauerhaft aktiv. Unter anderem ist STAT3 an der kooperativen Signaltransduktion von LIF und BMP2 beteiligt. Dabei werden STAT3 durch LIF und MADH1 durch BMP2 aktiviert. Zum kooperativen Signal kommt es durch Assoziation von STAT3 mit MADH1 über p300 (Vorwerk, 2002). STAT6 spielt eine wichtige Rolle bei der IL4-vermittelten Aktivierung und Differenzierung von B-Lymphozyten. STAT6 und der IL4-Rezeptor sind beide in den leukämischen Proben herunterreguliert.

Es konnten vier Transkripte identifiziert werden, die mit dem Proto-Onkogen MYC interagieren können. Die Expression von MYC selbst zeigte keine Unterschiede. Diese vier Gene sind alle in den Leukämien herunterreguliert. BIN1 gehört in die BAR-Familie. Die Mitglieder dieser Gruppe interagieren mit MYC und können dessen onkogene Eigenschaften inhibieren. BIN1 ist in vielen Zelltypen exprimiert und zeigt eindeutige Tumorsuppressoreigenschaften. In Tumorzellen von Patienten mit Leberkarzinomen ist die fehlende Expression von BIN1 regelmäßig nachzuweisen. Für BIN2 gibt es keine funktionellen Analysen, die die Eigenschaften als Tumorsuppressor belegen. BIN2 ist am höchsten in der Milz und in Leukozyten des peripheren Blutes exprimiert. MYCBP kodiert einen MYC-bindenden Transkriptionsfaktor und hat eine vermutete Rolle in der

Spermatogenese. RCL kann ebenfalls MYC binden und ist wahrscheinlich in die Proliferation von Zellen involviert.

Eine bemerkenswerte Gruppe an Transkripten kann die Steuerung zellulärer Vorgänge in B-Zellen und davon abgeleiteten leukämischen Blasten besonders verdeutlichen. Diese SONDENSÄTZE detektieren verschiedene Gene, die alle in einem engen regulatorischen Zusammenhang zueinander stehen: MADH1, MADH7, SNIP1, SMURF1, OAZ, API2, BIRC5, TANK, TIAF1, TRAF3, TRAF5, TRAF6, CTGF, TNFRSF5 (CD40) und TSG. Sämtliche Mitglieder dieser Gruppe sind in die Signaltransduktionswege von TGF $\beta$  (Transforming Growth Factor) und BMP (Bone Morphogenetic Protein) involviert. Diese Regulationswege steuern Zellwachstum, Apoptose, Antiapoptose, Differenzierung und Organentwicklung (Heldin, 1997) und stehen in einem engen Geflecht mit TNF (Tumor Nekrosis Faktor), NF- $\kappa$ B (Nuclear Factor  $\kappa$ -B) oder WNT (Wingless) vermittelten Signalwegen.

Die folgenden Gene zeigen in den leukämischen Zellen eine Überexpression. MADH1 und MADH7 (Mothers Against Decapentaplegic Homolog, synonym als SMADs bezeichnet) sind zwei Mitglieder einer Proteinfamilie, die als direkte Mediatoren der Signaltransduktion im Zytoplasma und im Kern agieren. SNIP1 (SMAD Nuclear Interacting Protein) und SMURF1 (SMAD Ubiquitination Regulatory Factor) sind in der Lage, die Funktion von MADH-Mitgliedern in der Zelle zu regulieren. OAZ ist ein DNA-bindender Transkriptionsfaktor, der durch MADH-Mitglieder gesteuert wird. Über TGF $\beta$  werden die beiden Mediatoren TIAF1 (TGF $\beta$  Induced Anti-Apoptotic Factor) und CTGF (Connective Tissue Growth Factor) induziert. Das induzierte TIAF1 inhibiert die durch TNF vermittelte Apoptose und CTGF inhibiert ein durch NF $\kappa$ B vermitteltes antiapoptotisches Signal. TSG (Twisted Gastrulation) gehört in die Gruppe der BMPs und steuert in der Embryonalentwicklung die dorsoventrale Ausrichtung von Organen. TRAF6 schließlich wird eine Schlüsselrolle in der Übermittlung NF $\kappa$ B vermittelter Überlebenssignale zugeschrieben (Mak, 2002). Neben diesen in den leukämischen Zellen heraufregulierten Genen gibt es in diesem Zusammenhang einige herunterregulierte Gene. CD40 kann als Mitglied der TNF-Rezeptorfamilie mit TRAF3 oder TRAF5 assoziiert sein und vermittelt Signale vorrangig in Immunreaktionen (Bradley, 2001). TANK ist ein Inhibitor verschiedener TRAF Proteine und steuert darüber deren Funktion. API2 bindet ebenfalls an TRAF-Mitglieder TRAF1 oder TRAF2 und kann auf diesem Wege vermutlich Apoptosesignale unterbinden.

Es sei in diesem Zusammenhang bereits ein weiteres Transkript erwähnt, das allerdings nur in den Zellen der Leukämiepatienten mit einer TEL/AML1-Translokation signifikant stärker exprimiert ist (vgl. 3.8). In den verbliebenen Leukämieproben ist die

Expression uneinheitlich, in den gesunden Kontrollzellen ist keine Expression nachzuweisen. Es handelt es sich dabei um BMP2, das am Anfang der hier dargestellten Signaltransduktionswege steht und das somit eine zentrale Bedeutung besitzt. Die Expression von BMP2 ist nicht exklusiv in den TEL/AML1-positiven Proben erhöht. Es findet sich auch in den meisten der TEL/AML1-negativen Leukämien eine erhöhte Expression. Insgesamt wird bei 17 der 25 mit HG-U133A untersuchten Leukämien BMP2 detektiert, jedoch in keiner der gesunden Kontrollproben. Aufgrund der Schwankungen im Expressionsmuster allein der leukämischen Proben ist BMP2 nicht als differentiell exprimiert detektiert worden.

#### 3.6.2.5 Regulation von Differenzierung und Entwicklung

Von den differentiell exprimierten Genen in dieser Gruppe stehen einige in einem direkten Zusammenhang mit der Entwicklung der B-Lymphozyten und wurden daher gesondert in Abschnitt 3.6.2.1 behandelt. Einige andere der hier genannten Gene regulieren die Entwicklung oder Differenzierung speziell in neuronalen Geweben oder sind bisher nur in der Spermatogenese als Vorgang der Zelldifferenzierung beschrieben worden. Die übrigen Gene zeigen keine besondere Gewebespezifität, sind aber alle an Prozessen der Organentwicklung beteiligt.

PPP2R1B ist ein regulatorisches Enzym, welches die mitogenaktivierte Proteinkinase (MAPK) Kaskade herunterreguliert und Signale zur Zellproliferation weiterleitet. Es ist bereits in Lymphom-Zelllinien als überexprimiert beschrieben, zeigt Überexpression in den hier untersuchten Leukämiezellen und scheint in die Karzinogenese involviert zu sein. Zu den in den Leukämieproben heraufregulierten Genen zählen SPRY1, SPRY2 und SPRY4, deren Homologe in der Embryonalentwicklung von Fruchtfliegen, Mäusen und Zebrafischen nachgewiesen werden konnten. Von den Mitgliedern der SPRY-Familie wird vermutet, dass sie in der Lage sind, den RAS/RAF-Kinase Signalweg zu inhibieren (Yusoff, 2002). Bemerkenswerterweise sind einige der mit diesem Signalweg verbundenen Transkripte in der Leukämie herunterreguliert.

Die Kinase MELK wird in der Embryonalentwicklung der Säugetiere in einem räumlich und zeitlich festgelegten Muster exprimiert. In der Maus wird MELK während der Reifung der Oozyten vor der Implantation und später dann bei der Spermatogenese nachgewiesen. Aufgrund dieses Expressionsmusters wird vermutet, dass MELK eine Rolle in der Embryonalentwicklung einnimmt. MSX1 kodiert ein weiteres Gen mit Bedeutung in der Organentwicklung. Es wird besonders im Zusammenhang mit der Entwicklung von Zähnen und Knochen beschrieben. Mutationen im MSX1 Gen führen zu Fehlbildungen im Gesichtsschädel. MSX1 wird über den WNT-Signalweg reguliert

und bildet eine Brücke zu BMP-Signalwegen. Kinasen der DYRK-Familie verfügen über ungewöhnliche Eigenschaften, da sie sowohl die Autophosphorylierung als auch die Phosphorylierung anderer Substrate katalysieren können. Vermutlich wirken sie auf Prozesse der Entwicklung und des Zellwachstums ein. DYRK2 ist ein Familienmitglied, das in den Zellen der Leukämie heraufreguliert ist. In der GNF-Expressionsdatenbank ist es praktisch in keinem der getesteten Gewebe nachweisbar. DYRK4 hingegen zeigt eine verringerte Expression in den leukämischen Zellen.

Eine verstärkte Expression in den Zellen der cALL zeigt ID2. Es inhibiert die Funktion verschiedener Transkriptionsfaktoren in einer dominant-negativen Art und Weise, indem es die Bindung von E2A Transkriptionsfaktoren an die DNA verhindert. Die Inhibition des Zellwachstums durch E2A fällt damit ebenso fort wie die positive Regulation der Differenzierung (Becker-Herman, 2002). Deutlich überexprimiert in den leukämischen Zellen ist schließlich der Insulinrezeptor INSR. Verschiedene Arbeiten konnten in der Vergangenheit die Bedeutung von Insulin und dem entsprechenden Rezeptor für das Wachstum leukämischer Zellen zeigen. In Zellkulturuntersuchungen wurde nachgewiesen, dass Insulin eine wichtige Rolle im Wachstum leukämischer B-Lymphozyten hat (Baier, 1992) und dass die proliferative Antwort der Zellen auf Insulin über den Insulinrezeptor vermittelt wird (Neely, 1992).

Mit HOXB7 ist ein Transkriptionsfaktor herunterreguliert, der ebenso wie das DNA-bindende RUNX3 (AML2) Protein hauptsächlich in verschiedenen Organen während der menschlichen fetalen oder embryonalen Entwicklung nachzuweisen ist. Nach antigener oder mitogener Stimulation beginnen reife B-Zellen in der Milz sich zu teilen und differenzieren sich in Richtung antikörperproduzierende B-Zellen. Als direkte Antwort auf die TGF $\beta$ -Stimulation interagiert RUNX3 hierfür mit MADH3 und ist an der transkriptionalen Kontrolle des Umschaltens (class switch) auf eine andere Ig-Klasse beteiligt (Cohen, 2001; Hanai, 1999; Shi, 1998). Interferonregulierende Faktoren bilden eine Proteinfamilie, die neben ihrer Funktion in virusvermittelten Signalwegen regulativ auf das Zellwachstum einwirkt. IRF2 und IRF5 sind zwei Mitglieder dieser Familie und zeigen beide eine geringere Expression in den Leukämiezellen. Die zentrale Bedeutung des ebenfalls herunterregulierten TRAF3 liegt in der Regulation des Zellwachstums und in der Weiterleitung von Signalen in Apoptosewegen über TNF (Tumor Necrosis Factor). Dafür bindet es spezifisch an den zytoplasmatischen Teil von CD40, einem Mitglied der TNF-Rezeptor Superfamilie, welches nicht differentiell exprimiert ist. HRY und DACH sind zwei Beispiele für herunterregulierte Gene, deren Homologe in *D. melanogaster* Funktionen in der Organentwicklung übernehmen. HRY ist im Menschen essentiell für die Entwicklung der Muskeln, des Nervensystems, des Blutsystems und für die Geschlechtsbestimmung. Der Stammzellfaktor KIT (SCFR) ist eine

membranständige Tyrosinkinase, die für die Entwicklung des Blutsystems ebenso wichtig ist wie für die Ausbildung der Gameten. SCFR wird zudem als Proto-Onkogen betrachtet, ist aber in den Zellen der cALL schwächer exprimiert als in den Kontrollzellen.

Es finden sich in den Leukämieproben vier Gene, die dort stärker als in den Kontrollen exprimiert sind und denen eine Funktion in der neuronalen Entwicklung zugeordnet wird. Das Neuropeptid NPY zeigt die stärkste Überexpression. NPY fördert die Proliferation neuronaler Zellen nach der Geburt und scheint auch die Proliferation neuronaler Vorläuferzellen im Erwachsenen zu regulieren. Die CDP-Diacylglycerolsynthase CDS2 ist normalerweise nur in Zellen des Zentralnervensystems während der Embryonalentwicklung nachzuweisen. CDS2 wird eine Aufgabe bei der neuronalen Zelldifferenzierung zugeschrieben. HCNP scheint ein multifunktionelles Protein zu sein, welches vorrangig Aufgaben während der neuronalen Entwicklung und Differenzierung übernimmt und mit der Alzheimer-Krankheit assoziiert ist. Die Expression von HCNP ist in diesen Patienten erhöht. Das Gen Down Syndrome Critical Region 1 DSCR1 ist ein Inhibitor von calcineurinvermittelten Signalen. In Patienten mit Trisomie 21 ist eine erhöhte Expression nachgewiesen worden, und neben der Entwicklung des Zentralnervensystems im gesunden Menschen scheint es bei Down Syndrom-Patienten an der Ausbildung der neurologischen, kardiologischen und immunologischen Defekte beteiligt zu sein (Casas, 2001). ARL4 ist als einziges Gen im Zusammenhang mit neuronaler Entwicklung in der cALL herunterreguliert. Dem Genprodukt wird neben der Funktion in der Neurogenese und in der Embryonalentwicklung eine Funktion bei der Spermatogenese zugeordnet.

Auf HG-U133A wurden drei weitere differentiell exprimierte Gene identifiziert, die eine Rolle bei der Spermatogenese spielen. In den Leukämien stärker exprimiert ist SPATA2, welches als Protein im Kern lokalisiert ist und bislang nur in Hoden detektiert wurde. Das ebenfalls in Hoden nachgewiesene SPAG1 ist in der Leukämie schwächer exprimiert. Hier wird eine Beteiligung an der Spermatogenese vermutet. Der Transkriptionsfaktor MYCBP hingegen ist ebenfalls regulativ an der Spermatogenese beteiligt, zeigt aber eine geringere Expression in den Zellen der Leukämie. Diese vier Gene zeigen bislang keinen erklärten Zusammenhang zur Leukämie oder zu anderen Krebserkrankungen. Es ist jedoch denkbar, dass diese Proteine in anderen Geweben Funktionen der Steuerung von Differenzierung und Entwicklung einnehmen.

### 3.6.2.6 Regulation von Zellzyklus und Zellproliferation

Diese relativ kleine Gruppe zeichnet sich ebenfalls dadurch aus, dass deutlich mehr Gene in den Leukämien stärker exprimiert sind als in den Kontrollen. Vergleichsweise wenig Gene unterliegen in den Leukämien einer negativen Regulation.

Die folgenden Gene sind in der cALL hoch reguliert. Der Antikörper KI67 reagiert mit dem nukleären Antigen MKI67, welches nur in proliferierenden Zellen, jedoch nicht in ruhenden Zellen nachzuweisen ist. Aufgrund dieser Tatsache wird MKI67 eine essentielle Rolle bei der Zellproliferation zugesprochen. Die stärkste Expression zeigte MKI67 in allen malignen Zelllinien der GNF-Expressionsdatenbank. Weiterhin gilt MKI67 als Markermolekül für Lymphome. Auch in den Zellen der cALL ist MKI67 stark exprimiert. SMC4 hat vermutlich Funktionen bei der Organisation und Kondensation der Chromosomen während der Mitose. Die folgenden sechs Proteine kontrollieren definierte Punkte des Zellzyklus: BUB1B und SAM68 den G2/M-Übergang und ESP1 den Wechsel von der Metaphase in die Anaphase. CDC2 ist die katalytische Untereinheit des M-Phasen Promoting Factors, der den Beginn der Mitose initiiert und in Eukaryoten universell vorkommt. CDC2 kontrolliert den Übergang von der G1- zur S-Phase sowie den Übergang von der G2- zur M-Phase. Die mit CDC2 verwandten Proteine CDC2L1 und CDC2L2 sind ebenfalls in den Leukämien höher exprimiert und haben vermutlich ähnliche Funktionen wie CDC2 selbst. Die Phosphatase CDC14A, ebenfalls in den Leukämien heraufreguliert, interagiert mit den Zentrosomen während der Interphase und ist an der Chromosomenverteilung und damit an der Stabilität des Genoms beteiligt. Die starke Expression dieser Gene deutet an, dass es sich bei den leukämischen Blasten um schnell proliferierende Zellen handelt. BUB1B ist zudem in die Pathogenese von Mastdarm Tumoren involviert. Auch diese Gene zeigten in der GNF-Expressionsdatenbank die höchste Expression in den malignen Zelllinien.

Das Antigen CD98, auch als SLC3A2 bezeichnet und in den leukämischen Blasten überexprimiert, wird von vielen sich teilenden Zellen exprimiert. So ist es zum Beispiel auf Zellen embryonaler Haut und der Lunge nachweisbar, wird in Rhabdomyosarkomen (Tumoren der quergestreiften Muskulatur) und Osteosarkomen (Knochentumoren) nachgewiesen. Für erwachsene Patienten mit einer ALL ist eine erhöhte Expression von CD98 ebenfalls nachgewiesen worden. CD98 ist ein lymphozytenaktivierendes Antigen mit vermuteten Funktionen in der Regulation von Zellwachstum und Teilung. BIRC5 (Survivin, API4) ist ein Inhibitor der Apoptose und spielt eine Rolle in der Zellzyklusprogression bei der Kontrolle des G2/M-Übergangs. Im Gegensatz zu den anderen verwandten Apoptoseinhibitoren der IAP-Familie (Inhibitor of Apoptosis) wird BIRC5 nicht in proliferierenden gesunden Zellen, zum Beispiel in Haut, Darm oder Knochenmark, nachgewiesen, dafür aber während der Embryonalentwicklung (Altieri,

1999). In vielen Tumoren konnte ohne Abweichung eine hohe Expression von BIRC5 gezeigt werden, während es in den gesunden Kontrollgeweben nicht nachgewiesen werden konnte (Velculescu, 1999). Der Transkriptionsfaktor E2F1 aktiviert am Übergang G1/S die Transkription einer Gruppe von Genen, die für die DNA Replikation in der S-Phase des Zellzyklus' benötigt werden.

Der Zellzyklus von Eukaryoten wird auch über cyclinabhängige Kinasen (CDK) reguliert, deren Aktivität selbst durch Cycline und CDK-Inhibitoren gesteuert wird. Die Mechanismen der Regulation beinhalten die Phosphorylierung und Dephosphorylierung von Serin-, Threonin- oder Tyrosinresten. Das in den leukämischen Proben hochregulierte CNK ist eine solche Serin/Threonin-Kinase. CDKN1A, CDKN2A und CDKN2C sind drei überexprimierte CDK-Inhibitoren, und GAK assoziiert mit Mitgliedern der CDK Familie. Mit CDKN1B ist ein cyclinabhängiger Kinase-Inhibitor in den leukämischen Zellen herunterreguliert, der als Inhibitor der Zellproliferation fungiert. In B-CLL Proben war CDKN1B stärker exprimiert als in den Kontrollen (Korz, 2002). Ebenso wie bei der Serin/Threonin-Kinase STK18 handelt es sich bei dem Cyclin CCNB2 um einen weiteren Regulator der Zellteilung. CCNB2 wird als spezifisches Transkript für Hoden beschrieben. In der GNF-Expressionsdatenbank lag die Expression für CCNB2 in einigen Lymphomen und Melanomen bis zu 50fach über dem Median aller getesteten Gewebe. Die Kinase STK18 wird im gesunden Menschen nur im Hoden und in der Schilddrüse exprimiert und konnte auch in einer T-Zell-Leukämie Zelllinie gezeigt werden.

Der in den Leukämiezellen stark herunterregulierte Transkriptionsfaktor E2F5 hat eine nicht näher bezeichnete Funktion in der Progression des Zellzyklus'. E2F5 ist nicht für die Progression des Zellzyklus' notwendig, aber unabdingbar für den Zellzyklusarrest sich teilender Zellen. PLAGL1 kodiert das Tumorsuppressorprotein ZAC1 (LOT1), dessen fehlende Expression bereits in verschiedenen malignen Zelllinien gezeigt wurde. Dieses Zinkfingerprotein kann wie das Homologe ZAC der Maus (und wie p53) die Proliferation von Tumorzellen durch Induktion von Apoptose und durch Zellzyklusarrest unterbinden.

#### 3.6.2.7 Allgemeiner Metabolismus

In diese Gruppe fallen Gene, die eine generelle Rolle bei verschiedenen Stoffwechselfunktionen übernehmen. Diese sind bisweilen schwer in einen Zusammenhang mit der Leukämie oder anderen Neoplasien zu stellen. Auf den ersten Blick ist jedoch die ungleiche Verteilung herauf- und herunterregulierter Gene in dieser Gruppe auffallend. Wesentlich mehr Gene sind in den leukämischen Zellen herunterreguliert als heraufreguliert.

Die Glutathion-S-Transferase GSTO1 wird als konstitutiv exprimiertes Gen beschrieben. Sie ist im Gegensatz zu anderen Glutathiontransferasen mit ihren häufig gewebespezifischen Expressionsmustern ubiquitär nachweisbar und in den Zellen der cALL schwächer exprimiert als in den Kontrollen. MGST1 kodiert eine zweite herunterregulierte Glutathion-S-Transferase. Ebenfalls herunterreguliert sind zwei Enzyme des NADH/NADPH Stoffwechsels: FLJ13052, eine NAD-Kinase, die die Synthese von NADP katalysiert, und NCF1, das Bestandteil des NADPH-Oxidasekomplexes ist. Die L-Kynurenin Hydrolase KYNU ist in den Kynurenin Stoffwechsel der Synthese von NAD-Kofaktoren aus Tryptophan involviert. Im Vergleich der Genexpression von hämatopoetischen Stammzellen mit der von prä-B-Zellen war KYNU in den B-Vorläuferzellen ebenfalls herunterreguliert (Muschen, 2002). OXCT ist ein Enzym des Ketonstoffwechsels, welches normalerweise speziell in Leukozyten und Fibroblasten nachzuweisen ist. Weiterhin ist ein Enzym SCD zur Synthese ungesättigter Fettsäuren herunterreguliert. Dieses ist in den Tumorzellen von Karzinomen des Darms oder der Speiseröhre stärker nachzuweisen als in den entsprechenden gesunden Zellen. Die Phosphoglukomutase PGM1 katalysiert den Transfer von Phosphatgruppen zwischen den 1- und 6-Positionen von Glucose. Das Enzym der Cholesterinbiosynthese FDFT1 ist ebenso in den Leukämien schwächer exprimiert wie der Katalysator des Embden-Meyerhof-Wegs der Glykolyse GPI.

Die Superoxid-Dismutase SOD2 ist ein Enzym der Mitochondrien und ist in Zellen des Prostatakarzinoms geringer exprimiert als in den umgebenden gesunden Epithelzellen (Bostwick, 2000). Bei SOD2 wird daher eine Funktion als Tumorsuppressor vermutet ebenso wie bei der Dipeptidase DPEP1, die im Wilms Tumor, einem Nierentumor des frühen Kindesalters, nicht nachgewiesen werden konnte. Beide Gene sind allerdings in den hier untersuchten Leukämien stärker exprimiert als in den Kontrollen. WSF1 ist ein transmembranes Protein, das im endoplasmatischen Retikulum lokalisiert ist und in vielen Geweben exprimiert wird. Eine Expression im gesunden Blut konnte nicht gezeigt werden. Drei hochregulierte Transkripte kodieren Proteasomuntereinheiten und sind damit am Proteinumsatz in den Zellen beteiligt. Im einzelnen sind das PSMC6, PSMD2 und PSMD11. USP11, USP14 und USP16 sind an der Deubiquitinierung von Proteinen beteiligt und dem Proteasomapparat assoziiert. Mit PSMD4 ist allerdings eine Proteasomuntereinheit in den leukämischen Zellen schwächer exprimiert.

Die folgenden fünf Gene sind insofern von besonderem Interesse, als dass sie Proteine kodieren, die einen Einfluss auf die Wirksamkeit verschiedener Chemotherapeutika haben können. Das Genprodukt von CDA katalysiert die Deaminierung von Cytosin-Nucleosid-Analoga, welche wegen ihrer zytotoxischen

Wirkung bei der Behandlung der AML in Form des Cytosin-Arabinosids eingesetzt werden. In den hier untersuchten pädiatrischen cALL-Proben ist CDA deutlich schwächer exprimiert als in den Kontrollen. Ebenfalls deutlich herunterreguliert sind ABCB1 (MDR1) und ABCB4 (MDR3). ABCB4 ist ein Multi-Drug-Resistance Gen, welches ein Protein kodiert, das aktiv, unter Verbrauch von ATP, Wirkstoffe aus der Zelle herustransportiert. Zu diesen Wirkstoffen zählen Chemotherapeutika. Die Expression von DGUOK (Deoxyguanosin-Kinase) zeigt bei pädiatrischen Patienten mit ALL oder AML große individuelle Schwankungen. Zur Behandlung dieser Krankheiten werden häufig Purin-Deoxyribonukleosid-Analoga eingesetzt. Damit diese zytotoxisch wirken können, müssen sie in der Zelle von dieser Deoxyguanosin-Kinase phosphoryliert werden. DGUOK ist in den hier untersuchten cALL-Patienten schwächer exprimiert als in den gesunden Kontrollen. FPGS kodiert eine Folylpolylglutamat-Synthase und steht ebenfalls in Zusammenhang mit therapieresistenten Leukämien. In den hier untersuchten cALL-Proben ist FPGS stärker exprimiert als in den Kontrollen. Eine Herunterregulation der Expression dieses Gens führt zur Therapieresistenz gegenüber Methotrexat, das bei verschiedenen Leukämien zum Einsatz kommt, unter anderem bei der Behandlung der cALL.

#### 3.6.2.8 Gene mit bekannten Zusammenhängen zu Krebserkrankungen

Eine Reihe von Genen, die hier als differentiell exprimiert identifiziert wurden, sind bereits in einem Zusammenhang mit Leukämien oder anderen Krebserkrankungen beschrieben worden. Hierunter fallen Tumorsuppressoren, Onkogene oder Gene, die Proteine kodieren, welche für Therapieresistenzen verantwortlich sein können. Ein Teil der hier dargestellten Gene ist bereits in den oben beschriebenen funktionellen Gruppen erläutert worden.

Zunächst zu den Genen, für die bereits Zusammenhänge mit verschiedenen lymphatischen oder myeloischen proliferativen Erkrankungen aufgedeckt wurden. An erster Stelle sollen hier nochmals CD10 und TdT genannt werden, als zwei der wichtigsten Marker zur Diagnose einer cALL. Die Metalloproteinase CD10 ist erstmals überhaupt als „Common Acute Lymphoblastic Antigen“ beschrieben worden, bevor sie in gesunden Zellen nachgewiesen werden konnte. CD10 ist auch auf gesunden B-Lymphozyten anwesend, nämlich auf den frühen Stadien der B-Zellreifungsreihe, den frühen und späten pro-B-Zellen und den großen prä-B-Zellen (vgl. Abbildung 1-3). Die Terminale Deoxynukleotidyltransferase TdT ist eine einmalige DNA-Polymerase, die ohne Matrize Deoxyribonukleotide an das Hydroxy-Ende von DNA-Fragmenten hängen kann. Die physiologische Funktion liegt im Einfügen von Nukleotiden an den Verbindungsstellen zwischen den V(H)-D und D-J(H) Segmenten der

Immunglobulingene. Dementsprechend ist TdT im gesunden Menschen nur in den Vorläuferzellen der B- und T-Lymphozyten während der Umlagerung der Immunglobulingene exprimiert, so zum Beispiel in den frühen und den späten pro-B-Zellen (vgl. Abbildung 1-4). Die Expression von CD10 und TdT macht den zellulären Ursprung der cALL Blasten noch einmal deutlich.

CTGF (Connective Tissue Growth Factor) ist ein Protein mit verschiedenen Funktionen, unter anderem in der Regulation von IGF-, TGF $\beta$ - und BMP4-vermittelten Signalen. Kürzlich konnte gezeigt werden, dass CTGF beim Vergleich des Blutes von ALL-Patienten mit Nabelschnurblut oder gesundem Blut nur in den leukämischen Proben nachweisbar war. Weiterhin korrelierte der Nachweis von CTGF im Blut der Patienten mit dem Verfehlen einer Remission der Erkrankung (Vorwerk, 2002; Vorwerk, 2000). Der Rezeptor für Insulin INSR ist in den Blasten der cALL stark exprimiert. Die Bedeutung von Insulin für das Wachstum leukämischer B-Lymphozyten ist bekannt (Baier, 1992; Neely, 1992; Zumkeller, 1999). Diese beginnen als Antwort auf Insulin zu proliferieren. HLX1 ist ein DNA-bindendes Protein mit einer Homöo-Box Domäne, das genregulatorische Eigenschaften besitzt. HLX1 wird eine Rolle bei der Regulation von Wachstum und Differenzierung hämatopoetischer Zellen zugeschrieben. Die Überexpression dieses Gens in Blasten der ALL und AML, das normalerweise in Zellen des Thymus und in CD34+ Stammzellen des Knochenmarks exprimiert ist, ist seit längerem bekannt (Deguchi, 1993a) und wird als Marker für die unreife hämatopoetischer Zellen betrachtet (Deguchi, 1993b).

CD98 ist in den leukämischen Zellen etwas stärker exprimiert als in den Kontrollzellen. Es ist in proliferierenden Zellen verschiedenen Ursprungs nachgewiesen worden, so unter anderem in der Lunge, in Keratinozyten, in embryonaler Haut und in einigen soliden Tumoren. Eine hohe Expression von CD98 korreliert mit dem Verfehlen einer kompletten Remission in pädiatrischen ALL-Patienten, und es konnte weiterhin gezeigt werden, dass die Expression in ALL-Zellen stärker ist als in AML-Zellen. Das in Zusammenhang mit Therapieresistenzen stehende FPGS, welches ebenfalls in den leukämischen Blasten überexprimiert ist, wurde bereits oben behandelt. Der Transkriptionsfaktor MN1 ist ein möglicher Fusionspartner bei Translokationen mit dem TEL-Locus in myeloischen Leukämien. Das entstandene Fusionstranskript, das durch die Translokation t(12;22)(p13;q11) kodiert wird, hat transformierende Eigenschaften. Dieselbe Translokation ist in den Blasten chronischer B-Zell Leukämien nachgewiesen worden. Die in dieser Studie untersuchten Patienten sind alle vom Typ cALL und tragen diese Translokation nicht. Lediglich die Expression des unveränderten MN1 ist hier heraufreguliert. Im gesunden Menschen ist MN1 nur im fetalen Gehirn nachzuweisen.

Die folgenden Gene sind alle in den Zellen der cALL schwächer exprimiert als in den Kontrollzellen oder nicht nachzuweisen. Der Transkriptionsfaktor ICSBP1 wird exklusiv in Zellen des Immunsystems oder in Zellen mit einem hämatologischen Ursprung exprimiert und spielt vermutlich eine regulative Rolle in Proliferation und Differenzierung. Das Fehlen des entsprechenden Transkripts ist bereits für myeloische Leukämien bekannt. CD37 ist ein Oberflächenantigen, welches auf B-Zellen stark und auf T-Zellen schwach exprimiert ist und in B-Vorläuferzellen nicht nachzuweisen ist. Vermutlich vermittelt CD37 Interaktionen zwischen B- und T-Zellen. Es wird bereits als schwach exprimiert in lymphatischen Leukämien beschrieben. BCL11A stellt eine allgemeine Integrationsstelle für Retroviren dar und zeigt die stärkste Expression im Gehirn, in Milz und Hoden. Es wird von den meisten hämatopoetischen Zellen gebildet und scheint in lymphatische Erkrankungen mit einem unbekanntem Mechanismus involviert zu sein. CD48, ein Zelladhäsionsmolekül, ist ein Marker für aktivierte B- und T-Zellen mit einer unbekanntem Funktion. Für Patienten mit einer Leukämie vom Typ CLL und einer Translokation am MLL-Locus 11q ist eine reduzierte Expression von CD48 bekannt. Das Oberflächenantigen CDw52 (CAMPATH1) wird im gesunden Menschen auf Lymphozyten und Monozyten exprimiert und ist verantwortlich für antikörperabhängige und zellvermittelte Zytotoxizität. Ein CAMPATH1 Antikörper wurde bereits zur Behandlung von chronischen B-Zell Leukämien (B-CLL) und einigen Lymphomen eingesetzt. In den hier untersuchten Proben ist CDw52 auf mRNA-Ebene jedoch nicht nachzuweisen.

CD23 ist ein Schlüssel-molekül für Wachstum, Entwicklung und Aktivierung von B-Zellen. Es ist in der Lage, IgE Moleküle zu binden und scheint das IgE-Niveau im Serum zu regulieren. Die Bildung von CD23 wird durch Interferon- $\gamma$  und IL4 gesteuert. In den hier untersuchten cALL Zellen ist CD23 mittels zwei unabhängiger SONDENSÄTZE auf HG-U133A als sehr stark herunterreguliert detektiert worden. Die fehlende Expression von CD23 in akuten Leukämien ist bereits beschrieben worden (Bruserud, 2002). Das Antigen CD21 (Complement Component Receptor, CR2) ist das Membranprotein der B-Lymphozyten, an das EBV während der Infektion bindet. Es ist bekannt, dass dieser bei Leukämiepatienten fehlen kann. Weiterhin ist CD21 Ligand von CD23. Die Interaktion beider Moleküle ist für B-Zellaktivierung und Proliferation wichtig. In der B-CLL können beide Gene heraufreguliert sein, und es wird vermutet, dass dies die Proliferation der B-CLL-Zellen bedingt (Lopez-Matas, 2000). CD23 wird bei der Diagnose von B-CLL und anderen lymphoiden Leukämien herangezogen. Der erhöhte Nachweis von löslichem CD23 im Serum von B-CLL-Patienten geht mit einer schlechten Prognose einher (Molica, 1996). Nebenbei bemerkt sei, dass der Rezeptor für

das CD23-regulierende Interleukin-4 in einigen Leukämiezellen herunterreguliert ist, wenngleich nicht signifikant für alle Proben.

Deutlich herunterreguliert sind drei Gene, deren Produkte mit Therapieresistenzen in akuten Leukämien in Verbindung gebracht werden. Wird CDA in Blasten von AML-Patienten exprimiert, kommt es zu einem Versagen der Therapie mit Cytosin-Nucleosid-Analoga. ABCB4 ist ein membranständiger Transporter, der aktiv Wirkstoffe aus der Zelle entfernen kann. Die Expression von ABCB4 in Zellen von Patienten mit ALL oder CLL ist verbunden mit einer Therapieresistenz. Purin-Deoxyribonucleosid-Analoga werden bei der Behandlung von AML und ALL eingesetzt. DGUOK phosphoryliert die als inaktive Vorstufe verabreichten Analoga zum zytotoxischen Wirkstoff. Kann in den Blasten die Expression von DGUOK nicht nachgewiesen werden, korreliert das mit einem Versagen dieser Therapieform. Die fehlende Expression von CDA und ABCB4 kann eine positive Prognose zum Behandlungserfolg liefern, die Abwesenheit von DGUOK hingegen eine negative.

Die nun folgenden Gene sind bereits in einem Zusammenhang mit weiteren Krebserkrankungen beschrieben worden, die nicht myeloischen oder lymphatischen Ursprungs sind. ID2 ist in der Lage, den antiproliferativen Effekt von Tumorsuppressoren der Retinoblastomfamilie (RB) aufzuheben und ermöglicht damit das Fortschreiten des Zellzyklus' (Lasorella, 2000). Die Unterbrechung dieses RB-abhängigen Regulationswegs ist ein Kennzeichen von Tumoren. ID2 ist auch in den Leukämiezellen stark exprimiert und könnte somit den RB-abhängigen Regulationsweg unterbrechen. BUB1B ist ein weiteres Protein, das in der Leukämie überexprimiert ist und das den Zellzyklus am Übergang von der G2- zur M-Phase reguliert. BUB1B scheint in die Pathogenese von Mastdarmkrebs involviert zu sein (Cahill, 1998). SPANXC wird normalerweise nur in Hoden exprimiert und scheint an der Spermatogenese beteiligt zu sein. Es wird in Zelllinien nachgewiesen, die von metastasierenden Melanomen abstammen. Zellen aus nicht metastasierenden Melanomen exprimieren SPANXC nicht (Zendman, 1999). In den cALL Zellen ist es überexprimiert. Einen Zusammenhang mit der Metastasierung von Tumoren zeigt auch das Zelladhäsionsprotein CTNNB1 (Beta-1-Catenin), welches in der cALL überexprimiert ist.

Der Apoptoseinhibitor BIRC5 zeigt neben der bekannten starken Expression in verschiedenen Adenokarzinomen von Lunge, Brust, Pankreas oder Prostata auch in der cALL verstärkte Expression. Der Transkriptionsfaktor TCF7L2 ist ebenfalls in verschiedenen Tumoren nachgewiesen worden. MG50 ist ein mit Melanomen assoziiertes Gen, das im gesunden Menschen in Herz, Lunge und Plazenta stark und in Leber, Dickdarm, Pankreas, Milz, Nieren und Hoden schwach exprimiert ist. In

gesunden Leukozyten ist MG50 nicht nachzuweisen. Die normale Funktion von MG50 liegt vermutlich in der Regulation der p53-vermittelten Apoptose. Neben einer erhöhten Expression in Zellen des Dickdarmkrebses oder von Melanomen ist MG50 auch in den Zellen der cALL heraufreguliert.

Es gibt eine Reihe von Genen, die in der cALL herunterreguliert sind, die aber im Zusammenhang mit anderen Tumoren als überexprimiert oder differentiell exprimiert beschrieben worden sind. Die Bedeutung dieses Genexpressionsmusters für die Pathogenese der cALL ist damit im einzelnen interessant. GASC1 ist für verschiedene Krebsarten, unter anderem für Speiseröhrenkrebs, ebenso wie der nukleäre Rezeptor NCOA3 für Brustkrebs als überexprimiert beschrieben. Die Expression des Haptoglobins HPR ist für Lymphome nachgewiesen, fehlt aber in den Zellen der cALL. Die Cystein-Proteinase Cathepsin Z (CTSZ), einer der Hauptbestandteile des lysosomalen proteolytischen Systems, ist in allen untersuchten menschlichen Geweben und in den meisten Tumoren nachzuweisen. In den Zellen der cALL fehlt CTSZ. Das Testin TES ist ein erstmals in Hoden nachgewiesenes Transkript und stellt möglicherweise ein Tumorsuppressorgen dar. In 22% der untersuchten Krebs-Zelllinien und in 44% der Zelllinien, die sich von bösartigen hämatopoetischen Erkrankungen ableiten, konnte TES nicht nachgewiesen werden. EVI2B ist bei Mäusen an der Ausbildung myeloischer Tumore nach Infektion mit einem Retrovirus beteiligt. Beim gesunden Menschen lässt sich das entsprechende Transkript in mononukleären Zellen des peripheren Blutes und in Fibroblasten nachweisen. Für Patienten mit einer CLL ist eine differentielle Überexpression von EVI2B bereits beschrieben worden (Aalto, 2001). Es ist in den hier untersuchten cALL-Proben allerdings herunterreguliert.

#### 3.6.2.9 Gene, die sich keiner der Gruppen zuordnen lassen

Den Großteil dieser Gruppe machen uncharakterisierte ESTs aus, über die keinerlei Informationen verfügbar sind. Das sind auf EOS-H 165 und auf HG-U133A 102 verschiedene Sequenzen des mit stringenten Kriterien reduzierten Datensatzes. Darin enthalten sind eine Reihe kodierender Sequenzen für unbekannte Proteine (z.B. KIAA1691, KIAA0176) und hypothetische Proteine (z.B. FLJ12442, FLJ20037), die in der cALL herauf- oder herunterreguliert sein können. Zum Teil sind die Expressionsunterschiede sehr deutlich. Die beiden Zelladhäsionsmoleküle CEACAM6 und CEACAM8 sind in den Zellen der Leukämie deutlich schwächer exprimiert als in den Kontrollen. Für andere Mitglieder der CEA-Familie ist eine Expression in cALL bereits beschrieben (Hananberg, 1994). Die Funktionen beider Proteine sind nicht bekannt. Sie zeichnen sich lediglich dadurch aus, dass sie mit einem Antikörper gegen

das Carcinoembryonic Antigen kreuzreagieren. Ein Zusammenhang beider Proteine mit Neoplasien ist nicht beschrieben.

Es finden sich in dieser Gruppe allerdings auch einige wenige Gene, die Proteine kodieren, die ubiquitär exprimiert sind und die keinen erklärbaren Zusammenhang zur cALL erkennen lassen. Das Glykoprotein CHI3L1 der Chondrozyten ist ebenso ein Beispiel wie TUFTIP, das eine wichtige Rolle bei der Entwicklung und Mineralisation von Zahnschmelz zu spielen scheint. Beta-Tropomyosin TMP2, Titin TTN und Troponin TNNC1 sind drei Proteine, die an der Regulation der Kontraktion quergestreifter Skelettmuskeln beteiligt sind. TNNC1 und TTN sind in den Leukämien schwächer exprimiert als in den Kontrollen. In der GNF-Expressionsdatenbank wurde TNNC1 ausschließlich im Herzen und dort mit sehr starker Expression detektiert. Es ist an der Kontraktion glatter Muskulatur beteiligt. TMP2 ist in den leukämischen Proben heraufreguliert. Bemerkenswerterweise zeigt die GNF-Expressionsdatenbank die höchste Expression von TMP2 in der Prostata, im Prostatakarzinom, im Uterus und den Ovarien. Allerdings wurden in dieser Studie keine quergestreiften Muskeln untersucht.

### 3.6.3 Gene, die mit beiden Chiptypen detektiert wurden

Als übereinstimmendes Ergebnis aus den Untersuchungen wurden 156 unterschiedliche Gene identifiziert, die mit Hilfe beider Chiptypen als differentiell exprimiert detektiert wurden. Davon sind 25 Gene in den leukämischen Zellen herauf-, 121 Gene herunterreguliert. Zu den heraufregulierten Genen zählen CTGF, LEF1, MADH7, SMURF1, TIAF1, BIRC5, SOD2 und SOX4. Unter den herunterregulierten Genen finden sich neben ARL4, BIN2, SYK, TRAF, SPYB, IL4R, LYN und BLR1 eine Reihe an Genen, die in azurophilen Granula neutrophiler Granulozyten exprimiert werden (vgl. 3.6.2.2). Diese doppelt detektierten Gene haben im Vergleich zu den nur mit einem Genechiptyp detektierten Genen keine herausgehobene Bedeutung für die Biologie der B-Lymphozyten oder für die Pathologie der cALL. Die mit beiden Genechips detektierten Gene unterscheiden sich nicht durch besondere Eigenschaften in den Messwerten von den anderen Genen, die in nur einer der beiden Analysen detektiert wurden. Es handelt sich weder um diejenigen Gene, welche die höchsten FC-Werte aufweisen, noch um die Gene mit den signifikantesten Unterschieden. Diese Gene sind innerhalb aller differentiell exprimierter Gene eines jeden Chiptyps gleichmäßig von den hohen bis zu den niedrigen FC-Werten verteilt und haben ganz unterschiedliche Signifikanzen. Eine herausgehobene Bedeutung dieser Gene lässt sich aus der doppelten Detektion mit unterschiedlichen Genechiptypen nicht ableiten.

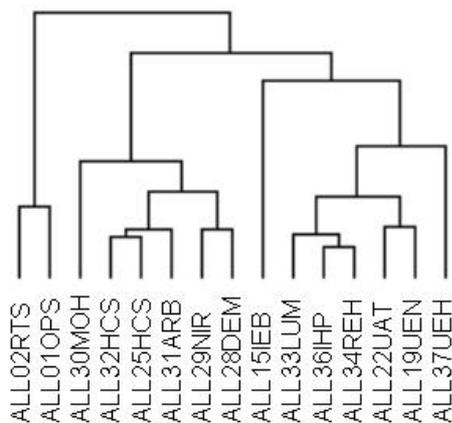
### 3.7 Untergruppen der Leukämien

Das Clustern von Proben oder Genen ohne die Verwendung zusätzlicher Informationen, die etwas über eine Eigenschaft der Proben aussagen (zum Beispiel Patient mit oder ohne TEL/AML1-Translokation), bezeichnet man als eine unüberwachte Clustermethode (vgl. 2.11.2.2). In einem solchen Ansatz sollen nun die leukämischen Proben geclustert werden mit dem Ziel, mögliche Subklassen der cALL aufzudecken. Wieder wurden die Daten der beiden Genechiptypen nur getrennt voneinander analysiert. Untersucht wurde mit zwei methodisch unterschiedlichen Verfahren, zum einen mit dem hierarchischen Clustern mit Hilfe des Programms von M. Eisen, zum anderen mit dem *k*-Means Clustern (in Zusammenarbeit mit Frau Dipl.-Inform. A. Thum, Institut für Informatik, Martin-Luther-Universität).

Grundsätzlich sollen in diesem Ansatz nur die Leukämieproben untersucht werden, also keine der Kontrollproben. Dabei wird keine Vorauswahl auf bereits identifizierte Gene getroffen, die einen Unterschied in der Expression zwischen cALL und gesunder Kontrolle zeigten. Diese Auswahl von Genen ist nicht sinnvoll, da hier bereits Informationen aus allen leukämischen Proben eingeflossen sind. Aber das sind genau die Gene, in denen sich alle Leukämieproben als eine gemeinsame Gruppe von den Kontrollen unterscheiden. Es sollen aber Untergruppen der cALL identifiziert werden, also Merkmale berücksichtigt werden, in denen sich die leukämischen Proben voneinander unterscheiden. Zum Clustern werden nur die Merkmale betrachtet, also Genexpressionswerte, die einen Unterschied zwischen den cALL-Proben zeigen. Gene, die durchgehend in keiner der cALL-Proben exprimiert sind, können also von der Analyse ausgeschlossen werden. Unter Berücksichtigung probenindividueller Schwankungen wurden alle Gene aus dem Datensatz entfernt, die in 90% der Proben nicht detektiert worden sind. Für die 25 mit HG-U133A untersuchten cALL-Proben bedeutet das die Entfernung von Genen, welche in mindestens 22 Proben nicht exprimiert waren. Dies reduzierte den Datensatz von 22.308 auf 11.188 Gene, mit denen geclustert wurde. Bei dem EOS-H Datensatz wurden alle Gene entfernt, die in mindestens 13 Proben nicht detektiert wurden. Hier reduzierte sich der Datensatz von 34.503 auf 19.144 Gene.

Beim Clustern der mit EOS-H untersuchten Proben sind keine konkreten Cluster auszumachen, die auf eine mögliche Unterteilung der cALL-Proben in Subklassen hindeuten (Abbildung 3-13). Dabei ist es gleichgültig, mit welcher der beiden Methoden geclustert wurde. Die Abbildung zeigt als Beispiel ein Ergebnis des hierarchischen Clusters. Einige der Proben zeigten eine auffällige Nähe zueinander. So clusterten die Proben ALL19UEN/ALL22UAT oder ALL28DEM/ALL29NIR immer in direkter Nachbarschaft. Die Proben ALL01OPS, ALL02RTS, ALL15IEB, ALL30MOH und

ALL37UEH hingegen stehen sehr weit entfernt von allen anderen Proben, ohne jedoch selbst eine Gruppe zu bilden. Die übrigen Proben verhalten sich uneinheitlich und bilden keine deutlichen Gruppen. Insgesamt führt dieser Ansatz nicht zu dem Ziel, die Proben in Subklassen einteilen zu können. Das Programm von M. Eisen bietet verschiedene Algorithmen zum Clustern an. Werden diese auf die vorliegenden Datensätze angewandt, ergeben sich für die Leukämieproben immer wieder neue Cluster. Es ist nicht davon auszugehen, dass diese Cluster in spezifischen Probeneigenschaften ihren Ursprung finden. Sie entstehen vermutlich als zufällige Cluster.



**Abbildung 3-13 Clustern der Leukämieproben**

Die in diesem Versuch angedeuteten Cluster der leukämischen Proben lassen sich mit keinem anderen Algorithmus bestätigen. Verwendet wurden die mit EOS-H ermittelten Daten.

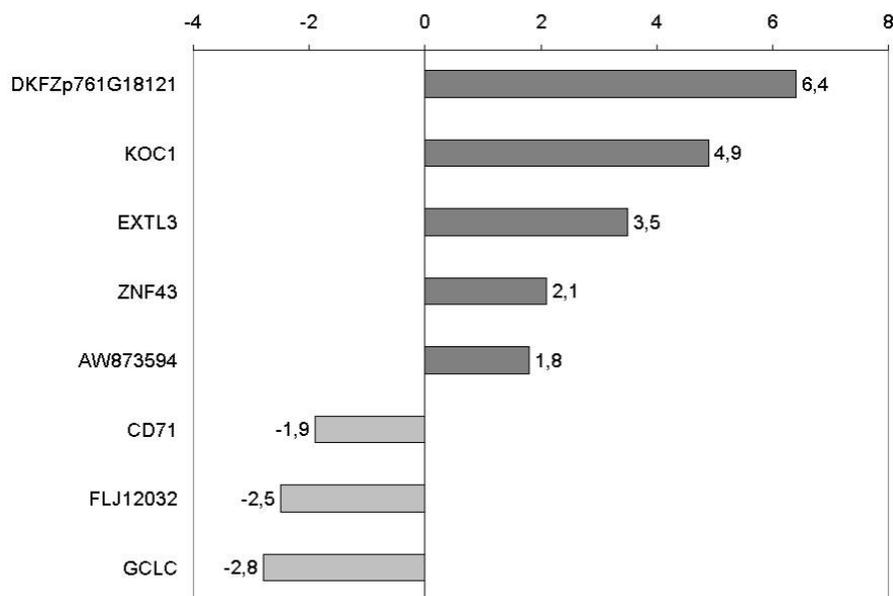
Das Clustern der mit HG-U133A untersuchten Proben liefert mit beiden Clustermethoden ebenfalls kein Ergebnis. Es gibt in diesem Datensatz keine deutlichen Ausreißer wie bei den EOS-H Daten, aber es lassen sich genauso deutlich auch keine Subklassen der cALL anhand der Expressionsdaten festlegen. Selbst die durch die Translokation TEL/AML1 genetisch eindeutig charakterisierten Patientenproben bildeten kein eigenes Cluster.

Die leukämischen Proben lassen sich anhand klinischer Daten in Gruppen zusammenfassen. Alle Informationen zu den klinischen Parametern sind in Tabelle 6-1 zusammengetragen. Das auffallendste Merkmal ist der Nachweis der TEL/AML1-Translokation in sieben Proben. Diese Proben sollen in 3.8 gesondert untersucht werden. Da der unüberwachte Ansatz des Clusters zur Identifikation von Untergruppen zu keinem Ergebnis führte, soll nun versucht werden, über klinische Parameter Gruppen zu definieren. Beim Vergleich diese Gruppen sollen dann für die jeweilige Gruppe spezifisch differentiell exprimierte Gene identifiziert werden. Hintergrund eines solchen Ansatzes ist unter anderem die Suche nach Angriffspunkten für Therapieformen, die maßgeschneidert für distinkte Formen der cALL sind. Die hier

dargestellten Analysen beschränken sich auf die mit HG-U133A erhobenen Daten. Als mögliche Gruppenparameter wurden untersucht: Alter bei Diagnose („jünger als 5 Jahre“ gegen „älter als 10 Jahre“), Geschlecht, Risikogruppen („Risikogruppe 1“ gegen „Risikogruppe 3“), Anzahl der Leukozyten im peripheren Blut zum Zeitpunkt der Diagnose („weniger als 10.000 Zellen/ $\mu\text{l}$ “ gegen „mehr als 50.000 Zellen/ $\mu\text{l}$ “), Anzahl der Leukozyten am Tag 8 („weniger als 1900 Zellen/ $\mu\text{l}$ “ gegen „mehr als 6000 Zellen/ $\mu\text{l}$ “), prozentualer Anteil der Blasten im Knochenmark am Tag 15 („weniger als 30%“ gegen „mehr als 30%“). Die verwendeten Grenzwerte für die Gruppeneinteilung wurden so gewählt, dass sich aus den zur Verfügung stehenden Proben möglichst eindeutige Gruppen bilden ließen. Dabei wurde die klinische Relevanz der betrachteten Parameter beachtet. Das bedeutet letztlich, daß bei der Bildung der Gruppen nicht immer alle Proben berücksichtigt wurden. Der Patient ALL31ARB war beispielsweise zum Zeitpunkt der Diagnose sieben Jahre alt und passte damit in keine der beiden Altersgruppen. Unter Verwendung der bekannten Abfragekriterien wurden bemerkenswerterweise nur bei zwei Gruppen signifikant differentiell exprimierte Gene identifiziert. Zum einen sind dies die „Altersgruppen“, zum anderen die Gruppen mit unterschiedlicher Anzahl an Leukozyten im peripheren Blut zum Zeitpunkt der Diagnose (White Blood Count, WBC).

Die beiden Gruppen „jünger als 5 Jahre“ (ALL25HCS, ALL33LUM, ALL36IHP, ALL38IER, ALL39AHS, ALL40EEL, ALL41APS, ALL43MED, ALL46ROY, ALL50GAJ, ALL53TAB, ALL54TEY, ALL56WET) und „älter als 10 Jahre“ (ALL45NUK, ALL49ERD, ALL51ERT, ALL52HCS, ALL57SUN) wurden so gewählt, dass sie sich möglichst eindeutig voneinander unterscheiden. Einige Proben bleiben bei dieser Abfrage unberücksichtigt (ALL31ARB, ALL34REH, ALL37UEH, ALL42ORK, ALL47HEM, ALL48PIL, ALL55ZAB). In verschiedenen Übersichtsarbeiten wurden Daten mehrerer Studien zusammengetragen, die die Einteilung der beiden Altersgruppen nahelegen (Hrusak, 2002; Pui, 2000). Dort wird von zwei Altersgruppen berichtet, in denen cALL-Fälle akkumulieren: bei Kindern im Vorschulalter und bei Kindern zwischen neun und zehn Jahren. Für die beiden Altersgruppen wurden mit dem vorliegenden Datensatz acht Gene identifiziert, die eindeutig zwischen diesen beiden Gruppen unterscheiden können (vgl. Abbildung 3-14). Im einzelnen sind das: GCLC, KOC1, ZNF43, CD71 (TFRC), EXTL3, und die drei nicht näher charakterisierten Transkripte EST (AW873594), FLJ12032 und DKFZp761G18121. GCLC ist die katalytische Untereinheit der Glutamat-Cysteinyligase, dem ersten limitierenden Enzym der Glutathionsynthese. KOC1 kodiert ein Protein, welches an die mRNA des Wachstumsfaktors IGF2 binden kann und so vermutlich dessen Biosynthese reguliert (Nielsen, 1999). CD71 (TFRC) kodiert den Transferrinrezeptor. EXTL3 steht für Exotosin-like-3 und ist vermutlich in die

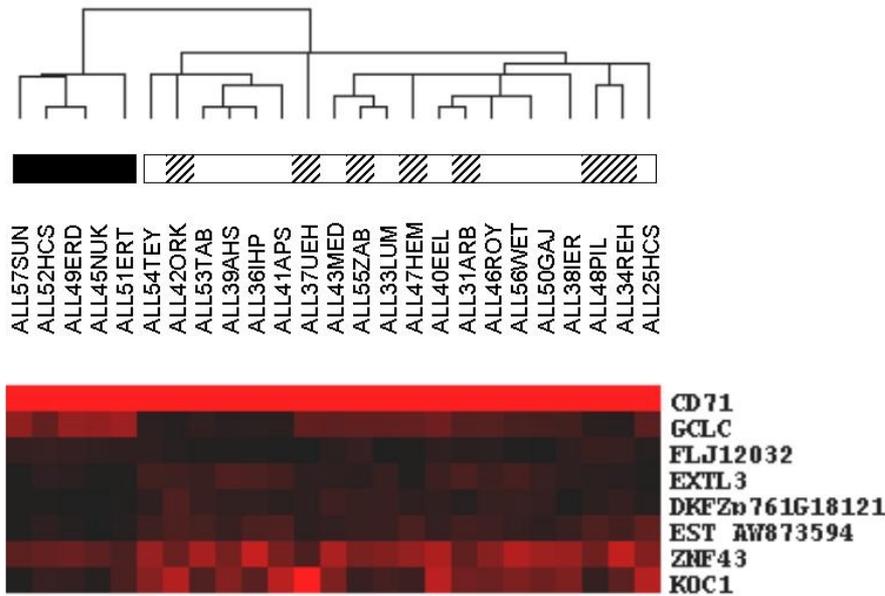
Heparinsynthese involviert. Für die verbleibenden Transkripte stehen keine weiteren Informationen zur Verfügung.



**Abbildung 3-14 Differenziell exprimierte Gene der „Altersgruppe“**

In der Expression dieser acht Gene unterscheiden sich die Patienten, die bei Diagnosestellung älter oder jünger als zehn Jahre waren. Angegeben sind die FC-Werte des Expressionsunterschieds. Positive Werte (dunkle Balken) bedeuten eine erhöhte Expression in der Gruppe der Patienten „jünger als zehn Jahre bei Diagnose“. Negative Werte (helle Balken) stehen für eine erhöhte Expression in der Gruppe „älter als zehn Jahre bei Diagnose“.

Diese acht ausgewählten Transkripte reichen aus, die leukämischen Proben entsprechend dem Alter der Patienten bei Diagnosestellung zu klassifizieren. In Abbildung 3-15 ist das Ergebnis der Klassifikation dargestellt. Auf der linken Seite stehen die Proben der Gruppe „älter als 10 Jahre“ korrekt alleine. Die Gruppe auf der rechten Seite vereint die Proben aller Patienten, die bei der Diagnose jünger als zehn Jahre waren. Das heißt die ursprünglich ausgewählte Gruppe der Patienten, die nicht älter als fünf Jahre waren, vermischt sich mit den Patienten, die zur Bestimmung der acht Gene nicht herangezogen worden sind, weil sie zum Zeitpunkt der Diagnosestellung zwischen fünf und zehn Jahren alt waren. Demnach lassen sich die Leukämiepatienten in die beiden Gruppen „jünger als 10 Jahre“ oder „älter als 10 Jahre“ einteilen. Bei den vorangegangenen Clusterexperimenten, bei denen alle Gene herangezogen wurden oder die Gene, in denen sich die leukämischen von den gesunden Proben unterschieden, sind diese beiden Altersgruppen nicht gesehen worden.

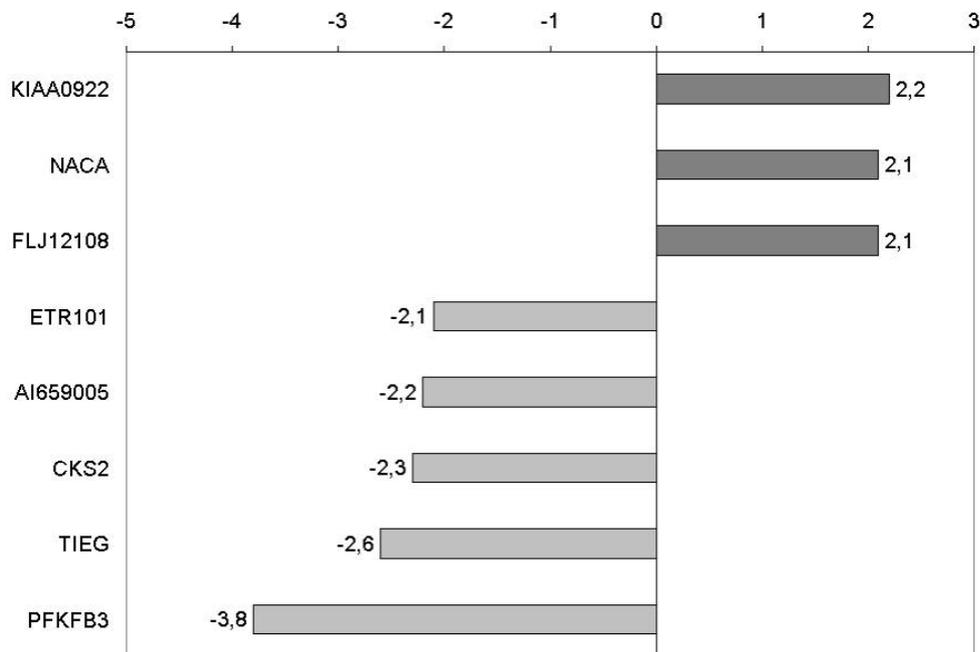


**Abbildung 3-15 Klassifikation der Leukämien anhand der „Altersgruppen-Gene“**

Mit Hilfe des reduzierten Datensatzes lassen sich die leukämischen Proben entsprechend dem Alter der Patienten bei Diagnosestellung klassifizieren. In der linken Gruppe sammeln sich die Patienten, die älter als zehn Jahre bei Diagnose waren (grauer Kasten); auf der rechten Seite stehen die jüngeren Patienten (weißer Kasten). Dies schließt die Patienten ein, die jünger als zehn Jahre waren (Markierung mit Schraffur), deren Proben aber nicht zur Identifikation der Gene herangezogen wurden. Diese Proben clustern entsprechend dem Alter der Patienten in der Gruppe „jünger als 10 Jahre“.

Als zweite Gruppe lieferte die Abfrage zwischen Patienten mit unterschiedlichem initialen WBC bei Diagnose ein Ergebnis. Zwischen den Gruppen „weniger als 10.000 Zellen/ $\mu$ l“ (ALL50GAJ, ALL51ERT, ALL33LUM, ALL38IER, ALL41APS, ALL42ORK, ALL43MED, ALL45NUK, ALL53TAB und ALL55ZAB) und „mehr als 50.000 Zellen/ $\mu$ l“ (ALL52HCS, ALL46ROY, ALL48PIL, ALL25HCS, ALL31ARB, ALL54TEY und ALL57SUN) können acht differentiell exprimierte Gene identifiziert werden, von denen drei in den Patienten mit initial hoher Leukozytenzahl und fünf in Patienten mit niedriger Leukozytenzahl stärker exprimiert sind (Abbildung 3-16). Ebenso wie der Nachweis bestimmter Translokationen oder das Alter der Patienten bei Diagnose, stellt der WBC einen prognostisch wichtigen Marker dar. Es gibt eine schlechtere Prognose für Patienten, die initial mehr als 50.000 Zellen/ $\mu$ l haben (Friedmann, 2000; Pui, 2000). Das Gen NACA ist in den Patienten mit hohem WBC stärker exprimiert. NACA ist speziell im Knochen während der Entwicklung exprimiert und funktioniert als Koaktivator der Transkription. In den Patienten mit niedrigerem WBC sind ETR101, CKS2 und TIEG stärker exprimiert. Der Transkriptionsfaktor ETR101 wird in vielen verschiedenen leukämischen Zelltypen nach Induktion mit dem Phorbolster TPA aktiviert. In unetrschiedlichen Zellkulturmodellen induziert TPA Proliferation und Differenzierung. CKS2 ist Bestandteil der CDC28 Proteinkinase und reguliert die Aktivität von CDKs.

TIEG ist ein Transkriptionsfaktor, der vermutlich in den TGF $\beta$ -abhängigen Signaltransduktionsweg involviert ist. Für KIAA0922, FLJ12108 und AI659005 liegen keine weiteren Informationen vor.



**Abbildung 3-16** Differenziell exprimierte Gene der „WBC-Gruppe“

Zwischen den Gruppen „weniger als 10.000 Zellen/ $\mu$ “ und „mehr als 50.000 Zellen/ $\mu$ “ können acht differenziell exprimierte Gene identifiziert werden, von denen drei in den Patienten mit initial hoher Leukozytenzahl (dunkelgraue Balken, positiver FC-Wert) und fünf in Patienten mit niedriger Leukozytenzahl (hellgraue Balken, negativer FC-Wert) stärker exprimiert sind.

Im Gegensatz zu den Genen der „Altersgruppe“ lassen sich mit den differenziell exprimierten Genen der „WBC-Gruppe“ die Proben nicht wieder klassifizieren. Sowohl beim hierarchischen Clustern als auch bei der PCA war es nicht möglich, die leukämischen Proben mit diesen Genen den Gruppen wieder korrekt zuzuordnen. Alle acht identifizierten Gene erfüllten die Kriterien eines mindestens zweifachen Unterschieds und eines *t*-Test Ergebnisses von  $p < 0,001$ . Dennoch war es weder unter Berücksichtigung aller Leukämieproben, noch unter Beschränkung auf die Proben, die für die Identifikation dieser Gene verwendet wurden, möglich korrekt zu klassifizieren.

### 3.8 *cALL* mit *t*(12;21)(p13;q22) *TEL/AML1*-Translokation

#### 3.8.1 Identifikation differenziell exprimierter Gene

Von den insgesamt 35 mit Hilfe der Genechips untersuchten Patienten hatten sieben Patienten eine nachgewiesene *TEL/AML1*-Translokation. Bei einer Patientin konnte eine Translokation am Gen *MLL* gezeigt werden. Für die übrigen Patienten wurden keine Translokationen oder anderen chromosomalen Aberrationen wie Hyperloidien

nachgewiesen oder untersucht. Mit den beiden Clusterverfahren allein (vgl. 3.7) war es nicht möglich, innerhalb der Leukämiegruppe Untergruppen zu definieren. Die Proben mit nachgewiesener TEL/AML1-Translokation bildeten bei diesem Ansatz ebenfalls keine klar erkennbare Untergruppe. S. Armstrong (Armstrong, 2001) hat bereits gezeigt, wie sich die Genexpressionsprofile von Leukämiepatienten mit einer MLL-Translokation von denen anderer Patienten ohne diese Translokation unterscheiden. Ihm ist es gelungen, mit Hilfe der Genexpressionsprofile eine neue ALL-Untergruppe zu definieren und zu zeigen, dass sich durch die Translokation bedingt die Genexpressionsprofile unterscheiden. Die Autoren schlagen sogar vor, aufgrund der unterschiedlichen Expressionsmuster, gestützt durch klinische Beobachtungen – Patienten mit MLL-Translokation unterscheiden sich in klinischen Parametern von den anderen ALL-Patienten, so zum Beispiel durch ein besseres Ansprechen auf eine Chemotherapie mit Antimetaboliten - neben den bekannten Formen der ALL und AML die Mixed-Lineage Leukemia (MLL) mit MLL-beteiligter Translokation als eigene Krankheitsentität zu definieren. Wenngleich es mit dem oben erwähnten unüberwachten Ansatz nicht gelungen ist, die TEL/AML1-positiven cALL-Proben zu clustern, lassen sich Gene identifizieren, in denen sich diese Leukämiezellen spezifisch von den gesunden Kontrollzellen unterscheiden.

Fünf der sieben Patienten mit TEL/AML1-Translokation wurden mit HG-U133A untersucht: ALL34REH, ALL39AHS, ALL43MED, ALL50GAJ, ALL55ZAB. Diese Gruppe kann getrennt von den anderen Patientenproben mit den sechs Kontrollen gesunder B-Lymphozyten auf diesem Chiptyp betrachtet werden (TEL/AML1-cALL vs CB). Ergänzend wurden die Leukämien ohne diese Translokation mit denselben Kontrollen verglichen (Nicht-TEL/AML1-cALL vs CB) und die Leukämieproben mit Translokation mit den Leukämieproben ohne Translokation (TEL/AML1-cALL vs Nicht-TEL/AML1-cALL). Die Idee für diese zusätzlichen Abfragen dabei ist, am Ende durch den Vergleich der Datensätze die Gene herauszufiltern, die exklusiv für die Leukämien mit TEL/AML1-Translokation sind. Sämtliche in diesem Abschnitt durchgeführten Analysen und daraus abgeleitete Aussagen beziehen sich nur auf die Experimente, die mit HG-U133A gemacht worden sind. Für die Identifikation differentiell exprimierter Gene wurden die folgenden bekannten Abfragen durchgeführt: die Mittelwerte für ein Gen aus beiden Gruppen sollen sich mindestens um den Faktor 2 unterscheiden ( $FC > 2$ ), weiterhin muss das Ergebnis des  $t$ -Tests mit  $p < 0,01$  sein. Mit dieser Abfrage wurden 854 Sondensätze identifiziert, die eine verstärkte Expression der entsprechenden Gene in den Leukämien zeigten. 897 Sondensätze detektierten in diesen Proben herunterregulierte Gene.

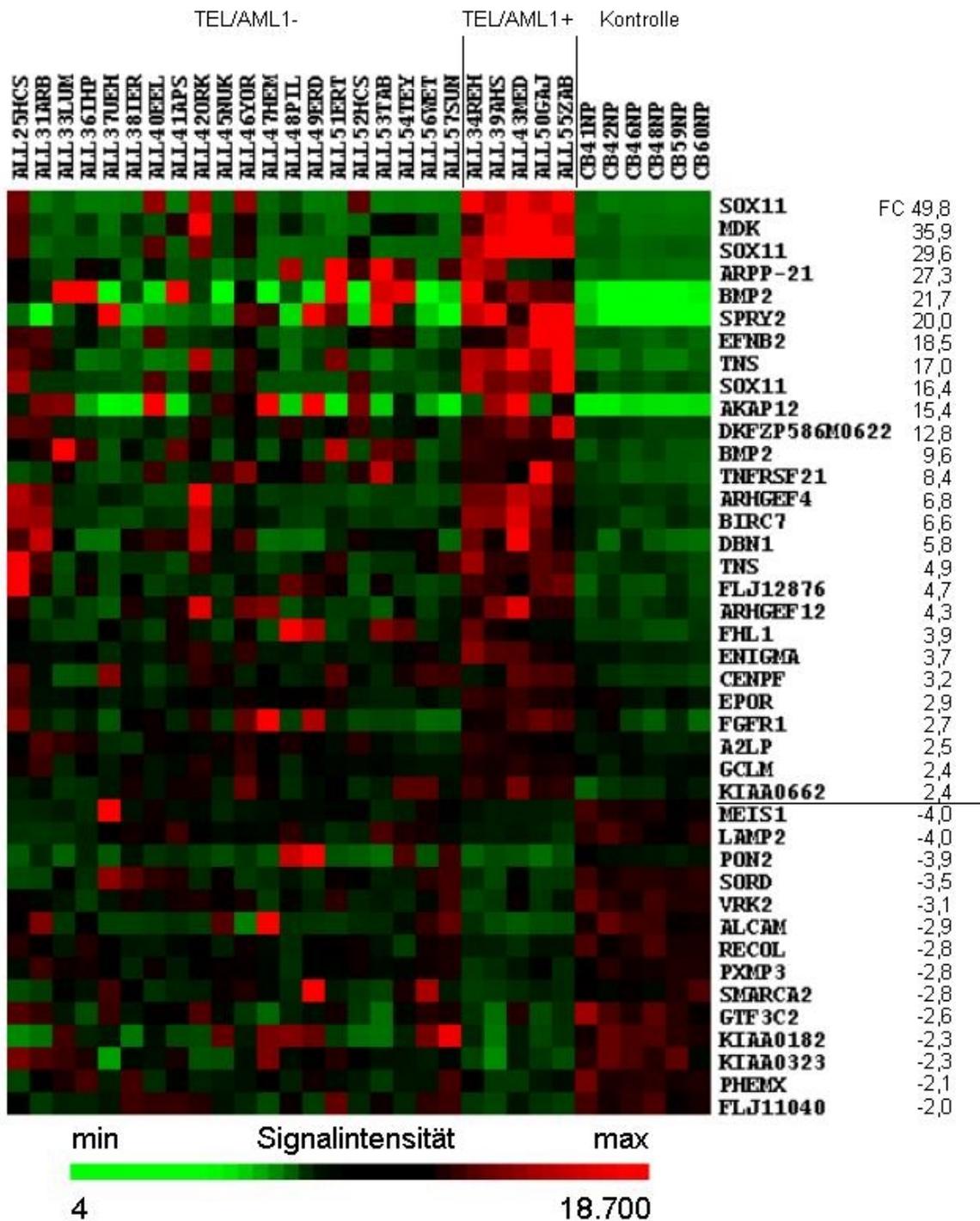


Abbildung 3-17 Differenziell exprimierte Gene TEL/AML1-positiver cALL und Kontrollen

In der Expression dieser 41 Gene unterscheiden sich die Patienten mit nachgewiesener TEL/AML1-Translokation exklusiv von den gesunden Kontrollen. Die gemessenen Signalwerte werden in einer Falschfarbendarstellung wiedergegeben (Minimum 4, Maximum 18.700, der hier betrachteten Gene).

Unter Verstärkung der Abfragekriterien (vgl. 3.4.2) wurde die Zahl detektierter Gene eingeschränkt. Ein differenziell exprimiertes Gen soll demnach in jeweils allen Proben

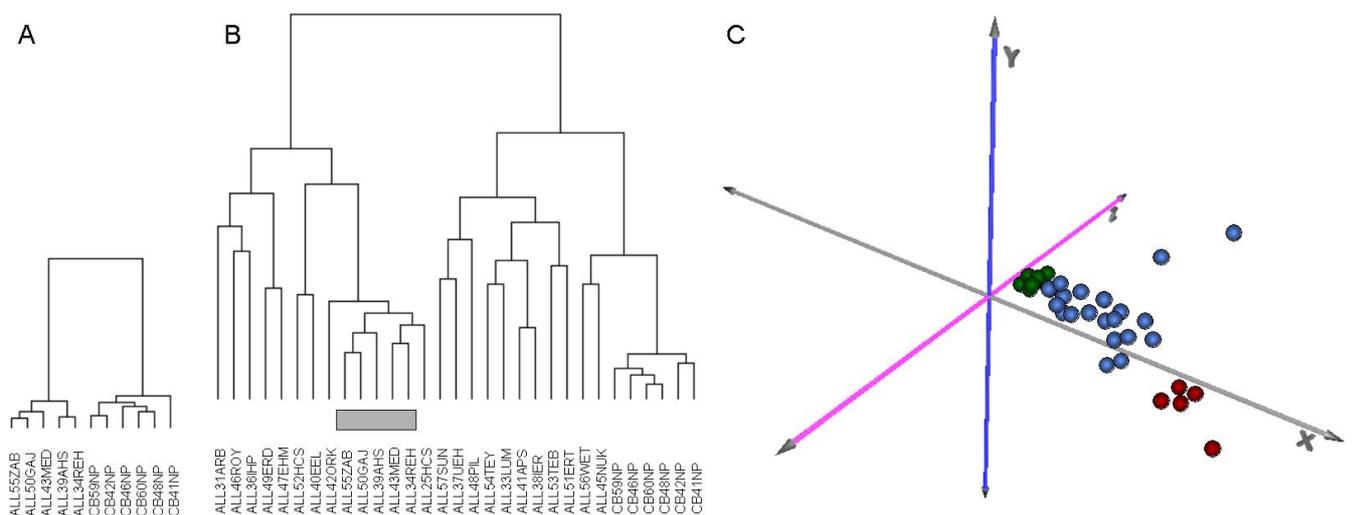
der einen Untersuchungsgruppe exprimiert und gleichzeitig in sämtlichen Proben der anderen Untersuchungsgruppe nicht exprimiert sein. Mit dieser Abfrage konnten 135 Gene beziehungsweise Sondensätze identifiziert werden, in denen sich die TEL/AML1-positiven Leukämien von den Kontrollen unterscheiden. Von diesen 135 sind 55 Gene in den Leukämien heraufreguliert (maximaler FC 86,1) und 80 Gene herunterreguliert (maximaler FC 273,8). Aus dem Vergleich sämtlicher Leukämieproben mit den Kontrollen (vgl. 3.4.2) resultieren 87 identifizierte Sondensätze, die in beiden Abfragen gefunden wurden. Somit verbleiben 48 Gene, die den Unterschied zwischen TEL/AML1-positiven Leukämien und den Kontrollen ausmachen. Die Untersuchung der Nicht-TEL/AML1-Leukämien mit den Kontrollen liefert 62 differentiell exprimierte Gene (25 herauf- und 37 herunterregulierte Gene, diese Daten werden hier nicht weiter diskutiert). Beim Vergleich mit den Daten aus der Analyse TEL/AML1-positiver Leukämien mit den Kontrollen erhält man 21 doppelt detektierte Gene, die allerdings bereits vollständig in den 86 Genen aus dem Vergleich aller Leukämien mit den Kontrollen als doppelt detektiert enthalten sind. Aus der Analyse der mit EOS-H Genechips ermittelten Daten wurden sieben Gene als doppelt detektiert gefunden. Nach Abzug dieser ebenfalls in den vorangegangenen Analysen als differentiell exprimiert identifizierten Gene verbleiben 41 Gene, die exklusiv den Unterschied zwischen cALL-Patientenproben mit TEL/AML1-Translokation und gesunden Kontrollen ausmachen (Abbildung 3-17). Hierbei handelt es sich um 27 in den Leukämien heraufregulierte (maximaler FC 49,8) und 14 herunterregulierte Gene (maximaler FC 4,0). Diese Gene werden in Abschnitt 3.8.3 näher untersucht.

Die beiden Gruppen leukämischer Proben (mit oder ohne TEL/AML1-Translokation) unterscheiden sich in der Expression von nur elf Genen (zwölf Sondensätze), von denen zwei Gene in den TEL/AML1-positiven Proben herauf- und neun herunterreguliert sind. Diese Gene werden im Detail in Abschnitt 3.8.4 betrachtet (vgl. dort auch Abbildung 3-20).

### 3.8.2 Klassifikation der Proben mit differentiell exprimierten Genen

Entsprechend Kapitel 3.4.3 soll mit diesen beiden Datensätzen klassifiziert werden. Mit dem ersten Datensatz der 41 differentiell exprimierten Gene, in denen sich die TEL/AML1-positiven Leukämieproben exklusiv von den Kontrollen unterscheiden, soll überprüft werden, ob sich die Proben wieder in die einzelnen Gruppen korrekt klassifizieren lassen. Unter Verwendung der sechs Leukämie- und fünf Kontrollproben, mit denen dieser Datensatz generiert wurde, lassen sich die Proben eindeutig voneinander trennen (Abbildung 3-18-A). Nimmt man die verbliebenen 20 Leukämieproben ohne die Translokation hinzu, trennen sich die gesunden Kontrollen

ebenso eindeutig von den leukämischen Proben ab. Die Trennung innerhalb der Leukämien hingegen ist nicht sehr scharf. Die Abbildung 3-18-B zeigt, dass die fünf Leukämien mit der Translokation eng zusammen fallen. Diese Proben zeigen die größte Ähnlichkeit untereinander. Sie bilden aber keine eindeutige Gruppe und trennen sich nicht sehr deutlich von den anderen Leukämien ab. Der geringe Unterschied zwischen den Leukämien mit und ohne TEL/AML1-Translokation wird bereits an der geringen Anzahl von Genen deutlich, in denen sich diese beiden Gruppen unterscheiden. Mit der PCA lässt sich der Unterschied in den drei Gruppen besser verdeutlichen (Abbildung 3-18-C). Die Leukämien mit der Translokation stehen als Gruppe abseits der übrigen Leukämien. Die Kontrollen zeigen den größten Unterschied zu den TEL/AML1-positiven Leukämien und haben im dreidimensionalen Achsenkreuz die größte Entfernung.

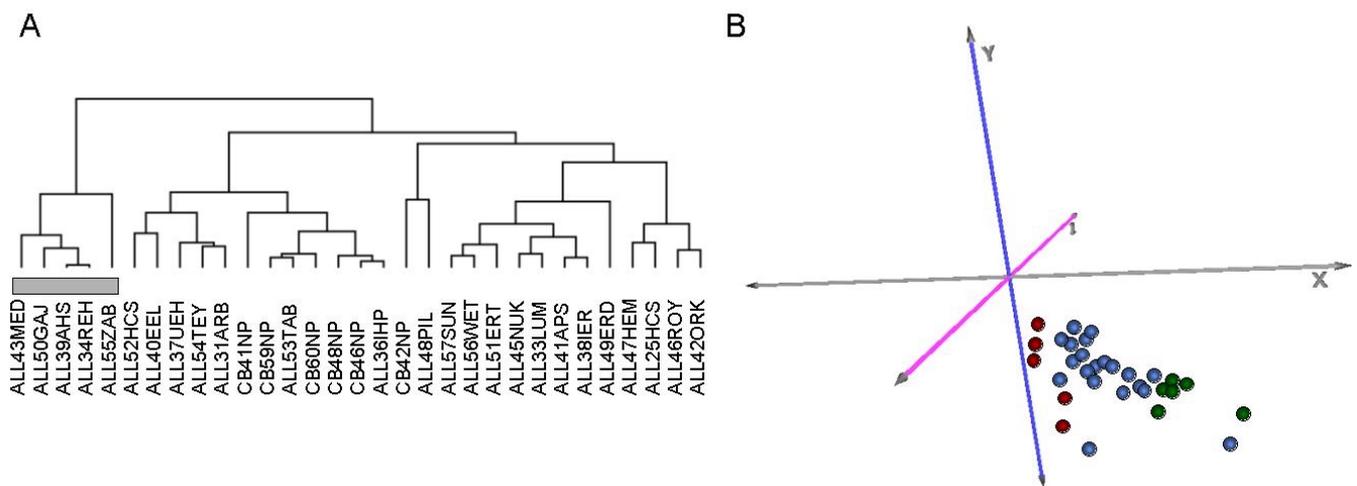


**Abbildung 3-18 Klassifikation aller Proben mit Genen, die in Leukämien mit einer TEL/AML1-Translokation differenziell exprimiert sind.**

A: Werden nur die Leukämieproben mit TEL/AML1-Translokation und die Kontrollen analysiert, werden diese eindeutig aufgeteilt. B: Die Leukämien mit der Translokation fallen eng zusammen (grauer Kasten). Diese Proben zeigen die größte Ähnlichkeit untereinander, bilden aber keine eigene Gruppe und trennen sich nicht sehr deutlich von den anderen Leukämien ab. Der geringe Unterschied zwischen den Leukämien mit und ohne TEL/AML1-Translokation wird bereits an der geringen Anzahl von Genen deutlich, in denen sich diese beiden Gruppen unterscheiden. C: In der PCA wird der Unterschied in beiden Gruppen deutlicher. Die Proben mit Translokation (rot) sondern sich von den übrigen Leukämien (blau) etwas ab. Die Kontrollen (grün) stehen ebenfalls konzentriert als Gruppe zusammen.

Auch die elf differenziell exprimierten Gene, in denen sich die Leukämien mit TEL/AML1-Translokation von den übrigen Leukämien ohne diese Translokation unterscheiden, reichen aus, um die Proben bei der Klassifikation korrekt zu sortieren. Die Abbildung 3-19-A zeigt, dass die fünf Proben eindeutig als Gruppe erkannt werden. Die verbleibenden Leukämien sind mit den Kontrollen vermischt. Das ist allerdings

insofern nicht verwunderlich, da in dem zum Klassifizieren verwendeten Datensatz nur Informationen zu Unterschieden zwischen Proben mit oder ohne TEL/AML1-Translokation stecken. Die gesunden Kontrollen sortierter B-Zellen sollten die Translokation nicht tragen. Die PCA stellt die Unterschiede in den drei Gruppen (Leukämien mit oder ohne TEL/AML1-Translokation und Kontrollen) deutlicher dar (Abbildung 3-19). Die gesunden Kontrollen zeigen in dieser Analyse eine größere Nähe zueinander.



**Abbildung 3-19 Klassifikation aller Proben mit Genen, die zwischen Leukämien mit und ohne TEL/AML1-Translokation unterscheiden**

A: Elf Gene reichen aus, um die Leukämien mit einer TEL/AML1-Translokation von den verbleibenden Proben abzutrennen. Diese Proben sind durch einen grauen Kasten hervorgehoben. B: In der PCA werden die Unterschiede in den drei Gruppen deutlicher (Leukämien mit TEL/AML1-Translokation sind rot, Leukämien ohne Translokation sind blau und die Kontrollen grün).

### 3.8.3 Differenziell exprimierte Gene der TEL/AML1-cALL

Wie im Kapitel 3.6 wurden die 41 differenziell exprimierten Gene hier ebenfalls in funktionelle Gruppen eingeteilt und näher betrachtet, um auf diesem Weg mögliche Zusammenhänge zwischen Genexpressionsprofil und Pathologie der cALL mit TEL/AML1-Translokation zu erkennen. Auf die verbliebenen, bereits in den vorangegangenen Analysen detektierten und besprochenen Gene soll am Ende des Abschnitts nur kurz eingegangen werden.

Interessanterweise gibt es in dieser Gruppe keine weiteren Gene, die allein in direkter Verbindung mit der Reifung von B-Lymphozyten stehen. Die Translokation scheint auf diesen Aspekt der leukämischen Pathologie keinen weiteren Einfluss zu nehmen.

Es wurden drei Gene mit einer Funktion in DNA-Reparatur, Transkription oder Proteinsynthese identifiziert, die alle herunterreguliert sind. Das entspricht den

Ergebnissen aus 3.6.2.3, wo die heraufregulierten Gene ihre Funktion in der Replikation von DNA hatten und die herunterregulierten Gene ebenfalls Aufgaben in Transkription und Translation übernahmen. Die DNA-Helikase RECQL ist hauptsächlich in die Reparatur UV-geschädigter DNA involviert. GTF3C2 ist ein allgemeiner Transkriptionsfaktor, der RNA-Polymerasen bei der Initiation der RNA-Synthese dient und VRK2 kodiert eine Kinase, die in die Replikation von DNA involviert ist.

Es sind zwölf Gene detektiert worden, deren Genprodukte Funktionen in der Signaltransduktion übernehmen, von denen elf in den Leukämiezellen herauf- und nur eines herunterreguliert sind. Entsprechend Kapitel 3.6.2.4 lassen sich die in Signaltransduktion involvierten Gene oder Genprodukte verschiedenen Signaltransduktionswegen zuordnen. Erythropoietin ist einer der wichtigsten Regulatoren von Wachstum und Differenzierung hämatopoetischer Zellen. EPOR kodiert den entsprechenden Rezeptor und ist in den Zellen der Leukämie stärker exprimiert als in den Kontrollen. Für 60% der Patienten mit AML und für 29% der Patienten mit ALL ist eine erhöhte Expression von EPOR bekannt (Takeshita, 2002). Im vorliegenden Datensatz ist EPOR in acht der 25 mit HG-U133A untersuchten Patientenproben nachweisbar. Die darin enthaltenen Patienten mit TEL/AML1-Translokation exprimieren alle EPOR. Damit ist EPOR eines der Gene, das einen Unterschied zwischen den cALL-Proben mit oder ohne TEL/AML1-Translokation macht. ARHGEF4 und ARHGEF12 sind zwei Proteine, die die Aktivität von Mitgliedern der RAS-Familie modulieren können und die in den leukämischen Proben ebenfalls heraufreguliert sind. Der potentielle Tumorsuppressor PHEMX, mit Funktionen in der Signaltransduktion, ist herunterreguliert. PHEMX wird wahrscheinlich in geringen Mengen ubiquitär exprimiert und ist besonders in hämatopoetischen Zellen nachzuweisen.

Bei den differentiell exprimierten Genen mit Funktionen in Differenzierung und Entwicklung besteht ein ähnliches Ungleichgewicht. Es finden sich sieben herauf- und nur drei herunterregulierte Gene. Auf die zentrale Bedeutung von BMP2 und die damit verbundenen TGF $\beta$ /BMP-Signaltransduktionswege für die Reifung der B-Lymphozyten ist bereits in Abschnitt 3.6.2.4 hingewiesen worden. BMP2 ist in den Proben mit TEL/AML1-Translokation deutlich stärker exprimiert als in den Kontrollen. Ein Inhibitor in RAS/RAF-Signalwegen ist SPRY2, welches in den Leukämien ebenso heraufreguliert ist. Auf diesem Weg ist SPRY2 an der Inhibition von FGF-Signalen (Fibroblast Growth Factor) während der Entwicklung beteiligt (Yusoff, 2002).

Es sind in diesem Datensatz zwei differentiell exprimierte Gene detektiert worden, die beide in einem engen Zusammenhang mit der Modulation von TNF-vermittelten Apoptosesignalen stehen. TNFRSF21 (DR6) und BIRC7 (Livin) sind beide

heraufreguliert. Das Expressionsmuster dieser Transkripte passt gut mit den Funktionen zusammen und deutet darauf hin, wie Apoptosesignale in den leukämischen Zellen unterbunden werden könnten. TNFRSF21 ist vermutlich an entzündlichen Reaktionen und an Immunreaktionen beteiligt und reguliert nach TNF-Stimulation den programmierten Zelltod. Eine Expression dieses Rezeptors konnte in verschiedenen Karzinomen und Melanomen gezeigt werden, jedoch nicht in malignen Zellen hämatopoetischen Ursprungs. Das ebenfalls heraufregulierte BIRC7 ist direkt in der Lage, die TNF/TNFRSF21-vermittelte Apoptose zu inhibieren.

Für drei weitere Transkripte ist eine Funktion speziell in der neuronalen Entwicklung beschrieben worden: SOX11, DBN1 und MDK. SOX11 ist ein Transkriptionsfaktor. Das Verwandte SOX4 ist beim Vergleich aller Leukämien mit den gesunden Kontrollen ebenfalls detektiert worden. DBN1 kodiert ein actinbindendes Protein, das in der neuronalen Entwicklung exprimiert wird und das eine Funktion in der Morphogenese zu haben scheint. Für MDK wurde die Expression in verschiedenen Tumoren bereits berichtet. Eine Überexpression von MDK scheint mit einer schlechten Prognose beim Blasenkrebs einherzugehen. EFNB2 ist nicht nur in die Neurogenese involviert, sondern auch an Angiogenese und der Bildung weiterer Organe beteiligt. EFNB2 zeigt eine Überexpression in Karzinomen des Magens. Die Hochregulation von EFNB2 in Zellen von Neuroblastomen korreliert mit einem günstigen Krankheitsverlauf (Tang, 2000).

Die beiden Gene, die in die funktionelle Gruppe von Zellzyklus und Proliferation einzuteilen sind, zeigen in den leukämischen Zellen eine stärkere Expression. Dies passt gut zu den Ergebnissen aus Abschnitt 3.6.2.6, wo zehn der 13 dort identifizierten Gene ebenfalls heraufreguliert waren. Zum einen handelt es sich hier um CENPF, einem Protein der Centromere, welches nur vom Beginn der Mitose bis zur Metaphase mit den Kinetochoren transient assoziiert ist. Zum anderen ist es CDC2, die katalytische Untereinheit eines Proteinkinasekomplexes, der als M-Phase Promoting Factor bezeichnet wird. Dieser induziert den Beginn der Mitose und ist bei den Eukaryoten universell vertreten. Das Holoenzym setzt sich aus den beiden Untereinheiten CDC2 und CCNB2 zusammen. Letzteres ist bereits mit Hilfe der EOS-H Genechips als in den cALL-Proben heraufreguliert identifiziert worden. Tendenziell sind beide Untereinheiten in allen leukämischen Proben stärker exprimiert als in den Kontrollzellen.

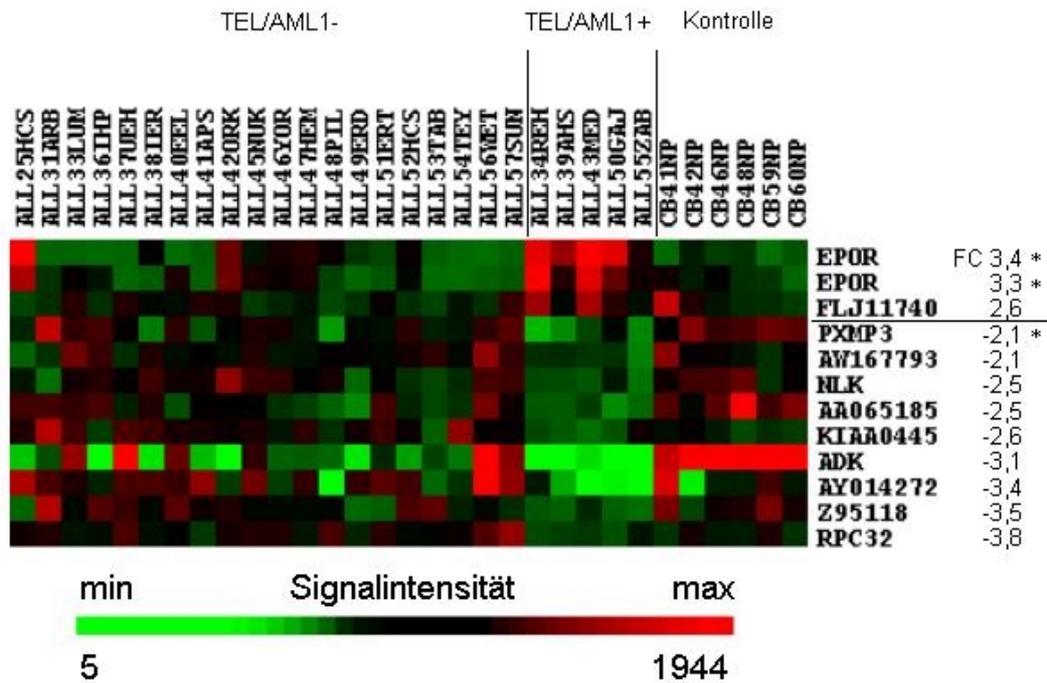
Es sind einige Gene identifiziert worden, die Proteine des Stoffwechsels kodieren und von denen lediglich eines in den Leukämien heraufreguliert ist. Hierbei handelt es sich um die Glutamat-Cysteinylgase GCLM, der regulatorischen Untereinheit der Glutamat-Cysteinsynthase. Zu den herunterregulierten Genen zählen die ubiquitär exprimierte Paraoxonase PON2 mit einer Funktion bei zellulären Reaktionen auf

oxidativen Stress, das peroxisomale Membranprotein PXP3 oder auch die Sorbitoldehydrogenase SORD.

Von den 135 Genen, die in den leukämischen Proben mit einer TEL/AML1-Translokation differentiell exprimiert sind, waren 93 bereits in den vorangegangenen Analysen der vollständigen Leukämiegruppe im Vergleich zu den Kontrollen detektiert und in den vorangegangenen Kapiteln beschrieben worden. Zu den wichtigsten Genen zählen CD10, SOX4, IL4R, BLR1 und CD23, die eine Rolle bei der Reifung von B-Lymphozyten spielen sowie LTF, SPAP1 und RNASE6, die immunologische Funktionen haben. ADAM28, MTX2, MAP3K5 und MELK sind an der Regulation von Differenzierung und Entwicklung beteiligt. E2F5, MKI67 und ESP1 übernehmen Funktionen beim Zellzyklus und in der Zellproliferation.

#### 3.8.4 Unterschiede der Leukämien mit und ohne TEL/AML1-Translokation

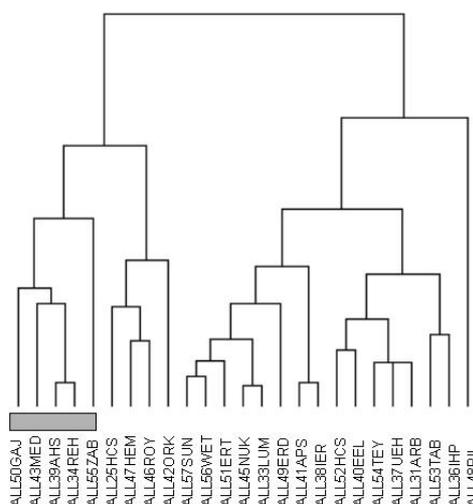
Die Leukämien mit und ohne TEL/AML1-Translokation unterscheiden sich in zwölf Sondenätzen (Abbildung 3-20). Detektiert wurden damit elf unterschiedliche Transkripte. Von diesen Transkripten sind drei bereits beim Vergleich der TEL/AML1-cALL mit den gesunden Kontrollen detektiert worden: EPOR (detektiert durch zwei unabhängige Sondenätze), ADK und PXP3. Unter den verbleibenden acht Transkripten kodieren sechs unbekannte Proteine oder sind ESTs. Die verbleibenden beiden Transkripte sind die nicht näher charakterisierte RNA-Polymerase RPC32 und die Kinase NLK. Beide Transkripte fehlen in den Proben mit der TEL/AML1-Translokation. In der GNF-Expressionsdatenbank wurde gezeigt, dass RPC32 in einigen Proben mit einem malignen Ursprung höher exprimiert ist als in den normalen Geweben. Die Nemo-Like Kinase NLK ist ein negativer Regulator in WNT-Signalwegen und phosphoryliert sowohl TCF (T-Cell Factor) als auch LEF (Lymphoid Enhancer Binding Factor) (Harada, 2002).



**Abbildung 3-20** Differentiell exprimierte Gene, in denen sich Leukämien mit und ohne TEL/AML1-Translokation unterscheiden

In der Expression dieser elf Gene unterscheiden sich die Patienten mit oder ohne nachgewiesener TEL/AML1-Translokation. Die gemessenen Signalwerte werden in einer Falschfarbendarstellung wiedergegeben (Minimum 5, Maximum 1944, der hier betrachteten Gene). Die mit Stern markierten Gene wurden bereits beim Vergleich der TEL/AML1-positiven cALL mit den Kontrollen detektiert.

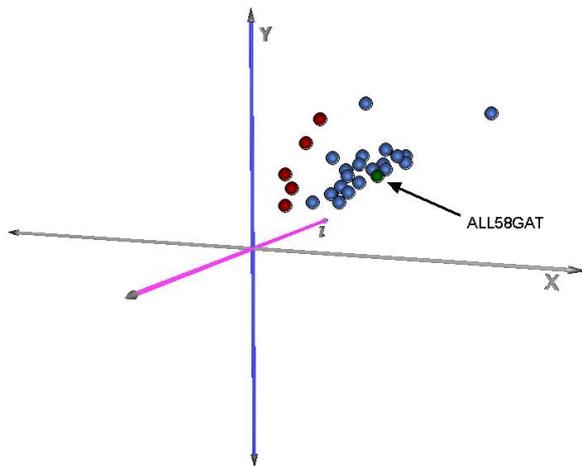
Diese elf Gene können vermutlich zur Identifikation leukämischer cALL Proben mit TEL/AML1-Translokation herangezogen werden. Werden die untersuchten Leukämien mit diesen elf Genen klassifiziert, so lassen sich die Proben entsprechend der Translokation gruppieren (Abbildung 3-21).



**Abbildung 3-21** Klassifikation aller Leukämieproben mit den Genen der TEL/AML1-Gruppe

Unter Verwendung der elf differentiell exprimierten Gene, die den Unterschied zwischen Leukämien mit oder ohne TEL/AML1-Translokation beschreiben, werden die Proben korrekt klassifiziert.

Nimmt man die Probe ALL58GAT hinzu, hierbei handelt es sich um die Lymphomprobe, bei der die TEL/AML1-Translokation nicht gezeigt worden ist, wird auch diese korrekt in die Gruppe der Proben ohne Translokation einsortiert. Zu bemerken ist, dass diese Probe nicht zur Identifikation differentiell exprimierter Gene verwendet worden ist (Abbildung 3-22).



**Abbildung 3-22 PCA der Lymphom- und cALL-Proben mit dem TEL/AML1-Gencluster**

In der PCA zeigen sich deutlich die beiden Gruppen der Leukämien mit (rot) und ohne (blau) TEL/AML1-Translokation. Die Lymphomprobe ALL58GAT (grün), bei der keine Translokation nachgewiesen wurde, wird den Leukämien ohne TEL/AML1-Translokation korrekt zugeordnet.

## 4 Diskussion

### 4.1 *Leukämische Blasten der cALL und gesunde B-Lymphozyten*

Die vorliegende Arbeit ist der erste Versuch, das Genexpressionsprofil leukämischer Zellen vom Typ cALL pädiatrischer Patienten mit gesunden B-Lymphozyten als Kontrollzellen zu vergleichen. Als Kontrollen dienten sortierte B-Lymphozyten aus dem Nabelschnurrestblut reifgeborener Kinder. Diese beiden Zelltypen haben ontogenetisch vermutlich denselben Ursprung. Mit Hilfe von zwei verschiedenen Genechiptypen wurden insgesamt 35 leukämische Proben, 15 Kontrollen sortierter B-Zellen aus Nabelschnurrestblut sowie acht weitere Kontrollen verschiedener Zellen aus Blut und Knochenmark erfolgreich analysiert und ausgewertet. Die bisher erschienenen Arbeiten zu Genexpressionsprofilen akuter lymphoblastischer Leukämien befassen sich vorrangig mit dem Versuch der Subklassifikation der ALL unter Berücksichtigung der unterschiedlichen chromosomalen Translokationen (Armstrong, 2001; Yeoh, 2002). Andere Arbeiten vergleichen direkt ALL und AML miteinander (Golub, 1999). In einer Veröffentlichung wurden gesunde Zellen zum Vergleich mit leukämischen Blasten herangezogen. Jedoch handelt es sich hierbei um komplettes Knochenmark und nicht um sortierte B-Lymphozyten. Die Detektion der Minimalen Resterkrankung (MRD, Minimal Residual Disease) nach erfolgter ALL-Therapie ist bedeutend für die Früherkennung und Behandlung von Rezidiven. Zu diesem Zweck wurden die Genexpressionsprofile des Knochenmarks von vier pädiatrischen cALL-Patienten mit dem Knochenmark zweier gesunder Menschen, einem Säugling und einem Erwachsenen verglichen (Chen, 2001b). Daraus wurden neun potentielle Transkripte abgeleitet, die sich zur Beobachtung der MRD eignen sollen.

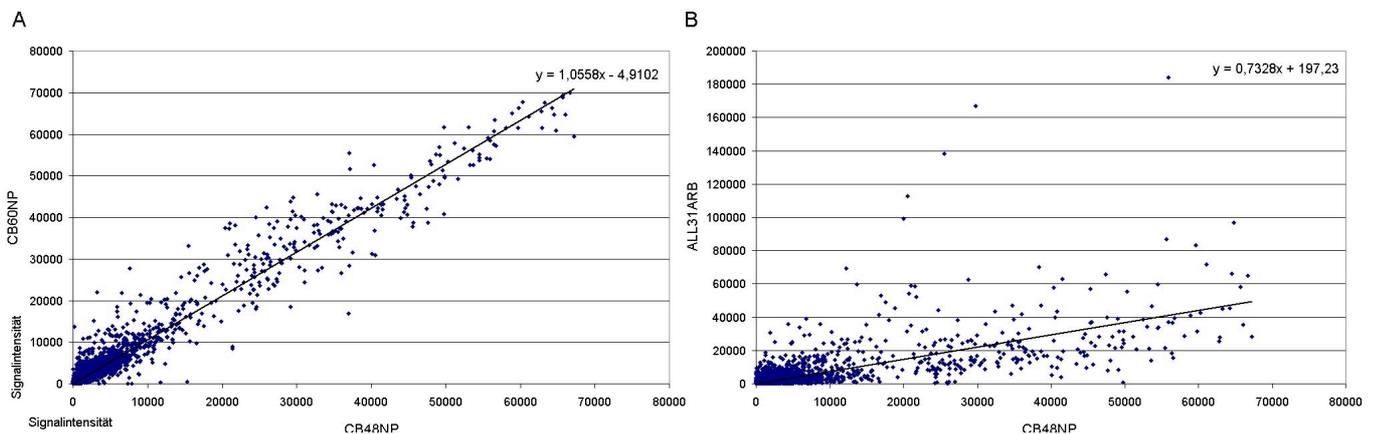
Der Vergleich der pathologisch veränderten Blasten mit gesunden Kontrollzellen wirft zwei Fragen auf. Erstens die Frage, welche Zellen im gesunden Menschen überhaupt geeignete Vergleichszellen zu den cALL-Blasten darstellen. Die leukämischen Zellen der cALL zeichnen sich besonders durch die Expression von CD19, CD10 und TdT aus. Die immunphänotypisch homologen Zellen der B-Zellreifungsreihe im gesunden Menschen sind hier nicht scharf einzugrenzen (vgl. Abbildung 1-3 und Abbildung 1-4). Sie reichen vom Stadium der frühen pro-B-Zelle bis zur großen prä-B-Zelle, so dass man von einem Fenster der B-Zellentwicklung sprechen kann, in das die leukämischen Blasten hineinpassen.

Die zweite Frage zielt darauf ab, wie sich diese Kontrollzellen in ausreichender Quantität und Qualität gewinnen lassen. Die Wahl der sortierten B-Lymphozyten aus Nabelschnurblut als gesunde Kontrollzellen hat verschiedene Vorteile. Die derzeit

gültigen Theorien zur Entstehung von Leukämien wie sie bereits in Kapitel 1.2 diskutiert wurden gehen von einer Entstehung des leukämischen Klons im Embryo aus. Daher mag es von Vorteil sein, die jungen und ursprünglichen Zellen des Nabelschnurbluts zum Vergleich heranzuziehen, wenngleich die cALL-Patienten selbst bereits etwas älter sind. Nabelschnurblut ist weiterhin die Quelle, in der bei leichter Verfügbarkeit die meisten unreifen B-Zellen vorkommen (Hoffkes, 1996; Rubinstein, 1995; Wu, 1999). Allein durch Depletion der Nicht-B-Zellen aus dem Nabelschnurblut konnte der Anteil unreifer B-Lymphozyten in den Proben im Durchschnitt auf über 25% gesteigert werden. Der Anteil der B-Zellen insgesamt lag nach Separation bei über 79%. Das Knochenmark gesunder Kinder als Quelle für B-Vorläuferzellen steht nicht zur Verfügung. Bei den B-Lymphozyten aus Tonsillen bleibt die Frage unbeantwortet, inwieweit diese Zellen wirklich als korrekte Kontrolle gelten können, wenn sie aus einem chronisch entzündlichen Gewebe stammen. Bei der Verwendung von Knochenmark gesunder Erwachsener als Quelle der B-Zellen besteht der Altersunterschied der Probanden im Vergleich zu den pädiatrischen cALL-Patienten als Problem (Hattenhorst, 2000a; Hattenhorst, 2000b). Die gesunden B-Vorläuferzellen wurden indirekt aus dem Nabelschnurblut isoliert, indem die Nicht-B-Zellen markiert und entfernt wurden. Das hat den großen Vorteil, dass die Zielzellen der Untersuchung unverändert bleiben. Eine mögliche Aktivierung oder Veränderung der Zellen durch Antikörperbindung, die eine Veränderung des Genexpressionsprofils zur Folge haben könnte, bleibt somit aus.

Die schließlich gewählte Separationsstrategie stellte einen Kompromiss zwischen der erforderlichen Ausbeute an Zellen und einer ausreichenden Reinheit dar. Im Vergleich zur direkten Markierung der unreifen B-Zellen mit CD19 und CD10 wie sie ursprünglich versucht wurde, lieferte die Verwendung des B-Cell Isolation Kits eine ausreichende Zellzahl mit RNA in einer ausreichenden Menge und Qualität. Dies ging allerdings auf Kosten der Reinheit, das heißt, das verwendete System depletierte alle B-Lymphozyten, also auch die CD19+/CD10- B-Zellen, und nicht nur die Vorläuferzellen. Aber, im verwendeten Nabelschnurblut befinden sich mehr unreife B-Zellen als in irgendeiner anderen verfügbaren Quelle. Wie die Analyse der differentiell exprimierten Gene gezeigt hat, sind die sortierten B-Lymphozyten mit Vorläuferzellen der Reifungsreihe neutrophiler Granulozyten kontaminiert (vgl. 3.6.2.2). Das zur Zellseparation verwendete System ist vom Hersteller zur Selektion von B-Lymphozyten aus adultem peripheren Blut konzipiert. Die Auswahl der dort vorhandenen Antikörper ist nicht in der Lage, unreife Granulozyten zu depletieren. Sie verbleiben unmarkiert in der Fraktion der separierten B-Lymphozyten. Bei der direkten Separation der B-Vorläuferzellen mit Antikörpern gegen CD19 und CD10 bestünde aber durch CD10-exprimierende Neutrophile Granulozyten ebenfalls ein Kontaminationsproblem.

Auf den EOS-H Genechips wurden neben den cALL-Proben und den Kontrollen sortierter B-Lymphozyten aus Nabelschnurrestblut (CB19NP-CB60NP) auch Proben unsortierten Knochenmarks (BM01NP–BM04NP) und Vollblut (PB01NP, PB02NP) erwachsener Probanden sowie aus Tonsillen isolierte B-Lymphozyten (TS01NP, TS02NP) untersucht. Mit nur zwei bis vier Proben sind diese Gruppen allerdings zu klein, als dass man sie bei der Identifikation differentiell exprimierter Gene berücksichtigen kann. Die Ergebnisse des Clusters in Abschnitt 3.2 haben gezeigt, wie sich diese Proben von den übrigen Kontrollen aus Nabelschnurblut unterscheiden. Die Abbildung 3-2 stellt das Ergebnis dar. Dementsprechend wurden diese Kontrollproben nicht als Untersuchungsgruppen zur Identifikation differentiell exprimierter Gene herangezogen. Diese Proben dienen vielmehr der Ergänzung der Datensätze, wenn es gilt, das Expressionsmuster einzelner Gene im Vergleich zu den leukämischen Proben zu betrachten.



**Abbildung 4-1 Korrelation der Expressionswerte zwischen Proben unterschiedlicher Gruppen**

A: Die Regressionsgerade verdeutlicht die Ähnlichkeit der beiden Kontrollproben zueinander. Zudem zeigt sich, wie wenig die Punktwolke um die Regressionsgerade streut. B: Die Unterschiede zwischen einer cALL- und einer Kontrollprobe werden deutlich. Im Vergleich zum Beispiel A streut die Punktwolke hier mehr. Die Regressionsgerade verdeutlicht zudem die geringere Ähnlichkeit der Proben zueinander. Die Signalwerte für jeden Sondensatz der jeweiligen Proben wurden gegeneinander aufgetragen und die sich ergebende Regressionsgerade berechnet. Die Intensität der Signale sind dimensionslose Zahlen.

Um zu einer Aussage zu kommen, worin der Unterschied in der Genexpression zwischen zwei Zelltypen besteht, reicht es nicht aus, lediglich zwei einzelne Proben miteinander zu vergleichen. Es muss eine genügend große Anzahl an Proben untersucht werden, um die Effekte individualspezifischer Unterschiede zu minimieren. Das bedeutet die Proben werden in die zwei Gruppen, „krank“ und „gesund“, eingeteilt beziehungsweise „cALL“ und „gesunde B-Lymphozyten“. Die Einteilung in Gruppen und der Vergleich dieser Gruppen setzen voraus, dass sich die Mitglieder einer Gruppe untereinander jeweils ähnlicher sind als zu den einzelnen Mitgliedern der anderen

Gruppe. Die Abbildung 4-1-A zeigt im Punktdiagramm exemplarisch die beiden Expressionsprofile der Kontrollproben CB48NP und CB60NP wie sie mit HG-U133A ermittelt wurden. Die einzelnen Messwertpaare streuen wenig, es gibt kaum Ausreißer und die entstandene Punktwolke gruppiert sich eng an eine Regressionsgerade, die den linearen Zusammenhang zwischen zwei Proben beschreibt. Die Regressionsgerade nähert sich in diesem Beispiel mit ihrer Funktion:  $y = 1,05x - 4,91$  der Idealform zweier identischer Proben:  $y = x$ . Wird hingegen eine Leukämieprobe (ALL31ARB) auf dieselbe Weise gegen eine Kontrollprobe (CB48NP) aufgetragen, so wird im Punktdiagramm der größere Unterschied zwischen beiden sichtbar (Abbildung 4-1-B). Die Form der Punktwolke wird breiter, es gibt mehr Ausreißer und die Lage der Punktwolke im Koordinatensystem ist drastisch verschoben. Exakter lässt sich das wiederum durch die Regressionsgerade beschreiben, die mit der Funktion:  $y = 0,73x + 197,23$  in stärkeren Maße von der Idealform abweicht als dies im oberen Beispiel der Fall ist.

Die Zugehörigkeit der untersuchten Proben zu bestimmten Zellreihen des hämatopoetischen Systems kann nicht nur durchflusszytometrisch gezeigt werden, wie es bei der Diagnose der leukämischen Zellen regelmäßig geschieht und wie es hier bei der Qualitätskontrolle der sortierten B-Lymphozyten gemacht worden ist. Auch auf mRNA-Ebene lässt sich dieser Nachweis führen, indem die entsprechenden Transkripte, deren Proteine im FACS dargestellt wurden, mit Hilfe der Genechips detektiert werden. In Abschnitt 3.3 konnte die Zugehörigkeit der Leukämieproben und der Kontrollen zur B-Lymphozytenreihe eindeutig bestätigt werden. Dabei sind die B-Zellmarker CD10, CD19, CD79a und TdT, die T-Zellmarker CD4, CD5 und CD8 und die Marker für myeloische Zellen CD13, CD15 und CD33 untersucht worden. Einzig die Expression von CD10 wich von den Erwartungen etwas ab. Die Kontrollproben sortierter B-Lymphozyten zeigten nicht in allen Fällen die Expression von CD10 auf der Ebene der mRNA. Dafür sind zwei Gründe anzuführen. Zum einen war der Gehalt an unreifen B-Zellen (ca. 25% CD19/CD10 doppelt-positive Zellen) in diesen Proben geringer als der Gehalt der ebenfalls CD10 exprimierenden Blasten (ca. 90%) in den Leukämieproben. Zum anderen konnte in anderen Arbeiten bereits gezeigt werden, dass leukämischen Blasten mehr CD10 Moleküle auf ihrer Oberfläche tragen als gesunde Zellen (Rego, 2001). Lymphatische Vorläuferzellen des fetalen Knochenmarks exprimieren maximal bis zu  $5 \times 10^4$  CD10 Moleküle. In den B-Zell Leukämien konnten bis zu  $1,8 \times 10^5$  Moleküle nachgewiesen werden. 33% der B-Zell Leukämien exprimieren signifikant mehr CD10 als Zellen des fetalen oder adulten Knochenmarks (Lavabre-Bertrand, 1994).

#### 4.2 EOS-H und HG-U133A. Zwei unterschiedliche Genechiptypen

Zu Beginn der Studie war es nicht geplant, die Untersuchungen auf zwei verschiedenen Typen von Genechips durchzuführen. Die begrenzte Verfügbarkeit von EOS-H Genechips erforderte es jedoch, im weiteren Verlauf der Arbeit einen Teil der Proben auf HG-U133A zu untersuchen. Dabei konnten immerhin sechs leukämische Proben und sechs Kontrollen der bereits mit EOS-H analysierten Proben auch mit HG-U133A betrachtet werden. Obwohl die gesamte Studie in ihrer Konzeption nicht darauf ausgelegt war, verschiedene Genechiptypen miteinander zu vergleichen, bestand somit zumindest die Möglichkeit, die mit beiden Chiptypen ermittelten Datensätze zu vergleichen und zu bewerten.

Die Untersuchung der verschiedenen Proben mit den beiden Genechiptypen lieferte, wie in Abschnitt 3.5 gezeigt wurde, für die mit EOS-H untersuchten Proben 362 und für die mit HG-U133A untersuchten Proben 1557 differentiell exprimierte Gene. Nach Abzug der Sondensätze, die nur auf einem der beiden Genechiptypen vorhanden sind, lassen sich die Ergebnisse gegenüberstellen (vgl. Tabelle 3-2). Im EOS-H Datensatz sind das 91 Sondensätze und in den mit HG-U133A ermittelten Ergebnissen 288 Sondensätze. Es sind 159 Sondensätze (58,7%) der mit EOS-H detektierten Gene auch mit HG-U133A detektiert, beziehungsweise 195 Sondensätze (15,1%) für den umgekehrten Vergleich. Das heißt diese Gene wurden mit beiden Genechiptypen als differentiell exprimiert erkannt. Die Differenz in der Anzahl der gleichwertig detektierten Sondensätze (159 und 195) liegt in der unterschiedlichen Redundanz der auf den Chiptypen für einzelne Gene vertretenen Sondensätzen begründet. Die verbleibenden Sondensätze der Ergebnisse (112 = 41,3% mit EOS-H und 1094 = 84,9% mit HG-U133A) sind mit beiden Genechiptypen unterschiedlich detektiert worden. Das bedeutet einer Detektion als differentiell exprimiert auf dem EOS-H Genechiptyp steht die Nicht-Detektion auf dem anderen Genechiptyp gegenüber und umgekehrt. Insgesamt wurden jedoch keine widersprüchlichen Ergebnisse in dem Sinne erzielt, dass ein Gen auf dem einen Chiptyp als herauf- und auf dem anderen als herunterreguliert identifiziert wurde.

Die naheliegendste Ursache für die unterschiedlichen Ergebnisse könnte in der ungleichen Zusammensetzung der jeweiligen Untersuchungsgruppen liegen. Aufgrund individualspezifischer Probeneigenschaften, also der natürlichen Varianz biologischer Proben in bezug auf Genom und Transkriptom, ist zu vermuten, dass ungleich zusammengesetzte Untersuchungsgruppen voneinander abweichende Ergebnisse liefern können. Zudem können die Gruppen nicht so groß gewählt werden, dass auftretende Varianzen nicht mehr stark ins Gewicht fallen. Die Verfügbarkeit von Patientenproben ist hier limitierend. Vergleicht man in Stichproben die Sondensätze, die nur mit EOS-H als differentiell exprimiert detektiert wurden, mit den entsprechenden

Sonden auf HG-U133A, sieht man, warum diese dort nicht detektiert wurden. In den meisten Fällen ist das Kriterium des  $t$ -Tests ( $p < 0,01$ ) nicht erfüllt (zum Beispiel NPY, MADH1, ID2 oder SAP30). Der  $t$ -Test betrachtet die Varianz innerhalb der beiden untersuchten Probengruppen. Ist die Varianz innerhalb der Gruppen zu groß, wird das Ergebnis des  $t$ -Tests größer. Damit wird angedeutet, dass ein Teil der abweichenden Ergebnisse in den individuellen Eigenschaften der Proben, also den unterschiedlich zusammengesetzten Gruppen begründet ist.

Für jeweils sechs leukämische Proben und sechs Kontrollproben bestand die Möglichkeit, die Ergebnisse direkt miteinander zu vergleichen. Diese Proben wurden parallel mit beiden Genechiptypen untersucht. Werden diese Proben gesondert ausgewertet, allerdings mit denselben Kriterien wie die übrigen Untersuchungen, stellt man fest, dass die Abweichungen der mit beiden Genechips ermittelten Ergebnisse in einer vergleichbaren Größenordnung liegen (vgl. 3.5.2 und Tabelle 3-3). Nur 20-22% der hier betrachteten Ergebnisse sind auf beiden Genechips identisch. Die übrigen SONDENSÄTZE wurden nur auf einem der beiden Genechiptypen als differentiell exprimiert detektiert. Damit deutet es sich an, dass die Ursachen für die voneinander abweichenden Ergebnisse nicht allein in der unterschiedlichen Zusammensetzung der Untersuchungsgruppen und damit in probenindividuellen Eigenschaften zu suchen sind.

Eine weitere Ursache für unterschiedliche Ergebnisse könnte in den nicht vollkommen identischen Abfragekriterien liegen (vgl. 3.4). Die zentralen Abfragekriterien wie  $t$ -Test und FC sind für beide Experimentreihen identisch. Bei den EOS-H Genechips musste auf eine Hilfskonstruktion zurückgegriffen werden, um festzulegen, welche Gene als exprimiert gelten können und um negative Werte zu korrigieren, die durch die Normalisierung entstanden sind. Die Festlegung, Signale, die kleiner als der Wert 50 sind, auf 50 zu setzen, entspricht den Erfahrungen unserer Kooperationspartner, lässt aber letztlich die individuellen Eigenschaften jedes Genechips unberücksichtigt. Die entsprechenden Angaben „A“ oder „P“ werden bei Affymetrix aus dem Verhältnis der PM- zu MM-Sonden berechnet und berücksichtigen die vollständigen Informationen eines jeden Arrays.

Es ist mit den vorliegenden Daten nicht möglich zu überprüfen, ob die bestehenden Unterschiede in den Ergebnissen auf die verschiedenen Normalisierungstechniken für die beiden Genechiptypen zurückzuführen sind (vgl. 2.11.1). Bei beiden Methoden handelt es sich um sogenannte „globale Normalisierungstechniken“, weil jeweils die vollständigen Informationen eines Genechips in die Normalisierung einfließen. Dennoch ist die Vergleichbarkeit eingeschränkt. Es ist zu vermuten, dass die Normalisierung der Rohdaten einen grundlegenden Einfluss auf die weitere Datenanalyse hat und damit

die Vergleichbarkeit unterschiedlich normalisierter Rohdaten eingeschränkt ist. Dennoch gibt es derzeit kaum Veröffentlichungen, die sich systematisch mit Normalisierungstechniken speziell für Oligonukleotid-Arrays auseinandergesetzt hätten (Li, 2001) oder Alternativen zur Normalisierung diskutieren (Eickhoff, 1999; Hill, 2001). Bei den beiden verwendeten Normalisierungsmethoden handelt es sich um Standardtechniken, die bereits auf den jeweiligen Chiptyp angewendet wurden (Alon, 1999; Glynne, 2000a).

Die beiden verwendeten Genechips unterscheiden sich in ihrer Konstruktion, in der Auswahl der zu untersuchenden Transkripte und in den entsprechenden Sonden zur Detektion. Damit sind drei weitere Ursachen benannt, die Einfluss auf die Ergebnisse nehmen können. Mit HG-U133A werden deutlich mehr Transkripte abgefragt als mit dem anderen Genechip. Im Gegensatz zum HG-U133A sind auf EOS-H keine MM-Sonden vorhanden. Das beinhaltet eine unterschiedliche Auswahl von Sonden für ein Transkript auf beiden Genechips. Damit sind direkt unterschiedliche Hybridisierungseigenschaften der Sonden verbunden, die wahrscheinlich einen Einfluss auf die Detektion haben. Weiterhin sind die Ursprungssequenzen der Transkripte, aus denen die Sonden entwickelt wurden, nicht immer identisch, auch wenn jeweils identische Gene abgefragt werden sollen. Das hat mitunter zur Folge, dass die SONDENSÄTZE unterschiedliche Bereiche eines Transkriptes abfragen (vgl. Tabelle 3-1). Vergleicht man in Stichproben die SONDENSÄTZE, die nur mit HG-U133A als differentiell exprimiert detektiert wurden, mit den entsprechenden Sonden auf EOS-H, also in der anderen Richtung als oben bereits geschildert, dann deutet sich an, dass die Eigenschaften der Genechips für unterschiedliche Ergebnisse mitverantwortlich sind. Ein Teil der mit HG-U133A sehr wohl als differentiell detektierten Gene wird von den SONDENSÄTZEN des EOS-H Genechips überhaupt nicht detektiert, also weder in den leukämischen Proben noch in den Kontrollen (zum Beispiel ESP1, H2BFL, E2F5 oder KYNU). Hier handelt es sich ganz deutlich um das Resultat unterschiedlicher Hybridisierungseigenschaften der SONDENSÄTZE. Ein anderer Teil nur von HG-U133A detektierter Gene erfüllt auf EOS-H das Kriterium des mindestens zweifachen Unterschieds im FC nicht (zum Beispiel TCN1, ITPKB, TMP2 oder DTR). Es zeigt sich, dass die FC-Werte bei den mit EOS-H ermittelten Daten durchweg geringer waren und sich in einem engeren Intervall bewegten als bei HG-U133A. Dieser Unterschied muß in den verschiedenen Hybridisierungseigenschaften der beiden Genechiptypen begründet sein. Vergleicht man allein die doppelt untersuchten Proben, werden die Differenzen in den FC-Bereichen deutlich. In diesem Zusammenhang ist das Phänomen zu betrachten, dass mit HG-U133A wesentlich mehr differentiell exprimierte Gene detektiert wurden als mit EOS-H. Unterschiedliche Varianzen in den Probengruppen

könnten zu diesem Befund beigetragen haben, aber es ist naheliegender, dafür unterschiedliche Hybridisierungseigenschaften verantwortlich zu machen. Dafür sprechen letztlich die weiten Bereiche, in denen sich FC-Werte bewegen. Die unterschiedlichen Chipeigenschaften tragen somit ebenfalls zu den nicht übereinstimmenden Ergebnissen bei.

Zu diskutieren sind in diesem Zusammenhang mögliche Fehler auf den Genechips. Für cDNA Arrays ist bereits diskutiert worden, dass 1% bis 5% der verwendeten Sonden nicht die Sequenz enthalten, die sie enthalten sollten. Dies ist als Resultat der Herstellung der Sonden aus bakteriell amplifizierten cDNA-Banken zu sehen (Knight, 2001). Eine mögliche Fehlerquelle bei oligonukleotidbasierten Genechips liegt in den Datenbanken, aus denen die Sequenzen entnommen werden. Hier können bei der Veröffentlichung von Sequenzierungsergebnissen Fehler entstehen wie die Eingabe von Sequenzen in der falschen Orientierung (Karlin, 2001; Knight, 2001). Letztlich ist mit Fehlern in der Annotation der Gene in den Datenbanken zu rechnen. Der Sondensatz HG-U133A 202864\_s gibt als Zielsequenz das Gen für den Transkriptions-Koaktivator SP110 an, während die zugehörigen Sequenzinformationen auf ein DNAJ Protein mit Funktionen in der Proteinfaltung verweisen.

Die Ergebnisse, die mit den beiden Genechiptypen ermittelt wurden, müssen nach den derzeit verfügbaren Analysemethoden als gleichberechtigte Ergebnisse nebeneinander stehen. Erstens sind sie nicht prinzipiell widersprüchlich, sondern können eher als einander ergänzend betrachtet werden. Zweitens sind sie das Resultat der unterschiedlich gestalteten Genechips und damit verbunden der unterschiedlichen Normalisierungstechniken. Aus den vorliegenden Daten ist jedoch nicht zu sagen, welches Modell nun besser oder korrekter sei. Die Frage, ob die unterschiedlichen Ergebnisse das Resultat der unterschiedlichen Zusammensetzung der Untersuchungsgruppen sind, deutet sich an, lässt sich aber aus den Datensätzen heraus nicht abschließend beantworten. Allein aus der Analyse der Daten lassen sich die mit Verwendung von zwei Chiptypen entstehenden Fragen nicht beantworten. Hierfür sind weiterführende Experimente notwendig, um die Expression der Gene, die unterschiedlich detektiert wurden, mit einer zweiten Methode zu überprüfen. Weiterhin ist zu bemerken, dass die Arbeit nicht darauf ausgelegt war, systematisch die verschiedenen Genechiptypen miteinander zu vergleichen. Die Versuchsanordnung ist dementsprechend nicht dazu geeignet, diese Frage zu beantworten.

Da die Inzidenz falsch-positiver Ergebnisse genügend klein ist, ist es nicht notwendig, jede Änderung in der Expression unabhängig und durch eine weitere Methode (zum Beispiel RT-PCR, Northern-Blot) zu validieren, erst recht nicht, wenn die Schlussfolgerungen aus den erhobenen Daten eher aus Gruppen oder Clustern von

Genen gezogen werden und nicht aus einzelnen Genen. Das entbindet aber nicht von der Verpflichtung zur detaillierten Validierung, wenn ein einzelnes Genprodukt als therapeutische Zielstruktur identifiziert wurde (Lockhart, 2000). Um abschätzen zu können, welchen experimentell bedingten Schwankungen die Ergebnisse der hier durchgeführten Genexpressionsanalyse unterworfen sind, wäre es sinnvoll, einen Teil der Ergebnisse im Vergleich einzelner Proben mit anderen Methoden zu verifizieren. Dies scheitert aber an der eingeschränkten Verfügbarkeit des von den Patienten stammenden Untersuchungsmaterials. Zur Untersuchung der differentiellen Genexpression in leukämischen Blasten kann man die Ergebnisse der Genechiptypen als Ausgangsbefunde für weiterführende Untersuchungen gleichberechtigt und einander ergänzend nebeneinander stehen lassen, zumal es keine direkt widersprüchlichen Ergebnisse gibt.

High-density Oligonucleotide Arrays sind hochspezifisch und extrem sensitiv. Die Falsch-positiv Rate der Detektion liegt bei <2%. Verwandte Gene mit einer hohen Homologie wie die Histon Gene HTA1 und HTA2 der Hefe *Saccharomyces cerevisiae* mit einer Sequenzidentität auf Ebene der DNA von 93% können zu 80% unterschieden werden. Unterschiede in der Genexpression eines Gens in zwei unterschiedlichen Proben, die nur den Faktor 2 betragen, können verlässlich gemessen werden. Die Sensitivität der Genechips bewegt sich unter idealen Voraussetzungen in einem Rahmen von 1:100.000 bis 1:300.000. Das bedeutet, dass RNA-Spezies, die mit nur ein bis drei Kopien pro Zelle vorkommen, detektiert werden können (Lipshutz, 1999; Lockhart, 1996). Unter normalen experimentellen Bedingungen ist eine verlässliche Detektion in einem Bereich von 1:300.000 bis 1:50.000 (0,5-1,5pM) möglich (Hill, 2000). In einer aktuellen Arbeit konnte E. Chudin diese Sachverhalte weitestgehend bestätigen (Chudin, 2002). Laut dieser Arbeit eignet sich das System dazu, Konzentrationsunterschiede der Transkripte mit dem Faktor 2 zu unterscheiden. Ein linearer Zusammenhang zwischen Transkriptmenge und Signal ist übereinstimmend im Bereich von 1pM bis 10pM Transkriptkonzentration zu beobachten. Jenseits des 10pM-Levels hört diese Linearität jedoch allmählich auf. Die Größe der beim Vergleich von Proben miteinander ermittelten FC-Werte ist mitunter Schwankungen unterworfen und steht bei geringer Transkriptkonzentration nicht immer im linearen Zusammenhang zu den RNA-Ausgangsmengen.

Die hohe Reproduzierbarkeit der Methode konnte experimentell nachgewiesen werden und ist in der Literatur beschrieben. In einer ersten Untersuchung wurden aus einer RNA-Probe, bestehend aus einem Pool von RNA aus sechs Gewebeproben von Patienten mit Colitis ulcerosa, unabhängig voneinander zwei Genechip-Experimente durchgeführt. Die ermittelten Fluoreszenzintensitäten aller SONDENSÄTZE beider

Genechips wurden einer linearen Regressionsanalyse unterzogen. Die beiden Experimente zeigten sich als hoch reproduzierbar: 99,16% der Gene zeigten vergleichbare Werte. Nur 50 der insgesamt 7.306 untersuchten Gene zeigten einen mehr als zweifachen Unterschied in der Fluoreszenzintensität (entspricht 0,68%); bei nur zwölf Genen war ein mehr als dreifacher Unterschied festzustellen (entspricht 0,16%). Der Korrelationskoeffizient zwischen beiden Experimenten betrug 0,97, was bedeutet, dass nur 3% der beobachteten Variationen experimentell bedingt sind. (Lawrance, 2001). In einer anderen Untersuchung wurde eine Kultur von *Saccharomyces cerevisiae* halbiert und unabhängig voneinander einer Genexpressionsanalyse unterzogen. Von den dort untersuchten 6.200 Genen zeigten 74 einen zweifachen Unterschied (entspricht 1,19%) und sechs einen dreifachen Unterschied (entspricht 0,09%) (Wodicka, 1997).

#### 4.3 Genexpressionsprofil der cALL im Kindesalter

Beim Vergleich der Genexpressionsprofile von leukämischen Blasten mit gesunden B-Lymphozyten aus Nabelschnurblut konnten mit den beiden Genechiptypen differentiell exprimierte Gene identifiziert werden. Mit EOS-H wurden 95 unterschiedliche Transkripte (99 Sondensätze) identifiziert, die in den leukämischen Proben heraufreguliert sind, und 256 herunterregulierte Transkripte (263 Sondensätze). Mit dem HG-U133A Genechip wurden 575 Transkripte identifiziert, die in den cALL-Proben stärker exprimiert sind als in den Kontrollen. 1.002 Transkripte zeigten eine schwächere Expression in den leukämischen Proben. Die mit HG-U133A detektierten Transkripte lassen sich für weitere Analysen (Klassifikation unbekannter Proben, Identifikation diagnostisch oder therapeutisch geeigneter Zielstrukturen) einschränken. Als zusätzliches Kriterium sollten die Gene jeweils exklusiv in der Gruppe der Leukämien oder der Kontrollgruppe exprimiert sein. Damit verbleiben 132 Sondensätze, die exklusiv in den leukämischen Zellen exprimierte Gene detektieren, und 138 Sondensätze für Gene, die in diesen Zellen im Vergleich zu den Kontrollen überhaupt nicht exprimiert sind. Nach der Diskussion im vorangegangenen Abschnitt können die beiden mit EOS-H und HG-U133A ermittelten Ergebnisdatensätze als gleichbedeutend und ergänzend nebeneinander betrachtet werden.

Die beiden Datensätze differentiell exprimierter Gene repräsentieren die spezifischen Expressionsmuster der cALL im Vergleich zu gesunden B-Lymphozyten. Das Muster dieser herauf- und herunterregulierter Gene ist spezifisch für die cALL und lässt eine Zuordnung von Proben zur Leukämie- oder zur Kontrollgruppe zu. Diese Gene lassen sich damit als cALL-Markergene benennen. In ihnen stecken genug Informationen, um die Proben unterschiedlicher Herkunft klassifizieren zu können. Mit

Hilfe der 362 Gene (EOS-H) beziehungsweise der 270 Gene (HG-U133A) lassen sich alle Proben korrekt der Gruppe der Leukämien oder der gesunden Kontrollen zuordnen. Will man die Proben hingegen mit den vollständigen Daten der Genechips clustern, so lassen sich die Gruppen nicht mehr eindeutig voneinander trennen. Die darin enthaltenen Informationen der weit über 20.000 nicht differentiell exprimierten oder nicht detektierten Gene überlagern vermutlich die aussagekräftigen Informationen der cALL-Markergene. Bei der nachträglich mit HG-U133A analysierten Probe ALL58GAT handelt es sich um die einzige Probe eines Patienten mit einem Lymphom. Die Genexpressionsdaten dieser Probe sind nicht in die Untersuchungen zur Identifikation der differentiellen Genexpression zwischen den cALL Blasten und den gesunden B-Lymphozyten eingeflossen. Damit eignet sich diese Probe als eine Art Kontrolle für die Leukämiespezifität des ermittelten Genexpressionsmusters der cALL. Werden die auf HG-U133A ermittelten 270 differentiell exprimierten Gene zur Klassifikation der Leukämieproben und der Kontrollen unter Hinzunahme der Lymphomprobe verwendet, so werden nicht nur Leukämien und Kontrollen eindeutig auf zwei getrennte Gruppen verteilt. Die neue Probe ALL58GAT wird dabei zusätzlich in die Nähe zu den übrigen Leukämien, aber abseits der Kontrollen sortiert. Dies liefert einen Hinweis darauf, dass die ausgewählten Gene dazu geeignet sind, vorher unbekannte Proben richtig zu klassifizieren und zwischen Leukämie und gesunder Probe zu unterscheiden. Weiterhin zeigt die „korrekte“ Klassifikation die große verwandtschaftliche Nähe dieses Lymphomtyps zur cALL auf.

Die differentiell exprimierten Gene scheinen dazu geeignet, Aussagen über die molekularen Grundlagen der pathologischen Veränderungen der cALL zu machen. Gene, die in einem funktionellen oder regulatorischen Zusammenhang zueinander stehen, haben vergleichbare Expressionsmuster und -stärken. Diese Annahme ist naheliegend und begründet. Damit sollte es möglich sein, aus gleichen Expressionsmustern verschiedener Gene funktionelle oder regulatorische Zusammenhänge abzuleiten, die bislang nicht bekannt waren. Mit Hilfe von Clusteranalysen konnte gezeigt werden, welche Gene im Verlauf der Reifungsreihe von der murinen prä-B-Zelle bis zur reifen B-Zelle ähnliche Expressionsmuster aufweisen (Hoffmann, 2002). Innerhalb dieser Gencluster konnten auch funktionelle Zusammenhänge aufgezeigt werden. In einer anderen Arbeit (Perou, 1999) wurde dargestellt, wie Clusteralgorithmen dazu genutzt werden können, Unterschiede in Genexpressionsmustern zwischen Zellen aus Brusttumoren und entsprechenden Kontrollen aufzudecken. Die so identifizierten Gencluster sollen funktionelle Informationen über die Interaktion von Tumorzellen mit dem sie umgebenden Gewebe enthalten.

Es scheint naheliegend, auf diesem Weg zu untersuchen, in welchem funktionellen Zusammenhang die differentiell exprimierten Gene stehen, die den Unterschied zwischen den lymphoiden Blasten der cALL und den gesunden B-Lymphozyten ausmachen. Das durchgeführte Clusterexperiment hat indes gezeigt, dass sich die beiden Datensätze so nicht sinnvoll clustern lassen. Die entstandenen Gencluster geben keine Informationen über funktionelle Zusammenhänge der Gene innerhalb einzelner Cluster. Wahrscheinlich ist die Anzahl der betrachteten Parameter - die beiden zellulären Zustände „krank“ und „gesund“ - zu gering. In die entstehenden Cluster werden vermutlich zu viele Gene sortiert, die nur zufällig ein entsprechendes Profil aufweisen, aber funktionell nicht miteinander in Verbindung stehen. Im Ansatz von R. Hoffmann wurde die Genexpression immerhin über fünf verschiedene zelluläre Stadien der B-Zellreifung hinweg untersucht. Damit können Muster gemeinschaftlich sich ändernder Expressionsprofile verschiedener Gene besser erkannt werden. Letztlich scheint das Clustern der Gene nur sinnvoll möglich zu sein, wenn man die Entwicklung und Veränderung der Genexpression über mehrere Schritte oder Entwicklungsstufen beobachten kann.

Da sich die Gene nicht sinnvoll clustern ließen, wurden sie zur detaillierten Auswertung in funktionelle Gruppen eingeteilt: 1. Entwicklung von B-Lymphozyten, 2. Immunsystem und Inflammation, 3. Replikation von DNA, Transkription, Proteinsynthese, 4. Signaltransduktion und Regulation von Genexpression, 5. Regulation von Differenzierung und Entwicklung, 6. Regulation von Zellzyklus und Zellproliferation und 7. allgemeiner Metabolismus. Das vorrangige Ziel dabei ist es, mögliche Zusammenhänge mit der Pathologie der cALL darzustellen und dabei auf zelluläre Eigenschaften der Leukämiezellen einzugehen. Die Schwerpunkte liegen entsprechend auf den Bereichen Wachstum, Entwicklung und Differenzierung von Zellen.

Darüberhinaus lassen sich auf diesem Wege die Gene in funktionelle und regulative Zusammenhänge stellen. Die Rolle der differentiell exprimierten Gene für die Pathologie der Krankheit kann so besser verstanden werden. Wenn für jedes einzelne Gen dessen Bedeutung für die Pathologie der cALL aus den Expressionsdaten heraus erklärt werden sollte, kommt man nicht umhin, die Expression mit einer alternativen Methode zu überprüfen. Können diese Gene allerdings sinnvoll in funktionelle Zusammenhänge gestellt werden, liefert bereits das Expressionsmuster der verschiedenen Gene einen Hinweis auf die Korrektheit des Erklärungsansatzes.

Es konnten Gene identifiziert werden, von denen bereits bekannt ist, dass sie in der cALL differentiell exprimiert sind. Dies waren sowohl heraufregulierte Gene (zum Beispiel CD10, TdT oder die Kinase ADRBK1) als auch herunterregulierte Gene (zum

Beispiel der EBV-Rezeptor CR2). Es wurden eine Reihe neuer Gene gefunden, deren differentielle Expression im Zusammenhang mit der cALL erstmalig beschrieben wird.

Die Patienten mit einer cALL zeichnen sich primär durch einen übereinstimmenden pathologischen Phänotyp aus. Hierzu gehören klinische Parameter (zum Beispiel die generell erhöhte Leukozytenzahl im Knochenmark) ebenso wie der Immunphänotyp der leukämischen Blasten (zum Beispiel der Nachweis von CD10 und TdT). Dennoch gibt es individuelle Unterschiede zwischen den Patienten, die auf die Existenz distinkter Untergruppen der cALL hinweisen. Diese Unterschiede machen zum Beispiel das Ansprechen auf die Therapie, die Beteiligung anderer Organe als dem Knochenmark oder der Zeitraum bis zum Auftreten eines Rezidivs aus. Die Einteilung der Patienten in verschiedene Risikogruppen mit therapeutischen Konsequenzen versucht dem Rechnung zu tragen (Schrappe, 1998; Schrappe, 2000b). Insbesondere die verschiedenen chromosomalen Veränderungen wie Translokationen oder Polyploidien deuten an, dass es mehr als eine Form der cALL geben könnte. Leukämien mit unterschiedlichen Translokationen wie TEL/AML1, MLL-AF4 oder BCR-ABL müssen vielleicht als eigenständige Krankheiten betrachtet werden. Sollten die Untergruppen der cALL genauer definiert werden können, könnten therapeutische Maßnahmen auf die Erfordernisse der einzelnen Untergruppen besser zugeschnitten werden. Die erstellten Genexpressionsprofile lassen prinzipiell eine Einteilung der Proben in Gruppen zu, was bereits mehrfach gezeigt worden ist. Die Klassifikation von Krebserkrankungen wird dabei durch die Expressionsmuster von Gruppen an Genen ermöglicht, die sich innerhalb der Klassen einheitlich verhalten (Golub, 1999). In einem solchen Ansatz wurden 78 Proben von Brusttumoren untersucht und diese in Subklassen mit klinischen Implikationen eingeteilt (Sorlie, 2001). Über einen vergleichbaren Ansatz konnten Nierenzellkarzinome klassifiziert und darüber hinaus prognostische Aussagen zu den gefundenen Untergruppen gemacht werden (Takahashi, 2001).

Die Identifikation von Untergruppen in den hier untersuchten cALL-Proben unter Verwendung der verschiedenen unüberwachten Clusteransätze führte mit beiden Datensätzen nicht zum Ziel (vgl. 3.7). Es ließen sich keine eindeutigen Subklassen aus den Genexpressionsprofilen ableiten. Für diesen Ansatz wurden nur die leukämischen Proben geclustert. In einem ersten Versuch wurden dafür alle auf den Genechips befindlichen SONDENSÄTZE verwendet, was zu keinem Ergebnis führte. In einem zweiten Versuch wurde die Anzahl der SONDENSÄTZE zum Clustern eingeschränkt. Es wurden alle Merkmale beziehungsweise Expressionswerte für Gene ausgesondert, die durchgehend in keiner der Leukämien exprimiert waren. Die Datensätze reduzierten sich damit für EOS-H von 22.308 auf 11.188 Gene und für HG-U133A von 34.503 auf

19.144 Gene. Auch mit den reduzierten Datensätzen ließen sich im unüberwachten Clustern keine Untergruppen identifizieren.

In zwei Arbeiten wurden die verschiedenen Clusteralgorithmen miteinander verglichen und bewertet (Gibbons, 2002; Yeung, 2001). Die Autoren konnten zeigen, dass der vom Programm „Cluster and TreeView“ (vgl. 2.11.2.2) verwendete Pearson Korrelationskoeffizient verlässliche Ergebnisse liefert (Gibbons, 2002). Zum Zusammenfügen der Proben, also zum eigentlichen hierarchischen Clustern bietet das Programm drei Optionen an. Es konnte gezeigt werden, dass mit allen Methoden vernünftige Ergebnisse erzielt werden, wobei das Single Linkage Clustering etwas schlechter abschnitt (Yeung, 2001). Im vorliegenden Datensatz wurde mit den verschiedenen Algorithmen getestet, ob sich die Proben in Untergruppen der cALL clustern lassen. Die Unterschiede in den Algorithmen bestehen in der Berechnung der Distanzen zwischen Merkmalen und der Erstellung der Distanzmatrix. Dementsprechend muß jeder Clusteralgorithmus leicht unterschiedliche Ergebnisse liefern (Quackenbush, 2001). Allerdings ließ sich mit den Leukämiedaten kein einziges Patientencluster mit einem der anderen Algorithmen reproduzieren. Dies deutet auf eine Homogenität und Ähnlichkeit der Proben hin, die eine Unterteilung in Subklassen verhindert. Es lassen sich aus den Genexpressionsdaten keine Untergruppen der cALL im Kindesalter definieren.

Die erhobenen klinischen Daten der Patienten wie sie in Tabelle 6-1 zusammengestellt sind lassen anhand einzelner Parameter wie zum Beispiel der initialen Zahl leukämischer Blasten bei der Diagnose, dem Ansprechen auf bestimmte Therapien, der Einteilung in Risikogruppen oder der Organbeteiligung auf einzelne Untergruppen schließen. Die überzeugendsten Merkmale zur Definition von Subklassen sind genetisch manifestierte Marker, die einen Einfluss auf die Genexpression ausüben, worunter bei den Leukämien besonders die chromosomalen Translokationen zu zählen sind. Translokationen am MLL-Locus treten relativ häufig in ALL-Patienten auf und in einem unüberwachten Clusteransatz konnte gezeigt werden, dass die Patienten mit einer solchen Translokation in eine eigene Subklasse der cALL einzuteilen sind (Armstrong, 2001). Unter den mit HG-U133A untersuchten Proben befanden sich fünf Patientenproben mit einer TEL/AML1-Translokation. Obwohl mit diesen Proben Gene identifiziert werden konnten, die charakteristisch für diese cALL-Form sind, und obwohl gezeigt werden konnte, dass diese Gene ausreichen, diese besonderen Leukämien von den übrigen Leukämien und von den gesunden Kontrollzellen zu unterscheiden, bildeten diese Leukämien mit beiden unüberwachten Clustermethoden - hierarchisches Clustern und *k*-Means Clustern - keine eigene Gruppe. Mit den vorliegenden Daten hat das unüberwachte Clustern der Patientenproben in keinem der Ansätze zu einem

Ergebnis geführt. Die Gründe hierfür könnten in der relativ kleinen Probengruppe liegen. Aber es deutet sich auch die große Ähnlichkeit der cALL-Proben untereinander an. Die nur sehr geringe Anzahl differentiell exprimierter Gene, die zwischen den Leukämien mit und ohne TEL/AML1-Translokation unterscheiden, spiegelt diese Ähnlichkeit wider (vgl. 3.8.4). Wenn die Ähnlichkeit zwischen zwei Gruppen groß ist, die Unterschiede in der Genexpression folglich gering sind, sind nicht genug Merkmale vorhanden, die zum Clustern ausreichen.

Die folgenden Untersuchungen wurden allein an den mit HG-U133A erhobenen Daten durchgeführt, da hier die meisten leukämischen Proben analysiert wurden. Die klinischen Daten der Patienten lassen eine Unterteilung der Patienten in Gruppen zu. Da die oben erwähnten unüberwachten Methoden nicht zu dem Ziel führten, Subklassen zu definieren, sollte mit diesem Ansatz überprüft werden, ob sich solche auf anderem Wege überhaupt finden lassen, sprich ob sich die in Gruppen eingeteilten Proben in ihren Genexpressionsprofilen unterscheiden. Als genetisch manifestierte Untergruppe werden die Leukämien mit einer TEL/AML1-Translokation in Abschnitt 4.4 gesondert beschrieben. Die anderen betrachteten Gruppen wurden wie folgt gebildet: Alter bei Diagnose („jünger als 5 Jahre“ gegen „älter als 10 Jahre“), Geschlecht der Patienten, Risikogruppen („Risikogruppe 1“ gegen „Risikogruppe 3“), Anzahl der Leukozyten im peripheren Blut zum Zeitpunkt der Diagnose („weniger als 10.000 Zellen/ $\mu$ l“ gegen „mehr als 50.000 Zellen/ $\mu$ l“), Anzahl der Leukozyten am Tag 8 („weniger als 1.900 Zellen/ $\mu$ l“ gegen „mehr als 6.000 Zellen/ $\mu$ l“), prozentualer Anteil der Blasten im Knochenmark am Tag 15 („weniger als 30%“ gegen „mehr als 30%“). Die Gruppen wurden im einzelnen so gewählt, dass sie sich möglichst eindeutig voneinander unterscheiden und gleichzeitig eine klinische Relevanz besitzen. Es ist nachgewiesen, dass Patienten mit einer ursprünglich geringen Zahl an Leukozyten im Blut (WBC) eine bessere Prognose haben (Friedmann, 2000; Schrappe, 2000a; Schrappe, 2000b). Auch das Geschlecht der Patienten hat einen Einfluss auf die Prognose, die für Mädchen besser ist als für erkrankte Jungen (Pui, 1999).

Mit Ausnahme der „Altersgruppe“ und der „WBC-Gruppe“ konnten für keine der betrachteten Gruppen spezifische Genexpressionsprofile ermittelt werden. Die beiden Patientengruppen mit unterschiedlichem Alter bei Diagnosestellung unterscheiden sich in der Expression von acht Genen. Mit Hilfe der Klassifikation konnte gezeigt werden, dass diese acht Transkripte ausreichen, im vorliegenden Datensatz alle älteren Patienten (älter als zehn Jahre) von den Jüngeren (jünger als zehn Jahre) zu unterscheiden. Welche Bedeutung die identifizierten Gene im einzelnen für pathologische Mechanismen und Prognose haben, bleibt offen. Fest steht jedoch, dass pädiatrische cALL-Patienten eine bessere Prognose haben, je jünger sie zum Zeitpunkt

der Diagnosestellung sind. Eine schlechtere Prognose haben Kinder, die jünger als ein Jahr oder älter als neun Jahre sind (Donadieu, 1998; Donadieu, 2000; Friedmann, 2000; Silverman, 2001).

Für die Patienten mit einem unterschiedlichen initialen WBC (WBC < 10.000 Zellen/ $\mu$ l oder WBC > 50.000 Zellen/ $\mu$ l) ließen sich neun unterschiedlich exprimierte Gene identifizieren. Die Prognose für Patienten mit mehr als 50.000 Leukozyten/ $\mu$ l im peripheren Blut ist schlechter als für Patienten mit geringeren Zellzahlen (Friedmann, 2000; Pui, 2000). Im Gegensatz zum „Alterscluster“ reichen diese Gene allerdings nicht aus, die Proben zu klassifizieren. Dies kann an den geringen Expressionsunterschieden zwischen den Gruppen liegen, die sich in kleinen FC-Werten äußern. Die Bedeutung der Expression einzelner Gene für die Prognose oder als Erklärung für die unterschiedliche Leukozytenzahl im Blut, bleibt zu klären.

Die Genexpressionsprofile der cALL-Patienten lassen in unüberwachten Clusteransätzen keine Unterteilung der Patienten in Untergruppen zu. Bei der Definition von Patientengruppen, die einige klinische Unterschiede zeigen, lassen sich aus dem vorliegenden Datensatz nur für die „Altersgruppe“ und die „WBC-Gruppe“ differentiell exprimierte Gene finden. Daraus läßt sich ableiten, dass es sich bei den hier definierten Gruppen vermutlich nicht um echte Subklassen der cALL handelt. Hierfür können zwei Gründe angeführt werden. Möglicherweise ist die Anzahl untersuchter Proben zu gering. Die definierten Untergruppen sind aus den klinischen Daten mit statistischen Methoden nicht auf ihre Signifikanz hin zu überprüfen. Dazu muss die Anzahl der Proben weiter erhöht werden. Es ist denkbar, dass die cALL des Kindesalters doch eine relativ homogene Erkrankung ist, in der das charakteristische Genexpressionsprofil von den meisten Patienten geteilt wird, unabhängig von individuellen klinischen Unterschieden. Einen positiven Hinweis darauf gibt die nur geringe Anzahl differentiell exprimierter Gene, die zwischen den Leukämien mit und ohne TEL/AML1-Translokation oder zwischen den älteren und jüngeren Patienten unterscheiden.

#### 4.4 cALL mit TEL/AML1-Translokation

Die chromosomale Translokation t(12;21)(p13;q22) fusioniert die Helix-Loop-Helix-Domäne des TEL (Translocation ETS Leukemia, ETV6) Gens, eines Transkriptionsfaktors der ETS-Familie, an die DNA-bindende und transaktivierende Domäne von AML1 (Acute Myeloid Leukemia 1), einem weiteren Transkriptionsfaktor (Golub, 1995). In pädiatrischen ALL-Patienten ist die TEL/AML1-Translokation (ETV6/CBFA2) am häufigsten anzutreffen (Romana, 1995). Diese Translokation besitzt besondere klinische Relevanz und bedeutet für die Patienten eine eher günstige

Prognose (Borkhardt, 1997). Patienten mit dieser Translokation sind typischerweise jünger als zehn Jahre und haben im Vergleich mit den anderen Patienten weniger Leukozyten in Blut und Knochenmark. Obwohl die Translokation mit einer guten Prognose in Verbindung gebracht wird, tragen ca. 20% der Patienten mit einem Rezidiv die TEL/AML1-Translokation. Dies ist vergleichbar mit der Inzidenz zum Diagnosezeitpunkt (Harbott, 1997). Zwar dauert es bei diesen Patienten länger bis zum Auftreten des Rezidivs, in der Regel mehr als zwei Jahre, aber weiterhin sind keine bedeutenden klinischen Unterschiede zwischen Patienten mit und ohne TEL/AML1-Translokation festzustellen (Harbott, 1997).

Beide Gene des Fusionstranskripts treten auch mit jeweils verschiedenen Fusionspartnern in unterschiedlichen Translokationen bei weiteren lymphatischen und myeloischen Leukämien in Erscheinung. Die TEL/AML1-Translokation ist jedoch auf B-Lymphozyten beschränkt. TEL ist ein Mitglied der ETS Familie von Transkriptionsfaktoren und funktioniert als sequenzspezifischer, DNA-bindender Transkriptionsfaktor (Rubnitz, 1999). Das reguläre TEL Transkript ist in verschiedenen hämatopoetischen und nicht-hämatopoetischen Zellen exprimiert. Die Funktion von TEL ist essentiell für die Hämatopoese im Knochenmark (Wang, 1998). Das AML1 Gen kodiert einen Transkriptionsfaktor mit Homologie zum *D. melanogaster* Segmentgen *runt*. Wahrscheinlich spielt AML1 eine essentielle Rolle in der Hämatopoese dadurch, dass es die Expression der Zielgene spezifisch für die verschiedenen Zellreihen des blutbildenden Systems steuert (Okuda, 1996). Im chimären Fusionstranskript sind die Helix-Loop-Helix-Domäne von TEL und die DNA-bindende *runt*-Domäne mit der transaktivierenden Domäne aus AML1 funktionstüchtig. Die Expression des Fusionstranskripts wird über den TEL Promotor gesteuert (Golub, 1995). In einer anderen Studie konnte festgestellt werden, dass es für jeden untersuchten Patienten individuelle Bruchpunkte in den Chromosomen gibt (Wiemels, 1999b), wenngleich AML1 fast vollständig und funktionstüchtig erhalten zu bleiben scheint. Es ist anzunehmen, dass die Expression der fusionierten Gene zu einer Änderung der Genexpressionsprofile führt.

Die cALL mit TEL/AML1-Translokation ist eine Leukämieform mit Blasten, die im Stadium der CD10+, CD34+, CD19+ prä-B-Zelle arretiert sind (Lemmers, 1999). Dies ist ein Stadium der B-Zellreifung, in dem die Zellen bei Abwesenheit von Signalen von der schweren Kette des prä-B-Zellrezeptors den programmierten Zelltod durchlaufen (Armstrong, 2002). Zur Expression von CD34 ist zu bemerken, dass dieser Marker hämatopoetischer Stammzellen in den mit HG-U133A untersuchten cALL-Proben heraufreguliert zu sein scheint (auf den EOS-H Genechips befinden sich zwei SONDENSÄTZE für CD34, die widersprüchliche Ergebnisse liefern). Der

Expressionsunterschied zwischen cALL und CB-Kontrolle auf HG-U133A ist nicht signifikant. CD34 ist in 21 von 25 cALL-Proben und in fünf von sechs CB-Proben exprimiert. Die Expression von CD34 scheint nach diesen Daten kein ausdrückliches Merkmal der TEL/AML1-positiven cALL zu sein.

In einer nicht veröffentlichten Arbeit von S. Armstrong wurden die Genexpressionsprofile von cALL-Patienten mit und ohne TEL/AML1-Translokation miteinander verglichen. Dabei deutete sich an, dass nur ca. 100 Gene einen signifikanten Unterschied zwischen beiden Probengruppen zeigten (Anmerkung in (Armstrong, 2002)). Bei der Untersuchung von Leukämien mit und ohne Translokationen am MLL-Locus wurden hingegen mehr als 1.200 differentiell exprimierte Gene identifiziert (Armstrong, 2001). Im vorliegenden Datensatz ist es ebenfalls auffällig, dass bei der Betrachtung der Leukämien mit TEL/AML1-Translokation weniger differentiell exprimierte Gene gefunden werden als bei den übrigen hier durchgeführten Analysen. Bei dem Vergleich zwischen den Leukämien mit und ohne diese Translokation konnten nur zwölf SONDENSÄTZE identifiziert werden, die differentiell exprimierte Gene detektieren. AML1 wird als schwacher Transkriptionsfaktor betrachtet, und in Tiermodellen konnte gezeigt werden, dass AML1-assoziierte Translokationen allein nicht ausreichen, eine Leukämie zu induzieren (Perry, 2002). Zusammengefasst ist es vor diesem Hintergrund erklärlich, warum sich die Leukämien beim unüberwachten Clustern nicht in deutliche Gruppen unterteilen lassen und warum beim Klassifizieren der Proben mit differentiell exprimierten Genen die TEL/AML1-positiven Proben nicht ganz eindeutig von den anderen Leukämien getrennt werden konnten (vgl. Abbildung 3-18).

Beim Vergleich der Expressionsprofile sämtlicher cALL-Proben mit den gesunden Kontrollen wurde AML1 mit zwei SONDENSÄTZEN als differentiell exprimiert detektiert. In den leukämischen Zellen ist AML1 leicht heraufreguliert. Im Detail betrachtet zeigt sich aber kein Unterschied zwischen den verschiedenen Leukämien. Verglichen mit den TEL/AML1-negativen Leukämien ist die Expression in den Zellen mit dieser Translokation nicht weiter erhöht. Zusammengefasst bedeutet das: die Expression von AML1 ist in der cALL erhöht, TEL ist in den Leukämien und in den Kontrollen nicht exprimiert, die Leukämien mit und ohne TEL/AML1-Translokation unterscheiden sich in der Expression dieser beiden Gene nicht auffallend.

Das Fusionstranskript steht unter der Kontrolle des TEL-Promotors. Damit bleibt die Frage, warum AML1 als exprimiert detektiert wird und für TEL selbst keine Expression in den TEL/AML1-positiven Proben zu verzeichnen ist, aus den Microarraydaten heraus unbeantwortet. Das Fusionstranskript als solches kann von den Genechips nicht detektiert werden. TEL1 ist aber auch weder in den Patientenproben ohne die

Translokation, noch in den Kontrollen nachzuweisen. Es lässt sich kein direkter Zusammenhang zwischen einem der beiden Transkriptionsfaktoren im Fusionstranskript und den differentiell exprimierten Genen feststellen. Nach heutigem Wissensstand wird die Expression der hier als differentiell exprimiert identifizierten Gene nicht von TEL oder AML1 gesteuert. Die Unterschiede zwischen den Leukämien mit und ohne die Translokation sind gering (vgl. 3.8.4). Es gibt nicht viele differentiell exprimierte Gene, in denen sich die beiden Gruppen von den Kontrollen unterscheiden (vgl. 3.8.1). Vielleicht sind die Veränderungen, hervorgerufen durch die Translokation, nicht so gravierend, weil beide reguläre Transkriptionsfaktoren in hämatopoetischen Zellen aktiv sind und die Fusion nicht zu einer tiefgreifenden Fehlregulation führt. Außerdem gibt es Hinweise auf die Funktion von TEL als starker Repressor der Transkription und auf eine veränderte Lokalisation des Fusionstranskripts im Zellkern, die sich beide in der Funktion des Fusionstranskripts ausdrücken könnten (Chakrabarti, 2000).

Die differentiell exprimierten Gene, die sich exklusiv in den cALL Proben mit TEL/AML1-Translokation nachweisen lassen, deuten mögliche Erklärungsansätze für klinische Besonderheiten dieser Patientengruppe an, die sich unter anderem durch eine günstigere Prognose auszeichnen. Das in diesen Zellen heraufregulierte EFNB2 ist nicht nur in die Neurogenese involviert, sondern auch an Angiogenese und der Bildung weiterer Organe beteiligt. Es zeigt eine Überexpression in Karzinomen des Magens und in Zellen von Neuroblastomen, wo die Überexpression mit einem günstigen Krankheitsverlauf korreliert (Tang, 2000). Der Rezeptor FGFR1 für den Fibroblasten-Wachstumsfaktor FGF ist an Organogenese und Onkogenese beteiligt und zeigt eine hohe Expression in diesen Leukämiezellen. Das Gen für FGFR1 ist Fusionspartner in verschiedenen Lymphomen und myeloproliferativen Erkrankungen. In der AML kann eine Überexpression von FGFR1 ohne die Beteiligung einer Translokation nachgewiesen werden (Larramendy, 2002). In den Zellen der TEL/AML1-positiven cALL ist aber auch SPRY2 überexprimiert, welches ein negativer Regulator von FGF in der Embryogenese von *Xenopus* (Mood, 2002), in der Entwicklung der Trachea von *Drosophila* und in den humanen Melanomzellen A375 ist (Glienke, 2000). Hier deutet sich ein spezieller Mechanismus der TEL/AML1-positiven Zellen zur Unterdrückung der durch FGFR1 vermittelten onkogenen Signale an.

Das Expressionsprofil dieser Zellen weist auch auf einen speziellen anti-apoptotischen Mechanismus hin. In diesen Zellen ist der Rezeptor TNFRSF21 (DR6) deutlich überexprimiert. Die Aktivierung von DR6 führt unter Beteiligung der Caspasen 3 und 7 in die Apoptose (Kasof, 2001a) und die Überexpression von DR6 induziert in manchen Karzinomzelllinien Apoptose. Neben DR6 ist in den TEL/AML1-positiven

Zellen auch BIRC7 (Livin) stark exprimiert, das in die Familie der Apoptoseinhibitoren (IAP) gehört. BIRC7 ist in Zellen der Plazenta, des fetalen Gehirns und in Melanomzellen nachweisbar. Der anti-apoptotische Effekt von BIRC7 beruht auf einer direkten Inhibition des Apoptose-initiiierenden DR6 (Kasof, 2001a; Kasof, 2001b). Zusätzlich exprimieren alle hier untersuchten Leukämiezellen BIRC5, das gezielt die durch DR6 aktivierten Caspasen 3 und 7 inhibiert (Kasof, 2001a) und verstärken die Anti-Apoptose.

#### *4.5 Vergleich der Ergebnisse mit anderen Genexpressionsstudien*

In den vergangenen Jahren sind einige wenige Studien veröffentlicht worden, die die Genexpressionsprofile verschiedener Leukämien und Lymphome untersucht haben. Diese Untersuchungen sind sowohl an pädiätrischen als auch an erwachsenen Patienten durchgeführt worden. Weiterhin gibt es einige Arbeiten zu Genexpressionsprofilen verschiedener normaler Gewebe mit dem Ziel, gewebespezifische Expressionsprofile gesunder Zellen aufzuklären (Su, 2002; Warrington, 2000). Die dort ermittelten Ergebnisse sind in der Betrachtung der in der cALL differentiell exprimierten Gene bereits berücksichtigt worden (vgl. 3.6). Im folgenden sollen die Ergebnisse einzelner Untersuchungen, bei allen Einschränkungen die dabei zu machen sind, mit den vorliegenden Ergebnissen verglichen werden. Wenngleich die methodischen Unterschiede zwischen den einzelnen Studien gering sind und prinzipiell dasselbe gemessen wird, ist mit Abweichungen in den Ergebnissen zu rechnen, die unterschiedliche Ursachen haben können. Das fängt mit dem Unterschied zwischen cDNA- und Oligonukleotid-Arrays an, die jeweils anders gestaltete Sonden tragen. Weiterhin gibt es Differenzen in den Ursprungssequenzen, aus denen die Sonden entwickelt wurden, obwohl sie letztlich dasselbe Gen oder dieselbe mRNA detektieren. Die Qualität der zu untersuchenden Proben kann schwanken. Auf der einen Seite stehen hier Proben aus der Zellkultur, die sehr homogen sind, was die Zellzusammensetzung betrifft. Andererseits wird Material aus Biopsien mit verschiedenen großen Anteilen kontaminierender Zellen aus angrenzenden Geweben untersucht. Letztlich spielt auch das Studiendesign eine Rolle. Aus einem Vergleich von ALL-Proben mit AML-Proben sind andere Ergebnisse zu erwarten als aus dem Vergleich von Leukämiezellen mit gesunden Kontrollen. Die Hindernisse im Vergleich der Ergebnisse anderer Arbeiten mit den vorliegenden Daten haben teilweise dieselben Ursachen oder Hintergründe wie diejenigen, welche den Vergleich der Daten von EOS-H mit HG-U133A erschwert haben.

In einigen Arbeiten wurden bereits Genexpressionsprofile maligner lymphatischer oder myeloischer Erkrankungen untersucht. Hierunter fallen Arbeiten zu akuten und

chronischen Leukämien und zu Lymphomen. A. Ferrando *et al.* (Ferrando, 2002) haben T-Zell Leukämien im Kindesalter untersucht und die Blasten mit gesunden T-Zellen verglichen. Dabei haben sie für die Krankheit charakteristische Signaltransduktionswege aufgedeckt. In einer anderen Arbeit (Larramendy, 2002) wurde gesundes Knochenmark mit den Zellen von AML-Patienten verglichen. Lymphome, das sind solide Tumoren aus B-Lymphozyten verschiedener Reifungsstadien, stehen in einem engen Zusammenhang zu den akuten B-Zell Leukämien. Sie gehen auf vergleichbare Zellen zurück und sind mit den bisherigen Möglichkeiten mitunter diagnostisch schwer voneinander zu trennen. Definitionsgemäß sind bei Lymphomen neben dem lokalen Tumor nicht mehr als 25% Blasten im Knochenmark zu finden. Unter Verwendung von Microarrays sind die B-Zell Lymphome in letzter Zeit untersucht worden. So konnte unter anderem gezeigt werden, dass sich Non-Hodgkin-Lymphome in zwei Gruppen unterteilen lassen (Alizadeh, 2000). Die eine Klasse ähnelt den B-Zellen in Keimzentren der Lymphknoten, die andere Klasse ähnelt aktivierten B-Zellen. In beiden Fällen handelt es sich um reife B-Zell Populationen und nicht um die Vorläuferzellen, die die akuten Leukämien ausmachen. In zwei neueren Arbeiten (Rosenwald, 2002; Shipp, 2002) wird versucht mit Hilfe der Genexpressionsprofile prognostische Aussagen über den Krankheitsverlauf zu machen. Zwei andere Arbeiten befassen sich mit den Genexpressionsprofilen der B-CLL, einer Erkrankung, die aufgrund des entarteten Zelltyps - auch hier handelt es sich um reife B-Zellen - den Lymphomen näher steht als der cALL (Klein, 2001; Rosenwald, 2001).

Die nun folgenden Arbeiten sollen näher diskutiert werden, da sie sich mit den Genexpressionsprofilen der ALL befassen. Eine der ersten Studien auf diesem Gebiet wurden ALL mit AML verglichen (Golub, 1999). Zum ersten Mal konnte hier gezeigt werden, wie allein mit Genexpressionsprofilen biologisch und klinisch relevante Unterscheidungen vorgenommen werden können. Das Anliegen der Arbeit war darzustellen, dass sich mit Hilfe krankheitsspezifischer Genexpressionsprofile vorher unbekannte Proben einer Gruppe zuordnen lassen (class prediction, Klassifikation). Weitere Aufmerksamkeit galt der Möglichkeit, aus Genexpressionsdaten verschiedener Patientenproben mit einer Leukämie neue Unterklassen definieren zu können (class discovery, Clustern). Beide Ansätze erfordern die Detektion differentiell exprimierter Gene. Die Autoren haben insgesamt 50 Gene benannt, von denen 25 jeweils in einer der beiden Erkrankungen überexprimiert sind. In den ALL-Proben waren unter anderem heraufreguliert E2A, ADA3 und MYB, die auch im Vergleich zu den gesunden Kontrollen hier eine höhere Expression zeigten. Für die Gene IRF2, CD79a, MYL2 gilt hingegen, dass sie in der ALL stärker exprimiert sind als in der AML, im Vergleich zu den gesunden Kontrollen dieser Untersuchung aber herunterreguliert sind. In der AML

überexprimiert und damit in der ALL unterexprimiert sind unter anderem AZU1, LYN und PRG1, die auch in den Zellen der hier untersuchten pädiatrischen cALL im Vergleich zu den gesunden B-Lymphozyten aus Nabelschnurrestblut herunterreguliert sind. Insbesondere E2A und CD79a verweisen auf den B-lymphozytären Ursprung der cALL-Blasten.

In einer kleinen Studie (Chen, 2001b) wurde versucht, mit Hilfe von Microarrays Marker zu finden, die sich dazu eignen, nach abgeschlossener ALL-Therapie Minimale Resterkrankungen zu detektieren. Zu diesem Zweck wurden Zellen von vier Patienten zum Zeitpunkt der Erstdiagnose mit Knochenmarkszellen von zwei gesunden Probanden verglichen, einem Säugling und einem Erwachsenen. Die Autoren benennen letztlich neun Gene, deren Expression in den leukämischen Blasten ihrer Studie heraufreguliert sind und von denen sie glauben, diese eigneten sich zur Überwachung einer MRD. Von diesen neun Genen zeigt lediglich die Kinase CKB auch in dieser Untersuchung eine geringfügige Überexpression. Indes finden sich im vorliegenden Datensatz zwei Gene, die in den hier untersuchten cALL-Proben deutlich und signifikant herunterreguliert sind. Das Differenzierungsantigen MNDA myeloischer Zellen ist ebenso herunterreguliert wie ANX6. Diese Diskrepanz in den Ergebnissen ist weniger in den verschiedenen verwendeten Genechips zu suchen, als vielmehr im unterschiedlichen Studiendesign.

In einem anderen Ansatz (Moos, 2002) wurde versucht, verschiedene Genexpressionsprofile zu identifizieren, mit deren Hilfe unterschiedliche Formen pädiatrischer Leukämien (AML und ALL, B-Zell- und T-Zell-ALL, Standard-Risiko und Hoch-Risiko ALL, ALL mit TEL/AML1-Translokation und ohne Translokation) klassifiziert werden können. Mit zwei unterschiedlichen statistischen Methoden wurden 29 Gene identifiziert, die sich in ihrer Expression zwischen den Gruppen mit und ohne TEL/AML1-Translokation unterscheiden. In der vorliegenden Arbeit konnte jedoch keines dieser Gene als Markergen für die Unterscheidung von Proben mit oder ohne TEL/AML1-Translokation gefunden werden (vgl. 3.8). Allerdings zeigten drei der Gene (CD45, GRHPR und GARS) eine verringerte Expression, wenn man die leukämischen Proben mit den gesunden B-Lymphozyten vergleicht.

E. Yeoh *et al.* (Yeoh, 2002) haben die Genexpressionsprofile von 360 pädiatrischen Leukämiepatienten miteinander verglichen. Darunter waren Gruppen von Patienten mit einer cALL und den Translokationen t(1;19)(q23;p13) E2A/PBX1, 11q23 MLL oder t(12;21)(p13;q22) TEL/AML1. Für jede der Patientengruppen mit definierter Translokation konnte ein spezifisches Expressionsprofil erstellt werden, in dem sich die einzelne Gruppe von den anderen unterscheidet. Weiterhin gelang es den Autoren, nicht zuletzt wegen der großen Probenanzahl, zu zeigen, wie die

Genexpressionsmuster mit dem Krankheitsverlauf korrelieren können. Aus dem Vergleich der Proben mit TEL/AML1-Translokation ergeben sich drei differentiell exprimierte Gene. MEIS1 und GS3955 sind im vorliegenden Datensatz in den Proben mit TEL/AML1-Translokation herunterreguliert. Bei E. Yeoh stehen sie als Marker für MLL- oder E2A/PBX1-positive ALL, was bedeutet, dass sie bei den TEL/AML1-positiven unterexprimiert sind. ARHGEF4 hingegen steht in beiden Datensätzen für eine deutliche Überexpression in den Zellen der TEL/AML1-positiven Leukämien. Diese Befunde machen die Heterogenität der hier untersuchten Leukämieproben deutlich, wenn man sie als eine phänotypisch einheitliche Gruppe auffasst und etwaige Translokationen nicht beachtet.

S. Armstrong *et al.* (Armstrong, 2001) haben ebenfalls Leukämien mit definierten Translokationen miteinander verglichen. Den 17 Proben mit einer Translokation am MLL-Locus 11q23 stehen als Kontrollen 20 Leukämien gegenüber, von denen zehn die TEL/AML1-Translokation trugen. Wichtigstes Ergebnis dieser Arbeit ist die Feststellung, dass sich die Genexpressionsprofile der Patienten mit MLL-Translokation von denen der anderen deutlich unterscheiden, dass dieses Genexpressionsprofil zwischen ALL und AML anzusiedeln ist und deshalb von einer eigenen Krankheitsentität zu sprechen sei, der MLL-ALL. Es wird eine Liste von 100 Genen angegeben, in denen sich die Leukämien mit der MLL-Translokation von den übrigen akuten Leukämien unterscheiden. Es sind einige Gene in der allgemeinen Leukämiegruppe heraufreguliert, die auch im Vergleich der hier untersuchten cALL-Proben mit den gesunden B-Zellen aus Nabelschnurblut heraufreguliert sind: CD10, TdT, MADH1, SMARCA4, DBN1, FHIT, MONDOA, FOXO1A, APP, SPTB. Im umgekehrten Fall, relative Überexpression in Leukämien mit MLL-Translokation beziehungsweise relative Überexpression in den hier untersuchten gesunden B-Zellen, lassen sich die folgenden Gene identifizieren: CD45, ELANH2 (SERPINB1), EBI2, PRG1, KIAA0027, ANXA1 und RNASE3. Diese Gene lassen somit nicht nur eine Unterscheidung zwischen cALL und gesunder Kontrolle zu, sondern auch zwischen cALL und ALL mit MLL-Translokation.

In einer großangelegten Studie (Warrington, 2000) wurden mit Hilfe von Genechips konstitutiv exprimierte Gene identifiziert, die in allen getesteten Geweben exprimiert werden. Dabei wurden sieben adulte Gewebe (Herz, Hirn, Lunge, Niere, Bauchspeicheldrüse, Gebärmutter und Hoden) und vier fetale Gewebe (Hirn, Lunge, Niere und Leber) untersucht. Die verwendeten Affymetrix HuGeneFL Genechips detektieren über 7.000 bekannte Gene. In diesem Ansatz konnten Gene identifiziert werden, die jeweils exklusiv in einem der elf Gewebe exprimiert waren und von den Autoren als gewebespezifische Markergene bezeichnet werden. Weiterhin wurden Gene ausfindig gemacht, deren mRNA nur in adulten oder nur in fetalen Geweben

nachzuweisen war. Als Haushalts-Gene bezeichnen die Autoren die 535 Gene, die in allen getesteten Proben nachzuweisen waren. Diese Gene werden offensichtlich mit der frühen Embryonalentwicklung angeschaltet und bleiben kontinuierlich im Erwachsenenalter exprimiert. Diese Gene scheinen essentiell für die Aufrechterhaltung allgemeiner zellulärer Funktionen zu sein. Unter den 535 identifizierten Genen befinden sich 47, die in allen elf Proben mit derselben Intensität exprimiert waren. Die restlichen Proben zeigten, wenngleich ubiquitär exprimiert, gewebespezifische Schwankungen in der Expressionsstärke. Der Satz der 47 ubiquitär und gleichmäßig stark exprimierten Gene eignet sich möglicherweise als interner, quantitativer Kontrollsatz für Genexpressionsanalysen (Warrington, 2000).

Vergleicht man diese 47 Haushalts-Gene mit den vorliegenden Ergebnissen, werden allerdings Unterschiede sichtbar. Von den 47 Genen konnten elf als SONDENSATZ auf den EOS-H Genechips nicht identifiziert werden, auf den HG-U133A Genechips sind acht der 47 Gene nicht durch entsprechende SONDENSATZE repräsentiert. Diese acht fehlenden Gene sind auch in den elf fehlenden Genen des EOS-H Genechips enthalten. Unter Berücksichtigung dieser nicht vertretenen Gene sind bei den mit EOS-H untersuchten Proben 28 Gene (entspricht 78%) als gleichmäßig exprimiert detektiert worden. Für die mit HG-U133A untersuchten Proben sind dies 27 Gene (entspricht 69%). Das bedeutet, dass es für acht Gene (entspricht 22%) auf EOS-H und für zwölf Gene (entspricht 31%) auf HG-U133A abweichende Ergebnisse gibt. Auffallend ist, dass es auf beiden Chiptypen nicht immer dieselben Gene sind, die dieses abweichende Ergebnis liefern, was an unterschiedlichen Hybridisierungseigenschaften der verschiedenen SONDENSATZE liegen mag. Lediglich für drei Gene (c-yes-1, SR31747 Bindungsprotein 1, 5-TH2c Rezeptor) ist es für beide Genechiptypen eindeutig, dass diese im vorliegenden Datensatz nicht als konstitutiv exprimiert gelten können. Die Autoren geben für die von ihnen detektierten Haushalts-Gene eine relative Größe für die Expressionsstärke an (von hoch exprimiert bis niedrig exprimiert in fünf Stufen). Es fällt auf, dass die Gene, die im vorliegenden Datensatz nicht eindeutig als exprimiert erkannt worden sind, von den Autoren durchgängig als mittel bis niedrig exprimierte Gene detektiert wurden. Die widersprüchlichen Ergebnisse könnten demnach auf unterschiedliche Bewertungen der Signalintensitäten im unteren Messwertbereich zurückzuführen sein.

Für diese voneinander abweichenden Ergebnisse mag es drei gewichtige Gründe geben. Zum einen könnte die pathologische Entartung der leukämischen Zellen dazu beitragen, dass manche konstitutiv exprimierten Gene nicht ubiquitär und mit einer gleichbleibenden Stärke exprimiert werden. Zum anderen können die Ursachen in der unterschiedlichen Gestaltung von Genechips und den verschiedenen SONDENSATZEN mit

ihren individuellen Hybridisierungseigenschaften liegen. Damit ist der Möglichkeit über die Expression der konstitutiv exprimierten Gene eine chipübergreifende Normalisierung durchführen zu wollen, eine klare Grenze gesetzt. In einer anderen Arbeit (Lee, 2002) wurde eben diese Möglichkeit der Normalisierung untersucht, mit dem Ergebnis, das Konzept der Haushalts-Gene insgesamt in Frage zu stellen. Die Autoren konnten mit Hilfe von Genechips große quantitative Schwankungen in der Expression von zwölf Genen feststellen, die üblicherweise zur Normalisierung von Northern-Blot oder RT-PCR Daten verwendet werden. Dabei wurden nicht einmal unterschiedliche Genechiptypen für die Studie verwendet. Für die Unterschiede in den Ergebnissen der Studie von J. Warrington und den hier vorgelegten Daten ist damit vermutlich nicht vorrangig die Pathologie der cALL-Proben verantwortlich zu machen. Insgesamt zeigen diese beiden Arbeiten die Schwierigkeiten beim Vergleich von Genexpressionsdaten auf, die mit unterschiedlichen Microarrays ermittelt wurden.

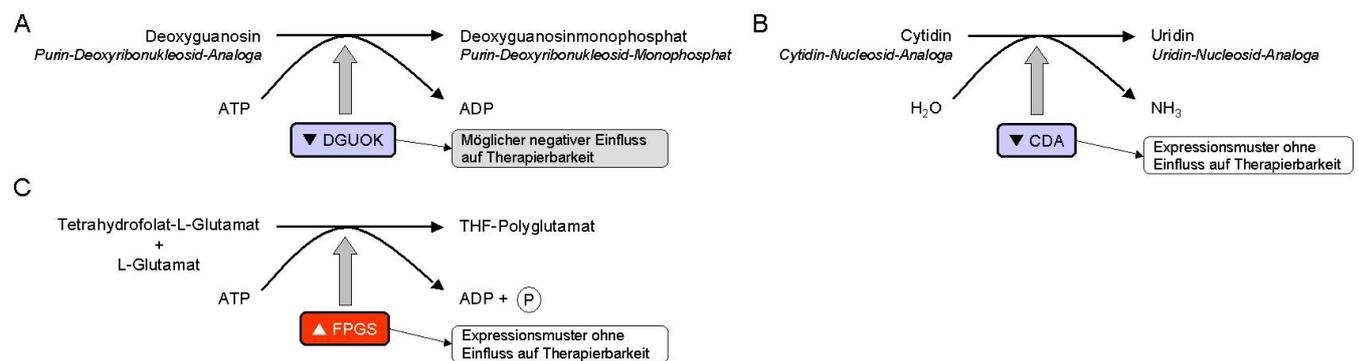
#### *4.6 Gene mit besonderem Einfluss auf die Pathologie der cALL*

Aus der Vielzahl differentiell exprimierter Gene in den Zellen der cALL sollen in diesem und den folgenden Kapiteln diejenigen Gene diskutiert werden, die vermutlich eine besondere Bedeutung für die Pathologie der cALL besitzen. Dabei erhebt diese Auswahl keinen Anspruch auf vollständige Berücksichtigung aller relevanten Gene. Aus den Daten heraus lässt sich zum Beispiel nicht ablesen, ob es sich bei der differentiellen Expression eines Gens um ein ursächlich pathogenes Phänomen oder um ein konsequentes Ereignis der Krankheit handelt.

Die fünf folgenden Gene sind von besonderem Interesse für die Therapierbarkeit der cALL. Die Genprodukte haben einen direkten Einfluss auf die Wirksamkeit verschiedener Chemotherapeutika. Das Gen ABCB4 (ATP-Binding Cassette, Subfamily B, Member 4) kodiert das Multi-Drug-Resistance Protein MDR3, welches aktiv unter Verbrauch von ATP hydrophobe Wirkstoffe, damit auch Chemotherapeutika, aus der Zelle heraustransportieren kann. In einer Studie (Arai, 1997) konnte gezeigt werden, dass sich in lediglich acht von 54 untersuchten Patienten (entspricht 15%) ABCB4 nachweisen ließ und dies mit dem Nichterreichen einer kompletten Remission korrelierte. Die Expression von ABCB1 (MDR1), einem weiteren Mitglied dieser Proteinfamilie, konnte in Zelllinien nachgewiesen werden, die von therapieresistenten Leukämiepatienten stammen. In den hier untersuchten Patientenproben war ABCB1 in keinem Fall exprimiert. ABCB1 ist integraler Bestandteil der Blut-Hirn-Schranke und transportiert Wirkstoffe aus dem Gehirn zurück ins Blut. Eine Überexpression von ABCB1 ist in verschiedenen Karzinomen mit multiplen Therapieresistenzen berichtet

worden. Beide MDR Gene sind in den leukämischen Proben herunterreguliert. Dieses Expressionsmuster hat keinen negativen Einfluss auf die Therapierbarkeit.

Zur Behandlung der ALL oder AML werden Purin-Deoxyribonucleosid-Analoga eingesetzt. Damit diese zytotoxisch wirken können, müssen sie in der Zelle von der Deoxyguanosin-Kinase DGUOK phosphoryliert werden (vgl. Abbildung 4-2-A). Die Expression von DGUOK zeigt bei pädiatrischen Patienten große individuelle Schwankungen, ist aber in den hier untersuchten cALL-Proben durchgehend nicht exprimiert, was einen Hinweis auf mögliche Schwierigkeiten der Therapierbarkeit dieser Patienten mit Purin-Deoxyribonucleosid-Analoga liefern könnte.



**Abbildung 4-2 Genexpressionsprofile und Therapieresistenz**

Das Expressionsprofil einzelner Gene lässt Aussagen über mögliche Therapieresistenzen zu. Die Fehlende Expression von DGUOK oder CDA kann einen negativen Einfluss auf die Therapierbarkeit mit verschiedenen Nucleosid-Analoga haben. Die erhöhte Expression von FPGS hingegen kann einen positiven Effekt auf die Therapierbarkeit mit Antifolaten haben.

Das Genprodukt von CDA, die Cytidin-Deaminase, kann neben ihrer natürlichen Funktion die Deaminierung von Cytosin-Nucleosid-Analoga katalysieren (vgl. Abbildung 4-2-B). Diese werden wegen ihrer zytotoxischen Wirkung bei der Behandlung der AML in Form des Cytosin-Arabinosids eingesetzt. Die Deaminierung der Analoga führt zum Verlust der gegen den Tumor gerichteten Wirkung. Die damit verbundene klinische Resistenz gegen Cytosin-Nucleosid-Analoga ist eine der Hauptursachen für ein Versagen der AML-Therapie. Die Cytosin-Nucleosid-Analoga werden bei erwachsenen Patienten mit Formen der ALL eingesetzt, die schwer zu behandeln, therapieresistent oder mit einer schlechten Prognose verbunden sind. In den hier untersuchten pädiatrischen cALL-Proben ist CDA deutlich schwächer exprimiert als in den Kontrollen und damit vermutlich ohne Einfluss auf die Therapie mit Cytosin-Nucleosid-Analoga. Es ist zu bemerken, dass diese Wirkstoffgruppe normalerweise bei der pädiatrischen cALL nicht eingesetzt wird.

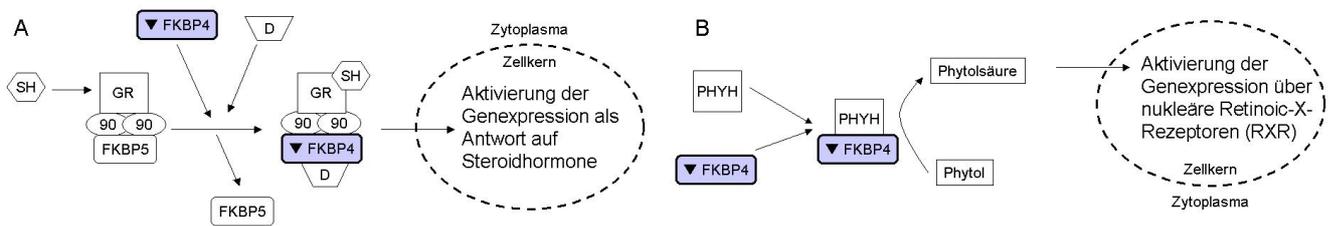
Die Wirksamkeit von Methotrexat als Therapeutikum zur Behandlung unter anderem von Krebserkrankungen ist abhängig von der enzymatischen Aktivität von FPGS (Folypolyglutamat-Synthetase), welches in methotrexatresistenten ALL-Patienten fehlt

(Gorlick, 1999). In den hier untersuchten Proben ist FPGS jedoch immer verstärkt exprimiert. Folsäure ist ein für das normale Zellwachstum essentielles Vitamin. Antifolate, zum Beispiel Methotrexat, werden als Medikamente zur Behandlung von bakteriellen Infektionen oder von Leukämien eingesetzt. Für die Verfügbarkeit der Folsäure und der Antifolate als Therapeutika in der Zelle ist die enzymatische Aktivität von FPGS notwendig (vgl. Abbildung 4-2-C).

Die Betrachtung des Expressionsmusters dieser vier Gene kann im einzelnen hilfreich sein, den Behandlungserfolg vor Therapiebeginn abzuschätzen, und deutet damit einen möglichen Einsatz von Genechips in der Diagnose der cALL an. Solche Ansätze werden bereits verfolgt. In einer großen Studie (Staunton, 2001) wurden die Genexpressionsprofile von 60 humanen Krebszelllinien bestimmt, deren Chemosensitivitätsprofil mit mehr als tausend Chemikalien bestimmt worden war. Damit konnten sie zeigen, dass aus dem Expressionsprofil der unbehandelten Zellen die Chemosensitivität vorhergesagt werden kann. Die Daten dieser fünf als differentiell exprimiert gefundene Gene deuten an, dass sich aus den Expressionsprofilen der Leukämien Aussagen zur Therapierbarkeit ableiten lassen und sich Genechips diagnostisch nutzen lassen.

Das Expressionsmuster zweier weitere Gene kann einen besonderen Einfluß auf die Pathologie der cALL besitzen. Die Wirkweise dieser Genprodukte liegt allerdings nicht in Resistenzen zur Chemotherapie, sondern in der Regulation von Genexpression. Das FK506-bindende Protein FKBP4 (FKBP52) ist in den Zellen der meisten untersuchten cALL-Patienten herunterreguliert. Es besitzt strukturelle und funktionelle Homologien zum N-Terminus des FKBP12. Als Mitglied der Immunophilin-Familie zeigt es Chaperoneigenschaften und besitzt eine Prolyl-Isomeraseaktivität. Wie die anderen Mitglieder dieser Proteinfamilie komplexiert es mit FK506, zeigt dabei aber keine immunsupprimierenden Eigenschaften. FKBP4 assoziiert mit dem Hitzeschockprotein HSP90 und ist damit Bestandteil von Steroidhormonrezeptoren (Davies, 2002; Peattie, 1992). Hierüber ist FKBP4 an intrazellulären Signalwegen beteiligt. Nach Bindung eines Steroidhormons an den Rezeptorkomplex wird in der Zelle FKBP5 gegen FKBP4 ausgetauscht. Der so aktivierte Komplex kann in den Zellkern wandern und die Expression der Zielgene in Antwort auf das Steroidhormon aktivieren (Abbildung 4-3-A). Durch das Fehlen von FKBP4 in den leukämischen Zellen könnte dieser Signalweg unterbrochen sein. In einem anderen Zusammenhang interagiert FKBP4 mit dem peroxisomalen Enzym Phytanoyl-CoA-Hydroxylase (PHYH). Auch hierbei greift FKBP4 in intrazelluläre Signalwege ein (Chambraud, 1999). Nach Assoziation von FKBP4 mit PHYH kann Phytol metabolisiert werden. Über nukleäre Retinoic-X-Rezeptoren kann das metabolisierte Phytol die Expression von Zielgenen aktivieren (Abbildung 4-3-B).

Auch dieser Signalweg könnte durch das Fehlen von FKBP4 in den leukämischen Zellen unterbrochen sein.



**Abbildung 4-3 FKBP4 im Zusammenhang intrazellulärer Signaltransduktion**

FKBP4 ist an verschiedenen intrazellulären Signalwegen beteiligt. Über den Glukokortikoidrezeptor gibt es eine Verbindung zur Steroidhormon-vermittelten Aktivierung der Genexpression (A). Durch die Bindung an PHYH greift es in die Phytol-vermittelte Aktivierung von Genexpression ein (B). SH: Steroidhormon, GR: Glukokortikoidrezeptor, 90: HSP90, D: Dynein, PHYH: Phytanoyl-CoA-Hydroxylase. Teil A verändert nach: Davies, 2002.

WHSC1 ist in den leukämischen Blasten der cALL heraufreguliert. Bei dem Wolf-Hirschhorn Syndrom, der namensgebenden Krankheit dieses Gens, handelt es sich um ein Fehlbildungssyndrom. Die Krankheit geht mit einer Deletion im kurzen Arm von Chromosom 4 einher, dem Bereich in dem WHSC1 liegt. In Myelomen konnte gezeigt werden, dass WHSC1 mit dem Immunglobulinen für IgH fusioniert ist (Stec, 1998). Im gesunden Menschen ist WHSC1 während der frühen Entwicklung ubiquitär exprimiert. Es verfügt über eine SET-Domäne, die in vielen Proteinen mit Funktionen in Entwicklung und Transkriptionskontrolle vorkommen. Das humane WHSC1 ist homolog zu SU(VAR)3-98B aus *D. melanogaster* (Schotta, 2002). Hierbei handelt es sich um ein Mitglied einer Proteinfamilie, die Heterochromatin-vermitteltes Gene-Silencing induzieren und darüber die Aktivität besonders von entwicklungsrelevanten Genen steuern. Für diese Funktion verfügen die Mitglieder dieser Familie über eine Methyltransferaseaktivität, die spezifisch für Histon H3 ist. Für WHSC1 ist ein weiterer Funktionsmechanismus beschrieben, über den die Aktivität von Genen inhibiert werden kann. Durch Bindung an den Promotor wird die Transkription von IL5 spezifisch unterbunden (Garlisi, 2001). Auf welchem Weg WHSC1 in den leukämischen Zellen die Expression bestimmter Gene herunterreguliert, muss experimentell überprüft werden.

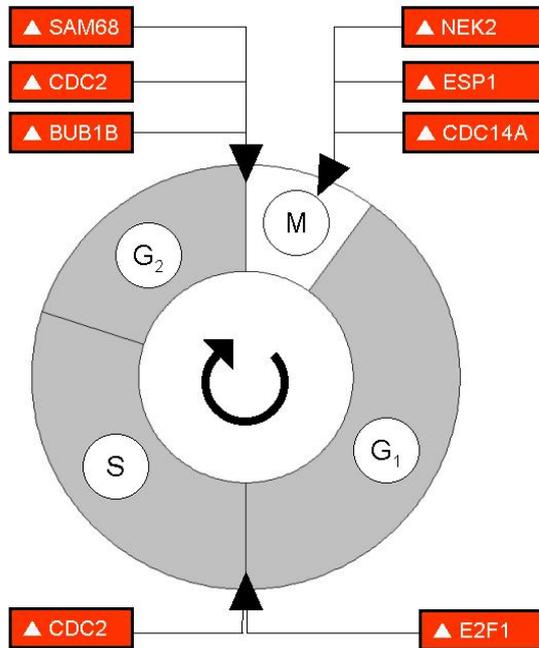
Die Überexpression von WHSC1 und die fehlende Expression von FKBP4 können auf verschiedenen Wegen an der Herunterregulation der Genexpression in den Lymphoblasten beteiligt sein. Auf diesem Wege könnte das Genexpressionsmuster dieser beiden Gene einen Erklärungsansatz für den Befund liefern, dass in den leukämischen Zellen im Vergleich zu den gesunden Kontrollen wesentlich mehr Gene herunter- als heraufreguliert sind.

#### 4.7 Leukämische Blasten sind Zellen mit hoher Proliferationsrate

Die leukämischen Blasten teilen sich schnell und scheinbar unkontrolliert. Sie vermehren sich rasch, verdrängen andere Zellen in Knochenmark und Blut und prägen damit das klinische Erscheinungsbild der Leukämie. In diesem Zusammenhang sind die differentiell exprimierten Transkripte zu sehen, deren Produkte Aufgaben in der Regulation des Zellzyklus' zukommen, die an der stabilen Verteilung der Chromosomen während der Mitose oder direkt an der Replikation von DNA beteiligt sind. Den in den leukämischen Proben stärker exprimierten Genen mit Funktionen im Zellzyklus stehen anzahlmäßig kaum Gene mit einer verringerten Expression gegenüber. Für fast alle hier heraufregulierten Gene ist bereits eine Überexpression in gesunden Geweben mit proliferierenden Zellen oder in anderen Krebserkrankungen beschrieben worden. Die Funktionsweisen und die Angriffspunkte dieser Proteine im Zellzyklus sind ganz unterschiedlich. Ebenso vielfältig sind die Signalwege, die auf die Zellteilung Einfluss nehmen können. Es ist daher nicht möglich, alle hier genannten differentiell exprimierten Gene in einen direkten funktionellen Zusammenhang zu stellen, so dass die Ergebnisse hier teilweise nur nebeneinander genannt werden können.

E2F1 und CDC2 kontrollieren zum Beispiel am G1/S-Übergang, BUB1B, SAM68 und CDC2 am G2/M-Übergang, ESP1, NEK2 und CDC14A am Ende der Mitose. In 70-80% der hepatozellulären Karzinomen ist CDC2 aktiviert und vermutlich an der Pathologie beteiligt (Li, 2002a). SAM68 ist unabhängig von CDC2 am G2/M-Übergang aktiv. SAM68-defiziente Zellen zeigen ein verringertes Zellwachstum, was durch die Verlängerung der G2/M-Phase verursacht wird (Li, 2002b). In der Hefe *Saccharomyces cerevisiae* ist die Phosphatase CDC14 ein Schlüsselprotein zur Beendigung der Mitose. Die Überexpression des humanen Homologs CDC14A verursacht Defekte in der Formation der mitotischen Spindeln und bei der Chromosomensegregation, was zu Fehlern in der Verteilung von Chromosomen und zu genetischer Instabilität führt (Kaiser, 2002). In die Regulation von CDC14 ist unter anderem ESP1 (Separase) involviert. Diese Separase ist ein Regulator der Tochterchromatidseparation, und Mutationen im Gen für ESP1 führen zu Defekten im Beenden der Mitose (Stegmeier, 2002). Die Proteinkinase NEK2 ist ein Regulator der Zentrosomenstruktur und der Zentrosomenfunktion. Die NEK2 Aktivität induziert die Teilung der verdoppelten Zentrosomen. Eine Überexpression bewirkt eine beachtlich gesteigerte Zentrosomenteilung (Fry, 1998). Die chromosomale Instabilität in kolorektalen Krebserkrankungen, die zur abweichenden Chromosomenzahlen führen, ist mit inaktivierenden Mutationen im Gen für BUB1B assoziiert (Cahill, 1998). Die beiden hochregulierten Gene CDC2L1 und CDC2L2 kodieren zwei nahezu identische Proteine, die vermutlich als Tumorsuppressor agieren können und antiproliferativ wirken. In

Abbildung 4-4 sind die einzelnen Proteine mit ihren Kontrollpunkten im Zellzyklus angegeben. Die Überexpression der Gene CDC14A, ESP1, NEK2 und BUB1B in Leukämiezellen ist bislang nicht beschrieben und zeigt mögliche Ursachen der regelmäßig zu beobachtenden Hyper- und Hypodiploidien bei der cALL auf.

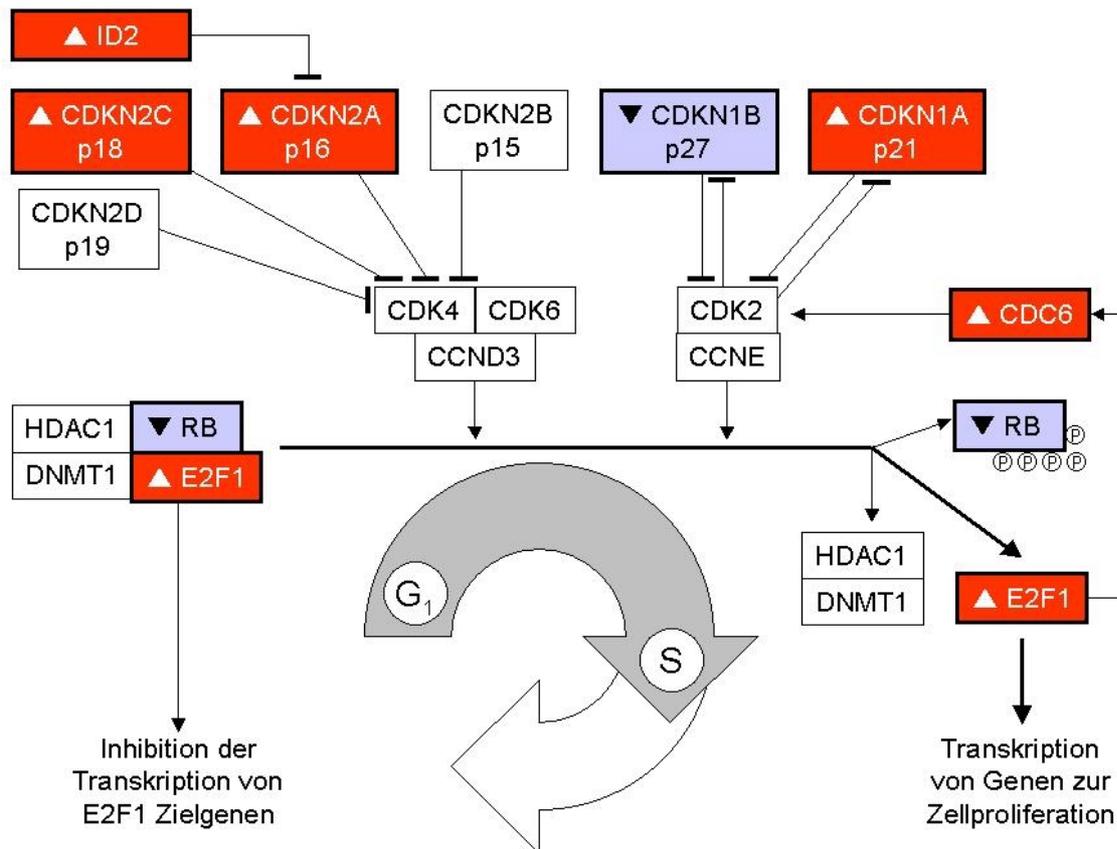


**Abbildung 4-4** Differenziell exprimierte Gene mit Kontrollfunktionen im Zellzyklus

Eine Reihe der in cALL-Zellen überexprimierten Gene sind an der Kontrolle des Zellzyklus' beteiligt. Sie regulieren das Fortschreiten an bestimmten Kontrollpunkten.

Der Übergang von G<sub>1</sub>- zur S-Phase des Zellzyklus', in der es zur Verdopplung des Erbmaterials kommt, ist der Beginn der Zellteilung und unterliegt einer besonderen Kontrolle (Abbildung 4-5). Nach passieren dieses Kontrollpunktes ist der nächste Zellzyklus unwiderruflich eingeleitet. Während der G<sub>1</sub>-Phase liegen die Transkriptionsfaktoren der E2F-Familie komplexiert mit anderen Proteinen vor. Dieser Komplex, bestehend aus DNMT1, HDAC1, RB und E2F, vermag Modifikationen an Chromatin zu inhibieren und damit die Transkription der Zielgene von E2F zu unterbinden (Robertson, 2000). Die von E2F gesteuerte Expression spezieller S-Phasen Gene ist durch den Komplex reprimiert. Durch Phosphorylierung des RB dissoziiert der gesamte Komplex und E2F kann als Transkriptionsfaktor aktiv werden. Die Transkription von E2F-Zielgenen führt zum Fortschreiten des Zellzyklus' in die S-Phase (Leone, 1998). In den Zellen der cALL ist E2F1 heraufreguliert. RB und das verwandte RBL2 (RB2, p130), die beide an E2F-Transkriptionsfaktoren binden, sind indes herunterreguliert. HDAC1 und DNMT1 konnten in allen untersuchten Proben nachgewiesen werden. Die Phosphorylierung des RB kann durch zwei Proteinkomplexe reguliert werden, welche die Schlüsselregulatoren des G<sub>1</sub>/S-Kontrollpunktes darstellen.

Der erste Komplex, bestehend aus den beiden Kinasen CDK4 und CDK6 und dem Cyclin CCND3, ist charakteristisch für B-Lymphozyten. In anderen Zellen wird Cyclin D3 durch andere Cycline ersetzt. Die Proteine CCND3, CDK4 und CDK6 sind in allen Stadien der normalen B-Zellentwicklung und in untersuchten leukämischen Zelllinien nachweisbar (Fink, 2001). Der zweite Proteinkomplex besteht aus Cyclin E CCNE und der Kinase CDK2. Die Bestandteile beider Proteinkomplexe sind in den leukämischen Zellen und den Kontrollen nachweisbar.



**Abbildung 4-5 Regulation des Zellzyklus' am Übergang der G<sub>1</sub>- zur S-Phase**

Der Übergang von G<sub>1</sub>- zur S-Phase des Zellzyklus', in der es zur Verdopplung des Erbmaterials kommt, ist der Beginn der Zellteilung und unterliegt einer besonderen Kontrolle. Viele an der Regulation beteiligte Faktoren sind in den Zellen der cALL differentiell exprimiert. Herunterregulierte Gene: ▼, heraufregulierte Gene▲.

Verschiedene Cyclin-abhängige Inhibitoren (CDKI) regulieren die Aktivität der Cyclin/CDK-Komplexe. Einige dieser Moleküle zeigen in den Zellen der cALL eine differentielle Expression. Heraufreguliert sind CDKN2C (p18, INK4C), CDKN2A (p16, INK4A) und CDKN1A (p21, CIP1). CDKN1B (p27, KIP1) ist als einziges Mitglied dieser Familie herunterreguliert. Die inhibitorischen CDKIs binden an die CDKs und verhindern deren Komplexierung mit den Cyclinen (Ekholm, 2000). Nach Aktivierung durch MYC konnte im Zellkulturmodell gezeigt werden, wie die CDKIs durch die Cycline D1 und D2

abgefangen werden und der Komplex CCNE/CDK2 ausgebildet werden kann (Perez-Roger, 1999). In der Folge wird RB phosphoryliert und dissoziiert aus dem Komplex mit E2F1 (Wyllie, 2002). E2F1 kann nun die Transkription von Genen aktivieren, die Proteine kodieren, welche bei der Replikation von DNA benötigt werden (z.B. DNA Polymerase  $\alpha$ ) oder zur Progression des Zellzyklus' (z.B. Cyclin A). Aus den Genexpressionsdaten ist abzulesen, wie sich die wichtige Regulation am Übergang von der G<sub>1</sub>- zur S-Phase im Ungleichgewicht befindet. Verschiedene Inhibitoren sind im Vergleich zu den Kontrollen über- oder unterexprimiert. RB ist in der cALL herunterreguliert; hingegen ist E2F1 heraufreguliert. Insgesamt zeigen die Daten jedoch auf, dass sich die leukämischen Blasten in einem Zustand hoher Proliferation befinden.

Seit kurzem wird die Überexpression von CD10 in den leukämischen Zellen der cALL mit der Proliferation dieser Zellen in Verbindung gebracht (Cutrona, 2002). Die Autoren konnten zeigen, dass CD10 positive Leukämiezellen der ALL eine erhöhte MYC Expression zeigen und dass sich diese Zellen im Zellzyklus befinden. In den hier untersuchten Proben konnte eine erhöhte Expression von MYC auf mRNA-Ebene nicht gezeigt werden. Die mit unterschiedlichen SONDENSÄTZEN detektierten Transkripte von MYC liefern allerdings widersprüchliche Ergebnisse und lassen keine Aussage zu.

Eine Reihe weiterer differentiell exprimierter Gene bringt die hohe Zellteilungsrate ebenfalls zum Ausdruck. In der Gruppe der Gene, deren Produkte Funktionen in Replikation von DNA, Transkription oder Proteinsynthese übernehmen, ist ein deutliches Expressionsmuster zu erkennen. Zu den heraufregulierten Genen gehören unter anderem die Topoisomerasen TOP1 und TOP2A, die verschiedenen Histonproteine sowie die beiden DNA-Polymerase-Untereinheiten POLE und POLQ. Laut GNF-Expressionsdatenbank zeigen diese Gene die stärkste Expression in den Geweben, die sich durch hohe Zellteilungsraten auszeichnen, oder in den dort untersuchten Zelllinien maligner Erkrankungen. Allerdings wurde bei einem Vergleich hämatopoetischer Stammzellen mit prä-B-Zellen gezeigt, dass TOP2A in den B-Vorläuferzellen schwächer exprimiert ist (Muschen, 2002). Es ist auffallend, dass die meisten heraufregulierten Gene dieser funktionellen Gruppe an der DNA-Replikation, DNA-Reparatur oder Chromosomenverdoppelung im Zuge der Zellteilung beteiligt sind.

Ebenso auffallend ist der Befund, dass die herunterregulierten Gene dieser Gruppe in Prozesse der Transkription und Translation involviert sind, zum Beispiel ELL2 (Elongationsfaktor der Transkription), ADARB1 (Editierung von RNA), RNGTT (Prozessierung von RNA, Capping) oder ribosomale Strukturproteine (unter anderem RPL5, RPL8, RPL12). Die leukämischen Blasten scheinen sich nicht durch eine allgemein erhöhte Genexpression oder Proteinsynthese auszuzeichnen. Ein Vergleich der Anzahl herauf- und herunterregulierter Gene unterstützt diese Sicht. Die

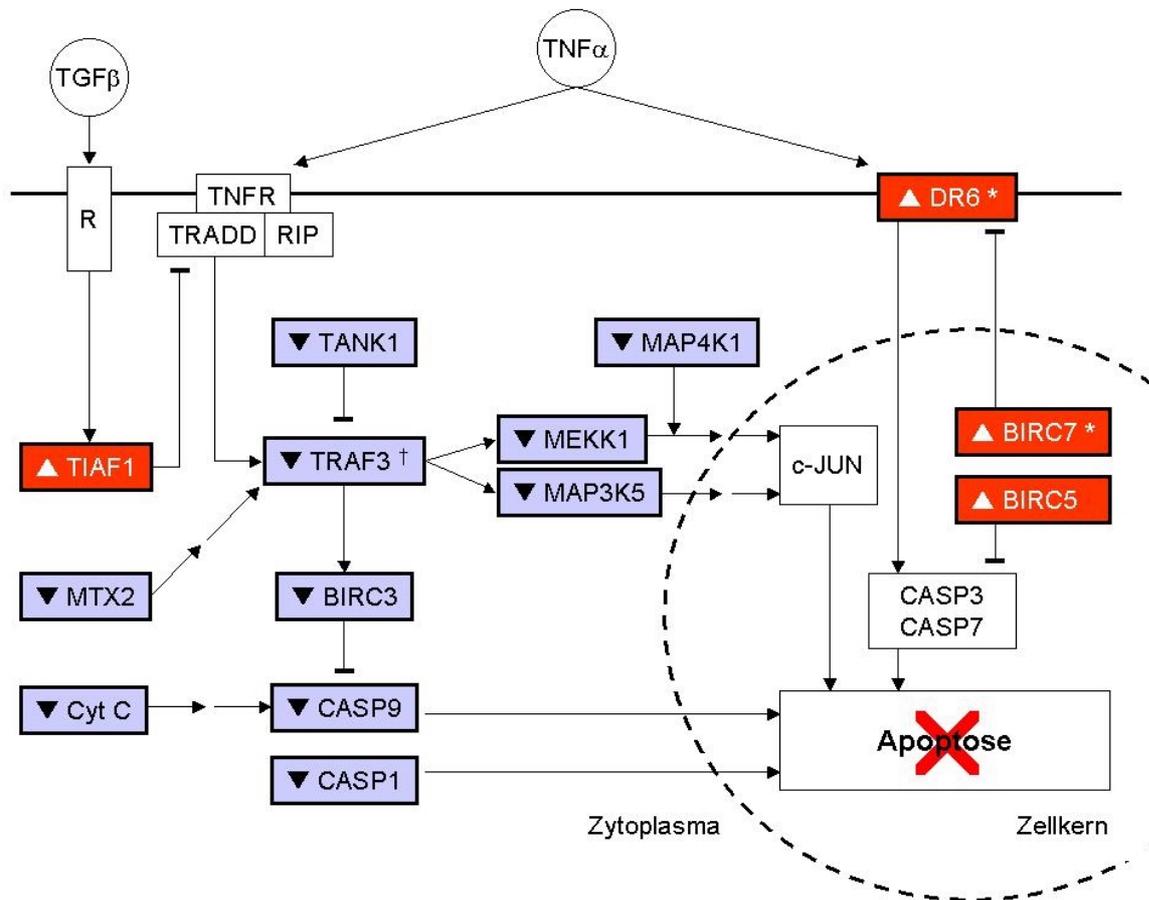
Datensätze differentiell exprimierter Gene, erhoben mit den beiden Genechiptypen, zeigen für die leukämischen Zellen eine deutlich geringere Anzahl heraufregulierter Gene als herunterregulierter Gene. In den leukämischen Zellen scheinen wesentlich mehr Gene abgeschaltet als angeschaltet zu werden. Über diesen interessanten Befund finden sich in der Literatur bislang keine weiteren Hinweise.

#### 4.8 *Überleben und programmierter Zelltod leukämischer Zellen*

Das Auslösen der Apoptose, also des programmierten Selbstmords der Zielzellen, ist eine der Aufgaben zytotoxischer T-Zellen in der Immunabwehr. Wichtige Mediatoren der Apoptose sind Perforin oder Granzyme, sowie verschiedene Chemokine. Mit Ausschüttung eines Zytokins wie  $\text{TNF}\alpha$  oder durch Aktivierung des Oberflächenantigens FAS (CD95) und Übertragung dieses Signals in das Innere der Zielzellen läuft in diesen ebenfalls ein Programm ab, das zum programmierten Selbstmord führt. Gegen Krebszellen scheint das Immunsystem allerdings nur sehr wenig ausrichten zu können. Die Krebszellen benutzen verschiedene Strategien, den Angriffen des Immunsystems auszuweichen. Eine Möglichkeit besteht darin, die Apoptosesignale intrazellulär zu umgehen oder ihnen gegenzusteuern. Die Anzahl der bekannten in die Apoptose involvierten Proteine wächst ständig, wenngleich die spezifische Funktion einzelner Proteine oder die verschiedenen Regulationswege noch unbekannt sind. Im Genexpressionsprofil der cALL-Zellen finden sich einige differentiell exprimierte Gene, die im Zusammenhang von Apoptose und Antiapoptose zu sehen sind. Die Abbildung 4-6 soll einen Überblick über beteiligte Proteine geben, wobei deren genaue Funktion nur angedeutet werden kann und deren Rolle für die Pathologie der leukämischen Blasten nicht immer ganz klar ist.

Einer der Signaltransduktionswege zum Auslösen der Apoptose geht von  $\text{TNF}\alpha$  aus. Der Tumornekrosefaktor alpha ist ein multifunktionelles proinflammatorisches Enzym, das den Lipidstoffwechsel, die Blutkoagulation oder den Insulinstoffwechsel beeinflusst. Das  $\text{TNF}\alpha$  Signal wird über Rezeptoren und Kofaktoren wie TRADD und RIP intrazellulär an Mitglieder der TRAF-Familie weitergeleitet. Die Übersicht zeigt einige der an diesem Weg beteiligte Kinasen, die allerdings in den leukämischen Zellen nicht exprimiert werden: MAP3K1 (MEKK1), MAP3K5 (ASK1) und MAP4K1 (HPK1). Die fehlende Expression dieser Kinasen in den Leukämiezellen kann nur als indirekter Hinweis dafür dienen, dass diese Apoptosewege nicht beschränkt werden oder unterbrochen sind. Das in der cALL heraufregulierte TIAF1 schließlich vermag in den Zellen der cALL über Inhibition der TNFR-Adaptoren TRADD, FADD oder RIP in die  $\text{TNF}\alpha$ -vermittelte Apoptose einzugreifen (Chang, 1998). Eine Überexpression von TIAF1

in den leukämischen Zellen und ein damit verbundener anti-apoptotischer Effekt ist bislang nicht beschrieben worden.



**Abbildung 4-6 An Apoptose oder Antiapoptose beteiligte Faktoren in Zellen der cALL.**

Verschiedene Genprodukte sind in den leukämischen Zellen differentiell exprimiert und an der Apoptose oder deren aktiven Verhinderung beteiligt. Einige Proteine, die zur Apoptose benötigt werden, sind in den Leukämien herunterreguliert (▼). Heraufreguliert sind hingegen Proteine, die antiapoptotisch wirken (▲). Signifikant differentiell exprimierte Gene sind grau hinterlegt, die mit \* gekennzeichneten Gene sind speziell in den Leukämien mit TEL/AML1-Translokation identifiziert worden. † An der Regulation der Apoptose sind verschiedene Mitglieder der TRAF-Familie beteiligt. TRAF3 ist als Mitglied dieser Proteinfamilie differentiell exprimiert. Die einzelnen Genbezeichnungen werden im Text erläutert.

Zwei heraufregulierte Gene der Leukämiezellen kodieren Mitglieder der IAP-Familie von Inhibitoren der Apoptose. BIRC5 (API4, Survivin) und BIRC7 (Livin) sind beide im Kern lokalisiert. Für BIRC5 sind Funktionen im Zellzyklus beschrieben worden. BIRC5 kontrolliert am G2/M-Übergang, assoziiert mit den Microtubuli des Spindelapparates und ist gleichzeitig ein Inhibitor der Apoptose (Altieri, 1999; Kasof, 2001a). Es konnte weiterhin gezeigt werden, dass die Überexpression eine Rolle in der Antiapoptose von Krebszellen spielt (Altieri, 1999; Ambrosini, 1997). Zur Inhibierung der Apoptose interagiert BIRC5 mit den beiden Caspasen CASP3 und CASP7. Die Überexpression in den

Leukämiezellen des ubiquitär exprimierten BIRC5 lässt eine Fehlregulation der Expression in den Krebszellen vermuten.

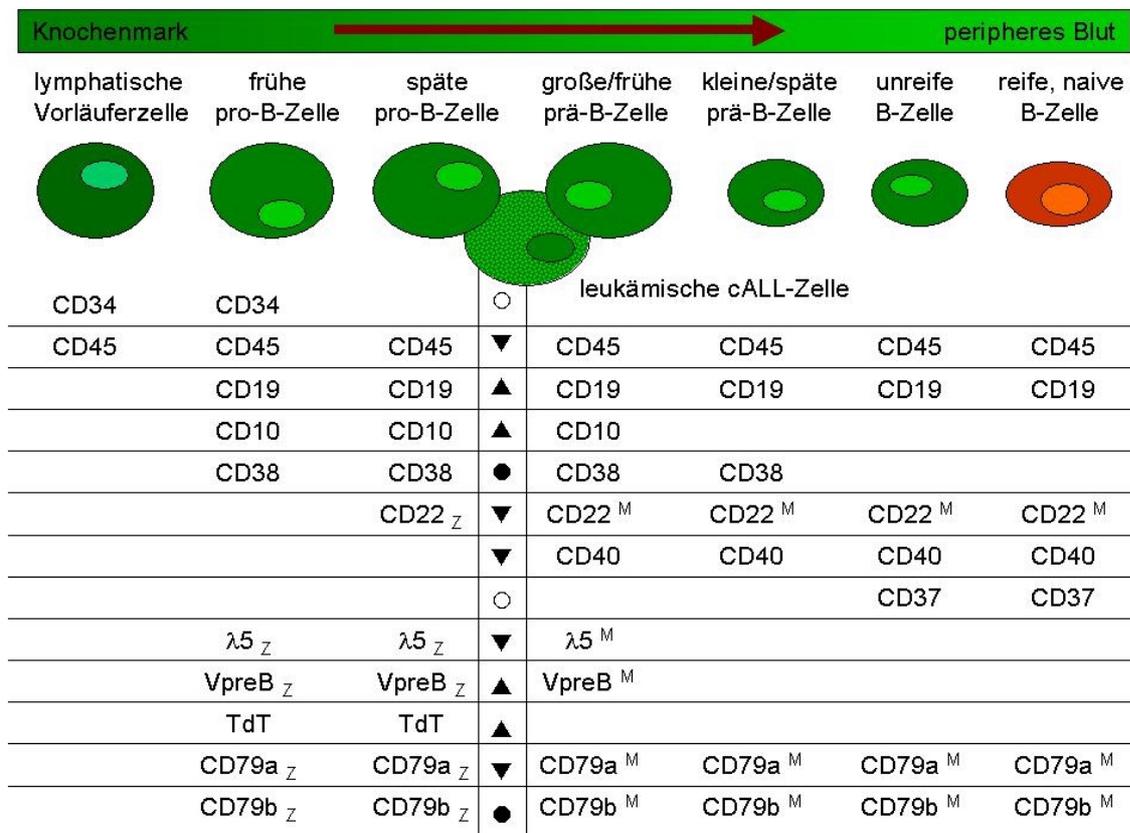
Zusätzlich zu BIRC5 ist in den Zellen mit TEL/AML1-Translokation BIRC7 heraufreguliert. Eine der antiapoptotischen Eigenschaften von BIRC7 geht auf die Hemmung des Rezeptors DR6 zurück (Kasof, 2001a; Kasof, 2001b). DR6 (TNFRSF21) ist Mitglied der TNF-Rezeptorfamilie. Eine Untergruppe dieser Familie bilden die sogenannten „Death“-Rezeptoren (DR). Die Aktivierung dieser Rezeptoren führt zur Einleitung der Apoptose. DR6 ist über die direkte Interaktion mit den Caspasen CASP3 und CASP7 an Apoptose beteiligt (Kasof, 2001a). Interessanterweise ist DR6 gerade in den Zellen mit einer TEL/AML1-Translokation ebenfalls heraufreguliert. Dies legt die Vermutung nahe, BIRC7 unterdrücke in diesen Zellen die DR6 vermittelte Apoptose direkt.

Für die Apoptose ist die Aktivität der mitochondrialen Atmungskette erforderlich. Nach Induktion der Apoptose durch verschiedene Stimuli steigt die Genexpression der Gene für Proteine der Atmungskette sehr rasch an. Eines dieser induzierten Gene ist Cytochrom C (CYTC), das Caspasen aktiviert und somit die Apoptose fördert (Chandra, 2002). Cytochrom C ist in den leukämischen Zellen deutlich vermindert exprimiert, was eine fehlende Apoptoseinduktion andeutet. Das Metaxin MTX2, ein Protein der äußeren Mitochondrienmembran mit Funktion im Proteinimport in die Mitochondrien, ist in den Zellen der cALL ebenfalls herunterreguliert und wird aber für die TNF-vermittelte Apoptose benötigt (Wang, 2001). Die leukämischen Zellen scheinen keine Apoptose zu machen. Die Daten deuten die Inhibition der Apoptose durch die leukämischen Zellen an und zeigen die verschiedenen Überlebensstrategien der Krebszellen auf.

#### *4.9 Die cALL ist eine Erkrankung von B-Vorläuferzellen*

Die Genprodukte mit einer Spezifität für einzelne B-Zellreifungsstadien lassen sich in drei Gruppen einteilen. Die erste Gruppe repräsentiert Gene, die in den sehr frühen Stadien der Reifungsreihe exprimiert sind. Zu diesen Proteinen gehören RAG1, RAG2, TdT,  $\lambda 5$  und VpreB. Diese Proteine spielen eine bedeutende Rolle in der frühen B-Zellentwicklung, werden für die Umlagerung der Immunglobulingene benötigt oder bilden Bestandteile des prä-B-Zellrezeptors. Die zweite Gruppe repräsentiert Proteine, die in allen Stadien der B-Zellreifung präsent sind, mit Ausnahme der endgültig differenzierten antikörperproduzierenden Plasmazellen. Zu diesen Proteinen gehören die Transmembranproteine  $Ig\alpha$ ,  $Ig\beta$ , CD40, CD72 und CD19. Weiterhin gehören in diese Gruppe zytoplasmatische Proteine wie die SRC-Familien-Proteine BTK oder BLK sowie nukleäre Proteine wie der B-Zell-spezifische Aktivator BSAP (PAX5). Die letzte Gruppe umfasst Proteine, die nicht auf reifen, ruhenden B-Zellen vorhanden sind,

sondern spezifisch nach der Aktivierung exprimiert werden. Zu diesen Proteinen gehören Syndecan und Blimp-1.



**Abbildung 4-7: Stellung der cALL Blasten innerhalb der B-Zellreifungsreihe**

Aus dem Genexpressionsprofil der leukämischen Blasten wird ersichtlich, dass diese in ihrem Differenzierungsstadium zwischen der späten pro-B- und der großen prä-B-Zelle stehen. Heraufregulierte Gene ▲, herunterregulierte Gene ▼, in cALL und Kontrollen gleichermaßen exprimiert ●, in keiner Probe nachweisbar ○.

Die leukämischen Blasten der cALL werden, nicht zuletzt aufgrund ihres Expressionsmusters entwicklungsrelevanter Oberflächenantigene, als unreife B-Zellen betrachtet (vgl. Abbildung 4-7). Diese Oberflächenantigene werden zur durchflusszytometrischen Diagnose der cALL herangezogen. In den Genexpressionsdaten findet sich dieses stadienspezifische Profil wieder. Das Oberflächenantigen aller B-Zellen, CD19, ist auch auf den Zellen der Leukämie zu finden. Die Expression von CD10, einem Marker der pro-B- und prä-B-Zellen, ist sogar deutlich erhöht. Ebenso sind CD38 und CD45 als Transkripte nachzuweisen, wengleich letzteres in den leukämischen Zellen deutlich erniedrigt ist. Die verringerte Expression von CD45 in leukämischen Blasten ist bereits untersucht (Inaba, 2000). Zwischen klinischen Parametern und einer fehlenden Expression von CD45 besteht keine erkennbare Korrelation. Aber in Kindern mit einer akuten B-Zell Leukämie konnte gezeigt werden, dass eine verringerte CD45-Expression mit einer erhöhten Expression

von CD34, CD10, TdT und CD24 einhergeht (Ratei, 1998). Tendentiell lässt sich dieser Befund auf Ebene der mRNA an den Genexpressionsdaten nachvollziehen. Zusammengenommen mit der reduzierten, aber nachweisbaren Expression von CD22, CD37 und CD40, lässt das den Schluss zu, bei den cALL-Blasten handelt es sich um maligne entartete Zellen im Stadium der späten pro-B- oder der großen prä-B-Zelle.

Dabei ist zum einen zu bemerken, dass die Kontrollen wegen der gewählten Separationsstrategie nicht ausschließlich aus CD19/CD10 doppelt-positiven B-Vorläuferzellen bestehen. Zum anderen ist diese Zellpopulation im gesunden Menschen nicht genau einem Reifungsstadium zuzuordnen. Die gleichzeitige Expression von CD19 und CD10 ist ein Charakteristikum der frühen pro-B-Zelle, der späten pro-B-Zelle und der großen prä-B-Zelle. Bei dieser „Auflösung“ ist eine exaktere Zuordnung der leukämischen Blasten zu einem Reifungsstadium nicht möglich. Zum anderen ist die Frage zu stellen, ob die leukämischen Blasten der cALL eine wirklich homogene Population darstellen, die exakt einem Differenzierungsstadium entspricht.

Die Reifung von der frühen pro-B-Zelle bis zur unreifen B-Zelle findet antigenunabhängig statt und ist nicht zuletzt der Bildung des Immunoglobulinrepertoires gewidmet. Die hierfür erforderlichen Gensegmentumlagerungen sind streng koordiniert. In diesem Prozess gibt es mindestens zwei Kontrollpunkte. Der erste ist abhängig vom prä-B-Zellrezeptor und liegt am Übergang von der großen zur kleinen prä-B-Zelle. Der zweite Kontrollpunkt liegt im Stadium der unreifen B-Zelle und ist dann vom B-Zellrezeptor abhängig. Die Umlagerung von Immunglobulingenen ist ein essentieller Bestandteil der B-Zellreifung, der schließlich zur großen Antikörperdiversität des menschlichen Immunsystems führt (Janeway, 2001; Melchers, 1998). Diese Rearrangements beginnen mit der Umlagerung der Segmente für die variablen Bereiche der schweren Ketten der Immunglobuline in der frühen pro-B-Zelle. Im Stadium der großen prä-B-Zelle werden nach erfolgreicher Umlagerung die weiteren Rearrangements gestoppt, und die Zellen beginnen sich zu teilen. Dies geschieht, nachdem die schwere Kette  $\mu$ H, bestehend aus ihrem konstanten (IGHM) und dem umgelagerten, variablen Teil, als Bestandteil des prä-B-Zellrezeptors transient an der Zelloberfläche exprimiert wird. Dieser prä-B-Zellrezeptor setzt sich aus der schweren Kette  $\mu$ H und zwei Proteinen zusammen, die zusammen die Surrogate Light Chain (SLC) der prä-B-Zellen bilden. Diese beiden Proteine sind IGLL1 ( $\lambda$ 5), vergleichbar dem konstanten Anteil der leichten  $\lambda$ -Kette, und VpreB mit einer ähnlichen Struktur wie der variable Bereich der leichten  $\lambda$ -Kette. Die Eigenschaften zur Signaltransduktion dieses Rezeptorkomplexes sind abhängig von der Assoziation mit zwei weiteren Oberflächenmolekülen, CD79a (Ig $\alpha$ ) und CD79b (Ig $\beta$ ). In seiner Struktur gleicht der prä-

B-Zellrezeptor dem späteren B-Zellrezeptor reifer Lymphozyten. Der prä-B-Zellrezeptor wird nur transient und in geringen Mengen an der Oberfläche der prä-B-Zellen exprimiert. Wenngleich der oder die Liganden nicht bekannt sind, ist die korrekte Formation des Rezeptors ein wichtiger Kontrollpunkt in der B-Zellreifung. Eine essentielle Funktion des prä-B-Zellrezeptors ist die Induktion der proliferativen Expansion derjenigen prä-B-Zellen, die eine intakte  $\mu$ H-Kette exprimieren können (Hess, 2001). Im späteren Stadium der kleinen prä-B-Zelle verschwindet der prä-B-Zellrezeptor mit der schweren Kette  $\mu$ H vollständig wieder von der Zelloberfläche, und die Umlagerungen der Gensegmente für die variablen Regionen der leichten Ketten  $\kappa$  und  $\lambda$  setzen ein. Auf der reifen B-Zelle, dem nachfolgenden Stadium, wird dann der reife B-Zellrezeptor mit der schweren Kette  $\mu$ H und der leichten Kette  $\kappa$  oder  $\lambda$  an der Oberfläche exprimiert. Hinzutreten die rezeptorassoziierten  $Ig\alpha$  und  $Ig\beta$  (Melchers, 2000).

In den cALL-Blasten sind Transkripte für alle Bestandteile des prä-B-Zellrezeptors nachzuweisen, allerdings nur mit einer deutlich verringerten Expression im Vergleich zu den gesunden Zellen. Das gilt für den konstanten Teil der schweren Kette (IGHM) und für IGLL1 ( $\lambda$ 5) der SLC und für die assoziierten  $Ig\alpha$  und  $Ig\beta$ . VpreB hingegen zeigt als einzige Komponente eine ganz deutliche Überexpression (vgl. Abbildung 4-7). Diese Überexpression von VpreB ist bekanntermaßen beschränkt auf maligne Zellen, die sich von B-Vorläuferzellen ableiten lassen (Bauer, 1991). Weiterhin sind eine ganze Reihe an Gensegmenten der variablen Regionen der leichten  $\kappa$  und  $\lambda$  Ketten nicht exprimiert, was andeutet, dass die Gensegmentumlagerungen noch nicht stattgefunden haben. Wenn aber die Blasten als entartete B-Zellen des pro-B- oder prä-B-Stadiums zu sehen sind, ist damit nicht zu rechnen, da die Umlagerung der Segmente für die leichten Ketten erst zu einem späteren Zeitpunkt stattfindet. Für die Umlagerung der Gensegmente von schweren und leichten Ketten der Immunglobulingene werden spezielle Rekombinasen, RAG1-RAG2 und TdT benötigt. Für den RAG1-RAG2-Komplex zeigen die Leukämiezellen keinen Unterschied im Vergleich zu den Kontrollen. Die Expression von TdT ist deutlich erhöht. Dieser Befund ist für die cALL bekannt, weswegen TdT bereits zur Diagnose eingesetzt wird. Aus den Genexpressionsdaten ist nicht abzulesen, ob die von den Transkripten exprimierten Bestandteile des prä-B-Zellrezeptors als Protein an der Oberfläche nachzuweisen sind. Beide Komponenten der SLC sowie die schwere IGHM sind bereits in den pro-B-Zellen intrazellulär nachweisbar, bevor der prä-B-Zellrezeptor mit den akzessorischen Immunglobulinen  $Ig\alpha$  und  $Ig\beta$  auf den nachfolgenden prä-B-Zellen nachzuweisen ist (Wang, 2002). Fasst man diese Ergebnisse zusammen, ergibt sich ein Bild, nach dem die leukämischen Zellen der cALL von ihrem Entwicklungsstadium her der späten pro-

B-Zelle entsprechen. Inwieweit die Expression des prä-B-Zellrezeptors vom gesunden Menschen abweicht, ist schwer zu entscheiden, nachdem gezeigt wurde, dass es nicht nur CD19+/SLC+/IGHM+ prä-B-Zellen, sondern auch CD19+/SLC+/IGHM- prä-B-Zellen zu geben scheint (Lemmers, 1999). Der für die Funktion des prä-B-Zellrezeptors erforderliche membranständige Partner, bei fehlendem IGHM müsste es eine „surrogate H chain“ geben, ist im Menschen bislang nicht identifiziert worden.

Verschiedene Transkriptionsfaktoren nehmen Schlüsselpositionen in der B-Zellreifungsreihe ein. Zu den wichtigsten sind E2A, EBF, PAX5, SOX4 und SYK zu zählen (Melchers, 1998) sowie ein Kofaktor der Transkription, OAZ. Wird E2A in Mäusen deletiert, entwickeln sich keine Lymphozyten. Das Gen E2A kodiert zwei Transkriptionsfaktoren mit unterschiedlichen basic-Helix-Loop-Helix-Domänen (bHLH). Diese beiden Proteine E47 und E12 sind das Resultat alternativen Spleißens. E2A ist in verschiedene Differenzierungsprozesse involviert, so zum Beispiel in die Myogenese oder Neurogenese. Allein in der Entwicklung der B-Lymphozyten ist E2A ein essentieller Faktor (Bain, 1994; Zhuang, 1994). Es gibt Hinweise darauf, dass die Steuerung der B-Zellreifung ein E2A-Dosis-abhängiger Prozess ist (Hanrahan, 2000). Die höchste Expression der E2A Proteine ist in pro-B-Zellen zu verzeichnen. Die Expression von E2A in den Leukämiezellen zeigt eine gewisse Überexpression. EBF zeigt ebenso wenig einen Unterschied wie die Expression von PAX5, das die Expression einer Reihe B-Zell-spezifischer Moleküle wie CD19, VpreB oder  $\lambda 5$  steuern kann. OAZ und SOX4 sind in den Leukämiezellen signifikant überexprimiert. Welchen Effekt dies auf die Reifung der B-Zellen haben könnte, bleibt unklar. Die Deletion dieser beider Transkriptionsfaktoren führt in der Maus jedenfalls zu einer gestörten B-Zellreifung. E2A als Transkriptionsfaktor steuert mit dem Kofaktor OAZ unter anderem die Transkription von VpreB,  $\lambda 5$ , und CD79a (Gisler, 2002; Persson, 1998; Sigvardsson, 2000). Die beiden letztgenannten sind allerdings beide nur schwach exprimiert. Hier deuten sich Fehlregulationen der cALL-Zellen an. Die Deletion von SYK in murinen Zellen führt zu Fehlsteuerungen in der B-Zellentwicklung. Allerdings fällt die Entwicklung der B-Lymphozyten in diesen Tieren nicht völlig aus. Die Zahl reifer B-Zellen ist stark vermindert. SYK ist in zwölf der mit HG-U133A untersuchten cALL-Proben nicht exprimiert in den anderen Zellen ist die Expression deutlich schwächer als in den Kontrollen.

Eine weitere Eigenschaft der pro-B-Zelle ist die Expression von KIT, dem Rezeptor für das Chemokin SCF (Stem Cell Factor). Neben IL3 und IL7 vermittelt SCF ein Proliferationssignal an normale pro-B-Zellen. Der IL7-Rezeptor zeigt keine Unterschiede zwischen den untersuchten Zellgruppen. Mit dem Übergang in das nächste Reifungsstadium, der großen prä-B-Zelle, geht die Expression von KIT und

TdT wieder verloren. Fasst man diese Ergebnisse zusammen, ergibt sich das Bild der Leukämiezellen, die mit ihrem Differenzierungsstadium den späten pro-B-Zellen ähneln.

Die Biologie gesunder, ruhender B-Lymphozyten und deren Veränderungen während der Aktivierung sowie die Frage nach Toleranz und Selbst-Toleranz ist in einigen Arbeiten an Mausmodellen mit Hilfe von Genechips untersucht worden (Glynne, 2000a; Glynne, 2000b). Mit Hilfe der SAGE-Methode wurde die Genexpression CD34+ hämatopoetischer Stammzellen mit CD19+/CD10+ prä-B-Zellen verglichen (Muschen, 2002). Mit Ausnahme der Topoisomerase TOP2A sind in den leukämischen Proben keine Gene exprimiert, die nach den Ergebnissen von Muschen den CD34+ Stammzellen zuzuordnen wären. Einige der Gene, die im Vergleich zu den CD34+ Stammzellen für prä-B-Zellen charakteristisch sind, zeigen in den hier untersuchten Leukämien Abweichungen vom Expressionsmuster gesunder B-Lymphozyten. Zu erwähnen sind unter anderem die Bestandteile des prä-B-Zellrezeptors, VpreB, IGLL1, IGMH, CD79a und CD79b oder die für die B-Zellreifung wichtigen Transkriptionsfaktoren E2A, POU2AF1 (OBF1) und POU2F2 (OCT2). Der hier durchgeführte Vergleich unterstreicht nochmals, in welchen Genen sich die Expression in leukämischen Blasten von der gesunden Zelle unterscheidet.

An murinen B-Zellen des Knochenmarks wurden eingehend die Veränderungen der Genexpression während der normalen B-Zellreifung in fünf aufeinander folgenden Stadien untersucht (Hoffmann, 2002). Die Reifungsreihe beginnt in der Maus mit der pro-B-Zelle (entspricht der frühen pro-B-Zelle im Menschen, vgl. Abbildung 1-3) und verläuft dann in entsprechenden Schritten wie beim Menschen über die prä-B-1-Zelle (späte pro-B-Zelle), die große prä-B-2-Zelle (große prä-B-Zelle), die kleine prä-B-2-Zelle (kleine prä-B-Zelle), die unreife B-Zelle zur reifen B-Zelle. Für fünf dieser Stadien, ausgenommen wurde nur die frühe pro-B-Zelle, wurden Genexpressionsprofile ermittelt und die Expression dieser stadienspezifischen Gene im Verlauf der gesamten Reifungsreihe dargestellt. Das ermöglicht den Vergleich mit den in den Leukämiezellen als differentiell exprimiert detektierten Genen und ebenso die Zuordnung der leukämischen Zellen zu einem Reifungsstadium. Eine exakte Übertragung der mit Mauszellen erzielten Ergebnisse auf menschliche Zellen ist wegen der Spezies-Unterschiede nicht ohne weiteres möglich. Dennoch liefert dieser Vergleich wertvolle Hinweise.

Es werden zwölf Gene benannt, die stadienspezifische Funktionen der murinen prä-B1-Zellen charakterisieren. Drei dieser Gene sind weder in den leukämischen Zellen noch in den Kontrollen exprimiert (IL3R, TBXA2R und TIAM1). Vier der Gene sind in den cALL-Blasten differentiell exprimiert. Die fehlende Expression von NCF1 (SOC2), FCER1G und KIT deutet an, dass cALL-Blasten nicht diesem sehr frühen

Reifungsstadium zuzuordnen sind. Lediglich CD105 (ENG) ist in den Blasten ebenfalls wie in den murinen prä-B1-Zellen überexprimiert. Dieses Membranglykoprotein ist Bestandteil des TGF $\beta$ -Rezeptorkomplexes und ist bereits auf aktivierten Monozyten, Erythroblasten und auf Lymphoblasten pädiatrischer Leukämien nachgewiesen worden. Die Liste der charakteristisch exprimierten Gene der murinen großen prä-B-2-Zellen ist mit 36 Transkripten wesentlich umfangreicher. Im vorliegenden Datensatz wurden davon zehn Gene detektiert, die eine Überexpression in den Leukämiezellen zeigen. Neben vier verschiedenen Histongenen sind das NEK2 und PCNA, mit Funktionen in der Mitose oder bei der Replikation, sowie STK18, CIT und MELK. Es wurden im Vergleich mit dieser Studie keine Gene gefunden, die typisch für große prä-B-2-Zellen sind und in den Leukämien herunterreguliert sind. Die wichtigste Veränderung der Zellen beim Schritt zum nächsten Entwicklungsstadium, der murinen kleinen prä-B-2-Zelle, ist der Zellzyklus-Arrest. Die kleinen prä-B-Zellen proliferieren nicht mehr. Dementsprechend sind die in diesem Stadium identifizierten Gene vorrangig an der Regulation dieser Funktion beteiligt. In den leukämischen Zellen konnten mit diesem Datensatz keine Parallelen gefunden werden. Es ist nicht zu entscheiden, ob das mit dem Reifungsstadium der Blasten oder deren ungesteuerter Proliferation als Krebszelle zusammenhängt.

Die letzte Gruppe der identifizierten Gene ist dazu geeignet, unreife von reifen B-Lymphozyten zu unterscheiden. Darüber hinaus zeigen diese Gene in allen davor liegenden B-Zellreifungsstadien eine geringe Expression. Unter diesen Genen befinden sich neun Gene, die in den leukämischen Zellen durchgängig nicht exprimiert sind. Neben den Oberflächenantigenen CD23, CD32, CD72, CR2, BLR1 und IFNGR2 sind das TANK und die Kinase MAP3K8. Die fehlende Expression dieser Gene in den Blasten ist möglicherweise keine spezifische Eigenschaft der leukämischen Zellen, sondern geht vielleicht auf die Kontrollproben zurück, die zwar angereichert, aber eben nicht nur aus B-Vorläuferzellen bestehen. Der Anteil der CD19+/CD10- B-Zellen macht sich hier bemerkbar.

Insgesamt bestätigen die Vergleiche mit dieser Studie sowie die in diesem Kapitel bereits dargestellten Ergebnisse, dass die Blasten der cALL als maligne entartete Zellen im Stadium der späten pro-B-Zellen zu betrachten sind. Die Leukämiezellen sind nicht in der Lage, sich weiter zu differenzieren. Stattdessen proliferieren die Zellen und prägen damit das Krankheitsbild.

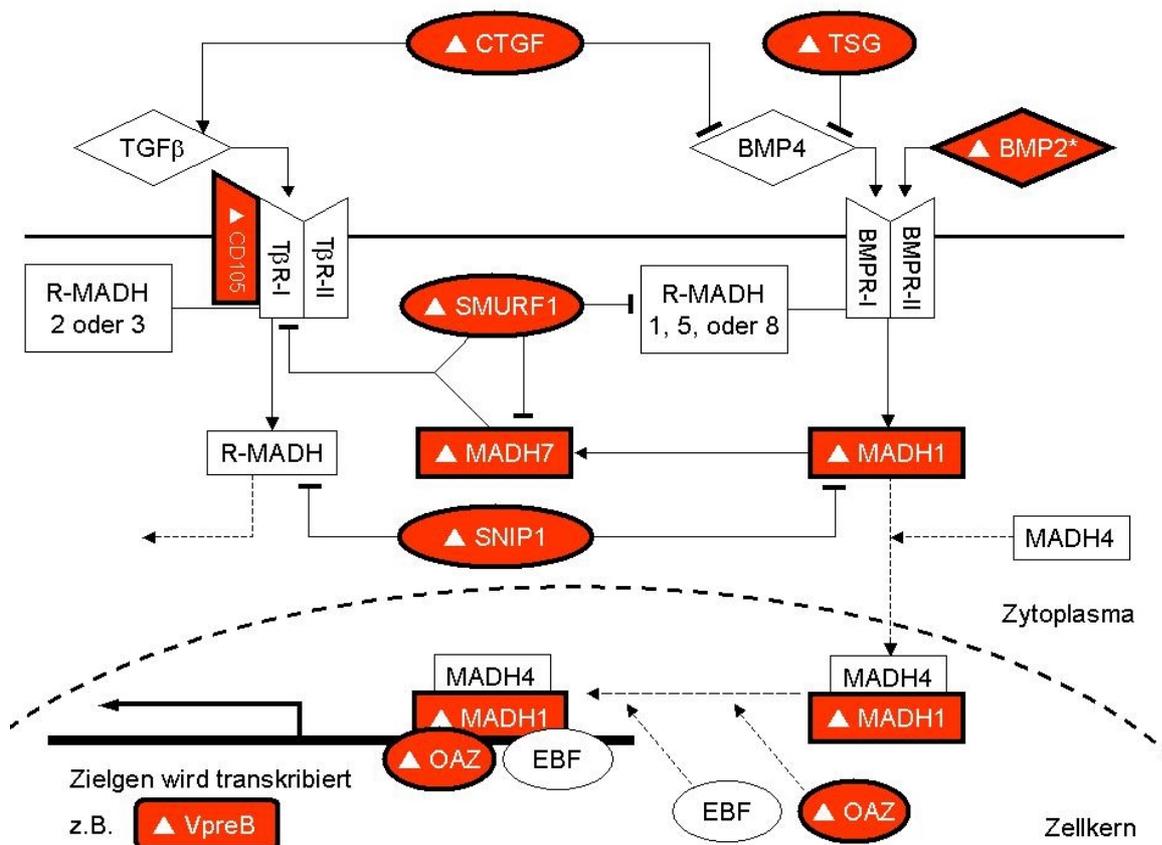
#### *4.10 TGF $\beta$ /BMP und WNT, kooperative Signalwege leukämischer Zellen*

Ein definiertes Netzwerk von Signaltransduktionswegen spielt in der Regulation zellulärer Vorgänge in den leukämischen Blasten eine herausgehobene Rolle. Die

Betrachtung der an Signaltransduktion beteiligten Gene mit einer differentiellen Expression in den Zellen zeigt eine intensive Beteiligung dieser Genprodukte an TGF $\beta$ /BMP- und an WNT-Signalwegen. Diese beiden Signalwege greifen zudem an verschiedenen Stellen ineinander. TGF $\beta$  ist ein multifunktionelles Protein mit Funktionen in der Kontrolle von Proliferation, Differenzierung, Apoptose, Modulation von Immunreaktionen und einer Vielzahl weiterer Aufgaben in verschiedenen Zelltypen, so zum Beispiel in der Entwicklung von Knochen, Knorpel, Zähnen, Herz und Nervenzellen (Heldin, 1997). In der Embryonalentwicklung der Bilateria wird die dorsoventrale Achse durch die Aktivität von BMP-Familienmitgliedern und ihren Antagonisten bestimmt. Ein BMP2-orthologes Protein konnte, wie zuvor bereits ein MADH-orthologes Protein, in der Embryonalentwicklung der riffbildenden Koralle *Acrophora millepora* nachgewiesen werden (Hayward, 2002). Hieran ist das entwicklungsgeschichtlich hohe Alter des Signalwegs ablesbar. Die Familienmitglieder BMP2 und BMP4 werden zudem im Menschen auch im Zahn exprimiert und sind Mediatoren der Knochen- und Zahnentwicklung (Li, 1998). In diesem Zusammenhang ist vielleicht die eigentümliche Überexpression von TUFTIP in den Zellen der Leukämie zu sehen, das eigentlich in der Regulation der Bildung von Zahnschmelz eine Rolle spielt. Welchen regulatorischen Effekt das TGF $\beta$ -Signal auf die Zellen hat, hängt vom zellulären Kontext ab. In einem einzigen Wundheilungsprozess kann es die Proliferation stimulieren oder inhibieren, je nachdem ob die Zielzelle ein Fibroblast oder ein Keratinozyt ist (Ashcroft, 1999). Der WNT-Signalweg, dessen Bezeichnung auf das Segmentpolaritätsgen *wg* (Wingless) in *Drosophila melanogaster* zurückgeht, ist ebenfalls an der embryonalen Achsenausbildung in der Entwicklung von Vertebraten beteiligt. In Mäusen konnte die Aktivität dieses Signalwegs bereits während der frühesten Manifestation der anterior-posterioren Achse im Mesoderm gezeigt werden. Neben Funktionen in der Embryonalentwicklung übernehmen beide Signalwege die Steuerung weiterer zellulärer Vorgänge, so in Zellen des blutbildenden und immunologischen Systems.

Die Abbildung 4-8 gibt einen vereinfachten Überblick über die am TGF $\beta$ /BMP-Signalweg beteiligten Mitspieler, beschränkt sich dabei aber auf die wichtigsten Komponenten, die insbesondere in hämatopoetischen Zellen beteiligt sind und in der cALL differentiell exprimiert sind. Die von außen an die Zellen herankommenden Signale, zum Beispiel TGF $\beta$  oder BMP4, werden über selektive Rezeptoren in das Innere der Zellen weitergeleitet. Neben verschiedenen Serin/Threonin-Kinasen sind Mitglieder der MADH-Familie (SMAD) im Innern der Zelle an der Weiterleitung des Signals beteiligt. Hier werden regulatorische, rezeptoraktivierte R-MADHs (z.B.

MADH1) und inhibitorische I-MADHs (MADH7) unterschieden, sowie Co-MADHs (MADH4). Die regulatorischen Familienmitglieder sind rezeptor- und damit signalabhängig. Die Proteine MADH2 und MADH3 sind spezifisch für TGF $\beta$ -Signale, MADH1, MADH5 und MADH8 sind spezifisch für BMP-Signale (Heldin, 1997). Im Kern wird dann über weitere DNA-bindende Kofaktoren, FAST für TGF $\beta$ -Signale und OAZ für BMP-Signale, und komplezierte Transkriptionsfaktoren, zum Beispiel EBF, die Expression der Zielgene gesteuert.



**Abbildung 4-8 Der TGF $\beta$ /BMP-Signaltransduktionsweg**

Die Abbildung gibt einen Ausschnitt des umfangreichen TGF $\beta$ /BMP-Signalwegs wieder mit den in den Leukämien differentiell exprimierten Genen. Signifikant differentiell exprimierte Gene sind grau hinterlegt, ▲ in den Leukämien heraufregulierte Gene, \* speziell in den Leukämien mit TEL/AML1-Translokation identifizierte Gene. Durchgezogene Linien stehen für Induktion bzw. Inhibition; gepunktete Linien bedeuten Assoziation von Proteinen bzw. deren Wanderung in den Zellkern. T $\beta$ R-I/-II und BMPR-I/-II stehen für die beiden spezifischen Rezeptorkomplexe der TGF $\beta$ - und BMP-Signalwege. Die übrigen Genbezeichnungen werden im Text erläutert.

Die Überexpression einiger Komponenten dieser Signalwege deutet an, dass diese in den Leukämiezellen eine besondere Rolle spielen. Als extrazellulär wirkende Proteine werden von den Blasten CTGF und TSG gebildet. CTGF (Connective Tissue Growth Factor), welches zum Beispiel in die Angiogenese oder die Wundheilung involviert ist, ist in der Lage, durch Bindung an die Zielmoleküle die Aktivität der Signale

von TGF $\beta$  positiv und von BMP4 negativ zu beeinflussen (Abreu, 2002). CTGF zeigt eine höhere Affinität zu BMP4 und verhindert die Bindung von BMP4 an seinen Rezeptor. Die Bindung von CTGF an TGF $\beta$  hingegen scheint das Signal zu verstärken. TSG (Twisted Gastrulation) wurde zuerst in *D. melanogaster* als weiterer Regulator der Embryonalentwicklung und als Mitglied der BMP-Familie identifiziert und hemmt die Wirkung von BMP4 spezifisch (Chang, 2001). Aus den Expressionsdaten ist nicht ersichtlich, ob die Leukämiezellen mit Bildung dieser beiden Peptide ihre eigene Funktion selbst steuern, aber ein autoregulativer Mechanismus ist denkbar. CD105 (ENG, Endoglin) ist Bestandteil des TGF $\beta$ -Rezeptorkomplexes und bindet TGF $\beta$  mit hoher Affinität. Bei der Untersuchung der B-Lymphozytenentwicklungsreihe wurde CD105 als charakteristisch für prä-B-Zellen beschrieben (Hoffmann, 2002).

MADH1 ist ein regulatorisches MADH-Mitglied, das in Verbindung mit dem nicht differentiell exprimierten, aber in den Proben nachweisbaren MADH4 die Expression der Zielgene steuert. MADH1 ist spezifisch für BMP-vermittelte Signale. Als inhibitorisches Familienmitglied im TGF $\beta$  Signalweg ist MADH7 heraufreguliert. MADH7 kann durch TGF $\beta$  induziert werden, assoziiert mit dem Typ I TGF-beta Rezeptor (T $\beta$ R-I) und ist damit ein Antagonist im TGF $\beta$ -Signalweg (Hayashi, 1997; Nakao, 1997). BMP2 ist besonders in den leukämischen Blasten mit TEL/AML1-Translokation überexprimiert. Als extrazelluläres Signalmolekül kann es über MADH1 regulatorisch auf MADH7 einwirken und moduliert möglicherweise so den gesamten Signalweg. Das inhibitorische MADH7 ist somit ein Effektor von Rückkopplungssignalen in diesem Signalweg. Das SMURF1 Genprodukt interagiert selektiv mit den durch Rezeptoren regulierten R-MADH-Proteinen, die spezifisch für den BMP-Signalweg sind. Die Wirkung von SMURF1 liegt in der Ubiquitinierung dieser R-MADHs, deren sich anschließender Degradation und Inaktivierung (Zhu, 1999). SMURF1 greift auch in den TGF $\beta$ -Signalweg ein. In COS7 Zellen der Meerkatze wurde nachgewiesen, dass SMURF1 mit MADH7 interagiert und dessen Ubiquitinierung induziert. Weiterhin assoziiert SMURF1 über MADH7 mit T $\beta$ R-I und beschleunigt damit den Umsatz und die Degradation von MADH7 und T $\beta$ R-I (Ebisawa, 2001). SNIP1 ist ein weiterer inhibitorischer Mediator dieser Signalwege. Es kann MADH1, MADH2 und MADH4 binden und die durch TGF $\beta$  und BMP vermittelte Expression von Zielgenen unterbinden (Kim, 2000). Zusammengenommen deuten diese Ergebnisse die Deregulation der BMP-Signalwege eventuell zugunsten der TGF $\beta$ -Signalwege an.

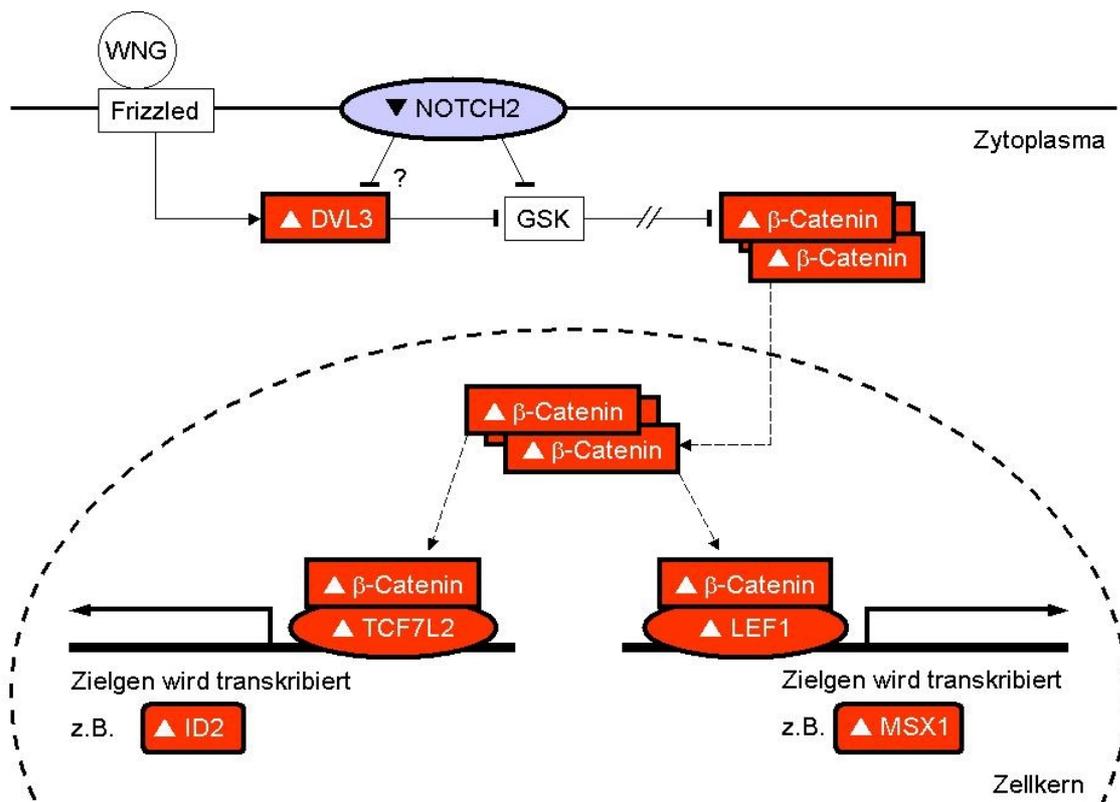
In neuronalen Progenitorzellen konnte gezeigt werden, wie BMP2, vermittelt über MADH1, und LIF (Leukämie Inhibierender Faktor), vermittelt über STAT3, synergistisch die Differenzierung dieser Zellen zu Astrozyten fördern (Nakashima, 1999). Ein

entsprechendes kooperatives Differenzierungssignal ist auch für hämatopoetische Vorläuferzellen vorstellbar. Unter Berücksichtigung einer möglichen Unterdrückung von BMP-Signalwegen und der deutlich verringerten Expression von STAT3 in den Leukämiezellen, ist ein Fehlen dieses Differenzierungssignals als Bestandteil der Deregulation zellulärer Vorgänge in den cALL Blasten durchaus denkbar.

Als letztes Mitglied dieses Signaltransduktionswegs ist der in den Leukämien heraufregulierte Kofaktor OAZ zu nennen. OAZ ist ein DNA-bindender Kofaktor, der mit MADH1 und MADH4 assoziiert und damit für den BMP-Signalweg spezifisch ist (Massague, 2000). Als Koaktivator der Transkription, der schließlich im Komplex aus MADH1, MADH4 und OAZ aktiv werden kann, scheint exklusiv EBF zu dienen (Tsai, 1998). EBF, das in den leukämischen Zellen nachweisbar aber nicht differentiell exprimiert ist, reguliert unter anderem die Expression von Bestandteilen des prä-B-Zellrezeptors, VpreB, IGLL1, Ig $\alpha$  und Ig $\beta$  (Gisler, 2000). Damit zeigt sich die direkte Verbindung des BMP-Signalwegs zur B-Zellentwicklung und aus den Expressionsdaten deutet sich weiterhin eine Deregulation dieses Signalwegs an.

Als zweiter intrazellulärer Signaltransduktionsweg, der offensichtlich in den leukämischen Zellen benutzt wird, ist der WNT-Signaltransduktionsweg zu nennen. WNT-Proteine gehören einer Familie sekretorischer Glykoproteine an, die in die Regulation von Entwicklungsprozessen verschiedener Organismen involviert sind. Der Prototyp dieser Familie ist *wg* (Wingless) aus *D. melanogaster*, ein Segmentpolaritätsgen der Embryogenese. In Abwesenheit des extrazellulären Signalmoleküls der WNT-Familie phosphoryliert die Kinase GSK3 konstitutiv  $\beta$ -Catenine (CTNNB), essentielle Mediatoren dieses Signalwegs. Phosphorylierte Catenine werden degradiert und akkumulieren nicht in den Zellen. Die anzusteuern den Zielgene dieses Signalwegs werden nicht transkribiert. Bindet WNT jedoch an den membranständige Rezeptor FZD (Frizzled), rekrutiert dieser intrazellulär den Inhibitor DVL (Dishevelled). Dieser Kinaseinhibitor blockiert die Aktivität von GSK3 und unphosphoryliertes  $\beta$ -Catenin akkumuliert im Zytosol, wandert in den Kern und bindet dort an Transkriptionsfaktoren wie TCF oder LEF. Die Transkription der Zielgene wird somit initiiert (Dale, 1998). Die Abbildung 4-9 gibt den Zustand des aktivierten Signaltransduktionswegs wieder wie er vermutlich in den leukämischen Zellen vorliegt. Die Genexpressionsanalyse hat gezeigt, dass eine Reihe essentieller Mitglieder dieses Wegs in ihrer Expression hochreguliert sind: DVL3 (ein Protein der Dishevelled-Familie), CTNNB1 (ein  $\beta$ -Catenin) und die beiden wichtigsten Transkriptionsfaktoren, die in diesem System angesteuert werden, TCF7L2 (TCF4) und LEF1. Für Mitglieder der NOTCH Familie gibt es Hinweise, dass sie durch Inhibition von DVL und GSK

regulativ auf den WNT-Signalweg Einfluss nehmen. NOTCH2 ist in den Leukämiezellen herunterreguliert und inhibiert den Signalweg vermutlich nicht. Das Catenin degradierende SIAH1 scheint ebenfalls modulierend in WNT Signalwege einzugreifen. Es wird vermutlich durch p53 aktiviert und kann dann als Proteinkomplex, bestehend aus SIAH1, SIP und EBI,  $\beta$ -Catenin degradieren. SIAH ist in den Zellen der Leukämie stark exprimiert. Für die beiden anderen Komponenten des Komplexes gibt es keine Expressionsunterschiede. Somit bleibt offen, in wie weit der Proteinkomplex regulativ aktiv ist. Die beiden wichtigsten Transkriptionsfaktoren, die durch den WNT-Signalweg angesprochen werden, LEF1 und TCF7L2, sind hochreguliert. Neben den Transkriptionsfaktoren sind zwei Gene in ihrer Expression heraufreguliert, die gerade durch diese Transkriptionsfaktoren gesteuert werden: ID2 und MSX1. Die Bedeutung der relativen Überexpression von ID2 und MSX1 für die Pathologie der cALL wird im folgenden Kapitel diskutiert. Die Aktivierung des WNT-Signalwegs geht in den cALL Blasten jedoch nicht mit einer Aktivierung von MYC einher, einem der prominentesten Zielmoleküle von  $\beta$ -Catenin/TCF.



**Abbildung 4-9 Der WNT-Signaltransduktionsweg**

Die Abbildung gibt einen Ausschnitt des WNT-Signalwegs mit den daran beteiligten und in den Leukämien differentiell exprimierten Genen wieder. Die inhibierende Funktion des NOTCH2 auf speziell DVL3 ist experimentell nicht belegt. Signifikant differentiell exprimierte Gene sind grau hinterlegt, ▲ in den Leukämien heraufregulierte Gene, ▼ in den Leukämien herunterregulierte Gene. Durchgezogene Linien stehen für Induktion bzw. Inhibition; gepunktete Linien bedeuten Assoziation von Proteinen bzw. deren Wanderung in den Zellkern.

Im WNT-Signalweg geschieht die Aktivierung von ID2 über TCF7L2 (TCF4). Zudem ist TCF7L2 in der Lage, LEF1 zu aktivieren. Hiermit handelt es sich um eine positive Rückkopplungsschleife innerhalb der WNT-Signaltransduktion (Hovanes, 2001). Es mehren sich die Hinweise auf eine kooperative Signaltransduktion der WNT- und der TGF $\beta$ /BMP-Signalwege. Mindestens drei Beispiele sind beschrieben, in denen diese beiden Signaltransduktionswege an einem Zielmolekül zusammenlaufen und synergistisch wirken können. WNT3A und BMP4 können, jeder für sich über die getrennten Signalwege, die Expression von MSX1 heraufregulieren. In Kooperation der beiden Signalwege kommt es zu einer weiteren Steigerung der MSX1 Expression (Willert, 2002) in den untersuchten embryonalen Karzinomzellen. MSX1 ist in den hier untersuchten Leukämiezellen verstärkt exprimiert. Ein ähnliches Modell wird in *Xenopus laevis* beschrieben. Über den Weg TGF $\beta$   $\rightarrow$  MADH3/MADH4  $\rightarrow$  LEF1 wird die Expression des Xtnw Gens aktiviert. Über den Weg WNT  $\rightarrow$   $\beta$ -Catenin  $\rightarrow$  LEF1 kann die Expression ebenfalls aktiviert werden. Die Kooperation beider Signalwege verstärkt die Xtnw Expression weiter (Labbe, 2000). Für ID2 gibt es unabhängige Beschreibungen. In Mastdarntumoren wurde die Aktivierung von ID2 durch WNT beschrieben (Rockman, 2001). Die Aktivierung von ID2 durch BMP4 ist in hämatopoetischen Stammzellen nachgewiesen worden (Hollnagel, 1999). Ein direkter Nachweis einer kooperativen Steuerung von ID2 in einem Zelltyp steht aus. Allerdings ist die Expression von ID2 in den leukämischen Blasten der cALL deutlich erhöht.

Diese Ergebnisse zeigen eindrücklich die Bevorzugung von TGF $\beta$ /BMP- und WNT-Signaltransduktionswegen in den leukämischen Blasten der cALL und deuten deren Bedeutung für die Pathologie der cALL an, die darin liegen könnte, dass die Expression B-Zell-spezifischer Gene verändert ist.

#### 4.11 Proliferation oder Differenzierung?

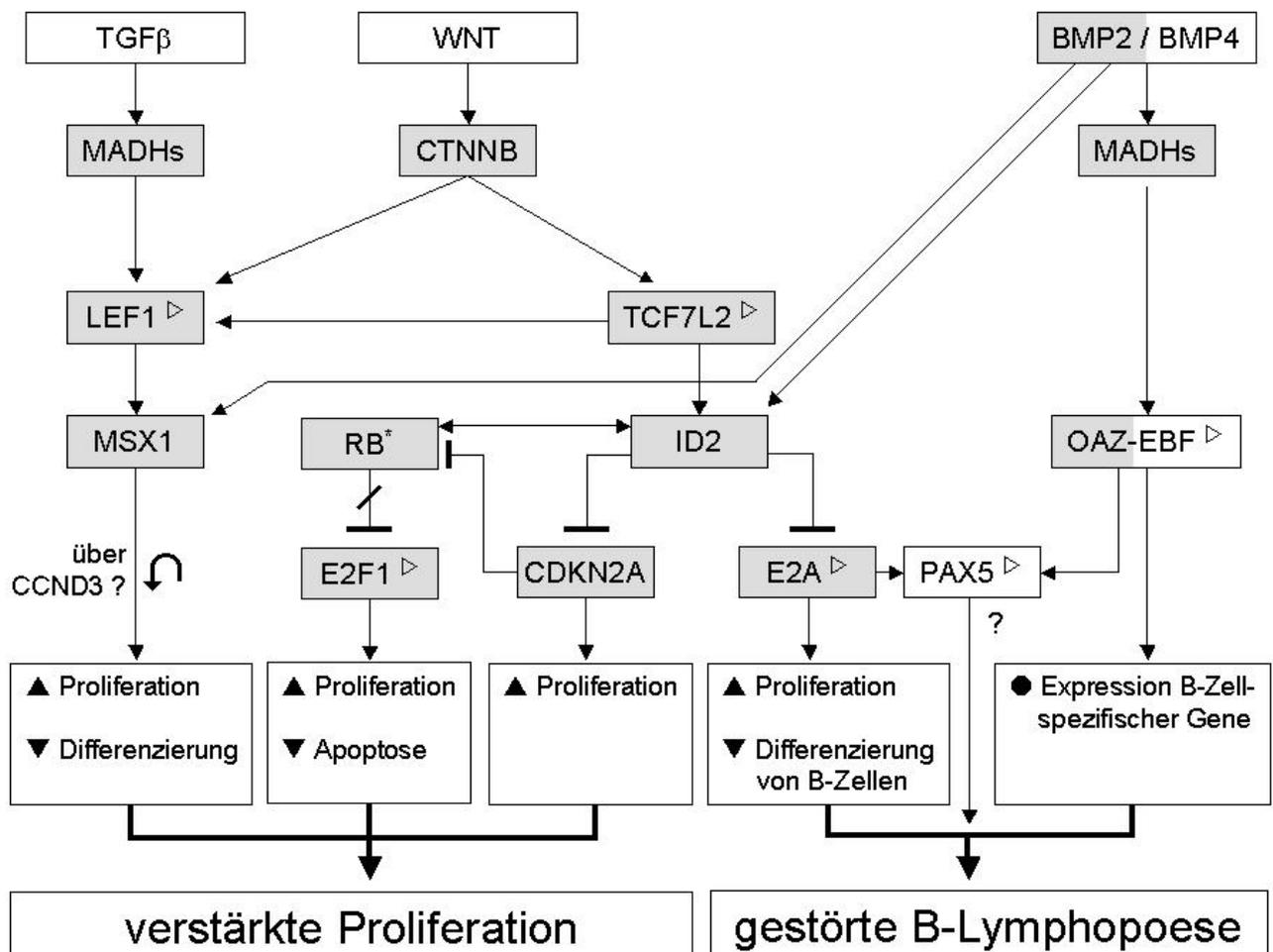
Die erhobenen Genexpressionsdaten vermitteln ein definiertes Bild der leukämischen Blasten, in dem sich ihre verschiedenen Eigenschaften widerspiegeln. Der Vergleich der Genexpressionsprofile von Leukämiezellen mit gesunden B-Lymphozyten zeigt die leukämischen Blasten als Zellen mit einer hohen Proliferationsrate einhergehend mit verstärkter DNA-Replikation und einer nicht verstärkten Proteinbiosynthese. Das Differenzierungsstadium, in dem die Reifung der leukämischen Zellen angehalten wurde, entspricht, nach Auswertung des Genexpressionsprofile, dem Reifungsstadium der späten pro-B-Zelle gesunder B-Vorläuferzellen. Um möglichen Abwehrmechanismen des Körpers auszuweichen, verfolgen die Leukämiezellen eigene Strategien zur Vermeidung der Apoptose, indem zum Beispiel Apoptoseinhibitoren wie

TIAF1 oder BIRC5 und BIRC7 heraufreguliert werden. In den Zellen der cALL werden weiterhin insbesondere zwei Signalwege benutzt, die eng miteinander vernetzt sind und die in die Regulation von Proliferation und Differenzierung, hier besonders die Reifung von B-Lymphozyten, involviert sind. Die Genexpressionsmuster zeigen eine bemerkenswerte Bevorzugung der Signalwege, an denen Mitglieder der TGF $\beta$ /BPM-Familie und der WNT-Familie beteiligt sind.

Mit Hilfe von Microarrays konnte in embryonalen Karzinomzellen kürzlich nachgewiesen werden, dass MSX1 (HOX7) und ID2 zwei bedeutende WNT-Zielgene sind (Willert, 2002). Die Expression dieser Gene ist in den Zellen der cALL deutlich erhöht. Zudem konnte bereits gezeigt werden, wie MSX1 und ID2 auch über BMP2 oder BMP4 aktiviert werden können (Hollnagel, 1999). Im WNT-Signalweg verläuft die Aktivierung von MSX1 über LEF1. Aber es kann zu einer kooperativen Aktivierung von LEF1 über WNT- und TGF $\beta$ -Signalwege kommen (Labbe, 2000). Ohne experimentellen Beleg ist diese kooperative Aktivierung von MSX1 in den Leukämiezellen bislang eine naheliegende Vermutung. MSX1 ist ein entwicklungsrelevantes Gen des Menschen, dessen Mutation zu Fehlbildungen am Gesichtsschädel und an den Zähnen führt. Nach einem Regulationsmodell, das in der Maus entwickelt wurde, greift MSX1 steuernd in den Zellzyklus ein. MSX1 scheint die Expression des Cyclins D1 indirekt hochzuregulieren. Dies führt dazu, dass der Zellzyklus nicht verlassen werden kann, die Zellen weiter proliferieren, nicht die G0-Phase erreichen und in der Folge sich nicht differenzieren können. Dies ist ein Bild, wie es die entarteten Leukämiezellen geben (Hu, 2001). Einschränkend ist jedoch zu erwähnen, dass in B-Zellen das Cyclin D3 (CCND3) an die Stelle von D1 tritt (Fink, 2001). CCND3 ist in allen untersuchten leukämischen und gesunden Proben nachzuweisen. Ein ähnliches Steuerungsmodell mit CCND3, in dem MSX1 die Proliferation fördert und die Differenzierung hemmt, ist für die Leukämie vorstellbar. Aus den Genexpressionsdaten ist allerdings nicht ablesbar, welche die Zielproteine einer MSX1-Aktivität sind.

Als weiteres Zielmolekül des WNT-Signalwegs wurde ID2 identifiziert (Willert, 2002). ID2 reguliert Proliferation und Differenzierung. Eine deregulierte Expression von ID2 ist an der malignen Transformation von Zellen beteiligt. ID2 ist in der Lage, den antiproliferativen Effekt von Tumorsuppressoren der Retinoblastomfamilie (RB) aufzuheben und ermöglicht damit das Fortschreiten des Zellzyklus' (Lasorella, 2000). Die Bindung von ID2 an RB führt zur Expression von Genen, die speziell für den Fortschritt der S-Phase des Zellzyklus' benötigt werden. In den leukämischen Zellen ist das möglicherweise nicht der Wirkmechanismus von ID2, da RB eine verringerte Expression in den Leukämiezellen aufweist. Möglicherweise inhibiert ID2 in den leukämischen Blasten CDKN2A. Dieser Kinase-Inhibitor CDKN2A ist in der Lage, Zellen

in der späten G1-Phase zu arretieren und somit die Zellteilung zu regulieren. Inaktivierende Mutationen im Gen für CDKN2A scheinen an der Entstehung von Tumoren beteiligt zu sein. In den Zellen der cALL wird CDKN2A stärker exprimiert als in den Kontrollen. Dennoch kommt es nicht zur Arretierung des Zellzyklus'. Dafür ist möglicherweise ID2 verantwortlich, welches in der Lage ist, den CDKN2A vermittelten Zellzyklusarrest vollständig aufzuheben (Lasorella, 1996). Eine weitere Funktion von ID2 ist die Hemmung des Transkriptionsfaktors E2A, der essentiell für die Differenzierung von B-Lymphozyten ist (Bain, 1994; Zhuang, 1994). Die Hemmung von E2A durch Überexpression von ID2 führt zur Inhibition der Differenzierung von B-Zellen (Becker-Herman, 2002) und zur Förderung der Proliferation (Rockman, 2001).



**Abbildung 4-10 Modell zur Fehlregulation von Differenzierung und Proliferation in der cALL.**

Über verschiedene Signaltransduktionswege wird in den leukämischen Blasten die Proliferation gefördert, die weitere Differenzierung jedoch verhindert. Signifikant differentiell exprimierte Gene oder Ausschnitte eines Signalwegs sind grau hinterlegt. ▲ in den Leukämien wahrscheinlich induzierter Prozess, ▼ in den Leukämien vermutlich inhibierter Prozess, ● wahrscheinliche Fehlregulation, ▷ Transkriptionsfaktor, ⊕ vermutlich beteiligter Faktor, \* herunterreguliert.

Wie bereits oben erwähnt greift das Retinoblastomprotein RB über verschiedene Mechanismen in die Zellteilung ein. Neben der Regulation über ID2 bestehen in der gesunden Zelle auch Wechselwirkungen zwischen Mitgliedern der RB-Familie und den Transkriptionsfaktoren der E2F-Familie. Diese Interaktion ist für die Entscheidung des Zellschicksals in Richtung Apoptose oder Proliferation von entscheidender Bedeutung. Derzeit ist jedoch nicht geklärt, über welche regulatorischen Netzwerke innerhalb der Zelle diese Entscheidung zustande kommt (Gorgoulis, 2002). Die Bindung von RB an E2F unterdrückt die Expression der E2F-Zielgene am Übergang von G1- zur S-Phase des Zellzyklus'. In den Blasten der cALL ist die Expression von RB und dem verwandten RBL2 vermindert, von E2F1 jedoch erhöht. Der Verlust der verschiedenen RB-Familienmitglieder führt zum Verlust der G1-Kontrolle im Zellzyklus und zu andauernden Zellteilungen (Dannenberg, 2000; Sage, 2000). Die fehlende Expression speziell von RBL2 in Retinoblastomen geht einher mit einer reduzierten Apoptose und einer verminderten Differenzierung der Zellen (Bellan, 2002). Andererseits führt die Überexpression von E2F1 in Lungenkarzinomen dazu, dass die Tumorzellen in die Proliferation getrieben werden (Gorgoulis, 2002). In zwei verschiedenen Zellkulturmodellen (synchronisierte murine Fibroblasten und humane Zellen eines Osteosarkoms) und unter Verwendung von Genechips wurden zudem Zielgene der Kontrolle durch E2F1 identifiziert (Ishida, 2001; Muller, 2001). Durch E2F1 werden demnach besonders Gene der Proliferation hochreguliert, zum Beispiel BMP2, BUB1, TOP2A, CDKN2C oder CDC2. In den Zellen der cALL scheint der Verlust der Kontrolle durch RB und die erhöhte Expression von E2F1 genau diese Gene positiv anzusteuern und die Zellen in die Proliferation zu treiben.

Drei Transkriptionsfaktoren sind für die frühe B-Zellentwicklung essentiell. Zu Beginn der B-Lymphopoese aktivieren die Transkriptionsfaktoren E2A und EBF koordiniert die Expression B-Zell-spezifischer Gene. Nachfolgend übernimmt PAX5 (BSAP) die Aufgabe, nicht-B-Zell-spezifische Gene in ihrer Expression zu reprimieren. Alle drei Transkriptionsfaktoren sind damit an der Festlegung der Lineage-Spezifität beteiligt (O'Riordan, 1999). PAX5 wird weiterhin für das Fortschreiten der Differenzierung über das pro-B-Zellstadium hinaus benötigt (Busslinger, 2000). Die Expression von PAX5 wird durch E2A und EBF reguliert (O'Riordan, 1999). Es ist in den hier untersuchten leukämischen Proben anwesend, aber leicht herunterreguliert, wenngleich der Unterschied mit beiden Genechiptypen und jeweils zwei Sondensätzen detektiert nicht signifikant ist. Die Fehlregulation von E2A und EBF könnte das durch PAX5 vermittelte Signal zur weiteren Differenzierung stören. Die Aktivität von E2A könnte in den leukämischen Blasten durch die Aktivität des ID2 inhibiert werden. EBF wird über den BMP-Signalweg gesteuert. Die Genexpressionsdaten deuten eine Fehlregulation

dieses Signalwegs an. Die Überexpression von E2A kann an der Onkogenese direkt beteiligt sein. In einem Mausmodell führt die viral induzierte Überexpression von E2A zur Entstehung eines T-Zell-Lymphoms (Mikkers, 2002). Die Abbildung 4-10 gibt ein mögliches Modell, wie die beteiligten Faktoren auf Proliferation und Differenzierung der maligne entarteten Zellen Einfluss nehmen können. Die Frage nach den Ursachen dieser Fehlregulation bleibt dabei unbeantwortet. Zusammengefasst zeigt das Genexpressionsprofil der leukämischen Blasten auf, wie in diesen Zellen die weitere Differenzierung angehalten wird und wie die Zellen in einem Zustand dauernder Zellteilung gehalten werden können.

## 5 Zusammenfassung

Die akute lymphoblastische Leukämie vom Typ common ALL (cALL), bei der B-Lymphozyten des hämatopoetischen Systems maligne entartet sind, ist die häufigste Krebserkrankung des Kindesalters. Im Knochenmark beginnen die unreifen B-Vorläuferzellen zu proliferieren ohne sich weiter zu differenzieren. Die leukämischen Blasten der cALL exprimieren auf ihrer Zelloberfläche die Antigene CD19 und CD10. Damit sind sie gesunden B-Vorläuferzellen im Stadium der späten pro-B-Zelle oder der großen prä-B-Zelle immunphänotypisch homolog, die als gesunde Referenzzellen herangezogen werden können. Unter Verwendung von „DNA high-density Oligonucleotide Arrays“ wird in der vorliegenden Arbeit zum ersten Mal das Genexpressionsprofil leukämischer Zellen der cALL mit gesunden Zellen eines entsprechenden Differenzierungsstadiums in einem Ansatz der funktionellen Genomik analysiert. Das vorrangige Ziel dabei ist die Aufklärung krankheitsspezifischer Eigenschaften auf molekularer Ebene der mRNA, die einen grundlegenden Einblick in kausale Zusammenhänge zwischen der Genexpression und der Biologie beziehungsweise der Pathologie der leukämischen Zellen zulassen. Im Vergleich der Genexpressionsprofile von 35 pädiatrischen cALL Proben mit 15 gesunden Kontrollen aus Nabelschnurrestblut separierten B-Vorläuferzellen wurden über 1.500 differentiell exprimierte Gene identifiziert und analysiert.

Verschiedene klinische Parameter wie die initiale Blastenzahl im Blut, das Ansprechen auf die Therapie oder das Alter der Patienten bei Diagnosestellung lassen eine Einteilung der Patienten in Gruppen zu. Allerdings lassen sich nur für die Patientengruppen „jünger als 10 Jahre bei Diagnose“ und „älter als 10 Jahre“ spezifische Genexpressionsprofile identifizieren. Weiterhin zeigt sich, dass leukämische Blasten mit einer  $t(12;21)(p13;q22)$  Translokation ein spezifisches Genexpressionsprofil besitzen, in dem sie sich von den übrigen cALL Proben, auch im Vergleich zu den gesunden Kontrollen, unterscheiden.

Für alle untersuchten leukämischen Proben gilt, in ihrem Differenzierungsstadium stehen die Blasten der cALL zwischen später pro-B- und großer prä-B-Zelle. Charakteristische Marker wie CD10, CD19 und CD38 lassen sich in allen Proben nachweisen. Marker unreifer Zellen (CD34) oder reifer B-Zellen (CD37) lassen sich weder in den Leukämien noch in den Kontrollen zeigen. Bestandteile des prä-B-Zellrezeptors sind unterschiedlich differentiell exprimiert.

Die leukämischen Blasten sind Zellen mit einer hohen Proliferationsrate. Entsprechend finden sich überexprimierte Gene, die an der Kontrolle des Zellzyklus beteiligt sind (z.B. CDC2, CDC6 oder E2F1) oder die in die Replikation von DNA

involviert sind (z.B. TOP2A oder POLE). Gene der Proteinsynthese hingegen sind häufig herunterreguliert (z.B. SLU7 oder ADARB1).

Apoptose wird in den leukämischen Zellen vermutlich aktiv unterbunden. Essentielle Gene der Apoptose wie zum Beispiel die Caspasen 9 oder 1 sind nicht exprimiert. Hingegen sind Inhibitoren der Apoptose (z.B. TIAF1, BIRC5 oder BIRC7) stark exprimiert. Zwei intrazelluläre Signaltransduktionswege spielen in den leukämischen Blasten eine herausgehobene Rolle. Der aktivierte WNT/ $\beta$ -Catenin-Signalweg führt über die ebenfalls hochregulierten Transkriptionsfaktoren LEF1 und TCF7L2 zur Expression von ID2 und MSX1, die beide an der Kontrolle von Proliferation und Differenzierung beteiligt sind. MSX1 fördert vermutlich über das B-Zell-spezifische Cyclin D3 (CCND3) die Proliferation der Zellen. Die Hemmung von E2A durch Überexpression von ID2 führt zur Inhibition der Differenzierung von B-Zellen und zur Förderung der Proliferation. Als weiteren Mechanismus kann ID2 den CDKN2A vermittelten Zellzyklusarrest vollständig aufheben und somit Proliferation fördern. Über den ebenfalls aktivierten BMP/MADH-Signalweg wird mittels der Transkriptionsfaktoren OAZ und EBF die Transkription B-Zell-spezifischer Gene, z.B. VpreB heraufreguliert. Die Inhibition von E2A durch ID2 kann in diesem Zusammenhang auch eine Unterdrückung von Differenzierungssignalen bedeuten.

Das Genexpressionsprofil der cALL-Blasten zeigt eine deutliche Fehlregulation im Vergleich zu den gesunden Kontrollzellen besonders in Genen, deren Produkte die Zellen eindeutig in Richtung Proliferation steuern und dabei die Differenzierung der unreifen B-Zellen arretieren.

## 6 Anhang

## 6.1 Tabellen

Tabelle 6-1 Übersicht der verwendeten cALL-Proben

Name	Quelle	Geschlecht	% Blasten im KM	Alter	BCR-ABL	MLL-X	TEL-AM1/1	RISIKO	Leuk-0	Leuk-8	Blast-15
ALL01OPS	KM	m	>90%	7,0	0	0	0	2	20600	2600	1,50%
ALL02RTS	KM	w	>95%	8,5	0	0	0	3	498000	211000	7,50%
ALL15IEB	KM	w	>89%	12,11	0	0	1	2	3400	1200	0,00%
ALL18YEM	KM	w	>64%	8,5	0	0	1	2	2400	3300	n.b.
ALL19UEN	KM	w	>40%	4,9	0	0	0	1	5800	5000	n.b.
ALL22AUT	KM	w	>90%	15,8	0	0	0	2	5900	7500	2,00%
ALL25HCS	KM	m	>90%	4,4	0	0	0	1	50000	700	n.b.
ALL28DEM	KM	m	>90%	3,11	0	0	0	2	32000	1600	3,00%
ALL29NIR	KM	w	>90%	14,2	0	0	0	2	28500	600	7,50%
ALL30MOH	KM	w	>90%	3,5	0	0	0	2	13900	1900	2,50%
ALL31ARB	KM	w	>90%	7,0	0	0	0	2	77400	3200	6,00%
ALL32HCS	KM	m	>90%	4,0	0	0	0	2	2900	2100	n.b.
ALL33LUM	KM	w	>97%	2,5	0	0	0	2	2100	3500	0,50%
ALL34REH	KM	w	>97%	5,1	0	0	1	1	38800	4700	12,00%
ALL36IHP	KM	m	>98%	2,9	0	0	0	2	11600	3100	0,00%
ALL37UEH	PB	w	>99%	0,6	0	1	0	3	n.b.	n.b.	n.b.
ALL38IER	KM	m	>90%	4,3	0	0	0	2	8600	1400	1,50%
ALL39AHS	KM	m	>90%	4,4	0	0	1	1	22600	1700	0,50%
ALL40EEL	KM	m	>90%	3,4	0	0	0	1	33900	2900	1,00%
ALL41APS	KM	m	>90%	4,3	0	0	0	2	6500	1800	n.b.
ALL42ORK	KM	m	>90%	5,7	0	0	0	1	2800	2200	n.b.
ALL43MED	KM	w	>90%	3,0	0	0	1	2	3700	4500	3,50%
ALL45NUK	KM	w	>90%	11,8	0	0	0	2	2000	6400	n.b.
ALL46ROY	KM	w	>90%	2,1	0	0	0	1	87200	3700	n.b.
ALL47HEM	KM	m	>90%	5,6	0	0	0	3	14500	8600	64,50%
ALL48PIL	KM	m	>90%	9,9	0	0	0	2	134000	2700	82,50%
ALL49ERD	KM	m	>90%	11,0	0	0	0	3	16700	14700	50,50%
ALL50GAJ	KM	w	>90%	4,0	0	0	1	1	1700	1000	0,00%
ALL51ERT	KM	w	>90%	15,2	0	0	0	2	8900	2900	39,00%
ALL52HCS	KM	w	>90%	14,9	0	0	0	2	57500	1400	53,50%
ALL53TAB	KM	w	>90%	4,1	0	0	0	1	4800	6200	1,00%
ALL54TEY	KM	m	>90%	3,5	0	0	0	1	56700	2700	0,00%
ALL55ZAB	KM	w	>90%	8,11	0	0	1	1	2100	1800	0,00%
ALL56WET	KM	w	>90%	4,11	0	0	0	2	39000	4500	6,50%
ALL57SUN	KM	m	>90%	10,6	0	0	0	2	443400	1700	7,00%
ALL58GAT	LQ	m	>95%	n.b.	n.b.	n.b.	n.b.	n.b.	n.b.	n.b.	n.b.

KM: Knochenmark, PB: peripheres Blut, LQ: Liquor, m: männlich, w: weiblich, Alter bei Diagnose (Jahre, Monate), Risiko: Einteilung in Risikogruppe nach GPOH, Leuk: Anzahl der Leukozyten im PB am Tag 0 oder Tag 8 [Zellen/ $\mu$ l], Blast: Anzahl leukämischer Blasten im KM am Tag 15, n.b.: nicht bestimmt

Tabelle 6-2 Übersicht der verwendeten Kontrollproben

Name	Volumen in ml	Fraktion	% lebende Zellen	Zellen gesamt	% CF $\alpha$ + an lebenden Zellen	Zellen gesamt	% CF an lebenden Zelle	%CF an lebenden Zellen
CB*1NP	n.b.	Lymphozyten	96,78	9,60	7,01	0,67	2,98	92,99
		B-Zellen	93,10	n.b.	n.b.	n.b.	n.b.	n.b.
CB*2NP	n.b.	Lymphozyten	98,90	9,30	10,56	0,98	3,72	89,43
		B-Zellen	93,78	1,50	80,75	1,21	n.b.	19,25
CB19NP	70,00	Lymphozyten	n.b.	16,20	n.b.	n.b.	n.b.	n.b.
		B-Zellen	97,51	1,90	70,09	1,33	15,37	28,98
CB22NP	50,00	Lymphozyten	n.b.	8,25	n.b.	n.b.	n.b.	n.b.
		B-Zellen	96,26	0,77	84,33	0,65	26,30	15,66
CB23NP	75,00	Lymphozyten	n.b.	16,00	n.b.	n.b.	n.b.	n.b.
		B-Zellen	98,24	0,96	58,41	0,56	19,35	40,06
CB24NP	75,00	Lymphozyten	93,34	13,80	14,48	2,00	5,58	85,51
		B-Zellen	98,64	0,63	76,29	0,48	15,16	23,71
CB26NP	60,00	Lymphozyten	93,74	10,00	25,35	2,54	8,44	71,26
		B-Zellen	97,84	1,90	84,27	1,60	32,02	15,73
CB36NP	65,00	Lymphozyten	96,59	20,00	12,18	2,44	5,70	87,82
		B-Zellen	98,77	1,35	90,73	1,22	29,55	9,20
CB38NP	60,00	Lymphozyten	88,77	10,00	4,90	0,49	1,03	95,10
		B-Zellen	97,35	0,87	77,23	0,67	28,17	22,77
CB41NP	60,00	Lymphozyten	87,17	12,00	3,03	0,36	0,46	96,97
		B-Zellen	98,01	0,60	62,34	0,37	18,73	37,65
CB42NP	60,00	Lymphozyten	95,88	9,90	9,28	0,92	4,71	90,72
		B-Zellen	98,34	0,96	71,44	0,69	26,68	28,55
CB46NP	55,00	Lymphozyten	95,61	6,90	12,29	0,85	5,06	87,71
		B-Zellen	98,27	0,90	87,53	0,79	37,41	12,26
CB48NP	90,00	Lymphozyten	80,73	9,50	5,79	0,55	1,82	94,20
		B-Zellen	92,66	1,20	93,01	1,12	32,09	6,98
CB59NP	65,00	Lymphozyten	97,16	15,00	9,67	1,45	3,29	90,32
		B-Zellen	98,15	1,17	85,01	0,99	18,48	14,99
CB60NP	42,00	Lymphozyten	97,63	6,90	15,09	1,04	5,47	84,90
		B-Zellen	99,41	0,75	88,09	0,66	27,85	11,91
TS01NP	n.b.	Lymphozyten	82,29	27,00	62,77	16,95	20,88	37,23
		B-Zellen	91,52	0,38	91,35	0,35	62,16	8,65
TS02NP	n.b.	Lymphozyten	98,88	51,30	52,79	27,08	12,69	41,97
		B-Zellen	97,59	0,35	94,15	0,33	n.b.	5,85

Volumen: Blutvolumen aus Nabelschnur entnommen, Lymphozyten: nach Ficoll-Aufreinigung, B-Zellen: nach Selektion, n.b.: nicht bestimmt

Tabelle 6-3 Übersicht der Ergebnisse der cALL-Proben

Name	Zellzahl x 10E7	RNA µg	IVT µg	EOS-H % Gene detektiert	HG-U133A pre AB		HG-U133A post AB	
					% Gene detektiert	SF	% Gene detektiert	SF
ALL01OPS	nd	2,5	11,6	12,1				
ALL02RTS	nd	20,2	23,6	10,0				
ALL15 IEB	nd	27,0	7,2	5,5				
ALL18YEM	nd	36,0	4,4	1,5				
ALL19UEN	nd	78,0	19,3	25,7				
ALL22AUT	nd	7,8	13,2	22,7				
ALL25HCS	8,0	25,4	61,4	25,5	21,2	9,5	23,9	3,9
ALL28DEM	8,5	11,3	44,4	16,6	7,7	28,0	11,2	12,6
ALL29NIR	8,5	11,9	48,7	19,1	9,5	24,1	12,9	10,4
ALL30MOH	10,0	40,7	58,1	17,4	15,4	13,9	19,0	5,6
ALL31ARB	2,0	33,1	72,5	21,4	19,9	10,7	25,3	3,9
ALL32HCS	6,0	27,1	55,4	23,9	18,5	12,1	22,8	4,4
ALL33LUM	2,5	10,6	27,9	27,9	29,3	4,1	29,1	2,2
ALL34REH	2,0	12,2	34,4	29,7	26,9	5,0	26,3	2,6
ALL36IHP	3,0	10,2	24,6	27,2	25,5	5,4	26,1	2,7
ALL37UEH	4,0	21,2	60,4	32,7	35,9	3,0	36,9	1,4
ALL38IER	3,6	5,0	66,4		30,3	4,5	36,3	1,0
ALL39AHS	2,0	2,3	70,4		28,2	5,3	34,5	1,2
ALL40EEL	2,0	1,8	77,7		31,2	4,4	37,3	1,1
ALL41APS	2,0	3,6	72,7		33,8	3,7	38,6	1,0
ALL42ORK	1,0	3,0	47,1		31,6	5,3	36,1	1,5
ALL43MED	2,0	43,8	84,4		35,2	3,4	39,2	0,9
ALL45NUK	2,0	3,4	44,5		34,5	3,6	38,7	1,2
ALL46ROY	10,0	20,9	51,5		36,6	3,0	39,5	1,0
ALL47HEM	4,0	36,7	53,7		34,0	3,2	37,2	1,0
ALL48PIL	9,0	34,6	51,0		28,5	5,3	33,7	1,9
ALL49ERD	5,5	10,6	55,1		32,2	2,8	35,3	1,0
ALL50GAJ	2,4	3,1	77,0		32,9	3,5	36,2	1,2
ALL51ERT	2,0	5,6	83,5		27,3	4,7	30,1	1,5
ALL52SCH	10,0	32,1	76,9		33,8	3,9	37,1	1,3
ALL53TAB	3,0	80,1	63,3		31,0	3,6	33,5	1,4
ALL54TEY	4,0	23,9	72,1		36,5	3,1	38,4	1,1
ALL55ZAB	3,0	12,6	106,0		32,5	2,9	34,5	1,0
ALL56WET	10,0	87,4	67,5		40,0	3,9	43,7	0,9
ALL57SUN	4,0	16,7	50,2		24,4	5,3	28,0	1,6
ALL58GAT	0,5	2,5	63,5		43,8	24,9	nd	nd

pre AB: vor Antikörper-Verstärkung, post AB: nach Antikörper-Verstärkung, SF: Skalierungsfaktor

Tabelle 6-4 Übersicht der Ergebnisse der Kontroll-Proben

Name	Quelle	% B-Zellen	Zellzahl x10E7	RNA µg	IVT µg	EOS-H % Gene detektiert	HG-U133A pre AB		HG-U133A post AB	
							% Gene detektiert	SF	% Gene detektiert	SF
CB*1NP	CB	n.b.	n.b.	8,6	10,9	6,3				
CB*2NP	CB	80,75	1,21	11,5	4,8	13,7				
CB19NP	CB	70,09	1,33	19,7	61,1	20,6				
CB22NP	CB	84,33	0,65	11,9	62,8	17,3				
CB23NP	CB	58,41	0,56	11,9	62,2	13,3				
CB24NP	CB	76,29	0,48	13,8	11,8	7,1				
CB26NP	CB	84,27	1,60	40,3	84,5	19,2				
CB36NP	CB	90,79	1,23	106,1	54,3	18,7	5,2	42,0	9,9	16,7
CB38NP	CB	77,23	0,67	15,6	32,1	28,6	19,0	12,3	26,8	4,4
CB41NP	CB	62,34	0,37	16,8	37,7	26,7	27,0	7,0	32,8	2,4
CB42NP	CB	45,00	0,43	21,2	35,0	33,8	28,7	6,2	34,9	1,9
CB46NP	CB	87,52	0,79	9,7	60,4	28,8	33,8	3,4	36,6	1,3
CB48NP	CB	93,01	1,12	9,1	65,9	22,0	37,2	2,3	36,2	1,1
CB59NP	CB	85,01	0,99	17,4	77,8	32,8	34,2	3,3	38,7	0,9
CB60NP	CB	88,09	0,66	16,3	83,3	33,2	39,7	2,5	41,0	0,9
TS01NP	TS	91,35	0,35	34,2	12,9	18,5				
TS02NP	TS	94,51	0,33	23,4	17,9	20,6				
BM01NP	KM	n.b.	n.b.	32,3	53,4	18,5				
BM02NP	KM	n.b.	n.b.	20,1	47,8	12,5				
BM03NP	KM	n.b.	n.b.	28,5	46,5	20,2				
BM04NP	KM			EOS-Kontrolle		34,4				
PB01NP	PB			EOS-Kontrolle		29,1				
PB02NP	PB			EOS-Kontrolle		24,4				

CB: Nabelschnurblut, TS: Tonsille, KM: Knochenmark, PB: peripheres Blut, pre AB: vor Antikörper-Verstärkung, post AB: nach Antikörper-Verstärkung, SF: Skalierungsfaktor, n.b.: nicht bestimmt

Tabelle 6-5 Qualitätsmerkmale der untersuchten Genechips

Name	EOS-H % Gene	HG-U133A pre AB		HG-U133A post AB	
		% Gene	SF	% Gene	SF
CB*1NP	6,3				
CB*2NP	13,7				
CB19NP	20,6				
CB22NP	17,3				
CB23NP	13,3				
CB24NP	7,1				
CB26NP	19,2				
CB36NP	18,7	5,2	42,0	9,9	16,7
CB38NP	28,6	19,0	12,3	26,8	4,4
CB41NP	26,7	27,0	7,0	32,8	2,4
CB42NP	33,8	28,7	6,2	34,9	1,9
CB46NP	28,8	33,8	3,4	36,6	1,3
CB48NP	22,0	37,2	2,3	36,2	1,1
CB59NP	32,8	34,2	3,3	38,7	0,9
CB60NP	33,2	39,7	2,5	41,0	0,9
ALL01OPS	12,1				
ALL02RTS	10,0				
ALL15 IEB	5,5				
ALL18YEM	1,5				
ALL19UEN	25,7				
ALL22AUT	22,7				
ALL25HCS	25,5	21,2	9,5	23,9	3,9
ALL28DEM	16,6	7,7	28,0	11,2	12,6
ALL29NIR	19,1	9,5	24,1	12,9	10,4
ALL30MOH	17,4	15,4	13,9	19,0	5,6
ALL31ARB	21,4	19,9	10,7	25,3	3,9
ALL32HCS	23,9	18,5	12,1	22,8	4,4
ALL33LUM	27,9	29,3	4,1	29,1	2,2
ALL34REH	29,7	26,9	5,0	26,3	2,6
ALL36IHP	27,2	25,5	5,4	26,1	2,7
ALL37UEH	32,7	35,9	3,0	36,9	1,4
ALL38IER		30,3	4,5	36,3	1,0
ALL39AHS		28,2	5,3	34,5	1,2
ALL40EEL		31,2	4,4	37,3	1,1
ALL41APS		33,8	3,7	38,6	1,0
ALL42ORK		31,6	5,3	36,1	1,5
ALL43MED		35,2	3,4	39,2	0,9
ALL45NUK		34,5	3,6	38,7	1,2
ALL46ROY		36,6	3,0	39,5	1,0
ALL47HEM		34,0	3,2	37,2	1,0
ALL48PIL		28,5	5,3	33,7	1,9
ALL49ERD		32,2	2,8	35,3	1,0
ALL50GAJ		32,9	3,5	36,2	1,2
ALL51ERT		27,3	4,7	30,1	1,5
ALL52HCS		33,8	3,9	37,1	1,3
ALL53TAB		31,0	3,6	33,5	1,4
ALL54TEY		36,5	3,1	38,4	1,1
ALL55ZAB		32,5	2,9	34,5	1,0
ALL56WET		40,0	3,9	43,7	0,9
ALL57SUN		24,4	5,3	28,0	1,6
ALL58GAT		43,8	24,9	n.b.	n.b.
TS01NP	18,5				
TS02NP	20,6				
BM01NP	18,5				
BM02NP	12,5				
BM03NP	20,2				
BM04NP	34,4				
PB01NP	29,1				

% Gene: Anteil der detektierten Gene, pre AB: vor Antikörperverstärkung, post AB: nach Antikörperverstärkung, SF: Skalierungsfaktor, n.b.: nicht bestimmt. Die Probe ALL58GAT wurde nicht zur Identifikation differentiell exprimierter Gene herangezogen, sondern wurde als Kontrolle verwendet.

Von der Analyse ausgeschlossene Arraydaten sind grau hinterlegt.

## 6.2 Abkürzungsverzeichnis

A	absent (ein Gen ist nicht exprimiert)
AK	Antikörper
ALL	Akute Lymphoblastische Leukämie
ALL-BFM	ALL-Studie Berlin-Frankfurt-Münster
AML	Akute Myeloische Leukämie
B-CLL	B-Zell Chronisch Lymphozytäre Leukämie
BMP	Bone Morphogenetic Protein
BSA	Bovines Serum Albumin
cALL	common ALL
CALLA	Common Acute Lymphoblastic Leukemia Antigen
CB	Cord Blood (Nabelschnurrestblut)
CD	Cluster of Differentiation
cDNA	Complementary DNA (komplementäre DNA)
CLL	Chronisch Lymphatische Leukämie
dNTP	Desoxynukleosidtriphosphat
DMSO	Dimethylsulfoxid
DNA	Desoxyribonucleic Acid (Desoxyribonukleinsäure)
DTT	Dithiothreitol
EBV	Epstein-Barr Virus
EDTA	Ethylendiamintetraacetat
EST	Expressed Sequence Tag
FACS	Fluorescence Activated Cell Sorting
FC	Fold Change (Faktor des Expressionsunterschieds)
FITC	Fluorescein-Isothiocyanat
g	Gravitationsbeschleunigung [ $g = 9,80665 \text{ m s}^{-2}$ ]
GAM	Goat Anti Mouse (Ziege gegen Maus)
GPOH	Gesellschaft für pädiatrische Onkologie und Hämatologie
h	Stunde
HSA	Humanes Serumalbumin
IVT	<i>in vitro</i> -Transkription
KM	Knochenmark
l	Liter
$\lambda$	Wellenlänge [m]
m	Meter
m	milli-, $10^{-3}$
M	marginal (ein Gen ist geringfügig exprimiert)
M	Mol
$\mu$	mikro-, $10^{-6}$
MACS	Magnetic Activating Cell Sorting (magnetisch aktivierte Zellsortierung)
MB	Micro Beads
MES	2-[N-Morpholino]ethansulfonsäure
MHC	Major Histocompatibility Complex (Haupthistokompatibilitätskomplex)
min	Minute
MM	mismatch
MRD	Minimal Residual Disease (Minimale Resterkrankung)
mRNA	messenger RNA
n	nano-, $10^{-9}$
n.b.	nicht bestimmt
NK	Natürliche Killerzellen
OD	optische Dichte
OMIM	Online Mendelian Inheritance in Man

P	present (ein Gen ist exprimiert)
PB	peripheres Blut
PBS	Phosphat-gepufferte Salzlösung
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction (Polymerase Kettenreaktion)
PE	Phycoerythrin
PerCP	Peridinchlorophyll
PM	perfect match
RNA	Ribonucleic Acid (Ribonukleinsäure)
RT	Raum-, Umgebungstemperatur
RT-PCR	Reverse Transkriptase PCR
sec	Sekunde
SF	Skalierungsfaktor
SLC	Surrogate Light Chain
SOM	Self-organizing Maps
TdT	Terminale Deoxynukleotidyltransferase
TGF $\beta$	Tissue Growth Factor beta-1
TNF	Tumor Nekrosis Faktor
TS	Tonsilla palatina, Gaumenmandel
U	Unit
WBC	White Blood Count (Anzahl der Leukozyten im Blut)

### 6.3 *Anmerkungen zur verwendeten Nomenklatur*

Die vorliegende Untersuchung betrachtet die Expression von Genen, erläutert die von Microarrays detektierten Gentranskripte und beschreibt die Funktion der Genprodukte im zellulären Zusammenhang. Aus dem Text wird jeweils deutlich, welches dieser drei Moleküle gerade behandelt wird. Die Bezeichnungen für Gene oder Proteine folgen streng der Nomenklatur der NCBI-Datenbanken, wo man um eine Standardisierung der Bezeichnungen bemüht ist. In den dort verfügbaren Datenbanken, insbesondere OMIM, werden die jeweiligen Gene oder Proteine in der Bezeichnung oder der Schreibweise nicht länger unterschieden. Die Namen werden dort einheitlich in nicht kursiv gesetzten Großbuchstaben geschrieben (z.B. BIRC5). In der vorliegenden Arbeit werden diese Nomenklatur und die Schreibweise übernommen. Sofern für Gene in der Literatur weitere Synonyme verwendet werden, sind diese zumeist verbreiteter als der standardisierte Name. Aus diesem Grund werden die häufig verwendeten Bezeichnungen zusätzlich genannt (z.B. BIRC5 = API4). Wenn von der Funktion eines Gens gesprochen wird, ist darunter strenggenommen die Funktion des Genproduktes, also des Proteins zu verstehen. Soweit Trivialnamen von Proteinen häufig verwendet werden, werden diese ebenfalls zusätzlich genannt (z.B. BIRC5 = Survivin).

## 6.4 Literaturverzeichnis

- Aalto, Y., El-Rifa, W., Vilpo, L., Ollila, J., Nagy, B., Vihinen, M. *et al.* Distinct gene expression profiling in chronic lymphocytic leukemia with 11q23 deletion. *Leukemia* 15, 1721-1728. (2001).
- Abreu, J. G., Ketpura, N. I., Reversade, B. & De Robertis, E. M. Connective-tissue growth factor (CTGF) modulates cell signalling by BMP and TGF-beta. *Nat Cell Biol* 4, 599-604. (2002).
- Alizadeh, A. A., Eisen, M. B., Davis, R. E., Ma, C., Lossos, I. S., Rosenwald, A. *et al.* Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 403, 503-511 (2000).
- Alon, U., Barkai, N., Notterman, D. A., Gish, K., Ybarra, S., Mack, D. *et al.* Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proc Natl Acad Sci U S A* 96, 6745-6750. (1999).
- Altieri, D. C., Marchisio, P. C. & Marchisio, C. Survivin apoptosis: an interloper between cell death and cell proliferation in cancer. *Lab Invest* 79, 1327-1333. (1999).
- Ambrosini, G., Adida, C. & Altieri, D. C. A novel anti-apoptosis gene, survivin, expressed in cancer and lymphoma. *Nat Med* 3, 917-921. (1997).
- Arai, Y., Masuda, M., Sugawara, I., Arai, T., Motoji, T., Tsuruo, T. *et al.* Expression of the MDR1 and MDR3 gene products in acute and chronic leukemias. *Leuk Res* 21, 313-319. (1997).
- Arfin, S. M., Long, A. D., Ito, E. T., Toller, L., Riehle, M. M., Paegle, E. S. *et al.* Global gene expression profiling in Escherichia coli K12. The effects of integration host factor. *J Biol Chem* 275, 29672-29684. (2000).
- Armstrong, S. A., Hsieh, J. J. & Korsmeyer, S. J. Genomic approaches to the pathogenesis and treatment of acute lymphoblastic leukemias. *Curr Opin Hematol* 9, 339-344. (2002).
- Armstrong, S. A., Staunton, J. E., Silverman, L. B., Pieters, R., den Boer, M. L., Minden, M. D. *et al.* MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. *Nat Genet* 30, 41-47. (2001).
- Ashcroft, G. S., Yang, X., Glick, A. B., Weinstein, M., Letterio, J. L., Mizel, D. E. *et al.* Mice lacking Smad3 show accelerated wound healing and an impaired local inflammatory response. *Nat Cell Biol* 1, 260-266. (1999).
- Baier, T. G., Jenne, E. W., Blum, W., Schonberg, D. & Hartmann, K. K. Influence of antibodies against IGF-I, insulin or their receptors on proliferation of human acute lymphoblastic leukemia cell lines. *Leuk Res* 16, 807-814. (1992).
- Bain, G., Maandag, E. C., Izon, D. J., Amsen, D., Kruisbeek, A. M., Weintraub, B. C. *et al.* E2A proteins are required for proper B cell development and initiation of immunoglobulin gene rearrangements. *Cell* 79, 885-892. (1994).
- Bauer, M., Heeg, K., Wagner, H. & Lipford, G. B. DNA activates human immune cells through a CpG sequence-dependent manner. *Immunology* 97, 699-705. (1999).
- Bauer, S. R., Kubagawa, H., Maclennan, I. & Melchers, F. VpreB gene expression in hematopoietic malignancies: a lineage- and stage-restricted marker for B-cell precursor leukemias. *Blood* 78, 1581-1588. (1991).
- Becker-Herman, S., Lantner, F. & Shachar, I. Id2 negatively regulates B cell differentiation in the spleen. *J Immunol* 168, 5507-5513. (2002).

- Bellan, C., De Falco, G., Tosi, G. M., Lazzi, S., Ferrari, F., Morbini, G. *et al.* Missing expression of pRb2/p130 in human retinoblastomas is associated with reduced apoptosis and lesser differentiation. *Invest Ophthalmol Vis Sci* 43, 3602-3608. (2002).
- Ben-Dor, A., Bruhn, L., Friedman, N., Nachman, I., Schummer, M. & Yakhini, Z. Tissue classification with gene expression profiles. *J Comput Biol* 7, 559-583 (2000).
- Ben-Dor, A. & Yakhini, Z. Statistical Methods: Identification of Differentially Expressed Genes and Class Prediction Using Mutual Information and Total Number of Misclassification Analyses. (*N Engl J Med*) (344), (539-548) (2001).
- Bene, M. C. & Faure, G. C. CD10 in acute leukemias. GEIL (Groupe d'Etude Immunologique des Leucemies). *Haematologica* 82, 205-210. (1997).
- Biondi, A., Cimino, G., Pieters, R. & Pui, C. H. Biological and therapeutic aspects of infant leukemia. *Blood* 96, 24-33. (2000).
- Borkhardt, A., Cazzaniga, G., Viehmann, S., Valsecchi, M. G., Ludwig, W. D., Burci, L. *et al.* Incidence and clinical relevance of TEL/AML1 fusion genes in children with acute lymphoblastic leukemia enrolled in the German and Italian multicenter therapy trials. Associazione Italiana Ematologia Oncologia Pediatrica and the Berlin-Frankfurt-Munster Study Group. *Blood* 90, 571-577. (1997).
- Borregaard, N. & Cowland, J. B. Granules of the human neutrophilic polymorphonuclear leukocyte. *Blood* 89, 3503-3521. (1997).
- Bostwick, D. G., Alexander, E. E., Singh, R., Shan, A., Qian, J., Santella, R. M. *et al.* Antioxidant enzyme expression and reactive oxygen species damage in prostatic intraepithelial neoplasia and cancer. *Cancer* 89, 123-134. (2000).
- Bradley, J. R. & Pober, J. S. Tumor necrosis factor receptor-associated factors (TRAFs). *Oncogene* 20, 6482-6491. (2001).
- Brantley, M. A., Jr. & Harbour, J. W. The molecular biology of retinoblastoma. *Ocul Immunol Inflamm* 9, 1-8. (2001).
- Brazma, A. & Vilo, J. Gene expression data analysis. *FEBS Lett* 480, 17-24. (2000).
- Brenner, S., Johnson, M., Bridgham, J., Golda, G., Lloyd, D. H., Johnson, D. *et al.* Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat Biotechnol* 18, 630-634. (2000).
- Bruserud, O., Gjertsen, B. T. & Ulvestad, E. Expression of Fc(epsilon)-receptors by human acute myelogenous leukemia (AML) blasts: studies of high- and low- (CD23) affinity receptor expression and the effects of IgE-mediated receptor ligation on functional AML blast characteristics. *Leuk Res* 26, 515-521. (2002).
- Burdach, S. in *Medizinische Therapie in Klinik und Praxis* (ed. Schoelmerich, P.) (, in press).
- Busslinger, M., Nutt, S. L. & Rolink, A. G. Lineage commitment in lymphopoiesis. *Curr Opin Immunol* 12, 151-158 (2000).
- Butte, A. J., Tamayo, P., Slonim, D., Golub, T. R. & Kohane, I. S. Discovering functional relationships between RNA expression and chemotherapeutic susceptibility using relevance networks. *Proc Natl Acad Sci U S A* 97, 12182-12186 (2000).
- Cahill, D. P., Lengauer, C., Yu, J., Riggins, G. J., Willson, J. K., Markowitz, S. D. *et al.* Mutations of mitotic checkpoint genes in human cancers. *Nature* 392, 300-303. (1998).
- Calado, R. T., Garcia, A. B. & Falcao, R. P. Age-related changes of immunophenotypically immature lymphocytes in normal human peripheral blood. *Cytometry* 38, 133-137 (1999).

- Carr, R. Neutrophil production and function in newborn infants. *Br J Haematol* 110, 18-28. (2000).
- Casas, C., Martinez, S., Pritchard, M. A., Fuentes, J. J., Nadal, M., Guimera, J. *et al.* Dscr1, a novel endogenous inhibitor of calcineurin signaling, is expressed in the primitive ventricle of the heart and during neurogenesis. *Mech Dev* 101, 289-292. (2001).
- Chakrabarti, S. R., Sood, R., Nandi, S. & Nucifora, G. Posttranslational modification of TEL and TEL/AML1 by SUMO-1 and cell- cycle-dependent assembly into nuclear bodies. *Proc Natl Acad Sci U S A* 97, 13281-13285. (2000).
- Chambraud, B., Radanyi, C., Camonis, J. H., Rajkowski, K., Schumacher, M. & Baulieu, E. E. Immunophilins, Refsum disease, and lupus nephritis: the peroxisomal enzyme phytanoyl-CoA alpha-hydroxylase is a new FKBP-associated protein. *Proc Natl Acad Sci U S A* 96, 2104-2109. (1999).
- Chandra, D., Liu, J. W. & Tang, D. G. Early mitochondrial activation and cytochrome c upregulation during apoptosis. *J Biol Chem* 277, 28 (2002).
- Chang, C., Holtzman, D. A., Chau, S., Chickering, T., Woolf, E. A., Holmgren, L. M. *et al.* Twisted gastrulation can function as a BMP antagonist. *Nature* 410, 483-487. (2001).
- Chang, N. S., Mattison, J., Cao, H., Pratt, N., Zhao, Y. & Lee, C. Cloning and characterization of a novel transforming growth factor-beta1-induced TIAF1 protein that inhibits tumor necrosis factor cytotoxicity. *Biochem Biophys Res Commun* 253, 743-749. (1998).
- Chen, J. J., Peck, K., Hong, T. M., Yang, S. C., Sher, Y. P., Shih, J. Y. *et al.* Global analysis of gene expression in invasion by a lung cancer model. *Cancer Res* 61, 5223-5230. (2001a).
- Chen, J. S., Coustan-Smith, E., Suzuki, T., Neale, G. A., Mihara, K., Pui, C. H. *et al.* Identification of novel markers for monitoring minimal residual disease in acute lymphoblastic leukemia. *Blood* 97, 2115-2120. (2001b).
- Chudin, E., Walker, R., Kosaka, A., Wu, S. X., Rabert, D., Chang, T. K. *et al.* Assessment of the relationship between signal intensities and transcript concentration for Affymetrix GeneChip arrays. *Genome Biol* 3, 5.1-5.10 (2002).
- Claverie, J. M. Computational methods for the identification of differential and coordinated gene expression. *Hum Mol Genet* 8, 1821-1832 (1999).
- Cohen, M. M., Jr. RUNX genes, neoplasia, and cleidocranial dysplasia. *Am J Med Genet* 104, 185-188. (2001).
- Comans-Bitter, W. M., de Groot, R., van den Beemd, R., Neijens, H. J., Hop, W. C., Groeneveld, K. *et al.* Immunophenotyping of blood lymphocytes in childhood. Reference values for lymphocyte subpopulations. *J Pediatr* 130, 388-393. (1997).
- Cutrona, G., Tasso, P., Dono, M., Roncella, S., Ulivi, M., Carpaneto, E. M. *et al.* CD10 is a marker for cycling cells with propensity to apoptosis in childhood ALL. *Br J Cancer* 86, 1776-1785. (2002).
- Dale, T. C. Signal transduction by the Wnt family of ligands. *Biochem J* 329, 209-223. (1998).
- Dannenberg, J. H., van Rossum, A., Schuijff, L. & te Riele, H. Ablation of the retinoblastoma gene family deregulates G(1) control causing immortalization and increased cell turnover under growth-restricting conditions. *Genes Dev* 14, 3051-3064. (2000).
- Davies, T. H., Ning, Y. M. & Sanchez, E. R. A new first step in activation of steroid receptors: hormone-induced switching of FKBP51 and FKBP52 immunophilins. *J Biol Chem* 277, 4597-4600. (2002).
- de Lau, W. & Clevers, H. LEF1 turns over a new leaf. *Nat Genet* 28, 3-4. (2001).

- Deguchi, Y., Agus, D. & Kehrl, J. H. A human homeobox gene, HB24, inhibits development of CD4<sup>+</sup> T cells and impairs thymic involution in transgenic mice. *J Biol Chem* 268, 3646-3653. (1993a).
- Deguchi, Y., Yamanaka, Y., Theodossiou, C., Najfeld, V. & Kehrl, J. H. High expression of two diverged homeobox genes, HB24 and HB9, in acute leukemias: molecular markers of hematopoietic cell immaturity. *Leukemia* 7, 446-451. (1993b).
- DeRisi, J. L., Iyer, V. R. & Brown, P. O. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* 278, 680-686. (1997).
- Donadieu, J., Auclerc, M. F., Baruchel, A., Leblanc, T., Landman-Parker, J., Perel, Y. *et al.* Critical study of prognostic factors in childhood acute lymphoblastic leukaemia: differences in outcome are poorly explained by the most significant prognostic variables. Fralle group. French Acute Lymphoblastic Leukaemia study group. *Br J Haematol* 102, 729-739. (1998).
- Donadieu, J., Auclerc, M. F., Baruchel, A., Perel, Y., Bordigoni, P., Landman-Parker, J. *et al.* Prognostic study of continuous variables (white blood cell count, peripheral blast cell count, haemoglobin level, platelet count and age) in childhood acute lymphoblastic leukaemia. Analysis Of a population of 1545 children treated by the French Acute Lymphoblastic Leukaemia Group (FRALLE). *Br J Cancer* 83, 1617-1622. (2000).
- Ebisawa, T., Fukuchi, M., Murakami, G., Chiba, T., Tanaka, K., Imamura, T. *et al.* Smurf1 interacts with transforming growth factor-beta type I receptor through Smad7 and induces receptor degradation. *J Biol Chem* 276, 12477-12480. (2001).
- Eickhoff, B., Korn, B., Schick, M., Poustka, A. & van der Bosch, J. Normalization of array hybridization experiments in differential gene expression analysis. *Nucleic Acids Res* 27, e33. (1999).
- Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* 95, 14863-14868 (1998).
- Eklholm, S. V. & Reed, S. I. Regulation of G(1) cyclin-dependent kinases in the mammalian cell cycle. *Curr Opin Cell Biol* 12, 676-684. (2000).
- Felix, C. A. & Lange, B. J. Leukemia in infants. *Oncologist* 4, 225-240 (1999).
- Ferrando, A. A., Neuberg, D. S., Staunton, J., Loh, M. L., Huard, C., Raimondi, S. C. *et al.* Gene expression signatures define novel oncogenic pathways in T cell acute lymphoblastic leukemia. *Cancer Cell* 1, 75-87 (2002).
- Fields, S., Kohara, Y. & Lockhart, D. J. Functional genomics. *Proc Natl Acad Sci U S A* 96, 8825-8826. (1999).
- Fink, J. R. & LeBien, T. W. Novel expression of cyclin-dependent kinase inhibitors in human B-cell precursors. *Exp Hematol* 29, 490-498. (2001).
- Fodor, S. P., Rava, R. P., Huang, X. C., Pease, A. C., Holmes, C. P. & Adams, C. L. Multiplexed biochemical assays with biological chips. *Nature* 364, 555-556. (1993).
- Fodor, S. P., Read, J. L., Pirrung, M. C., Stryer, L., Lu, A. T. & Solas, D. Light-directed, spatially addressable parallel chemical synthesis. *Science* 251, 767-773. (1991).
- Freeman, W. M., Robertson, D. J. & Vrana, K. E. Fundamentals of DNA hybridization arrays for gene expression analysis. *Biotechniques* 29, 1042-1046, 1048-1055. (2000).
- Friedmann, A. M. & Weinstein, H. J. The role of prognostic features in the treatment of childhood acute lymphoblastic leukemia. *Oncologist* 5, 321-328 (2000).
- Fry, A. M., Meraldi, P. & Nigg, E. A. A centrosomal function for the human Nek2 protein kinase, a member of the NIMA family of cell cycle regulators. *Embo J* 17, 470-481. (1998).

- Garber, M. E., Troyanskaya, O. G., Schluens, K., Petersen, S., Thaesler, Z., Pacyna-Gengelbach, M. *et al.* Diversity of gene expression in adenocarcinoma of the lung. *Proc Natl Acad Sci U S A* 98, 13784-13789. (2001).
- Garlisi, C. G., Uss, A. S., Xiao, H., Tian, F., Sheridan, K. E., Wang, L. *et al.* A unique mRNA initiated within a middle intron of WHSC1/MMSET encodes a DNA binding protein that suppresses human IL-5 transcription. *Am J Respir Cell Mol Biol* 24, 90-98. (2001).
- Gasch, A. P., Spellman, P. T., Kao, C. M., Carmel-Harel, O., Eisen, M. B., Storz, G. *et al.* Genomic expression programs in the response of yeast cells to environmental changes. *Mol Biol Cell* 11, 4241-4257. (2000).
- Gibbons, F. D. & Roth, F. P. Judging the quality of gene expression-based clustering methods using gene annotation. *Genome Res* 12, 1574-1581. (2002).
- Gisler, R., Jacobsen, S. E. & Sigvardsson, M. Cloning of human early B-cell factor and identification of target genes suggest a conserved role in B-cell development in man and mouse. *Blood* 96, 1457-1464. (2000).
- Gisler, R. & Sigvardsson, M. The human V-preB promoter is a target for coordinated activation by early B cell factor and E47. *J Immunol* 168, 5130-5138. (2002).
- Glienke, J., Fenten, G., Seemann, M., Sturz, A. & Thierauch, K. H. Human SPRY2 inhibits FGF2 signalling by a secreted factor. *Mech Dev* 96, 91-99. (2000).
- Glynne, R., Akkaraju, S., Healy, J. I., Rayner, J., Goodnow, C. C. & Mack, D. H. How self-tolerance and the immunosuppressive drug FK506 prevent B-cell mitogenesis. *Nature* 403, 672-676 (2000a).
- Glynne, R., Ghandour, G., Rayner, J., Mack, D. H. & Goodnow, C. C. B-lymphocyte quiescence, tolerance and activation as viewed by global gene expression profiling on microarrays. *Immunol Rev* 176, 216-246. (2000b).
- Golub, T. R. Genomic approaches to the pathogenesis of hematologic malignancy. *Curr Opin Hematol* 8, 252-261. (2001).
- Golub, T. R., Barker, G. F., Bohlander, S. K., Hiebert, S. W., Ward, D. C., Bray-Ward, P. *et al.* Fusion of the TEL gene on 12p13 to the AML1 gene on 21q22 in acute lymphoblastic leukemia. *Proc Natl Acad Sci U S A* 92, 4917-4921. (1995).
- Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P. *et al.* Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 286, 531-537 (1999).
- Gorgoulis, V. G., Zacharatos, P., Mariatos, G., Kotsinas, A., Bouda, M., Kletsas, D. *et al.* Transcription factor E2F-1 acts as a growth-promoting factor and is associated with adverse prognosis in non-small cell lung carcinomas. *J Pathol* 198, 142-156. (2002).
- Gorlick, R., Cole, P., Banerjee, D., Longo, G., Li, W. W., Hochhauser, D. *et al.* Mechanisms of methotrexate resistance in acute leukemia. Decreased transport and polyglutamylation. *Adv Exp Med Biol* 457, 543-550 (1999).
- Greaves, M. Molecular genetics, natural history and the demise of childhood leukaemia. *Eur J Cancer* 35, 173-185. (1999).
- Greaves, M. F. Aetiology of acute leukaemia. *Lancet* 349, 344-349. (1997).
- Hanai, J., Chen, L. F., Kanno, T., Ohtani-Fujita, N., Kim, W. Y., Guo, W. H. *et al.* Interaction and functional cooperation of PEBP2/CBF with Smads. Synergistic induction of the immunoglobulin germline C $\alpha$  promoter. *J Biol Chem* 274, 31577-31582. (1999).
- Hanenbergh, H., Baumann, M., Quentin, I., Nagel, G., Grosse-Wilde, H., von Kleist, S. *et al.* Expression of the CEA gene family members NCA-50/90 and NCA-160 (CD66) in child-

- hood acute lymphoblastic leukemias (ALLs) and in cell lines of B-cell origin. *Leukemia* 8, 2127-2133. (1994).
- Hanrahan, J., Pan, L., Greenbaum, S., Bradney, C., Hjelmeland, M., Dai, M. *et al.* A genetic investigation of E2A function in lymphocyte development. *Immunol Res* 22, 211-222 (2000).
- Harada, H., Yoshida, S., Nobe, Y., Ezura, Y., Atake, T., Koguchi, T. *et al.* Genomic structure of the human NLK (nemo-like kinase) gene and analysis of its promoter region. *Gene* 285, 175-182. (2002).
- Harbott, J., Viehmann, S., Borkhardt, A., Henze, G. & Lampert, F. Incidence of TEL/AML1 fusion gene analyzed consecutively in children with acute lymphoblastic leukemia in relapse. *Blood* 90, 4933-4937. (1997).
- Hattenhorst, U., Glynne, R., Murray, R. & Burdach, S. Differential Gene Expression Analysis Utilizing DNA-Microchips In Common Acute Leukemia Versus Normal Bone Marrow. *Acta Haematologica* 103, 43 (2000a).
- Hattenhorst, U., Tom, E., Glynne, R., Murray, R. & Burdach, S. Differential gene expression analysis in pediatric common acute lymphoblastic leukemia versus normal pre-B-cells and bone marrow utilizing DNA-microarrays. *Blood* 96, 107A (2000b).
- Hayashi, H., Abdollah, S., Qiu, Y., Cai, J., Xu, Y. Y., Grinnell, B. W. *et al.* The MAD-related protein Smad7 associates with the TGFbeta receptor and functions as an antagonist of TGFbeta signaling. *Cell* 89, 1165-1173. (1997).
- Hayward, D. C., Samuel, G., Pontynen, P. C., Catmull, J., Saint, R., Miller, D. J. *et al.* Localized expression of a dpp/BMP2/4 ortholog in a coral embryo. *Proc Natl Acad Sci U S A* 99, 8106-8111. (2002).
- Hedenfalk, I., Duggan, D., Chen, Y., Radmacher, M., Bittner, M., Simon, R. *et al.* Gene-expression profiles in hereditary breast cancer. *N Engl J Med* 344, 539-548. (2001).
- Heldin, C. H., Miyazono, K. & ten Dijke, P. TGF-beta signalling from cell membrane to nucleus through SMAD proteins. *Nature* 390, 465-471. (1997).
- Hess, J., Werner, A., Wirth, T., Melchers, F., Jack, H. M. & Winkler, T. H. Induction of pre-B cell proliferation after de novo synthesis of the pre-B cell receptor. *Proc Natl Acad Sci U S A* 98, 1745-1750. (2001).
- Hill, A. A., Brown, E. L., Whitley, M. Z., Tucker-Kellogg, G., Hunter, C. P. & Slonim, D. K. Evaluation of normalization procedures for oligonucleotide array data based on spiked cRNA controls. *Genome Biol* 2 (2001).
- Hill, A. A., Hunter, C. P., Tsung, B. T., Tucker-Kellogg, G. & Brown, E. L. Genomic analysis of gene expression in *C. elegans*. *Science* 290, 809-812. (2000).
- Hoffkes, H. G., Schmidtke, G., Uppenkamp, M. & Schmucker, U. Multiparametric immunophenotyping of B cells in peripheral blood of healthy adults by flow cytometry. *Clin Diagn Lab Immunol* 3, 30-36 (1996).
- Hoffmann, R., Seidl, T., Neeb, M., Rolink, A. & Melchers, F. Changes in gene expression profiles in developing B cells of murine bone marrow. *Genome Res* 12, 98-111. (2002).
- Hofmann, W. K., de Vos, S., Elashoff, D., Gschaidmeier, H., Hoelzer, D., Koeffler, H. P. *et al.* Relation between resistance of Philadelphia-chromosome-positive acute lymphoblastic leukaemia to the tyrosine kinase inhibitor ST1571 and gene-expression profiles: a gene-expression study. *Lancet* 359, 481-486. (2002).
- Hollnagel, A., Oehlmann, V., Heymer, J., Ruther, U. & Nordheim, A. Id genes are direct targets of bone morphogenetic protein induction in embryonic stem cells. *J Biol Chem* 274, 19838-19845. (1999).

- Holter, N. S., Mitra, M., Maritan, A., Cieplak, M., Banavar, J. R. & Fedoroff, N. V. Fundamental patterns underlying gene expression profiles: simplicity from complexity. *Proc Natl Acad Sci U S A* 97, 8409-8414. (2000).
- Hovanes, K., Li, T. W., Munguia, J. E., Truong, T., Milovanovic, T., Lawrence Marsh, J. *et al.* Beta-catenin-sensitive isoforms of lymphoid enhancer factor-1 are selectively expressed in colon cancer. *Nat Genet* 28, 53-57. (2001).
- Hrusak, O. & Porwit-MacDonald, A. Antigen expression patterns reflecting genotype of acute leukemias. *Leukemia* 16, 1233-1258. (2002).
- Hu, G., Lee, H., Price, S. M., Shen, M. M. & Abate-Shen, C. Msx homeobox genes inhibit differentiation through upregulation of cyclin D1. *Development* 128, 2373-2384. (2001).
- Inaba, T., Shimazaki, C., Sumikuma, T., Okamoto, A., Shimura, K., Takahashi, R. *et al.* CD45-negative acute leukemia in adulthood. *Eur J Haematol* 64, 66-67. (2000).
- Ishida, S., Huang, E., Zuzan, H., Spang, R., Leone, G., West, M. *et al.* Role for E2F in control of both DNA replication and mitotic functions as revealed from DNA microarray analysis. *Mol Cell Biol* 21, 4684-4699. (2001).
- Iyer, V. R., Eisen, M. B., Ross, D. T., Schuler, G., Moore, T., Lee, J. C. *et al.* The transcriptional program in the response of human fibroblasts to serum. *Science* 283, 83-87. (1999).
- Janeway, C. A., Travers, P., Walport, M. & Shlomchik, M. *Immunobiology* (Churchill Livingstone, 2001).
- Kaiser, B. K., Zimmerman, Z. A., Charbonneau, H. & Jackson, P. K. Disruption of centrosome structure, chromosome segregation, and cytokinesis by misexpression of human Cdc14A phosphatase. *Mol Biol Cell* 13, 2289-2300. (2002).
- Kapranov, P., Cawley, S. E., Drenkow, J., Bekiranov, S., Strausberg, R. L., Fodor, S. P. *et al.* Large-scale transcriptional activity in chromosomes 21 and 22. *Science* 296, 916-919. (2002).
- Karlin, S., Bergman, A. & Gentles, A. J. Genomics: Annotation of the Drosophila genome. *Nature* 411, 259-260. (2001).
- Kasof, G. M. & Gomes, B. C. Livin, a novel inhibitor of apoptosis protein family member. *J Biol Chem* 276, 3238-3246. (2001a).
- Kasof, G. M., Lu, J. J., Liu, D., Speer, B., Mongan, K. N., Gomes, B. C. *et al.* Tumor necrosis factor-alpha induces the expression of DR6, a member of the TNF receptor family, through activation of NF-kappaB. *Oncogene* 20, 7965-7975. (2001b).
- Kersey, J. H. Fifty years of studies of the biology and therapy of childhood leukemia. *Blood* 90, 4243-4251. (1997).
- Khan, J., Simon, R., Bittner, M., Chen, Y., Leighton, S. B., Pohida, T. *et al.* Gene expression profiling of alveolar rhabdomyosarcoma with cDNA microarrays. *Cancer Res* 58, 5009-5013. (1998).
- Kim, C. H., Pelus, L. M., Appelbaum, E., Johanson, K., Anzai, N. & Broxmeyer, H. E. CCR7 ligands, SLC/6CKine/Exodus2/TCA4 and CKbeta-11/MIP-3beta/ELC, are chemoattractants for CD56(+)/CD16(-) NK cells and late stage lymphoid progenitors. *Cell Immunol* 193, 226-235. (1999).
- Kim, R. H., Wang, D., Tsang, M., Martin, J., Huff, C., de Caestecker, M. P. *et al.* A novel smad nuclear interacting protein, SNIP1, suppresses p300-dependent TGF-beta signal transduction. *Genes Dev* 14, 1605-1616. (2000).
- Kinderkrebsregister, D. Jahresbericht 1999. (1999).

- Klein, U., Tu, Y., Stolovitzky, G. A., Mattioli, M., Cattoretti, G., Husson, H. *et al.* Gene expression profiling of B cell chronic lymphocytic leukemia reveals a homogeneous phenotype related to memory B cells. *J Exp Med* 194, 1625-1638. (2001).
- Knight, J. When the chips are down. *Nature* 410, 860-861. (2001).
- Knight, J. Minimum standards set out for gene-expression data. *Nature* 415, 946. (2002).
- Korz, C., Pscherer, A., Benner, A., Mertens, D., Schaffner, C., Leupolt, E. *et al.* Evidence for distinct pathomechanisms in B-cell chronic lymphocytic leukemia and mantle cell lymphoma by quantitative expression analysis of cell cycle and apoptosis-associated genes. *Blood* 99, 4554-4561. (2002).
- Labbe, E., Letamendia, A. & Attisano, L. Association of Smads with lymphoid enhancer binding factor 1/T cell-specific factor mediates cooperative signaling by the transforming growth factor-beta and wnt pathways. *Proc Natl Acad Sci U S A* 97, 8358-8363. (2000).
- Larramendy, M. L., Niini, T., Elonen, E., Nagy, B., Ollila, J., Vihinen, M. *et al.* Overexpression of translocation-associated fusion genes of FGFR1, MYC, NPM1, and DEK, but absence of the translocations in acute myeloid leukemia. A microarray analysis. *Haematologica* 87, 569-577. (2002).
- Lasorella, A., Iavarone, A. & Israel, M. A. Id2 specifically alters regulation of the cell cycle by tumor suppressor proteins. *Mol Cell Biol* 16, 2570-2578. (1996).
- Lasorella, A., Nosedà, M., Beyna, M., Yokota, Y. & Iavarone, A. Id2 is a retinoblastoma protein target and mediates signalling by Myc oncoproteins. *Nature* 407, 592-598. (2000).
- Lavabre-Bertrand, T., Janossy, G., Ivory, K., Peters, R., Secker-Walker, L. & Porwit-MacDonald, A. Leukemia-associated changes identified by quantitative flow cytometry: I. CD10 expression. *Cytometry* 18, 209-217. (1994).
- Lawrance, I. C., Fiocchi, C. & Chakravarti, S. Ulcerative colitis and Crohn's disease: distinctive gene expression profiles and novel susceptibility candidate genes. *Hum Mol Genet* 10, 445-456. (2001).
- Lee, P. D., Sladek, R., Greenwood, C. M. & Hudson, T. J. Control genes and variability: absence of ubiquitous reference transcripts in diverse mammalian expression studies. *Genome Res* 12, 292-297. (2002).
- Lemmers, B., Gauthier, L., Guelpa-Fonlupt, V., Fougereau, M. & Schiff, C. The human (PsiL+mu-) proB complex: cell surface expression and biochemical structure of a putative transducing receptor. *Blood* 93, 4336-4346. (1999).
- Leone, G., DeGregori, J., Yan, Z., Jakoi, L., Ishida, S., Williams, R. S. *et al.* E2F3 activity is regulated during the cell cycle and is required for the induction of S phase. *Genes Dev* 12, 2120-2130. (1998).
- Levy, O., Martin, S., Eichenwald, E., Ganz, T., Valore, E., Carroll, S. F. *et al.* Impaired innate immunity in the newborn: newborn neutrophils are deficient in bactericidal/permeability-increasing protein. *Pediatrics* 104, 1327-1333. (1999).
- Li, C. & Wong, W. H. Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc Natl Acad Sci U S A* 98, 31-36. (2001).
- Li, H., Bartold, P. M., Zhang, C. Z., Clarkson, R. W., Young, W. G. & Waters, M. J. Growth hormone and insulin-like growth factor I induce bone morphogenetic proteins 2 and 4: a mediator role in bone and tooth formation? *Endocrinology* 139, 3855-3862. (1998).
- Li, K. K., Ng, I. O., Fan, S. T., Albrecht, J. H., Yamashita, K. & Poon, R. Y. Activation of cyclin-dependent kinases CDC2 and CDK2 in hepatocellular carcinoma. *Liver* 22, 259-268. (2002a).

- Li, Q. H., Haga, I., Shimizu, T., Itoh, M., Kurosaki, T. & Fujisawa, J. Retardation of the G2-M phase progression on gene disruption of RNA binding protein Sam68 in the DT40 cell line. *FEBS Lett* 525, 145-150. (2002b).
- Lipshutz, R. J., Fodor, S. P., Gingeras, T. R. & Lockhart, D. J. High density synthetic oligonucleotide arrays. *Nat Genet* 21, 20-24 (1999).
- Liu, G., Loraine, A. E., Shigeta, R., Cline, M., Cheng, J., Valmeekam, V. *et al.* NetAffx: Affymetrix probesets and annotations. *Nucleic Acids Res* 31, 82-86. (2003).
- Lockhart, D. J., Dong, H., Byrne, M. C., Follettie, M. T., Gallo, M. V., Chee, M. S. *et al.* Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat Biotechnol* 14, 1675-1680 (1996).
- Lockhart, D. J. & Winzeler, E. A. Genomics, gene expression and DNA arrays. *Nature* 405, 827-836. (2000).
- Lopez-Matas, M., Rodriguez-Justo, M., Morilla, R., Catovsky, D. & Matutes, E. Quantitative expression of CD23 and its ligand CD21 in chronic lymphocytic leukemia. *Haematologica* 85, 1140-1145. (2000).
- Lucio, P., Gaipa, G., van Lochem, E. G., van Wering, E. R., Porwit-MacDonald, A., Faria, T. *et al.* BIOMED-I concerted action report: flow cytometric immunophenotyping of precursor B-ALL with standardized triple-stainings. BIOMED-1 Concerted Action Investigation of Minimal Residual Disease in Acute Leukemia: International Standardization and Clinical Evaluation. *Leukemia* 15, 1185-1192. (2001).
- Maia, A. T., Ford, A. M., Jalali, G. R., Harrison, C. J., Taylor, G. M., Eden, O. B. *et al.* Molecular tracking of leukemogenesis in a triplet pregnancy. *Blood* 98, 478-482. (2001).
- Mak, T. W. & Yeh, W. C. Signaling for survival and apoptosis in the immune system. *Arthritis Res* 4, S243-252. (2002).
- Massague, J. & Wotton, D. Transcriptional control by the TGF-beta/Smad signaling system. *Embo J* 19, 1745-1754. (2000).
- McGall, G., Labadie, J., Brock, P., Wallraff, G., Nguyen, T. & Hinsberg, W. Light-directed synthesis of high-density oligonucleotide arrays using semiconductor photoresists. *Proc Natl Acad Sci U S A* 93, 13555-13560. (1996).
- Melchers, F., ten Boekel, E., Seidl, T., Kong, X. C., Yamagami, T., Onishi, K. *et al.* Repertoire selection by pre-B-cell receptors and B-cell receptors, and genetic control of B-cell development from immature to mature B cells. *Immunol Rev* 175, 33-46. (2000).
- Melchers, M. & Rolink, A. in *Fundamental Immunology* (ed. Paul, W. E.) 183-224 (Lippincott-Raven Publishers, Philadelphia, 1998).
- Mikkers, H., Allen, J. & Berns, A. Proviral activation of the tumor suppressor E2a contributes to T cell lymphomagenesis in EmicroMyc transgenic mice. *Oncogene* 21, 6559-6566. (2002).
- Miltenyi, S., Muller, W., Weichel, W. & Radbruch, A. High gradient magnetic cell separation with MACS. *Cytometry* 11, 231-238 (1990).
- Molica, S., Levato, D., Dell'Olio, M., Matera, R., Minervini, M., Dattilo, A. *et al.* Cellular expression and serum circulating levels of CD23 in B-cell chronic lymphocytic leukemia. Implications for prognosis. *Haematologica* 81, 428-433. (1996).
- Mood, K., Friesel, R. & Daar, I. O. SNT1/FRS2 mediates germinal vesicle breakdown induced by an activated FGF receptor1 in *Xenopus* oocytes. *J Biol Chem* 277, 33196-33204. (2002).

- Moos, P. J., Raetz, E. A., Carlson, M. A., Szabo, A., Smith, F. E., Willman, C. *et al.* Identification of gene expression profiles that segregate patients with childhood leukemia. *Clin Cancer Res* 8, 3118-3130. (2002).
- Mori, H., Colman, S. M., Xiao, Z., Ford, A. M., Healy, L. E., Donaldson, C. *et al.* Chromosome translocations and covert leukemic clones are generated during normal fetal development. *Proc Natl Acad Sci U S A* 99, 8242-8247. (2002).
- Muller, H., Bracken, A. P., Vernell, R., Moroni, M. C., Christians, F., Grassilli, E. *et al.* E2Fs regulate the expression of genes involved in differentiation, development, proliferation, and apoptosis. *Genes Dev* 15, 267-285. (2001).
- Muschen, M., Lee, S., Zhou, G., Feldhahn, N., Barath, V. S., Chen, J. *et al.* Molecular portraits of B cell lineage commitment. *Proc Natl Acad Sci U S A* 99, 10014-10019. (2002).
- Nakao, A., Afrakhte, M., Moren, A., Nakayama, T., Christian, J. L., Heuchel, R. *et al.* Identification of Smad7, a TGFbeta-inducible antagonist of TGF-beta signalling. *Nature* 389, 631-635. (1997).
- Nakashima, K., Yanagisawa, M., Arakawa, H., Kimura, N., Hisatsune, T., Kawabata, M. *et al.* Synergistic signaling in fetal brain by STAT3-Smad1 complex bridged by p300. *Science* 284, 479-482. (1999).
- Nasedkina, T., Domer, P., Zharinov, V., Hoberg, J., Lysov, Y. & Mirzabekov, A. Identification of chromosomal translocations in leukemias by hybridization with oligonucleotide microarrays. *Haematologica* 87, 363-372. (2002).
- Neely, E. K., Rosenfeld, R. G., Illescas, A. & Smith, S. D. Mitogenic effects of human recombinant insulin on B-cell precursor acute lymphoblastic leukemia cells. *Leukemia* 6, 1134-1142. (1992).
- Nielsen, J., Christiansen, J., Lykke-Andersen, J., Johnsen, A. H., Wewer, U. M. & Nielsen, F. C. A family of insulin-like growth factor II mRNA-binding proteins represses translation in late development. *Mol Cell Biol* 19, 1262-1270. (1999).
- Notterman, D. A., Alon, U., Sierk, A. J. & Levine, A. J. Transcriptional gene expression profiles of colorectal adenoma, adenocarcinoma, and normal tissue examined by oligonucleotide arrays. *Cancer Res* 61, 3124-3130. (2001).
- O'Gorman, M. R., Millard, D. D., Lowder, J. N. & Yogev, R. Lymphocyte subpopulations in healthy 1-3-day-old infants. *Cytometry* 34, 235-241. (1998).
- Okuda, T., van Deursen, J., Hiebert, S. W., Grosveld, G. & Downing, J. R. AML1, the target of multiple chromosomal translocations in human leukemia, is essential for normal fetal liver hematopoiesis. *Cell* 84, 321-330. (1996).
- O'Riordan, M. & Grosschedl, R. Coordinate regulation of B cell differentiation by the transcription factors EBF and E2A. *Immunity* 11, 21-31. (1999).
- Peattie, D. A., Harding, M. W., Fleming, M. A., DeCenzo, M. T., Lippke, J. A., Livingston, D. J. *et al.* Expression and characterization of human FKBP52, an immunophilin that associates with the 90-kDa heat shock protein and is a component of steroid receptor complexes. *Proc Natl Acad Sci U S A* 89, 10974-10978. (1992).
- Perdomo, J., Holmes, M., Chong, B. & Crossley, M. Eos and pegasus, two members of the Ikaros family of proteins with distinct DNA binding activities. *J Biol Chem* 275, 38347-38354. (2000).
- Perez-Roger, I., Kim, S. H., Griffiths, B., Sewing, A. & Land, H. Cyclins D1 and D2 mediate myc-induced proliferation via sequestration of p27(Kip1) and p21(Cip1). *Embo J* 18, 5310-5320. (1999).

- Perou, C. M., Jeffrey, S. S., van de Rijn, M., Rees, C. A., Eisen, M. B., Ross, D. T. *et al.* Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. *Proc Natl Acad Sci U S A* 96, 9212-9217. (1999).
- Perry, C., Eldor, A. & Soreq, H. Runx1/AML1 in leukemia: disrupted association with diverse protein partners. *Leuk Res* 26, 221-228. (2002).
- Persson, C., Martensson, A. & Martensson, I. L. Identification of a tissue- and differentiation stage-specific enhancer of the VpreB1 gene. *Eur J Immunol* 28, 787-798. (1998).
- Pui, C. H. Childhood leukemias. *N Engl J Med* 332, 1618-1630. (1995).
- Pui, C. H. Acute lymphoblastic leukemia in children. *Curr Opin Oncol* 12, 3-12. (2000).
- Pui, C. H., Boyett, J. M., Relling, M. V., Harrison, P. L., Rivera, G. K., Behm, F. G. *et al.* Sex differences in prognosis for children with acute lymphoblastic leukemia. *J Clin Oncol* 17, 818-824. (1999).
- Pui, C. H. & Evans, W. E. Acute lymphoblastic leukemia. *N Engl J Med* 339, 605-615. (1998).
- Pullen, J., Shuster, J. J., Link, M., Borowitz, M., Amylon, M., Carroll, A. J. *et al.* Significance of commonly used prognostic factors differs for children with T cell acute lymphocytic leukemia (ALL), as compared to those with B-precursor ALL. A Pediatric Oncology Group (POG) study. *Leukemia* 13, 1696-1707. (1999).
- Quackenbush, J. Computational analysis of microarray data. *Nat Rev Genet* 2, 418-427. (2001).
- Ratei, R., Sperling, C., Karawajew, L., Schott, G., Schrappe, M., Harbott, J. *et al.* Immunophenotype and clinical characteristics of CD45-negative and CD45-positive childhood acute lymphoblastic leukemia. *Ann Hematol* 77, 107-114. (1998).
- Rego, E. M., Garcia, A. B., Carneiro, J. J. & Falcao, R. P. Immunophenotype of normal and leukemic bone marrow B-precursors in a Brazilian population. A comparative analysis by quantitative fluorescence cytometry. *Braz J Med Biol Res* 34, 183-194. (2001).
- Rego, E. M., Garcia, A. B., Viana, S. R. & Falcao, R. P. Age-related changes of lymphocyte subsets in normal bone marrow biopsies. *Cytometry* 34, 22-29 (1998).
- Robertson, K. D., Ait-Si-Ali, S., Yokochi, T., Wade, P. A., Jones, P. L. & Wolffe, A. P. DNMT1 forms a complex with Rb, E2F1 and HDAC1 and represses transcription from E2F-responsive promoters. *Nat Genet* 25, 338-342. (2000).
- Rockman, S. P., Currie, S. A., Ciavarella, M., Vincan, E., Dow, C., Thomas, R. J. *et al.* Id2 is a target of the beta-catenin/T cell factor pathway in colon carcinoma. *J Biol Chem* 276, 45113-45119. (2001).
- Romana, S. P., Poirel, H., Leconiat, M., Flexor, M. A., Mauchauffe, M., Jonveaux, P. *et al.* High frequency of t(12;21) in childhood B-lineage acute lymphoblastic leukemia. *Blood* 86, 4263-4269. (1995).
- Rosenwald, A., Alizadeh, A. A., Widhopf, G., Simon, R., Davis, R. E., Yu, X. *et al.* Relation of gene expression phenotype to immunoglobulin mutation genotype in B cell chronic lymphocytic leukemia. *J Exp Med* 194, 1639-1647. (2001).
- Rosenwald, A., Wright, G., Chan, W. C., Connors, J. M., Campo, E., Fisher, R. I. *et al.* The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *N Engl J Med* 346, 1937-1947. (2002).
- Rubinstein, P., Dobrila, L., Rosenfield, R. E., Adamson, J. W., Migliaccio, G., Migliaccio, A. R. *et al.* Processing and cryopreservation of placental/umbilical cord blood for unrelated bone marrow reconstitution. *Proc Natl Acad Sci U S A* 92, 10119-10122 (1995).

- Rubnitz, J. E., Pui, C. H. & Downing, J. R. The role of TEL fusion genes in pediatric leukemias. *Leukemia* 13, 6-13. (1999).
- Sage, J., Mulligan, G. J., Attardi, L. D., Miller, A., Chen, S., Williams, B. *et al.* Targeted disruption of the three Rb-related genes leads to loss of G(1) control and immortalization. *Genes Dev* 14, 3037-3050. (2000).
- Schadt, E. E., Li, C., Su, C. & Wong, W. H. Analyzing high-density oligonucleotide gene expression array data. *J Cell Biochem* 80, 192-202. (2000).
- Schotta, G., Ebert, A., Krauss, V., Fischer, A., Hoffmann, J., Rea, S. *et al.* Central role of Drosophila SU(VAR)3-9 in histone H3-K9 methylation and heterochromatic gene silencing. *Embo J* 21, 1121-1131. (2002).
- Schrapppe, M., Arico, M., Harbott, J., Biondi, A., Zimmermann, M., Conter, V. *et al.* Philadelphia chromosome-positive (Ph+) childhood acute lymphoblastic leukemia: good initial steroid response allows early prediction of a favorable treatment outcome. *Blood* 92, 2730-2741. (1998).
- Schrapppe, M., Reiter, A., Ludwig, W. D., Harbott, J., Zimmermann, M., Hiddemann, W. *et al.* Improved outcome in childhood acute lymphoblastic leukemia despite reduced use of anthracyclines and cranial radiotherapy: results of trial ALL-BFM 90. German-Austrian-Swiss ALL-BFM Study Group. *Blood* 95, 3310-3322. (2000a).
- Schrapppe, M., Reiter, A., Zimmermann, M., Harbott, J., Ludwig, W. D., Henze, G. *et al.* Long-term results of four consecutive trials in childhood ALL performed by the ALL-BFM study group from 1981 to 1995. Berlin-Frankfurt-Munster. *Leukemia* 14, 2205-2222. (2000b).
- Schubart, K., Massa, S., Schubart, D., Corcoran, L. M., Rolink, A. G. & Matthias, P. B cell development and immunoglobulin gene transcription in the absence of Oct-2 and OBF-1. *Nat Immunol* 2, 69-74. (2001).
- Schuler, G. D. Pieces of the puzzle: expressed sequence tags and the catalog of human genes. *J Mol Med* 75, 694-698. (1997).
- Schuler, G. D., Boguski, M. S., Stewart, E. A., Stein, L. D., Gyapay, G., Rice, K. *et al.* A gene map of the human genome. *Science* 274, 540-546. (1996).
- Sgroi, D. C., Teng, S., Robinson, G., LeVangie, R., Hudson, J. R., Jr. & Elkahoul, A. G. In vivo gene expression profile analysis of human breast cancer progression. *Cancer Res* 59, 5656-5661. (1999).
- Sherlock, G. Analysis of large-scale gene expression data. *Curr Opin Immunol* 12, 201-205 (2000).
- Shi, M. J. & Stavnezer, J. CBF alpha3 (AML2) is induced by TGF-beta1 to bind and activate the mouse germline Ig alpha promoter. *J Immunol* 161, 6751-6760. (1998).
- Shipp, M. A., Ross, K. N., Tamayo, P., Weng, A. P., Kutok, J. L., Aguiar, R. C. *et al.* Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. *Nat Med* 8, 68-74. (2002).
- Sigvardsson, M. Overlapping expression of early B-cell factor and basic helix-loop-helix proteins as a mechanism to dictate B-lineage-specific activity of the lambda5 promoter. *Mol Cell Biol* 20, 3640-3654. (2000).
- Silverman, L. B., Gelber, R. D., Dalton, V. K., Asselin, B. L., Barr, R. D., Clavell, L. A. *et al.* Improved outcome for children with acute lymphoblastic leukemia: results of Dana-Farber Consortium Protocol 91-01. *Blood* 97, 1211-1218. (2001).

- Simon, R., Radmacher, M. D., Dobbin, K. & McShane, L. M. Pitfalls in the use of DNA microarray data for diagnostic and prognostic classification. *J Natl Cancer Inst* 95, 14-18. (2003).
- Sorlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H. *et al.* Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* 98, 10869-10874. (2001).
- Staunton, J. E., Slonim, D. K., Coller, H. A., Tamayo, P., Angelo, M. J., Park, J. *et al.* Chemosensitivity prediction by transcriptional profiling. *Proc Natl Acad Sci U S A* 98, 10787-10792. (2001).
- Stec, I., Wright, T. J., van Ommen, G. J., de Boer, P. A., van Haeringen, A., Moorman, A. F. *et al.* WHSC1, a 90 kb SET domain-containing gene, expressed in early development and homologous to a Drosophila dysmorphia gene maps in the Wolf-Hirschhorn syndrome critical region and is fused to IgH in t(4;14) multiple myeloma. *Hum Mol Genet* 7, 1071-1082. (1998).
- Stegmeier, F., Visintin, R. & Amon, A. Separase, polo kinase, the kinetochore protein Slk19, and Spo12 function in a network that controls Cdc14 localization during early anaphase. *Cell* 108, 207-220. (2002).
- Sturn, A., Quackenbush, J. & Trajanoski, Z. Genesis: cluster analysis of microarray data. *Bioinformatics* 18, 207-208. (2002).
- Su, A. I., Cooke, M. P., Ching, K. A., Hakak, Y., Walker, J. R., Wiltshire, T. *et al.* Large-scale analysis of the human and mouse transcriptomes. *Proc Natl Acad Sci U S A* 99, 4465-4470. (2002).
- Su, G. H., Ip, H. S., Cobb, B. S., Lu, M. M., Chen, H. M. & Simon, M. C. The Ets protein Spi-B is expressed exclusively in B cells and T cells during development. *J Exp Med* 184, 203-214. (1996).
- Takahashi, M., Rhodes, D. R., Furge, K. A., Kanayama, H., Kagawa, S., Haab, B. B. *et al.* Gene expression profiling of clear cell renal cell carcinoma: Gene identification and prognostic classification. *Proc Natl Acad Sci U S A* 98, 9754-9759. (2001).
- Takeshita, A., Shinjo, K., Naito, K., Ohnishi, K., Higuchi, M. & Ohno, R. Erythropoietin receptor in myelodysplastic syndrome and leukemia. *Leuk Lymphoma* 43, 261-264. (2002).
- Tamayo, P., Slonim, D., Mesirov, J., Zhu, Q., Kitareewan, S., Dmitrovsky, E. *et al.* Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. *Proc Natl Acad Sci U S A* 96, 2907-2912 (1999).
- Tanaka, T. S., Jaradat, S. A., Lim, M. K., Kargul, G. J., Wang, X., Grahovac, M. J. *et al.* Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray. *Proc Natl Acad Sci U S A* 97, 9127-9132. (2000).
- Tang, X. X., Zhao, H., Robinson, M. E., Cnaan, A., London, W., Cohn, S. L. *et al.* Prognostic significance of EPHB6, EFNB2, and EFNB3 expressions in neuroblastoma. *Med Pediatr Oncol* 35, 656-658 (2000).
- Tanuma, N., Nakamura, K., Shima, H. & Kikuchi, K. Protein-tyrosine phosphatase PTPepsilon C inhibits Jak-STAT signaling and differentiation induced by interleukin-6 and leukemia inhibitory factor in M1 leukemia cells. *J Biol Chem* 275, 28216-28221. (2000).
- Thomas, J. G., Olson, J. M., Tapscott, S. J. & Zhao, L. P. An efficient and robust statistical modeling approach to discover differentially expressed genes using genomic expression profiles. *Genome Res* 11, 1227-1236. (2001).

- Tsai, R. Y. & Reed, R. R. Identification of DNA recognition sequences and protein interaction domains of the multiple-Zn-finger protein Roaz. *Mol Cell Biol* 18, 6447-6456. (1998).
- van den Berg, H. & van der Lelie, J. Acute lymphoblastic leukaemia in puberty and adolescence. *Ann Oncol* 11, 1375-1379. (2000).
- Van Gelder, R. N., von Zastrow, M. E., Yool, A., Dement, W. C., Barchas, J. D. & Eberwine, J. H. Amplified RNA synthesized from limited quantities of heterogeneous cDNA. *Proc Natl Acad Sci U S A* 87, 1663-1667. (1990).
- Velculescu, V. E., Madden, S. L., Zhang, L., Lash, A. E., Yu, J., Rago, C. *et al.* Analysis of human transcriptomes. *Nat Genet* 23, 387-388 (1999).
- Virtaneva, K., Wright, F. A., Tanner, S. M., Yuan, B., Lemon, W. J., Caligiuri, M. A. *et al.* Expression profiling reveals fundamental biological differences in acute myeloid leukemia with isolated trisomy 8 and normal cytogenetics. *Proc Natl Acad Sci U S A* 98, 1124-1129. (2001).
- Vorwerk, P., Wex, H., Hohmann, B., Mohnike, K., Schmidt, U. & Mittler, U. Expression of components of the IGF signalling system in childhood acute lymphoblastic leukaemia. *Mol Pathol* 55, 40-45. (2002).
- Vorwerk, P., Wex, H., Hohmann, B., Oh, Y., Rosenfeld, R. G. & Mittler, U. CTGF (IGFBP-rP2) is specifically expressed in malignant lymphoblasts of patients with acute lymphoblastic leukaemia (ALL). *Br J Cancer* 83, 756-760. (2000).
- Wang, L. C., Swat, W., Fujiwara, Y., Davidson, L., Visvader, J., Kuo, F. *et al.* The TEL/ETV6 gene is required specifically for hematopoiesis in the bone marrow. *Genes Dev* 12, 2392-2402. (1998).
- Wang, X., Ono, K., Kim, S. O., Kravchenko, V., Lin, S. C. & Han, J. Metaxin is required for tumor necrosis factor-induced cell death. *EMBO Rep* 2, 628-633. (2001).
- Wang, Y. H., Stephan, R. P., Scheffold, A., Kunkel, D., Karasuyama, H., Radbruch, A. *et al.* Differential surrogate light chain expression governs B-cell differentiation. *Blood* 99, 2459-2467. (2002).
- Warrington, J. A., Nair, A., Mahadevappa, M. & Tsyganskaya, M. Comparison of human adult and fetal expression and identification of 535 housekeeping/maintenance genes. *Physiol Genomics* 2, 143-147. (2000).
- Wiemels, J. L., Cazzaniga, G., Daniotti, M., Eden, O. B., Addison, G. M., Masera, G. *et al.* Prenatal origin of acute lymphoblastic leukaemia in children. *Lancet* 354, 1499-1503. (1999a).
- Wiemels, J. L. & Greaves, M. Structure and possible mechanisms of TEL-AML1 gene fusions in childhood acute lymphoblastic leukemia. *Cancer Res* 59, 4075-4082. (1999b).
- Willert, J., Epping, M., Pollack, J., Brown, P. & Nusse, R. A Transcriptional Response to Wnt protein in human embryonic carcinoma cells. *BMC Dev Biol* 2, 8. (2002).
- Wodicka, L., Dong, H., Mittmann, M., Ho, M. H. & Lockhart, D. J. Genome-wide expression monitoring in *Saccharomyces cerevisiae*. *Nat Biotechnol* 15, 1359-1367. (1997).
- Wu, A. G., Michejda, M., Mazumder, A., Meehan, K. R., Menendez, F. A., Tchabo, J. G. *et al.* Analysis and characterization of hematopoietic progenitor cells from fetal bone marrow, adult bone marrow, peripheral blood, and cord blood. *Pediatr Res* 46, 163-169. (1999).
- Wyllie, A. H. E2F1 selects tumour cells for both life and death. *J Pathol* 198, 139-141. (2002).

- Yeoh, E.-J., Ross, M. E., Shurtleff, S. A., Williams, W. K., Patel, D., Mahfouz, R. *et al.* Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. *Cancer Cell* 1, 133-143 (2002).
- Yeung, K. Y., Haynor, D. R. & Ruzzo, W. L. Validating clustering for gene expression data. *Bioinformatics* 17, 309-318. (2001).
- Yusoff, P., Lao, D. H., Ong, S. H., Wong, E. S., Lim, J., Lo, T. L. *et al.* Sprouty2 inhibits the Ras/MAP kinase pathway by inhibiting the activation of Raf. *J Biol Chem* 277, 3195-3201. (2002).
- Zendman, A. J., Cornelissen, I. M., Weidle, U. H., Ruiters, D. J. & van Muijen, G. N. CTp11, a novel member of the family of human cancer/testis antigens. *Cancer Res* 59, 6223-6229. (1999).
- Zhang, L., Zhou, W., Velculescu, V. E., Kern, S. E., Hruban, R. H., Hamilton, S. R. *et al.* Gene expression profiles in normal and cancer cells. *Science* 276, 1268-1272 (1997).
- Zhu, H., Kavsak, P., Abdollah, S., Wrana, J. L. & Thomsen, G. H. A SMAD ubiquitin ligase targets the BMP pathway and affects embryonic pattern formation. *Nature* 400, 687-693. (1999).
- Zhuang, Y., Soriano, P. & Weintraub, H. The helix-loop-helix gene E2A is required for B cell formation. *Cell* 79, 875-884. (1994).
- Zumkeller, W. & Burdach, S. The insulin-like growth factor system in normal and malignant hematopoietic cells. *Blood* 94, 3653-3657. (1999).

## 6.5 Publikationsliste

## Originalarbeiten

- Hattenhorst, U. E.**, A. Thum, M. S. Staeger, G. Hansen, S. Posch, G. Richter, S. Burdach. Gene Chip Assessment Of Differential Gene Expression in Pediatric cALL Versus Normal CD10+CD19+ Lymphocytes. Submitted.
- Staeger, M. S., V. Gorelov, A. Bulankin, U. Fischer, K. Dumon, L. Hohndorf, **U. E. Hattenhorst**, C. Kramm, S. Burdach. Stable Transgenic Expression of IL-2 and HSV1-tk by Single and Fusion Tumor Cell Lines Bearing EWS/FLI-1 Chimeric Genes. *Pediatr Hematol Oncol* 20, 119-140 (2003).
- Dirksen, U., **U. E. Hattenhorst**, P. Schneider, H. Schroten, U. Göbel, A. Böcking, K. M. Müller, R. Murray, S. Burdach. Defective expression of granulocyte-macrophage colony-stimulating factor/interleukin-3/interleukin-5 receptor common beta chain in children with acute myeloid leukemia associated with respiratory failure. *Blood* 92, 1097-1103 (1998).
- Dirksen, U., R. Nishinakamura, P. Groneck, **U. E. Hattenhorst**, L. Noguee, R. Murray, S. Burdach. Human pulmonary alveolar proteinosis associated with a defect in GM-CSF/IL-3/IL-5 receptor common beta chain expression. *J Clin Invest* 100, 2211-2217 (1997).
- Rad, M. R., B. Habbig, G. Jansen, **U. E. Hattenhorst**, M. Kroll, C. P. Hollenberg. Analysis of the DNA sequence of a 34,038 bp region on the left arm of yeast chromosome XV. *Yeast* 13, 281-286 (1997).
- Konferenzbeiträge
- Hattenhorst, U. E.**, S. Bergmann, M. Staeger, S. Burdach. Lymphoblast Biology And Leukemogenesis: New Insights From Comprehensive Gene Expression Profiling of Pediatric cALL. *44th Meeting of the American Society of Hematology, Philadelphia, USA. Blood* 100, 313a (2002).
- Hattenhorst, U. E.**, M. Staeger, A. Thum, S. Burdach. Mining Transcriptional Information for Immunotherapy of Childhood Leukemia. *Second Israeli-German Bi-National Conference: Current Concepts in Pediatric Oncology and Hematology, Berlin. Monatsschrift Kinderheilkunde* 150, 1291-1292 (2002).
- Staeger, M., **U. E. Hattenhorst**, M. Schrappe, S. Burdach. Leveraging Genome Information for Immunotherapy of Childhood Leukemia. *3rd biennial Hannover Symposium on Childhood Leukemia, Hannover. Tagungsband* (2002).
- Staeger, M., S. Körber, **U. E. Hattenhorst**, S. Burdach. Gene chip sustained accomplishments in tumor biology, molecular diagnosis and immunotherapy of Ewing tumors. *60. wissenschaftliche Halbjahrestagung der Gesellschaft für Pädiatrische Onkologie und Hämatologie, Berlin. Monatsschrift Kinderheilkunde* 150 (2002).
- Hansen, G., **U. E. Hattenhorst**, S. Burdach. Functional genomics and immune and gene therapy in pediatric oncology. *Gemeinsame Jahrestagung der DGHO und ÖGHO, Hannover. Tagungsband* (2001).

- Hattenhorst, U. E.**, R. Glynne, R. Murray, S. Burdach. Differential Gene Expression Analysis Utilizing DNA-Microchips In Common Acute Leukemia Versus Normal Bone Marrow. *13th International Symposium on Molecular Biology of Hematopoiesis, New York, USA. Acta Haematologica* 103, 43 (2000).
- Hattenhorst, U. E.**, E. Tom, R. Glynne, R. Murray, S. Burdach. Differential gene expression analysis in pediatric common acute lymphoblastic leukemia versus normal pre-B-cells and bone marrow utilizing DNA-microarrays. *42th Meeting of the American Society of Hematology, San Francisco, USA. Blood* 96, 107A (2000).
- Hattenhorst, U. E.**, S. Burdach. Novel Alterations in the Coding Sequence of the IL3/ IL5/ GM-CSF Receptors Common beta-Chain in Patients With Pulmonary Alveolar Proteinosis. *28<sup>th</sup> Annual Meeting of the International Society for Experimental Hematology, Monte Carlo, Monaco. Exp Hematol* 27, 101 (1999).
- Dirksen, U., **U. E. Hattenhorst**, H. Schrotten, A. Böcking, R. Nishinakamura, R. Murray, S. Burdach. Defective GM-CSF/IL3/IL5 receptor common beta chain expression with or without AML associated with respiratory insufficiency. *Med Pediatr Oncol* 30, 137 (1998).
- Hattenhorst, U. E.**, U. Dirksen, S. Burdach. Alternate Splicing in IL3/IL5/GM-CSF Receptor Common beta-Chain Expression. *39th Meeting of the American Society of Hematology, USA . Blood* 90 (1997).

## Danksagung

Mein besonderer Dank gilt Herrn Prof. Dr. S. Burdach, für die Aufnahme in seine Arbeitsgruppe und für die Überlassung des Themas, sowie für die anhaltenden Diskussionen und Anregungen. Weiterhin bedanke ich mich bei Prof. Dr. G. Reuter für das Interesse an meiner Arbeit, sowie für seine ständige Diskussionsbereitschaft und Unterstützung.

Der gesamten Arbeitsgruppe des Forschungszentrums für krebskranke Kinder gilt mein herzlicher Dank für das gute Arbeitsklima. Insbesondere danke ich Dr. G. Richter und Dr. M. Staeger für viele hilfreiche Diskussionen und Anregungen, Dr. F. Meyer-Wentrup und PD Dr. G. Hansen nicht zuletzt für die Unterstützung bei der Klärung vieler medizinischer Fragen, Dr. O. Diwan für die Hilfestellung bei Fragen zur Durchflusszytometrie, S. Bergmann und I. Volkmer schließlich für ihre ausdauernde Hilfsbereitschaft bei allen kleinen und großen Fragen rund um den Laboralltag.

Besonders bedanke ich mich bei A. Thum und Prof. Dr. S. Posch, die nicht müde wurden, mir die Bioinformatik näher zu bringen. Ich danke ihnen für die vielfältige Unterstützung bei der Auswertung der Genexpressionsdaten.

Bei Dr. E. Tom und Dr. R. Murray möchte ich mich für die gute Kooperation und den Aufenthalt in South San Francisco bedanken, bei dem ich die Möglichkeit hatte die Technik der DNA-Microarray-gestützten Genexpressionsanalyse von Grund auf zu erlernen.

Ich danke den Professoren F. Röpke, D. Körholz, M. Schrappe und der Studiengruppe ALL-BFM der GPOH für die Bereitstellung und Überlassung von Probenmaterial gesunder Spender und an Leukämie erkrankter Kinder.

Ganz besonders danke ich meinen Eltern, die mir mein Studium ermöglicht haben und mir immer geholfen haben, meinen Weg zu gehen. Ich danke meiner gesamten Familie und meinen Freunden, die auch in den weniger guten Zeiten nicht aufgehört haben, an mich zu glauben.

Ich danke Felix, dem Sturm in meinem Leben, *ma tempête*.

# Lebenslauf

## Persönliche Daten

Name, Vorname	Hattenhorst, Uwe Eduard
Anschrift	Johann-Andreas-Segner-Str. 1, 06108 Halle
Geburtsdatum	01.07.1966
Geburtsort	Wuppertal
Nationalität	deutsch
Familienstand	ledig

## Schulbildung

1972-1976	Besuch der Gemeinschaftsgrundschule Ricarda-Huch-Straße in Düsseldorf
1976-1986	Besuch des Gymnasiums Koblenzerstraße in Düsseldorf

## Wissenschaftlicher Werdegang

1988-1995	Studium der Biologie an der Heinrich-Heine-Universität in Düsseldorf
1995	Diplomarbeit „Sequenzierung und Charakterisierung eines 5,3kb großen Fragmentes vom linken Arm des Chromosoms XV der Hefe <i>Saccharomyces cerevisiae</i> “ im Rahmen des „European Yeast Genome Project“ am Institut für Mikrobiologie, Prof. Dr. C. Hollenberg
1996-1998	Wissenschaftlicher Angestellter im Rahmen eines Forschungsprojektes zur „Charakterisierung der gemeinsamen beta-Kette der Rezeptoren für IL3, IL5 und GM-CSF“ im Labor für experimentelle Hämatologie und Stammzelltransplantation in der Kinderklinik der Medizinischen Einrichtungen der Heinrich-Heine-Universität bei Prof. Dr. Stefan Burdach
seit 1998	Wissenschaftlicher Angestellter und Laborleiter im Forschungszentrum für Krebskranke Kinder der Universitätsklinik und Poliklinik für Kinder- und Jugendmedizin an der Martin-Luther-Universität Halle-Wittenberg, Direktor Prof. Dr. Stefan Burdach
2000-2003	Promotion im Forschungszentrum für Krebskranke Kinder der Universitätsklinik und Poliklinik für Kinder- und Jugendmedizin an der Martin-Luther-Universität Halle-Wittenberg, Direktor Prof. Dr. Stefan Burdach

Halle, den 10. März 2003

## Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Arbeit selbständig und ohne fremde Hilfe verfasst habe, nur die angegebenen Quellen und Hilfsmittel verwendet habe und die den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Halle, den 10. März 2003