Entwicklung und Implementierung von Auswertungswerkzeugen für Hochdurchsatz-DNA-Kopienzahl-Analysen und deren Anwendung auf Lymphomdaten

Dissertation

zur Erlangung des akademischen Grades
Dr. rer. med.

an der Medizinischen Fakultät
der Universität Leipzig

Eingereicht von:
Dipl. Inf. Markus Kreuz
Geb. 02.05.1980 Leipzig

Angefertigt am:
Institut für Medizinische Informatik, Statistik und Epidemiologie der Universität Leipzig

Betreuer:
Prof. Dr. Markus Löffler

Beschluss über die Verleihung des Doktorgrades vom: 24. 02. 2015

Bibliografische Beschreibung:

Kreuz, Markus

Entwicklung und Implementierung von Auswertungswerkzeugen für Hochdurchsatz
DNA-Kopienzahl-Analysen und deren Anwendung auf Lymphomdaten

Universität Leipzig, Dissertation

119 S., 73 Lit., 13 Abb., 1 Tab.

Referat:

Aberrationen in der DNA-Kopienzahl sind häufige genetische Veränderungen bei malignen Lymphomerkrankungen. Zugewinne sowie Deletionen stellen dabei Mechanismen zur Onkogen-Aktivierung sowie Tumorsuppressorgen-Inaktivierung dar und tragen somit zur Pathogenese der Erkrankung bei. Array-CGH und SNP-Array sind Messplattformen, die die genomweite Bestimmung von Kopienzahlaberrationen in einem Experiment ermöglichen. Die bei der Analyse entstehenden Datensätze sind komplex und erfordern automatische Methoden zur Unterstützung der Analyse und Interpretation der Messergebnisse.

In dieser Promotionsarbeit wurden Methoden entwickelt, welche die Analyse von Array-CGH- und SNP-Array-Messungen ermöglichen. Diese Methoden wurden für die Auswertung umfangreicher Datensätze von malignen Non-Hodgkin-Lymphomen verwendet. Dabei wurden Lymphome der Entitäten Burkitt-Lymphom, diffus großzelliges B-Zell-Lymphom, Mantelzelllymphom, primäres ZNS-Lymphom und peripheres T-Zell-Lymphom – nicht anderweitig spezifiziert – analysiert. Für die untersuchten Lymphom-Entitäten konnten hierbei zahlreiche neue rekurrente Kopienzahlaberrationen sowie uniparentale Disomien gezeigt werden, die neue Einblicke in die Pathogenese der jeweiligen Erkrankungen erlauben.

Darüber hinaus erfolgte ein Vergleich beider Messplattformen anhand eines Datensatzes mit gepaarten Array-CGH- und SNP-Array-Daten. Für die eingesetzten Plattformen (2800k-BAC-Array vs. Affymetrix 250k-Sty-SNP-Array) konnte eine circa zwölffach höhere effektive Auflösung der SNP-Array-Plattform gezeigt werden. Die wesentlichen Ergebnisse dieser Arbeit sind in sieben Publikationen eingeflossen.

Inhaltsverzeichnis

Abkürzungsverzeichnis

ABC	Eine Subgruppe diffus großzelliger B-Zell-Lymphome mit einem Genexpressionsmuster, das aktivierten B-Zellen ähnlich ist [1] (**a**ctivated **B-c**ell type)
aCGH	Array-CGH (Microarray-basierte komparative genomische Hybridisierung)
BL	Burkitt-Lymphom
Chr	Chromosom
CGH	komparative genomische Hybridisierung (**c**omparative **g**enomic **h**ybridization)
CNP	Kopienzahl-Polymorphismus (**c**opy **n**umber **p**olymorphism)
DLBCL	diffus großzelliges B-Zell Lymphom (**d**iffuse **l**arge **B-c**ell **l**ymphoma)
FDR	False Discovery Rate
FISH	Fluoreszenz-in-situ-Hybridisierung
GCB	Eine Subgruppe diffus großzelliger B-Zell-Lymphome mit einem Genexpressionsmuster, das Keimzentrums-B-Zellen ähnlich ist [1] (**g**erminal **c**enter **B**-cell type)
HMM	Hidden-Markov-Model
ICGC	International Cancer Genome Consortium [2] (Internationales, interdisziplinäres Projekt zur Untersuchung der molekularen Ursachen von 50 verschiedenen Krebserkrankungen)
IPI	Internationaler Prognostischer Index (International Prognostic Index)
Kb	Kilobasen
LOH	Verlust der Heterozygotie (**l**oss **o**f **h**eterozygosity)
Mb	Megabasen
mBL	Lymphom, das auf Basis seines Genexpressionsprofils als Burkitt-Lymphom klassifiziert wurde [3] (**m**olecular **B**urkitt **l**ymphoma)
MAD	Median der absoluten Abweichungen (**m**edian **a**bsolute **d**eviation)
MCL	Mantelzelllymphom (**m**antle **c**ell **l**ymphoma)

MCR	Bereich innerhalb einer rekurrent aberranten Genomregion, in der die Inzidenz der Aberration und damit die Anzahl der betroffenen Tumorproben am größten ist (**m**inimal **c**ommon **r**egion)
MMML	Molekulare Mechanismen in malignen Lymphomen (von der Deutschen Krebshilfe gefördertes Verbundprojekt)
NGS	Next Generation Sequencing
PAR	Pseudoautosomale Region
PAR1	Pseudoautosomale Region 1
PAR2	Pseudoautosomale Region 2
PCR	Polymerase-Kettenreaktion (**p**olymerase **c**hain **r**eaction)
PTCL-NOS	Peripheres T-Zell-Lymphom – nicht anderweitig spezifiziert – (**p**eripheral **T**-**c**ell **l**ymphoma – **n**ot **o**therwise **s**pecified –)
PZNSL	Primäres ZNS-Lymphom
SNP	Einzelnukleotid-Polymorphismus (**s**ingle **n**ucleotide **p**olymorphism)
TCGA	The Cancer Genome Atlas (Projekt zur Charakterisierung von Mutationen bei Krebserkrankungen)
UPD	Uniparentale Disomie
ZNS	Zentralnervensystem

Tabellenverzeichnis

Abbildungsverzeichnis

1 Einführung

Veränderungen der DNA-Kopienzahl spielen bei unterschiedlichen Erkrankungen, insbesondere bei der Pathogenese von malignen Tumorerkrankungen, eine wichtige Rolle [4]. Die Länge der betroffenen Genomregionen kann dabei von wenigen Basenpaaren bis zu ganzen Chromosomen reichen. Veränderungen der Kopienzahl können die Genexpression der betroffenen Gene beeinflussen. So können Deletionen zur Inaktivierung von Tumorsuppressorgenen und Zugewinne zur Aktivierung von Onkogenen führen [5]. Durch umfangreiche genomische Analysen an verschiedenen Tumorentitäten konnten Regionen identifiziert werden, die mit hoher Rekurrenz Kopienzahlaberrationen in den jeweiligen Tumorentitäten zeigten [6]. Teilweise führten diese Analysen zur Identifikation von Onko- und Tumorsuppressorgenen und ermöglichen somit perspektivisch neue, spezifischere Therapieoptionen [6].

Array-CGH und SNP-Array sind Messplattformen, die die genomweite Detektion von Kopienzahlveränderungen in einer einzigen Analyse ermöglichen. Die bei den Messungen erzeugten Datensätze sind umfangreich und die Messergebnisse können durch eine Vielzahl von Störgrößen beeinflusst werden. Daher werden automatische Methoden benötigt, die die Auswertung der Datensätze unterstützen und dabei für bekannte Störgrößen adjustieren. Neben der Analyse einzelner Messungen werden auch Methoden zur gemeinsamen Auswertung mehrerer Messexperimente und Methoden für die integrative Auswertung mit anderen klinischen und molekularen Daten, beispielsweise gepaarten Genexpressionsmessungen, benötigt.

Die Ergebnisse, die im Rahmen dieser Promotionsschrift präsentiert werden, umfassen die Entwicklung von Analysemethoden zur Auswertung von Kopienzahldatensätzen auf Basis von Array-CGH- und SNP-Array-Messungen. Die verschiedenen Methoden werden dabei auf umfangreiche Lymphomdatensätze angewendet. Es erfolgt die Beschreibung der daraus resultierenden Analyseergebnisse sowie deren biologische und klinische Bedeutung. Zusätzlich werden beide Messplattformen verglichen und es werden Methoden für die Assoziation der Kopienzahldaten mit anderen molekulargenetischen und klinischen Annotationen der Lymphome vorgestellt. Im ersten Teil der Arbeit (Abschnitt 1.1) erfolgt dazu eine kurze Einführung zur Thematik der Kopienzahlaberrationen und der untersuchten Tumorentität der Lymphome. Im Anschluss (Abschnitt 1.2) wird die Motivation für die Arbeit dargelegt. Die folgenden Abschnitte 1.3 und 1.4 beschreiben die Methoden für die Analyse der Array-CGH- und SNP-Array-Daten und deren Applikation auf Lymphomdatensätze. In Abschnitt 1.5 wird ein Vergleich beider Plattformen anhand der Ergebnisse von gepaarten Array-CGH- und SNP-Array-Messungen vorgenommen. Daran schließt sich eine Beschreibung von Methoden zur Assoziation von Kopienzahl- und Genexpressionsdaten sowie zur Assoziation mit dichotomen oder stetigen Merkmalen an (Abschnitt 1.6). Die Abschnitte 2.1– 2.7 beinhalten Publikationen, in die Ergebnisse dieser Promotionsarbeit eingeflossen sind. Die wichtigsten Ergebnisse der Publikationen sind jeweils auf einer einleitenden Seite zusammengefasst. Der geleistete Eigenanteil an diesen Publikationen ist in Abschnitt 2.8 dargestellt. Abschließend werden die Resultate in Kapitel 3 diskutiert und ein Ausblick in die zukünftige Rolle der Kopienzahlanalyse gewagt. Kapitel 4 fasst die wichtigsten Ergebnisse dieser Promotionsarbeit kurz zusammen.

1.1 Biologischer Hintergrund

1.1.1 Aberrationen der DNA-Kopienzahl und Tumorentstehung

Somatische Mutationen sind ungerichtete Veränderungen des Erbgutes, die spontan auftreten oder durch äußere Einflüsse verursacht werden. Sie können negative Auswirkungen haben und zum Tod der betroffenen Zelle führen oder Krankheiten des Organismus verursachen [7]. Es gilt als nachgewiesen, dass somatische Mutationen die Ursache von Tumorerkrankungen sind [8]. Die für die Tumorentstehung relevanten Gene werden in zwei Gruppen unterteilt. Proto-Onkogene können durch Mutationen in einen aktivierten Zustand („gain-of-function") versetzt werden und so zur Tumorentstehung beitragen. Die aktivierte Form wird als Onkogen bezeichnet [8]. Tumorsuppressorgene bilden die zweite Gruppe. Bei diesen Genen sind Mutationen, die zum Verlust der Funktion („loss-of-function") führen, mit der Tumorentstehung assoziiert. Während Mutationen in einem Proto-Onkogen dominante Effekte auf die betroffene Zelle haben können, sind die Mutationen in Tumorsuppressorgenen im allgemeinen rezessiv, sodass beide Allele betroffen sein müssen, damit eine Auswirkung auf den Phänotyp der Zelle sichtbar wird [8].

Je nach Art und Größe der von der Mutation betroffenen Genomregion unterscheidet man Gen-, Chromosomen- und Genommutationen [9]. Chromosomenmutationen führen zu Veränderungen in größeren Chromosomenabschnitten und werden im nachfolgenden Absatz näher beschrieben. Man unterscheidet Verluste (Deletionen) und Zugewinne (Duplikationen) von Chromosomenabschnitten. Beiden Mutationstypen ist eine Veränderung der vorhandenen Gendosis gemeinsam, die Auswirkungen auf den Phänotyp haben kann. Die Änderung der Gendosis kann beispielsweise zur Änderung der Expression und somit der Aktivität von betroffenen Genen führen. Folglich können diese Aberrationen zur Aktivierung von Proto-Onkogenen sowie zur Inaktivierung von Tumorsuppressorgenen führen. Sind beide Allele von einer Deletion betroffen, spricht man von einer homozygoten Deletion, beim Zugewinn vieler Kopien von einer (High-Level-)Amplifikation. Weitere Chromosomenmutationen sind Translokationen und Inversionen. Hier kommt es nicht zu einer Veränderung der Gendosis. Bei Translokationen erfolgt die Umlagerung eines Chromosomenabschnittes an ein anderes Chromosom, was zur Entstehung von Fusionsproteinen oder fehlerhafter Regulation der Genexpression führen kann. Bei Inversion erfolgt der inverse Einbau eines Chromosomenteils nach doppeltem Bruch des Chromosoms [9]. Der Verlust eines elterlichen Allels führt zum Verlust der Heterozygosität (Loss Of Heterozygosity, LOH). Für die betroffenen Gene ist folglich nur noch die Erbinformation eines Elternteils verfügbar. Die Deletion eines Allels führt automatisch auch zu LOH in der betroffenen Region, allerdings kann LOH auch ohne Veränderung der Kopienzahl auftreten, wenn gleichzeitig das verbleibende elterliche Allel dupliziert wird. In diesem Fall spricht man von uniparentaler Disomie (UPD). Deletionen und UPD stellen Mechanismen zur Inaktivierung von Tumorsuppressorgenen dar. Oftmals erfolgt die Deaktivierung dabei in zwei Schritten. Das Tumorsuppressorgen wird zunächst durch eine Punktmutation in einem Allel inaktiviert. Anschließend wird die intakte Genkopie durch UPD oder Deletion entfernt [8].

Die im Rahmen dieser Promotionsarbeit betrachteten Messplattformen, Array-CGH und SNP-Array, ermöglichen die Detektion von Chromosomenmutationen, die zu einer Veränderung der Gendosis führen: Deletionen und Duplikationen. Darüber hinaus ermöglicht die SNP-Array-Plattform die Detektion von uniparentalen Disomien, respektive kopienzahlneutralen LOH. Die Detektion dieser Aberrationen in Tumorproben kann die

Grundlage für die Identifikation von neuen Onko- und Tumorsuppressorgenen sein. Erschwerend ist jedoch, dass die Tumorzellen häufig genetisch instabil sind und dadurch viele genetische Veränderungen akkumulieren [8]. Die in einem Tumor entdeckten Mutationen sind daher teilweise zufällige Nebenprodukte der erhöhten genetischen Instabilität. Erst durch die Messungen von mehreren Tumorfällen können rekurrent betroffene Genomregionen identifiziert werden, die aller Voraussicht nach relevante Gene für die Tumorentstehung umfassen.

1.1.2 Lymphome

Maligne Lymphome sind Tumorerkrankungen, die durch Lymphozyten und deren Vorläuferzellen hervorgerufen werden [10]. Lymphozyten sind weiße Blutkörperchen und umfassen neben natürlichen Killerzellen auch B- und T-Zellen. Die beiden letzteren bilden die zelluläre Basis des adaptiven Immunsystems [11].

B- und T-Zellen entstehen beim Menschen aus Blutstammzellen im Knochenmark, wobei die Progenitorzellen der B-Lymphozyten im Knochenmark (**B**onemarrow) und die der T-Lymphozyten im **T**hymus reifen. B-Zellen ermöglichen über ihre Ausdifferenzierung zu Plasmazellen die Produktion von Antikörpern und somit die humorale Immunantwort. Es besteht auch die Möglichkeit der Differenzierung von B-Lymphozyten zu Gedächtniszellen, die bei erneutem Kontakt mit dem gleichen Antigen eine schnellere und stärkere Immunantwort ermöglichen. Zur Bildung hochaffiner Antikörper durchlaufen B-Zellen in ihrer Entwicklung mehrere Schritte. Zunächst wird über eine genetische Umlagerung, die sogenannte V(D)J-Rekombination, genetische Variabilität in die Region eingeführt, die den B-Zell-Rezeptor bzw. die Antikörper kodiert. Durch das Durchlaufen dieses Zufallsprozesses wird gewährleistet, dass im menschlichen Organismus circa 10^{12} verschiedene Antikörper produziert werden können, ohne das es Kontakt zu einem Antigen gab [11]. Im Rahmen der Affinitätsreifung im Keimzentrum wird diese Zahl noch deutlich erhöht. Dort erfolgt durch die somatische Hypermutation eine weitere Veränderung der genetischen Information der B-Zelle. Die Affinität des B-Zell-Rezeptors bzw. der Antikörper wird durch eine künstlich erhöhte Mutationsrate in der Genomregion, die den B-Zell-Rezeptor kodiert, verändert. Gekoppelt mit einem Selektionsprozess, der B-Zellen mit erhöhter Affinität selektiert, kann der Organismus so mit der Entwicklung und Produktion hochaffiner Antikörper auf bedrohliche Antigene reagieren.

T-Lymphozyten können sich zu verschiedenen Klassen von T-Zellen ausdifferenzieren, die verschiedene Aufgaben in der Immunantwort übernehmen. Sie können als Cytotoxische T-Zellen infizierte Zellen zerstören, als T-Helferzellen über die Ausschüttung von Zytokinen kostimulatorisch auf andere Zellen des Immunsystems wirken oder als Regulatorische T-Zellen die Aktivität anderer Lymphozyten unterdrücken und somit die Selbsttoleranz des Immunsystems gewährleisten [11]. Bei der Entwicklung der T-Zellen erfolgt ebenfalls eine V(D)J-Rekombination, wobei die Affinität des T-Zell-Rezeptors verändert wird, der den T-Zellen das Erkennen von Antigenen ermöglicht.

Sowohl B- als auch T-Zellen können neoplastische Transformation durchleben und so zur Entstehung von Leukämien oder Lymphomen führen [12]. Dabei spielen Fehler in den genetischen Prozessen zur Veränderung der B- und T-Zell Rezeptoren oftmals eine zentrale Rolle. Bei den Lymphomen unterscheidet man zunächst zwischen Hodgkin- und Non-Hodgkin-Lymphomen. Hodgkin-Lymphome sind durch einen bestimmten abnormalen Zelltyp, die sogenannten Sternberg-Reed-Riesenzellen gekennzeichnet. Bei Non-Hodgkin-Lymphomen (NHL) treten diese Zellen nicht auf. Im Rahmen dieser Arbeit wurden verschiedene Entitäten der Non-Hodgkin-Lymphome auf Kopienzahlveränderungen der DNA untersucht. Non-Hodgkin-Lymphome können in B-Zell- und T-Zell-Lymphome eingeteilt

werden, je nachdem welchen zellulären Ursprung sie haben. Die Charakteristiken der Tumorzellen spiegeln dabei häufig das Entwicklungsstadium der Zelle wieder, aus welcher der Tumor entstanden ist. Seit 2001 erfolgt die Klassifikation der Lymphomerkrankungen anhand der WHO-Klassifikation. Für die Klassifikation werden neben dem zellulären Ursprung, der Differenzierungsgrad der Zellen sowie morphologische, klinische, immunphänotypische, molekularbiologische und zytogenetische Merkmale berücksichtigt [13, 14]. Die aktuelle WHO-Klassifikation der Lymphome wurde 2008 in [14] publiziert. Im Rahmen dieser Arbeit werden B-Zell-Lymphome der Typen Mantelzelllymphom (Manuskript 2 (Abschnitt 2.2) und Manuskript 3 (Abschnitt 2.3)), Burkitt-Lymphom (Manuskript 6 (Abschnitt 2.6)), diffus großzelliges B-Zell-Lymphom (Manuskript 7 (Abschnitt 2.7)) und primäres ZNS-Lymphom (Manuskript 4 (Abschnitt 2.4)) sowie T-Zell-Lymphome des Typs peripheres T-Zell-Lymphom – nicht anderweitig spezifiziert – (Manuskript 5 (Abschnitt 2.5)) analysiert. Eine umfassende Charakterisierung dieser Lymphomentitäten kann im Rahmen dieser Arbeit nicht erfolgen. In den betreffenden Publikationen und in der zugehörigen Zusammenfassung sind jedoch die wesentlichen klinischen und molekulargenetischen Eigenschaften der untersuchten Lymphomentitäten beschrieben.

1.2 Motivation und Rationale für die Arbeit

Bei der Messung mit Hochdurchsatztechnologien wie Array-CGH und SNP-Arrays entstehen umfangreiche Datensätze mit Millionen von Messpunkten. Die Auswertung dieser Datensätze ist entsprechend komplex und erfordert die Unterstützung durch automatische Auswertewerkzeuge. Je nach verwendeter Messplattform gliedert sich die Auswertung dabei in verschiedene Analyseschritte, die mit entsprechenden Algorithmen unterstützt werden müssen.

Für beide Plattformen sind Methoden zur Qualitätsanalyse und Normalisierungsschritte zur Korrektur von systematischen Messfehlern erforderlich, um die Vergleichbarkeit der Messergebnisse über mehrere Fälle zu gewährleisten. Da die Kopienzahlaberrationen zusammenhängende Genomregionen umfassen, sind oftmals mehrere benachbarte Messsonden betroffen. Basierend auf der Annahme, dass die Messfehler für jede Sonde unabhängig und gleichverteilt sind, werden für beide Technologien Glättungs- oder Segmentierungsverfahren benötigt. Dadurch können zusammenhängende Regionen mit gemeinsamer zugrundeliegender Kopienzahl selektiert und so das Signal-zu-Rausch-Verhältnis verbessert werden [15]. Anschließend müssen die ermittelten Segmente klassifiziert werden, wobei die dabei verwendeten Schwellwerte von Faktoren wie Tumorzellgehalt oder Hybridisierungsqualität der Chips abhängen und in der Regel in Abhängigkeit von verwendeter Messplattform und analysierter Tumorentität gewählt werden. Bei der Analyse mehrerer Tumorfälle sollte für beide Plattformen die Häufigkeit der auftretenden Kopienzahlaberrationen analysiert werden. Regionen, die häufig von konkordanten Aberrationen betroffen sind, bezeichnet man als rekurrent. Diese sollten automatisch selektiert werden und können anschließend in weitere Analysen einfließen, beispielsweise in die Korrelation der DNA-Kopienzahl mit der Expression der Gene in der betroffenen Region.

Ziel dieser Arbeit ist es, Analysepipelines für Array-CGH- und SNP-Array-Daten bereitzustellen. Im Rahmen des Verbundprojektes „Molekulare Mechanismen in Malignen Lymphomen (MMML)" erfolgten umfangreiche Kopienzahl-Messungen an malignen B-Zell-Lymphomen. Dabei wurden für n=514 Lymphome CGH-Arrays hybridisiert. Für n=192 Tumore erfolgte die Messung mit Affymetrix-SNP-Arrays. Im Rahmen dieser Arbeit sollten diese Datensätze mit den entwickelten Analysepipelines ausgewertet werden. Darüber hinaus erfolgte ein Vergleich der Messergebnisse beider Plattformen anhand von n=96 Fällen mit gepaarten Messungen. Dabei wurde die Auflösung beider Plattformen unter Berücksichtigung der individuellen Messwertstreuung beider Plattformen verglichen.

Die Auswertepipeline für SNP-Array-Daten wurde zusätzlich auf umfangreiche Datensätze anderer Lymphomentitäten angewendet, darunter Mantelzelllymphome (n=32), Primäre ZNS-Lymphome (n=19) und periphere T-Zell-Lymphome – nicht anderweitig spezifiziert – (n=49).

Für die Analysen gilt es, neben der deskriptiven Beschreibung der auftretenden Kopienzahlaberrationen auch den Zusammenhang zwischen der Inzidenz der Aberrationen und weiteren Charakteristiken der Lymphome bzw. Patienten zu betrachten. Ein besonderer Schwerpunkt liegt dabei auf der gemeinsamen Analyse von DNA-Kopienzahl und Genexpression. Hierfür werden Methoden benötigt, die für rekurrente Kopienzahlaberrationen konkordant über- bzw. unterexprimierte Gene selektieren können. Transkripte, die einen entsprechenden Gendosiseffekt zeigen, sind als putative Onko- bzw. Tumorsuppressorgene geeignete Kandidaten für weitere funktionale Analysen.

1.3 Array-CGH-Analyse

Motivation

Vergleichende genomische Hybridisierung (comparative genomic hybridization, CGH) ist eine Methode zur Messung von Kopienzahlveränderungen im Genom. Dabei wird eine untersuchte Gewebeprobe, beispielsweise Tumorgewebe, mit einer euploiden Referenzprobe verglichen. Für Tumor- und Referenzprobe wird jeweils DNA extrahiert, unterschiedlich gefärbt und anschließend gemischt. Bei klassischen chromosomalen CGH-Messungen erfolgt die Hybridisierung des DNA-Gemischs in einer kompetitiven Reaktion gegen Metaphasechromosomen. Dabei konkurrieren DNA-Fragmente aus Tumor- und Referenzprobe um freie Bindungsstellen. Anschließend werden die Farbmuster der Chromosomen mittels Floureszensmikroskopie und CCD-Kamera bestimmt und mit geeigneter Software ausgewertet. Dabei ermöglicht das Farbverhältnis von Tumor- und Referenzfärbung Rückschlüsse auf Kopienzahlveränderungen im Tumor. Für Genomregionen mit normaler Kopienzahl im Tumor ist ein ausgeglichenes Verhältnis der beiden Farbstoffe zu erwarten. Für Regionen, bei denen der Tumor eine Deletion aufweist, ist jedoch weniger komplementäre Tumor-DNA im Probengemisch. Dadurch ist weniger passende Tumor-DNA vorhanden, die um die freien Bindungsstellen an den Chromosomen konkurrieren kann. Entsprechend ist der Anteil an bindender Referenz-DNA größer, was sich in einem Ungleichgewicht bei der Färbung der entsprechenden Genomregionen widerspiegelt, wobei der Farbstoff der Referenz-DNA überwiegt. Im umgekehrten Fall eines Kopien-Zugewinns im Tumor wird entsprechend der Farbstoff der Tumor-DNA überwiegen.

Bei der Matrix- oder Array-CGH-Technologie werden anstatt der Metaphasechromosomen DNA-Sonden auf einem Chip verwendet. Dabei werden DNA-Klone, die jeweils einer definierten Genomregion zugeordnet werden können, auf einem Chip fixiert (gespottet). Bei modernen Chips können auf engstem Raum eine Vielzahl von Sonden untergebracht werden. Für jeden dieser Spots kann nach der Reaktion das Farbverhältnis von gebundener Tumor- und Referenz-DNA bestimmt werden. Das Farbverhältnis ermöglicht Rückschlüsse auf die Kopienzahl des Tumors an der zugehörigen Genomposition. Durch die hohe Anzahl an Sonden pro Array kann so eine deutlich höhere Auflösung als bei chromosomaler CGH erreicht werden, sodass kleinere Aberrationen detektiert bzw. die Grenzen von Aberrationen genauer bestimmt werden können. Als Ergebnis eines Array-CGH-Experiments liegen für jeden gemessenen Tumorfall Farbverhältnisse für alle auf dem Array gespotteten Sonden vor. Die Farbintensitäten werden dabei quantifiziert und das Verhältnis liegt anschließend als Quotient von Tumor- und Referenzintensität vor bzw. wird mit log2 auf die logarithmische Skala transformiert. Die im Rahmen dieser Arbeit analysierten Chips umfassten circa 2800 Klonpositionen, wobei aktuelle Arrays deutlich höhere Sondenanzahlen aufweisen können.

Methodik

Im Rahmen der Arbeit wurde das Auswertewerkzeug *aCGHPipeline* konzipiert und mit der Statistiksoftware R [16] implementiert (vgl. Abschnitt 2.1, Publikation 1). Das Auswertewerkzeug erlaubt zunächst das Einlesen von Array-CGH-Daten sowie der zugehörigen Annotation für die gemessenen Fälle und CGH-Klone des verwendeten Arrays. Anschließend erfolgt eine Qualitätsanalyse, bei der für alle gemessenen Fälle und alle gemessenen CGH-Klone Auffälligkeiten bezüglich der Streuung sowie Häufungen von Fehlwerten deskriptiv analysiert werden. Hierbei kann auch ein automatischer Abgleich der Klone mit bekannten Kopienzahl-Polymorphismen (CNP) erfolgen. Kopienzahlveränderungen

in diesen Regionen stellen eine natürliche Varianz der Kopienzahl dar [17]. Entsprechende Veränderungen sind häufig nicht mit Tumorentstehung oder Progression assoziiert.

Durch Unterschiede in den Mengen bzw. der Qualität der für die Hybridisierung verwendeten Tumor- und Kontroll-DNA können chipspezifische Verzerrungen der Messwerte entstehen. Folgende Gleichung illustriert den Effekt, wobei T die Intensität des Tumorsignals, K die Intensität der euploiden Kontrolle und q der chipspezifische Verzerrungsfaktor ist:

$$\log_2\left(\frac{T}{q*K}\right) = \log_2\left(\frac{T}{K}\right) + \log_2\left(\frac{1}{q}\right) = \log_2\left(\frac{T}{K}\right) + c$$

aCGHPipeline erlaubt es, den Verzerrungsparameter c zu schätzen und die Daten chipweise zu normalisieren (vgl. Abschnitt 2.1, Publikation 1).

Nach der Normalisierung der Daten können die Messwerte $x_{i,j}$ für Array i und Klon j mit folgendem Modell beschrieben werden:

$$x_{i,j} = \log_2\left(\frac{(\Gamma_i(c_{i,j}^T - c_{i,j}^N) + c_{i,j}^N)p_{i,j}^{aber} + c_i^N(1 - p_{i,j}^{aber})}{c_{i,j}^N}\right) + \varepsilon_{i,j}$$

$$p_{i,j}^{aber} = \left(1 - a_i^N\right)\left(1 - t_{i,j}^N\right)$$

Hierbei ist $c_{i,j}^T$ die Kopienzahl des Tumors für Array i und Klon j, $c_{i,j}^N$ entsprechend die Kopienzahl der normalen Kontrollprobe an dieser Position. $p_{i,j}^{aber}$ ist der Anteil aberranter Zellen in der Tumorprobe, wobei der Anteil durch den chipspezifischen Anteil an normalen/euploiden Zellen in der Tumorprobe (a_i^N) sowie den Anteil an Tumorzellen, die die Aberration nicht aufweisen ($t_{i,j}^N$), bestimmt ist. Γ_i ist ein chipspezifischer Dämpfungsfaktor und $\varepsilon_{i,j}$ der Messfehler.

Das Signal-zu-Rausch-Verhältnis für Array-CGH-Messungen ist gering [15]. Die Ursache dafür liegt zum einen in den relativ hohen Messfehlern ($\varepsilon_{i,j}$). Zum anderen werden die Signalausschläge für Kopienzahlaberrationen häufig durch Beimengung von Normalzellen in den Tumorproben (a_i^N) und durch unspezifische Kreuzhybridisierung (Γ_i) gedämpft. Zusätzlich können die Tumore inhomogen sein, sodass der Tumor aus mehreren Subklonen besteht, die sich in ihrem Kopienzahlprofil unterscheiden. Trägt ein großer Anteil der Tumorzellen die Aberration nicht ($t_{i,j}^N$ ist entsprechend groß), dann wird das Signal ebenfalls gedämpft.

Da Kopienzahlaberrationen häufig mehrere benachbarte Klone umfassen, können Glättungsfilter oder Segmentierungsverfahren eingesetzt werden, um die Streuung der Messwerte zu reduzieren. Bei Segmentierungsverfahren erfolgt eine Einteilung der Messwerte in Regionen mit gleicher zugrunde liegender Kopienzahl. Für jedes ermittelte Segment wird die Kopienzahl einzeln robust geschätzt, beispielsweise über den Mittelwert oder Median aller zugehörigen Messwerte. In *aCGHPipeline* sind Schnittstellen zu zwei verschiedenen Segmentierungsverfahren implementiert: einmal auf Basis von Hidden-Markov-Modellen (Fridlyand et al.) [18], sowie auf Basis zirkulärer binärer Segmentierung (Olshen et al.) [19]. Die aus der Segmentierung resultierenden Segmente können anschließend mit *aCGHPipeline* bezüglich der vermuteten Kopienzahl im Tumor klassifiziert

werden. Dabei erfolgt die Einteilung in Segmente mit normaler Kopienzahl, mit Kopie-Zugewinnen und mit Deletionen.
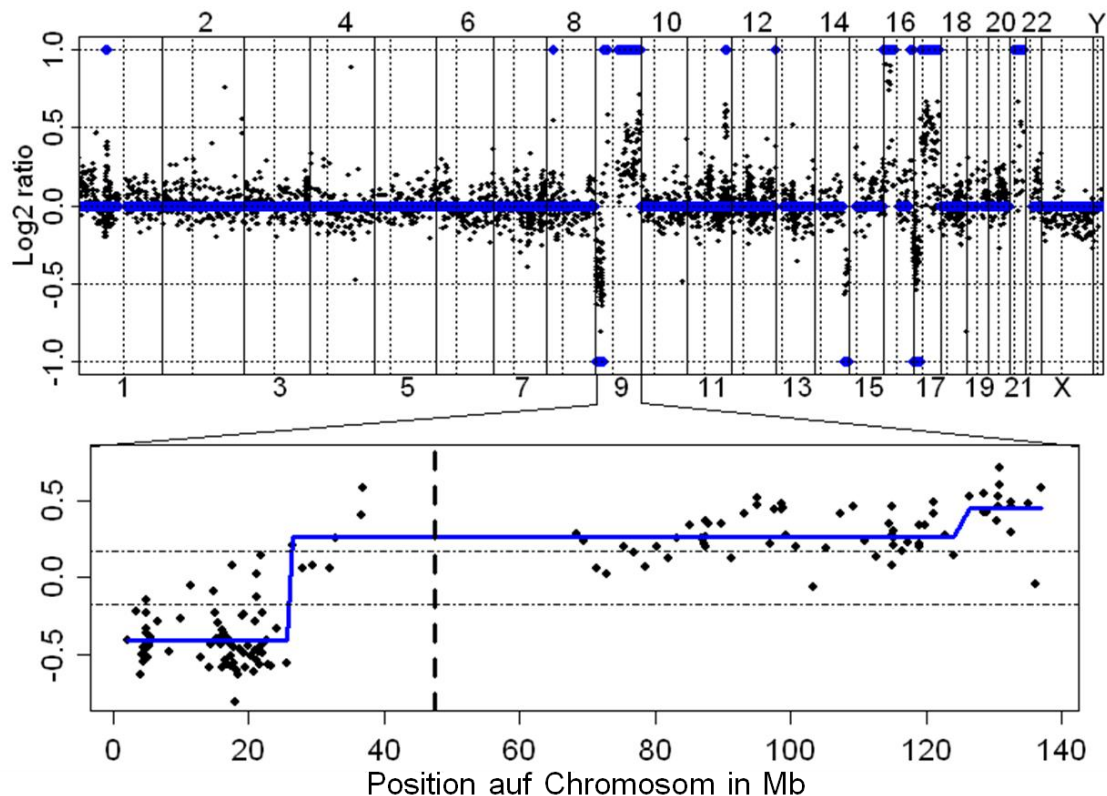

Abbildung 1: Darstellung des Kopienzahlprofils für ein Lymphom. Im oberen Teil der Grafik sind alle Messwerte in genomischer Reihenfolge auf der log2ratio-Skala dargestellt (schwarze Punkte). Die Nummerierung auf der X-Achse gibt dabei die zugehörigen Chromosomen an. Vertikale schwarze Linien zeigen die Chromosomengrenzen und vertikale Strichlinien die Positionen der Zentromere an. Blaue Markierungen bei 1 zeigen Zugewinne, bei –1 Deletionen an. Markierungen bei 0 stellen die normale Kopienzahl dar. Im unteren Teil der Grafik ist Chromosom 9 im Detail dargestellt. Schwarze Punkte zeigen die Rohmesswerte auf der log2ratio-Skala an. Die blaue Linie zeigt das Ergebnis der Segmentierung, das die Basis für die Klassifikation im oberen Teil der Grafik ist. Die horizontalen schwarzen Linien markieren die Schwellwerte für die Klassifikation. Segmente oberhalb der oberen Linie werden als Zugewinne, unterhalb der unteren Linie als Deletionen klassifiziert. Segmente zwischen beiden Linien werden als Regionen mit normaler Kopienzahl gewertet

Die Klassifikation kann dabei mittels festen Schwellwerten oder alternativ mit streuungsabhängigen Schwellwerten erfolgen. Bei letzterem Verfahren führt stärkere Streuung in den Messwerten zu höheren und damit konservativeren Schwellwerten. Das Ergebnis von Segmentierung und Klassifikation ist für einen Lymphomfall exemplarisch in Abbildung 1 dargestellt.

Für viele Fragestellungen erfolgt die Analyse mehrerer Tumore der gleichen Entität. Dabei soll die Inzidenz der Aberrationen in der Tumorentität bestimmt werden. Die Detektion von Regionen, die in der untersuchten Tumorentität häufig von gleichartigen Aberrationen betroffen sind, ist dabei von zentralem Interesse. Diese häufig aberranten oder auch rekurrenten Regionen umfassen vermutlich Gene, die eine entscheidende Rolle bei der Entstehung oder Progression der untersuchten Tumore spielen. Seltene Aberrationen treten im Gegensatz dazu meist zufällig auf und resultieren aus der erhöhten genomischen Instabilität der Tumorzellen. Das Paket *aCGHPipeline* umfasst eine Methode zur automatischen Detektion von rekurrenten Regionen. Dabei werden die Häufigkeiten von

Zugewinnen und Deletionen für alle gemessenen Tumore über alle Klone aggregiert. Die daraus entstehenden Häufigkeiten von Zugewinnen und Deletionen werden mittels einer Hidden-Markov-Modell-basierten Methode segmentiert. Anschließend werden die entstehenden Segmente basierend auf ihrer zugrunde liegenden Häufigkeit in rekurrente und sporadische Regionen klassifiziert. Abschließend kann für jede rekurrente Region in jedem untersuchten Tumor geprüft werden, ob der Tumor diese Aberration aufweist. Das erfolgt über die Analyse des Anteils an aberranten Klonen innerhalb der rekurrenten Regionen und individuell für jeden gemessenen Tumorfall. Die dabei ermittelte Klassifikation kann anschließend für weitere Analysen genutzt werden. Typische Fragestellungen sind dabei die prognostische oder diagnostische Bedeutung des Vorliegens bestimmter Aberration oder auch deren Einfluss auf die Genexpression der zur Region gehörigen Gene. Die automatische Detektion von rekurrenten Regionen und die Klassifikation der einzelnen Tumorfälle bezüglich des Vorliegens der Aberration ist exemplarisch in Abbildung 2 gezeigt. Abschließend bietet *aCGHPipeline* eine Reihe von Maßzahlen für die Quantifizierung der genomischen Instabilität der untersuchten Tumorfälle und umfasst Methoden, die eine gemeinsame Auswertung mit gepaarten Genexpressionsdatensätzen unterstützen.

Anwendung

aCGHPipeline ist inzwischen auf eine Vielzahl verschiedener Array-CGH-Datensätze angewendet worden. Im Rahmen des MMML-Verbundprojektes wurden mittels *aCGHPipeline* 514 Lymphomproben untersucht. Die Ergebnisse dieser Analysen sind dabei in eine Vielzahl von Publikationen eingeflossen (vgl. Manuskripte 1 und 7, Quellen [20-26]). Darüber hinaus erfolgte die Analyse eines Datensatzes von 53 klarzelligen Nierenzellkarzinomen mit *aCGHPipeline*. Hierbei lag der Fokus der Analyse auf dem Zusammenhang von Kopienzahlaberrationen und dem Metastasierungsverhalten der Tumore [27]. Im Rahmen des deutschen Gliomnetzwerkes (German Glioma Network) wird *aCGHPipeline* für Analysen verwendet, bei denen die Kopienzahlprofile von Primärtumoren und gepaartem Rezidivtumor verglichen werden [28].
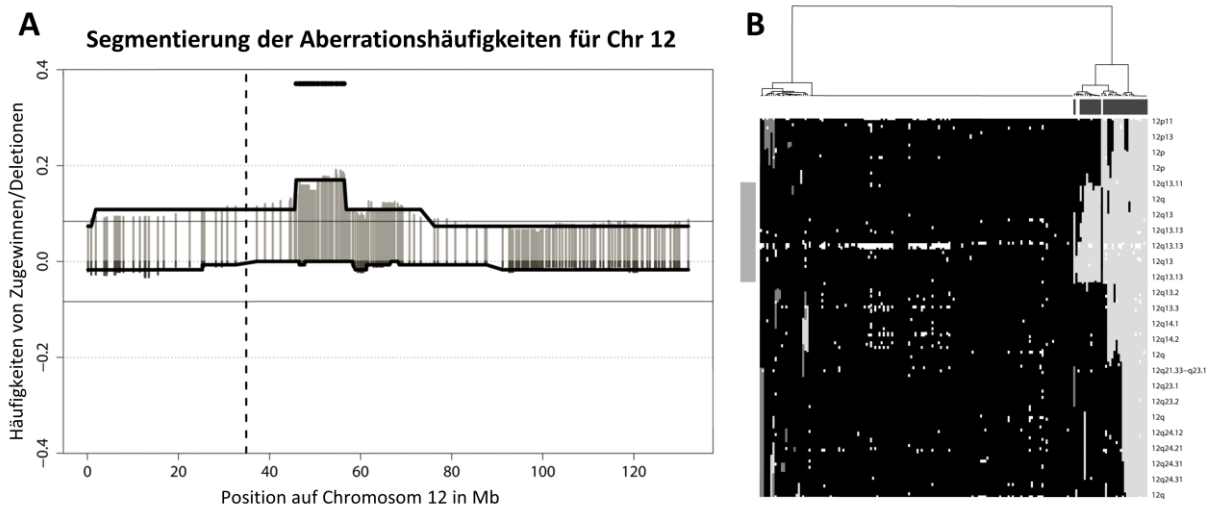
Abbildung 2: Automatische Detektion von rekurrenten Regionen exemplarisch für Chromosom 12 eines analysierten Lymphomdatensatzes (vgl. Manuskript 1).

(A) Für jeden Klon sind die Häufigkeiten von Zugewinnen (hellgrau) und Deletionen (dunkelgrau) bei allen 183 untersuchten Tumoren dargestellt. Die schwarzen Linien zeigen die Segmentierung in Regionen von benachbarten Klonen mit gleichen Aberrationshäufigkeiten, getrennt für Zugewinne und Deletionen. Die gestrichelte vertikale Linie gibt die Position des Zentromers an. Die horizontalen Geraden bei circa –0,1 und +0,1 geben die Schwelle für die Klassifikation in rekurrente und sporadische Aberrationen an. Während Deletionen auf Chromosom 12 selten sind, gibt es häufige Zugewinne. Eine Region bei circa 50 Mb zeigt in annähernd 20 % der untersuchten Lymphomproben Zugewinne (vgl. schwarze Markierung).

Diese Region ist in **(B)** für alle 183 Lymphome untersucht worden. Die Heatmap zeigt die Messwerte aller 183 Lymphome für Chromosom 12 nach erfolgter Klassifizierung in Deletion, normale Kopienzahl und Zugewinn. Jede Spalte zeigt einen Lymphomfall, wobei diese bezüglich Ähnlichkeit geclustert sind. Jede Zeile gibt die Messwerte eines Klones an, wobei die Klone entsprechend ihrer Position auf Chromosom 12 geordnet sind (von oben beginnend). Auf der rechten Seite der Heatmap ist zur Orientierung für einzelne Klone die Position auf dem untersuchten Chromosom angegeben. Hellgraue Färbung zeigt Zugewinne, dunkelgraue Färbung Deletionen an. Schwarze Färbung steht für normale Kopienzahl und weiße für Fehlwerte, beispielsweise durch schlechte Qualität für einzelne Klonmesswerte. Die Markierung am linken Rand der Heatmap gibt die Position der untersuchten rekurrenten Region an (vgl. schwarze Markierung in (A)). Die schwarze Markierung am oberen Rand kennzeichnet alle Lymphomfälle, bei denen der rekurrente Zugewinn tatsächlich vorliegt. Für die Klassifikation wurde gefordert, dass mindestens 50 % der Klone innerhalb der rekurrenten Region im jeweiligen Fall einen Zugewinn aufweisen mussten

1.4 SNP-Array-Analyse

Motivation

SNP-Chips sind DNA-Microarrays, die es ermöglichen, den Genotyp von Einzelnukleotid-Polymorphismen (Single Nucleotide Polymorphism, SNP) zu bestimmen. SNPs stellen Variationen einzelner Basenpaare an einer Genomposition innerhalb einer Spezies dar. Bei diploiden Genomen gibt es für jeden SNP in einem Individuum drei mögliche Ausprägungen: homozygot A/A, homozygot B/B oder heterozygot A/B. Dabei repräsentieren A und B jeweils eine der Basen Adenin (A), Guanin (G), Cytosin (C) und Thymin (T), die an der entsprechenden Genomposition vorkommen können. Die im Rahmen dieser Arbeit analysierten SNP-Arrays ermöglichen dabei die parallele Messung von mehreren 100.000 SNP-Positionen. Es kamen 100k (50k Xba und 50k Hind) und 500k (250k Sty und 250k Nsp) SNP-Arrays der Firma Affymetrix zum Einsatz.

Die Hybridisierung einer Probe auf dem SNP-Array und die nachfolgende Analyse läuft in mehreren Schritten ab (vgl. Abbildung 3) [29]. Zunächst wird die DNA der untersuchten Probe extrahiert und mittels Restriktionsenzymen verdaut. An die Schnittstellen der entstehenden Fragmente werden anschließend Adapter ligiert. Die dabei entstehenden Segmente werden danach mittels PCR amplifiziert, wobei präferentiell Fragmente bestimmter Länge amplifiziert werden. Im Anschluss werden die amplifizierten Segmente fragmentiert und gelabelt. Danach erfolgt die Hybridisierung auf dem SNP-Array. Auf dem SNP-Array sind für jede zu messende SNP-Position verschiedene Oligonukleotide gespottet, die jeweils 25 Basen umfassen und als Probeset bezeichnet werden. Die Probesets umfassen dabei Proben, die komplett komplementär zu je einer möglichen Basen-Ausprägung der SNP-Position sind. Diese werden PM-Proben (Perfect-Match) genannt. Darüber hinaus gibt es auch MM-Proben (Mismatch), die zu beiden möglichen Basenausprägungen mindestens eine Mismatchbase unter den 25 Nukleotiden aufweisen. Die MM-Proben sollen damit der Abschätzung von unspezifischer Kreuzhybridisierung dienen.

Nach mehreren Färbe- und Waschschritten [30] werden abschließend die Signalintensitäten ermittelt. Dazu werden die Arrays gescannt und die daraus resultierenden Bilddaten mittels geeigneter Software ausgewertet. Für jede SNP-Position können anschließend die Intensitäten der PM-Proben für Genoytp A und B und die Intensitäten der MM-Proben analysiert werden, um den Genotyp des zugehörigen SNPs in der gemessenen Probe zu bestimmen. Für die Kopienzahlanalyse wird eine Tumor-DNA-Probe auf einem SNP-Array hybridisiert und anschließend die Intensität aller Proben eines Probesets aggregiert. Die ermittelte Intensität wird anschließend mit den Intensitäten von mehreren euploiden Kontrollfällen verglichen. Die Messung dieser Kontrollen erfolgt dabei auf unabhängigen SNP-Arrays der gleichen Plattform. Deletionen im Tumor führen zu reduzierten Mengen an komplementärer DNA für alle Proben der betroffenen Probesets, Zugewinne führen zu einer erhöhten Menge an passenden DNA Strängen. Entsprechend beobachtet man für Deletionen reduzierte und für Zugewinne erhöhte Intensitäten für die jeweils betroffenen Probesets.
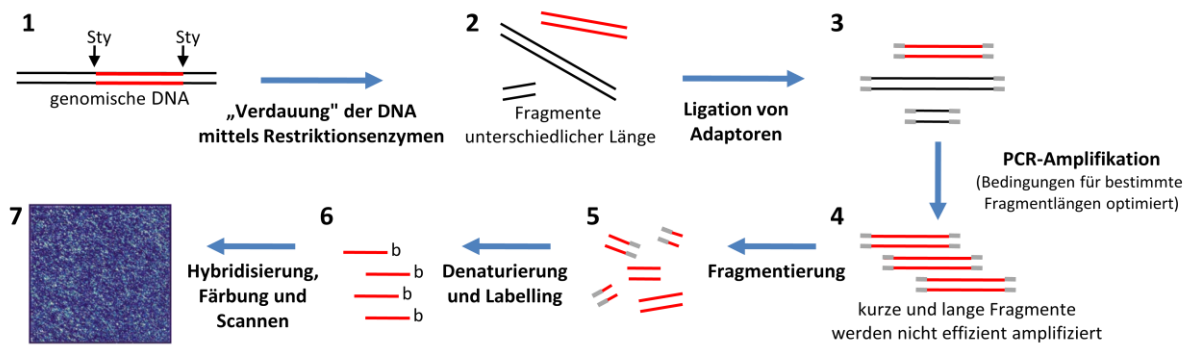
Abbildung 3: Ablauf einer SNP-Array-Messung: Mittels Restriktionsenzymen wird die DNA verdaut. An die entstehenden Fragmente werden anschließend Adaptoren ligiert. Mittels PCR erfolgt anschließend die Amplifikation. Abschließend werden die Stränge gelabelt, fragmentiert und auf einem Array hybridisiert

Methodik

Im Rahmen der Arbeit wurde eine Auswertepipeline für die Kopienzahlanalyse mit SNP-Array-Daten entwickelt und auf mehrere Lymphomdatensätze angewendet. Grundlage für die Analyse stellen gemessene SNP-Arrays der Tumorproben und Messungen von euploiden Kontrollen dar. Dabei sollten Tumorproben und euploide Kontrollen am besten im selben Labor gemessen werden, da ansonsten laborspezifische Messunterschiede zu erhöhter Streuung in den Kopienzahlsignalen führen (vgl. Abschnitt 1.5). Die komplette Analysepipeline ist in Abbildung 4 schematisch dargestellt.

In einem ersten Analyseschritt werden Tumor- und Kontrollproben gemeinsam genotypisiert. Dafür wurde im Rahmen dieser Arbeit der BRLMM-Algorithmus verwendet [31]. Die Call-Rate, der Anteil erfolgreich bestimmter Genotypen, ist dabei ein erstes wichtiges Kriterium für die Hybridisierungsqualität der Chips.

Basierend auf den ermittelten Genotypdaten erfolgt die Bestimmung von Regionen mit Loss of Heterozygosity (LOH). Dafür wird eine auf Hidden-Markov-Modellen basierte Methode eingesetzt [32], die in der Software dChip [33, 34] implementiert ist. Die Methode ermöglicht die Detektion von LOH-Regionen in Tumorfällen ohne gepaarte euploide Kontrollproben vom selben Patienten. Die Methode basiert auf der Selektion von zusammenhängenden Genomregionen, in denen keine heterozygoten Genotypen auftreten bzw. deren beobachtete Häufigkeit deutlich unter der erwarteten Häufigkeit liegt. Für die Segmentierung und Selektion putativer LOH-Regionen wird ein Hidden-Markov-Modell verwendet, das die genomischen Abstände von benachbarten SNPs, die SNP-spezifische Häufigkeit von heterozygoten Genotypen und die zugrundeliegende Haplotypstruktur im Genom berücksichtigt [32]. Die Sensitivität der Methode ist dabei abhängig von der Anzahl bzw. der Dichte der gemessenen SNP-Marker. Sie ist jedoch geringer als bei einem direkten Vergleich von Tumorgenotypen mit gepaarten Keimbahn-Daten vom selben Patienten. Hierbei sucht man nach Regionen, in denen heterozygote Genotypen in der Keimbahn-Kontrolle in der gepaarten Tumorprobe homozygot werden. Für viele Analysen werden jedoch aus Kostengründen und aufgrund der Materialverfügbarkeit keine gepaarten Keimbahn-Kontrollen gemessen, sodass die beschriebene Methode in diesen Fällen eine gute Alternative darstellt.
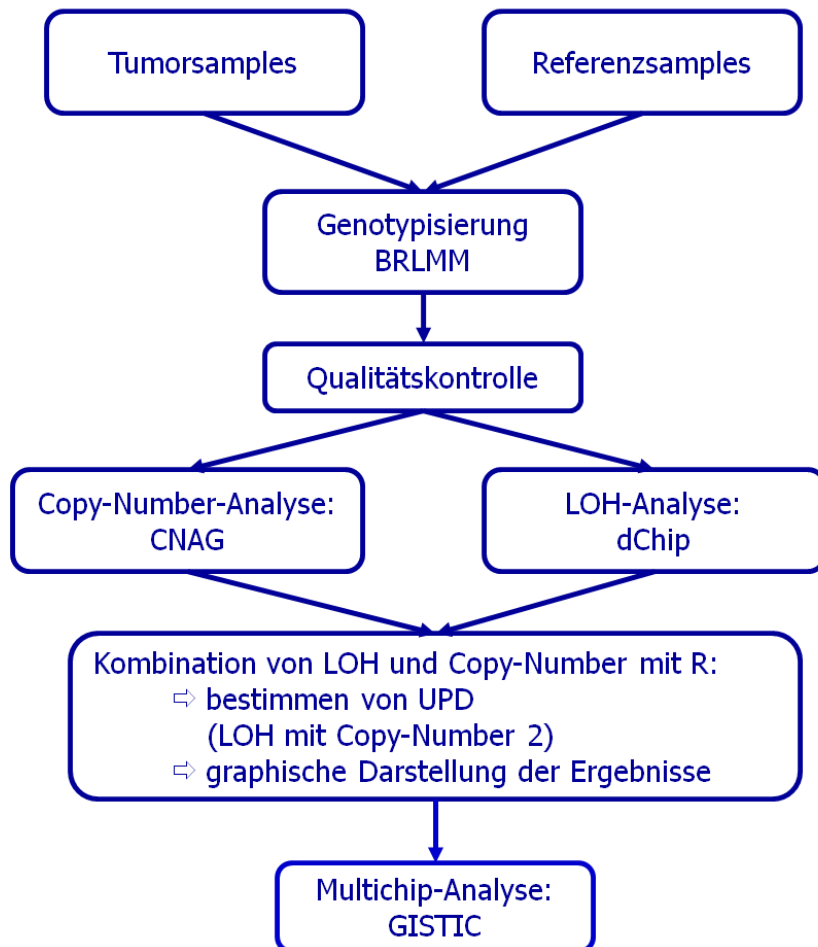
Abbildung 4: Schematischer Überblick über die einzelnen Schritte der SNP-Array-Analyse und die verwendeten Methoden für jeden Analyseschritt

Die Analyse der DNA-Kopienzahl erfolgt mit CNAG [35]. Diese Methode ermöglicht für jeden Tumorfall die Selektion eines individuellen Sets von bis zu 10 euploiden Kontrollen. Dabei erfolgt die Auswahl der Kontrollen mit dem Ziel, dass die Streuung des Kopienzahlsignals des Tumorfalles minimiert wird. Die Signalintensität hängt stark von der Länge der PCR-Fragmente und deren Anteil an Guanin und Cytosin (GC-Anteil) ab, da diese Parameter die Amplifikationskinetik der PCR-Reaktion beeinflussen [36]. Die Art des Einflusses kann je nach Laborbedingungen und PCR-Parametern variieren. Im Rahmen der Analyse eines Mantelzelllymphomdatensatzes (vgl. Manuskripte 2 und 3; Abschnitte 2.2 und 2.3) konnte dieser Effekt gezeigt werden (vgl. Abbildung 5 undAbbildung 6). Da in CNAG zusätzlich eine Normalisierung für die Fragmentlänge und den GC-Gehalt der gemessenen SNPs erfolgt, kann eine weitere Reduktion der Messwertstreuung erreicht werden. Die ermittelten Kopienzahlsignale werden anschließend mittels eines Hidden-Markov-Modells analysiert [35]. Die Zustände des Hidden-Markov-Modells bilden dabei die DNA-Kopienzahl im Tumor ab, wobei die Menge möglicher Zustände 0, 1, 2, 3, 4, 5 und ≥6 DNA-Kopien im Tumor umfasst. Durch die Anpassung des Hidden-Markov-Modells erfolgt für jeden analysierten Tumorfall die Segmentierung in zusammenhängende Segmente. Hierbei ist jedem Segment ein Zustand und somit eine geschätzte DNA-Kopienzahl zugeordnet. Als Ergebnis erhält man für jeden Tumorfall und jeden gemessenen SNP des Arrays einen geschätzten Rohwert für die Kopienzahl auf der log2ratio-Skala sowie ein zugehöriges Segment und damit eine Schätzung der Kopienzahl im Tumor an der Genomposition des SNPs.
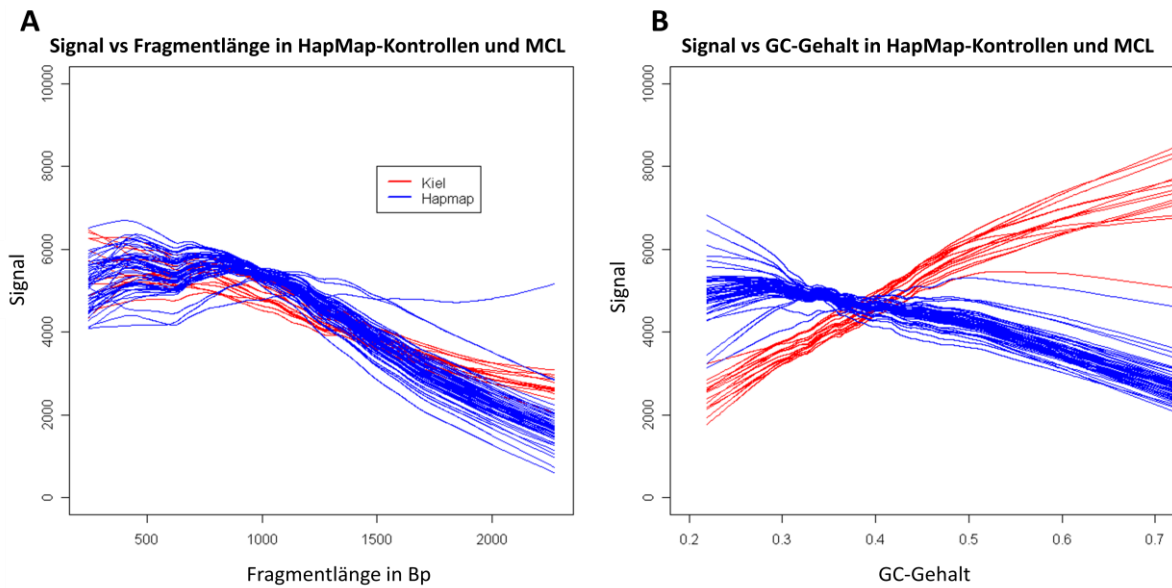
Abbildung 5: (A) Einfluss von Fragmentlänge und **(B)** GC-Gehalt auf die Signalintensität. Blaue Kurven markieren die HapMap-Kontrollen, rote Kurven die Mantelzelllymphome. Während der Zusammenhang zwischen Signalintensität und Fragmentlänge (A) in beiden Gruppen gleichförmig ist (längere Fragmente führen zu reduzierter Signalintensität), ist der Zusammenhang mit dem GC-Gehalt in den Gruppen gegenläufig (B). Das führt bei der Ermittlung der Kopienzahlprofile durch den Vergleich der Lymphomfälle mit den Kontrollen zu erhöhter Streuung (vgl. Abbildung 6)
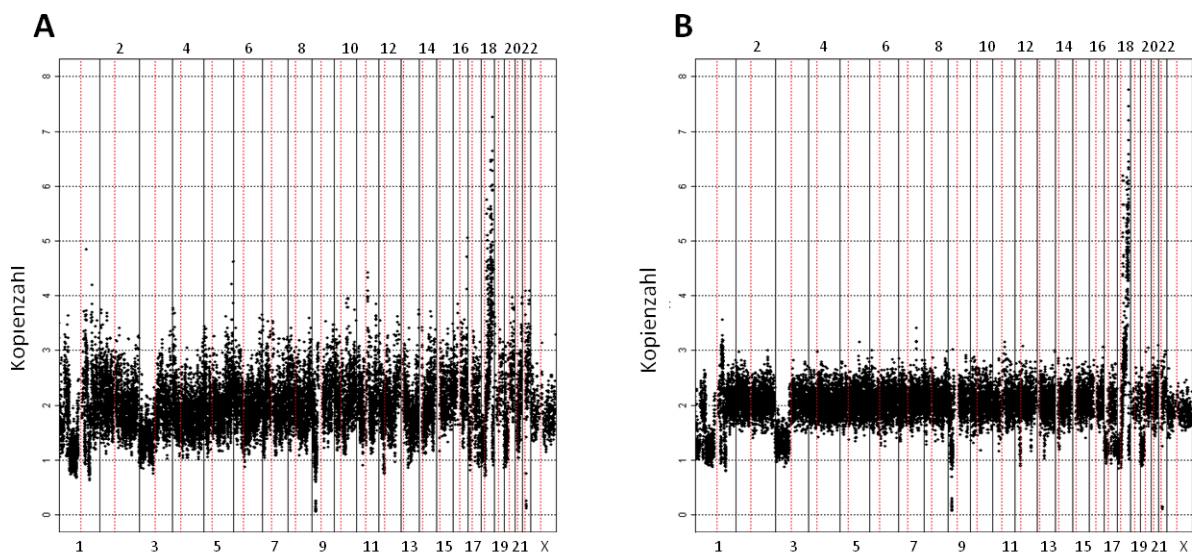
Abbildung 6: Kopienzahlanalyse für die Zelllinie Granta519. In Grafik **(A)** ist das Kopienzahlprofil der Zelllinie basierend auf der Analyse mit dChip [33, 34] dargestellt. Die robust mit medianer absoluter Abweichung (MAD) geschätzte Streuung beträgt dabei 0,46. Im Abschnitt **(B)** erfolgte die Korrektur für PCR-Fragmentlänge sowie GC-Gehalt. Dadurch konnte die Streuung um circa 40 % reduziert werden (MAD=0,27). Damit wird die anschließende Segmentierung und Klassifizierung deutlich erleichtert.

Die Ergebnisse von LOH und Kopienzahlanalyse werden anschließend in der Statistiksoftware R zusammengeführt. Das ermöglicht die Bestimmung von uniparentalen Disomien (UPD), also Regionen, die LOH aufweisen und gleichzeitig normale Kopienzahl zeigen. Des Weiteren kann auf Basis der Hidden-Markov-Modell-basierten Segmentierung der Daten eine Klassifikation in High-Level-Amplifikationen (Zugewinne von sehr vielen Kopien), einfache Zugewinne, Deletionen und homozygote Deletionen (Verluste von beiden Kopien) erfolgen.

14

Bei der Standardkonfiguration werden Segmente mit HMM Zustand=0 als homozygote Deletionen, Zustand=1 als einfache Deletionen und Zustand=2 als normal klassifiziert. Segmente mit den Zuständen 3 und 4 werden als einfache/schwache Zugewinne und Segmente mit den Zuständen 5 und 6 als High-Level-Amplifikationen zusammengefasst. Darüber hinaus können kurze aberrante Segmente, die nur wenige benachbarte SNPs umfassen, als mögliche technische Artefakte gefiltert werden. Abschließend können die detektierten Aberrationen und deren Überlapp mit bekannten Kopienzahlpolymorphismen in tabellarischer Form exportiert und die Kopienzahlprofile für Einzelfälle sowie aggregiert für alle Tumorfälle bzw. Subgruppen grafisch in HTML-Form dargestellt werden. Die Daten können entsprechend der Vorgaben der Software GISTIC [37] exportiert werden. Dadurch kann GISTIC verwendet werden, um rekurrente Aberrationen zu bestimmen.

Anwendung

Die hier beschriebene Auswertepipeline wurde im Rahmen des MMML-Verbundprojektes auf 39 Burkitt-Lymphome angewendet. Die Klassifikation der Tumore erfolgte dabei mittels Genexpressionsdaten. In vorherigen Arbeiten des Verbundes konnte gezeigt werden, dass die Abgrenzung von molekularen Burkitt-Lymphomen (mBL), d. h. mittels Transkriptomdaten klassifizierten Burkitt-Lymphomen, zuverlässig funktioniert und gut reproduzierbar ist [3]. Die Ergebnisse dieser Arbeit sind in Manuskript 6 (vgl. Abschnitt 2.6) dargestellt. Ein bedeutendes Ergebnis der Analyse war die Detektion von homozygoten Deletionen in der Region Chr1:23538204-23657473 für zwei mBL-Fälle. Diese Region umfasst zwei Gene, *E2F2* und *ID3*. In späteren Analysen konnte gezeigt werden, dass das Gen *ID3* in circa 70 % der Burkitt-Lymphome durch Mutationen oder homozygote Deletionen beschädigt ist [26, 38]. Dies deutet darauf hin, dass die Inaktivierung von *ID3* ein bis dahin unbekannter, zentraler Mechanismus bei der Entstehung des Burkitt-Lymphoms ist.

In einer weiteren Analyse mit der beschriebenen Pipeline wurden im Rahmen des MMML-Verbundes neben mBL auch diffus großzellige B-Zell-Lymphome untersucht und dabei eine bisher unbekannte rekurrente Deletion in der Region 19p13 aufgedeckt [39]. In gleicher Weise erfolgte die Analyse von SNP-Array-Daten für Mantelzelllymphome, primäre ZNS-Lymphome und periphere T-Zell-Lymphome (nicht anderweitig spezifiziert). Die Ergebnisse dieser Analysen sind in den Manuskripten 2, 3, 4 und 5 zusammengefasst (vgl. Abschnitte 2.2, 2.3, 2.4 und 2.5). Darüber hinaus wurden SNP-Array-Daten einer Familie mit Prädispositionssyndrom für Rhabdoid-Tumore untersucht [40]. In den untersuchten Fällen war das Gen *SMARCB1,* das für den Hauptteil der Erkrankungen verantwortlich ist, nicht inaktiviert. Es konnte jedoch eine LOH-Region mit normaler Kopienzahl auf Chromosom 19 gezeigt werden. Gleichzeitig konnte eine nonsense-Mutationen in dem Gen *SMARCA4/BRG1,* das in der entsprechenden Genomregion liegt, nachgewiesen werden. Beide Gene, *SMARCA4* und *SMARCB1*, sind an der Bildung des SWI/SNF Chromatin-Remodeling-Complex beteiligt, sodass mit der Inaktivierung von *SMARCA4* durch Mutation ein alternativer Mechanismus zur Entstehung der Tumore gezeigt werden konnte.

1.5 Vergleich von Array-CGH- und SNP-Array-Analyse

<u>Motivation</u>

Im Rahmen des von der deutschen Krebshilfe unterstützten Verbundprojektes „Molekulare Mechanismen in malignen Lymphomen" (MMML), wurden circa 900 Lymphomproben molekulargenetisch charakterisiert. Dabei erfolgte unter anderem die Analyse von Kopienzahlveränderungen. Innerhalb der ersten Förderperiode (Oktober 2003 bis April 2006) erfolgte die Messung der Kopienzahl für circa 500 Lymphomfälle mit CGH-Arrays. Dabei kam ein benutzerdefinierter Chip mit circa 2800 BAC/PAC-Klonen zum Einsatz (vgl. Abschnitte 1.3 und 2.1). In der 2. Förderperiode des Projektes sollte geprüft werden, ob die bisherigen Array-CGH-Messungen zukünftig durch Analysen mit 250k-SNP-Arrays der Firma Affymetrix abgelöst werden können. Diese SNP-Arrays umfassen deutlich mehr Marker als die CGH-Arrays und ermöglichen durch die Genotypinformationen darüber hinaus die Detektion von uniparentalen Disomien (UPD).

<u>Methodik</u>

Bei direktem Vergleich der Anzahl von gemessenen Markern pro Array und des medianen Abstandes von benachbarten Markern (vgl. Tabelle 1) zeigt sich eine deutliche Überlegenheit der SNP-Array-Plattform. Allerdings muss bei einem Vergleich auch die Qualität der einzelnen Messpunkte berücksichtigt werden, d. h., wie groß die Messfehler pro Messpunkt sind.

	Array-CGH (2,8k-Array)	Affymetrix 250k-Sty-SNP-Array	Verhältnis Array-CGH vs. SNP-Array
Anzahl Marker	2751	circa 238000	1:86,5
Medianer Abstand zwischen benachbarten Markern	665 Kb	5 Kb	133:1
Genotypisierung (LOH Detektion)	Nein	Ja	-

Tabelle 1: Vergleich der Eigenschaften von SNP- und CGH-Arrays

Um beide Methoden zu vergleichen, wurden 95 aggressive B-Zell-Lymphome und fünf Zelllinien mit Affymetrix 250k-Sty-SNP-Arrays gemessen. Für die Zelllinien wurden dabei Replikate in zwei verschiedenen Labors gemessen. Damit sollte geprüft werden, ob die zukünftige Messung weiterer Fälle auf mehrere Labors verteilt werden kann oder ob eine Aufteilung zu Verzerrungen bzw. Batcheffekten führt. Als euploide Kontrollproben wurden 48 frei verfügbare HapMap-Fälle von Affymetrix verwendet. Diese wurden durch 20 laboreigene Kontrollen ergänzt, die im gleichen Labor wie die 95 Primärfälle und ein Replikat der fünf Zelllinien gemessen wurden.

Für 92 der 95 Lymphomfälle und für vier von fünf Zelllinien lagen bereits Messungen mit dem 2,8k-CGH-BAC/PAC-Array vor. Die Analyse der CGH-Arrays erfolgte mit dem R-Paket *aCGHPipeline,* wie in den Abschnitten 1.3 und 2.1 beschrieben. Für die Auswertung der SNP-Array-Daten wurde, wie in Abschnitt 1.4 beschrieben, vorgegangen.

Für alle 92 Primärfälle, die mit beiden Plattformen analysiert wurden, erfolgte der Vergleich der Messwert-Streuung in Regionen mit normaler Kopienzahl. Dazu wurde der Median der absoluten Abweichungen (MAD) der Rohmesswerte auf der Log2-Skala ermittelt. Dabei wurden nur Regionen betrachtet, die in beiden Plattformen als „normal", d. h. Kopienzahl N=2, klassifiziert wurden. In diesen Regionen sollten die Rohmesswerte auf der Log2-Skala

zufällig um 0 streuen. Es zeigte sich, dass die Streuung bei den SNP-Chips etwa 1,8-fach über der Streuung der CGH-Arrays lag (vgl. Abbildung 7A).

Nachfolgend wurde untersucht, wie die höhere Anzahl der Messsonden der SNP-Arrays genutzt werden kann, um die höhere Streuung auszugleichen. Dazu wurden für jeden BAC-Klon des CGH-Arrays alle überlappenden SNP-Positionen des 250k-Sty-Arrays bestimmt. Für circa 5 % der BAC-Klone gab es keine überlappenden SNPs, 95 % der BAC-Klone wurden durch jeweils 1 bis 40 SNPs des 250k-Sty-Arrays repräsentiert. Bildet man für jeden BAC-Klon den Median aller überlappenden SNP-Rohmesswerte, ergibt sich eine geringere Streuung für die gemittelten SNP-Messwerte im Vergleich zu den Rohwerten aus der CGH-Messung (vgl. Abbildung 7B). Durch die Mittelung der SNP-Signale wird erwartungsgemäß deren Streuung deutlich reduziert.
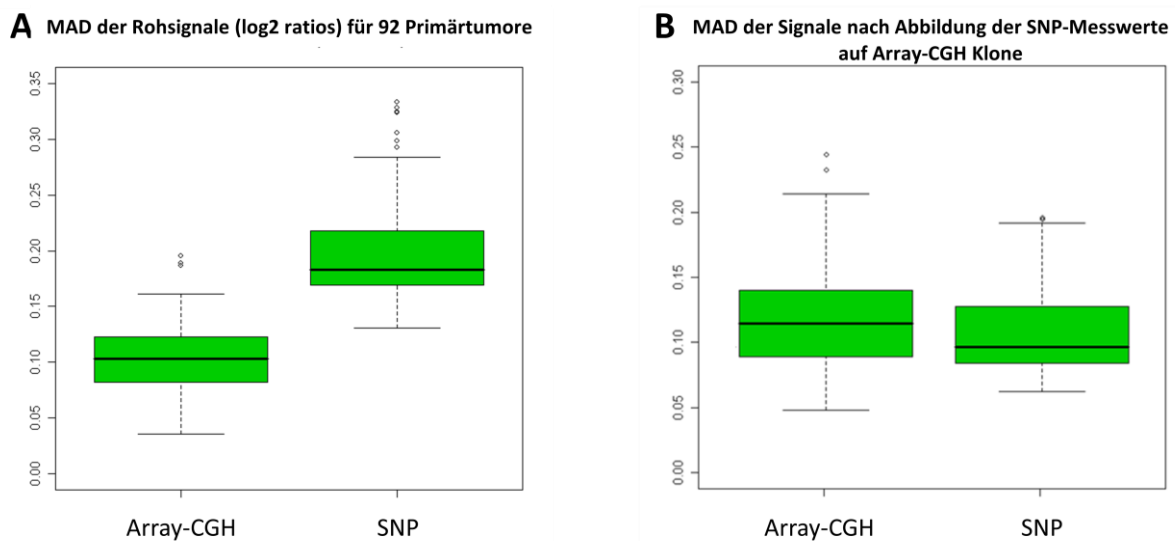
Abbildung 7: Vergleich der Streuung der Rohmesswerte für Array-CGH und SNP-Arrays
Für die Schätzung der Streuung wurde der Median der absoluten Abweichungen (MAD) aller Signale verwendet, deren Kopienzahl auf beiden Plattformen als „normal" (N=2) ermittelt wurde. **(A)** Die Rohsignale der SNP-Arrays zeigen im Vergleich zu den Array-CGH-Daten eine im Mittel circa 1,8-fach erhöhte Streuung. **(B)** Bildet man die SNP-Messwerte auf die CGH-Klone ab und mittelt alle zugehörigen SNP-Messwerte eines CGH-Klones mit dem Median, kann die Streuung der SNP-Messwerte reduziert werden. Es zeigt sich, dass die ermittelte Streuung etwas geringer als bei den Klonmesswerten der CGH-Arrays ist

Im Anschluss wurden die Klone entsprechend der Anzahl der überlappenden SNP-Messwerte eingeteilt. Für jede untersuchte Tumorprobe kann die Streuung der SNP-Messwerte mit der Streuung der CGH-Messwerte in Abhängigkeit der Anzahl gemittelter SNP-Messwerte verglichen werden (vgl. Abbildung 8A). Dabei zeigte sich erwartungsgemäß ein Abfallen der Streuung mit steigender Anzahl der gemittelten SNP-Messwerte, sodass ab einer bestimmten Anzahl gemittelter SNP-Werte die Streuung unter das Niveau der CGH-Messwerte fällt. Die Anzahl der gemittelten SNP-Positionen, die benötigt wird, damit das gemittelte SNP-Signal niedrigere Streuung als die CGH-Rohsignale zeigen, kann je nach betrachtetem Tumorfall variieren. Im Beispiel von Abbildung 8A wird dieser Effekt bei einer Mittelung von mindestens vier SNP-Signalen erreicht.

Um die Anzahl der SNP-Messwerte zu bestimmen, die im Mittel benötigt werden, um die gleiche bzw. niedrigere Streuung zu erhalten als die CGH-Messwerte, wurden alle 92 Primärtumore mit gepaarten Array-CGH und SNP-Array-Daten analysiert. Getrennt für die Anzahl überlappender SNP-Positionen wurde für jeden der 92 Fälle der Quotient aus der Streuung der gemittelten SNP-Signale und der Streuung der CGH-Signale bestimmt.

Anschließend erfolgte die Darstellung des Medians dieser Quotienten, getrennt für jede Anzahl an überlappenden SNPs. Dabei zeigt sich, dass im Mittel bei einer Mittelung von ≥7 SNP-Signalen die Streuung der gemittelten SNP-Signale niedriger ist, als die der CGH-Messungen (vgl. Abbildung 8B). Berücksichtigt man daher die erhöhte Anzahl an Messpunkten der SNP-Arrays und gleichzeitig die erhöhte Streuung der SNP-Signale, zeigt sich, dass die effektive Auflösung der SNP-Array-Plattform etwa das Zwölffache der CGH-Arrays beträgt. Dabei wird die 86,5-fache Anzahl an Messsonden durch die Anzahl der Sonden geteilt, die benötigt wird, um die gleiche Signalqualität zu erreichen (86,5/7 ≈ 12).
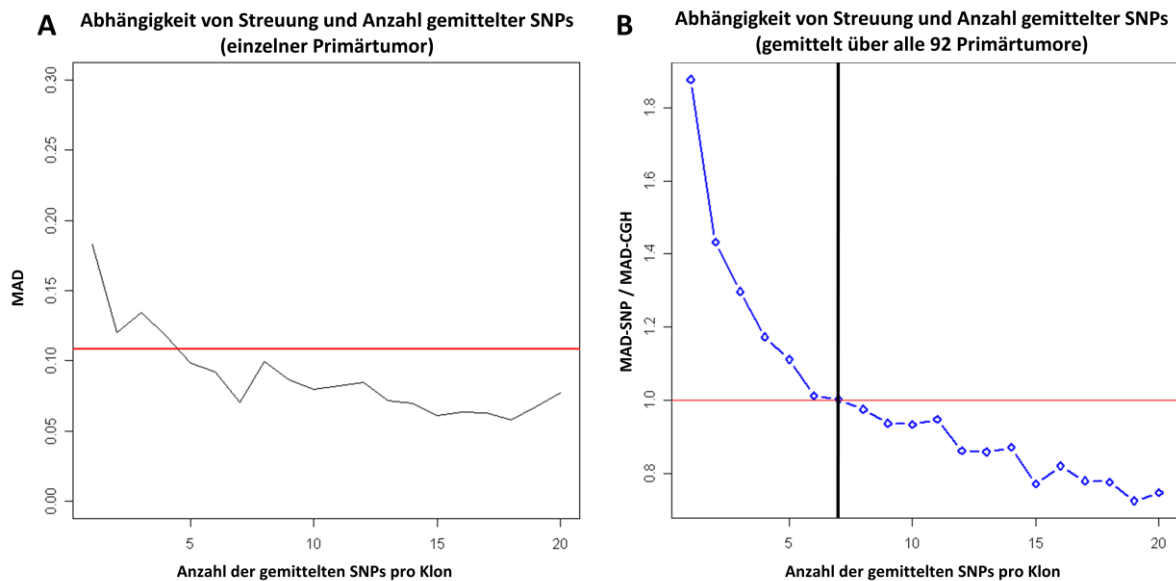
Abbildung 8: Zusammenhang zwischen der Anzahl gemittelter SNP-Signale und Streuung. **(A)** Vergleich der Streuung der Array-CGH-Signale und der gemittelten SNP-Signale für eine Lymphomprobe. Die rote Linie zeigt die Streuung der CGH-Signale an. Die schwarze Linie zeigt die Streuung der gemittelten SNP-Signale in Abhängigkeit zur Anzahl gemittelter SNPs an. **(B)** Vergleich der Streuung der Array-CGH-Signale und der gemittelten SNP-Signale, gemittelt über alle 92 Primärfälle. Die blaue Linie zeigt den Median des Quotienten der Streuung von gemittelten SNP-Signalen und Array-CGH-Signalen an. Die rote Linie bei 1 zeigt an, ab wann die gemittelten SNP-Signale geringere Streuung als die CGH-Messungen aufweisen, d. h., ab wann der Quotient <1 ist. Im untersuchten Datensatz wird dies bei einer Mittelung von ≥7 SNP-Signalen erreicht (vertikale Linie).

Im Anschluss wurde die Konkordanz der Array-CGH- und SNP-Array-Daten analysiert. Dazu wurden für die Daten der Zelllinien die Kopienzahlsignale der CGH-Arrays mit den SNP-Signalen verglichen. Um die Vergleichbarkeit zu gewährleisten, erfolgte für jeden CGH-Klon die Mittelung aller überlappenden SNP-Signale. Für die Zelllinie Su-DHL5 ist der Vergleich in

Abbildung 9 dargestellt. Insgesamt zeigte sich eine gute Konkordanz der Messergebnisse. Nur für eine Zelllinie (HAT) zeigte sich eine größere diskrepante Region auf Chromosom 1, wobei eine Nachmessung mittels Agilent 244k-Array-CGH am Institut für Humangenetik in Kiel das Ergebnis der SNP-Array-Analyse bestätigte. Für alle 92 mit beiden Technologien gemessenen Primärfälle ergab ein Vergleich der detektierten rekurrenten Aberrationen eine gute Übereinstimmung zwischen beiden Plattformen.
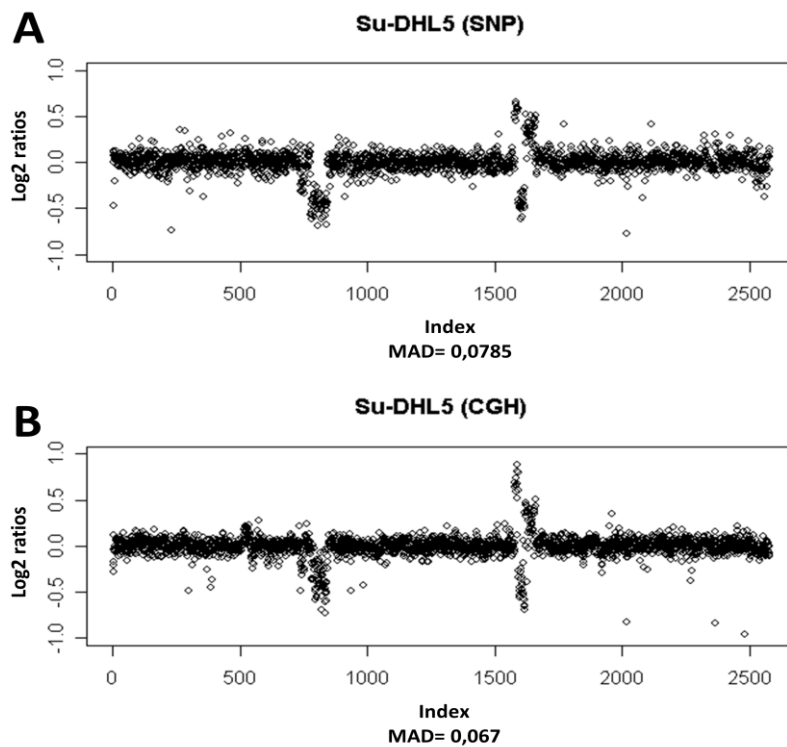
Abbildung 9: Vergleich der Kopienzahlsignale von SNP-Arrays **(A)** und Array-CGH **(B)** für die Zelllinie Su-DHL5. Für beide Plattformen sind die Kopienzahlsignale in genomischer Reihenfolge auf der Log2ratio-Skala dargestellt. Positive und negative Signalausschläge weisen auf Deletionen bzw. Zugewinne hin. Für den SNP-Array wurde die Messung auf die CGH-Klone abgebildet, wobei jeder CGH-Klon durch den Median aller überlappenden SNP-Signale repräsentiert wird.

Anhand der fünf gemessenen Zelllinien erfolgte eine Analyse der Reproduzierbarkeit der SNP-Array-Messungen. Die Zelllinien wurden jeweils in zwei verschiedenen Labors gemessen. Die Ergebnisse sind exemplarisch in Abbildung 10 dargestellt. Insgesamt zeigte sich eine sehr gute Konkordanz der Replikatmessungen. Im Mittel konnten für 98.5 % der Marker, die LOH zeigten, die LOH auch in der zweiten Messung bestätigt werden. Für >96 % aller Positionen, für die Kopienzahlveränderungen nachgewiesen werden konnten, zeigten die Replikationsmessungen konkordante Kopienzahlveränderungen an. Besonders homozygote Deletionen konnten aufgrund des starken Signalausschlages, trotz der geringen Größe der betroffenen genomischen Regionen, zuverlässig reproduziert werden.

In einem Labor wurden 20 euploide Kontrollproben gemessen. Diese wurden gemeinsam mit 48 unabhängigen Affymetrix-Hapmap-Kontrollen für die Kopienzahlanalyse verwendet. Bei Ausschluss der laboreigenen Referenzen zeigte sich eine um circa 35 % erhöhte Streuung in den Kopienzahlsignalen. Folglich kann die Qualität der Messung durch den Einsatz laboreigener Referenzen deutlich verbessert werden.
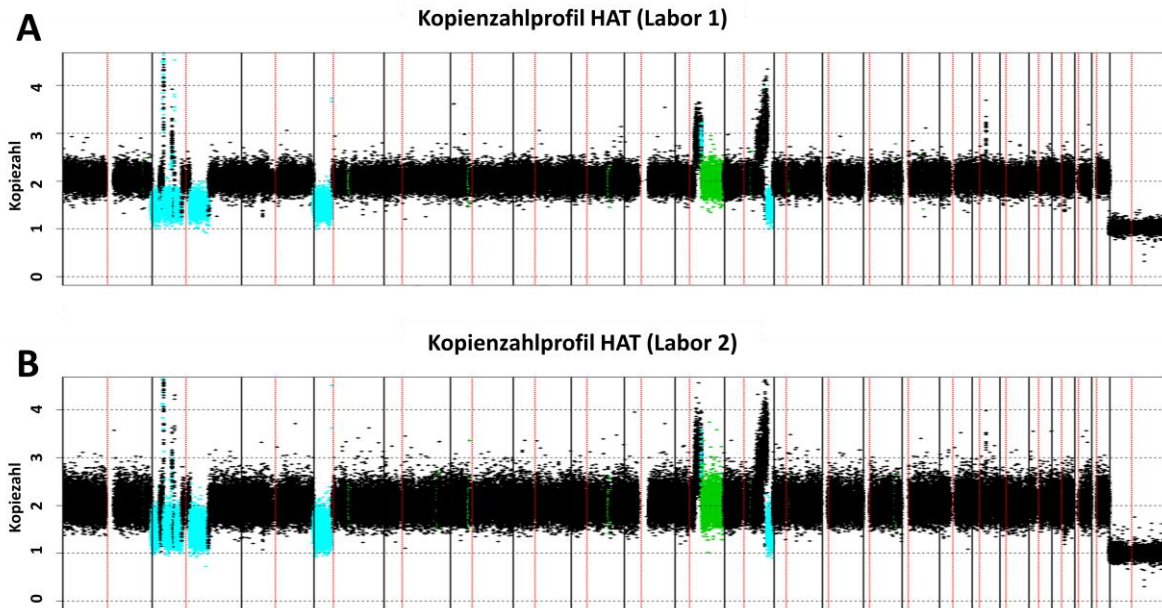
Abbildung 10: Reproduzierbarkeit der SNP-Array-Messung der Zelllinie HAT. Die Messungen erfolgten in zwei verschiedenen Labors. Die Punkte zeigen jeweils die rohen Kopienzahlsignale geordnet nach Genomposition an. Sie ergeben sich aus dem Vergleich der SNP-Signalintensitäten mit normalen Kontrollen. Grüne Punkte zeigen LOH-Regionen ohne Kopienzahlveränderungen an (UPD). Cyan gefärbte Regionen zeigen LOH bei gleichzeitigem Vorliegen von Kopienzahlaberrationen an

Ergebnisse

Zusammenfassend konnte im Rahmen dieser Analyse gezeigt werden, dass die effektive Auflösung der SNP-Array-Plattform das circa Zwölffache der CGH-Arrays beträgt. Da die SNP-Array-Plattform zusätzlich die Detektion von LOH ermöglicht, ist sie die überlegene Technologie. Es konnte darüber hinaus gezeigt werden, dass die Übereinstimmung von CGH- und SNP-Array-Messungen gut ist und die SNP-Array-Analysen auch bei der Messung in verschiedenen Labors gut reproduzierbar ist. Allerdings sollten für die Analyse jeweils laboreigene euploide Kontrollen bereitgestellt werden. Als Ergebnis der Analyse erfolgte die Kopienzahlanalyse im MMML-Verbundprojekt anschließend mit Affymetrix 250k-SNP-Arrays.

1.6 Assoziationen von DNA-Kopienzahlaberrationen mit RNA-Expression, Lymphomentität sowie klinischen und phänotypischen Faktoren

Motivation

Die deskriptive Beschreibung der auftretenden Kopienzahlaberrationen sowie deren Häufigkeiten in einer untersuchten Tumorentität sind oftmals nur der erste Schritt in der Auswertung von CGH- und SNP-Array-Datensätzen.

Rekurrente Aberrationen betreffen oftmals größere Regionen des Genoms, die eine Vielzahl von Genen und nichtkodierenden RNAs enthalten. Das erschwert die Identifikation von relevanten Zielgenen innerhalb dieser Regionen. Für die Auswahl plausibler Kandidatengene erfolgt oftmals ein Abgleich der Kopienzahldaten mit gepaarten Genexpressionsmessungen. Dabei soll unter Berücksichtigung des Problems des multiplen Testens untersucht werden, welche Gene/Transkripte innerhalb von rekurrenten Regionen einen konkordanten Gendosiseffekt zeigen, das heißt bei erhöhter bzw. reduzierter DNA-Kopienzahl eine erhöhte bzw. reduzierte Expression aufweisen. Diese sind vielversprechende Kandidaten für Onkogene (Kopien-Zugewinn und erhöhte Expression) oder Tumorsuppressorgene (Deletion und reduzierte Expression) und können anschließend in größeren Kollektiven auf Aberrationen oder mittels funktionaler Analysen untersucht werden.

Häufig werden bei einer Analyse die Aberrationshäufigkeiten für verschiedene Tumorsubgruppen getrennt betrachtet. Dabei soll untersucht werden, ob den Tumoren gleiche Aberrationen zugrunde liegen. Große Abweichungen in den beobachteten Aberrationshäufigkeiten weisen dabei auf unterschiedliche molekulare Mechanismen bei der Entstehung oder Progression der zugehörigen Tumorsubgruppen hin. Folglich können die ermittelten Aberrationshäufigkeiten in den betrachteten Subgruppen verwendet werden, um Einblicke in die genetischen Unterschiede der Entitäten zu gewinnen. Weiterhin können Regionen mit starken Unterschieden in den Aberrationshäufigkeiten geeignete diagnostische Marker sein, die es erlauben, die Subgruppen zu unterschieden. Für manche Fragestellungen ist es darüber hinaus erforderlich, den Zusammenhang zwischen der Inzidenz von Kopienzahlaberrationen in Abhängigkeit von stetigen Größen wie dem Erkrankungsalter des Patienten zu untersuchen. Dafür gilt es Methoden bereitzustellen, die es erlauben, Fragestellungen dieser Art zu untersuchen.

Methodik

Im Rahmen der Promotionsarbeit wurde eine Methode entwickelt, mit der eine integrative Analyse von Genexpressions- und Kopienzahldaten möglich ist. Sie nutzt als Eingabe die Grenzen der rekurrenten Regionen sowie eine Statustabelle in Form einer N x M Matrix, die für jeden gemessenen Tumorfall *n* angibt, ob er die entsprechende rekurrente Region *m* aufweist oder nicht. Diese Daten werden in *aCGHPipeline* automatisch generiert. Zusätzlich werden eine Matrix der gepaarten Expressionswerte und die Genompositionen der gemessenen Transkripte benötigt. Die Methode vergleicht anschließend, für jede rekurrente Region individuell, die Expression aller Transkripte in der betroffenen Genomregion zwischen den Tumoren mit und ohne Aberration. Als statistischer Test für die Differenz des Expressionsniveaus in Abhängigkeit der Kopienzahl wird der t-Test eingesetzt. Für jede rekurrente Region erfolgt die Adjustierung für multiples Testen über alle gemessenen Transkripte innerhalb der zugehörigen Genomregion mittels False Discovery Rate (FDR), wobei der q-value frei wählbar ist [41]. Die Ausgabe erfolgt für jede betrachtete Region grafisch bzw. im HTML-Format und ist beispielhaft in Abbildung 11 dargestellt.
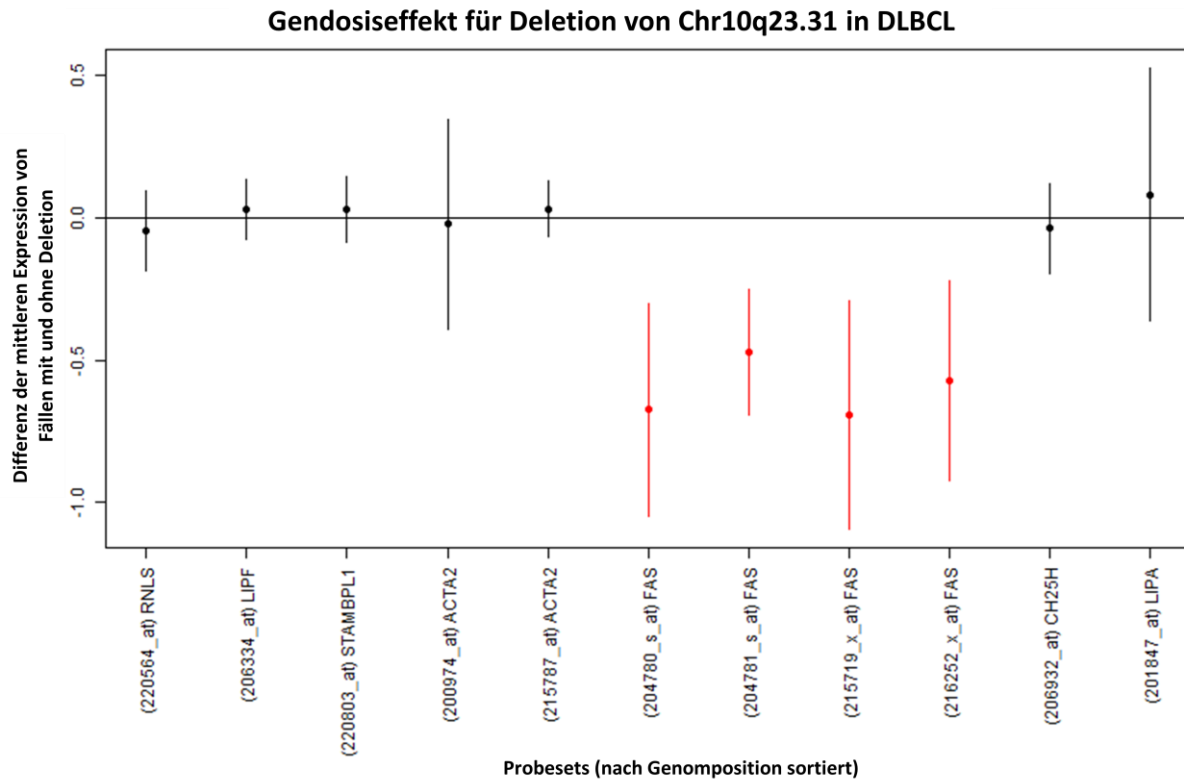
Abbildung 11: Analyse des Einflusses von Deletionen der Region Chr10q23.31 in DLBCL auf die Expression der betroffenen Gene. Auf der x-Achse sind die zugehörigen Probesets in genomischer Reihenfolge sortiert und mit Gensymbol sowie Affymetrix hgu133a Probeset-ID gekennzeichnet. Auf der y-Achse ist die Differenz der mittleren Expression von Fällen mit und ohne Deletion dargestellt. Dabei zeigen die Punkte die Differenzen der Mittelwerte und die Linien die 95 % Konfidenzintervalle an. Probesets, deren Expression sich signifikant in den Fällen mit und ohne Aberration unterscheidet, sind rot markiert. Als Schwellwert für Signifikanz wurde FDR = 10 % verwendet

Aberrationen der DNA-Kopienzahl können darüber hinaus die Expression von Genen außerhalb der betroffenen Genomregion beeinflussen (*trans*-Effekte). Das Aufspüren dieser Effekte erfordert aufgrund der komplexen Kombinatorik jedoch sehr große Datensätze von gepaarten Kopienzahl- und Expressionsdaten [42] und konnte daher im Rahmen dieser Promotionsarbeit nicht verfolgt werden.

Um die Unterschiede der Kopienzahlprofile zwischen zwei Subgruppen systematisch zu untersuchen, wurde eine Methode implementiert, die für jede Messsonde die Häufigkeit von Zugewinnen und Verlusten vergleicht. Für jede Sonde wird dazu in den Vergleichsgruppen die Anzahl der Fälle mit Kopien-Zugewinn sowie die Anzahl der Fälle ohne Zugewinn (normale Kopienzahl oder Deletion) ermittelt. Für die daraus entstehende Vierfeldertafel wird der p-Wert des Fisher-Tests bestimmt. Analog erfolgt der Test für die Häufigkeit von Deletionen. Anschließend werden die $-\log_{10}(p)$ Werte für Zugewinne und Verluste in ein gemeinsames Vergleichsdiagramm geplottet (vgl. Abbildung 12C). Hohe Ausschläge zeigen starke Differenzen in den Häufigkeiten von Zugewinnen bzw. Deletionen an, wobei Werte von 2 einem nominalen p-Wert von 0,01 entsprechen. In Abbildung 12 ist die Methode anhand des Array-CGH-Datensatzes des MMML-Verbundes illustriert. Dabei erfolgte die Analyse an einem Teildatensatz von n=330 Lymphomen, für den der Mutationsstatus des Gens *TP53* (Genomposition: 17p13.1) vorlag. Verglichen wurden hierbei Lymphome mit (n=87) und ohne *TP53*-Mutation (n=243). Die signifikantesten Unterschiede in den Häufigkeiten von Kopienzahl-Aberrationen zwischen beiden Subgruppen ergaben sich

hierbei für die Genomregion 17p. Die Analyse der Kopienzahlprofile beider Subgruppen zeigt für Fälle ohne *TP53*-Mutation in circa 5 % der Tumore eine Deletion von 17p an (Abbildung 12B), in den Tumoren mit *TP53*-Mutationen zeigen circa 40 % der Lymphome eine Deletion (Abbildung 12A). Die Analyse des Vergleichsprofils nach der oben geschilderten Methodik zeigt einen deutlichen Ausschlag für Verluste bei 17p, also der Genomposition von *TP53*. Der Ausschlag beträgt im Maximum circa 10, was einem nominalen p-Wert von $1*10^{-10}$ entspricht. *TP53* ist eines der bekanntesten Tumorsuppressorgene und spielt eine zentrale Rolle bei der DNA-Reparatur, der Regulation des Zellzyklus und der Steuerung der Apoptose. Deaktivierung von *TP53* ist ein zentraler Schritt in der Entwicklung von vielen humanen Tumorerkrankungen [43]. Eine naheliegende Interpretation dieses Befundes ist daher, dass bei Auftreten einer *TP53*-Mutation häufig das verbleibende, intakte Allel durch eine Deletion ausgeschaltet wird und dadurch beide Allele des Gens deaktiviert werden. Ein ähnliches Zusammenwirken von *TP53*-Mutationen und chromosomalen Aberrationen von Chr17p13.1 konnte auch für Mantelzelllymphome gezeigt werden (vgl. Manuskript 3; Abschnitt 2.3).
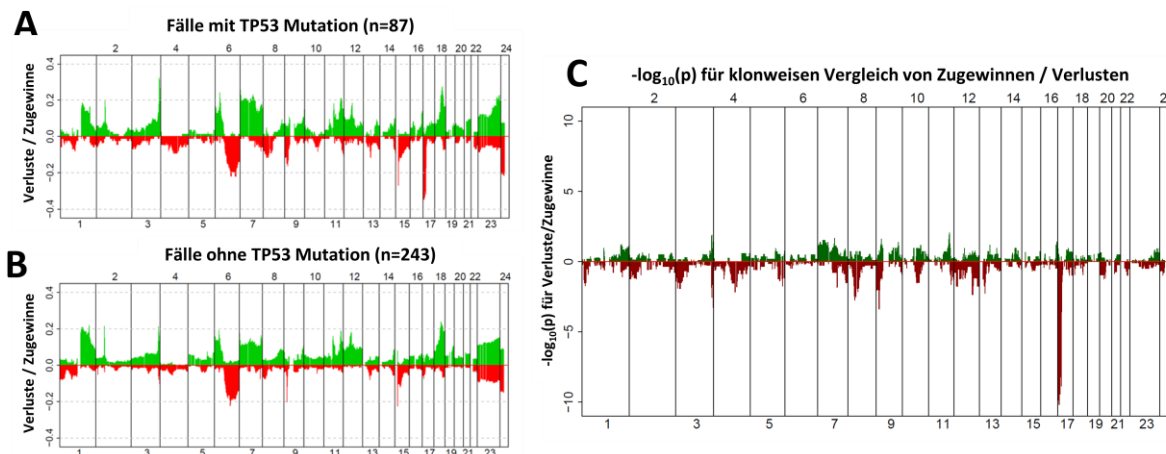
Abbildung 12: Vergleich der Kopienzahlprofile des MMML-Array-CGH-Datensatzes in Abhängigkeit des Mutationsstatus von *TP53* (Genomposition: 17p13.1). **(A)**, **(B)** Die beiden Profile zeigen die Häufigkeit der Kopienzahlaberrationen in den Lymphomen **(A)** mit *TP53*-Mutation (n=87) und **(B)** ohne *TP53*-Mutation (n=243) an.
(C) Vergleichsprofil. Für Zugewinne sind die Häufigkeiten zwischen Fällen mit und ohne Mutation klonweise mit Fisher-Test verglichen, wobei der resultierende p-Wert des Tests als –$\log_{10}$(p) dargestellt ist. Für Verluste erfolgt der Vergleich analog mit der Darstellung von $\log_{10}$(p). Verluste bei 17p zeigen den größte Ausschlag, wobei p-Werte von annähernd $1*10^{-10}$ erreicht werden

Für die Analyse der Kopienzahlprofile in Abhängigkeit von stetigen Einflussgrößen lässt sich die oben eingeführte Methode nicht einsetzen. Alternativ können jedoch die rekurrenten Regionen und die genetische Komplexität für jeden untersuchten Tumorfall bestimmt werden. Methoden zur automatischen Bestimmung dieser Größen sind beispielsweise auch im Paket *aCGHPipeline* implementiert (vgl. Abschnitt 1.3). Der Einfluss stetiger Größen auf die Inzidenz einzelner rekurrenter Regionen oder die genetische Komplexität im Allgemeinen kann anschließend mittels logistischer Regression bzw. im Fall der genetischen Komplexität mittels Poisson-Regression untersucht werden. Anschließend können die Zusammenhänge beispielsweise wie in Abbildung 13 dargestellt werden.
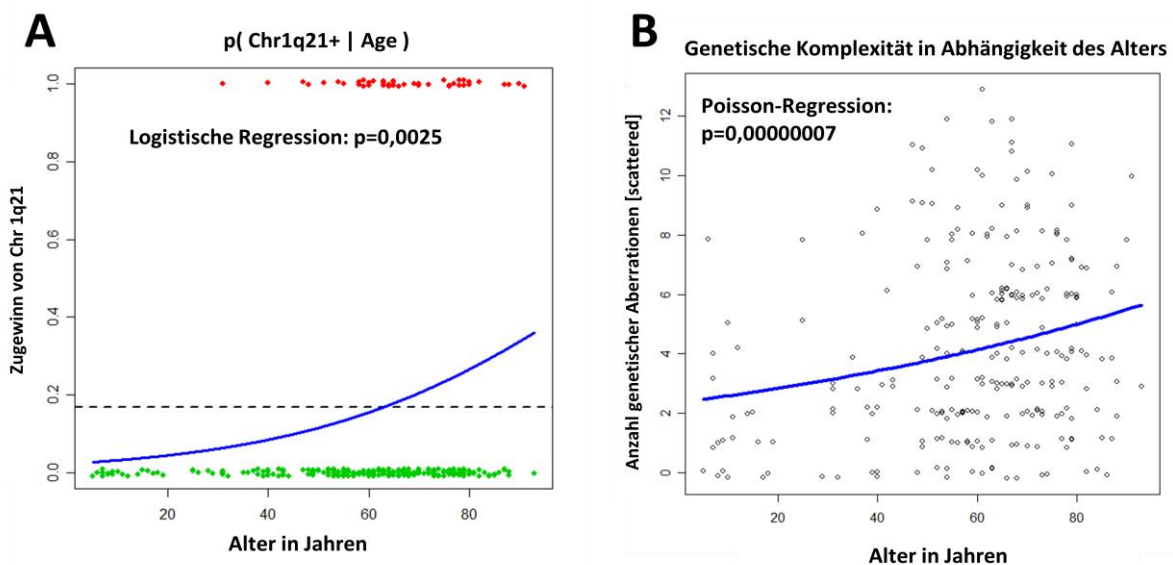
Abbildung 13: Zusammenhang zwischen dem Auftreten eines 1q21-Zugewinns sowie genetischer Komplexität und Erkrankungsalter.

(A) Zusammenhang zwischen 1q21-Zugewinnen und Erkrankungsalter. Die grünen Punkte zeigen Lymphome ohne 1q21-Zugewinn, die roten Punkte Lymphome mit 1q21-Zugewinn (die Punkte sind bezüglich der y-Richtung leicht verrauscht, um Überlappungen bei gleichem Erkrankungsalter zu vermeiden). Die horizontal gestrichelte Linie gibt die Häufigkeit von 1q21+ im gesamten untersuchten Kollektiv an, die blaue Linie zeigt das Ergebnis der logistischen Regression an und beschreibt die bedingte Wahrscheinlichkeit für das Auftreten von 1q21+ gegeben das Erkrankungsalter.

(B) Zusammenhang von genetischer Komplexität und Erkrankungsalter. Die genetische Komplexität ist dabei die Summe aller aufgetretenen Aberrationen (vgl. Abschnitt 2.7), die Punkte sind analog zur linken Teilgrafik gescattert, um Überlappungen zu vermeiden. Die blaue Linie stellt das Ergebnis der Poisson-Regression dar und zeigt somit das Ansteigen der genetischen Komplexität mit zunehmendem Alter

Anwendung

Die Anwendung der Methode zur Korrelation von Kopienzahl- und Genexpression zur Detektion von signifikanten Gendosiseffekten ist in Manuskript 1 (vgl. Abschnitt 2.1) und Manuskript 6 (vgl. Abschnitt 2.6) mit CGH- und SNP-Array-Datensätzen im Rahmen des MMML-Projekts gezeigt. Beim Einsatz der Methode gibt es jedoch eine Reihe von Limitationen. Zunächst sind gepaarte Kopienzahl- und Genexpressionsdaten erforderlich. Darüber hinaus können Zusammenhänge nur bei entsprechend großer Anzahl von Fällen mit und ohne Aberration aufgedeckt werden. Außerdem ist nicht auszuschließen, dass in Fällen ohne Aberration die Expression relevanter Gene über andere, beispielsweise epigenetische, Mechanismen dereguliert wird und daher kein Zusammenhang der Expression und der zugrundeliegenden Kopienzahl gezeigt werden kann. Weitere Anwendungen der beschriebenen Methode zur Korrelation von Genexpressions- und Kopienzahldaten sind in Manuskript 4 (Abschnitt 2.4) und in [26, 39] zu finden.

Die Methode zum Vergleich die Kopienzahlprofile zwischen zwei Subgruppen wurde ebenfalls in einer Reihe von Publikationen eingesetzt [20, 22, 24, 26]. Auch diese Methode ist durch die Fallzahl der untersuchten Vergleichskollektive limitiert.

Die genetische Charakteristik von diffus großzelligen B-Zell-Lymphomen (DLBCL) in Abhängigkeit des Alters wurde in Manuskript 7 (Abschnitt 2.7) anhand von n=364 Lymphomproben untersucht. Dabei kam die im aktuellen Abschnitt beschriebene Methode zur Assoziation von Kopienzahlaberrationen mit stetigen Einflussgrößen zum Einsatz. Neben

den numerischen Kopienzahlaberrationen wurde auch die Inzidenz von Translokationen und molekulargenetischen Klassifikatoren [20, 44] sowie die Ausprägungen von immunhistochemischen Markern analysiert. Dabei konnte für Zugewinne von 1q21, 18q21, 7p22 und 7q21 sowie chromosomale Veränderung (Zugewinne und Translokationen) des BCL6-Lokus auf 3q27 eine Assoziation mit höherem Erkrankungsalter nachgewiesen werden. Weiterhin zeigte sich eine erhöhte Häufigkeit des ABC-Subtyps sowie von BCL2-Expression bei hohem Erkrankungsalter. Die einzige Aberration, die bei jüngeren Patienten gehäuft vorkam, war die *IRF4*-Translokation. Diese Translokation ist jedoch charakteristisch für eine Subgruppe von Lymphomen, die sich auch in ihrem Genexpressionsprofil von den übrigen Lymphomen unterscheiden [24]. Lymphome dieses Typs treten vorwiegend bei Kindern und jungen Erwachsenen auf [24]. Insgesamt konnte ein Ansteigen der genomischen Komplexität mit zunehmendem Erkrankungsalter gezeigt werden (vgl. Abbildung 13). Dabei zeigt sich jedoch ein kontinuierlicher Anstieg, sodass keine plausiblen Altersgrenzen für die Unterscheidung von Lymphomen mit niedriger und hoher genetischer Komplexität festgelegt werden können. Die Daten deuten vielmehr auf ein dem DLBCL zugrunde liegendes Evolutionsmodell hin, bei dem mehrere genetische Aberrationen über die Zeit akquiriert werden. Dabei hat jede Aberration einen konstanten Hazard. Genetisch komplexe Lymphome treten daher gehäuft bei älteren Patienten auf, da durch die längere Zeitspanne die Wahrscheinlichkeit, viele, gegebenenfalls auch weniger schwerwiegende Aberrationen zu akkumulieren, erhöht ist.

2 Publikationen

2.1 Publikation 1: „Development and implementation of an analysis tool for array-based comparative genomic hybridization" Methods Inf Med. 2007;46(5):608-13

Kreuz M, Rosolowski M, Berger H, Schwaenen C, Wessendorf S, Loeffler M, Hasenclever D.

Bedingt durch die Komplexität und den Umfang der gemessenen Daten ist die manuelle Analyse von Array-CGH-Daten sehr zeitaufwendig und fehleranfällig. Ziel war es daher, ein Konzept für die automatische Auswertung von aCGH-Datensätzen zu erstellen. Basierend auf diesem Konzept sollten die benötigten Methoden entwickelt und ein Auswertewerkzeug implementiert werden.

Im Rahmen der Arbeit wurde das R-Paket *aCGHPipeline* entwickelt, welches die Analyse einzelner Arrays sowie die Aggregation der Daten mehrerer Arrays unterstützt. Das Paket ermöglicht zunächst das Einlesen und die Qualitätskontrolle der Daten. Mittels chipweiser Normalisierung können Verzerrungsquellen wie Unterschiede in der Menge bzw. Qualität von Tumor- und Referenz-DNA eliminiert werden. Das Paket ermöglicht die Segmentierung der normalisierten Daten in Genomregionen mit gleicher Kopienzahl. Dafür stehen zwei verschiedenen Methoden zur Verfügung [18, 19]. Anschließend ermöglicht die implementierte Klassifikationsmethode die Einteilung der Segmente in Zugewinne, Verluste und Regionen mit normaler Kopienzahl. Danach können die Ergebnisse der einzelnen CGH-Arrays aggregiert werden. Es besteht die Möglichkeit automatisch Regionen, in denen gehäuft konkordante Aberrationen auftreten (rekurrente Regionen), für weitere Auswertungen zu selektieren. *aCGHPipeline* unterstützt die Korrelation von Kopienzahl- mit Genexpressionsdaten und ermöglicht dadurch die Detektion von Gendosiseffekten. Im Rahmen der Arbeit wurde das entwickelte R-Paket auf einen Datensatz von 189 Lymphomen angewendet. Diese wurden jeweils mit einem Array gemessen, der 2799 BAC/PAC-Klone umfasste. Dabei erfolgte ein Vergleich zwischen automatischer Auswertung mittels *aCGHPipeline* und manueller Bewertung der Daten. Hierfür wurden 106 Lymphomfälle betrachtet. Es ergab sich eine Übereinstimmung für circa 97 % aller betrachteten Klon-Messwerte. Mehr als 83 % der manuell als aberrant gewerteten Klone wurden von der automatischen Methode konkordant bewertet. In Bereichen, in denen eine Methode einen Zugewinn und die andere Methode eine Deletion anzeigt, war stets ein Vorzeichenfehler in den manuell bewerteten Daten die Ursache. Weitere Unterschiede traten hauptsächlich an Segmentgrenzen und bei schwachen Signalausschlägen, d. h. quantitativ schwachen Aberrationen, auf. In diesen Fällen ist die Bewertung der Ergebnisse ohne weitere unabhängige Validierung schwierig.

Das in der Publikation beschriebene Kollektiv wurde im Rahmen der verbleibenden Laufzeit des MMML-Verbundes auf n=514 Lymphomproben erweitert. Die Ergebnisse dieser Analysen sind mittlerweile in zahlreiche Publikationen eingeflossen [20-26].

Das R-Paket *aCGHPipeline* wurde zudem eingesetzt, um einen Datensatz von 53 klarzelligen Nierenzellkarzinomen zu analysieren und den Zusammenhang zwischen auftretenden Kopienzahlaberrationen und klinischem Verlauf sowie Metastasierungsverhalten der Tumore zu untersuchen [27]. Aktuell wird *aCGHPipeline* für Analysen im Rahmen des deutschen Gliomnetzwerkes (German Glioma Network) verwendet. Hierbei werden Array-CGH-Messungen von Gliomen mit verschiedenen histologischen Graden und Paare von Primärtumoren und Rezidiven untersucht [28].

Development and Implementation of an Analysis Tool for Array-based Comparative Genomic Hybridization

M. Kreuz[1], M. Rosolowski[1], H. Berger[1], C. Schwaenen[2], S. Wessendorf[2], M. Loeffler[1], D. Hasenclever[1]

[1] Institute for Medical Informatics, Statistics and Epidemiology (IMISE), University of Leipzig, Leipzig, Germany
[2] Department of Internal Medicine III, University Hospital of Ulm, Ulm, Germany

Summary

Objectives: Array-comparative genomic hybridization (aCGH) is a high-throughput method to detect and map copy number aberrations in the genome. Multi-step analysis of high-dimensional data requires an integrated suite of bioinformatic tools. In this paper we detail an analysis pipeline for array CGH data.

Methods: We developed an analysis tool for array CGH data which supports single and multi-chip analyses as well as combined analyses with paired mRNA gene expression data. The functions supporting relevant steps of analysis were implemented using the open source software R and combined as package aCGHPipeline. Analysis methods were illustrated using 189 CGH arrays of aggressive B-cell lymphomas.

Results: The package covers data input, quality control, normalization, segmentation and classification. For multi-chip analysis aCGHPipeline offers an algorithm for automatic delineation of recurrent regions. This task was performed manually up to now. The package also supports combined analysis with mRNA gene expression data. Outputs consist of HTML documents to facilitate communication with clinical partners.

Conclusions: The R package aCGHPipeline supports basic tasks of single and multi-chip analysis of array CGH data.

Keywords

Array CGH, DNA copy number, gene expression, gene dosage

1. Introduction

Aberrations in copy number play an important role in various diseases, especially in the pathogenesis of malignant tumors [1]. The length of aberrant segments can range from a few base-pairs to whole chromosomes. Copy number aberrations can affect gene expression: Deletions may lead to tumor suppressor gene inactivation and copy number gains may cause activation of oncogenes [2].

Array comparative genomic hybridization (aCGH) is a method to detect and map these copy number aberrations in the genome. Several thousand known DNA clones or oligonucleotides are spotted on a chip [1]. Each clone represents a specific region of the genome. Resolution and coverage of the analysis depends on the number of spotted clones and their distribution in the genome.

DNA is isolated from test and reference tissue and differentially labeled using fluorescence dyes. A balanced mixture of labeled test and reference DNA is hybridized to the CGH array.

Test and reference DNA compete for free binding sites [3]. Signals of test and reference fluorescence intensity at each clone position are measured, preprocessed [4] and combined as a log2 ratio. This signal is assumed proportional to the log2 ratio of test and reference copy number in the corresponding genomic region. If the reference tissue is chosen euploid, information on copy number changes in the test sample can be obtained. Raw data from one CGH-array experiment consists of several thousand clone-specific log2 ratios.

2. Motivation

Due to data complexity, manual interpretation of array CGH data is time-consuming and error-prone. Automatic methods which facilitate the array CGH analysis are required.

We developed an analysis tool for array CGH data [5] which supports single and multi-chip analysis as well as combined analysis with paired mRNA gene expression data. The methods were implemented using the open source software R [6] and combined as package aCGHPipeline. The package is available upon request. Most of the currently available analysis programs are limited in the capability of multiple-chip analysis. For example, CGH-Plotter [7] and CGHPro [8] provide only a graphical comparison of different chips but there is no support for further statistical investigations of genetic differences. The aim of aCGHPipeline is to overcome these limitations.

Here we describe relevant steps in the analysis of array CGH data that can be performed using functions from our tool. Figure 1 gives a schematic overview over tasks addressed below which are supported by the package. Illustrations are taken from an analysis of array CGH data of 189 aggressive B-cell lymphomas [9]. Each array contained 2799 BAC/PAC clones. 1500 of these clones cover the whole genome at intervals

of approximately 2 Mb, the remaining 1299 clones span regions known to be frequently involved in B-cell neoplasms or contain proto-oncogenes or tumor suppressor genes [9].

3. Methods

3.1 Quality Control

As for other high-throughput methods a quality control step is required to detect and reject invalid data. For every chip and every clone aCGHPipeline calculates the median absolute deviation and the proportion of missing values. Summary statistics, histograms and lists of arrays and clones which are suspicious are summarized in a user-friendly HTML output file for further inspection.

Large-scale copy number polymorphisms also occur as natural variation of the normal human genome. The length of polymorphic segments ranges from kilo- to megabases [10]. To distinguish between normal variation and cancer-specific aberrations, it is important to compare the genomic positions of the used clone probes with known polymorphic sites. Therefore every clone on the array is compared with the "Database of Genomic Variants" (http://projects.tcag.ca/variation/) [11]. Clones which are located in regions with known copy number polymorphisms are marked, enabling the user to interpret the measured values cautiously.

3.2 Normalization

Imbalances in the amount or quality of used test and reference DNA may lead to systematic array-specific bias in the measured log2 ratios. Differences in labeling efficiencies of the fluorescence dyes may be eliminated by using a dye swap design [12]. The aim of normalization is to remove chip-specific bias. aCGHPipeline provides a global normalization for the correction of array CGH data: Assuming that the majority of clone positions is euploid in both tissues most of the log2 ratios should vary about

Fig. 1
Steps of analysis in array CGH data
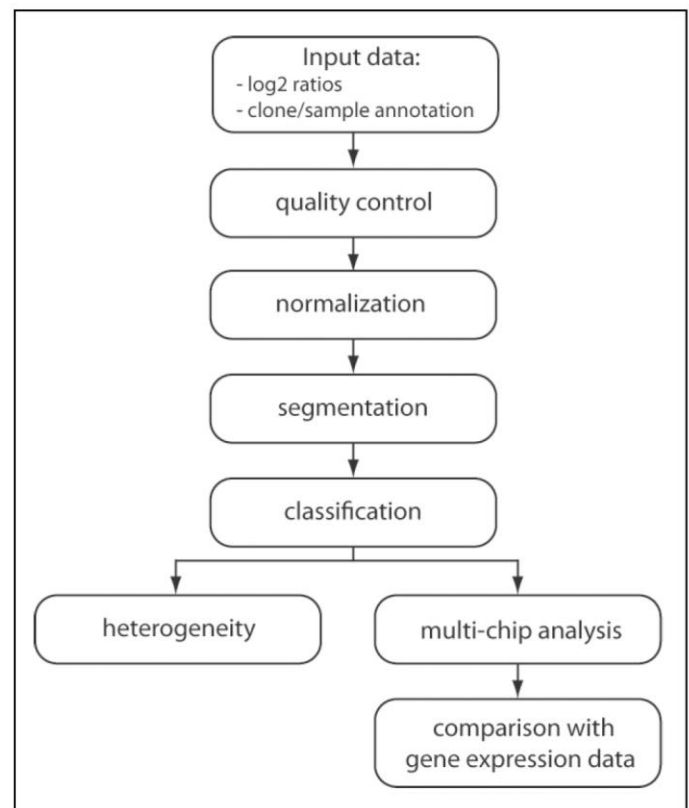
zero. Thus, a measure of location, like mean, median or mode, can be used to estimate the bias individually for each array. The bias is then removed by subtracting the measure of location from all measurements on this array. We take the 50% of the measurements with the highest values in the corresponding density function and calculate the mean of these values weighted with the appropriate density as an estimator for the mode of the distribution. We suggest using this mode estimator for normalization, because based on a simulation study it is more robust in cases with a relevant proportion of aberrant clones.

3.3 Segmentation

The goal of array CGH analysis is the detection of copy number aberrations in tumor DNA. Measured signals have to be classified in chromosomal gains, losses and regions with normal copy number. Copy numbers cannot be directly read off the signal ratios: The signal to noise ratio is rather low [13] and samples of primary tumors are typically contaminated with a certain amount of non-tumor DNA thus attenuating the ratio. A strategy for noise reduction is needed.

Copy number aberrations typically occur in segments comprising several clones. Segmentation methods aim at detecting segments of neighboring clones with the same copy number and to smooth the signal to reduce the noise.

Several segmentation algorithms for array CGH data are described in the literature [2, 7, 14-19]. We have implemented an interface to the Hidden Markov Model (HMM)-based segmentation method of Fridlyand et al. [17] and to the Circular Binary Segmentation (CBS) method of Olshen et al. [16]. Based on the results from other research groups [20, 21], and following our own experience, we currently recommend using the CBS method of Olshen.
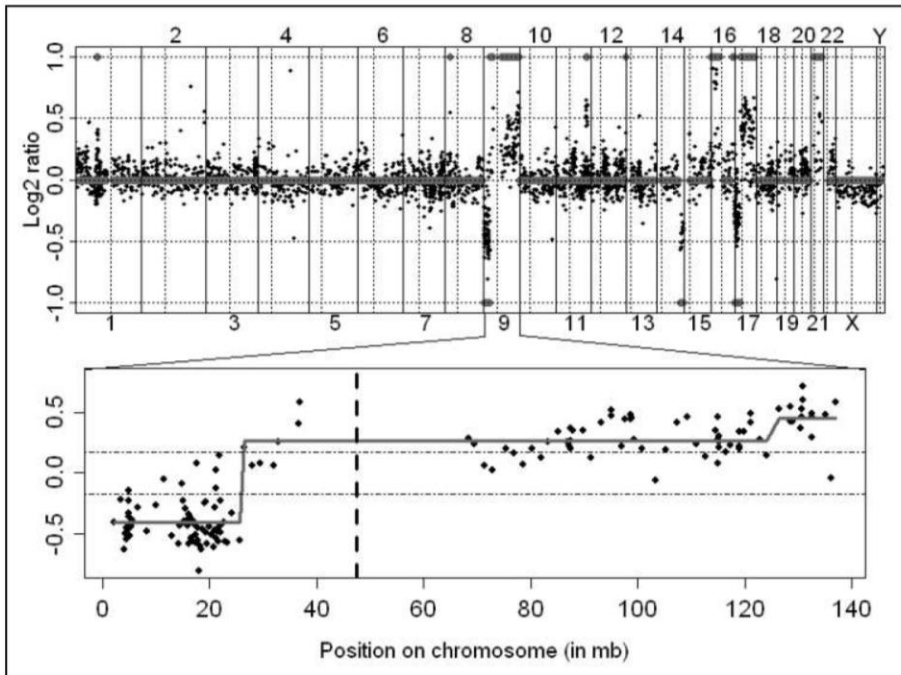
Fig. 2 The upper graphic illustrates an array CGH profile of a single chip. Black points show the normalized log2 ratios in genomic order. Grey bars indicate the classified data (0 = normal copy number; 1 = copy number gain; −1 = copy number loss). The lower graphic illustrates the segmentation for chromosome 9. The bold grey line marks the resulting segments. Thresholds for classification are indicated by dotted black lines.

3.4 Classification

After segmentation each array is represented by a set of segments and smoothed segment-specific signals. To ease interpretation the segments are classified into gains, losses and segments with normal copy number. Note that with current technology further differentiation within the loss or gain category is difficult except for high-level amplification peaks. Thresholds are required to achieve such a subdivision. If a segment level is above the threshold, the segment is classified as a gain. A segment which is below minus the threshold is counted as a loss. This process is also referred to as threshholding. aCGHPipeline supports choosing a fixed or a noise-dependent threshold. A fixed threshold is uniformly specified on the log2 ratio scale for all arrays. A noise-dependent threshold is specified in units of an array-specific noise estimate. For each array aCGHPipeline provides a robust noise estimator based on median absolute deviation of the differences between measured log2 ratios and the cor-

responding smoothed segment levels. Thus the noise estimate is not inflated by genetic heterogeneity. Choosing the noise-dependent approach leads to an increased threshold for chips with poor quality but not for chips with a high genetic heterogeneity.

Users have to specify the classification threshold manually considering information on the fraction of normal tissue contamination in tumor samples and the sensitivity of the array type used. For a given array platform the sensitivity can be assessed on the basis of data with known copy number aberrations or through the analysis of DNA with different numbers of X chromosomes [13, 22].

An example of classified array CGH data is shown in Figure 2. After classification the genetic heterogeneity of each array can be summarized, e.g. by calculating the number of aberrant segments or clones. These heterogeneity measures help to quantify the instability of the genome in different tumor entities or to analyze the impact of heterogeneity on the prognosis of patients.

3.5 Multi-chip Analysis

When analyzing a large set of tumor samples a major biological question is to detect regions which are commonly aberrant and therefore may be key events for tumor development or proliferation.

A major objective of multi-chip analysis is thus to delineate genomic regions which are characteristically aberrant in a specific type of cancer. These recurrent regions provide lists of putative candidate genes for biological follow-up research. Presence or absence of recurrent regions in single tissues define candidate variables for prognostic factor analyses.

Up to now these recurrent regions were visually determined e. g. by looking at heatmaps (Fig. 3b). We developed an algorithm to automatically propose recurrent regions. We first calculate and plot the frequencies of gains and losses for each clone along the genome (see Fig. 3a for a single chromosome). To delineate recurrent regions as "genomic regions which are characteristically aberrant in the data" we divide clones into segments with the same frequency of aberration and define recurrent regions as those resulting segments that are both above a certain threshold and dominate adjacent segments.

3.5.1 Segmentation of Multiple-chip Frequency Data

For segmentation we use a Hidden Markov Model (HMM) [23]. Each state has a binomial emission distribution with varying underlying probability p. HMMs with up to five states are fitted chromosome-wise to the frequencies of gains and losses of the clones because the number of recurrence levels is initially unknown. The model that fits best is selected via the Akaike information criterion (AIC). Smoothing is achieved through a state transition matrix giving high probability to staying in the same underlying state. The result of the segmentation procedure for gains and losses is illustrated in Figure 3a.

3.5.2 Threshold for Recurrence

As a next step we have to separate "characteristic" segments from sporadic ones. Every segment is therefore compared with a frequency threshold. Regions which show an aberration more frequently than the selected threshold were counted as recurrent segments. Recurrent segments that dominate neighboring segments are selected as recurrent regions.

The user can specify a frequency threshold or use a threshold that is suggested by our algorithm. aCGHPipeline uses a 2-means clustering of the frequencies of gains and losses of the clones to distinguish sporadic form characteristic aberrations. The mean of the center of the "sporadic aberrations" and the center of the "characteristic aberrations" centers is used as the threshold.

Automatically delineated recurrent regions should be checked manually by inspecting a heatmap because theoretically the identity of the cases contributing to a recurrent region may vary along the segment indicating that the region should be split.

3.5.3 Analysis of Recurrent Regions

Having delineated the recurrent regions in the data set one may want to decide on presence or absence of recurrent regions in single samples. We have implemented a voting algorithm for this purpose. A recurrent region consists of a set of clones. A recurrent region is called present in an individual sample if and only if a user-specified proportion (e.g. 50%) of its clone set is classified as concordant aberration.

The algorithm determines a matrix of CGH arrays by recurrent regions indicating presence and absence of the recurrent regions on the respective chip. The horizontal blue bar in Figure 3b indicates the result of the voting algorithm for the recurrent region marked by the vertical grey bar.

3.6 Analysis of Paired Array CGH and Gene Expression Data

If paired array CGH and gene expression data is available one can ask whether a re-
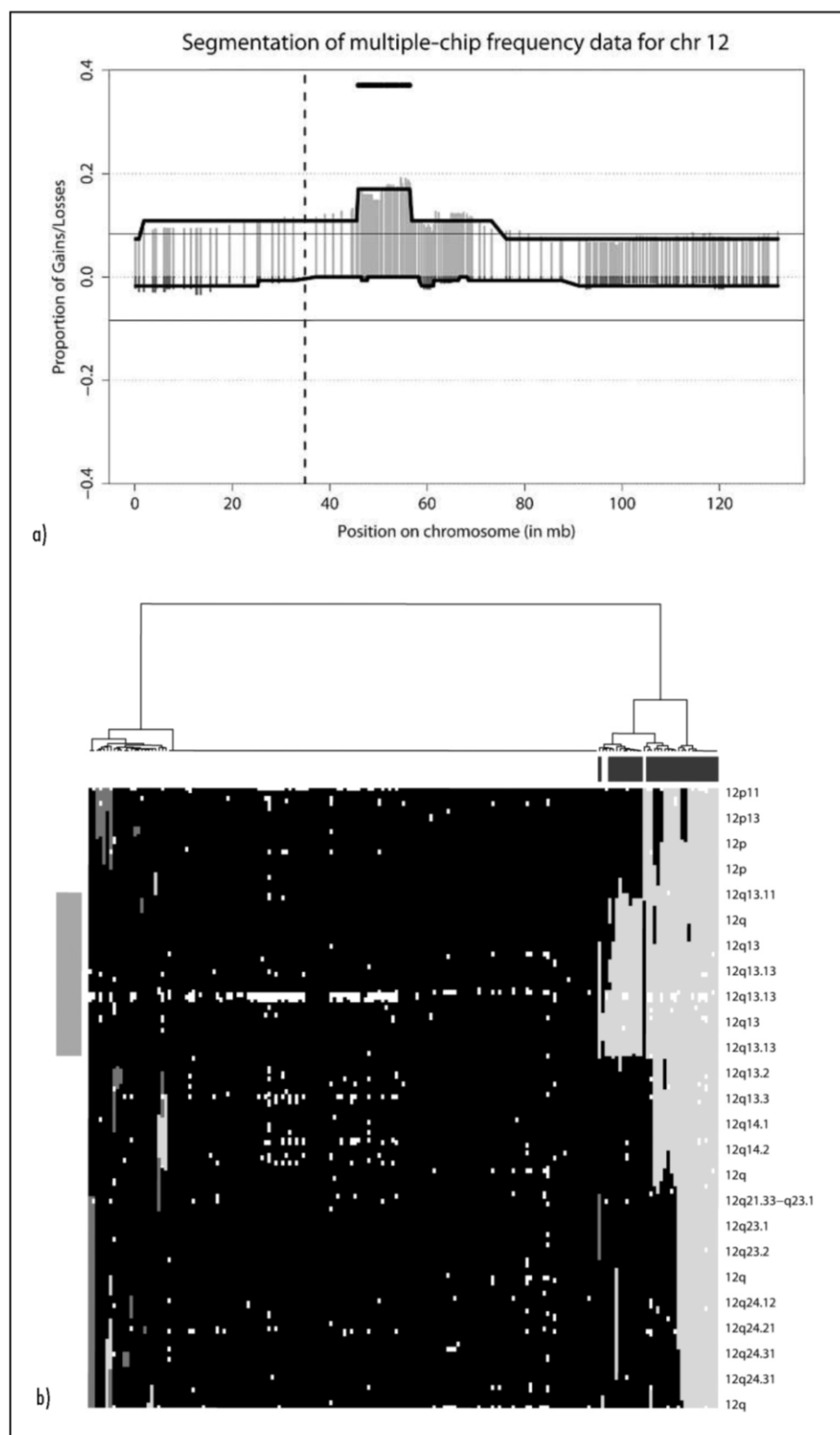
Fig. 3 a) Aggregated data of 183 lymphoma CGH arrays for chromosome 12. Light and dark grey bars indicate the proportion of gains and losses for each clone. Segmentation via HMM is shown by bold lines. Black horizontal lines mark the threshold for recurrence. The recurrent region (45.7-56.9 mb) is highlighted by a bold black bar. b) The heatmap of copy number aberrations on chromosome 12. Light grey lines indicate copy number gains, dark grey lines copy number losses. White points mark missing values. The bold bar on the left side shows the recurrent region. Black marks on the top indicate cases in which the recurrent gain is called present.
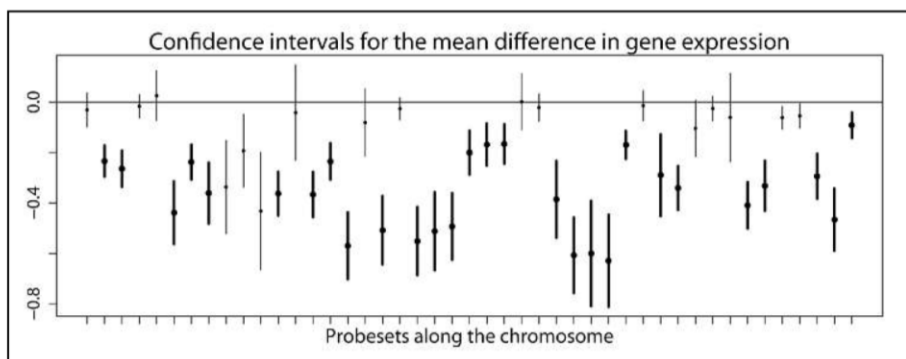
Fig. 4 Gene expression data in a recurrent loss on chromosome 6. Vertical lines illustrate the 95% confidence intervals for the differences in gene expression of cases with and without a loss. Bold lines mark probe sets with a significant gene dose effect.

spective aberration is associated with a gene dose effect on mRNA gene expression. Given the assignment of copy number aberrations in recurrent regions to individual samples as described above, mRNA expression in samples with the aberration can be compared to samples without this aberration.

Figure 4 depicts confidence intervals for the difference in mean gene expression comparing cases with and without a loss for genes mapping into a recurrent region on chromosome 6 in lymphoma data.

If a higher resolution is desired, a mapping assigning copy number information from individual clones to genes in the corresponding genomic region is required.

We have implemented an algorithm that generates a data pair for each gene expression probe set consisting of the respective mRNA expression level and a copy number state derived from the array CGH data. If a probe set maps on one or more CGH clones their copy number state is assigned if equivalent. If the probe set maps in a gap between clones on the CGH array the copy number state of the neighboring clones is returned if equivalent. Otherwise a missing value is assigned.

4. Application

The analysis pipeline was applied to a set of 189 arrays of aggressive B-cell lymphomas. Within the quality control step six arrays were rejected due to high noise. In addition 48 clones were excluded from the analysis because they were missing in more than 50 percent of the cases. 291 of the remaining clones are located in regions with known copy number polymorphisms.

Array CGH data was normalized using global normalization with mode estimator. Normalized log2 ratios were smoothed by applying the Circular Binary Segmentation (CBS) method of Olshen et al. with default parameters.

For a subset of 106 arrays, manual interpretation of the data was available. Best concordance between manual and automatic classification was achieved by using a fixed threshold of 0.17 for the classification of the segments resulting from CBS. In that case 96.8% of all clones were assessed consistently. More than 83% of the clones which were called aberrant by manual interpretation were concordantly detected by the automatic method. Investigation of discrepant results showed that in cases where one method assigned a copy number gain while the other assigned a loss mostly represented sign errors within the manual analysis. The bulk of discrepancies occurred at boundaries of aberrant regions and in cases where the putative aberration is quantitatively weak. In these cases a validation of the results is difficult.

Using the methods for multi-chip analysis 16 regions of recurrent gains and six regions of recurrent losses were delineated for further investigation.

5. Conclusion and Outlook

With aCGHPipeline we developed a comprehensive tool for the analysis of array CGH data. The package covers data input, quality control, normalization, segmentation and classification. The automatic analysis essentially reproduced the manual interpretation. It is less error-prone, less time-consuming and more reproducible, thus further analyses of our group will be based on it.

For multi-chip analysis aCGHPipeline proposes an algorithm for automatic delineation of recurrent regions. This task was performed manually up to now. The package also supports combined analysis with mRNA gene expression data. Results are presented as HTML documents to facilitate communication with clinical partners. aCGHPipeline is implemented in R in which further data analysis can easily be performed. The current version of the program provides no graphical user interface so that users need a basic knowledge of R programming. A planned integration of aCGHPipeline in the gene expression warehouse (GeWare) [24] will add this feature.

References

1. Albertson DG, Pinkel D. Genomic microarrays in human genetic disease and cancer. Hum Mol Genet 2003; 12: 145-152.
2. Picard F, Robin S, Lavielle M, Vaisse C, Daudin JJ. A statistical approach for array CGH data analysis. BMC Bioinformatics 2005; 11: 6-27.
3. Repsilber D, Ziegler A. Two-color Microarray Experiments. Methods Inf Med 2005; 44: 400-404.
4. Rahnenführer J. Image Analysis for cDNA Microarrays. Methods Inf Med 2005; 44: 405-407.
5. Kreuz M., Diplomarbeit: Entwicklung und Implementierung eines Auswertungswerkzeuges für Matrix-CGH-Daten. 2006.
6. Ihaka R, Gentleman RR. A Language for Data Analysis and Graphics. Journal of Computational and Graphical Statistics 1996; 5 (3): 299-314.
7. Autio R, Hautaniemi S, Kauraniemi P, Yli-Harja O, Astola J, Wolf M, Kallioniemi A. CGH-Plotter:

MATLAB toolbox for CGH-data analysis. Bioinformatics 2003; 19 (13): 1714-1715.

8. Chen W, Erdogan F, Ropers HH, Lenzner S, Ullmann R. CGHPRO – a comprehensive data analysis tool for array CGH. BMC Bioinformatics 2005; 6: 85.

9. Hummel M, Bentink S, Berger H, Klapper W, Wessendorf S, Barth TF, Bernd HW, Cogliatti SB, Dierlamm J, Feller AC, Hansmann ML, Haralambieva E, Harder L, Hasenclever D, Kühn M, Lenze D, Lichter P, Martin-Subero JI, Möller P, Müller-Hermelink HK, Ott G, Parwaresch RM, Pott C, Rosenwald A, Rosolowski M, Schwaenen C, Stürzenhofecker B, Szczepanowski M, Trautmann H, Wacker HH, Spang R, Loeffler M, Trümper L, Stein H, Siebert R; Molecular Mechanisms in Malignant Lymphomas Network Project of the Deutsche Krebshilfe. A biologic definition of Burkitt's lymphoma from transcriptional and genomic profiling. N Engl J Med 2006; 354 (23): 2419-2430.

10. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W, Cho EK, Dallaire S, Freeman JL, Gonzalez JR, Gratacos M, Huang J, Kalaitzopoulos D, Komura D, MacDonald JR, Marshall CR, Mei R, Montgomery L, Nishimura K, Okamura K, Shen F, Somerville MJ, Tchinda J, Valsesia A, Woodwark C, Yang F, Zhang J, Zerjal T, Zhang J, Armengol L, Conrad DF, Estivill X, Tyler-Smith C, Carter NP, Aburatani H, Lee C, Jones KW, Scherer SW, Hurles ME. Global variation in copy number in the human genome. Nature 2006; 444 (7118): 444-454.

11. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C. Detection of large-scale variation in the human genome. Nat Genetics 2004; 36 (9): 949-951.

12. Yang YH, Dudoit S, Luu P, Lin DM, Peng V, Ngai J, Speed TP. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. Nucleic Acids Res 2002; 30 (4): e15.

13. Bilke S, Chen QR, Whiteford CC, Khan J. Detection of low level genomic alterations by comparative genomic hybridization based on cDNA micro-arrays. Bioinformatics 2005; 21 (7): 1138-1145.

14. Wang P, Kim Y, Pollack J, Narasimhan B, Tibshirani R. A method for calling gains and losses in array CGH data. Biostatistics 2005; 6 (1): 45-58.

15. Hupé P, Stransky N, Thiery JP, Radvanyi F, Barillot E. Analysis of array CGH data: from signal ratio to gain and loss of DNA regions. Bioinformatics 2004; 20 (18): 3413-3422.

16. Olshen AB, Venkatraman ES, Lucito R, Wigler M. Circular binary segmentation for the analysis of array-based DNA copy number data. Biostatistics 2004; 5 (4): 557-572.

17. Fridlyand J, Snijders A, Pinkel D, Albertson D, Jain A. Hidden Markov models approach to the analysis of array CGH data. Journal of Multivariate Analysis 2004; 90 (1): 132-153.

18. Jong K, Marchiori E, Meijer G, Vaart AV, Ylstra B. Breakpoint identification and smoothing of array comparative genomic hybridization data. Bioinformatics 2004; 20 (18): 3636-3637.

19. Myers CL, Dunham MJ, Kung SY, Troyanskaya OG. Accurate detection of aneuploidies in array CGH and gene expression microarray data. Bioinformatics 2004; 20 (18): 3533-3543.

20. Willenbrock H, Fridlyand J. A comparison study: applying segmentation to array CGH data for downstream analyses. Bioinformatics 2005; 21 (22): 4084-4091.

21. Lai WR, Johnson MD, Kucherlapati R, Park PJ. Comparative analysis of algorithms for identifying amplifications and deletions in array CGH data. Bioinformatics 2005; 21 (19): 3763-3770.

22. Pollack JR, Sørlie T, Perou CM, Rees CA, Jeffrey SS, Lonning PE, Tibshirani R, Botstein D, Børresen-Dale AL, Brown PO. Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. Proc Natl Acad Sci U S A 2002; 99 (20): 12963-12968.

23. Rabiner LR, A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. Proceedings of the IEEE 1989; 77 (2): 257-286.

24. Rahm E, Kirsten T, Lange J. The GeWare data warehouse platform for the analysis of molecular-biological and clinical data. Journal of Integrative Bioinformatics 2007; 4 (1): 47

Correspondence to:
Markus Kreuz
University of Leipzig
Institute for Medical Informatics, Statistics and Epidemiology (IMISE)
Haertelstr. 16-18
04107 Leipzig
Germany
E-mail: markus.kreuz@imise.uni-leipzig.de

32

2.2 Publikation 2: „Recurrent loss of the Y chromosome and homozygous deletions within the pseudoautosomal region 1: association with male predominance in mantle cell lymphoma" Haematologica. 2008 Jun;93(6):949-50

Nieländer I, Martín-Subero JI, Wagner F, Baudis M, Gesk S, Harder L, Hasenclever D, Klapper W, Kreuz M, Pott C, Martinez-Climent JA, Dreyling M, Arnold N, Siebert R.

Das Mantelzelllymphom (MCL) ist ein malignes Lymphom aus der Gruppe der B-Zell-Non-Hodgkin-Lymphome. Es ist durch die Translokation zwischen den Chromosomen 11 und 14 (t(11;14)(q13;q32)) charakterisiert, welche zur Überexpression von *Cyclin-D1* führt und in nahezu allen MCL auftritt [45]. Männer sind deutlich häufiger betroffen als Frauen (Verhältnis: 2:1 oder höher [45]). Die Gründe für dieses Ungleichgewicht sind bisher nicht bekannt. Eine mögliche Ursache könnten genetische oder epigenetische Veränderungen auf den Geschlechtschromosomen sein, welche zur Tumorentstehung beitragen.

Im Rahmen der Studie wurden daher genetische Aberrationen auf den Geschlechtschromosomen untersucht. Es konnte gezeigt werden, dass häufig Verluste des Y-Chromosoms in Mantelzelllymphomen auftreten. In einem Datensatz, der am Institut für Humangenetik in Kiel untersucht wurde, traten Verluste des Y-Chromosoms in 22 von 80 (27,5 %) der untersuchten männlichen Patienten auf. Mittels Dreifarben-Fluoreszenz-in-situ-Hybridisierung (FISH) wurde das Auftreten der t(11;14)-Translokation und des Verlusts von Chromosom Y gemeinsam untersucht. Damit konnte gezeigt werden, dass in einem Großteil der untersuchten Proben die Verluste von Chromosom Y klonal in den t(11;14)-positiven Tumorzellen waren. In 15 von 18 untersuchten Proben trat der Chromosom-Y-Verlust in den Lymphomzellen deutlich häufiger auf als in den t(11;14)-negativen Bystander-Zellen.

Im Rahmen der Studie wurden auch 20 MCL-Proben von männlichen Patienten mittels Affymetrix 100k-SNP-Arrays untersucht (vgl. Manuskript 3). Für die Analyse der SNP-Array-Daten wurde die in Abschnitt 1.4 beschriebene Auswertepipeline verwendet. Das Y-Chromosom ist auf den Affymetrix 100k-Arrays nur durch die Pseudoautosomalen Regionen (PARs) repräsentiert. Diese Abschnitte des Genoms sind auf dem X- und Y- Chromosom gleichermaßen vorhanden. Die Pseudoautosomale Region 1 (PAR1) umfasst circa 2,6 Megabasen (Mb) [46] und ist auf dem 100k-SNP-Array durch 19 SNPs repräsentiert. Verluste von Chromosom Y zeigen sich durch einfache Deletionen in den 19 betreffenden SNPs. Mittels der SNP-Array-Analyse konnten für zwei Lymphomfälle homozygote Deletionen der PAR1-Region gezeigt werden. In diesen Fällen war die genetische Information sowohl auf dem X- als auch auf dem Y-Chromosom deletiert. Mittels FISH-Analyse wurden die homozygoten Deletionen für die PAR1-Gene *SHOX*, *CSF2RA* und *CRLF2* in beiden Fällen erfolgreich verifiziert und zwei weitere Fälle mit homozygoten Deletionen in einem unabhängigen Datensatz detektiert.

Zusammenfassend konnte gezeigt werden, dass die Auswertepipeline für SNP-Arrays (vgl. Abschnitt 1.4) auch zur Detektion von Kopienzahlaberrationen auf den Geschlechtschromosomen X und Y geeignet ist. Darüber hinaus konnte gezeigt werden, dass Verluste von Chromosom Y häufig als sekundäre Aberrationen in MCL auftreten und die Gene in der Region PAR1 für die Entstehung von MCL potenziell bedeutsame Tumorsuppressorgene sind.

Recurrent loss of the Y chromosome and homozygous deletions within the pseudoautosomal region 1: association with male predominance in mantle cell lymphoma

Mantle cell lymphoma (MCL) is a B-cell lymphoprolif-erative disorder which predominantly affects men. In a large retrospective survey of the European MCL Network including 304 patients, median age was 63 years at first diagnosis with a male preponderance of 76%.[1] The genet-ic hallmark of MCL is the translocation t(11;14)(q13;q32) which leads to overexpression of the *CCND1* gene encoding Cyclin D1. Although recent studies revealed a number of genomic alterations and differentially expressed genes in MCL, the causes for the male predom-inance are still unknown. Hormonal differences might contribute to this gender imbalance. Another hypothesis is that male predominance in MCL results from a sex chromosome linked genetic or epigenetic alteration. Interestingly, some cytogenetic studies on MCL reported recurrent loss of the whole chromosome Y.

Using the karyotype parsing software from the Progenetix project [*www.progenetix.net*] on data from the Mitelman database [available form URL: *http://cgap.nci.nih.gov/Chromosomes/Mitelman, October 2007 edition*], we show that 42 out of 365 (11.5%) cytogeneti-cally analyzed t(11;14)-positive B-cell non-Hodgkin's lymphoma (B-NHL) in male patients (excluding plasmo-cytoma/multiple myeloma, *NHL not otherwise specified* and immature NHL) harbor a deletion of the Y chomosome. Similarly, 22 out of 80 (27.5%) t(11;14)-positive lym-phoma in male patients cytogenetically analyzed in the Institute of Human Genetics (University Hospital Schleswig-Holstein, Campus, Kiel, Germany) showed loss of the Y chromosome. These findings indicate that the loss of the Y chromosome is a recurrent cytogenetic event in MCL.

Loss of the Y chromosome is known to be a common aging phenomenon in cells of elderly males,[2] but recent studies also suggest significance in tumor development, particularly in prostate cancer.[3,4] To differentiate between age-related random and clonal losses of the Y chromo-some, we analyzed whether loss of chromosome Y is restricted to the t(11;14)-positive MCL tumor cells or whether this change also occurs in the normal cells of the biopsies, e.g. due to an age-related effect. Therefore, we performed triple-color Fluorescence *in situ* hybridization (FISH) in 21 MCL of male patients harboring a chromo-some Y loss by conventional cytogenetic analysis. The commercially available locus-specific identifier (LSI) IGH/CCND1 dual color, dual fusion translocation probe (Abbott/Vysis, Downers Grove, IL, USA) was used to detect the translocation t(11;14)(q13;q32)/IGH-CCND1 fusion and combined with a probe for chromosome region Yq12 (CEP Y Sat III, spectrum aqua, Abbott/Vysis) (Figure 1 AB). FISH was performed using reported meth-ods[5] on fixed cells from cytogenetic analyses. The cytoge-netically described loss of the Y chromosome was con-firmed by FISH in all but 3 of the 21 MCL cases. Cytogenetic studies of these 3 MCL cases showed mark-er chromosomes, which presumably harbor the Yq12 chromosome material detected by the FISH probe. FISH analyses demonstrated that in 12 of the remaining 18 MCL, loss of chromosome Y was present in the clone with the translocation t(11;14) whereas a maximum of

3% (median 1.1%, range 0-3%) of the t(11;14)-negative cells showed Y chromosomal loss (*Online Supplementary Figure 1A*). In 3 other MCL cases, 9% (case 15), 12% (case 5) and 16% (case 13) of the t(11;14)-negative cells har-bored a chromosome Y deletion, but loss of chromosome Y was obviously more frequent (5.9–8.3 fold) in the t(11;14)-positive cells. Accordingly, loss of chromosome Y seems to be a clonal feature of the t(11;14)-positive tumor cells in 83% of the investigated MCL. In 5 of the 15 MCL cases with clonal chromosome Y loss, we also identified at least 3% t(11;14)-positive cells without chromosome Y deletion (cases 11-15). The presence of the Y chromo-some in these lymphoma cells indicates that loss of chro-mosome Y occurred secondary to t(11;14).

It is widely assumed that the Y chromosome predomi-nantly lacks genes involved in oncogenesis. Most of its few genes are localized at the pseudoautosomal regions (PARs) which show homology and recombine with two segments on the X chromosome. A recent array CGH study identified complete loss of the terminal short arm of the X chromosome including the Kallmann (KAL) gene locus in Xp22.31 and the PAR1 in one t(11;14)-positive primary MCL.[6] Here we evaluated recently obtained 100K GeneChip data (GeneChip® Human Mapping 100K Set, Affymetrix, Santa Clara, CA, USA) (Nieländer *et al.*, *in preparation*) of 20 primary MCL from male patients with regard to aberrations targeting the pseudoautosomal regions of the sex chromosomes. Among these were 5 cases with chromosome Y loss detected by FISH (cases 4, 6, 9, 14 and 18).

Probe preparation and array hybridization was per-formed according to the GeneChip® Human Mapping 100K assay protocol (Affymetrix, Santa Clara, CA, USA) (*http://www.affymetrix.com*). Copy number analy-sis was performed using the CNAG program v2.0.[7]

We used 90 Hapmap samples provided by Affymetrix (30 CEPH trios) as euploid reference arrays (*http://www.affymetrix.com/support/technical/sam-ple_data/hapmap_trio_data.affx*). Segmentation of raw copy number data was performed using the Hidden Markov Model approach provided by CNAG.

There are no tags for chromosome Y on the microarray except for the PAR. Remarkably, in 2 of the MCL cases with chromosome Y deletion, GeneChip analysis identi-fied a region of complete loss in the pseudoautosomal region 1 (PAR1) in Xp/Yp (*Online Supplementary Figure 1C*). According to the GeneChip data, the minimally deleted region spanned approximately 2.5 Mb from the Xp-telomere to the SNP "rs5982788", including 16 pseudoautosomal genes (NCBI Build 35) (*Online Supplementary Figure 1, Online Supplementary Table 1*). To confirm and extend these findings, FISH analyses were performed using locus-specific probes, which consisted of differentially labeled bacterial artificial chromosome (BAC/PAC) clones. In this way, homozygous loss of the *SHOX* (RP11-800K15), *CSF2RA* (RP4-674K6), and *CRLF2* (RP11-475E20) genes was confirmed in both MCL show-ing biallelic loss of the PAR1 in the GeneChip data. Furthermore, a FISH screening of the remaining 16 MCL with chromosome Y loss detected one additional case showing biallelic loss of all these gene loci. The case described by Rubio-Moscardo *et al.* was also confirmed by FISH to harbor biallelic loss of the PAR1.[6] To further confirm the border of the homozygously deleted region, we performed FISH analyses of the 3 MCL with Xp-dele-tion detected here using the LSI Kallmann Region Probe (Abbott/Vysis). In 2 of these cases, homozygous loss did not comprise the *KAL* gene locus. The amount of detect-

34

ed cells showing biallelic loss of the PAR1 in the samples studied here corresponded approximately to the tumor cell content which was calculated by t(11;14)-FISH (Figure 1D). The signal constellation in the putative non-tumorous cells indicated that both X-chromosomal and Y-chromosomal PAR1 were present, ruling out a constitutional copy number polymorphism (*Online Supplementary Figure 1E*). These findings suggest deletion of the X-chromosomal PAR1 to be restricted to the t(11;14)-positive cells in these MCL.

To summarize, the clonality of chromosome Y loss in t(11;14)-positive tumor cells provides evidence for its relevance as secondary aberration in MCL. It seems that its role in MCL development might have been underestimated. Our data may also rule out the hypothesis that the loss of chromosome Y in lymphoid cells may predispose these cells to the development of MCL, thus explaining male predominance. The detection of a deletion in Xp22.33 in addition to chromosome Y loss suggests the involvement of a pseudoautosomal tumor suppressor gene (TSG). Inactivation of a PAR-linked candidate TSG in combination with Y chromosome loss through an age-related effect in elderly cells could be a plausible explanation for the male predominance observed in MCL. Remarkably, pseudoautosomal linkage has also been proposed for Hodgkin's lymphoma.[8] It may be of particular interest that a cluster of cytokine-receptor genes resides on the PAR1 region.[9] Aberrant expression of chemokines and chemokine receptors has been reported to play a role in malignant hematopoietic cells.[10,11] Future studies, including mutation and expression analyses of genes located in PAR1, might lead to the identification of candidate genes involved in MCL lymphomagenesis.

Inga Nieländer,[1] José I. Martín-Subero,[1] Florian Wagner,[3] Michael Baudis,[4] Stefan Gesk,[1] Lana Harder,[1] Dirk Hasenclever,[5] Wolfram Klapper,[6] Markus Kreuz,[5] Christiane Pott,[7] José A. Martinez-Climent,[8] Martin Dreyling,[9] Norbert Arnold,[2] and Reiner Siebert[1]

[1]Institute of Human Genetics, University Hospital Schleswig-Holstein, Campus Kiel, Christian-Albrechts-University, Kiel, Germany; [2]Department of Gynecology and Obstetrics, University Hospital Schleswig-Holstein, Campus Kiel, Christian-Albrechts-University, Kiel, Germany; [3]German Resource Center for Genome Research (RZPD), Berlin, Germany; current address: ATLAS Biolabs GmbH, Berlin, Germany; [4]Institute of Molecular Biology, University of Zürich, Zürich, Switzerland; [5]Institute for Medical Informatics, Statistics and Epidemiology (IMISE), University of Leipzig, Leipzig, Germany; [6]Institute of Pathology, Section Hematopathology and LN Registry, University Hospital Schleswig-Holstein, Campus Kiel, Christian-Albrechts-University, Kiel, Germany; [7]Second Medical Department, University Hospital Schleswig-Holstein, Campus Kiel, Christian-Albrechts-University, Kiel, Germany; [8]Center for Applied Medical Research CIMA, University of Navarra, Pamplona, Spain; [9]Department of Medicine III, University Hospital Grosshadern/LMU, CCG Leukemia, GSF-National Research Center for Environment and Health, Munich, Germany

References

1. Tiemann M, Schrader C, Klapper W, Dreyling MH, Campo E, Norton A, et al. Histopathology, cell proliferation indices and clinical outcome in 304 patients with mantle cell lymphoma (MCL): a clinicopathological study from the European MCL Network. Br J Haematol 2005;131:29-38.
2. Anonymous. Loss of the Y chromosome from normal and neoplastic bone marrows. United Kingdom Cancer Cytogenetics Group (UKCCG). Genes Chromosomes Cancer 1992;5:83-8.
3. Vijayakumar S, Hall DC, Reveles XT, Troyer DA, Thompson IM, Garcia D, et al. Detection of recurrent copy number loss at Yp11.2 involving TSPY gene cluster in prostate cancer using array-based comparative genomic hybridization. Cancer Res 2006;66:4055-64.
4. Pyakurel P, Montag U, Castaños-Vélez E, Kaaya E, Christensson B, Tönnies H, et al. CGH of microdissected Kaposi's sarcoma lesions reveals recurrent loss of chromosome Y in early and additional chromosomal changes in late tumour stages. AIDS 2006;20:1805-12.
5. Martín-Subero JI, Harder L, Gesk S, Schlegelberger B, Grote W, Martinez-Climent JA, et al. Interphase FISH assays for the detection of translocations with breakpoints in immunoglobulin light chain loci. Int J Cancer 2002;98:470-4.
6. Rubio-Moscardo F, Climent J, Siebert R, Piris MA, Martín-Subero JI, Nieländer I, et al. Mantle-cell lymphoma genotypes identified with CGH to BAC microarrays define a leukemic subgroup of disease and predict patient outcome. Blood 2005;105:4445-54.
7. Nannya Y, Sanada M, Nakazaki K, Hosoya N, Wang L, Hangaishi A, et al. A robust algorithm for copy number detection using high-density oligonucleotide single nucleotide polymorphism genotyping arrays. Cancer Res 2005;65:6071-9.
8. Horwitz MS, Mealiffe ME. Further evidence for a pseudoautosomal gene for Hodgkin's lymphoma: Reply to 'The familial risk of Hodgkin's lymphoma ranks among the highest in the Swedish Family-Cancer Database' by Altieri A and Hemminki K. Leukemia 2007;21:351.
9. Kremer E, Baker E, D'Andrea RJ, Slim R, Phillips H, Moretti PA, et al. A cytokine receptor gene cluster in the X-Y pseudoautosomal region? Blood 1993;82:22-8.
10. Trentin L, Cabrelle A, Facco M, Carollo D, Miorin M, Tosoni A, et al. Homeostatic chemokines drive migration of malignant B cells in patients with non-Hodgkin lymphomas. Blood 2004;104:502-8.
11. Durig J, Schmucker U, Duhrsen U. Differential expression of chemokine receptors in B cell malignancies. Leukemia 2001; 15:752-6.

35

2.3 Publikation 3: „GeneChip analyses point to novel pathogenetic mechanisms in mantle cell lymphoma" Br J Haematol. 2009 Feb;144(3):317-31

Vater I, Wagner F, Kreuz M, Berger H, Martín-Subero JI, Pott C, Martinez-Climent JA, Klapper W, Krause K, Dyer MJ, Gesk S, Harder L, Zamo A, Dreyling M, Hasenclever D, Arnold N, Siebert R.

Basierend auf Mausexperimenten konnte gezeigt werden, dass die für MCL charakteristische Translokation t(11;14)(q13;q32) für die Lymphomentstehung nicht ausreicht [47], sondern weitere sekundäre Aberrationen für die Transformation notwendig sind.

Im Rahmen dieser Studie wurden daher 26 MCL-Proben sowie sechs MCL-Zelllinien mittels Affymetrix 100k-SNP-Arrays untersucht. Ziel war es, sekundäre Aberrationen in MCL zu beschreiben, wobei der Schwerpunkt auf der Detektion von UPDs und fokalen High-Level-Amplifikationen sowie homozygoten Deletionen lag. Diese konnten mit klassischen CGH- und Array-CGH-Analysemethoden bisher nur unzureichend charakterisiert werden. Aus Voruntersuchungen war bekannt, dass UPDs rekurrente genetische Aberrationen in MCL-Zelllinien sind [48]. Als euploide Kontrollen dienten 90 frei verfügbare HapMap-Fälle. Die Daten wurden wie in Abschnitt 1.4 beschrieben analysiert. Da keine laboreigenen Referenzen vorlagen, zeigten die Kopienzahlprofile vergleichsweise hohe Streuung. Um die Menge falsch positiver Aberrationen zu reduzieren, wurden daher Aberrationen mit weniger als 10 benachbarten SNPs verworfen.

Die dabei beobachteten Aberrationsmuster zeigten eine große Konkordanz bezüglich rekurrenter Aberrationen in vorherigen Kopienzahl-Analysen bei MCL [49]. Es konnten häufige Zugewinne auf den Chromosomenarmen 3q, 8q, 10p, sowie Verluste auf den Chromosomenarmen 1p, 6q, 8p, 9p, 9q, 10p, 11q, 13q und 17p gezeigt werden. Zusätzlich konnten fünf Regionen mit Zugewinnen in ≥4 Fällen und sieben Regionen mit Verlusten in ≥4 Fällen gezeigt werden, die in der Literatur bisher nicht beschrieben wurden. Für die Region 11q13.4–q13.5 konnte eine rekurrente High-Level-Amplifikation nachgewiesen werden (in drei Zelllinien detektiert), die zwei snoRNAs und neun Gene, unter anderem *MAP6*, umfasst. Für die Analyse der homozygoten Deletionen wurde zusätzlich zu den Standardkriterien (geschätzte Kopienzahl=0 für ≥10 benachbarte SNPs) ein sensitives Screening mit dem Kriterium: geschätzte Kopienzahl ≤0,6 für mindestens zwei benachbarte SNPs, von denen mindestens einer einen „No-Call" zeigte, durchgeführt. Alle elf homozygoten Deletionen, die mittels konservativen Kriterien in MCL-Zelllinien detektiert wurden, konnten mittels FISH-bzw. PCR-Analyse bestätigt werden. Unter anderem auch eine homozygote Deletion des Gens *MAP2*. Von den 17 zusätzlichen, mittels sensitiver Kriterien bestimmten homozygoten Deletionen konnten hingegen nur vier bestätigt werden. In den Primärtumoren konnte aufgrund der Kontamination durch Normalzellen keine Validierung erfolgen.

Es konnte gezeigt werden, dass die Chromosomenarme 11q und 13q zusätzlich zu Deletionen auch rekurrent von UPDs betroffen waren, wobei der Tumorzellgehalt der Primärfälle zwischen 21 % und 95 % (Mittelwert 75 %) schwankte. Die Sensitivität der UPD-Analyse ist bei niedrigem Tumorzellgehalt gering, sodass die Rekurrenz der UPD-Ereignisse unterschätzt wird [50]. Für einen Lymphomfall wurde neben einer UPD-Region auf 17p auch eine homozygote Mutation von *TP53* sowie normale Kopienzahl mittels FISH gezeigt. Somit wurde die Deaktivierung eines Tumorsuppressorgenes durch UPD nachgewiesen.

Im Rahmen der Studie wurde gezeigt, dass mit *TP53*, *MAP2* und *MAP6* mehrere Gene Aberrationen zeigen, die mit der Organisation von Mikrotubuli assoziiert sind. Mittels Expressions-, Methylierungs- und Mutationsanalyse konnte am Institut für Humangenetik in Kiel weitere Evidenz für die Bedeutung dieser Gene in MCL gewonnen werden.

GeneChip analyses point to novel pathogenetic mechanisms in mantle cell lymphoma

Inga Vater,[1] Florian Wagner,[2]* Markus Kreuz,[3] Hilmar Berger,[3] José I. Martín-Subero,[1] Christiane Pott,[4] Jose A. Martinez-Climent,[5] Wolfram Klapper,[6] Kristina Krause,[6] Martin J. S. Dyer,[7] Stefan Gesk,[1] Lana Harder,[1] Alberto Zamo,[8] Martin Dreyling,[9] Dirk Hasenclever,[3] Norbert Arnold[10] and Reiner Siebert[1]

[1]Institute of Human Genetics, Christian-Albrechts University Kiel, Kiel, Germany, [2]German Resource Centre for Genome Research (RZPD), Berlin, Germany, [3]Institute for Medical Informatics, Statistics and Epidemiology (IMISE), University of Leipzig, Leipzig, Germany, [4]Second Medical Department, Christian-Albrechts University Kiel, Kiel, Germany, [5]Centre for Applied Medical Research (CIMA), University of Navarra, Pamplona, Spain, [6]Institute of Pathology, Section Haematopathology and Lymph Node Registry, Christian-Albrechts University Kiel, Kiel, Germany, [7]MRC Toxicology Unit, University of Leicester, Leicester, UK, [8]Department of Pathology, University of Verona, Verona, Italy, [9]Department of Medicine III, University Hospital Grosshadern/LMU, CCG Leukaemia, GSF-National Research Centre for Environment and Health, Munich, Germany, and [10]Department of Gynaecology and Obstetrics, Christian-Albrechts University Kiel, Kiel, Germany

Summary

The translocation t(11;14)(q13;q32) is the genetic hallmark of mantle cell lymphoma (MCL) but is not sufficient for inducing lymphomagenesis. Here we performed genome-wide 100K GeneChip Mapping in 26 t(11;14)-positive MCL and six MCL cell lines. Partial uniparental disomy (pUPD) was shown to be a recurrent chromosomal event not only in MCL cell lines but also in primary MCL. Remarkably, pUPD affected recurrent targets of deletion like 11q, 13q and 17p. Moreover, we identified 12 novel regions of recurrent gain and loss as well as 12 high-level amplifications and eight homozygously deleted regions hitherto undescribed in MCL. Interestingly, GeneChip analyses identified different genes, encoding proteins involved in microtubule dynamics, such as *MAP2*, *MAP6* and *TP53*, as targets for chromosomal aberration in MCL. Further investigation, including mutation analyses, fluorescence *in situ* hybridisation as well as epigenetic and expression studies, revealed additional aberrations frequently affecting these genes. In total, 19 of 20 MCL cases, which were subjected to genetic and epigenetic analyses, and five of six MCL cell lines harboured at least one aberration in *MAP2*, *MAP6* or *TP53*. These findings provide evidence that alterations of microtubule dynamics might be one of the critical events in MCL lymphomagenesis contributing to chromosomal instability.

Keywords: mantle cell lymphoma, microarray, single nucleotide polymorphisms, microtubule-associated proteins, partial uniparental disomy.

Mantle cell lymphoma (MCL) is an aggressive B-cell non-Hodgkin lymphoma (B-NHL) genetically characterised by the translocation t(11;14)(q13;q32) that leads to overexpression of cyclin D1 (Williams *et al*, 1992). Based on studies with transgenic mice, it is well established that this chromosomal event alone is not sufficient to result in lymphomagenesis and that secondary genomic alterations are required for malignant transformation (Lovec *et al*, 1994; Gladden *et al*, 2006). In this multistep transformation process, tumour suppressor gene (TSG) inactivation has been shown to

be a key mechanism. In MCL, frequently targeted TSGs encode for proteins that inhibit malignant transformation by protecting the genome from DNA damage (*ATM*) or from deregulated cell cycle progression (*CDKN2A/P16, RB1*) or by inducing apoptosis in cells with a disrupted cell cycle control (*TP53*)(Nielander *et al*, 2007).

MCL has been the subject of several array-based comparative genomic hybridisation (arrayCGH) studies, and genomic aberrations in addition to t(11;14)(q13;q32) have been extensively characterised (Kohlhammer *et al*, 2004; de Leeuw *et al*, 2004; Rubio-Moscardo *et al*, 2005a,b; Tagawa *et al*, 2005; Mestre-Escorihuela *et al*, 2007). Except one, all these studies used microarrays consisting of bacterial artificial chromosome (BAC) or P1-artificial chromosome (PAC) clones for which maximal resolution is restricted to 50–200 kb. Microarrays consisting of short synthetic oligonucleotides provide higher resolution. Some platforms allow genotyping simultaneously, so that loss of heterozygosity (LOH) without changes in DNA copy number is also detectable. This chromosomal event, so called partial uniparental disomy (pUPD), has been recently reported as a frequent mechanism of TSG inactivation in haematological neoplasms and solid tumours (Bignell *et al*, 2004; Bruce *et al*, 2005). In 2006, we could show that pUPD is a recurrent genetic mechanism alternative to chromosomal deletion in MCL cell lines (Nielaender *et al*, 2006). By loss of (part of) one parental chromosome and duplication of the remaining one from the other parent, pUPD results in LOH without chromosomal deletion. As the gene dosage is not altered, pUPD cannot be detected by conventional arrayCGH.

In the present study, we used 100 K GeneChip arrays to determine genomic imbalances and pUPD in a series of 26 MCL patients and six MCL cell lines. Taking advantage of the high resolution of this array, we focused on detection of small homozygous deletions. Moreover, this microarray enabled the identification of pUPD. The detection of these two genetic alterations has been shown to be a promising strategy to identify novel TSGs (Nielander *et al*, 2007). We identified minimally altered regions targeted by homozygous deletions harbouring novel candidate TSG loci, which might be associated with MCL pathogenesis. Moreover, we could show that pUPD is a recurrent chromosomal event not only in MCL cell lines, but also in MCL primary cases and leads to inactivation of TSGs. This further confirms that pUPD has to be considered as an alternative mechanism of TSG inactivation in MCL. Finally, our study identified genes encoding microtubule-associated proteins to be frequently targeted by chromosomal and epigenetic aberrations in MCL.

Methods and materials

Tissue samples

DNA-samples and fixed cells were selected from the files of the Institute of Human Genetics, the Second Medical Department

and the Institute of Haematopathology, (University Hospital Schleswig-Holstein, Campus, Kiel, Germany). Genomic DNA was extracted from frozen dimethyl sulphoxide stocks derived from lymph node (LN), peripheral blood (PB), bone marrow (BM) and spleen (S) or from frozen tissue blocks of the affected LN using the QIAamp DNA Blood Mini Kit (Qiagen, Hilden, Germany). The MCL panel for the single nucleotide polymorphism (SNP) array comprised 26 patients with primary MCL. The median age of the patients at diagnosis was 66 years (range, 38–84 years). A subset of cases of this panel was also used for fluorescence *in situ* hybridisation (FISH) (*n* = 13), *MAP2* mutation analyses (*n* = 20), *TP53* mutation analyses (*n* = 26) and epigenetic studies (*n* = 20). For the FISH screens and denaturing high performance liquid chromatography (DHPLC) analyses of *MAP2*, an additional 14 and 20 t(11;14)-positive MCL cases from files of the Institute of Human Genetics (University Hospital Schleswig-Holstein, Campus, Kiel, Germany) were included, respectively. Table SI shows all MCL samples included in this study. Tonsils and PB of healthy donors were used as non-tumours controls. For real time RT-PCR experiments, LN cDNA pooled from 12 Caucasians (aged 20–59 years) and purchased from the human immune system multiple tissue cDNA panel (Clontech Laboratories, Mountain View, CA, USA) and a tonsil freshly prepared from the Institute of Haematopathology (University Hospital Schleswig-Holstein, Campus, Kiel, Germany) were used as normal controls. The study was performed in the framework of the 'European MCL Network', for which central and local ethics approval was obtained.

Cell lines

A panel of six MCL cell lines was studied: GRANTA-519, HBL-2, UPN-1, REC-1, MAVER-1 and JEKO-1. A compilation of studies characterising these cell lines is shown in Table SII.

Microarrays

The GeneChip Human Mapping 100 K array set (Affymetrix, Santa Clara, CA, USA) has been used according to the protocol provided by the manufacturer (Affymetrix) (http://www.affymetrix.com). Microarrays were washed and stained with the Fluidics Station 450 (Affymetrix) and scanned with the GeneChip Scanner 3000 (Affymetrix) using GeneChip Operating System (GCOS; Affymetrix) version 1.4. The BRLMM algorithm (http://www.affymetrix.com/support/technical/whitepapers/brlmm_whitepaper.pdf) was used with default parameters (score threshold = 0·5, prior size = 10 000 and DM threshold = 0·17) to genotype MCL tumour samples in combination with Hapmap reference arrays (http://www.affymetrix.com/support/technical/whitepapers/brlmm_whitepaper.pdf). The genotyping call rates of the hybridised SNP chips ranged from 96·06% to 99·88% (median = 98·59%) for the 50K XbaI array and from 94·54% to 99·77% (median = 98·92%) for the 50 K HindIII array. Ninety HapMap samples provided by

38

Affymetrix (30 CEPH trios) were used as normal reference arrays (http://www.affymetrix.com/support/technical/sample_data/hapmap_trio_data.affx). A complete list of Affymetrix reference samples is shown in Table SIII.

Copy number analysis

Copy number analysis was performed using the Copy Number Analyser for GeneChip (CNAG) program v2.0 (Nannya *et al*, 2005). The optimised reference selection method implemented in CNAG was used (Nannya *et al*, 2005). Thus, CNAG selected a gender-specific reference set out of the 90 controls individually for each array. XbaI and HindIII arrays were combined for the analysis. Segmentation of raw copy number data was performed using the Hidden Markov Model (HMM) approach provided by CNAG. HMM parameters were adjusted individually for each array due to differences in hybridisation quality and tumour cell content. Starting with default parameters, the mean levels of HMM states were adjusted manually to increase smoothing for samples with higher noise while increasing sensitivity for low noise samples. With regard to outliers and technical artefacts, HMM segments with aberrant copy number were considered as a copy number aberration only if they consisted of at least 10 consecutive SNPs. High-level amplifications were defined as aberrations with HMM copy number ≥5. For male cases, the call of high-level amplifications on chromosome X was copy numbers ≥4.

Liberal screening for homozygous deletions. Homozygous deletions were defined as aberrations with copy number = 0.

Sensitive screening for homozygous deletions. In this more sensitive approach for detecting homozygous deletions, the Copy Number Analysis Tool (CNAT; Affymetrix) version 2.0 was used to calculate the copy number (CN), applying a 0·5 Mb genome smoothing filter. The data set was screened for regions with a copy number ≤0·6 in a minimum of two adjacent SNPs of which at least one showed a 'NoCall'.

LOH analysis

A HMM-based method (Beroukhim *et al*, 2006) implemented in the dChip program (Lin *et al*, 2004; Zhao *et al*, 2004) (Build date: Apr 11 2007) was used to infer regions with LOH from tumour samples. The HMM considering haplotype (LD-HMM) (Beroukhim *et al*, 2006) method was selected for the LOH calculations to account for linkage disequilibrium (LD)-induced SNP dependencies. The LOH call threshold was set to the default value of 0·5. An empirical haplotype correction (Beroukhim *et al*, 2006) was applied. Thus, the genotypes of putative tumour-associated LOH regions were compared with the genotypes observed in euploid reference samples. If the genotypes in the respective region were highly concordant between the tumour sample and at least 5% of the normal reference samples the LOH region was rejected by dChip.

Partial uniparental disomy

Partial UPD regions represent genomic regions in which the LOH was not caused by altered copy number. A LOH region determined by dChip was called UPD if the copy number analysis using CNAG revealed no aberrations within that region. If a LOH region was partially affected by copy number aberrations, subregions with normal copy number were called UPD if they comprised at least 50 neighbouring SNPs.

Copy number polymorphisms

As an attempt to distinguish copy number aberrations from copy number variations (CNVs) present in healthy individuals, aberrant regions were compared to published data included in the 'Database of Genomic Variants' (http://projects.tcag.ca/variation/). Regions showing overlap ≥50% with known genomic variants were classified as CNVs.

Interphase FISH

The commercially available locus-specific identifier (LSI) IGH/CCND1 dual colour probe (Abbott/Vysis, Downers Grove, IL, USA) was applied to detect translocation t(11;14)(q13;q32). Differentially labeled bacterial artificial chromosome (BAC) clones, fosmid clones or commercial centromere probes (Abbott/Vysis) were applied to verify copy number results from 100K GeneChip analyses and to perform FISH screenings. A summary of the used FISH assays is shown in Table SIV. FISH was performed as published elsewhere (Martin-Subero *et al*, 2002). Hundred nuclei were evaluated per hybridisation whenever possible.

Delineation of homozygous deletions

Polymerase chain reaction (PCR)-based methods were used to confirm homozygous deletions detected in the SNP array data of the MCL cell lines. To identify novel candidate TSGs, regions of homozygous loss were delineated by PCR. Primer sequences, polymerase enzymes and PCR conditions are summarised in Table SV.

Mutation analyses

The coding exons and exon/intron-boundaries of *MAP2* were amplified and analysed by DHPLC as previously described (Arnold *et al*, 1999). Sequences and annealing temperatures of primer pairs are shown in Table SV. PCR products that showed aberrant chromatograms were subjected to direct sequencing using the Big Dye Terminator v1.1 Cycle Sequencing Kit and the Genetic Analyser ABI PRISM 310 system (Applied Biosystems, Foster City, CA, USA). *TP53* mutation analysis was performed as previously described (Gross *et al*, 2001). Sequence variations were checked and named according to the R12 release of the IARC TP53 Database (Petitjean *et al*,

39

2007) to determine whether they are rare polymorphisms or deleterious.

Methylation studies

Sodium bisulfite conversion of genomic DNA was performed with EpiTect Bisulfite Kit (Qiagen). The methylation status of *MAP2* was evaluated by methylation-specific PCR (MSP) (Herman *et al*, 1996) as well as bisulfite sequencing (Frommer *et al*, 1992). For bisulfite sequencing, the whole CpG island upstream of *MAP2* was amplified after DNA bisulfite treatment, ligated into a pCR2.1-TOPO vector and transformed into competent *Escherichia coli* (TOP10) using the TOPO TA Cloning Kit (Invitrogen, Carlsbad, CA, USA). Insertion of the DNA fragments was tested by PCR of the colonies. Cloned DNA was amplified from 10 positive colonies using vector-specific primers followed by direct sequencing using Big Dye Terminator v1.1 Cycle Sequencing Kit and the Genetic Analyser ABI PRISM 310 system (Applied Biosystems). Primer sequences for MSP and bisulfite sequencing, the used polymerase enzymes and PCR conditions used in the analyses are shown in Table SV. To control the result of bisulfite conversion and the specificity of the MSP reactions, we used CpGenome Universal Methylated and CpGenome Universal Unmethylated DNA (Millipore, Billerica, MA, USA).

Expression studies

MAP2 and *MAP6* expression in MCL cell lines was analysed by SYBR Green-based real-time RT-PCR using the iCycler iQ multi-color-real-time PCR detection system (Biorad, Hercules, CA, USA). Total RNA was isolated from cultured cells with NucleoSpin RNA/Protein Kit (Macherey-Nagel, Düren, Germany), high quality was confirmed with the Experion automated electrophoresis system (Biorad) and cDNA synthesis was performed using the QuantiTect Reverse Transcription Kit (Qiagen). For the real-time PCR of *MAP2* we used QuantiTect SYBR Green PCR Kit and the QuantiTect Primer Assays of *MAP2* (Hs_ MAP2_1_SG) and *MAP2* alternative transcripts (Hs_MAP2_va.1_SG) (Qiagen). These two primer assays are able to detect all known transcript variants of *MAP2*. For *MAP6* real time PCR we used the QuantiTect Primer Assay Hs_ MAP6_1_SG. Normalisation for the quantity of cDNA was done by performing simultaneous real-time RT-PCR for hypoxanthine phosphoribosyltranferase 1 (*HPRT1*), β-glucuronidase (*GUSB*) and glucose-6-phosphat-dehydrogenase (*G6PD*) with appropriate QuantiTect Primer Assays (Hs_HPRT1_1_SG, Hs_GUSB_1_SG, and Hs_G6PD_1_SG) (Qiagen). The thermal profile for the SYBR Green-based PCRs consisted of 15 min Taq polymerase activation at 95°C followed by 45 cycles of PCR at 94°C for 15 s (denaturation), 55°C for 30 s (annealing), and 72°C for 30 s (extension). Following amplification, a melting curve analysis was performed to verify the correct product by its specific melting temperature (Tm). Melting curve analysis consisted of a denaturation step at 95°C

for 1 min, lowered to 55°C for 30 s, and followed by 80 cycles of incubation in which the temperature is increased to 95°C at a rate of 0·5°C/10s/cycle with continuous reading of fluorescence. Results were analysed by iCycler iQ Optical Software Version 3.0a and ratios were calculated by $\Delta\Delta CT$ method.

Results

GeneChip array data

In this study, samples from 26 MCL patients (20 male and six female) were subjected to 100K GeneChip Mapping analyses. The characteristic translocation t(11;14)(q13;q32) was confirmed by cytogenetics and/or FISH. Based on FISH, the tumour cell content varied from 21% to 95% (mean 75%) (Table SI). In addition, a panel of six t(11;14)-positive MCL-derived cell lines was studied. Results of the LOH analysis based on this 100K GeneChip array data have already been published for five of these MCL cell lines (Nielaender *et al*, 2006).

Gains and losses. Copy number analyses identified a pattern of genomic imbalances typical for MCL, including low-level gain and monoallelic loss of regions that are known to be recurrently affected in MCL. A genomic overview of gains and losses detected in the investigated MCL primary cases is given in Fig 1. Two hundred thirty-six and 204 genetic events leading to monoallelic loss were identified in the primary MCL and MCL cell lines, respectively (mean of 9 per MCL and 34 per MCL cell line). In addition to deletions, 206 low-level copy gains were detected in the primary MCLs (mean of 8 per MCL) and 181 in the MCL cell lines (mean of 30 per MCL cell line). By overlapping genomic segments showing copy number changes, we delineated the minimal altered regions (MARs) as well as minimal peak regions within these MARs, which were determined by imbalances of single cases (Fig SI). Chromosomal regions that were affected in at least four primary MCL or MCL cell lines are summarised in Table SVI. Frequently affected regions of chromosomal gain were identified in 1q23.3-q25.1, 2p25.1, 2q32.3, 3q27.3-q28, 4q35.1-q35.2, 7p22.3-p15.3, 7p12.1, 7q21.11, 8p23.3, 8q22.1-q24.21, 10p15.3, 10p12.1-p11.22, 11q13.3-q21, 12q13.3-q14.1, 13q31.3, 14q11.2, 15q21.3, 17q23.2, 18q21.2-q22.3 and 19p12. Genomic loss was frequently detected in 1p33-p32.3, 1p31.1-p21.1, 2q13, 2q24.1-q31.2, 2q37.1, 3p14.2-p12.2, 6q23.3-q27, 8p23.2-p21.2, 9p24.3-p21.2, 9q13-q31.2, 10p14-p13, 11q22.3-q23.2, 12p13.2-p13.1, 13q12.3-q13.1, 13q14.2-q14.3, 13q22.2-q31.1, 13q34, 14q32.12, 15q11.2, 17p13.3-p12, 22q11.22, 22q13.2-q13.33, and Xp22.33. Some of these regions contain known CNVs, also present in healthy individuals, and are marked in Table SVI. In addition to common alterations previously reported in MCL, 100K GeneChip mapping identified five regions frequently targeted by chromosomal gain as well as seven regions of recurrent genomic loss, which have not been reported yet. Some of these novel regions contain candidate genes showing typical tumour suppressor or oncogene properties, such as *FHIT* in 3p13,
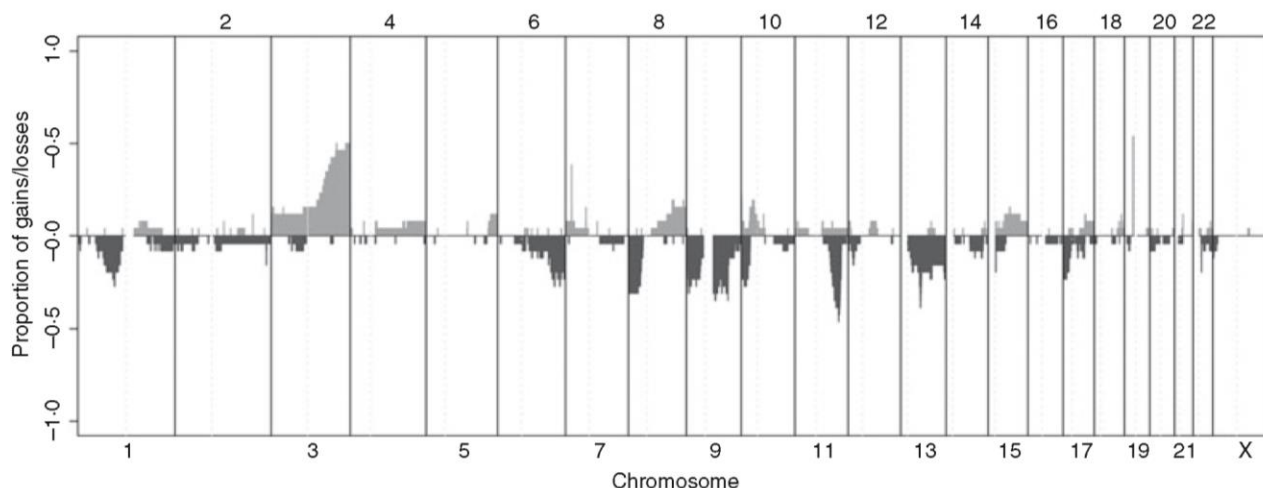
40

Fig 1. Genome-wide detection of copy number changes in MCL primary cases. Proportion of gains and losses analysing GeneChip data of 26 MCL primary cases are displayed from 1pter to Xqter. Light grey columns indicate chromosomal gain, whereas dark grey columns indicate loss of genetic material.

CDKN1B in 12p13.1, *MYCBP* in 13q22.3, or *CDKN2AIP* and *ING2* in 4q35.1 (Table SVI).

High-level amplifications. Regions showing a copy number ≥5 were classified as high-level amplified. Amplicons included genes previously described as being amplified or overexpressed in MCL, such as *BCL2* (18q21.33), *GPC5/MIRN17* (13q31.3) and *BMI1* (10p12.2) (Bea *et al*, 2001; de Leeuw *et al*, 2004; Rubio-Moscardo *et al*, 2005b). Novel detected regions of genomic amplification in MCL involved chromosomes 4p14, 6p21.2, 8p23.1, 8p23.1-p22, 8q24.21, 11q13.4-q13.5, 11q14.1, 11q22.3, 11q23.1-q23.2, 17p11.2, 18q12.2 and 18q12.2-q12.3 (Table SVII). Notably, most of the detected high-level amplifications (∼90%) were observed in the MCL cell lines. There were only two primary MCL cases showing this kind of aberration. Each of these newly detected amplified regions was observed in only one sample with the exception of 11q13.4-q13.5, which was detected in three MCL cell lines. This region harbours nine genes including *MAP6* (microtubule-associated protein 6), *SERPINH1* (serine proteinase inhibitor 1), *GDPD5* (glycerophosphodiester phosphodiesterase domain 5), *RPS3* (ribosomal Protein S3), *UVRAG* (UV radiation resistance associated gene), *WNT11* (wingless-type *MMTV* integration site family), *MOGAT2* and *DGAT2* (mono- and diacylglycerol O-acyltransferases), as well as the hypothetical protein *LOC283212*. In addition, this chromosomal region is coding for 2 snoRNAs (*SNORD15A and SNORD15B*).

Homozygous deletions. Two different approaches were used for the identification of candidate homozygously deleted regions in the MCL GeneChip data. In the liberal screening of the MCL cell lines, 11 homozygous deletions were identified in eight different chromosomal regions and, except the CNV, all of them were confirmed by PCR and/or FISH (Table I). Applying the sensitive screening method, 19 additional candidate homozygous deletions, all affecting different loci, were identified in MCL cell lines; four were validated by PCR and/ or FISH, two were consistent with known CNVs (Table I). The additional 13 regions suggestive for homozygous loss could not be confirmed. Notably, this evaluation criteria detected every previously reported homozygous deletion in MCL cell lines, including TSG loci in 1p32.3 (*CDKN2C/P18*), 2q13 (*BCL2L11*), 9p21.3 (*CDKN2A/P16*), and 13q14.2 (*RB1*) (Nielander *et al*, 2007). Nevertheless, 43% of the regions suggestive for homozygous deletion represented false positive findings. However, four PCR-proven homozygously deleted regions and two CNVs detected by the sensitive screening were not detected by the liberal approach, most likely due to smoothing of the data. Table I displays the verified regions of homozygous loss in the MCL cell lines detected by both approaches. Moreover, the borders of five novel deleted regions were delineated by PCR to identify potential TSGs (Table I).

In the 26 primary MCL cases, 15 and six regions met the criteria for homozygous deletion in the sensitive and liberal approach, respectively. Confirmation by PCR-based methods was not performed due to contamination of the tumour samples by normal surrounding tissue. FISH was only applied for verification of deletions exceeding ∼100 kb (i.e. the insert size of the used BAC, PAC or fosmid clones) that were not classified as CNVs. Homozygous deletion could be confirmed in 9p21.3 and in Xp22.33. In Table II, regions of homozygous loss, which were detected by GeneChip data analysis of the MCL primary cases, are summarised with regard to size, CNVs, target genes and FISH verification.

Loss of heterozygosity in regions with normal copy number. The 100K GeneChip data of the primary MCL cases was also subjected to LOH analyses in order to identify regions of

41

Table I. Homozygous deletions in MCL cell lines.

Locus*	Liberal screening [start-stop in bp*]	Sensitive screening [start-stop in bp*]	PCR-based confirmation and delineation [start-stop in bp*]	FISH confirmation	Affected cell lines	Target genes
1p32.3	Not detected	50 837 181-57 453 350	(Mestre-Escorihuela *et al*, 2007)	Not done	UPN-1	*CDKN2C*
2q13	111 616 112-112 182 931	111 616 112-112 155 057	(Tagawa *et al*, 2005)	Not done	JEKO-1	*BCL2L11*
2q34	210 048 358-210 472 004	210 048 358-210 472 004	210 079 601-210 462 202	Confirmed	UPN-1	*MAP2*
2q37.3	Not detected	237 396 643-237 486 004	237 272 264-237 637 712	Confirmed	UPN-1	No genes
9p21.3	21 948 524-22 102 599	21 948 524-22 090 176	(Kohlhammer *et al*, 2004; de Leeuw *et al*, 2004; Tagawa *et al*, 2005)	Not done	GRANTA-519, REC-1, MAVER-1	*CDKN2A*, *CDKN2B*
9p21.2	27 316 780-27 716 911	27 248 185-27 716 731	27 287 093-27 940 630	Not done	MAVER-1	*MOBKL2B*, *IFNK*, *C9orf72*
9p21.1	28 761 537-30 316 115	28 761 537-30 316 115	Confirmed but not delineated by PCR	Not done	MAVER-1	No genes
12p13.1	Not detected	13 617 759-13 618 078	13 616 272-13 652 667	Not done	REC-1	*GRIN2B*
13q14.2	Not detected	47 812 793-47 817 924	(Pinyol *et al*, 2007)	Not done	UPN-1	*RB1*
13q33.1	101 552 480-101 943 751	101 552 480-101 943 751	Confirmed by FISH not delineated by PCR	Confirmed	JEKO-1	*FGF14*
18q22.1	62 853 023-63 491 358	63 097 606-63 491 358	62 390 461-64 346 179	Confirmed	UPN-1	*DSEL*
22q11.22	20 994 635-21 479 136	20 994 635-21 479 136	CNV†	Not done	GRANTA-519, HBL-2	*IGL@*
Xp22.31	Not detected	6 492 343-6 543 186	CNV†	Not done	JEKO-1	No genes
Xq28	Not detected	153 981 072	CNV†	Not done	GRANTA-519	No genes

CNV, copy number variation.
*NCBI Build 35.
†According to database of genomic variants.

pUPD. A median number of 1.6 pUPDs (ranging from 0 to 5) was identified in the samples. As previously reported for MCL cell lines (Nielaender *et al*, 2006), regions frequently affected by pUPD in primary MCL are known to be commonly targeted by deletions in MCL, e.g. 11q and 13q (Fig 2A). Some of the detected regions showing pUPD are known to harbour common TSGs, like *TP53* in 17p13.1.

Mutation analyses of *TP53* in one MCL sample displaying pUPD in the short arm of chromosome 17 (Fig 2B) revealed a homozygous missense mutation of a single nucleotide (g.14490 T>A) in exon 8, affecting a DNA binding domain (Fig 2C). This mutation has been reported to be deleterious for *TP53*-DNA interaction (http://www-p53.iarc.fr). FISH analysis using a *TP53* locus-specific probe confirmed a normal gene dosage of two copies (Fig 2D).

Involvement of genes encoding microtubule-associated proteins in MCL

Remarkably, 100K GeneChip data and further investigation of candidate genes identified genes encoding different microtubule-associated proteins, such as *MAP2*, *MAP6* and *TP53*, to be recurrently affected by chromosomal aberrations in MCL. Although it is widely known that MCL karyotypes show high complexity, alterations in microtubule organisation have not been reported so far.

MAP6 gene. High-level amplification of *MAP6* was confirmed in all three MCL cell lines MAVER-1, JEKO-1 and HBL-2 by FISH using a BAC clone spanning the *MAP6* locus. A combined analysis of FISH and R-banding revealed the existence of two derivative chromosomes 6 harbouring the amplified 11q13.5 region in MAVER-1 (Fig 3). FISH screening of interphase nuclei from 27 primary MCL identified one case with five gene copies, although the control gene locus in 11q22.3 was also involved. To investigate whether gene dosage affected *MAP6* expression, we performed real-time RT-PCR in the six MCL cell lines. Higher expression of *MAP6* was detected in JEKO-1, GRANTA-519 and REC-1 compared to a freshly prepared tonsil, which was used as normal control (Fig S2).

MAP2 gene. Homozygous deletion of *MAP2* detected in the MCL cell line UPN-1 was confirmed by PCR and FISH. In addition, the REC-1 cell line showed a heterozygous deletion of part of the long arm of chromosome 2, including *MAP2*. FISH screening of the additional MCL cell lines and 27 primary cases failed to detect further cases with chromosomal loss affecting *MAP2*.

Mutation analysis of *MAP2* was performed in 40 primary MCL and five MCL cell lines by DHPLC of the coding exons and the exon/intron boundaries followed by sequencing aberrant fragments. DHPLC screening of *MAP2* identified a G>A transition in one MCL case (MCL 16). This mutation

42

Table II. Putative and confirmed homozygous deletions in MCL primary cases.

Locus*	Sensitive screening [start-stop in bp*]	Liberal screening [start-stop in bp*]	Number of affected MCL	FISH confirmation	Candidate genes
1p21.2	100 823 226-100 823 636	Not detected	1	Not done (<100 kb)	No genes
3p26.3	1 510 440-1 511 278	Not detected	1	Not done (CNV†)	No genes
4q22.1	90 356 219-90 356 586	Not detected	1	Not done (CNV†)	No genes
5q22.2	111 710 742-111 711 216	Not detected	1	Not done (CNV†)	EPB41L4A
9p24.1	7 501 166-7 503 589	Not detected	1	Not done (CNV†)	No genes
9p23	12 742 340-12 750 718	8 916 956-12 974 756	1	Not done (CNV†)	PTPRD, TYRP1
9p21.3	Not detected	21 762 317-22 801 336	2	Homozygous deletion verified	CDKN2A, CDKN2B
9p21.2	27 692 974-27 693 855	Not detected	1	Not done (<100 kb)	No genes
9p21.1	Not detected	30 425 441-31 302 460	1	Not done (CNV†)	No genes
11q22.2	100 654 097-100 773 657	100 546 098-102 615 349	1	No assay applicable due to repetitive sequences	BIRC2, BIRC3
11q22.3	109 749 455-109 749 570	Not detected	1	Not done (CNV†)	No genes
12p13.1	13 617 759-13 618 078	Not detected	1	Not done (<100 kb)	GRIN2B
22q11.22	21 099 653-21 337 561	21 099 653-21 479 136	1	Not done (CNV†)	IGL@
Xp22.33	677 050-2 528 646	677 050-2 561 008	2	Homozygous deletion verified	PAR1 genes
Xp21.2	29 303 376-29 351 134	Not detected	1	Not done (CNV†)	IL1RAPL1
Xp21.1	33 396 094-33 415 950	Not detected	1	Not done (CNV†)	No genes
Xq21.33	98 159 771-98 164 282	Not detected	2	Not done (<100 kb)	No genes

CNV, copy number variation; PAR1, pseudoautosomal region 1.

*NCBI Build 35.

†According to database of genomic variants.

affected the first position of a coding triplet in exon 5 causing an amino acid change from glutamic acid (E) to lysine (K) (Fig 4). There was no evidence for a SNP at this DNA sequence position in NCBI Build database and DHPLC analyses failed to detect this alteration in 50 DNA samples (100 alleles) derived from peripheral blood of healthy individuals (data not shown).

To investigate the DNA methylation status of the CpG island within the MAP2 promotor (Fig 5A), bisulfite-treated DNA probes of five MCL cell lines and 20 MCL cases were subjected to MSP. In 18 of the 20 MCL cases the pattern of PCR products indicated partial DNA hypermethylation, including the MCL 16 with the MAP2 mutation. The DNAs of the MCL cell lines JEKO-1 and HBL-2 were also partially hypermethylated and the MSP pattern of the REC-1 cell line indicated virtually complete DNA methylation of the MAP2 CpG island (Fig 5B). Notably, REC-1 had only one MAP2 allele due to a heterozygous deletion of part of chromosome 2. The MCL cell line GRANTA-519 showed only a weak PCR product of methylated DNA and the additional two MCL cases and the MCL cell line MAVER-1 seemed to be completely unmethylated. In contrast to the distribution in MCL, eight non-tumours controls (four tonsils and four PB) showed a MSP pattern indicating an unmethylated status of MAP2 promotor. These findings could be confirmed by bisulfite sequencing (Fig 5C and D, Fig S3).

To investigate how DNA methylation status of the MAP2 CpG island correlates with MAP2 gene expression in MCL cell lines we performed real-time RT-PCR. Two different primer assays were used to detect all known MAP2 transcript variants (Fig 5E). In UPN-1 and REC-1, no expression of MAP2 transcripts could be detected. Real-time RT-PCR results of HBL-2 indicated reduced expression of all MAP2 transcripts compared to control LN tissue. In GRANTA-519 and MAVER-1 the expression of MAP2 transcripts was similar or minimally reduced whereas JEKO-1 showed a considerable higher expression of all MAP2 transcripts compared to the LN tissue (Fig 5E). Similar results were obtained using a freshly prepared tonsil as reference (data not shown).

TP53 gene. Due to its influence on microtubule dynamics, TP53 was subjected to mutation analyses in all samples of the MCL GeneChip panel. One case showing a homozygous TP53 mutation had already been analysed as part of the pUPD investigation. In six of the additional 25 primary cases mutations within the coding sequence of TP53 were detected using DHPLC followed by sequencing aberrant fragments. Detected mutations include six different point mutations and one microdeletion of one single base pair (Table III). In all these cases copy number and LOH analyses of the 100K GeneChip data indicated genomic loss of the second allele (Table SVI). Three point mutations and one microdeletion affecting the coding sequence of TP53 were detected in MCL cell lines UPN-1, HBL-2, MAVER-1 and JEKO-1 (Table III). In all these cell lines the second TP53 allele was deleted. Both REC-1 and GRANTA-519 exhibited the wild type allele but the latter showed a heterozygous deletion of the TP53 locus. Some of the MCL cell line
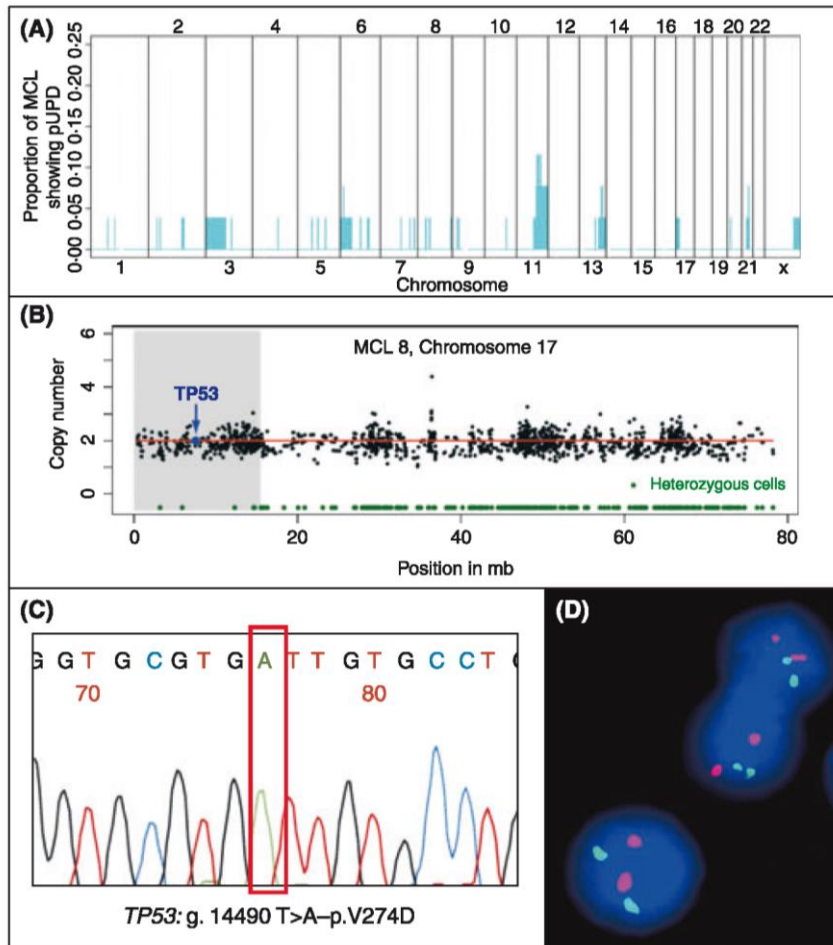
43

Fig 2. Partial uniparental disomy (pUPD) in MCL. (A) Genome-wide distribution of pUPD in 26 primary MCL cases detected by 100K GeneChip Mapping (Affymetrix). Proportion of MCL cases showing pUPD is displayed in turquoise. Chromosomes are shown from 1pter (left) to Xqter (right). (B) Detection of pUPD in the short arm of chromosome 17 in MCL 8. The profile of genomic imbalances is given as black dots with a value of 2 indicating a balanced status. Chromosome 17 is shown from pter (left) to qter (right). Heterozygous calls are given as green dots and the estimated LOH region is overlaid in grey. A blue arrow indicates chromosomal location of the *TP53* gene. (C) FISH analysis using a locus-specific probe consisting of BAC clones labelled in spectrum orange (*TP53*) and spectrum green (control locus in 17q21.2). Signal constellation indicates a normal gene dosage of two copies in the MCL 8. (D) Homozygous point mutation affecting exon 8 of *TP53* in the same MCL showing pUPD in 17p13.1.

data has been already reported elsewhere (Amin *et al*, 2003; M'Kacher *et al*, 2003; Camps *et al*, 2006; Zamo *et al*, 2006).

Discussion

100K GeneChip microarray data

In the present study, we performed 100K GeneChip microarray analyses of 26 primary MCL and six MCL cell lines. The use of short synthetic oligonucleotides as arrayed elements enabled

the detection of genomic imbalances with high resolution (approximately 24 kb) and genotyping was performed simultaneously to identify regions of pUPD. Overall, the identified genomic imbalance pattern was consistent with those previously described by cytogenetics or arrayCGH studies in MCL (Kohlhammer *et al*, 2004; de Leeuw *et al*, 2004; Rubio-Moscardo *et al*, 2005b; Tagawa *et al*, 2005; Mestre-Escorihuela *et al*, 2007; Pinyol *et al*, 2007).

Using the GeneChip mapping technique, we not only delineated previously reported alterations but also identified

Fig 3. High level amplification in 11q13.5. (A) Genomic profiles of part of chromosome 11 in three MCL cell lines. Copy number is given as black dots with a value of 2 indicating a balanced status. Chromosomal region in the long arm is shown from centromeric to telomeric. The minimal altered region is indicated by blue vertical bars. (B–D) Interphase FISH of the MCL cell lines using a locus-specific probe for *MAP6* (spectrum green) and a control locus in 11q22.3 (spectrum orange). (E–F) R banding followed by FISH on metaphases of MAVER-1 using the described *MAP6* probe. High level amplification was shown to affect two derivative chromosomes 6 [a:der(6)(6pter?6q22::11q13?::11q22?11q23::11q22->11q23::?), b:der(6)(?::6p21?6q22::11q13?11q14::11q23? amp)].

44

(A) Chromosome 11

MAVER-1

JEKO-1

HBL-2

(B) MAVER-1

(C) JEKO-1

(D) HBL-2

(E) chr. 11 — der(6) a — der(6) b

(F) chr. 11 — der(6) a — der(6) b

45
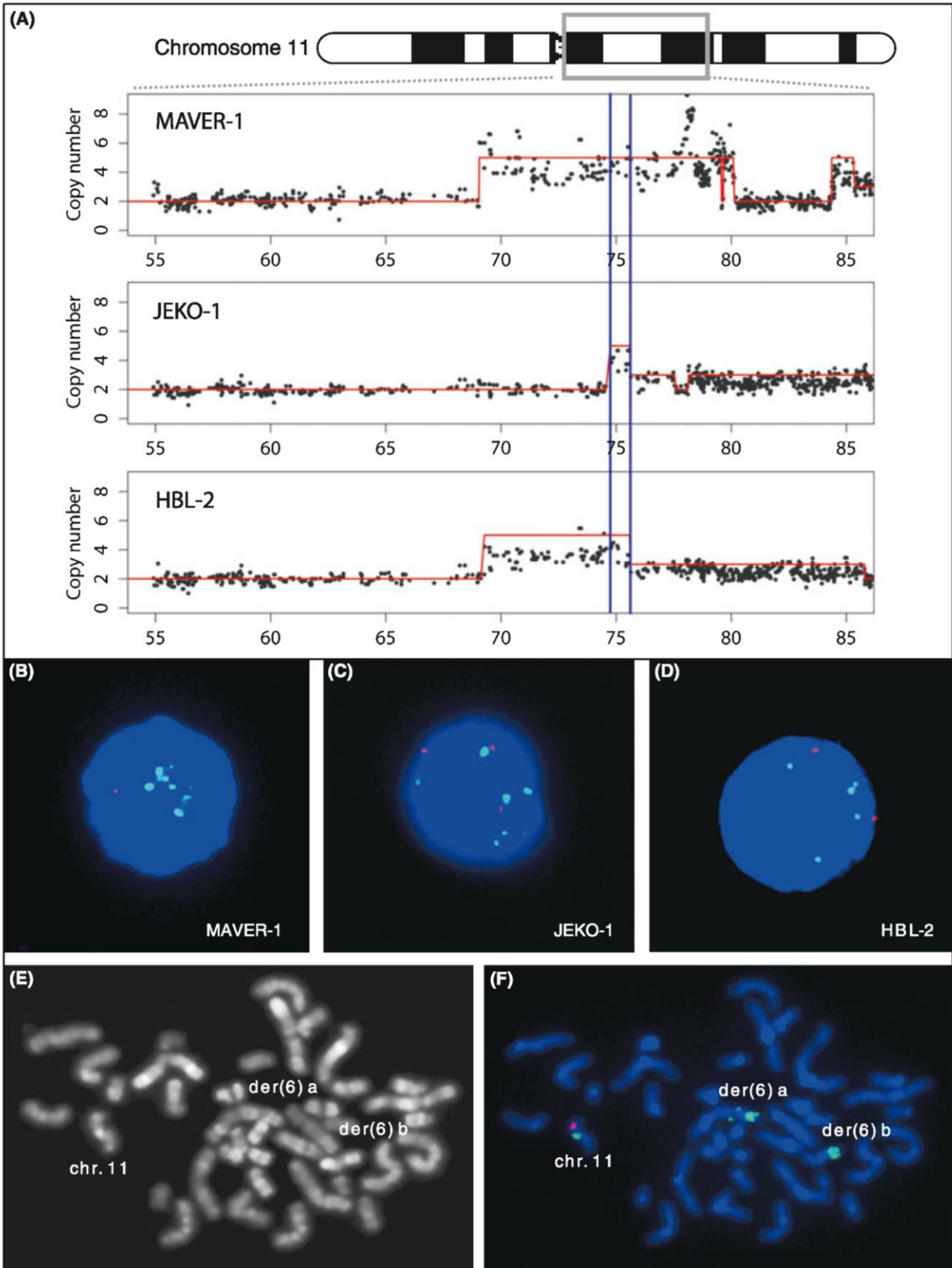
Fig 4. Homozygous deletion in 2q34. (A) Genomic profile of part of the chromosome 2 in the MCL cell line UPN-1. Copy number is given as black dots with a value of 2 indicating a balanced status. Chromosomal region in the long arm is shown from centromeric to telomeric. A red arrow indicates the homozygously deleted region. A Genome Browser extract (NCBI Build 35) displays gene content and the chromosomal position of primers (red bars) for delineation of the minimal affected region. (B) Multiplex-PCR of the target gene and a control locus confirmed the homozygous deletion of *MAP2* but not of the adjacent coding sequences of RNAz s59616 and C2orf21. (C) DHPLC chromatograms of the primary MCL 16 (red) and a healthy control (blue). (D) Sequencing of the aberrant fragment reveal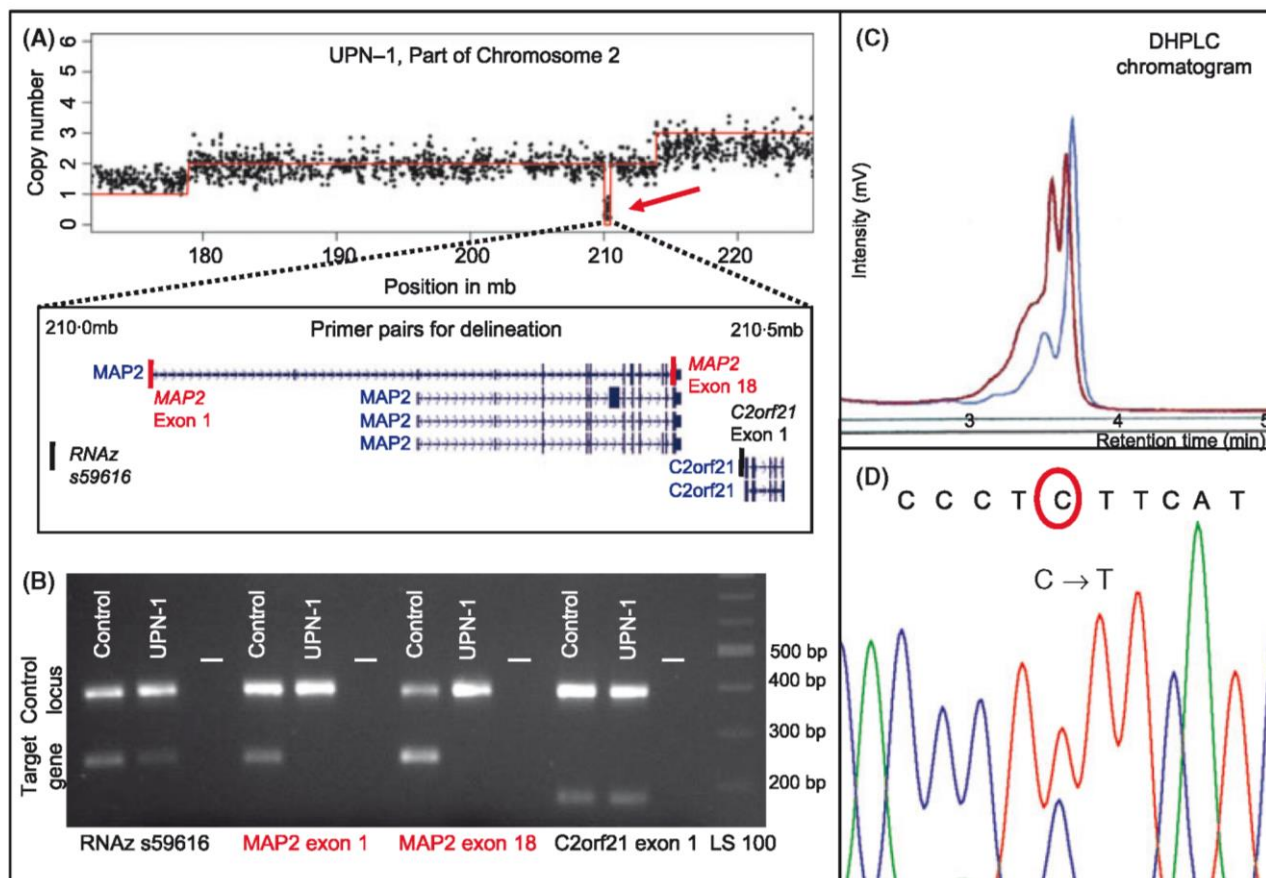ed a heterozygous mutation in exon 5 of *MAP2* changing amino acid composition of the protein (E163K). Here, the reverse strand is displayed showing a C>T transition.

novel regions of recurrent genomic imbalance in MCL. Taking advantage of the high resolution of the used SNP arrays, we focused on detecting high-level amplifications and small regions of homozygous loss. In this way, novel regions harbouring potential oncogenes or candidate TSG were identified that might be involved in tumourigenesis. Moreover, we introduced two useful bioinformatic approaches for identifying homozygous losses in 100K GeneChip data and point out advantages and disadvantages.

In line with a recent study of five MCL cell lines (Nielaender *et al*, 2006), LOH analysis of primary MCL tumor tissue demonstrated that pUPD is a recurrent genetic mechanism in MCL tumourigenesis. Genomic distribution of detected pUPD in the analysed primary MCL showed that recurrently affected regions, such as 11q and 13q, are commonly targeted by deletions in MCL. Furthermore, we explicitly demonstrated TSG inactivation by pUPD targeting the *TP53* locus in 17p13.1. A homozygous missense mutation

Fig 5. Epigenetic and expression studies of MAP2. (A) Genome Browser extract (NCBI Build 35) displaying location of *MAP2* CPG island. (B) Methylation specific PCR (MSP) of REC-1, two primary MCL and controls. Methylation specific amplification was verified using universal methylated (MC) and universal unmethylated (UC) DNA as controls. Genomic DNA (gC) was tested to exclude unspecific amplification. (C) Part of the analysed *MAP2* CpG island sequences investigating bisulfite-treated DNA of REC-1 and a tonsil. The included CpG dinucleotides 5–11 are marked by black bars. (D) Bisulfite genomic sequencing of the *MAP2* CpG island in the MCL cell line REC-1 and two primary MCL. Tonsil and peripheral blood (PB) samples from healthy donors were used as controls. Each column represents a CpG dinucleotide and each row represents one of 10 sequenced clones. Black and white spots indicate methylated and unmethylated CpGs, respectively. Grey spots indicate CpG dinucleotides that failed to be analysed. (E) Real-time Reverse Transcription (RT-) PCR. Two primer assays were applied to investigate all known *MAP2* transcripts. Primer position is indicated by black arrows. Expression of *MAP2* transcripts was examined in six MCL cell lines and compared to *MAP2* expression in lymph node tissue of healthy individuals. Each value has been normalised to the average expression of three housekeeping genes. In UPN-1 and REC-1 no *MAP2* expression was detected, thus $\log_2$ ratio was not calculable (N/C).
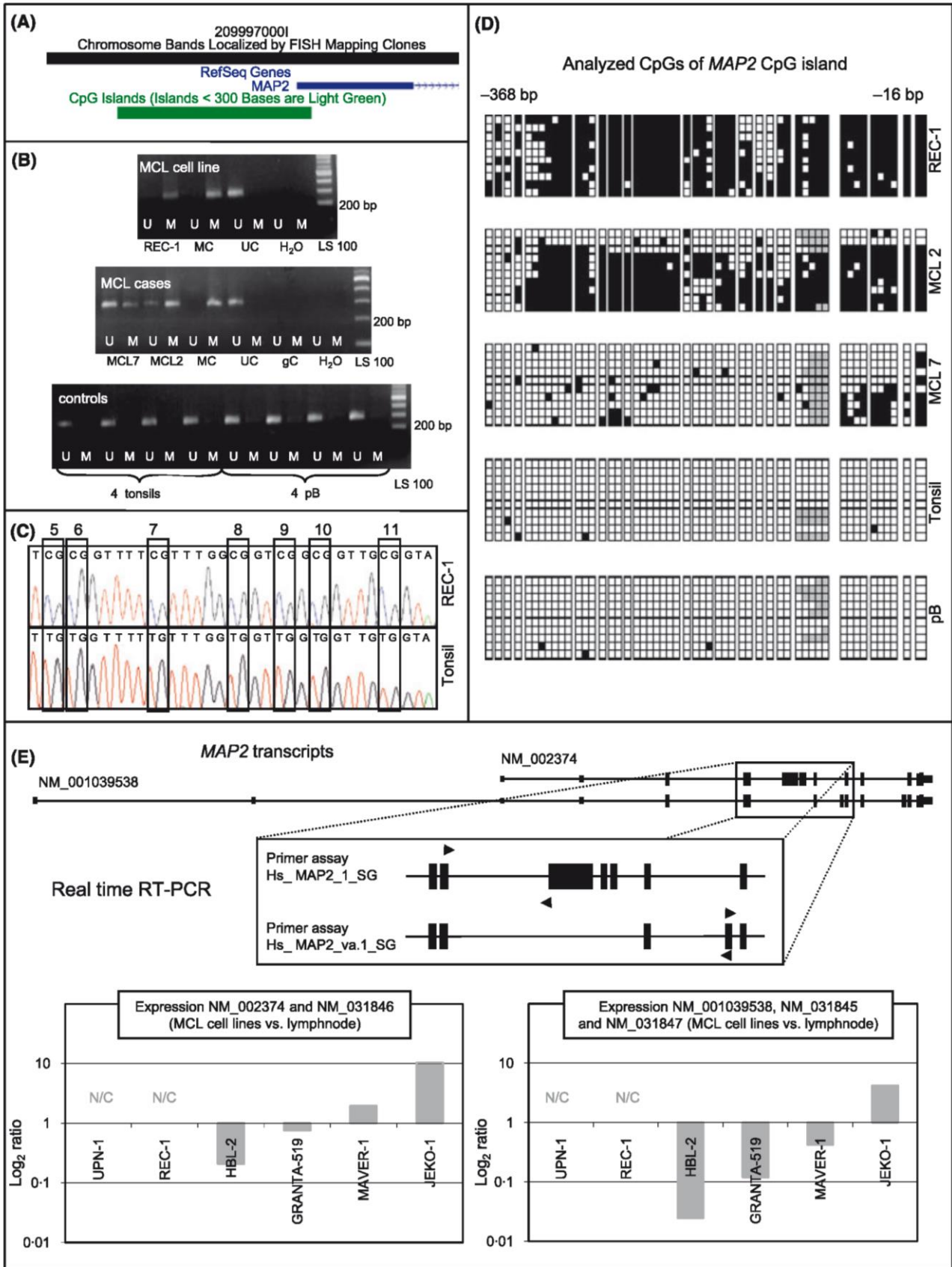
47

Table III. *TP53* mutation analyses in MCL.

MCL primary cases	*TP53* mutation	Second abberation in 17p13.1
MCL 4	g.12178 delC - p.122X	Deletion
MCL 6	g.14487 G>T - p.R273L	Deletion
MCL 8	g.14490 T>A - p.V274D	pUPD
MCL 15	g.13203 G>A - p.R175H	Deletion
MCL 17	g.14070 C>T - p.R248L	Deletion
MCL 19	g.12108 G>T - p.E62X	Deletion
MCL 21	g.13091 G>C - p.A138P	Deletion

MCL cell lines	*TP53* mutation	Second abberation in 17p13.1
UPN-1	g.14525 G>A - p.E286K	Deletion
MAVER-1	g.14511 A>G - p.D281G	Deletion
JEKO-1	g.12096 delC - p.58X	Deletion
HBL-2	g.14511 A>G - p.D281G	Deletion

g., genomic sequence position; p., protein sequence position (http://www-p53.iarc.fr).

affecting a DNA binding domain was detected. This mutation is frequent in lymphomas (28%) and has been reported to be deleterious for *TP53*-DNA interaction (http://www-p53.iarc.fr). *TP53* inactivation by chromosomal deletion is a common chromosomal event in MCL and is associated with poor prognosis (Rubio-Moscardo *et al*, 2005b). The present study showed that pUPD is an alternative mechanism to chromosomal deletion leading to homozygosity of a TSG inactivating mutation. Thus, pUPD seems to be a critical genetic event in MCL pathogenesis.

Genes encoding microtubule-associated proteins as targets of chromosomal aberrations

Interestingly, analysis of the 100K GeneChip data identified different genes encoding microtubule-associated proteins (MAPs) to be involved in chromosomal alterations in MCL. MAPs are cellular proteins that are associated with microtubules and alter their dynamics. Microtubule dynamic property is crucial for the assembly of the mitotic spindle and the attachment and movement of chromosomes along the spindle (Zhai *et al*, 1996). Microtubule-targeting drugs suppressing microtubule dynamics are widely used as cancer chemotherapeutic agents (Jordan & Wilson, 2004). In addition to their direct involvement in the physical process of mitosis, microtubules also serve as scaffolds for signalling molecules (Mollinedo & Gajate, 2003). The family of MAPs includes products of oncogenes, tumour suppressors and apoptosis regulators, suggesting that alteration of microtubule dynamics and changes in the scaffolding properties of microtubules may be critical events in tumourigenesis and tumour progression (Bhat & Setaluri, 2007). Until now, alterations in microtubule organisation have not been reported in MCL (Jares *et al*, 2007).

In this study, a homozygous deletion of the *MAP2* locus was identified in the MCL cell line UPN-1. Real time RT-PCR revealed absence of *MAP2* expression in UPN-1 and REC-1. REC-1 harbours a heterozygous deletion of part of the long arm of chromosome 2 and epigenetic studies showed complete DNA methylation of the CpG island of the remaining *MAP2* allele. Moreover, the DNAs of the MCL cell lines JEKO-1 and HBL-2 were also partially hypermethylated. MSP analysis demonstrated partial hypermethylation in 90% of 20 investigated primary MCL. In one of these cases showing partial hypermethylation, a point mutation affecting the coding sequence of *MAP2* was identified. These findings suggest that DNA hypermethylation is a frequent mechanism leading to *MAP2* gene inactivation in MCL. MAP2 protein participates in the stabilisation of microtubules and is predominantly expressed in neurons, where it is essential for the regulation of organelle transport within axon and dendrites (Sanchez *et al*, 2000). Moreover, MAP2 was the first protein shown to copurify and interact directly with the regulatory subunit of the protein kinase A (PKA), also known as cAMP-dependent protein kinase (cAPK) (Vallee *et al*, 1981; Theurkauf & Vallee, 1982). Among other cellular effects, PKA-catalysed phosphorylation modulates cell growth, cell division and actin cytoskeleton rearrangements. MAP2 protein operates as an A-kinase anchoring protein (AKAP) and targets the PKA to microtubules. Attachment to microtubules occurs through its tubulin-binding domain (Serrano *et al*, 1984; Hirokawa, 1994). MAP2 also harbours a conserved binding site for phosphatase PP2A, although direct binding of the phosphatase has yet to be reported. PP2A represents a family of heterotrimeric serine/threonine phosphatases implicated in the regulation of a plethora of cellular processes such as apoptosis, transcription, translation, DNA replication, signal transduction, protection against tumourigenesis and cell division (Janssens *et al*, 2005). Soltani *et al* (2005) reported that *MAP2* expression is associated with prognosis in melanoma. A five-year clinical follow-up study showed longer disease-free survival of patients whose primary tumors express abundant *MAP2* as compared with patients with weak or no *MAP2* expression. Moreover, exogenous expression by adenovirus leads to cell cycle arrest, growth inhibition and apoptosis in metastatic melanoma cells (Fang *et al*, 2001; Soltani *et al*, 2005). Thus, lack of *MAP2* expression might be also associated with MCL pathogenesis.

The p53 protein is also associated with microtubules *in vitro* and *in vivo* (Giannakakou *et al*, 2000) and has been reported to regulate other microtubule-associated proteins (Murphy *et al*, 1999; Johnsen *et al*, 2000; Mirza *et al*, 2002). *TP53* inactivation by chromosomal deletions or by mutations is a common genetic alteration in MCL. Galmarini *et al* (2003) demonstrated that microtubule protein composition was altered in *TP53* mutants (mut-p53) and dynamic instability of microtubules was significantly increased. Mutation analyses in this report identified seven primary MCL cases to harbour homozygous mutations in the coding sequence of *TP53*. In all these cases, the second allele

48

got lost by chromosomal deletion or pUPD. Similarly, *TP53* was shown to be homozygously mutated in four of six investigated MCL cell lines (Amin *et al*, 2003; M'Kacher *et al*, 2003; Camps *et al*, 2006; Zamo *et al*, 2006). Galmarini *et al* (2003) also reported that the MAP6 (also called STOP) protein and its corresponding mRNA-expression were increased in the mut-p53 cells, than in the wt-p53 cells suggesting negative transcriptional regulation of *MAP6* by p53 protein. Interestingly, our MCL GeneChip data showed high-level amplification of *MAP6* in the three MCL cell lines MAVER-1, JEKO-1 and HBL-2. Real-time RT-PCR detected higher expression of *MAP6* in JEKO-1, GRANTA-519 and REC-1 compared with non-tumours tonsil tissue. The apparent lack of correlation between *MAP6* dosage and expression in MAVER-1 and HBL-2 might be caused by putative non-neuronal transcripts variants that cannot be detected with the primers used. According to this, the murine homologue shows non-neuronal transcript variants, which lacks whole exons (Bosc *et al*, 2003). So far, no human non-neuronal transcript variant has been identified. High-level amplification of *MAP6* was frequently detected in MCL cell lines but in none of the investigated primary MCL. As cell lines are frequently derived from cases with advanced disease, this finding might indicate that *MAP6* amplification is associated with disease progression. In line with this hypothesis, a chromosomal rearrangement of another *MAP* gene, i.e. *MAP4*, was recently identified as secondary alteration in a large B-cell lymphoma (DLBCL), which was present at relapse but not at initial diagnosis (Murga Penas *et al*, 2006). According to our GeneChip data, it is widely assumed that cell lines generally harbour an increased number of chromosomal aberrations compared to the primary tumour cells.

Our study provides evidence that alterations of microtubule dynamics might be critical in MCL tumourigenesis and tumour progression. Nineteen of the 26 primary MCL from the SNP array panel and five of the six MCL cell lines harboured a genetic or epigenetic defect in at least one of the three microtubule-associated genes *MAP2*, *MAP6* and *TP53*. Six primary cases were not analysed by methylation-specific methods. Only one case and one MCL cell line, in which all three gene loci were analysed, did not show any of the investigated alterations.

The complexity of the mitotic spindle requires fine-tuning of the dynamics of all microtubules for proper function. MAPs can either stabilise or destabilise microtubules. Changes in levels of expression have been reported to correlate with aggressiveness of cancer cells or their sensitivity to microtubule-targeting agents. Although plenty of studies exist regarding microtubule-associated proteins in neurons, where they play a critical role in neurite outgrowth and dendrite development, their mechanisms of operation in mitotic spindle regulation are rather unclear. Mitotic spindle organisation is a fine-tuning process and investigation of the involved proteins is difficult due to low dosage. In contrast to abundant gene expression in brain tissue, expression of *MAP2* and *MAP6* is hardly detectable in other kinds of tissues. In our study,

genomic alterations, such as *MAP6* amplification or *MAP2* and *TP53* inactivation, provide evidence for the involvement of microtubule-associated genes in MCL tumourigenesis. Alterations in microtubule dynamics and mitotic spindle organisation might contribute to the karyotype complexity and chromosomal instability that are characteristic features of MCL. Supporting this hypothesis, a recent study underlined the high expression level of centrosome-associated gene products in blastoid MCL (Neben *et al*, 2007). MCL has one of the worst prognoses amongst all lymphomas. There is no therapy that can be considered as standard. Resistance against microtubule-targeting chemotherapeutic agents may be the consequence of changes in microtubule dynamics.

In conclusion, our study demonstrated that 100K GeneChip microarray analyses is a useful strategy to analyse MCL genomes with regard to genomic imbalances, particularly homozygous deletions, as well as pUPD. Moreover, we identified novel candidate TSG and oncogene loci that might harbour genes involved in MCL pathogenesis. Interestingly, different genes encoding microtubule-associated proteins could be identified as targets of chromosomal aberrations in MCL. Our findings suggest that alteration of microtubule dynamics is a critical genetic event in MCL.

Acknowledgements

References

Amin, H.M., McDonnell, T.J., Medeiros, L.J., Rassidakis, G.Z., Leventaki, V., O'Connor, S.L., Keating, M.J. & Lai, R. (2003) Characterization of 4 mantle cell lymphoma cell lines. *Archives of Pathology and Laboratory Medicine*, **127**, 424–431.

Arnold, N., Gross, E., Schwarz-Boeger, U., Pfisterer, J., Jonat, W. & Kiechle, M. (1999) A highly sensitive, fast, and economical technique for mutation analysis in hereditary breast and ovarian cancers. *Human Mutation*, **14**, 333–339.

Bea, S., Tort, F., Pinyol, M., Puig, X., Hernandez, L., Hernandez, S., Fernandez, P.L., van Lohuizen, M., Colomer, D. & Campo, E. (2001) BMI-1 gene amplification and overexpression in hematological malignancies occur mainly in mantle cell lymphomas. *Cancer Research*, **61**, 2409–2412.

Beroukhim, R., Lin, M., Park, Y., Hao, K., Zhao, X., Garraway, L.A., Fox, E.A., Hochberg, E.P., Mellinghoff, I.K., Hofer, M.D., Descazeaud, A., Rubin, M.A., Meyerson, M., Wong, W.H., Sellers, W.R. & Li, C. (2006) Inferring loss-of-heterozygosity from unpaired tumors using high-density oligonucleotide SNP arrays. *PLoS Computational Biology*, **2**, e41.

Bhat, K.M. & Setaluri, V. (2007) Microtubule-associated proteins as targets in cancer chemotherapy. *Clinical Cancer Research*, **13**, 2849–2854.

Bignell, G.R., Huang, J., Greshock, J., Watt, S., Butler, A., West, S., Grigorova, M., Jones, K.W., Wei, W., Stratton, M.R., Futreal, P.A., Weber, B., Shapero, M.H. & Wooster, R. (2004) High-resolution

analysis of DNA copy number using oligonucleotide microarrays. *Genome Research*, **14**, 287–295.

Bosc, C., Andrieux, A. & Job, D. (2003) STOP proteins. *Biochemistry*, **42**, 12125–12132.

Bruce, S., Leinonen, R., Lindgren, C.M., Kivinen, K., Dahlman-Wright, K., Lipsanen-Nyman, M., Hannula-Jouppi, K. & Kere, J. (2005) Global analysis of uniparental disomy using high density genotyping arrays. *Journal of Medical Genetics*, **42**, 847–851.

Camps, J., Salaverria, I., Garcia, M.J., Prat, E., Bea, S., Pole, J.C., Hernandez, L., Del Rey, J., Cigudosa, J.C., Bernues, M., Caldas, C., Colomer, D., Miro, R. & Campo, E. (2006) Genomic imbalances and patterns of karyotypic variability in mantle-cell lymphoma cell lines. *Leukemia Research*, **30**, 923–934.

Fang, D., Hallman, J., Sangha, N., Kute, T.E., Hammarback, J.A., White, W.L. & Setaluri, V. (2001) Expression of microtubule-associated protein 2 in benign and malignant melanocytes: implications for differentiation and progression of cutaneous melanoma. *American Journal of Pathology*, **158**, 2107–2115.

Frommer, M., McDonald, L.E., Millar, D.S., Collis, C.M., Watt, F., Grigg, G.W., Molloy, P.L. & Paul, C.L. (1992) A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proceedings of the National Academy of Sciences of the United States of America*, **89**, 1827–1831.

Galmarini, C.M., Kamath, K., Vanier-Viornery, A., Hervieu, V., Peiller, E., Falette, N., Puisieux, A., Ann Jordan, M. & Dumontet, C. (2003) Drug resistance associated with loss of p53 involves extensive alterations in microtubule composition and dynamics. *British Journal of Cancer*, **88**, 1793–1799.

Giannakakou, P., Sackett, D.L., Ward, Y., Webster, K.R., Blagosklonny, M.V. & Fojo, T. (2000) p53 is associated with cellular microtubules and is transported to the nucleus by dynein. *Nature Cell Biology*, **2**, 709–717.

Gladden, A.B., Woolery, R., Aggarwal, P., Wasik, M.A. & Diehl, J.A. (2006) Expression of constitutively nuclear cyclin D1 in murine lymphocytes induces B-cell lymphoma. *Oncogene*, **25**, 998–1007.

Gross, E., Kiechle, M. & Arnold, N. (2001) Mutation analysis of p53 in ovarian tumors by DHPLC. *Journal of Biochemical and Biophysical Methods*, **47**, 73–81.

Herman, J.G., Graff, J.R., Myohanen, S., Nelkin, B.D. & Baylin, S.B. (1996) Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *Proceedings of the National Academy of Sciences of the United States of America*, **93**, 9821–9826.

Hirokawa, N. (1994) Microtubule organization and dynamics dependent on microtubule-associated proteins. *Current Opinion in Cell Biology*, **6**, 74–81.

Janssens, V., Goris, J. & Van Hoof, C. (2005) PP2A: the expected tumor suppressor. *Current Opinion in Genetics and Development*, **15**, 34–41.

Jares, P., Colomer, D. & Campo, E. (2007) Genetic and molecular pathogenesis of mantle cell lymphoma: perspectives for new targeted therapeutics. *Nature Reviews Cancer*, **7**, 750–762.

Johnsen, J.I., Aurelio, O.N., Kwaja, Z., Jorgensen, G.E., Pellegata, N.S., Plattner, R., Stanbridge, E.J. & Cajot, J.F. (2000) p53-mediated negative regulation of stathmin/Op18 expression is associated with G(2)/M cell-cycle arrest. *International Journal of Cancer*, **88**, 685–691.

Jordan, M.A. & Wilson, L. (2004) Microtubules as a target for anticancer drugs. *Nature Reviews Cancer*, **4**, 253–265.

Kohlhammer, H., Schwaenen, C., Wessendorf, S., Holzmann, K., Kestler, H.A., Kienle, D., Barth, T.F., Moller, P., Ott, G., Kalla, J.,

Radlwimmer, B., Pscherer, A., Stilgenbauer, S., Dohner, H., Lichter, P. & Bentz, M. (2004) Genomic DNA-chip hybridization in t(11;14)-positive mantle cell lymphomas shows a high frequency of aberrations and allows a refined characterization of consensus regions. *Blood*, **104**, 795–801.

de Leeuw, R.J., Davies, J.J., Rosenwald, A., Bebb, G., Gascoyne, R.D., Dyer, M.J., Staudt, L.M., Martinez-Climent, J.A. & Lam, W.L. (2004) Comprehensive whole genome array CGH profiling of mantle cell lymphoma model genomes. *Human Molecular Genetics*, **13**, 1827–1837.

Lin, M., Wei, L.J., Sellers, W.R., Lieberfarb, M., Wong, W.H. & Li, C. (2004) dChipSNP: significance curve and clustering of SNP-array-based loss-of-heterozygosity data. *Bioinformatics*, **20**, 1233–1240.

Lovec, H., Grzeschiczek, A., Kowalski, M.B. & Moroy, T. (1994) Cyclin D1/bcl-1 cooperates with myc genes in the generation of B-cell lymphoma in transgenic mice. *EMBO Journal*, **13**, 3487–3495.

M'Kacher, R., Bennaceur, A., Farace, F., Lauge, A., Plassa, L.F., Wittmer, E., Dossou, J., Violot, D., Deutsch, E., Bourhis, J., Stoppa-Lyonnet, D., Ribrag, V., Carde, P., Parmentier, C., Bernheim, A. & Turhan, A.G. (2003) Multiple molecular mechanisms contribute to radiation sensitivity in mantle cell lymphoma. *Oncogene*, **22**, 7905–7912.

Martin-Subero, J.I., Harder, L., Gesk, S., Schlegelberger, B., Grote, W., Martinez-Climent, J.A., Dyer, M.J., Novo, F.J., Calasanz, M.J. & Siebert, R. (2002) Interphase FISH assays for the detection of translocations with breakpoints in immunoglobulin light chain loci. *International Journal of Cancer*, **98**, 470–474.

Mestre-Escorihuela, C., Rubio-Moscardo, F., Richter, J.A., Siebert, R., Climent, J., Fresquet, V., Beltran, E., Agirre, X., Marugan, I., Marin, M., Rosenwald, A., Sugimoto, K.J., Wheat, L.M., Karran, E.L., Garcia, J.F., Sanchez, L., Prosper, F., Staudt, L.M., Pinkel, D., Dyer, M.J. & Martinez-Climent, J.A. (2007) Homozygous deletions localize novel tumor suppressor genes in B-cell lymphomas. *Blood*, **109**, 271–280.

Mirza, A., McGuirk, M., Hockenberry, T.N., Wu, Q., Ashar, H., Black, S., Wen, S.F., Wang, L., Kirschmeier, P., Bishop, W.R., Nielsen, L.L., Pickett, C.B. & Liu, S. (2002) Human survivin is negatively regulated by wild-type p53 and participates in p53-dependent apoptotic pathway. *Oncogene*, **21**, 2613–2622.

Mollinedo, F. & Gajate, C. (2003) Microtubules, microtubule-interfering agents and apoptosis. *Apoptosis*, **8**, 413–450.

Murga Penas, E.M., Kawadler, H., Siebert, R., Frank, M., Ye, H., Hinz, K., Becher, C., Hummel, M., Barth, T.F., Bokemeyer, C., Stein, H., Trumper, L., Moller, P., Marynen, P., Du, M.Q., Yang, X., Hansmann, M.L. & Dierlamm, J. (2006) A novel fusion of the *MALT1* gene and the microtubule-associated protein 4 (*MAP4*) gene occurs in diffuse large B-cell lymphoma. *Genes, Chromosomes and Cancer*, **45**, 863–873.

Murphy, M., Ahn, J., Walker, K.K., Hoffman, W.H., Evans, R.M., Levine, A.J. & George, D.L. (1999) Transcriptional repression by wild-type p53 utilizes histone deacetylases, mediated by interaction with mSin3a. *Genes and Development*, **13**, 2490–2501.

Nannya, Y., Sanada, M., Nakazaki, K., Hosoya, N., Wang, L., Hangaishi, A., Kurokawa, M., Chiba, S., Bailey, D.K., Kennedy, G.C. & Ogawa, S. (2005) A robust algorithm for copy number detection using high-density oligonucleotide single nucleotide polymorphism genotyping arrays. *Cancer Research*, **65**, 6071–6079.

Neben, K., Ott, G., Schweizer, S., Kalla, J., Tews, B., Katzenberger, T., Hahn, M., Rosenwald, A., Ho, A.D., Muller-Hermelink, H.K.,

Lichter, P. & Kramer, A. (2007) Expression of centrosome-associated gene products is linked to tetraploidization in mantle cell lymphoma. *International Journal of Cancer*, **120**, 1669–1677.

Nielaender, I., Martin-Subero, J.I., Wagner, F., Martinez-Climent, J.A. & Siebert, R. (2006) Partial uniparental disomy: a recurrent genetic mechanism alternative to chromosomal deletion in malignant lymphoma. *Leukemia*, **20**, 904–905.

Nielander, I., Bug, S., Richter, J., Giefing, M., Martin-Subero, J.I. & Siebert, R. (2007) Combining array-based approaches for the identification of candidate tumor suppressor loci in mature lymphoid neoplasms. *Acta pathologica, Microbiologica, et Immunologica Scandinavica*, **115**, 1107–1134.

Petitjean, A., Mathe, E., Kato, S., Ishioka, C., Tavtigian, S.V., Hainaut, P. & Olivier, M. (2007) Impact of mutant p53 functional properties on TP53 mutation patterns and tumor phenotype: lessons from recent developments in the IARC TP53 database. *Human Mutation*, **28**, 622–629.

Pinyol, M., Bea, S., Pla, L., Ribrag, V., Bosq, J., Rosenwald, A., Campo, E. & Jares, P. (2007) Inactivation of RB1 in mantle-cell lymphoma detected by nonsense-mediated mRNA decay pathway inhibition and microarray analysis. *Blood*, **109**, 5422–5429.

Rubio-Moscardo, F., Blesa, D., Mestre, C., Siebert, R., Balasas, T., Benito, A., Rosenwald, A., Climent, J., Martinez, J.I., Schilhabel, M., Karran, E.L., Gesk, S., Esteller, M., deLeeuw, R., Staudt, L.M., Fernandez-Luna, J.L., Pinkel, D., Dyer, M.J. & Martinez-Climent, J.A. (2005a) Characterization of 8p21.3 chromosomal deletions in B-cell lymphoma: TRAIL-R1 and TRAIL-R2 as candidate dosage-dependent tumor suppressor genes. *Blood*, **106**, 3214–3222.

Rubio-Moscardo, F., Climent, J., Siebert, R., Piris, M.A., Martin-Subero, J.I., Nielander, I., Garcia-Conde, J., Dyer, M.J., Terol, M.J., Pinkel, D. & Martinez-Climent, J.A. (2005b) Mantle-cell lymphoma genotypes identified with CGH to BAC microarrays define a leukemic subgroup of disease and predict patient outcome. *Blood*, **105**, 4445–4454.

Sanchez, C., Diaz-Nido, J. & Avila, J. (2000) Phosphorylation of microtubule-associated protein 2 (MAP2) and its relevance for the regulation of the neuronal cytoskeleton function. *Progress in Neurobiology*, **61**, 133–168.

Serrano, L., Avila, J. & Maccioni, R.B. (1984) Controlled proteolysis of tubulin by subtilisin: localization of the site for MAP2 interaction. *Biochemistry*, **23**, 4675–4681.

Soltani, M.H., Pichardo, R., Song, Z., Sangha, N., Camacho, F., Satyamoorthy, K., Sangueza, O.P. & Setaluri, V. (2005) Microtubule-associated protein 2, a marker of neuronal differentiation, induces mitotic defects, inhibits growth of melanoma cells, and predicts metastatic potential of cutaneous melanoma. *American Journal of Pathology*, **166**, 1841–1850.

Tagawa, H., Karnan, S., Suzuki, R., Matsuo, K., Zhang, X., Ota, A., Morishima, Y., Nakamura, S. & Seto, M. (2005) Genome-wide array-based CGH for mantle cell lymphoma: identification of homozygous deletions of the proapoptotic gene BIM. *Oncogene*, **24**, 1348–1358.

Theurkauf, W.E. & Vallee, R.B. (1982) Molecular characterization of the cAMP-dependent protein kinase bound to microtubule-associated protein 2. *Journal of Biological Chemistry*, **257**, 3284–3290.

Vallee, R.B., DiBartolomeis, M.J. & Theurkauf, W.E. (1981) A protein kinase bound to the projection portion of MAP 2 (microtubule-associated protein 2). *Journal of Cell Biology*, **90**, 568–576.

Williams, M.E., Swerdlow, S.H., Rosenberg, C.L. & Arnold, A. (1992) Characterization of chromosome 11 translocation breakpoints at the bcl-1 and PRAD1 loci in centrocytic lymphoma. *Cancer Research*, **52**, 5541s–5544s.

Zamo, A., Ott, G., Katzenberger, T., Adam, P., Parolini, C., Scarpa, A., Lestani, M., Menestrina, F. & Chilosi, M. (2006) Establishment of the MAVER-1 cell line, a model for leukemic and aggressive mantle cell lymphoma. *Haematologica*, **91**, 40–47.

Zhai, Y., Kronebusch, P.J., Simon, P.M. & Borisy, G.G. (1996) Microtubule dynamics at the G2/M transition: abrupt breakdown of cytoplasmic microtubules at nuclear envelope breakdown and implications for spindle morphogenesis. *Journal of Cell Biology*, **135**, 201–214.

Zhao, X., Li, C., Paez, J.G., Chin, K., Janne, P.A., Chen, T.H., Girard, L., Minna, J., Christiani, D., Leo, C., Gray, J.W., Sellers, W.R. & Meyerson, M. (2004) An integrated view of copy number and allelic alterations in the cancer genome using single nucleotide polymorphism arrays. *Cancer Research*, **64**, 3060–3071.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Fig S1. Heatmaps and genomic distribution of gain and loss in 3q27.3-3q28 and 11q22.3-q23.2 in MCL.

Fig S2. Real-time RT-PCR for detecting *MAP6* transcripts.

Fig S3. Bisulfite sequencing of the *MAP2* CpG island.

Table SI. Characteristics and applications of the investigated MCL cases.

Table SII. Cell line references.

Table SIII. Affymetrix reference samples.

Table SIV. BAC, PAC and fosmid clones used for FISH assays.

Table SV. Primer sequences and PCR conditions.

Table SVI. Regions of recurrent gain and loss in MCL detected by 100K GeneChip analyses.

Table SVII. High level amplifications in MCL.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

51

2.4 Publikation 4: „Chromosomal imbalances and partial uniparental disomies in primary central nervous system lymphoma." Leukemia. 2009 Oct;23(10):1875-84

Schwindt H, Vater I, Kreuz M, Montesinos-Rongen M, Brunn A, Richter J, Gesk S, Ammerpohl O, Wiestler OD, Hasenclever D, Deckert M, Siebert R.

Primäre ZNS-Lymphome (PZNSL) sind aggressive B-Zell-Lymphome, die auf das zentrale Nervensystem begrenzt sind. Sie zeigen eine schlechte Prognose und werden als diffus großzellige B-Zell-Lymphome (DLBCL) klassifiziert [14]. Die Frage, inwieweit sich PZNSL von systemischen DLCBL bezüglich ihrer Pathogenese und ihrer molekulargenetischen Eigenschaften abgrenzen, ist Gegenstand aktueller Forschung.

Im Rahmen der Studie wurden Proben von 19 PZNSL mit Affymetrix 100k-SNP-Arrays untersucht. Als euploide Referenzen wurden 90 frei verfügbare HapMap-Proben sowie sieben laboreigene Kontrollfälle verwendet. Die Analyse der SNP-Arrays erfolgte wie in Abschnitt 1.4 beschrieben. Für 18 der untersuchten PZNSL-Proben lagen Genexpressionsmessungen mit Affymetrix hgu95av2-Chips vor.

Die Analyse der Kopienzahl-Aberrationen ergab elf Regionen mit rekurrenten Deletionen und zehn Regionen mit rekurrenten Zugewinnen. Die dabei betrachteten Regionen mit maximaler Überlappung (MCR) waren in mindestens vier untersuchten PZNSL-Fällen aberrant (>20 %). Für alle rekurrenten Regionen wurde die Genexpression der Gene innerhalb der MCR zwischen Fällen mit und ohne die entsprechende Aberration verglichen. Die dabei beobachtete Korrelation zwischen Gendosis und Expression war gering. Nur in vier Regionen konnten signifikante Gendosiseffekte für zugehörige Probesets gezeigt werden. Allerdings war die untersuchte Fallzahl mit 18 PZNSL-Proben klein. Zusätzlich wurden rekurrente UPD-Regionen bestimmt. Hier waren am häufigsten Regionen auf den Chromosomenarmen 6p (>40 %) und 9p (>20 %) betroffen.

Die im untersuchten Datensatz am häufigsten betroffene Region war 6p21.32. Insgesamt traten sieben einfache Deletionen, zwei homozygote Deletionen und sieben UPDs auf, sodass für die Region in 14 von 19 (74 %) untersuchten Fällen genomische Aberrationen gezeigt werden konnten. Die homozygoten Deletionen wurden mit FISH-Analysen bestätigt. Die zugehörige MCR umfasst die Gene *HLA-DRB*, *HLA-DQA* und *HLA-DQB* und hatte eine Länge von 0,3 Mb. In zehn von 19 untersuchten Fällen war die Region 6q21 von Verlusten betroffen, die unter anderem das Gen *PRDM1* enthält. Für dieses Gen konnte bereits Inaktivierung in PZNSL und DLBCL gezeigt werden [51, 52]. Die Region 9p21.3 war in sechs Fällen deletiert (in einem Fall homozygot) und zeigte in drei weiteren Fällen UPD. Sie umfasste die Gene *MTAP* sowie *CDKN2A/p16* und *CDKN2B/p15.* Auch für diese Region konnten bereits gehäuft Aberrationen in PZNSL gezeigt werden [53]. Methylierungsanalysen in acht PZNSL-Fällen zeigte in sechs Fällen eine Hypermethylierung der Promotorregion von *CDKN2A/p16* an.

Die Region 19q13.43 war am häufigsten von Kopien-Zugewinnen betroffen (47 %) und umfasst ebenso wie die Region 19q13.31 (in 37 % der Fälle zugewonnen) mehrere Gene, die Zinkfingerproteine codieren. Die Region 18q21.33–q23 zeigte in 43 % der untersuchten PZNSL Zugewinne. Mögliche Zielgene der Aberrationen sind hierbei *BCL2* und *MALT1*.

Zusammenfassend konnten in der Publikation 19 PZNSL-Fälle bezüglich ihres genomischen Aberrationsprofils charakterisiert werden. Dabei konnten sowohl bekannte als auch neue rekurrente Aberrationen beschrieben werden. Darüber hinaus konnte gezeigt werden, dass verschiedene genetische und epigenetische Mechanismen an der Deaktivierung bestimmter Zielgene beteiligt sind und eine ausschließliche Betrachtung der Kopienzahl unzureichend ist.

ORIGINAL ARTICLE

Chromosomal imbalances and partial uniparental disomies in primary central nervous system lymphoma

H Schwindt[1,5], I Vater[2,5], M Kreuz[3], M Montesinos-Rongen[1], A Brunn[1], J Richter[2], S Gesk[2], O Ammerpohl[2], OD Wiestler[4], D Hasenclever[3], M Deckert[1,6] and R Siebert[2,6]

[1]Department of Neuropathology, University Hospital of Cologne, Cologne, Germany; [2]Institute of Human Genetics, Christian-Albrechts-University Kiel, University Hospital Schleswig-Holstein, Campus Kiel, Kiel, Germany; [3]Institute for Medical Informatics, Statistics and Epidemiology, (IMISE), University of Leipzig, Leipzig, Germany and [4]German Cancer Research Center, Heidelberg, Germany

To determine the pattern of genetic alterations in primary central nervous system lymphomas (PCNSL), 19 PCNSL were studied by high-density single-nucleotide polymorphism arrays. Recurrent losses involved 6p21.32, 6q21, 8q12–12.2, 9p21.3, 3p14.2, 4q35.2, 10q23.21 and 12p13.2, whereas gains involved 18q21–23, 19q13.31, 19q13.43 and the entire chromosomes X and 12. Partial uniparental disomies (pUPDs) were identified in 6p and 9p21.3. These genomic alterations affected the HLA locus, the CDKN2A/p16, CDKN2B/p15 and MTAP, as well as the PRDM1, FAS, MALT1, and BCL2 genes. Increased methylation values of the CDKN2A/p16 promoter region were detected in 75% (6/8) PCNSL. Gene expression profiling showed 4/21 (20%) minimal common regions of imbalances to be associated with a differential mRNA expression affecting the FAS, STAT6, CD27, ARHGEF6 and SEPT6 genes. Collectively, this study unraveled novel genomic imbalances and pUPD with a high resolution in PCNSL and identified target genes of potential relevance in the pathogenesis of this lymphoma entity.

Introduction

Primary central nervous system lymphomas (PCNSL) are aggressive B-cell lymphomas confined to the central nervous system associated with a poor prognosis.[1,2] They are classified as diffuse large B-cell lymphoma (DLBCL).[3,4] However, it is still a matter of debate whether they differ from systemic DLBCL with respect to their molecular features and pathogenesis.

Studies addressing the genetic characteristics of PCNSL identified recurrent chromosomal abnormalities.[5–7] In addition to translocations of the immunoglobulin (IG) and the BCL6 loci present in a subset of PCNSL, gains and losses of genetic material are recurrent in PCNSL.[8,9] Gains of 18q were the most frequent alteration (38%) in one series of 13 PCNSL analyzed by interphase fluorescence in situ hybridization (FISH).[5] High-level amplifications were mapped to 18q21–23 by comparative genomic hybridization to chromosomes in two tumors of a series of 19 PCNSL.[7] In addition, gains of 12q (63%, 12/19), 1q, 9q, 16p, and 17q (26% each, 5/19) were identified by comparative genomic

hybridization and/or FISH.[7] Furthermore, deletion of 6q was reported in 47% (9/19) PCNSL.[7] Finally, deletion of the human leukocyte antigen (HLA) class II region in 6p21.32 including homozygous loss was reported to affect more than half of all PCNSL and to be associated with loss of MHC expression.[10–12]

Although these studies provided interesting and important insights into the genetic abnormalities in PCNSL, they were limited by a low resolution and the inability to detect partial uniparental disomy (pUPD). Therefore, a series of 19 PCNSL was studied by high-density single-nucleotide polymorphism (SNP) arrays. This high-resolution technology allowed narrowing down chromosomal regions of interest and, furthermore, identification of novel imbalances and regions with pUPD.

Materials and methods

Clinical data
Stereotactic biopsies of 19 immunocompetent patients (11 females and 8 males) were included in this study. All studies were approved by local ethics committees. Informed consent was provided according to the Declaration of Helsinki Principles. Systemic lymphoma manifestation was excluded by extensive staging. All PCNSL were histopathologically classified as CD20 + DLBCL according to the World Health Organization (WHO) classification.[3] The diagnoses were based on a combination of routine morphology and immunohistochemistry using monoclonal mouse antibodies against CD20, CD45, Ki-67 (MIB-1), IgM, IgG, BCL6, and MUM1/IRF4 as reported earlier.[13,14] The clinical characteristics are summarized in Table 1.

GeneChip 100k mapping array

DNA extraction was carried out as reported previously.[15] For each tumor sample, a total of 500 ng DNA were processed according to the Affymetrix GeneChip Mapping 100k Assay Manual (Affymetrix, Santa Clara, CA, USA). Briefly, 250 ng of genomic DNA was digested with XbaI or HindIII and ligated to GeneChip Human Mapping 50k adaptors. Adaptor-ligated restriction fragments were amplified by PCR, purified, fragmented and labeled as described.[16] Samples were hybridized to GeneChip Human Mapping 50k Xba240 or Hind240 arrays (Affymetrix) using the Affymetrix Hybridization Oven 640 (Affymetrix). Washes and staining of the arrays were performed with an Affymetrix Fluidics Station 450 (Affymetrix), and images were obtained using an Affymetrix GeneChip Scanner 3000 (Affymetrix).

Table 1 Patients' data

PCNSL no.	Gender	Age at diagnosis	Affected MCRs	DLBCL subtype significance
01	M	81	1, 5, 6, 10, 12, 13, 16, 19, and 20	GCB
02	F	61	5, 9, 14, and 18	GCB
03	F	68	2, 3, 4, 5, 6, 8, 9, 11, 18, and 20	ABC
04	F	67	1, 5, 6, 10, 12, 18, 19, and 20	GCB
05	F	75	3, 5, 6, 7, 12, 19, 20, and 21	ABC
06	M	66	2, 3, 5, 6, 10, 11, 13, 17, 18, 19, and 20	'type 3'
07	F	77	5, 9, 11, 16 and 21	'type 3'
08	F	78	17	'type 3'
09	M	56	—	'type 3'
10	F	75	21	GCB
11	F	65	17 and 20	ABC
12	F	82	6, 11, 14, 18, 19, and 21	'type 3'
13	M	52	1, 2, 4, 5, 11, 15, 16, and 20	'type 3'
14	F	47	5, 11, and 14	GCB
15	F	76	4, 5, 6, 9, 13, 14, 18, 19, 20, and 21	GCB
16	M	68	5, 7, 8, 16 and 18	GCB
17	M	74	2, 3, 4, 5, 6, 7, 8, 9, 11, 13, 15, and 17	ABC
18	M	55	1, 5, 6, 9, 10, 11, 12, 15, 18 and 21	ABC
19	M	54	5, 6, 7, 8, 11, 14, 15, 16, 19, and 20	NA

Abbreviations: ABC, GCB, 'type 3', activated B-cell-like, germinal center B-cell-like and 'type 3' (non-ABC/non-GCB) of DLBCL classified by gene expression profiling; DLBCL, diffuse large B-cell lymphoma; F, female; M, male; MCR, minimal common region, note that only regions with alterations in more than 20% are assigned as described in Table 2; NA, not analyzed; PCNSL, primary central nervous system lymphoma; pUPD, partial uniparental disomy.

Data analysis

Reference samples: As reference, 90 HapMap samples (30 Centre d'Etude du Polymorphism Humain (CEPH) trios) provided by Affymetrix were used (http://www.affymetrix.com/support/technical/sample_data/hapmap_trio_data.affx). The reference set was completed by a set of seven samples with diploid karyotypes hybridized in the same laboratory as the tumor samples.

Genotyping: The Bayesian robust linear model with Mahalanobis distance classifier (BRLMM) algorithm[17] was used with default parameters (score threshold $= 0.5$, prior size $= 10\,000$, and dynamic model (DM) threshold $= 0.17$) to genotype tumor samples together with euploid reference arrays.

Copy number analysis: Copy number was analyzed using the CNAG program v2.0 using the optimized reference selection method implemented in CNAG.[18] Thus, CNAG selected a gender-specific reference set out of the 97 controls individually for each array. *Xba*240 and *Hind*240 arrays were combined for the analysis. Segmentation of raw copy number data was carried out using the Hidden Markov Model (HMM) approach provided by CNAG. HMM parameters were adjusted individually for each array because of the differences in hybridization quality and tumor cell content (at least 80% as assured morphologically). Starting with default parameters, the mean levels of HMM states were manually adjusted to optimize segmentation results for each sample.

With regard to outliers and technical artifacts, HMM segments with aberrant copy number were considered as copy number aberration only if they consisted of at least 10 consecutive SNPs. The given breakpoints refer to the first neighboring SNPs without alteration. Homozygous deletions were defined as aberrations with copy number $= 0$. Aberrant regions were compared to the published data of the 'Database of Genomic variants.'[19] Regions showing overlap $> 50\%$ with known genomic variants were classified as copy number polymorphisms (CNV).

Detection of loss of heterozygosity/pUPD: The HMM-based method[20] implemented in the dChip program (build date: 11 April 2007)[21,22] was used to infer regions with loss of heterozygosity (LOH). The HMM considering haplotype (linkage disequilibrium (LD)-HMM)[20] method was selected for the LOH calculations to account for LD-induced SNP dependencies. The LOH call threshold was set to the default value of 0.5. An empirical haplotype[20] correction was applied. Thus, the genotypes of putative tumor LOH regions were compared with those observed in euploid reference samples. If the genotypes in the respective region were highly concordant between the tumor sample and at least 5% of the normal reference samples, the LOH region was rejected by dChip.

pUPD regions represent genomic regions in which LOH is not caused by altered copy number. LOH regions determined by dChip were called pUPD if copy number analysis using CNAG did not show aberrations within that region. If an LOH region was partially affected by copy number aberrations, subregions with normal copy number were called pUPD if they comprised at least 50 neighboring SNPs.

Correlation of minimal common regions with gene expression profiles

For all PCNSL (i.e., case nos. 1–18), except for PCNSL 19, gene expression profiles were available (GEO database[23] GSE6047[24]). For all mRNAs in the minimal common regions (MCR) detected by SNP, corresponding tags on the HG-U95Av2 platform (Affymetrix) were evaluated for mRNA expression. Normalized gene expression profiles[24] were analyzed in geWorkbench (http://wiki.c2b2.columbia.edu/workbench/index.php/Home). After replacing all values below 100 by 100 (background noise correction) and log2 transformation, mRNAs of tumors harboring these MCRs were compared with those without genetic alterations in these loci as evidenced by SNP analysis. Statistically significant differential expression was determined using Student's *t*-test with α-correction.
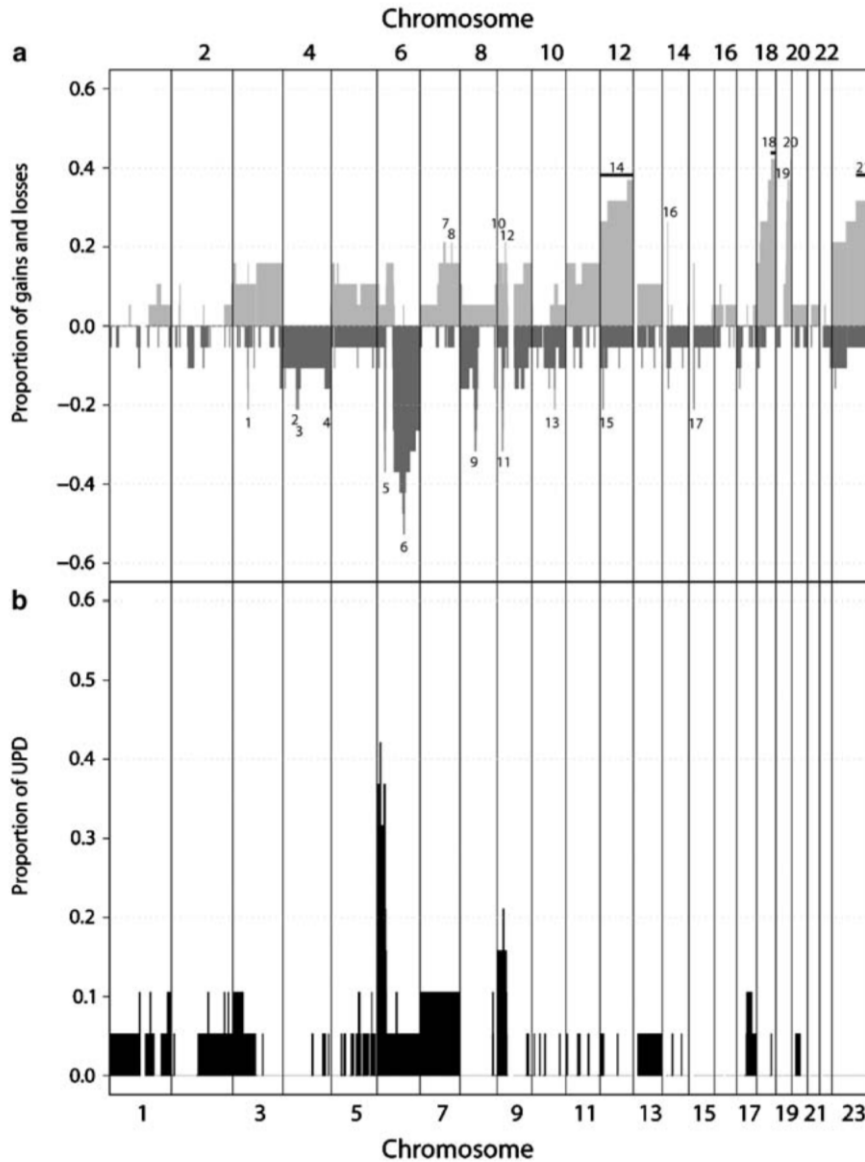
Figure 1 Genome-wide detection of copy number changes and regions with pUPD analyzing 100k GeneChip data of 19 PCNSL. (**a**) Proportion of gains and losses are displayed from 1pter (left) to Xqter (right). Light gray columns indicate chromosomal gain, whereas dark gray columns indicate loss of genetic material. The numbers indicate MCRs of frequent imbalances described in Table 2. (**b**) Proportions of pUPD from 1pter (left) to Xqter (right).

Sequencing of the *PRDM1* gene

Sequencing of the *PRDM1* gene was carried out as described previously.[25] In brief, all exons of the *PRDM1* gene were amplified by PCR. Amplicons were separated by agarose gel electrophoresis, subsequently cleaned up and directly sequenced from both sides.

FISH

To verify copy number results from 100k GeneChip analyses, dual-color FISH was applied. FISH probes consisted of spectrum orange (so) or spectrum green (sg) (Abbott/Vysis, Downers Grove, IL, USA) labeled P1-derived or bacterial artificial chromosome (PAC/BAC) clones (Invitrogen, Karlsruhe, Germany). The clones RP1-93N13 (so) and RP1-172K2 (sg)[12] were used to detect deletions in the HLA region in 6p21.32, and

clones RP11-140A9 (sg) and RP11-1081N13 (so) to detect deletions in 8q12. In addition, two commercial assays were applied to confirm copy number changes of *MALT1* and *ETV6* (LSI *MALT1* and LSI *ETV6* dual-color break apart probes, Abbott/Vysis). Probe preparation and FISH were performed as published.[26] Whenever possible, at least 100 interphase nuclei were evaluated per hybridization.

Bisulfite pyrosequencing

Bisulfite pyrosequencing of *CDKN2A/p16* was performed according to standard protocols with slight modifications in eight PCNSL (case nos. 02, 03, 06, 10, 11, 14, 17 and 19), in which sufficient material was available.[27] Briefly, genomic DNA was bisulfite converted using the EpiTect Bisulfite Conversion Kit (Qiagen, Hilden, Germany). In a subsequent PCR amplification, locus-specific primers were used with one primer

Table 2 MCR of frequent imbalances in 19 PCNSL

MCR	Cytoband	Physical mapping (Mb)[a]	Frequency	Candidate genes[b]
	Loss:			
1	3p14.2	60.7–60.8	4/19 (21%)	FHIT
2	4q12	53.6–54.0	4/19 (21%)	SCFD2
3	4q13.1	60.2–61.0	4/19 (21%)	Unknown
4	4q35.2	187.8–189.3	4/19 (21%)	FAT, ZFP42
5	6p21.32	32.5–32.7	7/19 (37%) 2 HZD, +7 pUPD (7/19; 37%)	HLA-DRB, HLA-DQA, HLA-DQB
6	6q21	106.4–108.4	10 /19 (52%)	PRDM1
9	8q12.1–q12.2	59.8–62.2	6 /19 (32%)	TOX; CA8; RAB2A
11	9p21.3	21.7–22.7	6/19 (32%) 1 HZD+3 pUPD (3/19; 16%)	MTAP, CDKN2A/p16, CDKN2B/p15
13	10q23.21	90.7–90.8	4/19 (21%)	FAS[d], AMSH-LP
15	12p13.2	11.7–11.9	4/19 (21%) 1 HZD	ETV6
17	15q11.2	19.2[c]–20.6	4/19 (21%)	(CNV)
	Gain:			
7	7q21.3–q22.1	93.8–100.4	4/19 (21%)	Unknown
8	7q31.33	125.6–125.7	4/19 (21%)	(CNV)
10	9p24.3	0.2[c]–0.5	4/19 (21%)	(CNV)
12	9p13.3	33.0–34.5	4/19 (21%)	NFX1, BAG1
14	12	Pter-Qter	5/19 (26%)	Several (STAT6, CD27)[d]
16	14q11.2	21.6–22.0	5/19 (26%)	(CNV)
18	18q21.33–q23	58.8 (47.0)–76.1	8/19 (43%)	BCL2, (MALT1)
19	19q13.31	48.2–49.3	7/19 (37%)	ZNF cluster (ETHE1[d])
20	19q13.43	61.5–63.8	9/19 (47%)	ZNF cluster
21	Xq21.33–q28	95.8–154.8	6/19 (32%)	Unknown (ARHGEF6, SEPT6)[d]

Abbreviations: CNV, copy number variation according to the Database of Genomic Variants; HZD, homozygous deletion; MCR, minimal common region; pUPD, partial uniparental disomy.
[a]According to NCBI Build 35.
[b]Common genes with typical tumor suppressor or oncogene properties.
[c]Position of first SNP represented on the microarray.
[d]Differentially expressed genes as determined by gene expression profiling.

biotinylated at the 5′-end (5′biotin-GAGGGGTTGGTTGGTTAT TAGA-3′; 5′-CTACAAACCCTCTACCCACCTAA-3′). Amplification reactions contained ~75 ng bisulfite-converted DNA, primers, AccuPrime Taq Polymerase and buffer II (Invitrogen, Karlsruhe, Germany), 50 mM $MgCl_2$ and 5 mM of each dNTP in a final volume of 50 µl. After initial denaturation, PCR consisted of 45 cycles of each 95 °C for 30 s, 60 °C for 30 s and 68 °C for 30 s, followed by a final synthesis at 68 °C for 2 min. PCR products were verified by gel electrophoresis. Single strands were prepared using the VacuumPrep Tool (Biotage, Uppsala, Sweden), followed by a denaturation step at 85 °C for 2 min and final sequencing primer (5′-CCCTCTACCCACCTAAAT-3′) hybridization. Pyrosequencing was performed using the Pyrosequencer ID and the DNA methylation analysis software Pyro Q-CpG 1.0.9 (Biotage), which was also used to quantify the ratio T:C (mC:C) at the CpG sites analyzed. Assays were validated using an in vitro methylated DNA (Millipore, Billerica, MA, USA) and pooled DNA was isolated from 20 peripheral blood samples from 10 men and 10 women. The latter was also used as a normal control.

Results

Frequency of genomic imbalances and pUPD

All PCNSL harbored at least three regions of chromosomal imbalances or pUPD. The median of aberrant regions per PCNSL was 18. A genomic overview of gains and losses detected in the 19 investigated PCNSL is presented in Figure 1a. A total of 188 genomic losses (median per PCNSL: 9, range: 0–32) and 126 genomic gains (median per PCNSL: 6, range: 0–13) were identified. Overlapping genomic segments of copy number changes delineated MCRs. In each of the cases, between 0 and 12 MCRs were affected. Chromosomal regions that were affected in at least 4 (21%) PCNSL are summarized in Table 2.
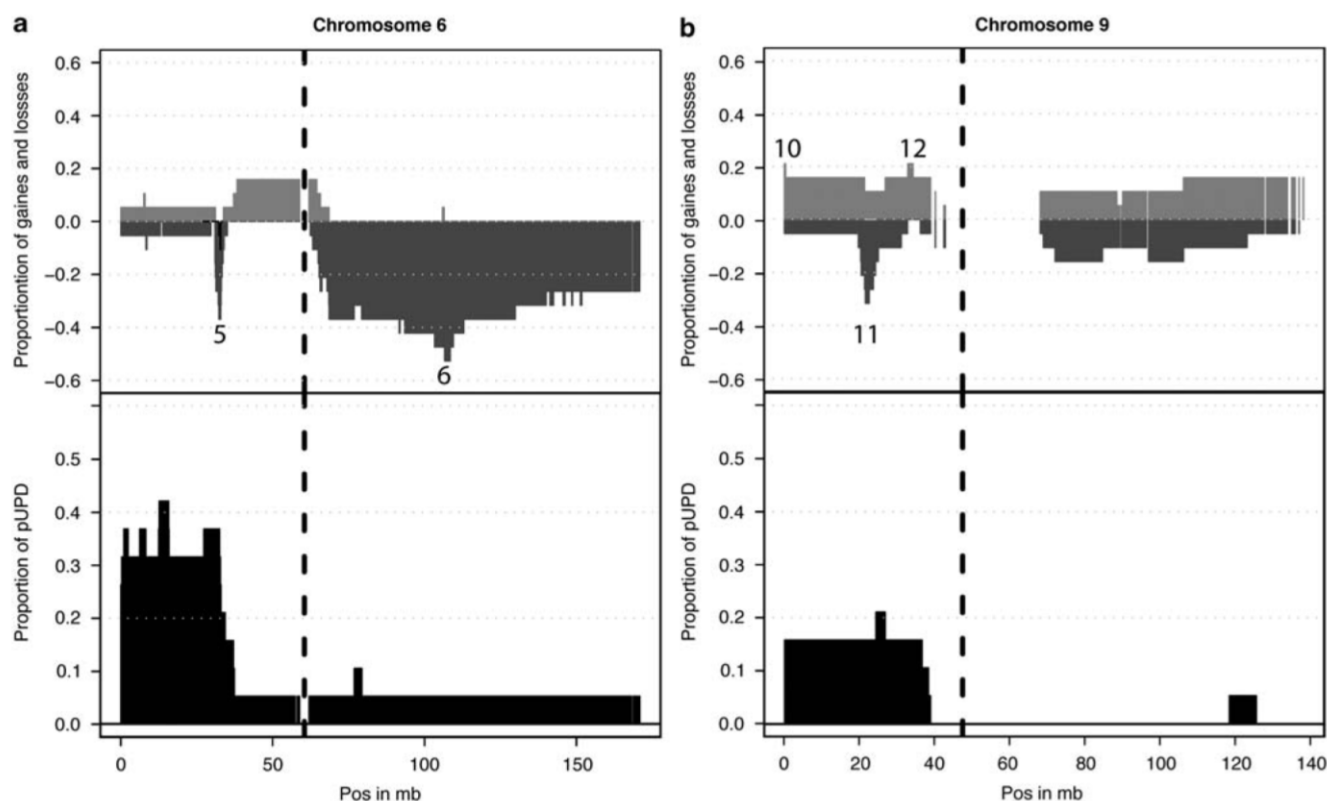
Figure 2 Imbalances and regions with pUPD at chromosome 6 (**a**) and chromosome 9 (**b**) analyzing 100k GeneChip data of 19 PCNSL. The chromosomes are displayed from pter (left) to qter (right). Proportion of deletions is indicated as dark gray columns and proportion of gains as light gray columns. Proportion of pUPD is separately displayed in black color. The numbers indicate MCR of frequent imbalances described in Table 2. 5: MCR of deletion in 6p21.32; 6: MCR of deletion in 6q21; 10: MCR of 9p24.3 (common CNV); 11: MCR of deletion in 9p21.3; and 12: MCR of gain in 9p13.3.

Gains were identified in 7q21.3–q22.1, 7q31.33, 9p24.3, 9p13.3, 12pter-qter, 14q11.2, 18q21.33–q23, 19q13.31, 19q13.43, and Xq21.33–q28. Genomic losses were frequently detected in 3p14.2, 4q12, 4q13.1, 4q35.2, 6p21.32, 6q21, 8q12.1–q12.2, 9p21.3, 10q23.21, 12p13.2 and 15q11.2. Some of these regions contain known CNVs, also present in healthy individuals (Table 2). The 100k GeneChip data of the PCNSL were also subjected to LOH analysis in order to detect regions of pUPD. Overall, we identified 78 regions of pUPD in 17 of these PCNSL with a median number of 4 pUPDs per PCNSL (range: 0–12). The genome-wide distribution of pUPD is shown in Figure 1b.

Loss of genetic material and pUPD

The region most commonly affected by loss or pUPD (14/19, 74%) was located in 6p21.32 (Figure 2a, Table 2, MCR 5). In two of these PCNSL (case nos. 04 and 05), homozygous loss was detected and subsequently confirmed by FISH. The MCR comprised 0.3 Mb and harbored the MHC class II encoding genes *HLA-DRB*, *HLA-DQA*, and *HLA-DQB*. Remarkably, the region most commonly affected by pUPD extended from the HLA region to the telomeric region of 6p.

Of the 19 PCNSL, 10 (53%) harbored a deletion in the long arm of chromosome 6 with an MCR of 2.0 Mb in 6q21 (Figure 2a, Table 2, MCR 6). The MCR encompassed 11 genes, among them the *PRDM1* gene. In this region, a pUPD was not detected. Sequencing of all exons of the *PRDM1* gene failed to detect any mutations of the second allele in the tumors with heterozygous deletion of *PRDM1*.

Six (32%) PCNSL carried a deletion in 8q12.1–q12.2 encompassing an MCR of 2.4 Mb, where *TOX*, *CA8* and *RAB2A* are located (Figure 3d, Table 2, MCR 9). In five of these cases, heterozygous deletion was confirmed by FISH. Six (32%) PCNSL showed a small (1.0 Mb) MCR of a deletion in 9p21.3 involving *CDKN2A/p16*, *CDKN2B/p15*, and *MTAP* (Figure 2b, Table 2, MCR 11). In one of these PCNSL, the deletion was homozygous. Further three PCNSL showed a pUPD in this region.

A small minimally deleted region affecting only one SNP within the 3p14.2 region was detected in 4 (21%) PCNSL, encompassing the *FHIT* gene locus (Figure 3a, Table 2, MCR 1). In 4 (21%) PCNSL, we detected a deletion of 0.4 Mb deletion in 4q12 involving the *SCFD2* gene locus (Figure 3b, Table 2, MCR 2). Another 0.8 Mb MCR of deletion was detected in 4q13.1 (Figure 3b, Table 2, MCR 3). In addition, 4 (21%) PCNSL showed a 1.5-Mb-sized MCR in 4q35.2, where the *FAT* gene is located (Figure 3b, Table 2, MCR 4). Two small (0.2 Mb) MCRs, one in 10q23.21 involving the *FAS* and *AMSH-LP* genes (Figure 3e, Table 2, MCR 13) and another in 12p13.2 affecting the *ETV6* gene (Figure 3f, Table 2, MCR 15), were observed in four (4/19, 21%) PCNSL. One PCNSL harboring a 12p13.2 deletion (PCNSL no. 13) even showed a homozygous loss, which was confirmed by FISH.

Gains of genetic material

Chromosomal gains affected regions ranging in size from 84 kb to whole chromosomes. Eight (43%) PCNSL showed a gain in 18q21.1–q23. In seven of these cases both *MALT1* and *BCL2*
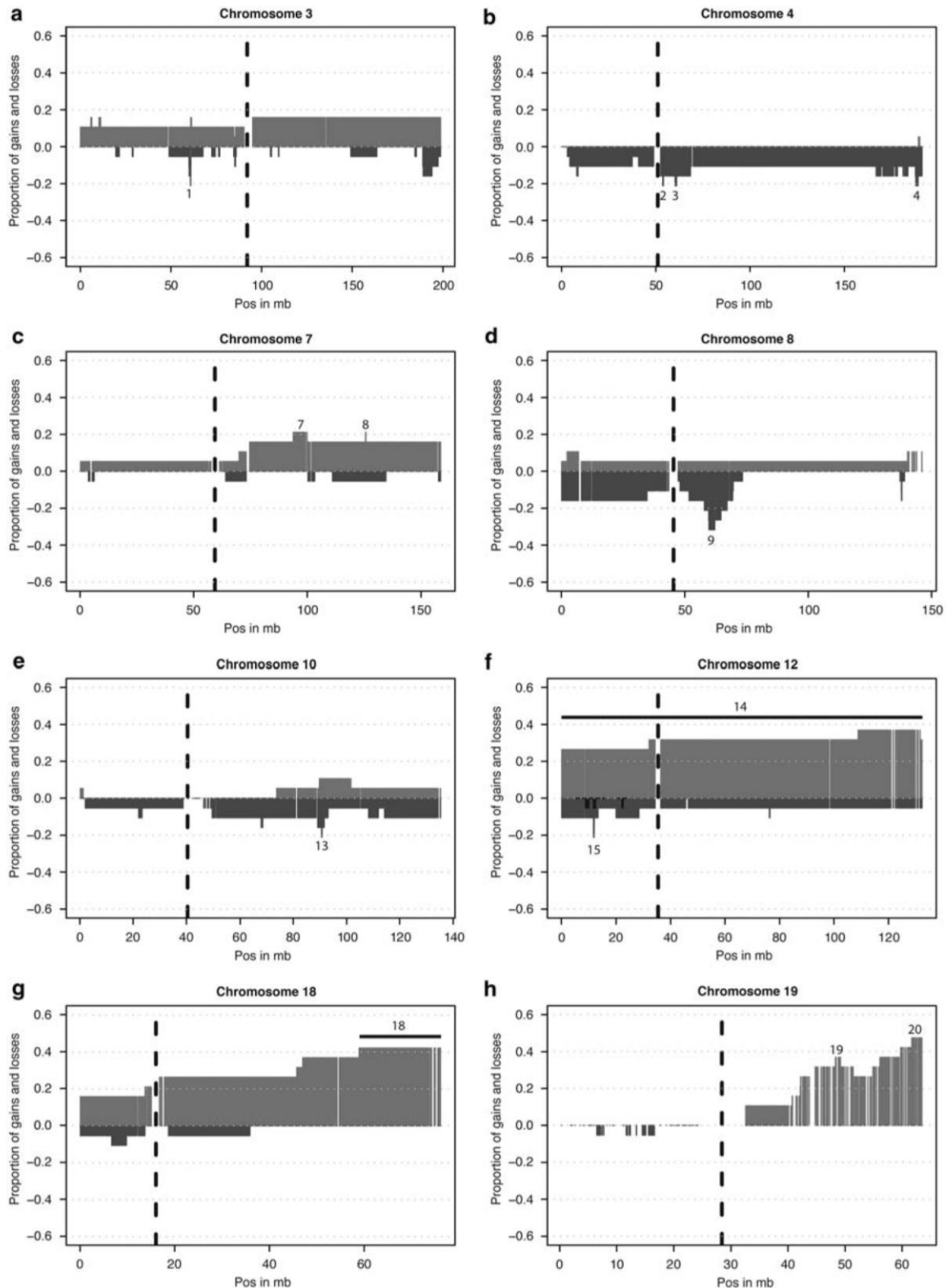
Figure 3 Detailed mapping of chromosomal imbalances in 19 PCNSL. Chromosomes 3 (**a**), 4 (**b**), 7 (**c**), 8 (**d**), 10 (**e**), 12 (**f**), 18 (**g**) and 19 (**h**) are displayed from pter (left) to qter (right). Proportions of deletions and gains are indicated as dark gray and light gray columns, respectively. The numbers indicate frequently altered MCR of frequent imbalances described in Table 2. 1: MCR of deletion in 3p14.2; 2: MCR of deletion in 4q12; 3: MCR of deletion in 4q13.1; 4: MCR of deletion in 4q35.2; 7: MCR of gain in 7q21.3–q22.1; 8: MCR in 7q31.33 (common CNV); 9: MCR of deletion in 8q12.1–q12.2; 13: MCR of deletion in 10q23.21; 14: gain of the entire chromosome 12; 15: MCR of deletion in 12p13.2; 18: MCR of gain in 18q21.33–q23; 19: MCR of gain in 19q13.31; and 20: MCR of gain in 19q13.43.

58
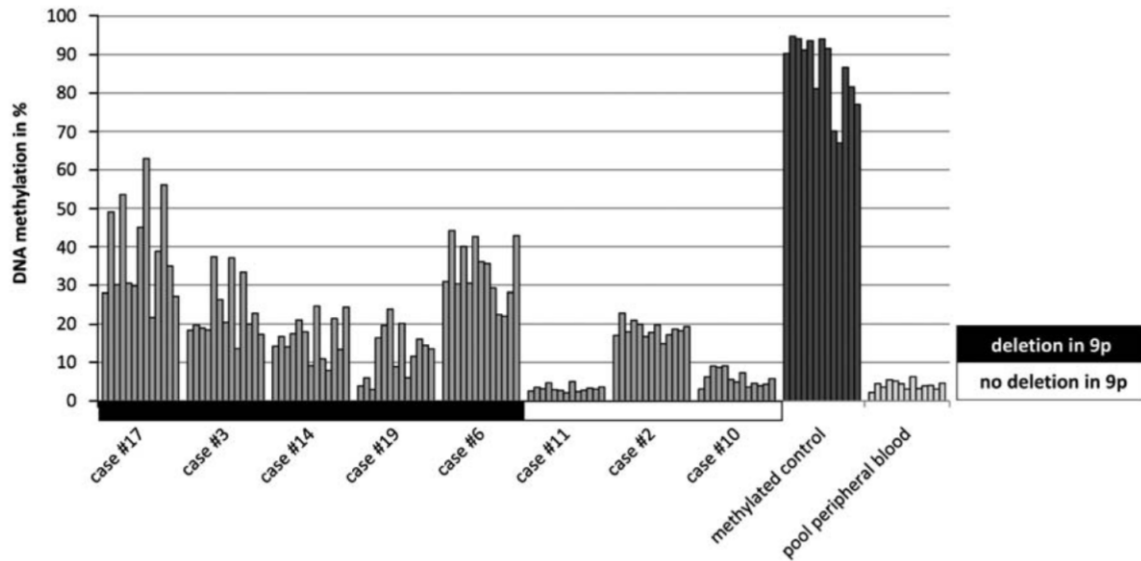
Figure 4 Pyrosequencing analysis of the CDKN2A/p16 locus. Quantification of DNA methylation of 13 CpG dinucleotides within the promoter region of *CDKN2A/p16* by pyrosequencing. Methylation values of each CpG are shown as gray vertical bars for each analyzed sample. PCNSL are arranged according to their chromosomal aberrations in 9p affecting the *CDKN2A/p16* locus (black horizontal bar = heterozygous deletion in 9p, white horizontal bar = no deletion in 9p). PCNSL with 9p deletion showed higher methylation values compared to those without 9p deletion.

were affected, whereas in a single case only the *BCL2* gene was gained (Figure 3g, Table 2, MCR 18). These results were confirmed by FISH. Two small MCRs of chromosomal gain were detected in 19q13.31 (7/19, 37%, Figure 3h, MCR 19) and in 19q13.43 (9/19, 47%, Figure 3h, MCR 20). Both regions harbor several genes encoding zinc finger proteins. Five (26%) PCNSL showed a gain of the entire chromosome 12. In two additional cases, gain was restricted to bands 12p11.21–q24.33 and 12q24.11–q24.33, respectively (Figure 3f, MCR 14). Six (32%) PCNSL showed a gain of Xq21.33–q28 (MCR 21; whole chromosome X: 4/19, Table 2). Four of 19 (21%) PCNSL showed gains in 7q21.3–q22.1 (Figure 3c, MCR 7, Table 2). Four additional tumors harbored gains in 9p13.3 (Table 2), the latter containing the *NFX1* and *BAG1* genes (Figure 2b, MCR 12).

Comparison of MCR with gene expression profiles

To investigate the effect of chromosomal imbalances on expression of genes within the MCR, corresponding tags on the HG-U95Av2 platform were further studied (Table 2).[24] In 17 of 21 MCRs, at least one mRNA could be analyzed with respect to a differential expression between PCNSL with and without any alteration in the respective MCR. Only four MCRs showed a differential expression of at least one mRNA.

In MCR 13 (loss of 10q23.21), only one of seven mRNA was significantly differentially expressed ($P = 1.9E-03$). As depicted in Supplementary Figure 1, FAS expression was very low in cases with alterations in this MCR but also in 5 of 14 PCNSL without alterations in MCR 13.

For MCR 14 (gain of chromosome 12), 668 tags were analyzed; of these, 203 of these tags were differentially expressed, whereas expression of 186 tags (92%) was below 100, that is, the defined background threshold. Seventeen tags, corresponding to 16 genes, were significantly differentially expressed at higher levels, including STAT6 and CD27 (Supplementary Figure 1).

In MCR 19 (gain of chromosome 19q13.31), one tag, corresponding to the gene *ETHE1*, was significantly differentially expressed ($P = 9.8E-03$, Supplementary Figure 1).

In MCR 21 (gain of chromosome Xq21.33–q28), 87 of 217 tags showed a significantly differential expression. Expression values were below 100, that is, the defined background threshold, in 83 of 87 (95%) of these tags. The four tags with a differential expression corresponded to two genes, *ARHGEF6* and *SEPT6* (Supplementary Figure 1).

Methylation status of the *CDKN2A/p16* locus in PCNSL

As shown in Figure 4, six of eight PCNSL showed increased methylation at the *CDKN2A/p16* locus compared with controls. Methylation in PCNSL with heterozygous deletion at the *CDKN2A/p16* locus was increased as compared with PCNSL without deletion in 9p, although this did not reach significance owing to the limited number of samples in each group. Thus, both deletions and DNA methylation may be considered as mechanisms of *CDKN2A/p16* inactivation in PCNSL.

Discussion

In this study, genomic imbalances and LOH were addressed in a series of 19 PCNSL using high-density oligonucleotide arrays. PCNSL showed multiple chromosomal imbalances. In addition to novel recurrent regions of genomic imbalances, unknown regions of pUPD were identified, and homozygous and heterozygous deletions in PCNSL involving the *HLA* locus and the tumor suppressor gene loci *MTAP*, *CDKN2A/p16*, and *CDKN2B/p15* were confirmed.

Alterations in 6p21.32 by deletion or pUPD were the most frequent abnormality (14/19, 73%). The altered region harbors the *HLA-DRB*, *HLA-DQA*, and *HLA-DQB* genes. These findings are in line with previous reports on homozygous and heterozygous deletions at the HLA class II locus, which were associated with a lack of MHC class II antigen expression by

the tumor cells of PCNSL. The progressive loss of MHC proteins may facilitate tumor cell escape from the immune response.[11,12,28] This may be particularly relevant for lymphomas of immunoprivileged organs such as brain and testis, which are characterized by a physiologically downregulated immunophenotype, including low numbers of lymphocytes. In fact, in extracerebral DLBCL, decreased CD8 T-cell numbers have been noticed in MHC class II-negative as compared with MHC class II-positive tumors (median: 2.8 vs 11%), and MHC class II-negative DLBCL were correlated with a poor outcome.[29,30]

Ten (53%) PCNSL showed a deletion involving 6q21, being in accordance with data reported previously.[7,10] The chromosomal region 6q21 is frequently deleted in various B-cell malignancies, including DLBCL, mantle cell lymphoma, acute lymphatic leukemia and immunoblastic lymphoma.[7,31–35] This region has been suggested to harbor a tumor suppressor gene. In fact, 6q21 contains the PRDM1 gene encoding BLIMP1, which has recently been reported to function as tumor suppressor gene in PCNSL as well as in systemic DLBCL of the activated B-cell type.[25,34,36] In a significant fraction (4/21, 21%) of PCNSL, inactivating mutations of the PRDM1 gene were associated with loss of BLIMP1 protein expression.[25] These mutations may contribute to tumorigenesis by blocking terminal B-cell differentiation. In this context, it is of note that the tumor cells of PCNSL are impaired in their differentiation as they do not perform Ig class switch recombination and, thus, are characterized by an IgM + IgD + phenotype.[37] However, in this study, sequencing analysis of 8 of the 10 identified PCNSL with 6q21 deletion failed to detect any PRDM1 mutation, which could suggest that either biallelic BLIMP1 inactivation is a secondary event in PCNSL or that alternative means of BLIMP1 inactivation exist in PCNSL.

SNP chip analysis also showed heterozygous deletions in 8q with a 2.4-Mb MCR between 8q12.1 and 8q12.2 in 6 of 19 (32%) PCNSL. This MCR harbors a gene encoding the nuclear factor TOX (KIAA0808), which is required for CD4 T-cell development.[38] Is it interesting that the IgG + B-cell population was reduced in the spleen of TOX-deficient mice. However, early steps of B-cell development apparently were normal in these animals. It is intriguing to speculate that TOX may contribute to the observed arrest of B-cell differentiation in PCNSL. TOX has been reported to be also deleted in 4% (8/205) of childhood acute lymphatic leukemia.[39] Further candidate genes within the 8q12.1–8q12.2 region are the potential oncogene CA8, which promotes growth, proliferation and invasion of carcinoma cells,[40–42] and the RAB2A gene encoding the ras-related rab2.

The tumor suppressor genes, CDKN2A/p16 and CDKN2B/p15, at 9p21.3 as well as MTAP were targeted by recurrent deletions and pUPD present in 6 and 3 of 19 (32 and 16%) PCNSL, respectively. One PCNSL even showed a biallelic loss.[43] Furthermore, CDKN2A/p16 has been shown to be methylated in 64% (16/25) in PCNSL.[44] P16, that is, one of the encoded proteins, specifically inhibits CDK4 and CDK6 and potentially blocks G1 cell-cycle progression through the dephosphorylation of retinoblastoma through its inhibitory effect on the CDK4/cyclin D1 complex activity.[45] Compared with previous studies, in which DNA methylation of the CDKN2A/p16 locus in PCNSL was mostly analyzed by methylation-specific PCR, we accurately quantitated methylation of 13 CpG dinucleotides by bisulfite pyrosequencing of DNA from eight PCNSL of this series. Our identification of recurrent CDKN2A/p16 promoter methylation, which likely leads to its invactivation, is in accordance with Chu et al.[44] Moreover, the frequent

methylation of the gene in PCNSL with heterozygous deletions suggests that biallelic inactivation in these tumors may result from two different mechanisms, that is DNA methylation and chromosomal deletion.

Additional recurrent deletions present in at least 4 (21%) PCNSL were in 3p14.2, 4q12, 4q13.1, 4q35.2, 10q23.21, and 12p13.2. In 3p14.2, the FHIT gene was partially deleted. FHIT, which is supposed to have a key function in tumor suppression,[46,47] has been shown to be either deleted or mutated in 33% (19/57) of systemic DLBCL. Clinically relevant, decreased FHIT expression has been reported to correlate with an inferior prognosis in systemic DLBCLs.[48] In 4q12, a 200-kb MCR of deletion involving the SCFD2 gene, a potential p53 target gene, was detected.[49] In 4q35.2, the FAT tumor suppressor gene was affected by deletion. Another small MCR of deletion affected the tumor suppressor gene FAS gene at 10q23.21. FAS (CD95), a transmembrane receptor of the tumor necrosis factor superfamily, is an important mediator of apoptosis.[50] FAS mutations, which may impair apoptosis, have been identified in 20% (2/10) PCNSL.[15] In 12p13.2, a 200-kb MCR of deletion was detected with one PCNSL showing biallelic loss; in another PCNSL, the pattern of imbalances suggested a deletion to have occurred before amplification of an allele. The deletions affected the ETV6 gene, which is frequently translocated in hematopoietic tumors.[51]

In PCNSL, recurrent gains in 18q have been described by various groups.[1,5–7,10] 18q21 harbors the BCL2 and MALT1 genes. MALT1 complexes with BCL10 and a CARD protein to activate the nuclear factor-κB pathway.[52] In fact, both MALT1 and BCL10 are expressed by the tumor cells of PCNSL, which also show evidence for an activation of the nuclear factor-κB pathway.[53] In our study, 8/19 PCNSL and 7/19 PCNSL showed gains of BCL2 and MALT1, respectively. Microarray data identified two small MCRs frequently gained in 19q13.31 and 19q13.43. Both regions harbor several genes encoding zinc finger proteins, which might be involved in transcriptional regulation. This region partly overlaps with a region of gain in 19q13.12–13.43 detected in 2 (22%) PCNSL recently.[10] In accordance with previous observations,[5–7] two PCNSL showed gains of 12p11.21–q24.33 and 12q24.11–12q24.33, and five further tumors showed triploidy of chromosome 12. Furthermore, PCNSL had two small MCRs of copy number gains in 7q21.3–q22.1 and 9p21.1–p21.3 and one large MCR of recurrent copy number gain in Xq21.33–q28. The region 7q21.3–q22.1 harbors the candidate genes NFX1 and BAG1. NFX1 is a nuclear transcription factor, which downregulates HLA class II antigen expression.[54] Bag1 interacts with Bcl2 to increase its antiapoptoic activity.[55]

It is tempting to speculate that PCNSL corresponding to the ABC and GCB subgroups as defined for systemic[56–58] may be characterized by different patterns of genomic aberrations indicating particular and distinct pathways relevant for pathogenesis. In this series of PCNSL, MCR 2 (loss 4q12), MCR 3 (loss 4q13.1), and MCR 15 (loss 12p13.2) were exclusively associated with the ABC-type, whereas MCR 14 (gain chromosome 12) was only observed in the GCB-subtype of PCNSL. However, this observation should be interpreted with caution due to the low number of PCNSL available for such studies, which precludes such an analysis.

In conclusion, the identification of novel candidate tumor suppressor genes and oncogene loci through high-density oligonucleotide SNP array analysis may provide the basis for further studies aiming at the identification of genes involved in PCNSL tumorigenesis.

Conflict of Interest

Acknowledgements

References

1 Montesinos-Rongen M, Siebert R, Deckert M. Primary lymphoma of the central nervous system: just DLBCL or not? *Blood* 2009; **113**: 7–10.

2 Schlegel U, Schmidt-Wolf IG, Deckert M. Primary CNS lymphoma: clinical presentation, pathological classification, molecular pathogenesis and treatment. *J Neurol Sci* 2000; **181**: 1–12.

3 Deckert M, Paulus W. Malignant lymphomas. In: Louis DN, Ohgaki H, Wiestler OD, Cavenee WK (eds). *WHO Classification of Tumors Pathology and Genetics of Tumours of the Nervous System*, 4th edn. IRAC: Lyon, 2007, pp. 188–192.

4 Kluin P, Deckert M, Ferry JA. Primary diffuse large B-cell lymphoma of the CNS. In: Swerdlow SH, Campo E, Harris NL, Jaffe ES, Pileri SA, Stein H, et al. (eds). *WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues*. IARC: Lyon, 2008, pp. 240–241.

5 Montesinos-Rongen M, Zühlke-Jenisch R, Gesk S, Martin-Subero JI, Schaller C, Van Roost D et al. Interphase cytogenetic analysis of lymphoma-associated chromosomal breakpoints in primary diffuse large B-cell lymphomas of the central nervous system. *J Neuropathol Exp Neurol* 2002; **61** (10): 926–933.

6 Rickert CH, Dockhorn-Dworniczak B, Simon R, Paulus W. Chromosomal imbalances in primary lymphomas of the central nervous system. *Am J Pathol* 1999; **155**: 1445–1451.

7 Weber T, Weber RG, Kaulich K, Actor B, Meyer-Puttlitz B, Lampel S et al. Characteristic chromosomal imbalances in primary central nervous system lymphomas of the diffuse large B-cell type. *Brain Pathol* 2000; **10**: 73–84.

8 Montesinos-Rongen M, Akasaka T, Zühlke-Jenisch R, Schaller C, Van Roost D, Wiestler OD et al. Molecular characterization of BCL6 breakpoints in primary diffuse large B-cell lymphomas of the central nervous system identifies GAPD as novel translocation partner. *Brain Pathol* 2003; **13**: 534–538.

9 Schwindt H, Akasaka T, Zühlke-Jenisch R, Hans V, Schaller C, Klapper W et al. Chromosomal translocations fusing the BCL6 gene to different partner loci are recurrent in primary central nervous system lymphoma and may be associated with aberrant somatic hypermutation or defective class switch recombination. *J Neuropathol Exp Neurol* 2006; **65**: 776–782.

10 Booman M, Szuhai K, Rosenwald A, Hartmann E, Kluin-Nelemans H, de Jong D et al. Genomic alterations and gene expression in primary diffuse large B-cell lymphomas of immune-privileged sites: the importance of apoptosis and immunomodulatory pathways. *J Pathol* 2008; **216**: 209–217.

11 Jordanova ES, Riemersma SA, Philippo K, Giphart-Gassler M, Schuuring E, Kluin PM. Hemizygous deletions in the HLA region account for loss of heterozygosity in the majority of diffuse large B-cell lymphomas of the testis and the central nervous system. *Genes Chromosomes Cancer* 2002; **35**: 38–48.

12 Riemersma SA, Jordanova ES, Schop RF, Philippo K, Looijenga LH, Schuuring E et al. Extensive genetic alterations of the HLA region, including homozygous deletions of HLA class II genes in B-cell lymphomas arising in immune-privileged sites. *Blood* 2000; **96**: 3569–3577.

13 Brunn A, Montesinos-Rongen M, Strack A, Reifenberger G, Mawrin C, Schaller C et al. Expression pattern and cellular sources of chemokines in primary central nervous system lymphoma. *Acta Neuropathol (Berl)* 2007; **114**: 271–276.

14 Montesinos-Rongen M, Küppers R, Schlüter D, Spieker T, Van Roost D, Schaller C et al. Primary central nervous system lymphomas are derived from germinal-center B cells and show a preferential usage of the V4–34 gene segment. *Am J Pathol* 1999; **155**: 2077–2086.

15 Montesinos-Rongen M, Van Roost D, Schaller C, Wiestler OD, Deckert M. Primary diffuse large B-cell lymphomas of the central nervous system are targeted by aberrant somatic hypermutation. *Blood* 2004; **103**: 1869–1875.

16 Kennedy GC, Matsuzaki H, Dong S, Liu WM, Huang J, Liu G et al. Large-scale genotyping of complex DNA. *Nat Biotechnol* 2003; **21**: 1233–1237.

17 Affymetrix. BRLMM: an improved genotype calling method for the GeneChip human mapping 500k array set. 2006, White Paper: Affymetrix Inc.

18 Nannya Y, Sanada M, Nakazaki K, Hosoya N, Wang L, Hangaishi A et al. A robust algorithm for copy number detection using high-density oligonucleotide single nucleotide polymorphism genotyping arrays. *Cancer Res* 2005; **65**: 6071–6079.

19 Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y et al. Detection of large-scale variation in the human genome. *Nat Genet* 2004; **36**: 949–951.

20 Beroukhim R, Lin M, Park Y, Hao K, Zhao X, Garraway LA et al. Inferring loss-of-heterozygosity from unpaired tumors using high-density oligonucleotide SNP arrays. *PLoS Comput Biol* 2006; **2**: e41.

21 Lin M, Wei LJ, Sellers WR, Lieberfarb M, Wong WH, Li C. dChipSNP: significance curve and clustering of SNP-array-based loss-of-heterozygosity data. *Bioinformatics* 2004; **20**: 1233–1240.

22 Zhao X, Li C, Paez JG, Chin K, Janne PA, Chen TH et al. An integrated view of copy number and allelic alterations in the cancer genome using single nucleotide polymorphism arrays. *Cancer Res* 2004; **64**: 3060–3071.

23 Barrett T, Troup DB, Wilhite SE, Ledoux P, Rudnev D, Evangelista C et al. NCBI GEO: mining tens of millions of expression profiles—database and tools update. *Nucleic Acids Res* 2007; **35**: D760–D765.

24 Montesinos-Rongen M, Brunn A, Bentink S, Basso K, Lim WK, Klapper W et al. Gene expression profiling suggests primary central nervous system lymphomas to be derived from a late germinal center B cell. *Leukemia* 2008; **22**: 400–405.

25 Courts C, Montesinos-Rongen M, Brunn A, Bug S, Siemer D, Hans V et al. Recurrent inactivation of the PRDM1 gene in primary central nervous system lymphoma. *J Neuropathol Exp Neurol* 2008; **67**: 720–727.

26 Martin-Subero JI, Harder L, Gesk S, Schlegelberger B, Grote W, Martinez-Climent JA et al. Interphase FISH assays for the detection of translocations with breakpoints in immunoglobulin light chain loci. *Int J Cancer* 2002; **98**: 470–474.

27 Tost J, Gut IG. Analysis of gene-specific DNA methylation patterns by pyrosequencing technology. *Methods Mol Biol* 2007; **373**: 89–102.

28 Booman M, Douwes J, Glas AM, Riemersma SA, Jordanova ES, Kok K et al. Mechanisms and effects of loss of human leukocyte antigen class II expression in immune-privileged site-associated B-cell lymphoma. *Clin Cancer Res* 2006; **12**: 2698–2705.

29 List AF, Spier CM, Miller TP, Grogan TM. Deficient tumor-infiltrating T-lymphocyte response in malignant lymphoma: relationship to HLA expression and host immunocompetence. *Leukemia* 1993; **7**: 398–403.

30 Rimsza LM, Roberts RA, Miller TP, Unger JM, LeBlanc M, Braziel RM et al. Loss of MHC class II gene and protein expression in diffuse large B-cell lymphoma is related to decreased tumor immunosurveillance and poor patient survival regardless of other prognostic factors: a follow-up study from the Leukemia and Lymphoma Molecular Profiling Project. *Blood* 2004; **103**: 4251–4258.

31 Boonstra R, Koning A, Mastik M, van den Berg A, Poppema S. Analysis of chromosomal copy number changes and oncoprotein expression in primary central nervous system lymphomas: frequent loss of chromosome arm 6q. *Virchows Arch* 2003; **443**: 164–169.

32 Bosga-Bouwer AG, Kok K, Booman M, Boven E, van der Vlies P, van den Berg A et al. Array comparative genomic hybridization reveals a very high frequency of deletions of the long arm of chromosome 6 in testicular lymphoma. *Genes Chromosomes Cancer* 2006; **45**: 976–981.

33 Bosga-Bouwer AG, van den Berg A, Haralambieva E, de Jong D, Boonstra R, Kluin P et al. Molecular, cytogenetic, and immuno-

phenotypic characterization of follicular lymphoma grade 3B; a separate entity or part of the spectrum of diffuse large B-cell lymphoma or follicular lymphoma? *Hum Pathol* 2006; **37**: 528–533.

34 Pasqualucci L, Compagno M, Houldsworth J, Monti S, Grunn A, Nandula SV *et al.* Inactivation of the PRDM1/BLIMP1 gene in diffuse large B cell lymphoma. *J Exp Med* 2006; **203**: 311–317.

35 Thelander EF, Ichimura K, Corcoran M, Barbany G, Nordgren A, Heyman M *et al.* Characterization of 6q deletions in mature B cell lymphomas and childhood acute lymphoblastic leukemia. *Leuk Lymphoma* 2008; **49**: 477–487.

36 Tam W, Gomez M, Chadburn A, Lee JW, Chan WC, Knowles DM. Mutational analysis of PRDM1 indicates a tumor-suppressor role in diffuse large B-cell lymphomas. *Blood* 2006; **107**: 4090–4100.

37 Montesinos-Rongen M, Schmitz R, Courts C, Stenzel W, Bechtel D, Niedobitek G *et al.* Absence of immunoglobulin class switch in primary lymphomas of the central nervous system. *Am J Pathol* 2005; **166**: 1773–1779.

38 Aliahmad P, Kaye J. Development of all CD4T lineages requires nuclear factor TOX. *J Exp Med* 2008; **205**: 245–256.

39 Bardet V, Couque N, Cattolico L, Hetet G, Devaux I, Duprat S *et al.* Molecular analysis of nonrandom 8q12 deletions in acute lymphoblastic leukemia: identification of two candidate genes. *Genes Chromosomes Cancer* 2002; **33**: 178–187.

40 Lu SH, Takeuchi T, Fujita J, Ishida T, Akisawa Y, Nishimori I *et al.* Effect of carbonic anhydrase-related protein VIII expression on lung adenocarcinoma cell growth. *Lung Cancer* 2004; **44**: 273–280.

41 Morimoto K, Nishimori I, Takeuchi T, Kohsaki T, Okamoto N, Taguchi T *et al.* Overexpression of carbonic anhydrase-related protein XI promotes proliferation and invasion of gastrointestinal stromal tumors. *Virchows Arch* 2005; **447**: 66–73.

42 Nishikata M, Nishimori I, Taniuchi K, Takeuchi T, Minakuchi T, Kohsaki T *et al.* Carbonic anhydrase-related protein VIII promotes colon cancer cell growth. *Mol Carcinog* 2007; **46**: 208–214.

43 Haller F, Lobke C, Ruschhaupt M, Cameron S, Schulten HJ, Schwager S *et al.* Loss of 9p leads to p16INK4A down-regulation and enables RB/E2F1-dependent cell cycle promotion in gastrointestinal stromal tumours (GISTs). *J Pathol* 2008; **215**: 253–262.

44 Chu LC, Eberhart CG, Grossman SA, Herman JG. Epigenetic silencing of multiple genes in primary CNS lymphoma. *Int J Cancer* 2006; **119**: 2487–2491.

45 Serrano M, Hannon GJ, Beach D. A new regulatory motif in cell-cycle control causing specific inhibition of cyclin D/CDK4. *Nature* 1993; **366**: 704–707.

46 Fujishita T, Doi Y, Sonoshita M, Hiai H, Oshima M, Huebner K *et al.* Development of spontaneous tumours and intestinal lesions in Fhit gene knockout mice. *Br J Cancer* 2004; **91**: 1571–1574.

47 Kameoka Y, Tagawa H, Tsuzuki S, Karnan S, Ota A, Suguro M *et al.* Contig array CGH at 3p14.2 points to the FRA3B/FHIT common fragile region as the target gene in diffuse large B-cell lymphoma. *Oncogene* 2004; **23**: 9148–9154.

48 Chen PM, Yang MH, Hsiao LT, Yu IT, Chu CJ, Chao TC *et al.* Decreased FHIT protein expression correlates with a worse prognosis in patients with diffuse large B-cell lymphoma. *Oncol Rep* 2004; **11**: 349–356.

49 Krieg AJ, Hammond EM, Giaccia AJ. Functional analysis of p53 binding under differential stresses. *Mol Cell Biol* 2006; **26**: 7030–7045.

50 Oehm A, Behrmann I, Falk W, Pawlita M, Maier G, Klas C *et al.* Purification and molecular cloning of the APO-1 cell surface antigen, a member of the tumor necrosis factor/nerve growth factor receptor superfamily. Sequence identity with the Fas antigen. *J Biol Chem* 1992; **267**: 10709–10715.

51 Wlodarska I, Mecucci C, Baens M, Marynen P, van den Berghe H. ETV6 gene rearrangements in hematopoietic malignant disorders. *Leuk Lymphoma* 1996; **23**: 287–295.

52 Ferch U, zum Buschenfelde CM, Gewies A, Wegener E, Rauser S, Peschel C *et al.* MALT1 directs B cell receptor-induced canonical nuclear factor-kappaB signaling selectively to the c-Rel subunit. *Nat Immunol* 2007; **8**: 984–991.

53 Courts C, Montesinos-Rongen M, Martin-Subero JI, Brunn A, Siemer D, Zühlke-Jenisch R *et al.* Transcriptional profiling of the nuclear factor-kappaB pathway identifies a subgroup of primary lymphoma of the central nervous system with low BCL10 expression. *J Neuropathol Exp Neurol* 2007; **66**: 230–237.

54 Song Z, Krishna S, Thanos D, Strominger JL, Ono SJ. A novel cysteine-rich sequence-specific DNA-binding protein interacts with the conserved X-box motif of the human major histocompatibility complex class II genes via a repeated Cys–His domain and functions as a transcriptional repressor. *J Exp Med* 1994; **180**: 1763–1774.

55 Takayama S, Sato T, Krajewski S, Kochel K, Irie S, Millan JA *et al.* Cloning and functional analysis of BAG-1: a novel Bcl-2-binding protein with anti-cell death activity. *Cell* 1995; **80**: 279–284.

56 Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A *et al.* Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 2000; **403**: 503–511.

57 Lenz G, Wright GW, Emre NC, Kohlhammer H, Dave SS, Davis RE *et al.* Molecular subtypes of diffuse large B-cell lymphoma arise by distinct genetic pathways. *Proc Natl Acad Sci USA* 2008; **105**: 13520–13525.

58 Rosenwald A, Wright G, Chan WC, Connors JM, Campo E, Fisher RI *et al.* The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *N Engl J Med* 2002; **346**: 1937–1947.

Supplementary Information accompanies the paper on the Leukemia website (http://www.nature.com/leu)

2.5 Publikation 5: „High resolution SNP array genomic profiling of peripheral T cell lymphomas, not otherwise specified, identifies a subgroup with chromosomal aberrations affecting the REL locus" Br J Haematol. 2010 Feb;148(3):402-12

Hartmann S, Gesk S, Scholtysik R, Kreuz M, Bug S, Vater I, Döring C, Cogliatti S, Parrens M, Merlio JP, Kwiecinska A, Porwit A, Piccaluga PP, Pileri S, Hoefler G, Küppers R, Siebert R, Hansmann ML.

Das periphere T-Zell-Lymphom – nicht anderweitig spezifiziert – (PTCL-NOS) ist eine seltene Erkrankung und macht circa 9 % aller aggressiven Lymphome aus. PTCL-NOS sind äußerst aggressiv und prognostisch ungünstig (ereignisfreies 5-Jahres-Überleben 20–30 %) [14].

In der Studie wurden 49 periphere T-Zell-Lymphome (PTCL-NOS) mit Affymetrix 250k-Sty-SNP-Arrays untersucht. Als Kontrollen dienten 39 HapMap-Fälle von Affymetrix sowie 20 laboreigene euploide Referenzfälle. Für zwei Lymphomfälle konnten weniger als 90 % der Genotypen bestimmt werden. Aus Qualitätsgründen wurden diese Fälle daher von der weiteren Analyse ausgeschlossen. Die „Call-Rate" der übrigen Lymphomfälle lag im Median bei 97 %. Die Analyse der SNP-Arrays erfolgte wie in Abschnitt 1.4 beschrieben. Aberrationen, die weniger als fünf zusammenhängende SNPs umfassten, wurden als mögliche Ausreißer bzw. falsch positive Aberrationen verworfen.

25 der analysierten 47 PTCL-NOS zeigten keinerlei genomische Aberrationen, mit Ausnahme von Kopienzahl-Polymorphismen (CNP) und Deletionen des T-Zell-Rezeptor Genlocus. Diese Beobachtung ist konsistent mit den Ergebnissen vergleichbarer Studien [54, 55]. Mittels Analyse der T-Zell-Rezeptor β-Rekombination konnte gezeigt werden, dass Tumorproben von Fällen mit komplexen genetischen Aberrationsmustern weniger infiltrierende Bystander-Zellen umfassten. Ein geringer Tumorzellgehalt der Proben verhindert daher vermutlich in einigen Fällen die Detektion von genetischen Aberrationen.

Chromosom 7 war am häufigsten von Zugewinnen betroffen, unter anderem die Region 7q21. In dieser Region liegt das Gen *CDK6,* für das bereits eine Bedeutung für die Regulation des Zellzyklus in T-Zell-Lymphomen gezeigt wurde [56]. Für die Region 2p15–16 konnten vier einfache Zugewinne sowie eine High-Level-Amplifikation gezeigt werden, die das Gen *REL* umfassen. Amplifikation von *REL* wurden für PTCL-NOS bisher nicht beschrieben, konnten aber bereits für andere Lymphome z. B. DLBCL gezeigt werden [57]. Für drei Fälle wurden die Zugewinne mit FISH erfolgreich validiert und in einem unabhängigen Datensatz von 18 PTCL-NOS konnten drei weitere Fälle mit *REL*-Zugewinnen nachgewiesen werden.

Die Region 10p11 zeigte in der Analyse am häufigsten Deletionen. Insgesamt wurden in sechs Fällen einfache und in einem Fall eine homozygote Deletion detektiert. Die Region umfasst die Gene *ZEB1*, *ARHGAP12*, *KIF5B*, *EPC1* und *CCDC7.*In anderen Studien wurden Verluste der Region in circa 15 % der untersuchten PTCL-NOS-Fälle gezeigt [54, 58]. Für adulte T-Zell-Leukämien konnten bereits Translokationen für das Gen *EPC1* gezeigt werden [59]. Die Deletionen auf Chromosom 10 konnten nur in drei von fünf getesteten Fällen mit FISH validiert werden. Eine mögliche Erklärung für die Diskrepanz der Ergebnisse ist der teilweise erhebliche Anteil an infiltrierenden Bystander-Zellen in den Tumorproben. Auf Chromosom 9p21 konnte in zwei Fällen eine homozygote Deletion gezeigt werden, die die Gene *CDKN2A*, *CDKN2B* und *MTAP* umfasst und bereits für PTCL-NOS beschrieben wurde [54]. UPD-Regionen zeigten im analysierten Datensatz nur geringe Rekurrenz.

bjh research paper

High resolution SNP array genomic profiling of peripheral T cell lymphomas, not otherwise specified, identifies a subgroup with chromosomal aberrations affecting the *REL* locus

Sylvia Hartmann,[1] Stefan Gesk,[2] René Scholtysik,[3] Markus Kreuz,[4] Stefanie Bug,[2] Inga Vater,[2] Claudia Döring,[1,5] Sergio Cogliatti,[6] Marie Parrens,[7] Jean-Philippe Merlio,[7] Anna Kwiecinska,[8] Anna Porwit,[8] Pier Paolo Piccaluga,[9] Stefano Pileri,[9] Gerald Hoefler,[10] Ralf Küppers,[3] Reiner Siebert[2] and Martin-Leo Hansmann[1]

[1]*Senckenberg Institute of Pathology, University of Frankfurt, Frankfurt,* [2]*Institute of Human Genetics, Christian-Albrechts-University Kiel, University Hospital Schleswig-Holstein, Campus Kiel,* [3]*Institute of Cell Biology, University of Duisburg-Essen, Medical School Essen, Essen,* [4]*Institute of Medical Informatics, Statistics and Epidemiology, University of Leipzig, Leipzig,* [5]*Institute for Informatics, University of Frankfurt, Frankfurt, Germany,* [6]*Institute of Pathology, Kantonsspital St.Gallen, St.Gallen, Switzerland,* [7]*Department of Pathology and Tumour Genetics, CHU Bordeaux and EA2406, Bordeaux, France,* [8]*Department of Pathology, Karolinska Institutet and Karolinska University Hospital Solna, Stockholm, Sweden,* [9]*Institute of Haematology and Medical Oncology "L. and A. Seràgnoli,", Sant'Orsola-Malpighi Hospital, University of Bologna, Bologna, Italy, and* [10]*Institute of Pathology, Medical School University of Graz, Graz, Austria*

Correspondence: Sylvia Hartmann, MD, Senckenberg Institute of Pathology, University of Frankfurt, Theodor-Stern-Kai 7, D-60590 Frankfurt, Germany.
E-mail: s.hartmann@em.uni-frankfurt.de

Summary

Little is known about genomic aberrations in peripheral T cell lymphoma, not otherwise specified (PTCL NOS). We studied 47 PTCL NOS by 250k GeneChip single nucleotide polymorphism arrays and detected genomic imbalances in 22 of the cases. Recurrent gains and losses were identified, including gains of chromosome regions 1q32–43, 2p15–16, 7, 8q24, 11q14–25, 17q11–21 and 21q11–21 (≥5 cases each) as well as losses of chromosome regions 1p35–36, 5q33, 6p22, 6q16, 6q21–22, 8p21–23, 9p21, 10p11–12, 10q11–22, 10q25–26, 13q14, 15q24, 16q22, 16q24, 17p11, 17p13 and Xp22 (≥4 cases each). Genomic imbalances affected several regions containing members of nuclear factor-kappaB signalling and genes involved in cell cycle control. Gains of 2p15–16 were confirmed in each of three cases analysed by fluorescence *in situ* hybridization (FISH) and were associated with breakpoints at the *REL* locus in two of these cases. Three additional cases with gains of the *REL* locus were detected by FISH among 18 further PTCL NOS. Five of 27 PTCL NOS investigated showed nuclear expression of the REL protein by immunohistochemistry, partly associated with genomic gains of the *REL* locus. Therefore, in a subgroup of PTCL NOS gains/rearrangements of *REL* and expression of REL protein may be of pathogenetic relevance.

Keywords: non-Hodgkin-lymphoma, peripheral T cell lymphoma NOS, SNP array, genomic profiling, *REL*.

Peripheral T cell lymphomas (PTCL) represent only approximately 9% of aggressive lymphomas (Rüdiger *et al*, 2002). Although transformation of mature T cells is hence an uncommon event (Newrzela *et al*, 2008), if transformed, they behave highly aggressively and patients show poor response to therapy and short survival. The largest subgroup among PTCL consists of those 'not otherwise specified' (NOS) (Swerdlow *et al*, 2008). These lymphomas are characterized by a heterogeneous morphology. At present, PTCL NOS cannot be further subdivided in subgroups based on morphology or

bjh british journal of haematology

immunohistochemistry. Recently, gene expression profiling studies were performed to identify subgroups showing different clinical behaviour (Ballester *et al*, 2006; Martinez-Delgado, 2006; de Leval *et al*, 2007; Piccaluga *et al*, 2007). However, these results may be difficult to interpret because PTCL NOS frequently show an abundant reactive infiltrate.

Several studies evaluating genomic aberrations in PTCL have been conducted by the use of classical cytogenetic analysis as well as metaphase comparative genomic hybridization (CGH), showing complex karyotypes in most cases (Schlegelberger *et al*, 1994; Lepretre *et al*, 2000; Renedo *et al*, 2001; Melendez *et al*, 2004; Zettl *et al*, 2004). Translocations affecting the T cell receptor gene loci were rarely observed (Lepretre *et al*, 2000; Gesk *et al*, 2003). Most frequent gains that were recurrent in more than one study were located on chromosomes 7q32-ter (Schlegelberger *et al*, 1994; Lepretre *et al*, 2000; Renedo *et al*, 2001; Melendez *et al*, 2004; Zettl *et al*, 2004), 19 (Lepretre *et al*, 2000; Renedo *et al*, 2001; Melendez *et al*, 2004), 17cen-q21 (Renedo *et al*, 2001; Zettl *et al*, 2004), 11q13 (Renedo *et al*, 2001; Zettl *et al*, 2004) and 9q34 (Renedo *et al*, 2001; Zettl *et al*, 2004). Losses were most frequently detected on chromosomes 6q (Schlegelberger *et al*, 1994; Lepretre *et al*, 2000; Melendez *et al*, 2004; Zettl *et al*, 2004) and 13q (Schlegelberger *et al*, 1994; Lepretre *et al*, 2000; Renedo *et al*, 2001; Melendez *et al*, 2004; Zettl *et al*, 2004). Both of these deletions have been reported in other lymphoma entities (Zhang *et al*, 1997; Hartmann *et al*, 2008; Schmitz *et al*, 2009) and may be associated with lymphoma progression. Recently, two BAC array CGH studies with 2460 and 2304 BAC probes were carried out on 20 and 51 PTCL NOS respectively (Thorns *et al*, 2007; Nakagawa *et al*, 2009). Commonly altered regions in both studies included gains of the chromosome region 4q as well as losses of 9p21, 13q14 and 13q21–q33.

Single nucleotide polymorphism (SNP) arrays were originally designed for genotyping, but they are currently also used for screening of copy number aberrations (Nieländer *et al*, 2007; Vater *et al*, 2009). In addition to copy number data they also provide information about the genotype and regions with loss of heterozygosity (LOH) or partial uniparental disomy (pUPD) (Matsuzaki *et al*, 2004; Vater *et al*, 2009). Recently, a GeneChip 50k SNP array study of 33 PTCL NOS and 40 angioimmunoblastic T cell lymphomas identified commonly gained regions on chromosomes 8q24, 9p23 and 19q13 as well as losses of chromosomes 9p21 and 3q26 (Fujiwara *et al*, 2008). In the present study, we analysed 47 PTCL NOS with a higher resolution 250k GeneChip SNP array, which offers a theoretical median inter-SNP-distance of 5·1 kb.

Materials and methods

Patient samples

Fresh frozen samples of 49 patients with PTCL NOS, diagnosed according to the World Health Organization criteria (Swerdlow *et al*, 2008) were included in this study. All cases were CD3-positive and 21 cases showed scattered CD30-positive cells. Angioimmunoblastic T cell lymphomas and anaplastic large cell lymphomas were excluded from the study [CD3, CD4, CD10, CD23 (Novocastra, Newcastle upon Tyne, UK), CD8, CD20, CD21, ALK1 and CD30 (DAKO, Glostrup, Denmark)]. Overall survival (OS) data were available for 16 cases. One of 26 cases tested was positive for Epstein–Barr virus encoded small RNA (EBER). Reference samples were obtained from 10 healthy male and female donors each. Four reference samples were of consanguineous origin. Informed consent was obtained in accordance with the Declaration of Helsinki and approval of the ethics committee of the University Hospital of Frankfurt as well as from the Institutional Review Boards of the institutions providing cases with clinical data was obtained.

GeneChip Mapping 250k SNP arrays

DNA of 49 PTCL NOS was extracted from frozen tissue using the QIAamp DNA Mini kit (Qiagen, Hilden, Germany) according to the manufacturers instructions. Two hundred and fifty nanogram DNA per case underwent a StyI restriction enzyme digest, adaptor-ligation, polymerase chain reaction (PCR)-amplification, clean-up and hybridization onto Affymetrix GeneChip Mapping 250k StyI-SNP-arrays according to the Affymetrix GeneChip Mapping 500k Assay Manual. Arrays with call rates <90% were excluded from further analyses. GeneChip SNP array data are available at http://www.ncbi.nlm.nih.gov/geo/ (accession number GSE15842).

Statistical analysis of GeneChip Mapping 250k SNP arrays

Thirty-nine Hapmap samples (NA10851, NA10855, NA10863, NA11831, NA11832, NA12056, NA12057, NA12234, NA12264, NA12707, NA12716, NA12717, NA12801, NA12812, NA12813, NA18503, NA18504, NA18505, NA18506, NA18507, NA18508, NA18515, NA18516, NA18517, NA18532, NA18545, NA18558, NA18605, NA18612, NA18959, NA18967, NA18969, NA18997, NA19137, NA19138, NA19139, NA19152, NA19153, NA19154, provided by Affymetrix, http://www.affymetrix.com/support/technical/sample_data/500k_data.affx) and 20 laboratory-specific samples were used as normal reference arrays. Copy number analysis was performed using the CNAG (Copy Number Analyzer for Genechip) programme v2.0 (Nannya *et al*, 2005). CNAG was configured to select an optimal reference set individually for each array (Nannya *et al*, 2005). Gender-specific reference sets were used. Segmentation of raw copy number data was performed using the Hidden Markov Model (HMM) approach provided by CNAG. HMM parameters were adjusted individually for each array due to differences in hybridization quality.

With regard to outliers and technical artefacts, HMM segments with aberrant copy number were considered as copy

65

number aberration only if they consisted of at least five consecutive SNPs. High level amplifications were defined as aberrations with HMM copy number ≥5, homozygous deletions as aberrations with copy number = 0. For the male chromosome X (except of pseudoautosomal region 1), interpretation of log2-ratios had to be adjusted with respect to the gender specific (single copy) reference. Therefore segments with estimated HMM copy number of 1–3 copies were assumed as normal (copy number n = 1). Segments with estimated copy number ≥4 were selected as copy number gains and segments with estimated copy number 0 as losses. Pseudoautosomal region 1 was treated in the same way as the autosomal regions. Pseudoautosomal region 2 was not covered by the 250k GeneChip SNP array. Copy number variations/polymorphisms (CNVs) were identified by the use of web databases (http://projects.tcag.ca/variation/ and http://genome.ucsc.edu/cgi-bin/hgGateway) and excluded from further analyses.

For genotyping analysis the BRLMM algorithm was used with default parameters (*score threshold* = 0·5, *prior size* = 10 000 and *DM threshold* = 0·17) to genotype PTCL samples in combination with the reference arrays (http://www.affymetrix.com/support/technical/whitepapers/brlmm_whitepaper.pdf). A HMM-based method (Beroukhim *et al*, 2006) implemented in the dChip programme (Lin *et al*, 2004), (build date: April 11 2007) was used to infer regions with LOH from tumour samples. The HMM considering haplotype method (Beroukhim *et al*, 2006) was selected for the LOH calculations to account for linkage disequilibrium induced SNP dependencies. The LOH call threshold was set to the value of 0·99. An empirical haplotype correction was applied (Beroukhim *et al*, 2006). Thus putative LOH regions were excluded if there was a 95% concordance of the homozygous genotypes of the candidate LOH region with respective regions of more than 5% of the reference samples. LOH regions were designated pUPDs if no copy number aberrations were present in the region. In LOH regions, partially affected by copy number aberrations subregions without copy number aberrations were classified as pUPD if they comprised at least 50 neighbouring SNPs.

T cell receptor β (TRB@) rearrangement analysis

T cell receptor β rearrangement analysis was conducted with DNA from the frozen tissue blocks using the commercially available BIOMED2 primers (van Dongen *et al*, 2003) (IdentiClone^TM TRB@ Gene Clonality Assay, #9-205-0021; InVivoscribe Technologies, La Ciotat, France). The length of the fluorescence-labelled PCR products was determined on a DNA sequencer (ABI 3100; Applied Biosystems, Darmstadt, Germany) and analysed by the GeneScan Software 3.1 (Applied Biosystems). Peak ratios were determined between the height of the clonal peak and the average of the two adjacent peaks (1–4 bp smaller and larger respectively) as described previously (Luo *et al*, 2001; Geissinger *et al*, 2005).

Fluorescence in situ hybridization (FISH)

Fluorescence in situ hybridization was performed according to routine protocols (Martin-Subero *et al*, 2006a,b; Ventura *et al*, 2006). For confirmation purposes, cryosections of three and five PTCL NOS cases investigated by GeneChip SNP arrays were analysed by FISH using probes for the *REL* locus (2p16·1) and a region on chromosome 10p11, respectively. Twenty further formalin-fixed paraffin embedded PTCL NOS cases on tissue microarray format were additionally studied by FISH for both chromosomal regions. A *REL* break apart probe was designed using the BAC clones RP11-373L24, covering the *REL* locus and extending to centromere, and RP11-498O5 mapping telomeric to *REL* (Hartmann *et al*, 2008). Recently established as well as commercially available FISH probes for *TRA@/TRD@*, *TRB@* and *TRG@* (Abbott/Vysis, Abbott Park, Illinois, USA) were applied. For FISH of 10p11 the differentially labelled BAC clones RP11-195O1 and RP11-476F14 were used.

Immunohistochemistry

Twenty-seven formalin-fixed paraffin embedded PTCL NOS cases were stained for REL protein either on whole tissue sections or tissue microarray format using a commercially available, previously published (Rodig *et al*, 2007) polyclonal rabbit anti-REL antibody (PC139 1:200; Calbiochem, Merck Chemicals, Nottingham, UK). To detect antibody binding, the DAKO Envision System (DAKO) was used. Images were acquired using a Nikon Eclipse E1000 microscope and Nikon digital camera DXM1200 (Nikon, Düsseldorf, Germany).

Results

250k SNP array analyses

To identify chromosomal imbalances at a genome-wide level in 49 PTCL NOS, DNAs from frozen tissue were hybridized onto Affymetrix GeneChip Mapping 250k SNP arrays. Two arrays showed call rates <90% and were therefore excluded from further analyses. Array call rates of the remaining 47 cases ranged from 90·3% to 99·4% with a median of 97·0%. The 20 laboratory-specific reference arrays showed a minimum call rate of 91·2%, maximum call rate of 98·4% and a median of 96·8%. Following the criteria described in Materials and methods, in 25 PTCL NOS cases no genomic imbalances were detected apart from known CNVs and deletions in the T cell receptor gene loci (Fig 1A), which were detectable in 17/25 cases and most likely represent physiological *TCR* gene rearrangements in the monoclonal T cell populations. The remaining 22 PTCL NOS showed a mean of 15·9 gains (median 13·5), 12·7 losses (median 8·5) and 6·5 pUPDs (median 2·0) per case (Table SI). A tendency was observed for inferior survival in patients with genomic imbalances. However, this was not statistically significant (Fig 2A).
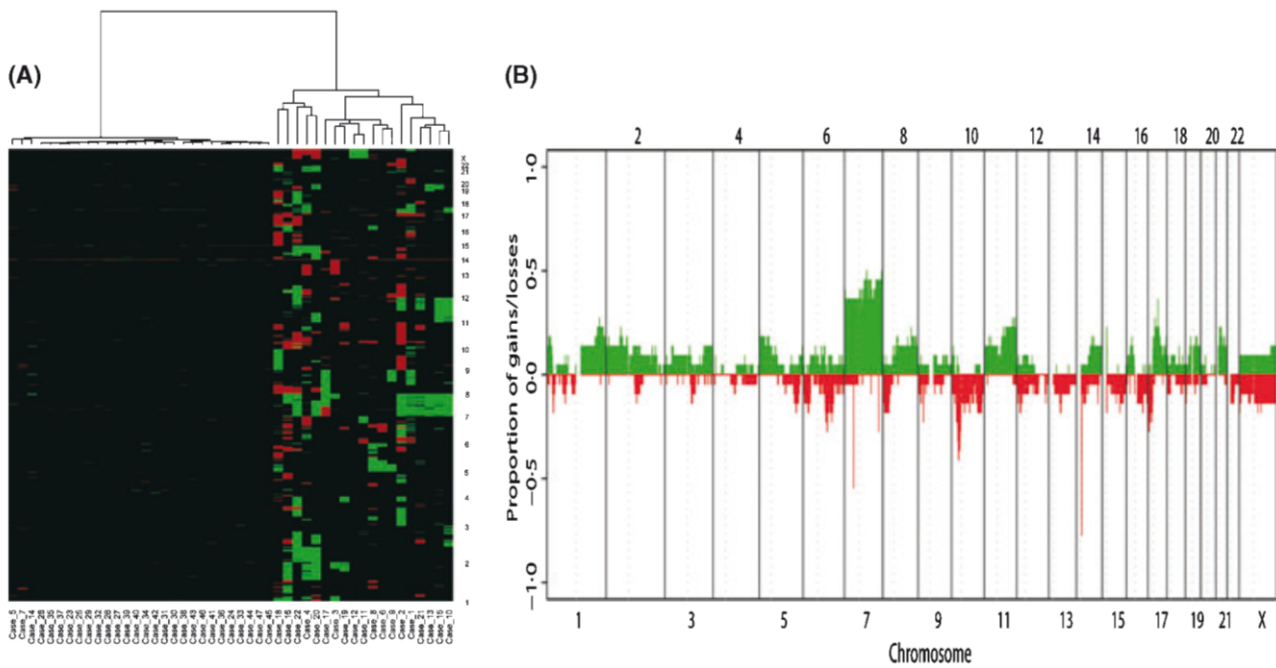
66

Fig 1. (A) Unsupervised hierarchical clustering of 47 PTCL NOS by genomic imbalances. Copy numbers are colour coded: gains green, losses red. (B) Profile of genomic gains and losses. Proportion of gains and losses in the 22 PTCL presenting a complex genomic aberration pattern, highlighted in green and red, respectively.

TRB@ rearrangement analysis

T cell receptor β rearrangement analysis was performed in 46/47 cases with successful SNP array hybridization. To estimate the amount of the clonal T cell population in relation to the polyclonal T cell background infiltrate, *TRB@* rearrangement patterns were analysed by visual inspection for the height of the clonal peak and the peaks of the reactive T cell background. A ratio between the height of the clonal peak and

the two adjacent peaks was determined, as published previously (Luo *et al*, 2001; Geissinger *et al*, 2005). In 10 of 15 PTCL with a heavy polyclonal background infiltrate or lack of a detectable clonal *TRB@* rearrangement no genomic imbalances were detected by SNP Chip analysis, probably because of the too high polyclonal T cell background (Table SII). A significant non-tumour cell infiltrate in cases without detection of chromosomal imbalances most probably also explains why biallelic T cell receptor gene rearrangements
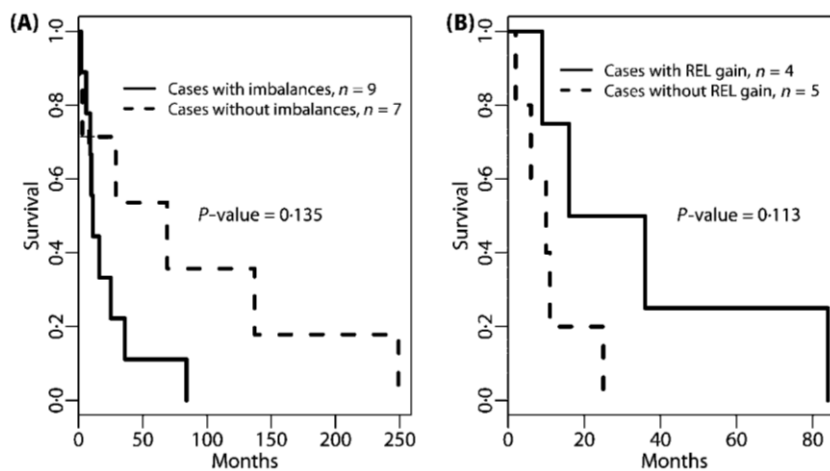
Fig 2. Overall survival in cases with available clinical data. (A) The survival of 16 patients of the two groups with and without detectable imbalances was compared by Kaplan–Meier analysis, with the *P*-value calculated by the Logrank test. (B) The survival of patients presenting gains of the *REL* locus was compared to patients without aberrations of the *REL* locus (all nine patient samples showing genomic imbalances) (Kaplan–Meier analysis, with the *P*-value calculated by the Logrank test).

(i.e. homozygous deletions of the T cell receptor gene loci in SNP array data) were detected significantly less frequent in these cases as compared to cases with genomic imbalances (not shown). Cases without genomic imbalances presented instead more frequently heterozygous deletions of T cell receptor gene loci. Interestingly, there were eight cases lacking genomic imbalances in SNP array data, which were also lacking a reactive T cell background infiltrate in *TRB@* rearrangement analysis and four of these cases presenting B cell contents <30% in CD20 immunostaining, suggesting the existence of a subgroup of PTCL NOS without chromosomal imbalances.

Detection of genomic imbalances in PTCL NOS by 250k SNP arrays

Two PTCL showed a total of five high level amplifications (copy number ≥5). One of these cases displayed four amplifications, mapping to chromosome 1p12–q21, 12p11–12, 12q12 and 12q12–13 (Table I), whereas the other case presented a single amplification of 2p15–16. Genomic gains in additional PTCL cases, overlapping with the genomic amplifications, were detected in one case each for the regions on chromosome 12, while four additional cases showed gains overlapping with the 491 kb amplified region on chromosome 2p15–16 (60 974 940–61 465 940 bp, see below). Gains (copy number ≥3) occurred most frequently on chromosome 7. According to statistical analysis, 65 regions on chromosome 7 were discontinuously gained in 7–11 PTCL cases, suggesting trisomy of chromosome 7 in seven cases (Fig 1). Further frequently gained regions (≥5 cases) were located on chromosome 1q32–43, 2p15–16, 8q24, 11q14–25, 17q11–21 and 21q11–21 (Table SIII).

Homozygous deletions were detected in *TRA@*, *TRB@* and *TRG@* in 12, 2 and 12 cases respectively, most likely reflecting biallelic T cell receptor gene rearrangements. Eleven further homozygous deletions were identified in five PTCL, two of them overlapping at chromosome 9p21 (Table II). Three

additional cases showed overlapping heterozygous deletions at chromosome 9p21, covering *CDKN2A*, *CDKN2B* and *MTAP*. Other homozygous deletions were located on chromosome 1p21, 5q21–22, 6q21, 6q23–24, 10p11, 14q32, Xp11 and Xq21 (one case each, Table II). Heterozygous deletions were observed most frequently on chromosome 10p11–12 (several small regions, 4–9 cases, Table SIV), 9p21, 17p11 and 17p13 (five cases each), 1p35–36, 5q33, 6p22, 6q16, 6q21–22, 8p21–23, 10q11–22, 10q25–26, 13q14, 15q24, 16q22, 16q24 and Xp22 (four cases each). Large pUPDs were mostly detected in cases with detectable genomic imbalances, most likely due to a strong reactive background infiltrate in the cases without detectable genomic aberrations (Yamamoto *et al*, 2007). pUPDs occurred most frequently on chromosome 1p21–22 and 9p (three cases each, Fig S1, Table SV). For the frequently lost region 1p35–36 two cases presented overlapping pUPD, for 5q33, 9p21, 10q11–21, 13q14 Xp11 and Xp21 one case each was detected with an overlapping pUPD (Tables II and SV).

Deletions of chromosome 10p11

Deletions other than those of the T cell receptor gene loci were most frequently detected on chromosome 10p. The homozygous deletion on chromosome 10p11 (linear position 31 643 666–33 014 706 bp) in one case was overlapping with heterozygous deletions in six further PTCL. FISH analysis was performed on cryosections of five of these cases available and deletions of 10p11 were confirmed in three of five cases, including the case with the biallelic loss. Twenty further PTCL NOS cases were investigated for deletions of 10p11 by FISH and a heterozygous deletion was detected in one additional case.

The affected region on chromosome 10p11 contains the five genes *ZEB1* (previously *TCF8*), *ARHGAP12*, *KIF5B*, *EPC1* and *CCDC7*. Given that this genomic region has been described to be deleted in human T-cell lymphotropic virus type 1

Table I. Amplifications detected by GeneChip Mapping 250k SNP array in 47 PTCL NOS.

Chromosome band	Sample with amplification (and samples with overlapping gains)	Startbase	Stopbase	Genes
1p12–q21	16	120 003 522	142 545 761	*HMGCS2, REG4, ADAM30, NOTCH2, PPIAL4A, PDE4DIP*
2p15–16	8 (4, 16, 20, 22)	60 974 940	61 465 940	**REL**, *PEX13, AHSA2*
12p11–12	16 (22)	26 009 784	32 724 615	*BHLHE41, SSPN, ITPR2, FGFR1OP2, TM7SF3, MED21, STK38L, ARNTL2, PPFIBP1, MRPS35, KLHDC5, ERGIC2, FAR2, TMTC1, PTHLH, CAPRIN2, DDX11, DENND5B, C12orf72, FAM60A, IPO8, BICD1, FGD4*
12q12	16 (22)	38 530 268	40 491 596	*SLC2A13, CNTN1, LRRK2, PDZRN4*
12q12–13	16 (22)	41 453 448	46 691 389	*IRAK4, TWF1, NELL2, DBX2, ARID5, SFRS2IP, PLEKHA9, SLC38A2, AMIGO2, P11, RAPGEF3, HDAC7, VDR, COL2A1*

Amplifications were defined as aberrations with copy numbers ≥5. Up to four additional cases showed gains overlapping with the amplified region (in brackets). Startbase and Stopbase derive from the first and last SNP affected by the amplification. In affected regions no miRNAs or pUPDs were found. (Genes, miRNAs and linear positions according to UCSC Genome Browser on Human – May 2004 Assembly.)

68

Table II. Homozygous deletions detected by GeneChip Mapping 250k SNP array in 47 PTCL NOS.

Chromosome band	Sample identity of case(s) with homozygous deletion	Number of additional cases with heterozygous deletions	Number of additional cases with pUPD	Startbase	Stopbase	Startbase extended	Stopbase extended	Genes	miRNAs
1p21	Case 17	–	2	104 102 295	105 338 839	103 835 683	105 371 094	Amylases family	–
5q21–22	Case 8	1	–	107 522 936	110 083 019	107 454 799	110 093 410	FBXL17, FER, MAN2A1, PJA2	–
6q21	Case 8	3	–	106 222 488	106 459 559	106 197 943	106 475 582	–	–
6q23–24	Case 2	1	–	138 642 529	142 254 496	138 589 964	142 286 201	HEBP2, REPS1, HECA, CITED2	–
7p14	Cases 1, 2, 3, 4, 6, 9, 15, 16, 20, 21, 22, 31	5	–	38 096 725	38 131 968	38 011 359	38 158 081	TRG@	–
7q34	Cases 9, 16	4	–	141 825 007	141 962 328	141 811 883	142 017 309	TRB@	–
9p21	Case 2	1	1	20 168 680	21 850 428	20 149 175	21 880 326	MLLT3, IFNA family, KLHL9, MTAP	miR-491 miR-31
9p21	Case 18	3	1	21 171 454	22 389 693	21 123 232	22 402 948	IFNA family, KLHL9, MTAP, CDKN2A, CDKN2B	miR-31
10p11	Case 8	6	–	31 643 666	33 014 706	31 605 627	33 070 321	ZEB1, ARHGAP12, KIF5B, EPC1, CCDC7	–
14q11	Cases 1, 2, 3, 6, 8, 9, 15, 16, 20, 21, 22, 31	22	–	21 929 710	22 012 459	21 840 943	22 016 220	TRA@	–
14q32	Case 8	2	–	88 843 295	88 906 978	88 832 561	88 915 440	FOXN3	–
Xp11	Case 8	2	1	41 354 957	46 225 034	41 329 820	46 300 801	CASK, MAOA, MAOB, NDP, EFHC2, FUNDC1, DUSP21, KDM6A, ZNF673, CHST7, SLC9A7	miR-221 miR-222
Xq21	Case 6	3	1	77 833 694	78 178 717	77 590 583	78 215 792	ZCCHC5, LPAR4, P2RY10	–
Xq21	Case 17	3	–	91 283 448	91 382 791	91 259 041	91 757 761	PCDH11X	–

Homozygous deletions were defined as aberrations with copy number = 0 after smoothing by the HMM approach. Startbase and Stopbase derive from the first and last SNP affected by the homozygous deletion. Startbase and Stopbase extended include adjacent chromosome regions, which are not covered by a SNP probe on the array. (Genes, miRNAs and linear positions according to UCSC Genome Browser on Human – May 2004 Assembly.)

69

(HTLV-1)-associated adult T cell leukaemias/lymphomas (ATLs) (Hidaka *et al*, 2008), a PCR-based diagnostic test for HTLV-1 proviral load was performed and HTLV-1 infection of PTCL NOS presenting deletions at chromosome 10p11 could be excluded. As *ZEB1*, which codes for zinc finger E-box binding homeobox 1, has been shown to be mutated in few ATLs (Hidaka *et al*, 2008) and *EPC1* shows reduced gene expression in PTCL (Piccaluga *et al*, 2007), the coding exons of these two potential target genes were sequenced in the six PTCL cases displaying heterozygous deletions of this region, but no mutations were detected. mRNA expression levels of *ZEB1* and *EPC1* were not significantly reduced in five PTCL showing a loss of 10p11 compared to 10 PTCL with balanced copy number of 10p11 and CD4$^+$ T-cells from peripheral blood. This may potentially be due to high expression of the genes in the reactive background infiltrate. Likewise, no methylation of *TCF8* or *EPC1* promoter sites was detected in PCR fragments amplified from bisulphite treated DNA (data not shown). No tumour suppressive function has yet been assigned to the other three genes *CCDC7*, *KIF5B* and *ARHGAP12* mapping to the deleted region nor did they show reduced gene expression in PTCL NOS (Piccaluga *et al*, 2007). Therefore they were not further investigated.

Gains of chromosome 2p15–16 covering REL and immunohistochemical detection of REL protein expression

In the GeneChip SNP array data, five cases showed gains of chromosome 2p15–16, with one amplification among these gains. Three of these cases were investigated by FISH and gains could be confirmed in each of the cases using probes flanking the *REL* locus. In addition, two of these cases showed a split signal indicating a chromosomal breakpoint at the *REL* locus. These two cases were further investigated by FISH with fusion probes for *REL* and *TRA@/TRD@*, *TRB@* and *TRG@*. However, no juxtaposition of *REL* next to one of the T cell receptor gene loci was detected. Twenty additional PTCL NOS cases were investigated by FISH for genomic changes of the *REL* locus. Three of 18 evaluable cases showed gains of *REL* (3–4 signals), but no additional chromosomal breakpoint at the *REL* locus was observed.

Immunohistochemical stainings for REL were performed for 27 cases of PTCL NOS, including five cases with genomic gains of 2p15–16. Five PTCL cases displayed nuclear positivity, among these two cases with gains of the *REL* locus. The case with the genomic amplification of *REL* in SNP array data and an additional breakpoint at the *REL* locus, presented the strongest immunohistochemical reaction of all cases, which was, however, mainly confined to the cytoplasm of the tumour cells (Fig 3). Reactive tonsils and Hodgkin lymphoma lymph nodes were stained as controls. As expected, cytoplasmatic REL positivity in tonsillar germinal centre B cells as well as strong cytoplasmatic and nuclear positivity of Hodgkin- and Reed-Sternberg cells of classical Hodgkin lymphoma was seen (Barth *et al*, 2003).

Discussion

In the present study, 250k SNP arrays were applied to 49 PTCL NOS. Forty-seven cases were suitable for further analysis

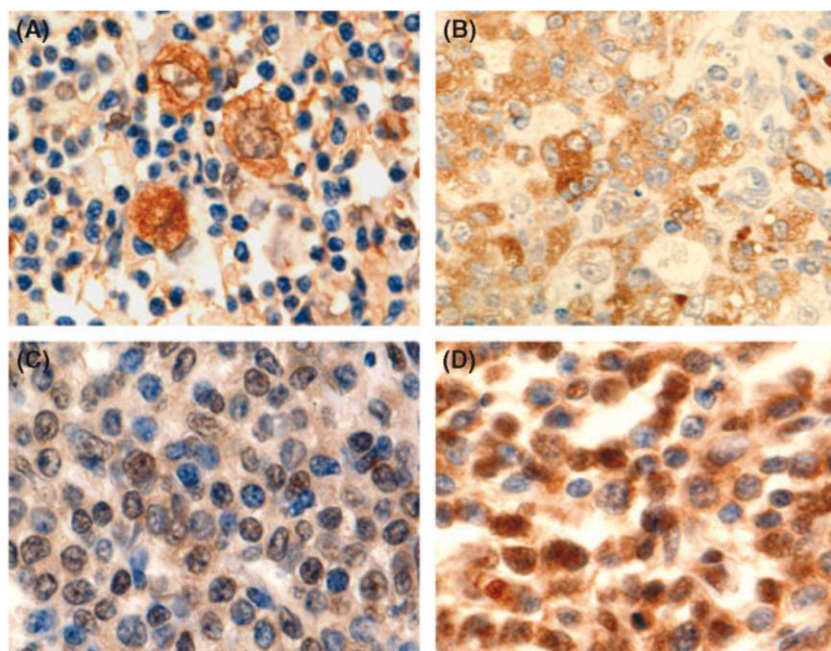**Fig 3.** REL immunostaining. (A) Classical Hodgkin lymphoma, positive control for nuclear and cytoplasmatic staining. (B) PTCL NOS with genomic amplification and breakpoint at the *REL* locus and cytoplasmatic staining pattern in the tumour cells. (C) PTCL NOS with nuclear staining pattern and gain of the *REL* locus. (D) PTCL NOS with nuclear staining pattern and balanced copy number of 2p15–16. 400× Original magnification.

70

according to quality control parameters. Gains of chromosome 2p15–16 detected by SNP array data in five cases could be confirmed in all three cases available for FISH analysis. Losses of 10p11 could be confirmed in three of five cases available for FISH analysis. These data demonstrate the validity of the array results, particularly concerning the detection of gains. Heterozygous losses are difficult to confirm by FISH in tissue sections of tumours with a reactive background infiltrate and therefore GeneChip SNP arrays may provide a more sensitive platform for the detection of losses. This may also apply to the detection of further heterozygous deletions of 10p11 by FISH in tissue sections.

In 25 PTCL NOS no chromosomal imbalances apart from CNVs were detected, whereas 22 PTCL NOS presented a complex aberration pattern. A similar fraction of PTCL NOS lacking detectable genomic imbalances was also observed in other studies in 11/27 cases (Renedo *et al*, 2001), in 57/73 cases (Fujiwara *et al*, 2008) and 22/51 cases (Nakagawa *et al*, 2009) and it was reported that these cases have a better overall survival (Fujiwara *et al*, 2008; Nakagawa *et al*, 2009). In the *TRB@* rearrangement analysis (Luo *et al*, 2001), cases with complex copy number aberrations displayed a lower reactive background infiltrate as the cases without detectable aberrations, in agreement with the study by Nakagawa *et al* (2009). Therefore, lack of detectable chromosomal imbalances was, in most cases, probably the result of a strong background infiltrate. However, eight cases without detectable copy number aberrations displayed a monoclonal rearrangement pattern without T cell background infiltrate, which would explain the lack of detectable copy number aberrations. Thus, it is likely that the malignant clones in these cases indeed lacked clonal genomic imbalances detectable at the resolution of our SNP-chip analysis. PTCL NOS with a lack of copy number aberrations and a balanced translocation as sole abnormality were previously reported by (Nelson *et al*, 2008). In conclusion, there is very likely a fraction of PTCL that has a mainly balanced copy number status. It remains to be determined whether such cases define a distinct subgroup of PTCL NOS in terms of pathogenesis and clinical behaviour.

In the present study, gains of chromosome 7 were the most frequent aberration observed, in line with previous studies (Renedo *et al*, 2001; Melendez *et al*, 2004; Zettl *et al*, 2004; Nelson *et al*, 2008; Nakagawa *et al*, 2009). Eleven of the 22 cases with aberrations presented gains of the region 7q21, containing *CDK6*, which has been shown to be one target gene on chromosome 7 in PTCL (Nagel *et al*, 2008). Losses of chromosome 9p21 were observed in five cases, including two homozygous deletions, and one additional case with an pUPD. Losses of 9p21, covering *CDKN2A* and *CDKN2B*, were frequently observed in previous array CGH studies of PTCL NOS (Thorns *et al*, 2007; Nakagawa *et al*, 2009) and seven of these deletions were homozygous. In addition to gains of 7q21, covering *CDK6*, and losses of 9p21, including *CDKN2A* and *CDKN2B*, other cell cycle genes mapped to frequently deleted regions: deletions of 13q14 (47 447 838–50 689 816 bp)

including *RB1* in four PTCL and deletions of 17p13 (6888–8 195 094 bp) covering *TP53* in five PTCL. The finding of recurrent genomic aberrations affecting several factors involved in cell cycle control in a total of 15 PTCL suggests an important role of cell cycle genes in PTCL NOS.

Gains of 8q24 were previously reported in some PTCL (Zettl *et al*, 2004; Thorns *et al*, 2007). Interestingly, we observed a 163 kb minimal common gained region on chromosome 8q24 (117 750 212–117 913 160 bp, five cases), containing *EIF3H*, whereas *MYC* (128 817 686–128 822 853 bp) was gained in only two of the five cases. *EIF3H* shows frequently genomic amplifications in prostate cancer and was shown to promote cell growth *in vitro* (Saramaki *et al*, 2001; Savinainen *et al*, 2006). Closely located to the common gained region on chromosome 8q24 in our study, a different region (118 306 152–118 306 326 bp) was observed to be frequently gained in PTCL NOS (11/33 cases) (Fujiwara *et al*, 2008), which was also gained in four of the five cases with gains on 8q24 in the present data set. Therefore, in PTCL NOS the target gene of gains affecting chromosome 8q24 may be different from *MYC*.

Deletions were most frequently detected on chromosome 10p11–12. The 1·97 Mb deletion on chromosome 10p11 covering the five genes *ZEB1*, *ARHGAP12*, *KIF5B*, *EPC1* and *CCDC7* occurred in seven PTCL, in one of these as a homozygous deletion. Deletions on chromosome 10p11–12 have been reported in metaphase CGH as well as array CGH studies in approximately 15% of PTCL NOS (Zettl *et al*, 2004; Nakagawa *et al*, 2009). In other T cell neoplasms, including Sézary syndrome, T – prolymphocytic leukaemia as well as ATLs, deletions of this particular region have also been described (Dürig *et al*, 2007; Hidaka *et al*, 2008; Vermeer *et al*, 2008). Therefore, in different subsets of mature T cell neoplasms similar mechanisms of transformation may be present. *EPC1*, enhancer of polycomb 1, one of the genes located in this region and displaying histone acetyltransferase activity (Attwooll *et al*, 2005), showed reduced mRNA expression in PTCL NOS compared to reactive T cell subsets (Piccaluga *et al*, 2007). mRNA expression of *ZEB1*, also mapping to this region, was downregulated in samples of leukaemic ATL isolated from peripheral blood compared to reactive CD4[+] T cell subsets (Hidaka *et al*, 2008). In the same study, missense mutations of *ZEB1* were detected in four of 44 ATL samples and cell lines. However, as mutations or reduced mRNA expression of *EPC1* or *ZEB1* were not found in the present study, their role in PTCL NOS remains unclear.

Gains of chromosome 2p15–16 (60 974 940–61 465 940 bp) including *REL* were detected in five PTCL, one of these cases displaying a genomic amplification. Genomic gains of the *REL* locus have – to our knowledge – so far not been reported in PTCL NOS, possibly due to a lower coverage of array platforms formerly applied. *REL* is one of the five nuclear factor (NF)-κB family members, which form various homo- and heterodimers and which translocate to the nucleus upon activation (Jost & Ruland, 2007). Gains of 2p15–16 could be confirmed in three

71

cases available for FISH and two of these cases showed, in addition to the gain of 2p15–16, a breakpoint at the *REL* locus. In an independent series of 18 PTCL NOS, three further cases presented gains of 2p15–16. Immunohistochemistry revealed nuclear expression of REL protein in five of 27 PTCL NOS, including two cases with gains of 2p15–16. Therefore, a small subgroup of PTCL NOS is characterized by nuclear REL expression, which is, in some cases, related to genomic aberrations of the *REL* locus. This should be studied on a larger series of cases – possibly in conjunction with the determination of other key markers of both the canonical and alternative NF-κB pathway. We described previously one PTCL NOS case presenting a translocation of *REL* to *TRA@/TRD@* (Gesk *et al*, 2003). In the current series none of the two PTCL NOS with *REL* breakpoints showed a translocation of *REL* to one of the *TCR* loci. This is in line with the finding that translocations of oncogenes to one of the *TCR* loci rarely occur in T cell lymphomas (Gesk *et al*, 2003; Leich *et al*, 2007). Genomic gains of the *REL* locus and nuclear expression of REL protein have been described in Hodgkin lymphoma (Joos *et al*, 2002; Barth *et al*, 2003) and primary mediastinal B cell lymphoma (Rodig *et al*, 2007; Weniger *et al*, 2007). NF-κB activity was reported in a subgroup of PTCL NOS in a gene expression study and was defined as an indicator of favourable prognosis (Martinez-Delgado *et al*, 2005). We observed a tendency for superior survival in patients with gains of *REL*. However, this did not reach statistical significance (Fig 2B). This finding needs further investigations in a larger cohort under a controlled clinical setting. In the present study, three cases each also displayed heterozygous deletions of 14q12–21 and 16q11–21 containing the two NF-κB signalling factors *NFKBIA* and *CYLD*. Overall, a total of 10 PTCL showed genomic imbalances affecting members of the NF-κB pathway, of which nine cases also displayed chromosomal aberrations containing cell cycle genes.

In conclusion, genomic aberrations detected in the present study are largely in line with previous data and aberrant genomic regions could be delineated to several candidate genes. Gains and translocations of *REL* as well as expression of REL protein were observed in a subgroup of PTCL. Several of the candidate genes identified are involved in cell cycle control or NF-κB signalling, which suggests that imbalances of both may be important survival factors in subgroups of PTCL NOS.

Acknowledgements

Authors' contributions

SH: Hybridization of GeneChip arrays, analysis and interpretation of array data, immunohistochemistry, drafting the manuscript; SG, SB: FISH assay design and analysis, interpretation of data; RSch: hybridization of GeneChip arrays; MK, IV, CD: development of evaluation algorithms, statistical analysis of data; SC, MP, JPM, AK, AP, PPP, SP, GH: pathological evaluation, supplied essential material; RK, RS: experimental design, analysis and interpretation of data, revising the manuscript; MLH: pathological evaluation, experimental design, revising the manuscript.

Conflict of interest

The authors report no potential conflicts of interest.

References

Attwooll, C., Oddi, S., Cartwright, P., Prosperini, E., Agger, K., Steensgaard, P., Wagener, C., Sardet, C., Moroni, M.C. & Helin, K. (2005) A novel repressive E2F6 complex containing the polycomb group protein, EPC1, that interacts with EZH2 in a proliferation-specific manner. *The Journal of Biological Chemistry*, **280**, 1199–1208.

Ballester, B., Ramuz, O., Gisselbrecht, C., Doucet, G., Loi, L., Loriod, B., Bertucci, F., Bouabdallah, R., Devilard, E., Carbuccia, N., Mozziconacci, M.J., Birnbaum, D., Brousset, P., Berger, F., Salles, G., Briere, J., Houlgatte, R., Gaulard, P. & Xerri, L. (2006) Gene expression profiling identifies molecular subgroups among nodal peripheral T-cell lymphomas. *Oncogene*, **25**, 1560–1570.

Barth, T.F., Martin-Subero, J.I., Joos, S., Menz, C.K., Hasel, C., Mechtersheimer, G., Parwaresch, R.M., Lichter, P., Siebert, R. & Mooller, P. (2003) Gains of 2p involving the REL locus correlate with nuclear c-Rel protein accumulation in neoplastic cells of classical Hodgkin lymphoma. *Blood*, **101**, 3681–3686.

Beroukhim, R., Lin, M., Park, Y., Hao, K., Zhao, X., Garraway, L.A., Fox, E.A., Hochberg, E.P., Mellinghoff, I.K., Hofer, M.D., Descazeaud, A., Rubin, M.A., Meyerson, M., Wong, W.H., Sellers, W.R. & Li, C. (2006) Inferring loss-of-heterozygosity from unpaired tumors using high-density oligonucleotide SNP arrays. *PLoS Computational Biology*, **2**, e41.

van Dongen, J.J., Langerak, A.W., Bruggemann, M., Evans, P.A., Hummel, M., Lavender, F.L., Delabesse, E., Davi, F., Schuuring, E., Garcia-Sanz, R., van Krieken, J.H., Droese, J., Gonzalez, D., Bastard, C., White, H.E., Spaargaren, M., Gonzalez, M., Parreira, A., Smith, J.L., Morgan, G.J., Kneba, M. & Macintyre, E.A. (2003) Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 Concerted Action BMH4-CT98-3936. *Leukemia*, **17**, 2257–2317.

72

Dürig, J., Bug, S., Klein-Hitpass, L., Boes, T., Jons, T., Martin-Subero, J.I., Harder, L., Baudis, M., Duhrsen, U. & Siebert, R. (2007) Combined single nucleotide polymorphism-based genomic mapping and global gene expression profiling identifies novel chromosomal imbalances, mechanisms and candidate genes important in the pathogenesis of T-cell prolymphocytic leukemia with inv(14)(q11q32). *Leukemia*, **21**, 2153–2163.

Fujiwara, S.I., Yamashita, Y., Nakamura, N., Choi, Y.L., Ueno, T., Watanabe, H., Kurashina, K., Soda, M., Enomoto, M., Hatanaka, H., Takada, S., Abe, M., Ozawa, K. & Mano, H. (2008) High-resolution analysis of chromosome copy number alterations in angioimmunoblastic T-cell lymphoma and peripheral T-cell lymphoma, unspecified, with single nucleotide polymorphism-typing microarrays. *Leukemia*, **22**, 1891–1898.

Geissinger, E., Bonzheim, I., Krenacs, L., Roth, S., Strobel, P., Ott, G., Reimer, P., Wilhelm, M., Muller-Hermelink, H.K. & Rudiger, T. (2005) Identification of the tumor cells in peripheral T-cell lymphomas by combined polymerase chain reaction-based T-cell receptor beta spectrotyping and immunohistological detection with T-cell receptor beta chain variable region segment-specific antibodies. *The Journal of Molecular Diagnostics*, **7**, 455–464.

Gesk, S., Martin-Subero, J.I., Harder, L., Luhmann, B., Schlegelberger, B., Calasanz, M.J., Grote, W. & Siebert, R. (2003) Molecular cytogenetic detection of chromosomal breakpoints in T-cell receptor gene loci. *Leukemia*, **17**, 738–745.

Hartmann, S., Martin-Subero, J.I., Gesk, S., Husken, J., Giefing, M., Nagel, I., Riemke, J., Chott, A., Klapper, W., Parrens, M., Merlio, J.P., Kuppers, R., Brauninger, A., Siebert, R. & Hansmann, M.L. (2008) Detection of genomic imbalances in microdissected Hodgkin and Reed-Sternberg cells of classical Hodgkin's lymphoma by array-based comparative genomic hybridization. *Haematologica*, **93**, 1318–1326.

Hidaka, T., Nakahata, S., Hatakeyama, K., Hamasaki, M., Yamashita, K., Kohno, T., Arai, Y., Taki, T., Nishida, K., Okayama, A., Asada, Y., Yamaguchi, R., Tsubouchi, H., Yokota, J., Taniwaki, M., Higashi, Y. & Morishita, K. (2008) Down-regulation of TCF8 is involved in the leukemogenesis of adult T-cell leukemia/lymphoma. *Blood*, **112**, 383–393.

Joos, S., Menz, C.K., Wrobel, G., Siebert, R., Gesk, S., Ohl, S., Mechtersheimer, G., Trumper, L., Moller, P., Lichter, P. & Barth, T.F. (2002) Classical Hodgkin lymphoma is characterized by recurrent copy number gains of the short arm of chromosome 2. *Blood*, **99**, 1381–1387.

Jost, P.J. & Ruland, J. (2007) Aberrant NF-kappaB signaling in lymphoma: mechanisms, consequences, and therapeutic implications. *Blood*, **109**, 2700–2707.

Leich, E., Haralambieva, E., Zettl, A., Chott, A., Rudiger, T., Holler, S., Muller-Hermelink, H.K., Ott, G. & Rosenwald, A. (2007) Tissue microarray-based screening for chromosomal breakpoints affecting the T-cell receptor gene loci in mature T-cell lymphomas. *Journal of Pathology*, **213**, 99–105.

Lepretre, S., Buchonnet, G., Stamatoullas, A., Lenain, P., Duval, C., d'Anjou, J., Callat, M.P., Tilly, H. & Bastard, C. (2000) Chromosome abnormalities in peripheral T-cell lymphoma. *Cancer Genetics Cytogenetics*, **117**, 71–79.

de Leval, L., Rickman, D.S., Thielen, C., Reynies, A., Huang, Y.L., Delsol, G., Lamant, L., Leroy, K., Briere, J., Molina, T., Berger, F., Gisselbrecht, C., Xerri, L. & Gaulard, P. (2007) The gene expression profile of nodal peripheral T-cell lymphoma demonstrates a

molecular link between angioimmunoblastic T-cell lymphoma (AITL) and follicular helper T (TFH) cells. *Blood*, **109**, 4952–4963.

Lin, M., Wei, L.J., Sellers, W.R., Lieberfarb, M., Wong, W.H. & Li, C. (2004) dChipSNP: significance curve and clustering of SNP-array-based loss-of-heterozygosity data. *Bioinformatics*, **20**, 1233–1240.

Luo, V., Lessin, S.R., Wilson, R.B., Rennert, H., Tozer, C., Benoit, B. & Leonard, D.G. (2001) Detection of clonal T-cell receptor gamma gene rearrangements using fluorescent-based PCR and automated high-resolution capillary electrophoresis. *Molecular Diagnosis*, **6**, 169–179.

Martinez-Delgado, B. (2006) Peripheral T-cell lymphoma gene expression profiles. *Hematological Oncology*, **24**, 113–119.

Martinez-Delgado, B., Cuadros, M., Honrado, E., Ruiz de la Parte, A., Roncador, G., Alves, J., Castrillo, J.M., Rivas, C. & Benitez, J. (2005) Differential expression of NF-kappaB pathway genes among peripheral T-cell lymphomas. *Leukemia*, **19**, 2254–2263.

Martin-Subero, J.I., Klapper, W., Sotnikova, A., Callet-Bauchu, E., Harder, L., Bastard, C., Schmitz, R., Grohmann, S., Hoppner, J., Riemke, J., Barth, T.F., Berger, F., Bernd, H.W., Claviez, A., Gesk, S., Frank, G.A., Kaplanskaya, I.B., Moller, P., Parwaresch, R.M., Rudiger, T., Stein, H., Kuppers, R., Hansmann, M.L. & Siebert, R. (2006a) Chromosomal breakpoints affecting immunoglobulin loci are recurrent in Hodgkin and Reed-Sternberg cells of classical Hodgkin lymphoma. *Cancer Research*, **66**, 10332–10338.

Martin-Subero, J.I., Wlodarska, I., Bastard, C., Picquenot, J.M., Hoppner, J., Giefing, M., Klapper, W. & Siebert, R. (2006b) Chromosomal rearrangements involving the BCL3 locus are recurrent in classical Hodgkin and peripheral T-cell lymphoma. *Blood*, **108**, 401–402. Author reply 402–403.

Matsuzaki, H., Dong, S., Loi, H., Di, X., Liu, G., Hubbell, E., Law, J., Berntsen, T., Chadha, M., Hui, H., Yang, G., Kennedy, G.C., Webster, T.A., Cawley, S., Walsh, P.S., Jones, K.W., Fodor, S.P. & Mei, R. (2004) Genotyping over 100,000 SNPs on a pair of oligonucleotide arrays. *Nature Methods*, **1**, 109–111.

Melendez, B., Diaz-Uriarte, R., Cuadros, M., Martinez-Ramirez, A., Fernandez-Piqueras, J., Dopazo, A., Cigudosa, J.C., Rivas, C., Dopazo, J., Martinez-Delgado, B. & Benitez, J. (2004) Gene expression analysis of chromosomal regions with gain or loss of genetic material detected by comparative genomic hybridization. *Genes, Chromosomes and Cancer*, **41**, 353–365.

Nagel, S., Leich, E., Quentmeier, H., Meyer, C., Kaufmann, M., Drexler, H.G., Zettl, A., Rosenwald, A. & MacLeod, R.A. (2008) Amplification at 7q22 targets cyclin-dependent kinase 6 in T-cell lymphoma. *Leukemia*, **22**, 387–392.

Nakagawa, M., Nakagawa-Oshiro, A., Karnan, S., Tagawa, H., Utsunomiya, A., Nakamura, S., Takeuchi, I., Ohshima, K. & Seto, M. (2009) Array comparative genomic hybridization analysis of PTCL-U reveals a distinct subgroup with genetic alterations similar to lymphoma-type adult T-cell leukemia/lymphoma. *Clinical Cancer Research*, **15**, 30–38.

Nannya, Y., Sanada, M., Nakazaki, K., Hosoya, N., Wang, L., Hangaishi, A., Kurokawa, M., Chiba, S., Bailey, D.K., Kennedy, G.C. & Ogawa, S. (2005) A robust algorithm for copy number detection using high-density oligonucleotide single nucleotide polymorphism genotyping arrays. *Cancer Research*, **65**, 6071–6079.

Nelson, M., Horsman, D.E., Weisenburger, D.D., Gascoyne, R.D., Dave, B.J., Loberiza, F.R., Ludkovski, O., Savage, K.J., Armitage, J.O. & Sanger, W.G. (2008) Cytogenetic abnormalities and clinical correlations in peripheral T-cell lymphoma. *British Journal of Haematology*, **141**, 461–469.

73

Newrzela, S., Cornils, K., Li, Z., Baum, C., Brugman, M.H., Hartmann, M., Meyer, J., Hartmann, S., Hansmann, M.L., Fehse, B. & von Laer, D. (2008) Resistance of mature T cells to oncogene transformation. *Blood*, **112**, 2278–2286.

Nieländer, I., Bug, S., Richter, J., Giefing, M., Martin-Subero, J.I. & Siebert, R. (2007) Combining array-based approaches for the identification of candidate tumor suppressor loci in mature lymphoid neoplasms. *Acta Pathologica, Microbiologica et Immunulogica Scandinavica*, **115**, 1107–1134.

Piccaluga, P.P., Agostinelli, C., Califano, A., Rossi, M., Basso, K., Zupo, S., Went, P., Klein, U., Zinzani, P.L., Baccarani, M., Dalla Favera, R. & Pileri, S.A. (2007) Gene expression analysis of peripheral T cell lymphoma, unspecified, reveals distinct profiles and new potential therapeutic targets. *Journal of Clinical Investigation*, **117**, 823–834.

Renedo, M., Martinez-Delgado, B., Arranz, E., Garcia, M., Urioste, M., Martinez-Ramirez, A., Rivas, C., Cigudosa, J.C. & Benitez, I. (2001) Chromosomal changes pattern and gene amplification in T cell non-Hodgkin's lymphomas. *Leukemia*, **15**, 1627–1632.

Rodig, S.J., Savage, K.J., LaCasce, A.S., Weng, A.P., Harris, N.L., Shipp, M.A., Hsi, E.D., Gascoyne, R.D. & Kutok, J.L. (2007) Expression of TRAF1 and nuclear c-Rel distinguishes primary mediastinal large cell lymphoma from other types of diffuse large B-cell lymphoma. *American Journal of Surgical Pathology*, **31**, 106–112.

Rüdiger, T., Weisenburger, D.D., Anderson, J.R., Armitage, J.O., Diebold, J., MacLennan, K.A., Nathwani, B.N., Ullrich, F. & Muller-Hermelink, H.K. (2002) Peripheral T-cell lymphoma (excluding anaplastic large-cell lymphoma): results from the Non-Hodgkin's Lymphoma Classification Project. *Annals of Oncology*, **13**, 140–149.

Saramaki, O., Willi, N., Bratt, O., Gasser, T.C., Koivisto, P., Nupponen, N.N., Bubendorf, L. & Visakorpi, T. (2001) Amplification of EIF3S3 gene is associated with advanced stage in prostate cancer. *American Journal of Pathology*, **159**, 2089–2094.

Savainainen, K.J., Helenius, M.A., Lehtonen, H.J. & Visakorpi, T. (2006) Overexpression of EIF3S3 promotes cancer cell growth. *Prostate*, **66**, 1144–1150.

Schlegelberger, B., Himmler, A., Gödde, E., Grote, W., Feller, A.C. & Lennert, K. (1994) Cytogenetic findings in peripheral T-cell lymphomas as a basis for distinguishing low-grade and high-grade lymphomas. *Blood*, **83**, 505–511.

Schmitz, R., Hansmann, M.L., Bohle, V., Martin-Subero, J.I., Hartmann, S., Mechtersheimer, G., Klapper, W., Vater, I., Giefing, M., Gesk, S., Stanelle, J., Siebert, R. & Kuppers, R. (2009) TNFAIP3 (A20) is a tumor suppressor gene in Hodgkin lymphoma and primary mediastinal B cell lymphoma. *Journal of Experimental Medicine*, **206**, 981–989.

Swerdlow, S.H., International Agency for Research on Cancer & World Health Organization (2008) *WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues*. International Agency for Research on Cancer, Lyon, France.

Thorns, C., Bastian, B., Pinkel, D., Roydasgupta, R., Fridlyand, J., Merz, H., Krokowski, M., Bernd, H.W. & Feller, A.C. (2007) Chromosomal aberrations in angioimmunoblastic T-cell lymphoma and peripheral T-cell lymphoma unspecified: A matrix-based CGH approach. *Genes, Chromosomes and Cancer*, **46**, 37–44.

Vater, I., Wagner, F., Kreuz, M., Berger, H., Martin-Subero, J.I., Pott, C., Martinez-Climent, J.A., Klapper, W., Krause, K., Dyer, M.J.,

Gesk, S., Harder, L., Zamo, A., Dreyling, M., Hasenclever, D., Arnold, N. & Siebert, R. (2009) GeneChip analyses point to novel pathogenetic mechanisms in mantle cell lymphoma. *British Journal of Haematology*, **144**, 317–331.

Ventura, R.A., Martin-Subero, J.I., Jones, M., McParland, J., Gesk, S., Mason, D.Y. & Siebert, R. (2006) FISH analysis for the detection of lymphoma-associated chromosomal abnormalities in routine paraffin-embedded tissue. *The Journal of Molecular Diagnostics*, **8**, 141–151.

Vermeer, M.H., van Doorn, R., Dijkman, R., Mao, X., Whittaker, S., van Voorst Vader, P.C., Gerritsen, M.J., Geerts, M.L., Gellrich, S., Soderberg, O., Leuchowius, K.J., Landegren, U., Out-Luiting, J.J., Knijnenburg, J., Ijszenga, M., Szuhai, K., Willemze, R. & Tensen, C.P. (2008) Novel and highly recurrent chromosomal alterations in Sezary syndrome. *Cancer Research*, **68**, 2689–2698.

Weniger, M.A., Gesk, S., Ehrlich, S., Martin-Subero, J.I., Dyer, M.J., Siebert, R., Moller, P. & Barth, T.F. (2007) Gains of REL in primary mediastinal B-cell lymphoma coincide with nuclear accumulation of REL protein. *Genes, Chromosomes and Cancer*, **46**, 406–415.

Yamamoto, G., Nannya, Y., Kato, M., Sanada, M., Levine, R.L., Kawamata, N., Hangaishi, A., Kurokawa, M., Chiba, S., Gilliland, D.G., Koeffler, H.P. & Ogawa, S. (2007) Highly sensitive method for genomewide detection of allelic composition in nonpaired, primary tumor specimens by use of affymetrix single-nucleotide-polymorphism genotyping microarrays. *American Journal of Human Genetics*, **81**, 114–126.

Zettl, A., Rudiger, T., Konrad, M.A., Chott, A., Simonitsch-Klupp, I., Sonnen, R., Muller-Hermelink, H.K. & Ott, G. (2004) Genomic profiling of peripheral T-cell lymphoma, unspecified, and anaplastic large T-cell lymphoma delineates novel recurrent chromosomal alterations. *American Journal of Pathology*, **164**, 1837–1848.

Zhang, Y., Weber-Matthiesen, K., Siebert, R., Matthiesen, P. & Schlegelberger, B. (1997) Frequent deletions of 6q23-24 in B-cell non-Hodgkin's lymphomas detected by fluorescence in situ hybridization. *Genes, Chromosomes and Cancer*, **18**, 310–313.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Fig S1. pUPDs detected in 47 PTCL NOS by GeneChip Mapping 250k SNP array.

Table SI. Number of imbalances per case detected by GeneChip Mapping 250k SNP array.

Table SII. *TCRB* rearrangement patterns.

Table SIII. Gains other than chromosome 7 in 47 PTCL NOS detected by GeneChip 250k SNP array.

Table SIV. Heterozygous deletions detected by GeneChip 250k SNP array.

Table SV. pUPDs detected by GeneChip 250k SNP array.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

74

2.6 Publikation 6: „Detection of genomic aberrations in molecularly defined Burkitt's lymphoma by array-based, high resolution, single nucleotide polymorphism analysis" Haematologica. 2010 Dec;95(12):2047-55

Scholtysik R, Kreuz M, Klapper W, Burkhardt B, Feller AC, Hummel M, Loeffler M, Rosolowski M, Schwaenen C, Spang R, Stein H, Thorns C, Trümper L, Vater I, Wessendorf S, Zenz T, Siebert R, Küppers R; Molecular Mechanisms in Malignant Lymphomas Network Project of Deutsche Krebshilfe.

Burkitt-Lymphome (BL) sind durch die sogenannte „Burkitt-Translokation" t(8;14)(q24;q32)) bzw. deren Varianten t(8;22) und t(2;8) charakterisiert [60]. Diese kann in fast allen Burkitt-Lymphomen, aber auch in anderen Lymphomen, z. B. DLBCL, nachgewiesen werden. Durch die Translokation gerät der Transkriptionsfaktor *MYC* unter den Einfluss von Immunglobulingenen und wird überexprimiert, wodurch die Proliferation der Zellen stark angeregt wird. Zusätzlich führt die Überexpression von *MYC* zur Stimulation von Apoptose-Signalwegen [60], der durch weitere Aberrationen wie die Mutation des Apoptose-Genes *TP53* entgegen gewirkt werden kann. Im Vergleich zu DLBCL zeichnen sich BL durch geringe genetische Komplexität aus [3], wobei ein Großteil der bisherigen Erkenntnisse auf chromosomaler CGH bzw. aCGH basierte, die nur geringe Auflösung ermöglichen.

Ziel der Arbeit war es, mittels hochauflösender Kopienzahl-Analyse weitere sekundäre genetische Aberrationen in Burkitt-Lymphomen zu detektieren und deren Auswirkung auf die Genexpression zu untersuchen. Im Rahmen dieser Studie wurden 39 molekular definierte BL (mBL) [3] mittels Affymetrix SNP-Arrays untersucht. 30 Fälle wurden dabei mit 500k-Arrays (250k Sty + 250k Nsp) untersucht, neun weitere mit 250k-Sty-Arrays. Die Analyse der SNP-Daten erfolgte mittels der in Abschnitt 1.4 beschriebenen Pipeline. Als euploide Referenzen wurden 20 laboreigene Sty, zehn laboreigene Nsp-Arrays sowie 39 HapMap-Fälle verwendet. Für die neun mit Sty-Arrays gemessenen Fälle wurden die fehlenden Nsp-Messwerte interpoliert, um eine gemeinsame Auswertung der Messungen zu gewährleisten.

Die Kopienzahl-Analyse bestätigte eine geringe genetische Komplexität für einen Großteil der untersuchten mBLs. Bei 29 von 39 untersuchten Tumoren zeigten >95 % der Marker normale Kopienzahl. Die Software GISTIC [37] wurde verwendet, um Regionen zu selektieren, die rekurrent von Aberrationen betroffen waren. Dabei wurden 13 rekurrente Zugewinne und 16 rekurrente Deletionen detektiert. Überlappende Uniparentale Disomien (UPD) konnten für Chromosom 1 und 17 sowie für die Chromosomenarme 6p und 13q gezeigt werden, wobei deren Rekurrenz mit circa 10 % jeweils gering war.

Für alle 39 untersuchten Lymphomfälle lagen Genexpressionsmessungen mit Affymetrix hgu133a-Arrays vor. Innerhalb der rekurrenten Regionen wurde untersucht, für welche Gene sich ein signifikanter Gendosiseffekt zeigt. Nach Korrektur für multiples Testen konnte für 17,5 % der Probesets in rekurrenten Regionen ein signifikanter Gendosiseffekt gezeigt werden. Bei acht mBL-Fällen wurden insgesamt 32 High-Level-Amplifikationen detektiert. In fünf mBL-Fällen wurden homozygote Deletionen nachgewiesen, die nicht durch Kopienzahl-Polymorphismen (CNP) erklärt werden konnten. Insgesamt waren sechs Regionen betroffen, wobei die Region Chr1:23538204–23657473 als einzige in zwei Fällen betroffen war. Diese Region umfasst die Gene *E2F2* und *ID3*. Im Rahmen des ICGC-Projektes Molecular Mechanisms in Malignant Lymphoma by Sequencing (MMML-Seq) konnte inzwischen gezeigt werden, dass *ID3* in 68 % (36/53) der mBL-Fälle mutiert ist [26]. Die Deaktivierung von *ID3* stellt daher vermutlich einen zentralen Mechanismus in der Entstehung von BL dar [26, 38].

Detection of genomic aberrations in molecularly defined Burkitt's lymphoma by array-based, high resolution, single nucleotide polymorphism analysis

René Scholtysik,[1] Markus Kreuz,[2] Wolfram Klapper,[3] Birgit Burkhardt,[4] Alfred C. Feller,[5] Michael Hummel,[6] Markus Loeffler,[2] Maciej Rosolowski,[2] Carsten Schwaenen,[7] Rainer Spang,[8] Harald Stein,[6] Christoph Thorns,[5] Lorenz Trümper,[9] Inga Vater,[10] Swen Wessendorf,[7] Thorsten Zenz,[7] Reiner Siebert,[10] and Ralf Küppers,[1] for the "Molecular Mechanisms in Malignant Lymphomas" Network Project of the Deutsche Krebshilfe

[1]Institute of Cell Biology (Cancer Research), Medical School, University of Duisburg-Essen, Essen; [2]Institute of Medical Informatics, Statistics and Epidemiology (IMISE), University of Leipzig, Leipzig; [3]Department of Pathology, Hematopathology Section and Lymph Node Registry, University Hospital Schleswig-Holstein, Campus Kiel/Christian-Albrechts-University, Kiel; [4]NHL-BFM Study Center, Department of Pediatric Hematology and Oncology, Justus-Liebig-University, Giessen; [5]Institute of Pathology, University of Lübeck, Lübeck; [6]Institute of Pathology, Charité, Campus Benjamin Franklin, Berlin; [7]Internal Medicine III, University Hospital of Ulm, Ulm; [8]Institute of Functional Genomics, University of Regensburg, Regensburg; [9]Department of Hematology/Oncology, University Hospital Göttingen, Göttingen and [10]Institute of Human Genetics, Christian-Albrechts University Kiel & University Hospital Schleswig-Holstein, Campus Kiel, Kiel, Germany

ABSTRACT

Background

Knowledge about the genetic lesions that occur in Burkitt's lymphoma, besides the pathognomonic *IG-MYC* translocations, is limited.

Design and Methods

Thirty-nine molecularly-defined Burkitt's lymphomas were analyzed with high-resolution single-nucleotide polymorphism chips for genomic imbalances and uniparental disomy. Imbalances were correlated to expression profiles and selected micro-RNA analysis. Translocations affecting the *MYC* locus were studied by fluoresence *in situ* hybridization.

Results

We detected 528 copy number changes, defining 29 recurrently imbalanced regions. Five hundred and eighteen regions of uniparental disomy were found, but these were rarely recurrent. Combined imbalance mapping and expression profiling revealed a strong correlation between copy number and expression. Several recurrent imbalances affected the MYC pathway: the micro-RNA-supercluster 17-92 was frequently gained and the transcription factor E2F2 was recurrently deleted. Molecular Burkitt's lymphoma lacking *MYC* translocations showed *MYC* gains. Amplifications of the polymerase iota gene were associated with increased frequency of positions scored as aberrant.

Conclusions

The present findings suggest that uniparental disomies do not play a major role in the pathogenesis of Burkitt's lymphoma, whereas some genes may contribute to the development of this lymphoma through gene dosage effects. Amplifications of the polymerase iota gene may be functionally linked with increased genomic alterations in Burkitt's lymphoma. The pattern and rarity of chromosomal changes detectable, even at the high resolution employed here, together with aberrations of genes regulating MYC activity, support the hypothesis that deregulation of the MYC pathway is the major force driving the pathogenesis of Burkitt's lymphoma, but show that this deregulation is more complex than previously known.

Key words: Burkitt's lymphoma, genomic imbalance, MYC, uniparental disomy.

76

Introduction

Burkitt's lymphoma (BL) is an aggressive mature B-cell lymphoma. Left untreated, it is fatal within months. There are three subtypes of BL: sporadic BL, mainly found in the Western world, endemic BL, mostly found in tropical Africa, and BL associated with immunosuppression.[1] About 90% of patients with endemic BL and 30% of those with sporadic BL are infected by Epstein-Barr virus (EBV), suggesting a pathogenic role of the virus, although only few viral genes are expressed in EBV⁺ BL.[2]

The hallmark genetic lesion of BL is the so-called "Burkitt translocation" t(8;14)(q24;q32) and its variants t(8;22) and t(2;8), which juxtapose the *MYC* oncogene to one of the three immunoglobulin (*IG*)-loci.[2] One of these Burkitt translocations is present in almost all cases of BL investigated so far,[2] but is not specific because it is also found in other lymphoma types.[3] *IG* locus-driven *MYC* expression leads to strong proliferation signals, allowing the cells to grow rapidly.[2] However, strong expression of *MYC* also stimulates apoptotic pathways.[2] The BL clone must, therefore, acquire additional lesions to disrupt this signaling to benefit from the growth-enhancing effects of MYC. Some of the additional mutations described so far disrupt the p53-mediated pathway of apoptosis, by targeting *TP53* itself or components of its signaling cascade, allowing BL cells to evade external and internal death signals.[2] Compared to most other mature B-cell lymphomas, BL is characterized by the rarity of chromosomal aberrations secondary to the *IG-MYC* fusion. The most frequent secondary changes in BL detected by conventional cytogenetics are gains in 1q and in chromosomes 7 and 12.[3]

The current knowledge on secondary chromosomal changes in BL predominantly relies on analyses performed by conventional cytogenetics and comparative genomic hybridization (CGH) to chromosomes or low-resolution arrays.[4-8] These techniques are not suitable for detecting chromosomal imbalances at high-resolution and might, therefore, have failed to detect pathogenically important changes. Moreover, they are also not suitable for detecting chromosomal changes not causing structural alterations, such as uniparental disomies (UPD). UPD were recently detected in various cancers[9] and have been proposed to be an alternative mechanism to down-regulate tumor suppressor genes by duplication of inactivating mutations or deletions on one allele.[10,11] Finally, previous studies have focused on cases not classified by gene expression. As we recently showed, a considerable number of mature aggressive B-cell lymphomas would have to be re-classified if molecular signatures are applied.[8] We, therefore, performed high-resolution single nucleotide polymorphism (SNP) chip analysis on 39 sporadic BL cases collected by the "Molecular Mechanisms of Malignant Lymphomas" (MMML) consortium and defined as "molecular Burkitt's lymphoma (mBL)".

Design and Methods

Collection and extraction of DNA

Biopsies were diagnosed by experienced pathologists from the MMML consortium according to WHO criteria.[1] Five cases of BL were used as "core" BL in the classification based on gene expression,[8] 29 were in the "atypical" group. A tumor cell content of at least 70% was an inclusion criterion. Whole tissue DNA of tumor sections was extracted with the QiaAmp DNA Blood Kit according to the manufacturer's manual (Qiagen, Hilden, Germany). Central approval for the MMML network was obtained through the institutional review board of the University of Göttingen (D403/05).

Single nucleotide polymorphism chips

All 39 samples were analyzed on 250k Sty GeneChips (Affymetrix, Santa Clara, CA, USA), and 30 of these with enough material were additionally hybridized to Nsp-chips for a combined resolution of approximately 500k SNP. Each 250k GeneChip was prepared and hybridized according to the Affymetrix manual. The GeneChips were scanned by a GeneChip scanner 3000 with G7 update (Affymetrix). The sample files were genotyped with the BRLMM-algorithm (*see below*). The SNP-chip files have been submitted to the GEO database under accession number GSE21597.

Genotyping and copy number analysis

The BRLMM algorithm[12] was applied with default parameters (*score threshold*=0.5, *prior size*=10000 and *DM threshold*=0.17) to genotype mBL and 63 non-mBL tumor samples (*unpublished data*) using 39 Hapmap samples provided by Affymetrix (*http://www.affymetrix.com/support/technical/sample_data/500k_data.affx*) as a reference. The reference set was complemented by 20 laboratory-specific reference samples (11 female, 9 male) for the Sty array and 10 (6 female, 4 male) for the Nsp array. We supplemented the Hapmap samples with laboratory-specific samples because the Copy Number Analyzer for GeneChips (CNAG) software selects the reference samples that minimize the signal variance, so that a larger collection of references results in a lower "noise", and the laboratory-specific references should better eliminate a laboratory-specific signature. Median call rates of the mBL tumor samples were 96.23% and 98.96% for Sty- and Nsp-arrays, respectively (range, 90.44-98.63% and 90.26-99.65%).

Copy number analysis was performed using the CNAG program v2.0,[13] employing the same reference samples as for genotyping. CNAG was configured to select an optimal gender-specific reference set individually for each array.[13] This resulted in selection of 97.6% of the laboratory-specific references for Sty-arrays and 81.6% for Nsp-arrays. The lower proportion for Nsp possibly reflects the lower number of references for the Nsp-samples, giving the software less variance in references to choose from. For samples with Sty and Nsp arrays available, data from both chips were combined. Segmentation of raw copy number data was performed using the hidden Markov model (HMM) approach provided by CNAG.

HMM parameters were adjusted individually for each array to adapt the segmentation to differences in hybridization quality and tumor cell content of the analyzed samples. Starting with default parameters, the mean levels of HMM states were adjusted to optimize the segmentation results, i.e. to avoid missing clearly aberrant regions, as well as to prevent frequent successive alternation between neighboring HMM states.

With regard to outliers and technical artifacts, HMM segments were considered as copy number aberration only if they consisted of at least five consecutive imbalanced SNP. A preliminary analysis with a more stringent limit of ten consecutive SNP resulted in the identification of the same 33 recurrent regions, and only in small shifts in the borders of six of these regions (see Genomic Identification of Significant Targets in Cancer [GISTIC] analysis), due to exclusion of one imbalanced case for each of these six regions. This shows the robustness of the approach. High level amplifications were defined as aberrations with an HMM copy number of at least five, homozygous deletions as aberrations with a copy number of zero.

77

For the chromosome X in males (except for pseudo-autosomal region 1), interpretation of log₂ ratios had to be adjusted with respect to the gender-specific (single copy) reference. Therefore, segments with an estimated HMM copy number of one to three copies were assumed to be normal (copy number n=1). Segments with an estimated copy number of at least four were selected as copy number gains and segments with an estimated copy number of zero as losses.

Loss of heterozygosity and uniparental disomy analysis

An HMM-based method[14] implemented in the dChip program[15,16] (Build date: Apr 11, 2007) was used to infer regions with loss of heterozygosity (LOH). The "HMM considering haplotype" (LD-HMM)[14] method was selected for the LOH calculations to account for linkage disequilibrium-induced SNP dependencies. The LOH call threshold was set at 0.99, applying an empirical haplotype correction.[14] Thus, putative LOH regions were excluded if there was 95% concordance of the homozygous genotypes of the candidate LOH region with respective regions of more than 5% of the reference samples. For samples for which both Sty and Nsp arrays were available, the combined Sty/Nsp set of 39 Affymetrix Hapmap samples was used as the normal reference set. The complete set of Sty reference arrays was selected as the reference for tumor samples without Nsp array data.

LOH regions were called UPD if no copy number aberrations were present in the region. In LOH regions partially affected by copy number aberrations, subregions without copy number aberrations were classified as UPD if they comprised 50 or more neighboring SNP.

Test for recurrence using Genomic Identification of Significant Targets in Cancer

For each imbalance detected by CNAG, the mean log₂ ratio was determined. For gains and losses, the segment level was set to 0.1 and -0.1 if the mean log₂ ratio was less than 0.1 or greater than -0.1, respectively. The segment level of balanced regions was set to 0. Using these segment values we selected recurrent gains and losses using GISTIC with standard parameters and thresholds.[17] Regions with known copy number polymorphisms were filtered by GISTIC using the Database of Genomic Variants (Version: June 2008).[18] Each recurrent aberration identified by GISTIC was represented by a region with the highest G-score (peak region) and a robustified wide peak region.[17] If the majority of the markers showed a concordant copy number status for the peak or the wide peak region, the respective region in each of the affected cases was classified as present.

Correlation of copy number and gene expression

We compared the mRNA expression level of genes within a recurrent region detected by GISTIC between cases with aberrant and balanced genomic status using Student's t-test, employing Hg17 (NCBI Build 35) for annotation, considering only regions with more than two contributing samples. Gene expression profiles of BL have been published.[8] Due to the small sample size we combined gains and high level amplifications. Adjustment for multiple testing within each recurrent region was performed with the step-down minP method implemented in the R-package multtest[19] using a family wise error rate of 5%.[20]

RNA extraction, reverse transcription and real-time polymerase chain reaction quantification of micro-RNA

RNA extraction, reverse transcription and real-time polymerase chain reaction (PCR) quantification were performed as described previously.[21] Briefly, the RecoverAll kit (Ambion, Austin, Texas, USA) was used for RNA-extraction from four 20 μm sections of formalin-fixed, paraffin-embedded tissues. The TaqMan® universal PCR master mix, No AmpErase® UNG-kit and the TaqMan® microRNA (miRNA) reverse transcription kit from Applied Biosystems (Foster City, California, USA) were used for cDNA synthesis and quantitative PCR. The reactions were performed in a 384-well format for 377 different miRNA (U6 in quadruplicate and miR16 in quintuplicate). The profiles were measured on the LightCycler® 480 instrument (Roche, Basel, Switzerland). To normalize the obtained raw CT values, we shifted the values of each sample such that the means of the probes of U6 and miR-16 were constant across all samples. That constant was chosen to be the mean of the means of U6 and miR-16.

Results

We generated 250k Sty SNP chips from 39 mature, aggressive B-cell lymphomas that had an mBL index ≥0.95 according to gene expression. For 30 of the tumors additional 250k Nsp chips were generated for a combined interrogation of 500k SNP, employing a resolution not previously attained for BL. Thirty-five of 38 evaluated tumors had an *IG-MYC* translocation detectable by fluorescence *in situ* hybridization (FISH).[8,22] Histological diagnoses by experienced pathologists did not always agree with the gene expression signature of mBL, with the series encompassing 27 BL, 6 centroblastic diffuse large B-cell lymphomas (DLBCL), one follicular lymphoma grade 3b, and 5 unclassifiable lymphomas. Twenty-four patients were 16 years or younger at the time of diagnosis.

As signals from single SNP probes are rather noisy, we only considered aberrations detected by at least five consecutive SNP for copy number analysis, resulting in a median detection window size of 33.8 kb and 16.6 kb for 250k and 500k arrays, respectively. Overall, the analysis detected 484 gains, 388 losses and 518 UPD before filtering for known copy number variations, combining the datasets from both arrays. Filtering left 308 gains and 220 losses that had less than 50% marker overlap with anno-

Table 1. Overview of the aberrations found in 39 cases of BL, after filtering for copy number polymorphisms.

Sample set	9 BL (Sty chips, 250k)			30 BL (Sty+Nsp chips, 500k)			All 39 BL
	Total	Mean	Median	Total	Mean	Median	Median size (bp)
Copy number changes	93	10.3	6	435	14.5	8.5	523,233
- Gains	63	7	3	245	8.2	4	362,547
- Losses	30	3.3	3	190	6.3	2.5	1 041,924
Uniparental disomies	58	6.4	4	460	15.3	14.5	770,662

A region was defined as aberrant if the HMM assigned a concordant copy number other than two to at least five consecutive SNP. Filtering for copy number polymorphisms was done using the Database of Genomic Variants (Version: June 2008).[40] As no corresponding germline DNA was available, UPD were detected by a statistical approach. The smallest detected aberration was a gain of 886 bases.

78

tated copy number polymorphisms (Table 1, Figure 1). Of these, 32 were high level amplifications and 6 were homozygous deletions (*Online Supplementary Tables S1 and S2*). Concerning the size of the copy number changes, the smallest aberrant region was 886 bases long and the median size was 522,475 bases (Table 1). The fraction of SNP detected in the normal copy number state of two was calculated to be more than 99% for 15 of the 39 tumors, and more than 95% for an additional 14 samples.

After filtering, data processing and testing for recurrent events, we detected 13 gains and 16 losses passing the significance threshold given by GISTIC (Tables 2 and 3; *Online Supplementary Figure S1*; complete annotation of genes in *Online Supplementary Tables S3 and S4*). It was not possible to calculate recurrence for UPD, as the tumors had too few overlapping UPD to reach a meaningful threshold. Regions with at least a few overlapping broad UPD encompass major parts of both arms of chromosomes 1 and 17, and the chromosomal arms 6p and 13q (Figure 1). The highest recurrence in gains (peak limit region aberrant in ≥7 cases) was detected for 1q25.1, 1q31.3, 3q27.2, 6q15, 11q24.3, and 13q31.3 (Table 2). The most frequent regions for losses of genetic material (peak limit region aberrant in ≥6 cases) were 3q13.13, 17p13.1,

19q13.42 and Xp22.33 (Table 3). A recent conventional CGH study which also employed a molecular definition of BL,[6] also described gains on 1q, 8q24-ter and 13q31-q32, and a recurrent loss on 17p (Tables 2 and 3).

Some regions contained genes with already known relevance for BL or cancer in general. For example, the gains at 3q27.3 (7 cases) always involved *BCL6*. Although a deregulation by translocation for this oncogene was described for large B-cell lymphomas,[23] none of the affected BL samples showed a translocation detectable by FISH.[8] Increased expression of *BCL6* was seen in affected cases (Figure 2B). Thus, BCL6 might be up-regulated solely by genomic gains in BL. Deleted regions involving or near previously known tumor suppressors were 9p21.3 (in only 2 cases) and 17p13.1, as these regions harbor *CDKN2B* and *TP53*, respectively, two well described tumor suppressor genes in BL.[24,25] While the recurrent loss on 9p21.3 narrowed exactly down to the *CDKN2B* gene, the target in the region 17p13.1 is unclear: four cases had lost one copy of the entire arm of chromosome 17, thereby deleting a copy of *TP53*. We also detected a partial UPD overlapping with the entire chromosomal region in three further cases. However, three additional cases had smaller heterozygous deletions, approximately 2 Mb away from the *TP53* gene,
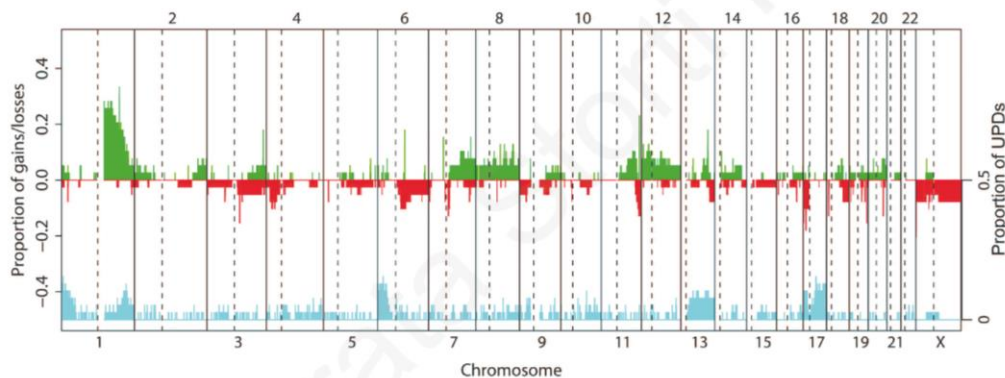

Figure 1. Proportion of gains, losses and UPD along chromosomes. y-axis: proportion of cases with the respective aberration, to the left copy number, to the right partial UPD; green: gains; red: losses; blue: UPD; dotted vertical lines depict centromeres. SNP that mapped to regions of known copy number variations and partial UPD that were also present in normal reference samples were excluded (see *Design and Methods* section).

Table 2. Recurrent gains in 39 BL as defined by GISTIC.

Cytoband	q-value[1]	Wide peak[2]	Peak limit[2]	N. wide peak	N. peak limit	Selected genes[3]
1q25.1	6.38E-06	chr1:166 155 454-171 863 618	chr1:171 851 737-171 863 618	10	10	*FASLG*
1q31.3	7.93E-06	chr1:195 253 751-195 672 280	chr1:195 369 624-195 638 957	12	13	*PTPRC*
3q27.3	0.058	chr3:188 956 472-189 175 357	chr3:189 056 204-189 167 011	6	7	*BCL6*
6q15	0.095	chr6:91 249 512-91 370 379	chr6:91 269 731-91 295 723	6	7	*MAP3K7*
7p12.2	0.12	chr7:49 995 112-50 127 527	chr7:49 995 112-50 127 527	6	6	*IKZF1*
7q34	0.11	chr7:141 162 879-142 244 396	chr7:141 162 879-142 244 396	6	6	
8q24.13	0.16	chr8:121 784 261-129 821 355	chr8:126 106 216-126 249 512	4	5	*MYC*
11q23.3	0.11	chr11:115 828 212-118 580 638	chr11:117 008 605-118 321 365	3	3	
11q24.3	0.012	chr11:127 825 348-127 931 093	chr11:127 825 348-127 919 072	9	9	*ETS1*
12q15	0.0051	chr12:61 724 591-82 263 925	chr12:67 278 531-68 393 145	3	4	
12q24.33	0.19	chr12:128 465 362-132 449 811	chr12:131 860 026-132 008 369	2	3	
13q31.3	1.16E-05	chr13:90 362 259-90 843 579	chr13:90 766 137-90 811 413	7	7	*hsa-mir-17-92*
18q21.2	7.75E-05	chr18:49 301 605-55 260 237	chr18:49 999 396-53 164 071	2	3	*POLI*

[1]significance as reported by GISTIC; [2]for the meaning of different regions see the Design and Methods section; [3]for a complete list of all annotated genes see Online Supplementary Table S3.

79

each including *STX8*, *USP43* and *WDR16* as the only annotated genes. In our gene expression data, only *STX8* is represented on the chip and is significantly down-regulated in cases with deletion (Figure 2A). Deletions involving 13q14.3 and/or 13q34 were recently observed by FISH in over 40% of childhood BL.[26] We, however, detected a recurrent loss centered at 13q32 only in 3/39 (7.7%) cases.

Concerning *MYC*, we detected a recurrent gain in five cases at 8q24.13, the location of this gene. Notably, the expression of *MYC* is not significantly different between cases with the respective gain and cases with an *IG-MYC* translocation, which is known to cause strong MYC over-expression (Figure 2A). It is interesting that three out of 39 BL cases had no detectable *MYC* translocation in FISH,[8,22] although this does not exclude cryptic insertions of *MYC* into an *IG* locus or vice versa. Nevertheless, they were scored as mBL by gene expression.[8] Two of these are among the cases with *MYC* gain.

The gain in 1q31.3, found in 13 cases, contained the gene *PTPRC*, a regulator of B-cell receptor and cytokine signaling,[27] and two annotated miRNA genes (hsa-mir-181b-1 and -213). The high recurrence of this gain hints at its importance for the tumor. Because the gene expression data showed no significant up-regulation of *PTPRC* in affected cases, and the gains often involved not the complete coding sequence of this gene (*data not shown*), the miRNA genes are strong target candidates.

The region gained on 13q31.3 in seven BL harbored the miRNA-17-92 supercluster. Only one of these cases showed a high level amplification at this location. For some of the BL analyzed (but not for the single case showing the amplification), miRNA expression profiles for their mature forms were recorded. The miRNA contained in the supercluster were consistently expressed at higher levels in cases with the respective gain (*Online Supplementary Table S5*), though this failed to be statistically significant after correcting for multiple testing. Three cases harbored

losses at 1p36.12 (and 4 further cases had large UPD involving the same region) which contained the gene *E2F2*, a transcription factor that binds to the promoter of the miRNA-17-92 cluster.

Three BL showed a high level amplification of 18q21.2. Remarkably, these three are first, third and seventh in a ranked list of all 39 BL ordered according to the proportion of aberrant SNP (Figure 3). The gene expression data for this region (*Online Supplementary Figure S2*) show that the *POLI* gene was significantly up-regulated, as were three other genes with unclear relevance.

Using expression data available for all BL,[8] we compared the expression of genes in a recurrent region in cases with an aberration detected by SNP-chip to cases without the respective aberration, to pinpoint gene dosage effects of possible oncogenes or tumor suppressors, if present. Some exemplary regions are shown in Figure 2. We detected a clear concordant effect of copy number on expression strength: a gain led preferentially to a stronger expression of a large fraction of gained genes, with 12 of 108 probe-sets even reaching statistical significance after correction for multiple testing (and not a single significantly down-regulated gene). Likewise, the majority of genes in a heterozygously lost region displayed reduced expression (52 of 258 probe-sets significant after correction, 2 inversely correlated). A statistically significant positive correlation was, therefore, detected in 17.5% of all probe-sets in aberrant regions.

Discussion

Cytogenetic and low-resolution CGH studies previously performed on BL revealed few consistent genomic imbalances.[4,5] A caveat regarding these studies was the definition of BL on non-molecular grounds. These earlier studies probably included unrecognized molecular DLBCL in their

Table 3. Recurrent losses in 39 cases of BL as defined by GISTIC.

Cytoband	q-value[1]	Wide peak[2]	Peak limit[2]	N. wide peak	N. peak limit	Selected genes[3]
1p36.12	0.014	chr1:23,457,835-23,714,048	chr1:23,535,005-23,673,234	2	3	*E2F2*
3p14.2	0.15	chr3:60,396,160-60,637,030	chr3:60,452,472-60,580,245	3	3	*FHIT*
3q13.13	0.085	chr3:110,456,383-110,657,226	chr3:110,482,177-110,611,242	6	6	*DPPA4, DPPA2*
4p15.32	0.15	chr4:17,969,802-29,966,659	chr4:17,991,350-18,456,461	3	4	
6q14.3	0.094	chr6:76,830,186-107,898,353	chr6:78,151,877-96,248,104	4	4	
7q11.22	0.15	chr7:66,573,850-67,281,794	chr7:66,810,706-67,240,659	5	5	
10p11.21	0.22	chr10:36,983,732-37,133,784	chr10:37,007,511-37,106,346	3	3	
11q24.3	0.085	chr11:122,263,650-134,452,384	chr11:126,328,752-132,876,984	4	5	
13q33.2	0.20	chr13:96,008,275-114,142,980	chr13:96,913,194-112,193,013	3	3	
17p13.1	0.014	chr17:9,125,548-9,654,733	chr17:9,125,548-9,598,824	7	7	*STX8, TP53*
18p11.23	0.094	chr18:7,079,886-7,359,571	chr18:7,125,227-7,359,571	5	5	
18q22.3	0.15	chr18:55,355,976-76,117,153	chr18:55,355,976-76,117,153	3	3	
19q13.42	0.085	chr19:58,746,599-59,029,391	chr19:58,758,793-58,990,786	6	6	
20q13.2	0.094	chr20:49,856,082-50,168,587	chr20:49,958,057-50,144,808	5	5	
Xp22.33	0.0042	chrX:1-1,499,465	chrX:1-1,499,465	8	8	
Xp11.3	0.085	chrX:1-154,824,264	chrX:46,975,370-47,150,005	3	3	

[1]*significance as reported by GISTIC;* [2]*for the meaning of different regions, see the Design and Methods section;* [3]*for a complete list of all annotated genes, see Online Supplementary Table S4.*

80

BL set, which contributed changes that newer studies do not recognize in mBL. The single high-resolution oligonucleotide array-CGH analysis in BL included, besides cell lines, only 13 primary tumors which were analyzed by a 44k array,[7] limiting resolution and sensitivity. Furthermore, lymphomas were selected as BL based only on histological diagnosis.[7] In our approach using molecularly-defined BL, we detected several of the aberrations that this paper found in only one to three histological BL. In most case we were able (due to the larger number of samples and the higher resolution) to set narrower limits to these regions and confirm them as recurrent events in primary tumors.

Concerning our published CGH analysis,[8,22] at a resolution of 3000 probes of a subset of cases also analyzed here, we note a good concordance between the two methods. As expected, the SNP analysis detected multiple additional aberrations that the CGH could not detect because of res-
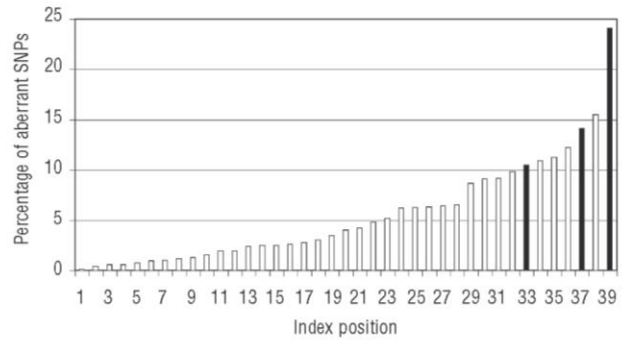

Figure 3. Correlation of the recurrent gain at chr18 involving *POLI* and the proportion of aberrant SNP. In a list ordered according to the proportion of aberrant SNP detected (y-axis), the three cases with an amplification involving *POLI* rank first, third and seventh (columns marked in black; *P*=0.0035 in a Mann-Whitney U test).
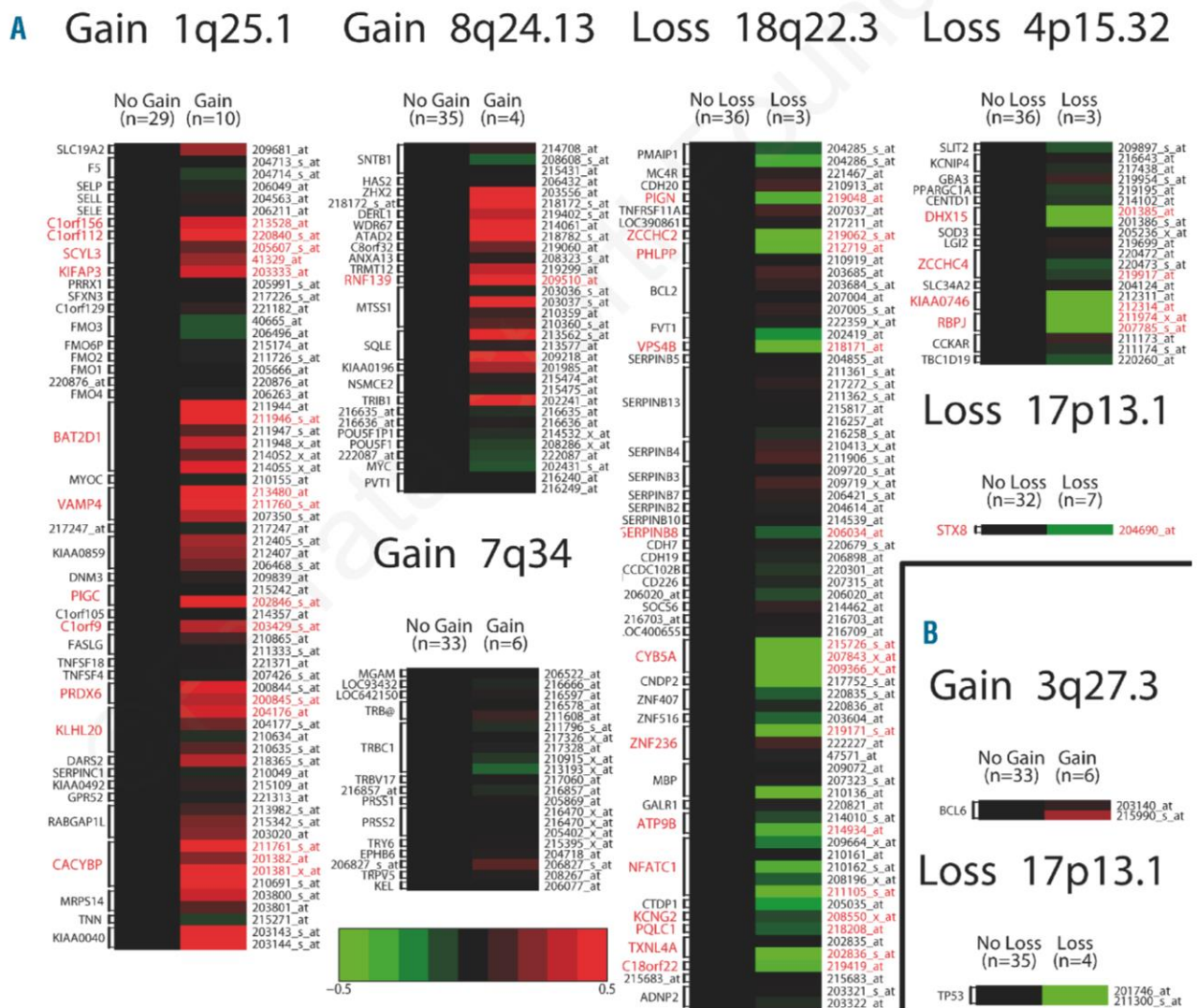
Figure 2. Gene dosage effects for genes in recurrently imbalanced regions. Green: reduced expression; red: stronger expression. The difference in expression between aberrant and non-aberrant cases is displayed on a log-scale. For all regions analyzed except 7q34, a clear gene dosage effect is visible. Gene names marked in red indicate significant differences in expression after correcting for multiple testing. (A) whole recurrently aberrant regions depicted; (B) single genes of interest out of a larger region.

81

olution constraints, but the raw signal values for the CGH clones gave the correct tendency for each aberration detected with SNP chips (*unpublished data*). Regarding high level amplifications and homozygous losses before filtering for copy number polymorphisms, 22 concordant regions were identified with both platforms. Importantly, however, 35 additional imbalances were only seen with the SNP-chips as the respective regions were not covered on the CGH-array, and in ten further instances of imbalances detected by SNP-chip analysis, CGH signals classified the regions as balanced.

Regarding the extent of genomic lesions in BL as compared to in other aggressive lymphomas, e.g. DLBCL, a direct comparison is currently not possible due to lack of data from identical high-resolution platforms. Nevertheless, an array CGH study with 3000 BAC clones detected a median of 8.5 copy number changes for DLBCL,[8] while, with our higher resolutions, the median copy number changes we found in BL was 6.0 with the 250k array and 8.5 with the 500k array (Table 1). Recent publications on comparable high-resolution SNP-chip approaches to mostly solid tumors reported more copy number changes per case (e.g. a median of 12 losses and 12 gains; although it should be considered that a change of >0.1 copies was already considered aberrant in one of these studies).[28,29] We, therefore, confirm that BL has a relatively stable genome.

An important advantage of the SNP chip method compared to CGH is the former's ability to detect UPD, with the caveat that the exact detection of these disomies was hampered in this study as no normal tissue DNA was available for germline comparison. Nevertheless, using a statistical approach, robust pinpointing of larger stretches of LOH is possible. In several tumors, UPD are an additional way to inactivate tumor suppressor genes,[9,10] beside deletions and/or mutations. Thus, the integration of losses and UPD may allow better detection of new candidate tumor suppressors. Instead, we show that in BL there is little congruency between UPD and regions recurrently lost and no strong accumulation of UPD in specific regions (Figure 1), suggesting that UPD do not play a major role in BL. Nevertheless, the few UPD detected here may still be of functional relevance in the affected cases, as exemplified for *TP53*, for which cases with UPD were found besides cases with deletions.

Regarding the differential gene expression analysis, we detected a clear gene dosage effect for the recurrently aberrant regions (Figure 2). The only exception to this is the recurrent gain at 7q34, which includes nine genes, but not a single one with a strong dosage effect. This underlines that the gain and loss of genetic material can be a potent mechanism to deregulate certain genes in BL, in line with the results of a recent study comparing array-CGH data with gene expression levels.[6]

Four cases that had a large deletion on chromosome 17 retained only a single copy of *TP53*, encoding the p53 tumor suppressor. These four cases showed lower *TP53* gene expression compared with the 35 unaffected cases (Figure 2B). Notably, three further cases showed UPD encompassing *TP53*, suggesting that these events may also have pathogenic relevance. Indeed, two of two cases with deletion and two of two cases with UPD that were sequenced had inactivating *TP53* mutations (*data not shown*). The exact target of smaller deletions on chromosome 17 in three additional cases was more ambiguous:

the deletion was approximately 2 Mb away from *TP53* and included three annotated genes (*STX8*, *USP43* and *WDR16*) that were heterozygously lost in each affected case. However, as none of these genes shows differential expression between B-cell subsets (*data not shown*) and there were no previous reports about their functional involvement in cancer, the relevance of these deletions is unclear.

A significantly up-regulated gene in the amplified region 18q21.2, *POLI*, codes for DNA polymerase iota which plays a role as an error-prone polymerase in somatic hypermutation[30] and some forms of DNA-damage response.[31] We speculate that over-expression of POLI shifted the balance of DNA repair to more error-prone repairs, which would explain the high percentage of aberrant SNP detected in these cases (Figure 3). The expression of *POLI* was also positively correlated with an increased percentage of aberrant SNP in our set of tumors (Spearman's correlation=0.3; *P*=0.06). In breast cancer cells, a higher amount of POLI leads to a higher mutation burden,[32] suggesting a similar role in BL. A recent study in mice actually implicated *POLI* as a new oncogene in a colorectal cancer model.[33]

We found that the *MYC* gene was not only targeted by translocation events, but additionally by gains of genetic material in five of the 39 BL. The gene expression was not different in mBL with the gain, but this is explainable by the fact that, to become a mBL, the up-regulation of *MYC* is obligatory.[8] Indeed, we identified three mBL that had no detectable *MYC* translocation by FISH ([8] and *data not shown*), but two of these showed a low-level gain including *MYC*. These cases were nevertheless classified as mBL.[8] The most frequent deregulation of *MYC* is, therefore, by translocation to another enhancer (such as *IG-MYC*), but we suppose that the gain of genomic material is an additional mechanism by which the expression of this important driver gene can be increased. The three BL with a gain in addition to the translocation probably represent unbalanced translocation events. For the remaining single case without translocation and without gain, the mechanism of deregulation is unknown. *TP53* deletions, which may hamper the pro-apoptotic activity of MYC, were not specifically associated with BL harboring *MYC* gains.

Regarding MYC deregulation, gain of the miRNA-17-92 supercluster in seven cases is of note, because it is directly transactivated by MYC.[34] This cluster has been shown to be frequently over-expressed in several types of solid tumors[35] and also B-cell malignancies,[36] in which it is also often amplified.[37] Over-expression of this cluster in mouse lymphocytes reduces apoptosis and accelerates lymphoma development,[38] defining it as an oncogene.[37,39,40] In addition, it was recently shown that these miRNA act synergistically with MYC in the development of aggressive cancer.[36] They provide the tumor with a means to counteract activation-induced apoptosis caused by the high *MYC* expression in BL, by reducing the expression of tumor suppressor genes, for example *TGFBR2*,[36] *BIM* and *PTEN*.[38,40]

The miRNA-17-92 supercluster is controlled by an autoregulatory loop with E2F-family proteins.[41] In our series, three cases had heterozygous deletion of the *E2F2* gene on 1p36.12, which directly binds the promoter of the supercluster and increases its expression. The three cases with *E2F2* deletion had lower mRNA levels of the gene than the other cases, although the difference did not reach

statistical significance (*data not shown*). The deletion of a positive regulator of the miRNA cluster at first seems counterintuitive, as other cases in our series had specifically amplified miRNA polycistron, hence over-expression of these miRNA seems to be advantageous for the clone. Besides regulating the miRNA cluster 17-92, *E2F2* thus presumably has other effects on the tumor which promoted the deletion. It was indeed shown that *E2F2* functions as a tumor suppressor in B cells of mice over-expressing *MYC*.[42] The mice with heterozygous *E2F2* deletions show strong haploinsufficiency, resulting in accelerated development of *MYC*-driven B-cell lymphomas, suggesting that also in human BL, the heterozygous losses could have a strong effect on the tumor clone. We thus propose that the heterozygous deletion of *E2F2* would indeed tend to decrease the expression of the miRNA-17-92 supercluster, but this potential down-regulation is overcompensated in BL by the high amount of MYC, a direct transactivator of this supercluster. In this way, the tumor can benefit from the remaining effects of reduced *E2F2* discussed above. In our dataset,

gain of the miRNA-cluster 17-92 and deletion of *E2F2* were mutually exclusive, hinting at the possibility that either event is sufficient to deregulate the feedback loop.

In summary, we refined the make-up of known genetic events in BL with high resolution, and found many new recurrent aberrations, whose functional significance is mostly unclear so far. Some of these newly identified lesions affect the MYC pathway, suggesting that this pathway is deregulated in a more complex fashion than previously thought.

Authorship and Disclosures

References

1. Swerdlow SH. WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues (4th ed). Lyon: International Agency for Research on Cancer; 2008.
2. Hecht JL, Aster JC. Molecular biology of Burkitt's lymphoma. J Clin Oncol. 2000;18(21):3707-21.
3. Boerma EG, Siebert R, Kluin PM, Baudis M. Translocations involving 8q24 in Burkitt lymphoma and other malignant lymphomas: a historical review of cytogenetics in the light of todays knowledge. Leukemia. 2009;23(2):225-34.
4. Barth TF, Muller S, Pawlita M, Siebert R, Rother JU, Mechtersheimer G, et al. Homogeneous immunophenotype and paucity of secondary genomic aberrations are distinctive features of endemic but not of sporadic Burkitt's lymphoma and diffuse large B-cell lymphoma with MYC rearrangement. J Pathol. 2004;203(4):940-5.
5. Garcia JL, Hernandez JM, Gutierrez NC, Flores T, Gonzalez D, Calasanz MJ, et al. Abnormalities on 1q and 7q are associated with poor outcome in sporadic Burkitt's lymphoma. A cytogenetic and comparative genomic hybridization study. Leukemia. 2003;17(10):2016-24.
6. Salaverria I, Zettl A, Bea S, Hartmann EM, Dave SS, Wright GW, et al. Chromosomal alterations detected by comparative genomic hybridization in subgroups of gene expression-defined Burkitt's lymphoma. Haematologica. 2008;93(9):1327-34.
7. Toujani S, Dessen P, Ithzar N, Danglot G, Richon C, Vassetzky Y, et al. High resolution genome-wide analysis of chromosomal alterations in Burkitt's lymphoma. PLoS One. 2009;4(9):e7089.
8. Hummel M, Bentink S, Berger H, Klapper W, Wessendorf S, Barth TF, et al. A biologic definition of Burkitt's lymphoma from transcriptional and genomic profiling. N Engl J Med. 2006;354(23):2419-30.
9. Tuna M, Knuutila S, Mills GB. Uniparental disomy in cancer. Trends Mol Med. 2009;15(3):120-8.
10. Nielaender I, Martin-Subero JI, Wagner F, Martinez-Climent JA, Siebert R. Partial uniparental disomy: a recurrent genetic mechanism alternative to chromosomal deletion in malignant lymphoma. Leukemia. 2006;20(5):904-5.
11. Vater I, Wagner F, Kreuz M, Berger H, Martin-Subero JI, Pott C, et al. GeneChip analyses point to novel pathogenetic mechanisms in mantle cell lymphoma. Br J Haematol. 2009;144(3):317-31.
12. Affymetrix. BRLMM: an improved genotype calling method for the genechip human mapping 500k array set. Affymetrix, Inc White Paper; 2006.
13. Nannya Y, Sanada M, Nakazaki K, Hosoya N, Wang L, Hangaishi A, et al. A robust algorithm for copy number detection using high-density oligonucleotide single nucleotide polymorphism genotyping arrays. Cancer Res. 2005;65(14):6071-9.
14. Beroukhim R, Lin M, Park Y, Hao K, Zhao X, Garraway LA, et al. Inferring loss-of-heterozygosity from unpaired tumors using high-density oligonucleotide SNP arrays. PLoS Comput Biol. 2006;2(5):e41.
15. Lin M, Wei LJ, Sellers WR, Lieberfarb M, Wong WH, Li C. dChipSNP: significance curve and clustering of SNP-array-based loss-of-heterozygosity data. Bioinformatics. 2004;20(8):1233-40.
16. Zhao X, Li C, Paez JG, Chin K, Janne PA, Chen TH, et al. An integrated view of copy number and allelic alterations in the cancer genome using single nucleotide polymorphism arrays. Cancer Res. 2004;64(9):3060-71.
17. Beroukhim R, Getz G, Nghiemphu L, Barretina J, Hsueh T, Linhart D, et al. Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. Proc Natl Acad Sci USA. 2007;104(50):20007-12.
18. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, et al. Detection of large-scale variation in the human genome. Nat Genet. 2004;36(9):949-51.
19. Pollard KS, Ge Y, Dudoit S. Multtest: Resampling-based multiple hypothesis testing, R package version 1.12.0, http://CRAN.R-project.org/package=multtest; 2006.
20. Kreuz M, Rosolowski M, Berger H, Schwaenen C, Wessendorf S, Loeffler M, et al. Development and implementation of an analysis tool for array-based comparative genomic hybridization. Methods Inf Med. 2007;46(5):608-13.
21. Roehle A, Hoefig KP, Repsilber D, Thorns C, Ziepert M, Wesche KO, et al. MicroRNA signatures characterize diffuse large B-cell lymphomas and follicular lymphomas. Br J Haematol. 2008;142(5):732-44.
22. Klapper W, Szczepanowski M, Burkhardt B, Berger H, Rosolowski M, Bentink S, et al. Molecular profiling of pediatric mature B-cell lymphoma treated in population-based prospective clinical trials. Blood. 2008;112(4):1374-81.
23. Sanchez-Beato M, Sanchez-Aguilera A, Piris MA. Cell cycle deregulation in B-cell lymphomas. Blood. 2003;101(4):1220-35.
24. Gaidano G, Ballerini P, Gong JZ, Inghirami G, Neri A, Newcomb EW, et al. p53 mutations in human lymphoid malignancies: association with Burkitt lymphoma and chronic lymphocytic leukemia. Proc Natl Acad Sci USA. 1991;88(12):5413-7.
25. Klangby U, Okan I, Magnusson KP, Wendland M, Lind P, Wiman KG. p16/INK4a and p15/INK4b gene methylation and absence of p16/INK4a mRNA and protein expression in Burkitt's lymphoma. Blood. 1998;91(5):1680-7.
26. Nelson M, Perkins SL, Dave BJ, Coccia PF, Bridge JA, Lyden ER, et al. An increased frequency of 13q deletions detected by fluorescence in situ hybridization and its impact on survival in children and adolescents with Burkitt lymphoma: results from the Children's Oncology Group study CCG-5961. Br J Haematol. 2010;148(4):600-10.
27. Hermiston ML, Zikherman J, Zhu JW. CD45, CD148, and Lyp/Pep: critical phosphatases regulating Src family kinase signaling networks in immune cells. Immunol Rev. 2009;228(1):288-311.
28. Beroukhim R, Mermel CH, Porter D, Wei G, Raychaudhuri S, Donovan J, et al. The

83

landscape of somatic copy-number alteration across human cancers. Nature. 2010;463(7283):899-905.

29. Leary RJ, Lin JC, Cummins J, Boca S, Wood LD, Parsons DW, et al. Integrated analysis of homozygous deletions, focal amplifications, and sequence alterations in breast and colorectal cancers. Proc Natl Acad Sci USA. 2008;105(42):16224-9.

30. Faili A, Aoufouchi S, Flatter E, Gueranger Q, Reynaud CA, Weill JC. Induction of somatic hypermutation in immunoglobulin genes is dependent on DNA polymerase iota. Nature. 2002;419(6910):944-7.

31. Petta TB, Nakajima S, Zlatanou A, Despras E, Couve-Privat S, Ishchenko A, et al. Human DNA polymerase iota protects cells against oxidative stress. EMBO J. 2008;27 (21):2883-95.

32. Yang J, Chen Z, Liu Y, Hickey RJ, Malkas LH. Altered DNA polymerase iota expression in breast cancer cells leads to a reduction in DNA replication fidelity and a higher rate of mutagenesis. Cancer Res. 2004;64(16):5597-607.

33. Starr TK, Allaei R, Silverstein KA, Staggs RA, Sarver AL, Bergemann TL, et al. A transposon-based genetic screen in mice identifies genes altered in colorectal cancer. Science. 2009;323(5922):1747-50.

34. O'Donnell KA, Wentzel EA, Zeller KI, Dang CV, Mendell JT. c-Myc-regulated microRNAs modulate E2F1 expression. Nature. 2005;435(7043):839-43.

35. Volinia S, Calin GA, Liu CG, Ambs S, Cimmino A, Petrocca F, et al. A microRNA expression signature of human solid tumors defines cancer gene targets. Proc Natl Acad Sci USA. 2006;103(7):2257-61.

36. Tagawa H, Karube K, Tsuzuki S, Ohshima K, Seto M. Synergistic action of the microRNA-17 polycistron and Myc in aggressive cancer development. Cancer Sci. 2007;98(9):1482-90.

37. Ota A, Tagawa H, Karnan S, Tsuzuki S, Karpas A, Kira S, et al. Identification and characterization of a novel gene, C13orf25, as a target for 13q31-q32 amplification in malignant lymphoma. Cancer Res. 2004;64 (9):3087-95.

38. Xiao C, Srinivasan L, Calado DP, Patterson HC, Zhang B, Wang J, et al. Lymphoproliferative disease and autoimmunity in mice with increased miR-17-92 expression in lymphocytes. Nat Immunol. 2008;9(4):405-14.

39. He L, Thomson JM, Hemann MT, Hemando-Monge E, Mu D, Goodson S, et al. A microRNA polycistron as a potential human oncogene. Nature. 2005;435(7043): 828-33.

40. Mu P, Han YC, Betel D, Yao E, Squatrito M, Ogrodowski P, et al. Genetic dissection of the miR-17~92 cluster of microRNAs in Myc-induced B-cell lymphomas. Genes Dev. 2009;23(24):2806-11.

41. Sylvestre Y, De Guire V, Querido E, Mukhopadhyay UK, Bourdeau V, Major F, et al. An E2F/miR-20a autoregulatory feedback loop. J Biol Chem. 2007;282(4):2135-43.

42. Rempel RE, Mori S, Gasparetto M, Glozak MA, Andrechek ER, Adler SB, et al. A role for E2F activities in determining the fate of Myc-induced lymphomagenesis. PLoS Genet. 2009;5(9):e1000640.

84

2.7 Publikation 7: „Patient age at diagnosis is associated with the molecular characteristics of diffuse large B-cell lymphoma" Blood. 2012 Feb 23;119(8):1882-7

Klapper W, Kreuz M, Kohler CW, Burkhardt B, Szczepanowski M, Salaverria I, Hummel M, Loeffler M, Pellissery S, Woessmann W, Schwänen C, Trümper L, Wessendorf S, Spang R, Hasenclever D, Siebert R; Molecular Mechanisms in Malignant Lymphomas Network Project of the Deutsche Krebshilfe.

Das diffus großzellige B-Zell-Lymphom (DLBCL) ist eine heterogene Erkrankung [14]. Mittels Genexpressionsanalysen konnten verschiedene molekulare Subgruppen identifiziert werden [1], die sich teilweise prognostisch unterscheiden [44]. Die Inzidenz von DLBCL steigt mit zunehmendem Alter stark an [13]. Bei erwachsenen Patienten sind DLBCL die häufigste Lymphomerkrankung und machen circa 30–40 % aller Non-Hodgkin-Lymphome aus [61]. Bei Kindern und Jugendlichen treten sie seltener auf. Für Kinder und Jugendliche bis zu einem Alter von 18 Jahren erfolgt die Therapie nach Protokollen der pädiatrischen Lymphom-Studiengruppen. Die verwendeten Behandlungsprotokolle für erwachsene Patienten unterscheiden sich recht deutlich von den pädiatrischen. Zusätzlich erfolgt bei adulten Patienten noch eine Einteilung in Patienten ≤60 und >60 Jahren entsprechend dem International Prognostic Index (IPI) [62]. Das Alter bei Erkrankung ist ein starker prognostischer Faktor bei DLCBL, wobei ein höheres Erkrankungsalter prognostisch ungünstig ist [62]. Darüber hinaus gibt es Hinweise auf molekulargenetische Unterschiede zwischen pädiatrischen und adulten DLBCL. Pädiatrische DLBCL sind im Wesentlichen vom Typ GCB und zeigen zentroblastische Morphologie. t(14;18)-Translokationen und hohe Proteinexpression von *BCL2* und *MUM1* treten hingegen selten auf.

Im Rahmen der Studie wurden 364 DLBCL aus dem gesamten Altersspektrum molekular-genetisch charakterisiert. Im Gegensatz zu den vorhergehenden Publikationen stand nicht die Detektion der genetischen Aberrationen im Mittelpunkt, sondern die Analyse der Inzidenz ihres Auftretens bei DLBCL in Abhängigkeit des Erkrankungsalters. Der Großteil aller molekulargenetischen Studien bei DLBCL erfolgte aufgrund der altersabhängigen Inzidenz der Erkrankung ausschließlich an adulten Patienten. Daher war Ziel der Studie zu prüfen, ob sich pädiatrische und adulte DLBCL molekulargenetisch unterscheiden lassen und ob es Hinweise auf plausible Altersgrenzen für die Einteilung der Patienten gibt.

Mittels logistischer Regressionsanalyse konnte gezeigt werden, dass Zugewinne der Chromosomenregionen 1q21, 18q21, 7p22 und 7q21 sowie Translokationen und Zugewinne des *BCL6*-Genlocus (3q27) mit höherem Erkrankungsalter assoziiert sind. Zusätzlich treten der ABC-Subtyp und die hohe Proteinexpression von BCL2 in älteren Patienten gehäuft auf. *IRF4*-Translokationen treten gehäuft bei pädiatrischen Patienten auf. Allerdings handelt es sich bei Lymphomen mit *IRF4*-Bruch vermutlich um eine eigene Krankheitsentität [24]. Mittels Poisson-Regression konnte gezeigt werden, dass die genetische Komplexität, d. h. die Anzahl der aufgetretenen genetischen Aberrationen, mit zunehmendem Erkrankungsalter ansteigt. Die Analyse der klinischen Daten ergab für ABC-Subtyp und *BCL2*-Proteinexpression, trotz Adjustierung für das Erkrankungsalter, eine prognostische Bedeutung.

Die Regressionsanalysen zeigen einen kontinuierlichen Anstieg der genetischen Komplexität und der altersassoziierten Marker. Hinweise auf eine feste Altersgrenze ergaben sich nicht. Stattdessen deutet der kontinuierliche Anstieg der genetischen Komplexität auf ein Tumorevolutionsmodell hin, bei dem genetische Aberrationen über die Zeit aggregiert werden.

Patient age at diagnosis is associated with the molecular characteristics of diffuse large B-cell lymphoma

Wolfram Klapper,[1] *Markus Kreuz,[2] *Christian W. Kohler,[3] *Birgit Burkhardt,[4] *Monika Szczepanowski,[1] *Itziar Salaverria,[5] Michael Hummel,[6] Markus Loeffler,[2] Shoji Pellissery,[5] Wilhelm Woessmann,[4] Carsten Schwänen,[7] Lorenz Trümper,[8] Swen Wessendorf,[7] Rainer Spang,[3] Dirk Hasenclever,[2] and Reiner Siebert,[5] for the Molecular Mechanisms in Malignant Lymphomas Network Project of the Deutsche Krebshilfe

[1]Department of Pathology, Hematopathology Section and Lymph Node Registry, University Hospital Schleswig-Holstein, Campus Kiel/University of Kiel, Kiel, Germany; [2]Institute for Medical Informatics, Statistics and Epidemiology, University of Leipzig, Leipzig, Germany; [3]Institute of Functional Genomics, University of Regensburg, Regensburg, Germany; [4]NHL-BFM Study Center, Department of Pediatric Hematology and Oncology, Justus-Liebig-University, Giessen, Germany; [5]Department of Human Genetics, University Hospital Schleswig-Holstein, Campus Kiel/University of Kiel, Kiel, Germany; [6]Institute of Pathology, Campus Benjamin Franklin, Charité-Universitätsmedizin Berlin, Berlin, Germany; [7]Cytogenetic and Molecular Diagnostics, Internal Medicine III, University Hospital of Ulm, Ulm, Germany; and [8]Department of Hematology and Oncology, Georg-August University of Göttingen, Göttingen, Germany

Diffuse large B-cell lymphoma is the most frequent type of B-cell lymphoma in adult patients but also occurs in children. Patients are currently assigned to therapy regimens based on arbitrarily chosen age limits only (eg, 18 or 60 years) and not biologically justified limits. A total of 364 diffuse large B-cell lymphomas and related mature aggressive B-cell lymphomas other than Burkitt lymphoma from all age groups were analyzed by comprehensive molecular profiling. The probability of several bio-logic features previously reported to be associated with poor prognosis in diffuse large B-cell lymphoma, such as ABC subtype, BCL2 expression, or cytogenetic complexity, increases with age at diagnosis. Similarly, various genetic features, such as *IRF4* translocations, gains in 1q21, 18q21, 7p22, and 7q21, as well as changes in 3q27, including gains and translocations affecting the *BCL6* locus, are significantly associated with patient age, but no cut-offs between age groups could be defined. If age was incorporated in multivariate analyses, genetic complexity lost its prognostic significance, whereas the prognostic impact of ABC subtype and age were additive. Our data indicate that aging is a major determinant of lymphoma biology. They challenge current concepts regarding both prognostic biomarkers and treatment stratification based on strict age cut-offs. (*Blood*. 2012;119(8): 1882-1887)

Introduction

Diffuse large B-cell lymphoma (DLBCL) is a heterogeneous disease composed of different histopathologic and genetic subtypes.[1] Moreover, gene expression profiling has identified several molecular subgroups, of which the activated (ABC) and germinal center (GCB) subtypes of DLBCL are clinically the most relevant.[2,3] Several studies have shown that the GCB subtype is associated with a more favorable prognosis than the ABC subtype.[3-5] Additional molecular biomarkers for prognosis have been identified, including BCL2 protein expression,[6] chromosomal aberrations,[7-9] and genetic complexity.[10] However, these biomarkers for DLBCL were established almost exclusively for lymphomas in adults. Studies investigating biomarkers usually incorporated age as part of the clinical International Prognostic Index (IPI) because age is the strongest predictor of outcome in DLBCL. In the IPI, age is used as a dichotomous variable with a cut-off at 60 years at diagnosis and higher age being an unfavorable risk factor.[11]

DLBCL accounts for approximately 30% of adult lymphomas and 10% of lymphomas diagnosed before the age of 18 years.[2] In clinical practice, children and adolescents up to the age of 18 years are usually treated according to protocols of pediatric lymphoma study groups, whereas patients older than 18 years are treated according to protocols of adult lymphoma study groups.[12] Over the past decades, chemotherapy strategies in pediatric and adult DLBCL study groups have developed very differently, making a direct comparison of clinical results difficult.[12] Nevertheless, the clinical outcome of DLBCL in children is much better than in adults.[12,13] The treatment outcome has been greatly improved in adult patients by adding rituximab to CHOP-like regimens (cyclophosphamide, anthracyclines like doxorubicin, vincristine, and prednisolone every 21 or 14 days).[14] The resulting survival rates for low-risk adult patients (who are predominantly young) are comparable with those for pediatric patients.[15] The latter, however, are treated completely differently, with pediatric regimens, which are short, 5- to 6-day dose-intense courses, including steroids, vincristine, high-dose methotrexate, cyclophosphamide or ifosfamide, doxorubicin, cytarabine, etoposide, and intrathecal therapy for some risk groups.[16,17]

Although the favorable prognosis of childhood and young adult DLBCL may be in part the result of the differences in the treatment protocols or so far unknown host factors, several lines of evidence suggest that DLBCLs in children differ biologically from their adult counterparts: (1) pediatric DLBCLs are almost exclusively of

86

the centroblastic subtype; (2) in childhood, primarily the GCB subtype of DLBCL is observed; (3) the chromosomal translocation t(14;18) involving the *BCL2* gene, which is present in 20% to 30% of adult DLBCL of GCB subtype, is virtually absent before the age of 18 years; and (4) BCL2 or IRF4/MUM1 protein expression is found less frequently in children than adults.[18,19] Overall, DLBCL in children seems to be molecularly more homogeneous than in adults, supporting the hypothesis of an age-dependent pathogenesis and biology. However, the age cut-off of 18 years used in clinical practice seems rather arbitrary and does not reflect DLBCL biology.[20]

In line with the incidence pattern of DLBCL, all large profiling studies published so far on this lymphoma have focused on adult and elderly patients and contain hardly any young adult or even pediatric patients. Thus, they do not allow us to conclude reliably whether differences in treatment strategies, host characteristics, or tumor biology are key to the strong association between age and prognosis. Therefore, we performed a comprehensive molecular characterization of DLBCL covering all age groups to determine possible age cut-offs (eg, between pediatric and adult DLBCL), which might serve as a biologic basis for assigning patients to treatment regimens.

Methods

Study population

From a cohort of 742 samples from the network project Molecular Mechanisms in Malignant Lymphoma (MMML), 364 cases with DLBCL and related mature aggressive B-cell lymphomas other than Burkitt lymphoma were entered in this study based on a selection algorithm outlined in supplemental Figure 1 (available on the *Blood* Web site; see the Supplemental Materials link at the top of the online article). All biopsy specimens were evaluated by a panel of hematopathologists according to the WHO lymphoma classification[1] by a process of independent analysis on single microscopes followed by discussion at a multihead microscope to find a consensus for discrepant cases, as recently described.[5] All cases were part of previous publications with a different focus.[5,21] The study cohort included the histopathologic diagnosis DLBCL, DLBCL in combination with a follicular lymphoma, atypical Burkitt lymphoma, and aggressive mature B-cell lymphoma not otherwise specified. Molecularly defined Burkitt lymphomas were excluded (see supplemental Figure 1 for case selection). All biopsy specimens obtained at relapse were excluded. Supplemental Table 1 and supplemental Figure 1 contain a detailed description of the study population. Because of the retrospective nature of the study and the broad range of age groups incorporated, the therapy applied was heterogeneous.[5,19] The protocols of the MMML network were approved by ethics committees of all participating institutions. A previously published series of DLBCL homogeneously treated with R-CHOP was used as a control cohort (supplemental Figure 2B).[4]

Immunohistochemistry

Immunohistochemical staining was performed on paraffin slides using standard techniques and antibodies against CD20, CD10, BCL2, BCL6, MUM1, and Ki67, as described previously.[5] The stainings were analyzed and scored semiquantitatively by at least 2 observers. In cases where no paraffin-embedded material was available, the stainings were performed on frozen sections. For all analyses, recently published guidelines for internal positive controls were applied.[22]

Gene expression profiling and matrix (array) comparative genomic hybridization

RNA and DNA were extracted from frozen sections (QIAGEN). Affymetrix U133A GeneChip hybridization was performed in accordance with the

manufacturer's recommendations using 5 μg of total RNA, as previously described.[5] The gene expression data from the previously published series[5,19] are available at http://www.ncbi.nlm.nih.gov/geo (GEO accession nos. GSE4475, GSE10172, and GSE22470, respectively). The cell-of-origin signature (ABC, GCB),[3] molecular Burkitt signature,[5] and PAP[23] were assigned to each case based on the gene expression data.

Array-based comparative genomic hybridization was performed in 273 cases by applying a BAC/PAC array containing 2799 DNA fragments.[24,25] The selection of recurrent copy number aberrations is described in supplemental Methods and was performed according to Kreuz et al.[26]

Interphase FISH

Interphase FISH was performed on frozen or paraffin-embedded tissue sections with the use of probes for *IGH, IGK, IGL, MYC, BCL6, IRF4,* and *BCL2* loci.[5,21] Tumor-biopsy specimens in which *MYC* was fused to *IGH, IGK,* or *IGL* (*IG-MYC*) were distinguished from lymphomas with *MYC* breakpoints without fusion of *MYC* to an immunoglobulin locus (non–IG-MYC), as previously described,[5,27] but all lymphomas with *MYC* breaks were combined for the analysis in the current study.

EZH2 mutational analysis

Replacement of a single tyrosine in the SET domain of the EZH2 protein (Tyr641) was tested as previously described.[28] In brief, exon 15 of EZH2 was amplified by PCR using the primers EZH2_6812 (5′-tttgtccccagtccattttc-3′) and EZH2_6813 (5′-tggcaattcatttccaatca-3′), and amplicons were subjected to direct sequencing using the same primers.

Statistical analyses

Differences in the incidence of biologic features were analyzed by 2 × 2 tables and tested using Fisher exact test. This analysis was done to compare children (younger than 15 years) and adults (18 years of age or older). The conditional probability that a biologic marker will be displayed, given the age at diagnosis, was analyzed using logistic regression.

To analyze the association between overall genomic instability and age, a score for genetic complexity was designed. Genetic complexity was calculated as the sum of all detected genomic aberrations per lymphoma sample taking into account: t(14;18) *IGH/BCL2* juxtaposition, *BCL6* translocation, *MYC* translocation, and all copy number aberrations listed in supplemental Table 2. Correlation of age at diagnosis and genetic complexity was analyzed using Poisson regression. Overall survival was defined as time from first day of therapy to death from any cause. Patients without an event in overall survival were censored at the last day with valid information. Overall survival was estimated by the Kaplan-Meier method and compared using the log-rank test. Multivariate analyses were done using Cox proportional-hazard models (see also supplemental Methods).

Results

Different molecular characteristics of pediatric and adult mature aggressive B-cell lymphoma other than Burkitt lymphoma

To determine which molecular features of DLBCL are associated with age, a cohort of 364 mature aggressive B-cell lymphomas diagnosed as DLBLCL, composite follicular lymphoma/DLBCL, or other high-grade B-cell lymphoma other than Burkitt lymphoma was analyzed for molecular features. In a first step, we stratified our cohort into children younger than 15 years (n = 15), adolescents of 15 to 18 years (n = 5), and adults 18 years of age or older (n = 344), as is commonly done in clinical practice, and tested for differences between children and adults in the incidence of biologic features of DLBCL. As already described, chromosomal translocations affecting the *IRF4* locus were significantly more frequent in
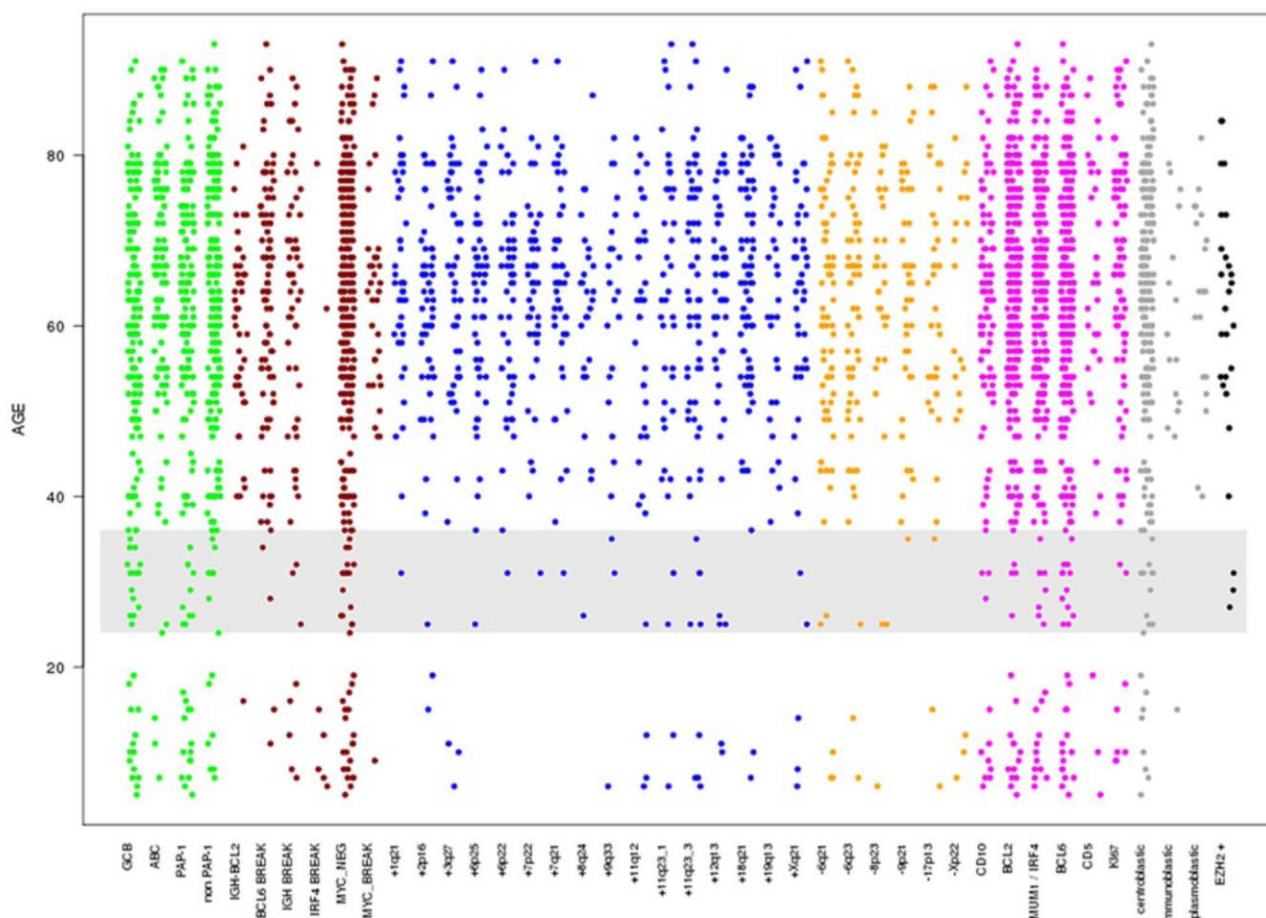
Figure 1. Scatter plot showing the phenotypic variables on the x-axis, with each dot being an event-positive representative at its corresponding age level. The y-axis represents age at diagnosis in years. The variables are not exclusive because a case with an IgH break might also be displayed in as being IgH-BCL2 translocation positive. Note that this visualization is distorted by the skewed distribution of age at diagnosis. The color code indicates membership of each variable to an individual group: green represents gene expression; red, chromosomal translocation/breaks; blue, chromosomal copy number gains; orange, chromosomal copy number losses; magenta, immunohistochemistry markers; gray, morphology; and black, EZH2 Tyr 641 mutation status. The span from 24 to 36 years in which visual inspection reveals a strong increase in several molecular features is highlighted in gray.

children than adults ($P < .001$).[21] In contrast, gains of 2p16 and 6p25 were never detected in children, but in 17% and 15% of adult DLBCLs, respectively ($P = .047$ and $P = .048$, supplemental Table 2 for all variables tested)

Continuous change of molecular characteristics in mature aggressive B-cell lymphoma (other than Burkitt lymphoma) with patient age

To identify whether a clear cut-off can be defined between lymphomas derived from young (pediatric) and old (adult) patients based on the presence of molecular features can be defined, we plotted the different molecular variables according to the age at diagnosis (Figure 1). Visual inspection of Figure 1 suggested that several features additional to that identified compared with that described above may have age-dependent frequencies of occurrence. Moreover, the increase in these features seems to be continuous throughout all age groups and appears to be strongest between 24 to 36 years rather than 18 years (Figure 1). However, the visualization has its pitfalls because the incidence of lymphomas in elderly patients is much higher and because the predominance of elderly patients in our cohort, which reflects the age distribution known for Germany (http://www.krebsregister-sh.de/),

distorts the plot. Therefore, we modeled the conditional probability that a molecular feature will be displayed at a given age at diagnosis using logistic regression analysis (LRA), which eliminates the distorting effect of the underlying age distribution. This analysis determines the likelihood that a lymphoma will show a molecular feature in relation to the patient age at the time of diagnosis. LRA revealed a statistically significant association between increasing age and the cell-of-origin signature of the ABC subtype, BCL2 protein expression, absence of *IRF4* translocations, gains in 1q21, 18q21, 7p22, and 7q21 as well as changes in 3q27, including gains and translocations affecting the *BCL6* locus (Figure 2A; and supplemental Table 3 for all variables tested). Indeed, the genetic complexity, measured as the sum of detectable genomic aberrations, increased continuously with age (Figure 2B). LRA confirmed that there is a significant continuous increase in the ABC probability not only in our dataset but also in an independent case collection ($P = .042$[4]; supplemental Figure 2). Poisson regression showed that the number of chromosomal aberrations in a lymphoma increased significantly with age (Figure 2A) with the shape of the curves of the LRA analyses and the Poisson regression suggesting a continuous increase in the probability of molecular features with patient age rather than a strict border between age groups (Figure 2).
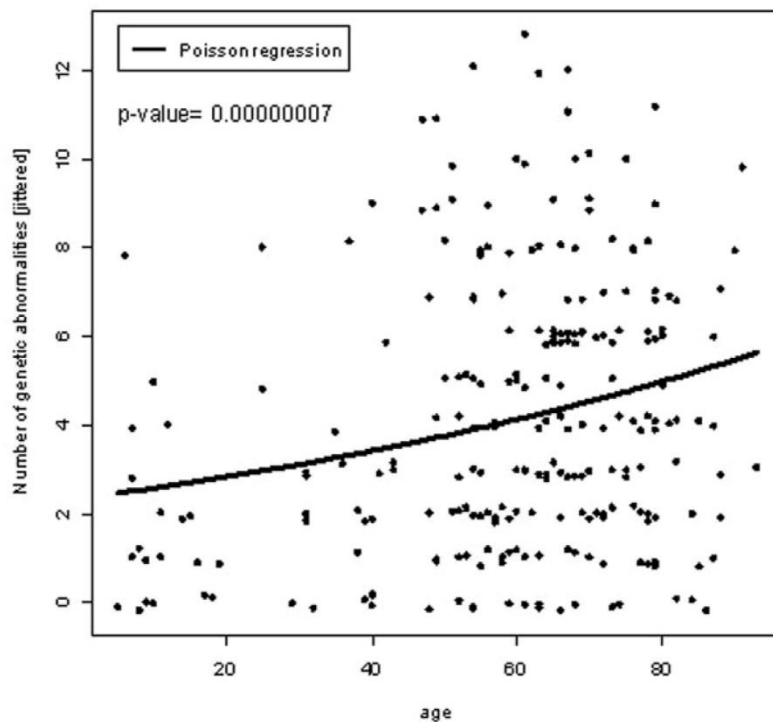
88
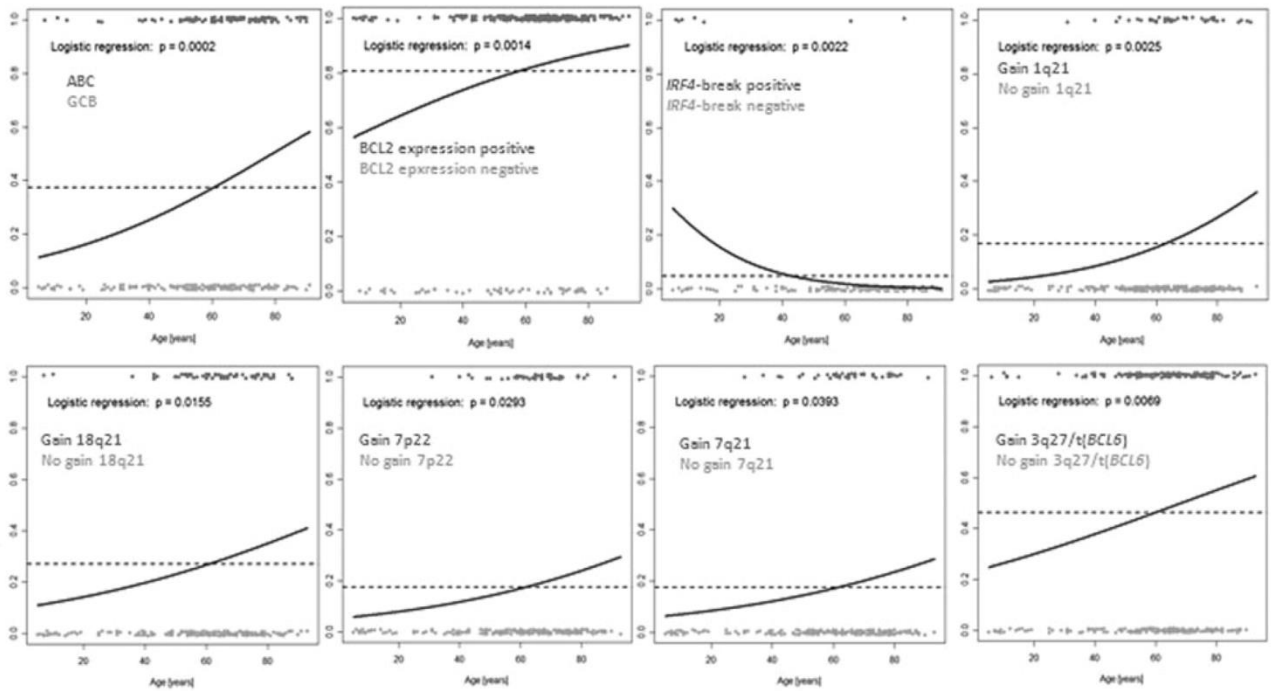
Figure 2. Logistic regression analysis of molecular features and age. (A) Logistic regression analyses of ABC/GCB gene expression groups, immunohistochemical BCL2 expression, IRF4 translocations, 1q21 gains, changes in 3q27 (including gains and translocations of BCL6), 18q21 gains, 7p22 gains, and 7q21 gains. Patients with the respective feature are plotted in red scattered around the horizontal at 1. Patients lacking the respective feature are given in green at the bottom scattered around 0. The dashed line indicates the overall frequency of the feature irrespective of age. The blue logistic regression curve represents the estimated conditional probability of having the feature at a given age at diagnosis. (B) Scatter plot of age at diagnosis (in years) versus genetic complexity calculated as the sum of all detected genomic aberrations per lymphoma sample taking into account: t(14;18) IGH/BCL2 fusion, BCL6 translocation, MYC translocation, and all copy number aberrations listed in supplemental Table 2. The line indicates the Poisson regression curve representing the estimated average number of abnormalities as a function of age at diagnosis.

89

Table 1. Prognostic impact of features associated with age in a univariate analysis and after adjustment for age

Marker	Univariate result		With adjustment for age	
	Relative risk	P	Relative risk	P
ABC GE	2.76	1.6×10^{-7}	2.14	.00013
BCL2 IHC	2.61	.00048	2.36	.0019
IRF4 break	0.18	.09	0.33	.28
1q21+	1.44	.11	1.20	.44
18q21+	1.45	.06	1.23	.29
7p22+	1.63	.02	1.50	.05
7q21+	1.43	.09	1.28	.25
3q aberration	1.43	.05	1.19	.34
Genetic complexity*	1.08	.009	1.05	.12

GE indicates gene expression; and IHC, immunohostochemistry.
*Genetic complexity as a continuous variable.

Interdependence of age and molecular characteristics in relation to prognosis

We tested for prognostic significance in a Cox model combining each molecular feature that was identified by LRA, with patient age. As Table 1 shows, the prognostic impact of the molecular ABC subtype ($P = .0001$) and BCL2 protein expression ($P = .0019$) was maintained in a Cox model, including age. In contrast, all genetic markers lost the significant prognostic impact if age was taken into account. We further validated the prognostic significance of the molecular subtype in a second and completely independent cohort homogeneously treated with immunochemotherapy[4] and confirmed the prognostic information of age and ABC/GCB subtype to be independent and additive (supplemental Figure 2B). Further molecular data, such as BCL2 protein expression, were not accessible for analysis in this cohort.

Discussion

The age cut-off of 18 years is commonly used to stratify patients for pediatric and adult therapy regimens. However, therapeutic approaches to DLBCL in pediatric and adult lymphoma study groups have developed very differently over the past decades. Children and adolescents up to the age of 18 years are treated with regimens that are also applied for Burkitt lymphoma, whereas anthracycline-based therapy in combination with anti-CD20 antibody is the current gold standard for adult patients.[12] This study showed that molecular features of DLBCL are associated with age and that prognostically unfavorable molecular features, such as the ABC subtype and BCL2 expression, increase with patient age at diagnosis. Our data are in line with previously published immuno-phenotypes of pediatric DLBCL.[29] However, our data do not support a pathogenetic dichotomy between DLBCL in children and adults ("2 lymphoma model"), as initially hypothesized. For all features associated with increasing age (ABC subtype, BCL2 protein expression, absence of IRF4 translocations, gains in 1q21, 18q21, 7p22, and 7q21 as well as changes in 3q27, including gains and translocations affecting the BCL6 locus, genetic complexity), there was no clear age cut-off between pediatric and adult DLBCL. Remarkably, the genetic features that increased in likelihood with age in the LRA included changes associated with both ABC (changes 3q27/BCL6, +18q21) and GCB type DLBCL (+1q21, +7p22, +7q21),[30] suggesting that there is an increase in genetic complexity with age that is to some extent independent of the molecular lymphoma subtype. The continuous increase in

genomic complexity with age might suggest an "age evolution model" characterized by a stochastic risk of acquiring genetic aberrations with age. Probably the type of genetic changes that accumulated in a lymphoma-initiating cell determines the manifestation of a certain molecular subtype of DLBCL. In relation to the choice of therapy, our data strongly argue for the development of common treatment protocols for DLBCL in adolescents and young adults because there is no biologic rationale for a clearly definable age limit and the therapy would be tolerated equally well.

Interestingly, differences in molecular features of pediatric and adult disease are not restricted to DLBCL but have recently been demonstrated for lymphoblastic leukemia[31] and follicular lymphoma.[32] However, despite differences in clinical presentation, such as sex distribution, we have failed to demonstrate molecular differences between pediatric and adult Burkitt lymphoma so far.[19] Nevertheless, molecular characterization of malignant tumors occurring in children and adults might be a very fruitful approach for gaining insights into disease biology.

The clinical course after chemotherapy for adult DLBCL varies substantially between individual patients. As a consequence, efforts have been made to establish prognostic biomarkers by analyzing lymphoma biology. However, the IPI, which is based on the clinical features age, lactate dehydrogenase value, performance status, Ann Arbor stage, and number of extranodal involvements, still seems to be the most powerful prognostic tool in DLBCL.[33] In the IPI, age is reflected by categorizing the patients as younger or older than 60 years. Most studies of biomarkers correlate with the IPI as the current prognostic gold standard. The cohort presented here is of limited value for survival analysis because of the heterogeneous treatment applied. Nevertheless, our data suggest that in future studies prognostic biomarkers analyzed in homogeneously treated cohorts should be very carefully analyzed for age association. Because the age at disease onset influences the biology of the lymphoma, prognostic markers might lose their prognostic significance if age is depicted in survival analysis as a continuous parameter. However, we also demonstrated that the prognostic significance of the molecular subtypes of DLBCL is independent of the patient age.

In conclusion, the present study, which represents the largest cohort of DLBCL analyzed by comprehensive molecular profiling so far, challenges current pathogenetic and clinical concepts in DLBCL. It questions the biologic rationale for the use and level of age cut-offs for stratification in clinical trials as well as the application of some age-associated biomarkers.[31,34]

Acknowledgments

Authorship

Contribution: W.K. and R. Siebert designed the project, analyzed the data, and wrote the manuscript; M.K., C.W.K., M.L., D.H., and R. Spang performed statistical analysis; M.S., I.S., M.H., S.P., C.S., and S.W. performed molecular analysis and analyzed the data; and B.B., W.W., and L.T. provided samples and clinical data.

Correspondence: Wolfram Klapper, Department of Pathology, Hematopathology Section and Lymph Node Registry, University Hospital Schleswig-Holstein, Campus Kiel/Christian-Albrecht University Kiel, Michaelisstrasse 11, D-24105 Kiel, Germany; e-mail: wklapper@path.uni-kiel.de.

References

1. Swerdlow SH, Campo E, Harris N, et al. *WHO Classification of Tumors of the Haematopoietic and Lymphoid Tissues.* Lyon, France: IARC; 2008.

2. Shipp MA, Ross KN, Tamayo P, et al. Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. *Nat Med.* 2002;8(1):68-74.

3. Rosenwald A, Wright G, Chan WC, et al. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *N Engl J Med.* 2002;346(25):1937-1947.

4. Lenz G, Wright G, Dave SS, et al. Stromal gene signatures in large-B-cell lymphomas. *N Engl J Med.* 2008;359(22):2313-2323.

5. Hummel M, Bentink S, Berger H, et al. A biologic definition of Burkitt's lymphoma from transcriptional and genomic profiling. *N Engl J Med.* 2006; 354(23):2419-2430.

6. Kramer MH, Hermans J, Parker J, et al. Clinical significance of bcl2 and p53 protein expression in diffuse large B-cell lymphoma: a population-based study. *J Clin Oncol.* 1996;14(7):2131-2138.

7. Klapper W, Stoecklein H, Zeynalova S, et al. Structural aberrations affecting the MYC locus indicate a poor prognosis independent of clinical risk factors in diffuse large B-cell lymphomas treated within randomized trials of the German High-Grade Non-Hodgkin's Lymphoma Study Group (DSHNHL). *Leukemia.* 2008;22(12):2226-2229.

8. Savage KJ, Johnson NA, Ben Neriah S, et al. MYC gene rearrangements are associated with a poor prognosis in diffuse large B-cell lymphoma patients treated with R-CHOP chemotherapy. *Blood.* 2009;114(17):3533-3537.

9. Copie-Bergman C, Gaulard P, Leroy K, et al. Immuno-fluorescence in situ hybridization index predicts survival in patients with diffuse large B-cell lymphoma treated with R-CHOP: a GELA study. *J Clin Oncol.* 2009;27(33):5573-5579.

10. Offit K, Wong G, Filippa DA, Tao Y, Chaganti RS. Cytogenetic analysis of 434 consecutively ascertained specimens of non-Hodgkin's lymphoma: clinical correlations. *Blood.* 1991;77(7):1508-1515.

11. International Non-Hodgkin's Lymphoma Prognostic Factors Project. A predictive model for aggressive non-Hodgkin's lymphoma. *N Engl J Med.* 1993;329(14):987-994.

12. Reiter A, Klapper W. Recent advances in the understanding and management of diffuse large B-cell lymphoma in children. *Br J Haematol.* 2008;142(3):329-347.

13. Woessmann W, Seidemann K, Mann G, et al. The impact of the methotrexate administration schedule and dose in the treatment of children and adolescents with B-cell neoplasms: a report of the BFM group study NHL-BFM95. *Blood.* 2004; 105(3):948-958.

14. Coiffier B, Lepage E, Briere J, et al. CHOP chemotherapy plus rituximab compared with CHOP alone in elderly patients with diffuse large-B-cell lymphoma. *N Engl J Med.* 2002;346(4):235-242.

15. Pfreundschuh M, Trumper L, Osterborg A, et al. CHOP-like chemotherapy plus rituximab versus CHOP-like chemotherapy alone in young patients with good-prognosis diffuse large-B-cell lymphoma: a randomised controlled trial by the MabThera International Trial (MInT) Group. *Lancet Oncol.* 2006;7(5):379-391.

16. Reiter A, Schrappe M, Tiemann M, et al. Improved treatment results in childhood B-cell neoplasms with tailored intensification of therapy: a report of the Berlin-Frankfurt-Munster Group Trial NHL-BFM 90. *Blood.* 1999;94(10):3294-3306.

17. Patte C, Auperin A, Michon J, et al. The Societe Francaise d'Oncologie Pediatrique LMB89 protocol: highly effective multiagent chemotherapy tailored to the tumor burden and initial response in 561 unselected children with B-cell lymphomas and L3 leukemia. *Blood.* 2001;97(11):3370-3379.

18. Oschlies I, Klapper W, Zimmermann M, et al. Diffuse large B-cell lymphoma in pediatric patients belongs predominantly to the germinal-center type B-cell lymphomas: a clinicopathologic analysis of cases included in the German BFM (Berlin-Frankfurt-Munster) Multicenter Trial. *Blood.* 2006; 107(10):4047-4052.

19. Klapper W, Szczepanowski M, Burkhardt B, et al. Molecular profiling of pediatric mature B-cell lymphoma treated in population-based prospective clinical trials. *Blood.* 2008;112(4):1374-1381.

20. Burkhardt B, Zimmermann M, Oschlies I, et al. The impact of age and gender on biology, clinical features and treatment outcome of non-Hodgkin lymphoma in childhood and adolescence. *Br J Haematol.* 2005;131(1):39-49.

21. Salaverria I, Philipp C, Oschlies I, et al. Translocations activating IRF4 identify a subtype of germinal center-derived B-cell lymphoma affecting predominantly children and young adults. *Blood.* 2011;118(1):139-147.

22. de Jong D, Xie W, Rosenwald A, et al. Immunohistochemical prognostic markers in diffuse large B-cell lymphoma: validation of tissue microarray as a prerequisite for broad clinical applications (a study from the Lunenburg Lymphoma Biomarker Consortium). *J Clin Pathol.* 2009;62(2):128-138.

23. Bentink S, Wessendorf S, Schwaenen C, et al. Pathway activation patterns in diffuse large B-cell lymphomas. *Leukemia.* 2008;22(9):1746-1754.

24. Fiegler H, Carr P, Douglas EJ, et al. DNA microarrays for comparative genomic hybridization based on DOP-PCR amplification of BAC and PAC clones. *Genes Chromosomes Cancer.* 2003; 36(4):361-374.

25. Schwaenen C, Nessling M, Wessendorf S, et al. Automated array-based genomic profiling in chronic lymphocytic leukemia: development of a clinical tool and discovery of recurrent genomic alterations. *Proc Natl Acad Sci U S A.* 2004; 101(4):1039-1044.

26. Kreuz M, Rosolowski M, Berger H, et al. Development and implementation of an analysis tool for array-based comparative genomic hybridization. *Methods Inf Med.* 2007;46(5):608-613.

27. Ventura RA, Martin-Subero JI, Jones M, et al. FISH analysis for the detection of lymphoma-associated chromosomal abnormalities in routine paraffin-embedded tissue. *J Mol Diagn.* 2006; 8(2):141-151.

28. Morin RD, Johnson NA, Severson TM, et al. Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. *Nat Genet.* 2010;42(2):181-185.

29. Miles RR, Raphael M, McCarthy K, et al. Pediatric diffuse large B-cell lymphoma demonstrates a high proliferation index, frequent c-Myc protein expression, and a high incidence of germinal center subtype: report of the French-American-British (FAB) international study group. *Pediatr Blood Cancer.* 2008;51(3):369-374.

30. Lenz G, Wright GW, Emre NC, et al. Molecular subtypes of diffuse large B-cell lymphoma arise by distinct genetic pathways. *Proc Natl Acad Sci U S A.* 2008;105(36):13520-13525.

31. Szczepanski T, Harrison CJ, van Dongen JJ. Genetic aberrations in paediatric acute leukaemias and implications for management of patients. *Lancet Oncol.* 2010;11(9):880-889.

32. Oschlies I, Salaverria I, Mahn F, et al. Pediatric follicular lymphoma: a clinico-pathological study of a population-based series of patients treated within the NHL-BFM (Berlin-Frankfurt-Munster) multicenter trials. *Haematologica.* 2010;95(2): 253-259.

33. Ziepert M, Hasenclever D, Kuhnt E, et al. Standard International Prognostic Index remains a valid predictor of outcome for patients with aggressive CD20+ B-cell lymphoma in the rituximab era. *J Clin Oncol.* 2010;28(14):2373-2380.

34. Thunberg U, Amini RM, Linderoth J, Roos G, Enblad G, Berglund M. BCL2 expression in de novo diffuse large B-cell lymphoma partly reflects normal differences in age distribution. *Br J Haematol.* 2009;146(6):683-684.

91

2.8 Kennzeichnung des Eigenanteils für alle eingeschlossenen Publikationen

Im folgenden Abschnitt sind die geleisteten Eigenanteile für alle eingeschlossenen Publikationen stichpunktartig zusammengefasst.

Manuskript 1:
- Konzeption und Entwicklung der Analysemethoden
- Implementierung der Methoden als R-Paket *aCGHPipeline*
- Analyse der Array-CGH-Datensatzes mit *aCGHPipeline*
- Vergleich der Ergebnisse der automatischen Analyse mit der manuellen Interpretation von CS und SW
- Schreiben des Manuskripts (gemeinsam mit DH)

Manuskript 2:
- Analyse des SNP-Array-Datensatzes wie in Abschnitt 1.4 dargestellt
- Anpassung der Qualitäts- und Glättungsparameter an die Charakteristiken des Datensatzes
- Anpassung der HMM-Interpretation des X-Chromosoms von männlichen Patienten zur Optimierung der Detektion von Deletionen des Y-Chromosoms via PAR1

Manuskript 3:
- Analyse des SNP-Array-Datensatzes wie in Abschnitt 1.4 dargestellt
- Anpassung der Qualitäts- und Glättungsparameter an die Charakteristiken des Datensatzes
- Selektion von homozygoten Deletionen und HL-Amplifikationen sowie der rekurrenten Regionen und deren Abgleich mit bekannten Copy-Number-Polymorphismen
- grafische Darstellung der Ergebnisse und Erstellung des zugehörigen Methodenteils sowie der Grafiken für die Publikation

Manuskript 4:
- Analyse des SNP-Array-Datensatzes wie in Abschnitt 1.4 dargestellt
- Anpassung der Qualitäts- und Glättungsparameter an die Charakteristiken des Datensatzes
- Selektion von homozygoten Deletionen und HL-Amplifikationen sowie der rekurrenten Regionen und deren Abgleich mit bekannten Copy-Number-Polymorphismen
- Bestimmung von Gendosiseffekten in rekurrenten Regionen mittels gepaarten Genexpressionsdaten (Affymetrix hgu95av2)
- grafische Darstellung der Ergebnisse und Erstellung des zugehörigen Methodenteils sowie der Grafiken für die Publikation

Manuskript 5:
- Analyse des SNP-Array-Datensatzes wie in Abschnitt 1.4 dargestellt
- Anpassung der Analysepipeline auf die neue Generation der SNP-Arrays (Affymetrix 250k/500k)
- Anpassung der Qualitäts- und Glättungsparameter an die Charakteristiken des Datensatzes
- Selektion von homozygoten Deletionen und HL-Amplifikationen sowie der rekurrenten Regionen und deren Abgleich mit bekannten Copy-Number-Polymorphismen
- grafische Darstellung der Ergebnisse und Erstellung des zugehörigen Methodenteils sowie der Grafiken für die Publikation

Manuskript 6:
- Analyse des SNP-Array-Datensatzes wie in Abschnitt 1.4 dargestellt
- Anpassung der Qualitäts- und Glättungsparameter an die Charakteristiken des Datensatzes
- Zusammenführung der Messergebnisse mit unterschiedlicher Auflösung (500k Sty+Nsp vs. 250k Sty-only)
- Selektion von homozygoten Deletionen und HL-Amplifikationen sowie der rekurrenten Regionen und deren Abgleich mit bekannten Copy-Number-Polymorphismen
- Bestimmung von Gendosiseffekten in rekurrenten Regionen mittels gepaarten Genexpressionsdaten (Affymetrix hgu133a)
- grafische Darstellung der Ergebnisse und Erstellung des zugehörigen Methodenteils sowie der Grafiken für die Publikation

Manuskript 7:
- Analyse der Array-CGH-Daten mit *aCGHPipeline* und Bestimmung der rekurrenten Regionen sowie deren Status in den untersuchten Fällen
- Analyse des Zusammenhanges der Marker (rekurrente Regionen, IHC, molekulare Klassifikationen) und des Erkrankungsalters als dichotome und stetige Variable (gemeinsam mit DH)
- Überlebenszeitanalyse zur Bestimmung der prognostischen Bedeutung der molekularen Marker in Abhängigkeit des Erkrankungsalters (gemeinsam mit DH)
- Zusammenstellung der Ergebnisse und des zugehörigen Methodenteils sowie der Grafiken für das Manuskript

3 Diskussion und Ausblick

Somatische Veränderungen in der DNA-Kopienzahl treten in einer Vielzahl von Tumorerkrankungen auf und spielen eine wichtige Rolle bei der Tumorgenese [63]. Für die Identifikation von Genen, die die Pathogenese dieser Erkrankungen treiben, stellt die Suche nach rekurrenten Kopienzahl-Aberrationen daher einen vielversprechenden Ansatz dar [6]. Die dabei gewonnenen Einblicke in die molekularen Mechanismen, die den Erkrankungen zugrunde liegen, können langfristig auch zu neuen Therapieansätzen führen [6].

Spezifische Aberrationen oder Aberrationsmuster können eine Bedeutung als diagnostische oder prognostische Marker haben [64]. Sie ermöglichen es darüber hinaus Tumore zu klassifizieren oder Vorhersagen über deren biologische Eigenschaften, beispielsweise das Metastasierungsverhalten, zu treffen [27, 65].

Die Untersuchung von Kopienzahlaberrationen in Datensätzen mit großer Fallzahl ermöglicht die Identifikation von typischen evolutionären Mustern in der Pathogenese. Das erlaubt die Abschätzung der Reihenfolge, in welcher die Tumore Aberrationen typischerweise akquirieren [66]. Folglich kann für individuelle Proben der Grad der Progression des zugehörigen Tumors abgeschätzt werden. Dieser ist für viele Tumorentitäten von prognostischer Bedeutung.

Die im Rahmen dieser Arbeit entwickelten Methoden ermöglichen die Auswertung Array-basierter Messungen von Kopienzahlaberrationen. Die mittels Array-CGH (vgl. Abschnitt 1.3) oder SNP-Arrays (vgl. Abschnitt 1.4) gemessenen Rohdaten werden zunächst vorverarbeitet und normalisiert. Anschließend wird durch Segmentierung und Klassifizierung die Detektion von Kopienzahlaberrationen ermöglicht. Die Auswertung von SNP-Arrays erlaubt zusätzlich die Detektion von UPDs. Darüber hinaus umfassen die entwickelten Auswertewerkzeuge die Funktionalität für die gemeinsame Betrachtung mehrerer Tumorproben sowie die Analyse von Regionen, die rekurrent von Aberrationen betroffen sind. Zusätzlich werden umfangreiche Methoden zur grafischen und tabellarischen Präsentation der Analyseergebnisse bereitgestellt. Abschließend wurden Methoden implementiert, mit denen der Zusammenhang von DNA-Kopienzahlaberrationen und RNA-Expression sowie die Assoziation mit klinischen und phänotypischen Eigenschaften der Tumore untersucht werden kann.

Die Auswertewerkzeuge wurden auf Datensätze von Mantelzelllymphomen, Burkitt-Lymphomen, diffus großzelligen B-Zell-Lymphomen, primären ZNS-Lymphomen und peripheren T-Zell-Lymphomen – nicht anderweitig spezifiziert – angewendet. Dabei konnten die untersuchten Lymphomentitäten bezüglich der Muster auftretender Kopienzahlaberrationen charakterisiert werden. Es gelang, zahlreiche genomische Regionen zu detektieren, die rekurrent von Aberrationen betroffen waren. Somit konnten neue Kandidaten für Onko- und Tumorsuppressorgene identifiziert werden, die Evidenz für neue Pathomechanismen in diese Erkrankungen liefern.

Es erfolgten umfangreiche Validierungen der wesentlichen Ergebnisse mittels FISH und für die Array-CGH-Auswertungen eine parallele manuelle Bewertung der experimentellen Daten. Dabei zeigte sich jeweils eine sehr gute Reproduzierbarkeit der Ergebnisse, die mittels der hier entwickelten Werkzeuge bestimmt wurden. Ein systematischer Vergleich von 95 B-Zell-Lymphomen und fünf Zelllinien, die jeweils parallel mit Array-CGH und SNP-Arrays untersucht wurden, ergab eine hohe Übereinstimmung zwischen beiden Messplattformen

und liefert dadurch weitere Evidenz für die Validität der Analysemethoden (vgl. Abschnitt 1.5).

Zusätzlich konnte für eine Reihe von genomischen Regionen, bei denen mittels SNP-Array-Messungen rekurrente Aberrationen detektiert wurden, inzwischen das vermeintliche Zielgen durch Mutationsanalysen identifiziert werden. Beispielhaft sei an dieser Stelle das Gen *ID3* genannt, für das rekurrente Deletionen in Burkitt-Lymphomen gezeigt werden konnten (Manuskript 6; Abschnitt 2.6). Spätere Analysen ergaben, dass *ID3* bei Burkitt-Lymphomen in circa 68 % der Fälle mutiert ist [26, 38, 67]. Mit der Inaktivierung von *ID3* konnte damit ein bisher unbekannter Pathomechanismus für das Burkitt-Lymphom beschrieben werden [68]. Ein weiteres Beispiel stellt die genomische Region Chromosom 19p13.3 dar, für die mittels SNP-Array-Messungen und der hier beschriebenen Auswertemethodik rekurrente homozygote Deletionen in Burkitt-Lymphomen und diffus großzelligen B-Zell-Lymphomen gezeigt werden konnten [39]. Durch exomweite Sequenzierung konnte *TNFSF7* (auch als *CD70* bezeichnet) als mögliches Zielgen der Deletion identifiziert werden, da es rekurrente Mutationen in DLBCL zeigte [69]. Beide Beispiele liefern Evidenz für die Validität der Analyseergebnisse und illustrieren darüber hinaus das Potenzial der Array-basierten Analyse von Kopienzahlveränderungen zur Exploration und Detektion von Genen, die Treiber der Karzinogenese sind. Es zeigt sich jedoch auch, dass die alleinige Betrachtung von strukturellen Aberrationen unzureichend ist. Einerseits sind die aberranten Genomregionen oftmals so groß, dass sie mehrere Gene umfassen. Nicht alle dieser betroffenen Gene haben eine Bedeutung für die Pathogenese und die Identifikation des oder der Kandidaten erfordert oftmals die Betrachtung weiterer molekulargenetischer oder experimenteller Datenebenen. Andererseits wird die Inzidenz, mit der ein Gen in einer bestimmten Tumorentität dereguliert ist, unterschätzt, da Mutationen oder epigenetische Aberrationen alternative Mechanismen für die Deregulation darstellen können.

Die Einsatzmöglichkeiten der hier vorgestellten Auswertewerkzeuge sind nicht auf Lymphome beschränkt. So wurden die Methoden bereits für die Analyse von klarzelligen Nierenzellkarzinomen [27], Glioblastomen [28] und Rhabdoid-Tumoren [40] erfolgreich eingesetzt. Übertragung auf weitere Tumorentitäten oder andere Erkrankungen, bei denen Veränderungen der DNA-Kopienzahl bedeutsam sind, wären ohne größere Anpassungen möglich. Eine Reihe von Kriterien beeinflussen jedoch die Qualität und Zuverlässigkeit der Auswertung.

Kritische Parameter für die Analyse sind eine hohe DNA-Qualität sowie ein hoher Tumorzellgehalt der Proben, da ansonsten ein schlechtes Signal-Rausch-Verhältnis die Analyse beeinträchtigt. Schlechte DNA-Qualität resultiert in einer hohen Streuung der Messwerte, bei niedrigem Tumorzellgehalt werden die durch Kopienzahländerungen hervorgerufenen Signalausschläge zu gering um die Aberrationen zuverlässig zu detektieren. Einen positiven Einfluss auf das Signal-Rausch-Verhältnis bei der Analyse von SNP-Array-Daten hat auch die Verwendung von laboreigenen euploiden Referenzen (vgl. Abschnitt 1.5), da dies zu niedrigerer Streuung der Messwerte führt. Biologische Charakteristiken der Tumore können die Analyse der DNA-Kopienzahl ebenfalls erschweren. Für polyklonale Tumore können Aberrationen, die ausschließlich in einem seltenen Tumorklon auftreten, nicht detektiert werden. Bei polyploiden Tumoren gestaltet sich die Normalisierung schwierig, da die Annahme, dass ein Großteil der Messpunkte normale Kopienzahl (N=2) hat, verletzt ist. Diese Einschränkungen betreffen jedoch alle derzeit verfügbaren Algorithmen.

Abschließend möchte ich an dieser Stelle einen kleinen Ausblick wagen und diskutieren, wie sich die Analyse von DNA-Kopienzahländerungen in Tumoren verändert und welche Rolle sie in Zukunft spielen könnte. Die Entwicklung moderner Hochdurchsatz-Sequenzierung, sogenannter „Next Generation Sequencing"-Verfahren (NGS), ermöglicht neue Einblicke in die genetischen Veränderungen, die Tumorerkrankungen zugrunde liegen [70]. Neben der Detektion von Punktmutationen sowie Insertationen und Deletionen können auch strukturelle Veränderungen wie Kopienzahlveränderungen oder Translokationen aufgedeckt werden. Dabei ermöglicht Next Generation Sequencing eine im Vergleich zu Array-CGH und SNP-Arrays höhere genomische Auflösung [71]. Dies erlaubt die Detektion kleinerer Kopienzahlveränderungen und eine genauere Bestimmung der Segmentgrenzen der Aberrationen.

Die meisten Methoden zur Bestimmung von Kopienzahlveränderungen mittels NGS-Daten basieren dabei auf der Betrachtung der Anzahl überlappender Sequenzen („Reads") pro Genomposition. Diese werden entweder zwischen Tumor- und gepaarter euploider Referenzprobe verglichen oder es erfolgt die reine Betrachtung der Tumorproben, wobei eine aufwendige Normalisierung für den GC-Gehalt und regionale Unterschiede in der Effizienz des Alignments erforderlich ist [72]. Danach erfolgt in der Regel eine Segmentierung mit anschließender Klassifikation der Segmente. Mit zunehmender Verbreitung der NGS-Analysen wurden eine Reihe von Analysemethoden für die Detektion von Kopienzahlaberrationen in NGS-Datensätzen etabliert. Ein aktueller Überblick über die verfügbaren Methoden und ein Vergleich deren Performance sind in Duan et al. zusammengestellt [73].

Es ist davon auszugehen, dass NGS bei der Detektion von Kopienzahlaberrationen in Zukunft eine immer größere Rolle spielen wird. Dies liegt neben der höheren Auflösung und der gleichzeitigen Bestimmung von Mutationen und Translokationen auch an den stetig sinkenden Kosten der NGS-Analysen. Während in aktuellen Studien NGS-Analysen häufig durch SNP-Array-Experimente zur Validierung sowie Qualitätskontrolle ergänzt werden, wird die Bedeutung von Array-basierten Methoden vermutlich langfristig schwinden. Trotzdem ist davon auszugehen, dass Array-basierte Methoden in näherer Zukunft weiterhin große Bedeutung sowohl für Validierungsexperimente, als auch für die Translation von explorativen Analyseergebnissen in die klinische bzw. diagnostische Routine haben werden.

Mit den im Rahmen dieser Arbeit entwickelten Auswertewerkzeugen kann daher auch weiterhin ein relevanter Beitrag zur Aufklärung von somatischen Kopienzahlaberrationen in Tumorerkrankungen geleistet werden. Des Weiteren können die in Abschnitt 1.6 beschriebenen Methoden zur Korrelation von DNA-Kopienzahlaberrationen und RNA-Expression sowie die Assoziation mit klinischen und phänotypischen Faktoren auch auf NGS-basierte Kopienzahldaten übertragen werden.

Große internationale Verbünde wie das „International Cancer Genome Consortium" (ICGC) oder „The Cancer Genome Atlas" (TCGA) haben sich zum Ziel gesetzt, neben den Mutationspektren und den strukturellen Aberrationen auch das Transkriptom und das Methylom von Tumoren zu untersuchen. Dabei werden beispielsweise im ICGC-Konsortium für 50 Tumorentitäten jeweils 500 Tumorproben mittels NGS analysiert [2]. Eine zentrale Fragestellung dieser Untersuchungen ist, wie unterschiedliche genetische und epigenetische Veränderungen in den Tumorentitäten zusammenwirken und zur Karzinogenese führen.

Die gemeinsame Betrachtung mehrerer molekulargenetischer Datenebenen erlaubt es, das Zusammenwirken mehrerer Aberrationen, beispielsweise die Mutation eines Genes bei

gleichzeitiger Deaktivierung der zweiten intakten Kopie durch Deletionen oder Uniparentale Disomie, zu analysieren. Des Weiteren kann die Aktivierung oder Deaktivierung von an der Karzinogenese beteiligten Genen in den Tumoren über unterschiedliche genetische oder epigenetische Mechanismen erfolgen. Eine weitere Herausforderung ist, dass die Tumorzellen relevante Regulationskreisläufe an mehreren Stellen aushebeln können, wodurch eine Betrachtung einzelner Gene unzureichend ist und stattdessen die Betrachtung von Regulationspathways erfolgen muss. Die im Rahmen dieser Arbeit erarbeiteten Methoden für die integrative Analyse mehrerer Datenebenen stellen nur einen kleinen, ersten Schritt dar. Die Entwicklung von bioinformatischen Methoden, die es erlauben, alle gemessenen molekulargenetischen Datenebenen grundlegend zusammenzuführen und zu integrieren, stellt eine der größten zukünftigen Herausforderungen dar.

4 Zusammenfassung

Dissertation zur Erlangung des akademischen Grades:
Dr. rer. med.

Titel:
Entwicklung und Implementierung von Auswertungswerkzeugen für Hochdurchsatz-DNA-Kopienzahl-Analysen und deren Anwendung auf Lymphomdaten

Eingereicht von:
Dipl. Inf. Markus Kreuz

Angefertigt am:
Institut für medizinische Informatik, Statistik und Epidemiologie der medizinischen Fakultät der Universität Leipzig

Betreut von:
Prof. Dr. Markus Löffler

Datum der Einreichung:
24. April 2014

Veränderungen in der DNA-Kopienzahl spielen eine wichtige Rolle bei verschiedenen Erkrankungen, insbesondere bei der Pathogenese und Progression von malignen Tumoren. Änderungen der DNA-Kopienzahl können die Expression von betroffenen Genen beeinflussen, Zugewinne können zur Aktivierung von Onkogenen, Deletionen zur Inaktivierung von Tumorsuppressorgenen führen. Die Detektion von Aberrationen der DNA-Kopienzahl in Tumorproben kann somit Aufschlüsse über die molekulargenetischen Mechanismen bei der Krankheitsentstehung liefern. Darüber hinaus können detektierte Kopienzahlveränderungen als diagnostische oder prognostische Marker eingesetzt werden und die Grundlage für neue therapeutische Ansätze bilden.

Array-CGH- und SNP-Array-Messungen sind zwei Technologien, die die genomweite Messung der DNA-Kopienzahl in einem einzigen Experiment ermöglichen. Die bei der Messung entstehenden Datensätze sind komplex und deren Auswertung erfordert Unterstützung durch automatische Methoden.

Im Rahmen dieser Promotionsarbeit wurden Methoden für die Kopienzahlanalyse auf Basis von Array-CGH- und SNP-Array-Daten entwickelt. Die entwickelten Methoden wurden für die Auswertung umfangreicher Datensätze von Non-Hodgkin-Lymphomen eingesetzt. Die Resultate dieser Arbeit sind unter anderem in sieben Publikationen eingeflossen, die die Grundlage der Promotionsarbeit bilden.

Die Ergebnisse der Promotionsarbeit umfassen die entwickelten Methoden (technische Ergebnisse) und biologischen Ergebnisse aus der Applikation der Methoden auf die Lymphomdatensätze (biologische und klinische Ergebnisse). Die wichtigsten Resultate sind im folgenden Abschnitt in Thesenform zusammengestellt:

Die wichtigsten technischen Ergebnisse dieser Arbeit sind:

1) Es wurde ein neues Auswertewerkzeug für Array-CGH-Daten konzipiert und entwickelt, welches alle Analyseschritte vom Einlesen der Daten über die Analyse der Einzelchips bis zur gemeinsamen Analyse von multiplen Chips unterstützt. Das Auswertewerkzeug wurde in der Statistiksoftware R implementiert und in der Zeitschrift „Methods of Information in Medicine" veröffentlicht (vgl. Manuskript 1; Abschnitt 2.1).

2) Für die Kopienzahlanalyse mit SNP-Arrays wurde eine neue Auswertepipeline entwickelt. Sie umfasst die Genotypisierung, Qualitätsanalyse, die Detektion von LOH und die Kopienzahlanalyse für Einzelchips und die Aggregation und integrative Analyse multipler Arrays.

3) Messungen mit der Array-CGH- und SNP-Array-Plattform wurden an einem Kollektiv von 96 paarweise gemessenen Lymphomproben verglichen. Für die Array-CGH-Messung kam dabei ein BAC-Array (circa 2800 Messpunkte), für die SNP-Array-Messungen der Affymetrix 250k-Sty-Array (circa 238.000 Messpunkte) zum Einsatz. Die wichtigsten Ergebnisse dieses Vergleichs waren:
   i. Gute Übereinstimmung der Array-CGH- und SNP-Array-Messung.
   ii. Deutlich höhere Markeranzahl für SNP-Array-Messungen, aber gleichzeitig erhöhte Streuung der Signale. Die effektive, für Streuung adjustierte Auflösung war für die SNP-Plattform circa zwölffach höher als für die CGH-Plattform. Folglich ist die SNP-Plattform überlegen, da mittels SNP-Array-Messung zusätzlich auch die Detektion von LOH möglich ist.
   iii. Die Reproduzierbarkeit der Ergebnisse bei SNP-Array-Replikations-hybridisierungen in verschiedenen Labors war gut. Laboreigene euploide Kontrollen reduzieren die Streuung der Messungen jedoch deutlich, sodass bei Aufteilung der Messungen auf mehrere Labors jeweils getrennte laboreigene Referenzfälle hybridisiert werden sollten.

4) Es wurden neue Methoden entwickelt, die den Vergleich von Kopienzahlaberrationen zwischen zwei Gruppen und die Analyse des Zusammenhangs von Kopienzahlaberrationen und stetigen Einflussgrößen, beispielsweise dem Erkrankungsalter, unterstützen (vgl. Manuskript 7; Abschnitt 2.7). Darüber hinaus wurden Methoden entwickelt, die eine gemeinsame Analyse von Kopienzahl- und Genexpressionsdaten unterstützen (vgl. Manuskripte 1 und 6; Abschnitte 2.1 und 2.6). Dies erlaubt die Identifikation von Genen, deren Expression durch Kopienzahlaberrationen dereguliert wird.

Die wichtigsten biologischen und klinischen Ergebnisse der Arbeit sind:

1) Mittels Affymetrix 100k-SNP-Arrays wurden 26 MCL-Proben sowie sechs MCL-Zelllinien auf Kopienzahlaberrationen untersucht. Die dabei beobachteten Aberrationsmuster zeigten gute Konkordanz bezüglich der rekurrenten Aberrationen zu vorherigen Kopienzahlstudien beim MCL. Im Rahmen der Studie wurde gezeigt, dass mit *TP53*, *MAP2* und *MAP6* mehrere Gene Aberrationen zeigen, die mit der Organisation von Mikrotubuli assoziiert sind. Darüber hinaus konnte durch die

Kopienzahlanalyse der PAR1-Region indirekt der Verlust von Chromosom Y als häufige sekundäre Aberrationen im MCL gezeigt werden (vgl. Manuskripte 2 und 3; Abschnitte 2.2 und 2.3).

2) 19 PZNSL-Fälle wurden mit Affymetrix 100k-SNP-Arrays untersucht. Die Analyse der Kopienzahlaberrationen ergab elf rekurrente Deletionen und zehn rekurrente Zugewinne. Darüber hinaus zeigten sich rekurrente UPD-Regionen in den Chromosomenarmen 6p (>40 %) und 9p (>20 %). Im untersuchten Datensatz war die Region 6p21.32 am häufigsten betroffen. Insgesamt traten sieben einfache Deletionen, zwei homozygote Deletionen und sieben UPDs auf, sodass für die Region in 14 von 19 (74 %) untersuchten Fällen genomische Aberrationen gezeigt werden konnten. Das illustriert auch, dass verschiedene genetische Mechanismen an der Deaktivierung bestimmter Tumorsuppressorgene beteiligt sein können und eine ausschließliche Betrachtung der Kopienzahl unzureichend ist (vgl. Manuskript 4; Abschnitt 2.6).

3) In einer weiteren Studie wurden 49 periphere T-Zell-Lymphome mit Affymetrix 250k-Sty-SNP-Arrays untersucht. 25 der analysierten Proben zeigten kaum bzw. keine Aberrationen, was konsistent mit den Ergebnissen vergleichbarer Studien ist. Für die Region 2p15–16 konnten vier einfache Zugewinne und eine High-Level-Amplifikation gezeigt werden, die das Gen *REL* umfassen. Amplifikationen von *REL* wurden für PTCL-NOS bisher nicht beschrieben, allerdings für andere Lymphome, beispielsweise DLBCL, gezeigt. Weitere rekurrente Aberrationen waren Zugewinne von Chromosom 7 sowie die Deletion von 10p11. UPD-Regionen zeigten hingegen nur geringe Rekurrenz.

4) Für 39 Lymphome, die mittels Genexpressionsdaten als Burkitt-Lymphome klassifiziert wurden, erfolgte mit hochauflösenden SNP-Arrays die Beschreibung der Kopienzahlaberrationen. Mittels gepaarter Genexpressionsdaten konnte für 17,5 % der Probesets in rekurrenten Regionen ein signifikanter Gendosiseffekt gezeigt werden. Insgesamt konnten 32 High-Level-Amplifikationen und sechs homozygote Deletionen nachgewiesen werden. Die Region Chr1:23538204–23657473, die unter anderem auch das Gen *ID3* umfasst, war als einzige mehrfach von homozygoten Deletionen betroffen. Im Rahmen des ICGC-Projektes Molecular Mechanisms in Malignant Lymphoma by Sequencing (MMML-Seq) konnte in nachfolgenden Analysen gezeigt werden, dass *ID3* in 68 % (36/53) der mBL-Fälle mutiert ist [26] (vgl. Manuskript 6; Abschnitt 2.6).

5) In einer Analyse von Array-CGH-Daten von 364 DLBC-Lymphomen konnte gezeigt werden, dass in DLBCL Zugewinne der Chromosomenregionen 1q21, 18q21, 7p22 und 7q21 sowie Translokationen und Zugewinne des *BCL6*-Genlokus (3q27) mit höherem Erkrankungsalter assoziiert sind. Darüber hinaus zeigte sich ein signifikanter Anstieg der genetischen Komplexität mit zunehmendem Erkrankungsalter. Die Regressionsanalysen zeigen dabei einen kontinuierlichen Anstieg der genetischen Komplexität und der altersassoziierten Marker. Hinweise auf eine feste Altersgrenze ergaben sich nicht. Stattdessen deutet der kontinuierliche Anstieg der genetischen Komplexität auf ein Tumorevolutionsmodell hin, bei dem genetische Aberrationen über die Zeit aggregiert werden (vgl. Manuskript 7; Abschnitt 2.7).

5 Referenzen

1. Alizadeh, A.A., Eisen, M.B., Davis, R.E., Ma, C., Lossos, I.S., Rosenwald, A., Boldrick, J.C., Sabet, H., Tran, T., Yu, X., Powell, J.I., Yang, L., Marti, G.E., Moore, T., Hudson, J., Jr., Lu, L., Lewis, D.B., Tibshirani, R., Sherlock, G., Chan, W.C., Greiner, T.C., Weisenburger, D.D., Armitage, J.O., Warnke, R., Levy, R., Wilson, W., Grever, M.R., Byrd, J.C., Botstein, D., Brown, P.O., and Staudt, L.M., *Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling.* Nature, 2000. **403**(6769): p. 503-11.

2. Hudson, T.J., Anderson, W., Artez, A., Barker, A.D., Bell, C., Bernabe, R.R., Bhan, M.K., Calvo, F., Eerola, I., Gerhard, D.S., Guttmacher, A., Guyer, M., Hemsley, F.M., Jennings, J.L., Kerr, D., Klatt, P., Kolar, P., Kusada, J., Lane, D.P., Laplace, F., Youyong, L., Nettekoven, G., Ozenberger, B., Peterson, J., Rao, T.S., Remacle, J., Schafer, A.J., Shibata, T., Stratton, M.R., Vockley, J.G., Watanabe, K., Yang, H., Yuen, M.M., Knoppers, B.M., Bobrow, M., Cambon-Thomsen, A., Dressler, L.G., Dyke, S.O., Joly, Y., Kato, K., Kennedy, K.L., Nicolas, P., Parker, M.J., Rial-Sebbag, E., Romeo-Casabona, C.M., Shaw, K.M., Wallace, S., Wiesner, G.L., Zeps, N., Lichter, P., Biankin, A.V., Chabannon, C., Chin, L., Clement, B., de Alava, E., Degos, F., Ferguson, M.L., Geary, P., Hayes, D.N., Johns, A.L., Kasprzyk, A., Nakagawa, H., Penny, R., Piris, M.A., Sarin, R., Scarpa, A., van de Vijver, M., Futreal, P.A., Aburatani, H., Bayes, M., Botwell, D.D., Campbell, P.J., Estivill, X., Grimmond, S.M., Gut, I., Hirst, M., Lopez-Otin, C., Majumder, P., Marra, M., McPherson, J.D., Ning, Z., Puente, X.S., Ruan, Y., Stunnenberg, H.G., Swerdlow, H., Velculescu, V.E., Wilson, R.K., Xue, H.H., Yang, L., Spellman, P.T., Bader, G.D., Boutros, P.C., Flicek, P., Getz, G., Guigo, R., Guo, G., Haussler, D., Heath, S., Hubbard, T.J., Jiang, T., Jones, S.M., Li, Q., Lopez-Bigas, N., Luo, R., Muthuswamy, L., Ouellette, B.F., Pearson, J.V., Quesada, V., Raphael, B.J., Sander, C., Speed, T.P., Stein, L.D., Stuart, J.M., Teague, J.W., Totoki, Y., Tsunoda, T., Valencia, A., Wheeler, D.A., Wu, H., Zhao, S., Zhou, G., Lathrop, M., Thomas, G., Yoshida, T., Axton, M., Gunter, C., Miller, L.J., Zhang, J., Haider, S.A., Wang, J., Yung, C.K., Cros, A., Liang, Y., Gnaneshan, S., Guberman, J., Hsu, J., Chalmers, D.R., Hasel, K.W., Kaan, T.S., Lowrance, W.W., Masui, T., Rodriguez, L.L., Vergely, C., Bowtell, D.D., Cloonan, N., deFazio, A., Eshleman, J.R., Etemadmoghadam, D., Gardiner, B.B., Kench, J.G., Sutherland, R.L., Tempero, M.A., Waddell, N.J., Wilson, P.J., Gallinger, S., Tsao, M.S., Shaw, P.A., Petersen, G.M., Mukhopadhyay, D., DePinho, R.A., Thayer, S., Shazand, K., Beck, T., Sam, M., Timms, L., Ballin, V., Lu, Y., Ji, J., Zhang, X., Chen, F., Hu, X., Yang, Q., Tian, G., Zhang, L., Xing, X., Li, X., Zhu, Z., Yu, Y., Yu, J., Tost, J., Brennan, P., Holcatova, I., Zaridze, D., Brazma, A., Egevard, L., Prokhortchouk, E., Banks, R.E., Uhlen, M., Viksna, J., Ponten, F., Skryabin, K., Birney, E., Borg, A., Borresen-Dale, A.L., Caldas, C., Foekens, J.A., Martin, S., Reis-Filho, J.S., Richardson, A.L., Sotiriou, C., Thoms, G., van't Veer, L., Birnbaum, D., Blanche, H., Boucher, P., Boyault, S., Masson-Jacquemier, J.D., Pauporte, I., Pivot, X., Vincent-Salomon, A., Tabone, E., Theillet, C., Treilleux, I., Bioulac-Sage, P., Decaens, T., Franco, D., Gut, M., Samuel, D., Zucman-Rossi, J., Eils, R., Brors, B., Korbel, J.O., Korshunov, A., Landgraf, P., Lehrach, H., Pfister, S., Radlwimmer, B., Reifenberger, G., Taylor, M.D., von Kalle, C., Majumder, P.P., Pederzoli, P., Lawlor, R.A., Delledonne, M., Bardelli, A., Gress, T., Klimstra, D., Zamboni, G., Nakamura, Y., Miyano, S., Fujimoto, A., Campo, E., de Sanjose, S., Montserrat, E., Gonzalez-Diaz, M., Jares, P., Himmelbauer, H., Bea, S., Aparicio, S., Easton, D.F., Collins, F.S., Compton, C.C., Lander, E.S., Burke, W., Green, A.R., Hamilton, S.R., Kallioniemi, O.P., Ley, T.J., Liu, E.T. and Wainwright, B.J., *International network of cancer genome projects.* Nature. **464**(7291): p. 993-8.

3. Hummel, M., Bentink, S., Berger, H., Klapper, W., Wessendorf, S., Barth, T.F., Bernd, H.W., Cogliatti, S.B., Dierlamm, J., Feller, A.C., Hansmann, M.L., Haralambieva, E., Harder, L., Hasenclever, D., Kuhn, M., Lenze, D., Lichter, P., Martin-Subero, J.I., Moller, P., Muller-Hermelink, H.K., Ott, G., Parwaresch, R.M., Pott, C., Rosenwald, A., Rosolowski, M., Schwaenen, C., Sturzenhofecker, B., Szczepanowski, M., Trautmann, H., Wacker, H.H., Spang, R., Loeffler, M., Trumper, L., Stein, H., and Siebert, R., *A biologic definition of Burkitt's*

lymphoma from transcriptional and genomic profiling. N Engl J Med, 2006. **354**(23): p. 2419-30.

4.   Albertson, D. and Pinkel, D., *Genomic microarrays in human genetic disease and cancer.* Hum Mol Genet, 2003. **12**: p. 145-152.

5.   Picard, F., Robin, S., Lavielle, M., Vaisse, C., and Daudin, J.-J., *A statistical approach for array CGH data analysis.* BMC Bioinformatics, 2005. **6** (:27).

6.   Beroukhim, R., Mermel, C.H., Porter, D., Wei, G., Raychaudhuri, S., Donovan, J., Barretina, J., Boehm, J.S., Dobson, J., Urashima, M., Mc Henry, K.T., Pinchback, R.M., Ligon, A.H., Cho, Y.J., Haery, L., Greulich, H., Reich, M., Winckler, W., Lawrence, M.S., Weir, B.A., Tanaka, K.E., Chiang, D.Y., Bass, A.J., Loo, A., Hoffman, C., Prensner, J., Liefeld, T., Gao, Q., Yecies, D., Signoretti, S., Maher, E., Kaye, F.J., Sasaki, H., Tepper, J.E., Fletcher, J.A., Tabernero, J., Baselga, J., Tsao, M.S., Demichelis, F., Rubin, M.A., Janne, P.A., Daly, M.J., Nucera, C., Levine, R.L., Ebert, B.L., Gabriel, S., Rustgi, A.K., Antonescu, C.R., Ladanyi, M., Letai, A., Garraway, L.A., Loda, M., Beer, D.G., True, L.D., Okamoto, A., Pomeroy, S.L., Singer, S., Golub, T.R., Lander, E.S., Getz, G., Sellers, W.R., and Meyerson, M., *The landscape of somatic copy-number alteration across human cancers.* Nature. **463**(7283): p. 899-905.

7.   Knippers, R., *Molekulare genetik*. 2001: Thieme Verlag.

8.   Alberts, B., Bray, D., and Lewis, J., *Molecular Biology of the Cell*. 4th ed. 2002: Taylor & Francis.

9.   Munk, K., *Grundstudium Biologie - Genetik*. 2001: Spektrum Verlag.

10.   Magrath, I., *The Non-Hodgkin's Lymphomas*. 1997: Arnold.

11.   Alberts, B., Johnson, A., and Lewis, J., *Molecular Biology of the Cell*. 5th ed. 2008: Taylor & Francis.

12.   Murphy, K., Travers, P., and Walport, M., *Janeway's Immunobiology*. 7th ed. 2008: Taylor & Francis.

13.   Morton, L.M., Wang, S.S., Devesa, S.S., Hartge, P., Weisenburger, D.D., and Linet, M.S., *Lymphoma incidence patterns by WHO subtype in the United States, 1992-2001.* Blood, 2006. **107**(1): p. 265-76.

14.   Swerdlow SH, I.A.f.R.o.C., World Health Organization, *WHO classification of tumours of haematopoietic and lymphoid tissues*. 2008.

15.   Bilke, S., Chen, Q.R., Whiteford, C.C., and Khan, J., *Detection of low level genomic alterations by comparative genomic hybridization based on cDNA micro-arrays.* Bioinformatics, 2005. **21**(7): p. 1138-45.

16.   Ihaka, R.G., R.R., *A Language for Data Analysis and Graphics.* Journal of Computational and Graphical Statistics, 1996. **5 (3)**: p. 299-314.

17.   Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., Andrews, T.D., Fiegler, H., Shapero, M.H., Carson, A.R., Chen, W., Cho, E.K., Dallaire, S., Freeman, J.L., Gonzalez, J.R., Gratacos, M., Huang, J., Kalaitzopoulos, D., Komura, D., MacDonald, J.R., Marshall, C.R., Mei, R., Montgomery, L., Nishimura, K., Okamura, K., Shen, F., Somerville, M.J., Tchinda, J., Valsesia, A., Woodwark, C., Yang, F., Zhang, J., Zerjal, T., Armengol, L., Conrad, D.F., Estivill, X., Tyler-Smith, C., Carter, N.P., Aburatani, H., Lee, C., Jones, K.W., Scherer, S.W., and Hurles, M.E., *Global variation in copy number in the human genome.* Nature, 2006. **444**(7118): p. 444-54.

18.   Fridlyand, J., Snijders, A.M., Pinkel, D., Albertson, D.G., and Jain, A.N., *Hidden Markov models approach to the analysis of array CGH data.* Journal of Multivariate Analysis, 2004. **90**(1): p. 132-153.

19.   Olshen, A.B., Venkatraman, E.S., Lucito, R., and Wigler, M., *Circular binary segmentation for the analysis of array-based DNA copy number data.* Biostatistics, 2004. **5**(4): p. 557-72.

20.   Bentink, S., Wessendorf, S., Schwaenen, C., Rosolowski, M., Klapper, W., Rosenwald, A., Ott, G., Banham, A.H., Berger, H., Feller, A.C., Hansmann, M.L., Hasenclever, D., Hummel, M., Lenze, D., Moller, P., Stuerzenhofecker, B., Loeffler, M., Truemper, L., Stein, H., Siebert, R., and Spang, R., *Pathway activation patterns in diffuse large B-cell lymphomas.* Leukemia, 2008. **22**(9): p. 1746-54.

21.   Klapper, W., Kreuz, M., Kohler, C.W., Burkhardt, B., Szczepanowski, M., Salaverria, I., Hummel, M., Loeffler, M., Pellissery, S., Woessmann, W., Schwanen, C., Trumper, L., Wessendorf, S., Spang, R., Hasenclever, D., and Siebert, R., *Patient age at diagnosis is associated with the molecular characteristics of diffuse large B-cell lymphoma.* Blood, 2012. **119**(8): p. 1882-7.

22.   Klapper, W., Szczepanowski, M., Burkhardt, B., Berger, H., Rosolowski, M., Bentink, S., Schwaenen, C., Wessendorf, S., Spang, R., Moller, P., Hansmann, M.L., Bernd, H.W., Ott, G., Hummel, M., Stein, H., Loeffler, M., Trumper, L., Zimmermann, M., Reiter, A., and Siebert, R., *Molecular profiling of pediatric mature B-cell lymphoma treated in population-based prospective clinical trials.* Blood, 2008. **112**(4): p. 1374-81.

23.   Martin-Subero, J.I., Kreuz, M., Bibikova, M., Bentink, S., Ammerpohl, O., Wickham-Garcia, E., Rosolowski, M., Richter, J., Lopez-Serra, L., Ballestar, E., Berger, H., Agirre, X., Bernd, H.W., Calvanese, V., Cogliatti, S.B., Drexler, H.G., Fan, J.B., Fraga, M.F., Hansmann, M.L., Hummel, M., Klapper, W., Korn, B., Kuppers, R., Macleod, R.A., Moller, P., Ott, G., Pott, C., Prosper, F., Rosenwald, A., Schwaenen, C., Schubeler, D., Seifert, M., Sturzenhofecker, B., Weber, M., Wessendorf, S., Loeffler, M., Trumper, L., Stein, H., Spang, R., Esteller, M., Barker, D., Hasenclever, D., and Siebert, R., *New insights into the biology and origin of mature aggressive B-cell lymphomas by combined epigenomic, genomic, and transcriptional profiling.* Blood, 2009. **113**(11): p. 2488-97.

24.   Salaverria, I., Philipp, C., Oschlies, I., Kohler, C.W., Kreuz, M., Szczepanowski, M., Burkhardt, B., Trautmann, H., Gesk, S., Andrusiewicz, M., Berger, H., Fey, M., Harder, L., Hasenclever, D., Hummel, M., Loeffler, M., Mahn, F., Martin-Guerrero, I., Pellissery, S., Pott, C., Pfreundschuh, M., Reiter, A., Richter, J., Rosolowski, M., Schwaenen, C., Stein, H., Trumper, L., Wessendorf, S., Spang, R., Kuppers, R., Klapper, W., and Siebert, R., *Translocations activating IRF4 identify a subtype of germinal center-derived B-cell lymphoma affecting predominantly children and young adults.* Blood, 2011. **118**(1): p. 139-47.

25.   Salaverria, I., Martin-Guerrero, I., Burkhardt, B., Kreuz, M., Zenz, T., Oschlies, I., Arnold, N., Baudis, M., Bens, S., García-Orad, A., Lisfeld, J., Schwaenen, C., Szczepanowski, M., Wessendorf, S., Pfreundschuh, M., Trümper, L., Klapper, W., and Siebert, R., *High resolution copy number analysis of IRF4 translocation-positive diffuse large B-cell and follicular lymphomas.* Genes Chromosomes Cancer, 2013. **52**(2):150-5.

26.   Richter, J., Schlesner, M., Hoffmann, S., Kreuz, M., Leich, E., Burkhardt, B., Rosolowski, M., Ammerpohl, O., Wagener, R., Bernhart, S.H., Lenze, D., Szczepanowski, M., Paulsen, M., Lipinski, S., Russell, R.B., Adam-Klages, S., Apic, G., Claviez, A., Hasenclever, D., Hovestadt, V., Hornig, N., Korbel, J.O., Kube, D., Langenberger, D., Lawerenz, C., Lisfeld, J., Meyer, K., Picelli, S., Pischimarov, J., Radlwimmer, B., Rausch, T., Rohde, M., Schilhabel, M., Scholtysik, R., Spang, R., Trautmann, H., Zenz, T., Borkhardt, A., Drexler, H.G., Moller, P., Macleod, R.A., Pott, C., Schreiber, S., Trumper, L., Loeffler, M., Stadler, P.F., Lichter, P., Eils, R., Kuppers, R., Hummel, M., Klapper, W., Rosenstiel, P., Rosenwald, A., Brors, B., and Siebert, R., *Recurrent mutation of the ID3 gene in Burkitt lymphoma identified by integrated genome, exome and transcriptome sequencing.* Nat Genet, 2012. **44**(12):1316-20

27.   Sanjmyatav, J., Junker, K., Matthes, S., Muehr, M., Sava, D., Sternal, M., Wessendorf, S., Kreuz, M., Gajda, M., Wunderlich, H., and Schwaenen, C., *Identification of genomic alterations associated with metastasis and cancer specific survival in clear cell renal cell carcinoma.* J Urol, 2011. **186**(5): p. 2078-83.

28.   Gietzelt. J, Kreuz. M, and Hasenclever. D, *Comparing copy number aberrations in pairs of tumor samples from the same patient.* Mainz / / 2011. 56. Jahrestagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie (gmds), 6. Jahrestagung der Deutschen Gesellschaft für Epidemiologie (DGEpi). Mainz, 26.-29.09.2011. Düsseldorf: German Medical Science GMS Publishing House; 2011. Doc 11 gmds 061, 2011.

29.   Kennedy, G.C., Matsuzaki, H., Dong, S., Liu, W.M., Huang, J., Liu, G., Su, X., Cao, M., Chen, W., Zhang, J., Liu, W., Yang, G., Di, X., Ryder, T., He, Z., Surti, U., Phillips, M.S., Boyce-Jacino, M.T.,

Fodor, S.P., and Jones, K.W., *Large-scale genotyping of complex DNA.* Nat Biotechnol, 2003. **21**(10): p. 1233-7.

30.   Matsuzaki, H., Loi, H., Dong, S., Tsai, Y.Y., Fang, J., Law, J., Di, X., Liu, W.M., Yang, G., Liu, G., Huang, J., Kennedy, G.C., Ryder, T.B., Marcus, G.A., Walsh, P.S., Shriver, M.D., Puck, J.M., Jones, K.W., and Mei, R., *Parallel genotyping of over 10,000 SNPs using a one-primer assay on a high-density oligonucleotide array.* Genome Res, 2004. **14**(3): p. 414-25.

31.   *Affymetrix. BRLMM: an improved genotype calling method for the genechip human mapping 500k array set.* Affymetrix, Inc White Paper, 2006.

32.   Beroukhim, R., Lin, M., Park, Y., Hao, K., Zhao, X., Garraway, L.A., Fox, E.A., Hochberg, E.P., Mellinghoff, I.K., Hofer, M.D., Descazeaud, A., Rubin, M.A., Meyerson, M., Wong, W.H., Sellers, W.R., and Li, C., *Inferring loss-of-heterozygosity from unpaired tumors using high-density oligonucleotide SNP arrays.* PLoS Comput Biol, 2006. **2**(5): p. e41.

33.   Lin, M., Wei, L.J., Sellers, W.R., Lieberfarb, M., Wong, W.H., and Li, C., *dChipSNP: significance curve and clustering of SNP-array-based loss-of-heterozygosity data.* Bioinformatics, 2004. **20**(8): p. 1233-40.

34.   Zhao, X., Li, C., Paez, J.G., Chin, K., Janne, P.A., Chen, T.H., Girard, L., Minna, J., Christiani, D., Leo, C., Gray, J.W., Sellers, W.R., and Meyerson, M., *An integrated view of copy number and allelic alterations in the cancer genome using single nucleotide polymorphism arrays.* Cancer Res, 2004. **64**(9): p. 3060-71.

35.   Nannya, Y., Sanada, M., Nakazaki, K., Hosoya, N., Wang, L., Hangaishi, A., Kurokawa, M., Chiba, S., Bailey, D.K., Kennedy, G.C., and Ogawa, S., *A robust algorithm for copy number detection using high-density oligonucleotide single nucleotide polymorphism genotyping arrays.* Cancer Res, 2005. **65**(14): p. 6071-9.

36.   Mallona, I., Weiss, J., and Egea-Cortines, M., *pcrEfficiency: a Web tool for PCR amplification efficiency prediction.* BMC Bioinformatics. **12**: p. 404.

37.   Beroukhim, R., Getz, G., Nghiemphu, L., Barretina, J., Hsueh, T., Linhart, D., Vivanco, I., Lee, J.C., Huang, J.H., Alexander, S., Du, J., Kau, T., Thomas, R.K., Shah, K., Soto, H., Perner, S., Prensner, J., Debiasi, R.M., Demichelis, F., Hatton, C., Rubin, M.A., Garraway, L.A., Nelson, S.F., Liau, L., Mischel, P.S., Cloughesy, T.F., Meyerson, M., Golub, T.A., Lander, E.S., Mellinghoff, I.K., and Sellers, W.R., *Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma.* Proc Natl Acad Sci U S A, 2007. **104**(50): p. 20007-12.

38.   Schmitz, R., Young, R.M., Ceribelli, M., Jhavar, S., Xiao, W., Zhang, M., Wright, G., Shaffer, A.L., Hodson, D.J., Buras, E., Liu, X., Powell, J., Yang, Y., Xu, W., Zhao, H., Kohlhammer, H., Rosenwald, A., Kluin, P., Muller-Hermelink, H.K., Ott, G., Gascoyne, R.D., Connors, J.M., Rimsza, L.M., Campo, E., Jaffe, E.S., Delabie, J., Smeland, E.B., Ogwang, M.D., Reynolds, S.J., Fisher, R.I., Braziel, R.M., Tubbs, R.R., Cook, J.R., Weisenburger, D.D., Chan, W.C., Pittaluga, S., Wilson, W., Waldmann, T.A., Rowe, M., Mbulaiteye, S.M., Rickinson, A.B., and Staudt, L.M., *Burkitt lymphoma pathogenesis and therapeutic targets from structural and functional genomics.* Nature, 2012. **490**(7418): p. 116-20.

39.   Scholtysik, R., Nagel, I., Kreuz, M., Vater, I., Giefing, M., Schwaenen, C., Wessendorf, S., Trumper, L., Loeffler, M., Siebert, R., and Kuppers, R., *Recurrent deletions of the TNFSF7 and TNFSF9 genes in 19p13.3 in diffuse large B-cell and Burkitt lymphomas.* Int J Cancer, 2011. **131**(5):E830-5.

40.   Schneppenheim, R., Fruhwald, M.C., Gesk, S., Hasselblatt, M., Jeibmann, A., Kordes, U., Kreuz, M., Leuschner, I., Martin Subero, J.I., Obser, T., Oyen, F., Vater, I., and Siebert, R., *Germline nonsense mutation and somatic inactivation of SMARCA4/BRG1 in a family with rhabdoid tumor predisposition syndrome.* Am J Hum Genet, 2010. **86**(2): p. 279-84.

41.   Benjamini, Y. and Hochberg, Y., *Controlling the false discovery rate: a practical and powerful approach to multiple testing.* J. Roy. Statist. Soc. Ser. B, 1995. **57**: p. 289-300.

42.   Curtis, C., Shah, S.P., Chin, S.F., Turashvili, G., Rueda, O.M., Dunning, M.J., Speed, D., Lynch, A.G., Samarajiwa, S., Yuan, Y., Graf, S., Ha, G., Haffari, G., Bashashati, A., Russell, R., McKinney, S., Langerod, A., Green, A., Provenzano, E., Wishart, G., Pinder, S., Watson, P.,

Markowetz, F., Murphy, L., Ellis, I., Purushotham, A., Borresen-Dale, A.L., Brenton, J.D., Tavare, S., Caldas, C., and Aparicio, S., *The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups.* Nature, 2012. **486**(7403): p. 346-52.

43. Lane, D.P., *Cancer. p53, guardian of the genome.* Nature, 1992. **358**(6381):15-6

44. Rosenwald, A., Wright, G., Chan, W.C., Connors, J.M., Campo, E., Fisher, R.I., Gascoyne, R.D., Muller-Hermelink, H.K., Smeland, E.B., Giltnane, J.M., Hurt, E.M., Zhao, H., Averett, L., Yang, L., Wilson, W.H., Jaffe, E.S., Simon, R., Klausner, R.D., Powell, J., Duffey, P.L., Longo, D.L., Greiner, T.C., Weisenburger, D.D., Sanger, W.G., Dave, B.J., Lynch, J.C., Vose, J., Armitage, J.O., Montserrat, E., Lopez-Guillermo, A., Grogan, T.M., Miller, T.P., LeBlanc, M., Ott, G., Kvaloy, S., Delabie, J., Holte, H., Krajci, P., Stokke, T., and Staudt, L.M., *The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma.* N Engl J Med, 2002. **346**(25): p. 1937-47.

45. Swerdlow Steven H and International Agency for Research on Cancer, W.H.O., *WHO classification of tumours of haematopoietic and lymphoid tissues*. 2008.

46. Mangs, A.H. and Morris, B.J., *The Human Pseudoautosomal Region (PAR): Origin, Function and Future.* Curr Genomics, 2007. **8**(2): p. 129-36.

47. Lovec, H., Grzeschiczek, A., Kowalski, M.B., and Moroy, T., *Cyclin D1/bcl-1 cooperates with myc genes in the generation of B-cell lymphoma in transgenic mice.* EMBO J, 1994. **13**(15): p. 3487-95.

48. Nielaender, I., Martin-Subero, J.I., Wagner, F., Martinez-Climent, J.A., and Siebert, R., *Partial uniparental disomy: a recurrent genetic mechanism alternative to chromosomal deletion in malignant lymphoma.* Leukemia, 2006. **20**(5): p. 904-5.

49. Rubio-Moscardo, F., Climent, J., Siebert, R., Piris, M.A., Martin-Subero, J.I., Nielander, I., Garcia-Conde, J., Dyer, M.J., Terol, M.J., Pinkel, D., and Martinez-Climent, J.A., *Mantle-cell lymphoma genotypes identified with CGH to BAC microarrays define a leukemic subgroup of disease and predict patient outcome.* Blood, 2005. **105**(11): p. 4445-54.

50. Yamamoto, G., Nannya, Y., Kato, M., Sanada, M., Levine, R.L., Kawamata, N., Hangaishi, A., Kurokawa, M., Chiba, S., Gilliland, D.G., Koeffler, H.P., and Ogawa, S., *Highly sensitive method for genomewide detection of allelic composition in nonpaired, primary tumor specimens by use of affymetrix single-nucleotide-polymorphism genotyping microarrays.* Am J Hum Genet, 2007. **81**(1): p. 114-26.

51. Courts, C., Montesinos-Rongen, M., Brunn, A., Bug, S., Siemer, D., Hans, V., Blumcke, I., Klapper, W., Schaller, C., Wiestler, O.D., Kuppers, R., Siebert, R., and Deckert, M., *Recurrent inactivation of the PRDM1 gene in primary central nervous system lymphoma.* J Neuropathol Exp Neurol, 2008. **67**(7): p. 720-7.

52. Pasqualucci, L., Compagno, M., Houldsworth, J., Monti, S., Grunn, A., Nandula, S.V., Aster, J.C., Murty, V.V., Shipp, M.A., and Dalla-Favera, R., *Inactivation of the PRDM1/BLIMP1 gene in diffuse large B cell lymphoma.* J Exp Med, 2006. **203**(2): p. 311-7.

53. Nakamura, M., Sakaki, T., Hashimoto, H., Nakase, H., Ishida, E., Shimada, K., and Konishi, N., *Frequent alterations of the p14(ARF) and p16(INK4a) genes in primary central nervous system lymphomas.* Cancer Res, 2001. **61**(17): p. 6335-9.

54. Nakagawa, M., Nakagawa-Oshiro, A., Karnan, S., Tagawa, H., Utsunomiya, A., Nakamura, S., Takeuchi, I., Ohshima, K., and Seto, M., *Array comparative genomic hybridization analysis of PTCL-U reveals a distinct subgroup with genetic alterations similar to lymphoma-type adult T-cell leukemia/lymphoma.* Clin Cancer Res, 2009. **15**(1): p. 30-8.

55. Fujiwara, S.I., Yamashita, Y., Nakamura, N., Choi, Y.L., Ueno, T., Watanabe, H., Kurashina, K., Soda, M., Enomoto, M., Hatanaka, H., Takada, S., Abe, M., Ozawa, K., and Mano, H., *High-resolution analysis of chromosome copy number alterations in angioimmunoblastic T-cell lymphoma and peripheral T-cell lymphoma, unspecified, with single nucleotide polymorphism-typing microarrays.* Leukemia, 2008. **22**(10): p. 1891-8.

56. Nagel, S., Leich, E., Quentmeier, H., Meyer, C., Kaufmann, M., Drexler, H.G., Zettl, A., Rosenwald, A., and MacLeod, R.A., *Amplification at 7q22 targets cyclin-dependent kinase 6 in T-cell lymphoma.* Leukemia, 2008. **22**(2): p. 387-92.

57. Rao, P.H., Houldsworth, J., Dyomina, K., Parsa, N.Z., Cigudosa, J.C., Louie, D.C., Popplewell, L., Offit, K., Jhanwar, S.C., and Chaganti, R.S., *Chromosomal and gene amplification in diffuse large B-cell lymphoma.* Blood, 1998. **92**(1): p. 234-40.

58. Zettl, A., Rudiger, T., Konrad, M.A., Chott, A., Simonitsch-Klupp, I., Sonnen, R., Muller-Hermelink, H.K., and Ott, G., *Genomic profiling of peripheral T-cell lymphoma, unspecified, and anaplastic large T-cell lymphoma delineates novel recurrent chromosomal alterations.* Am J Pathol, 2004. **164**(5): p. 1837-48.

59. Nakahata, S., Saito, Y., Hamasaki, M., Hidaka, T., Arai, Y., Taki, T., Taniwaki, M., and Morishita, K., *Alteration of enhancer of polycomb 1 at 10p11.2 is one of the genetic events leading to development of adult T-cell leukemia/lymphoma.* Genes Chromosomes Cancer, 2009. **48**(9): p. 768-76.

60. Hecht, J.L. and Aster, J.C., *Molecular biology of Burkitt's lymphoma.* J Clin Oncol, 2000. **18**(21): p. 3707-21.

61. Shipp, M.A., Ross, K.N., Tamayo, P., Weng, A.P., Kutok, J.L., Aguiar, R.C., Gaasenbeek, M., Angelo, M., Reich, M., Pinkus, G.S., Ray, T.S., Koval, M.A., Last, K.W., Norton, A., Lister, T.A., Mesirov, J., Neuberg, D.S., Lander, E.S., Aster, J.C., and Golub, T.R., *Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning.* Nat Med, 2002. **8**(1): p. 68-74.

62. *A predictive model for aggressive non-Hodgkin's lymphoma. The International Non-Hodgkin's Lymphoma Prognostic Factors Project.* N Engl J Med, 1993. **329**(14): p. 987-94.

63. Baudis, M., *Genomic imbalances in 5918 malignant epithelial tumors: an explorative meta-analysis of chromosomal CGH data.* BMC Cancer, 2007. **7**: p. 226.

64. Bejjani, B.A. and Shaffer, L.G., *Application of array-based comparative genomic hybridization to clinical diagnostics.* J Mol Diagn, 2006. **8**(5): p. 528-33.

65. Bruin, S.C., Klijn, C., Liefers, G.J., Braaf, L.M., Joosse, S.A., van Beers, E.H., Verwaal, V.J., Morreau, H., Wessels, L.F., van Velthuysen, M.L., Tollenaar, R.A., and Van't Veer, L.J., *Specific genomic aberrations in primary colorectal cancer are associated with liver metastases.* BMC Cancer, 2010. **10**: p. 662.

66. Rahnenfuhrer, J., Beerenwinkel, N., Schulz, W.A., Hartmann, C., von Deimling, A., Wullich, B., and Lengauer, T., *Estimating cancer survival and clinical outcome based on genetic tumor progression scores.* Bioinformatics, 2005. **21**(10): p. 2438-46.

67. Love, C., Sun, Z., Jima, D., Li, G., Zhang, J., Miles, R., Richards, K.L., Dunphy, C.H., Choi, W.W., Srivastava, G., Lugar, P.L., Rizzieri, D.A., Lagoo, A.S., Bernal-Mizrachi, L., Mann, K.P., Flowers, C.R., Naresh, K.N., Evens, A.M., Chadburn, A., Gordon, L.I., Czader, M.B., Gill, J.I., Hsi, E.D., Greenough, A., Moffitt, A.B., McKinney, M., Banerjee, A., Grubor, V., Levy, S., Dunson, D.B., and Dave, S.S., *The genetic landscape of mutations in Burkitt lymphoma.* Nat Genet, 2012. **44**(12): p. 1321-5.

68. Campo, E., *New pathogenic mechanisms in Burkitt lymphoma.* Nat Genet, 2012. **44**(12): p. 1288-9.

69. Lohr, J.G., Stojanov, P., Lawrence, M.S., Auclair, D., Chapuy, B., Sougnez, C., Cruz-Gordillo, P., Knoechel, B., Asmann, Y.W., Slager, S.L., Novak, A.J., Dogan, A., Ansell, S.M., Link, B.K., Zou, L., Gould, J., Saksena, G., Stransky, N., Rangel-Escareno, C., Fernandez-Lopez, J.C., Hidalgo-Miranda, A., Melendez-Zajgla, J., Hernandez-Lemus, E., Schwarz-Cruz y Celis, A., Imaz-Rosshandler, I., Ojesina, A.I., Jung, J., Pedamallu, C.S., Lander, E.S., Habermann, T.M., Cerhan, J.R., Shipp, M.A., Getz, G., and Golub, T.R., *Discovery and prioritization of somatic mutations in diffuse large B-cell lymphoma (DLBCL) by whole-exome sequencing.* Proc Natl Acad Sci U S A, 2012. **109**(10): p. 3879-84.

70. Schuster, S.C., *Next-generation sequencing transforms today's biology.* Nat Methods, 2008. **5**(1): p. 16-8.

71. Medvedev, P., Stanciu, M., and Brudno, M., *Computational methods for discovering structural variation with next-generation sequencing.* Nat Methods, 2009. **6**(11 Suppl): p. S13-20.

72.   Boeva, V., Popova, T., Bleakley, K., Chiche, P., Cappo, J., Schleiermacher, G., Janoueix-Lerosey, I., Delattre, O., and Barillot, E., *Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data.* Bioinformatics, 2012. **28**(3): p. 423-5.

73.   Duan, J., Zhang, J.G., Deng, H.W., and Wang, Y.P., *Comparative studies of copy number variation detection methods for next-generation sequencing technologies.* PLoS One, 2013. **8**(3): p. e59128.

6 Eigene Publikationen

[1] **Kreuz M**, Rosolowski M, Berger H, Schwaenen C, Wessendorf S, Loeffler M, Hasenclever D; Development and implementation of an analysis tool for array-based comparative genomic hybridization. Methods Inf Med. 2007;46(5):608–13.

[2] Bittenbring J, Parisot F, Wabo A, Mueller M, Kerschenmeyer L, **Kreuz M**, Truemper L, Landt O, Menzel A, Pfreundschuh M, Roemer K; MDM2 gene SNP309 T/G and p53 gene SNP72 G/C do not influence diffuse large B-cell non-Hodgkin lymphoma onset or survival in central European Caucasians. BMC Cancer. 2008 Apr 23;8:116.

[3] Nieländer I, Martín-Subero JI, Wagner F, Baudis M, Gesk S, Harder L, Hasenclever D, Klapper W, **Kreuz M**, Pott C, Martinez-Climent JA, Dreyling M, Arnold N, Siebert R; Recurrent loss of the Y chromosome and homozygous deletions within the pseudoautosomal region 1: association with male predominance in mantle cell lymphoma. Haematologica. 2008 Jun;93(6):949–50.

[4] Schwaenen C, Viardot A, Berger H, Barth TF, Bentink S, Döhner H, Enz M, Feller AC, Hansmann ML, Hummel M, Kestler HA, Klapper W, **Kreuz M**, Lenze D, Loeffler M, Möller P, Müller-Hermelink HK, Ott G, Rosolowski M, Rosenwald A, Ruf S, Siebert R, Spang R, Stein H, Truemper L, Lichter P, Bentz M, Wessendorf S; Molecular Mechanisms in Malignant Lymphomas Network Project of the Deutsche Krebshilfe; Microarray-based genomic profiling reveals novel genomic aberrations in follicular lymphoma which associate with patient survival and gene expression status. Genes Chromosomes Cancer. 2009 Jan;48(1):39–54

[5] Vater I, Wagner F, **Kreuz M**, Berger H, Martín-Subero JI, Pott C, Martinez-Climent JA, Klapper W, Krause K, Dyer MJ, Gesk S, Harder L, Zamo A, Dreyling M, Hasenclever D, Arnold N, Siebert R; GeneChip analyses point to novel pathogenetic mechanisms in mantle cell lymphoma. Br J Haematol. 2009 Feb;144(3):317–31.

[6] Martín-Subero JI, **Kreuz M***, Bibikova M, Bentink S, Ammerpohl O, Wickham-Garcia E, Rosolowski M, Richter J, Lopez-Serra L, Ballestar E, Berger H, Agirre X, Bernd HW, Calvanese V, Cogliatti SB, Drexler HG, Fan JB, Fraga MF, Hansmann ML, Hummel M, Klapper W, Korn B, Küppers R, Macleod RA, Möller P, Ott G, Pott C, Prosper F, Rosenwald A, Schwaenen C, Schübeler D, Seifert M, Stürzenhofecker B, Weber M, Wessendorf S, Loeffler M, Trümper L, Stein H, Spang R, Esteller M, Barker D, Hasenclever D, Siebert R; Molecular Mechanisms in Malignant Lymphomas Network Project of the Deutsche Krebshilfe; New insights into the biology and origin of mature aggressive B-cell lymphomas by combined epigenomic, genomic, and transcriptional profiling. Blood. 2009 Mar 12;113(11):2488–97.

[7] Schwindt H, Vater I, **Kreuz M**, Montesinos-Rongen M, Brunn A, Richter J, Gesk S, Ammerpohl O, Wiestler OD, Hasenclever D, Deckert M, Siebert R; Chromosomal imbalances and partial uniparental disomies in primary central nervous system lymphoma. Leukemia. 2009 Oct;23(10):1875–84

[8] Hartmann S, Gesk S, Scholtysik R, **Kreuz M**, Bug S, Vater I, Döring C, Cogliatti S, Parrens M, Merlio JP, Kwiecinska A, Porwit A, Piccaluga PP, Pileri S, Hoefler G, Küppers R, Siebert R, Hansmann ML; High resolution SNP array genomic profiling of peripheral T cell lymphomas, not otherwise specified, identifies a subgroup with chromosomal aberrations affecting the REL locus. Br J Haematol. 2010 Feb;148(3):402–12.

[9] Schneppenheim R, Frühwald MC, Gesk S, Hasselblatt M, Jeibmann A, Kordes U, **Kreuz M**, Leuschner I, Martin Subero JI, Obser T, Oyen F, Vater I, Siebert R; Germline nonsense mutation and somatic inactivation of SMARCA4/BRG1 in a family with rhabdoid tumor predisposition syndrome. Am J Hum Genet. 2010 Feb 12;86(2):279–84.

[10] Hoffmann M, Schirmer MA, Tzvetkov MV, **Kreuz M**, Ziepert M, Wojnowski L, Kube D, Pfreundschuh M, Trümper L, Loeffler M, Brockmöller J; German Study Group for High-Grade Non-Hodgkin Lymphoma: A functional polymorphism in the NAD(P)H oxidase subunit CYBA is related to gene expression, enzyme activity, and outcome in non-Hodgkin lymphoma. Cancer Res. 2010 Mar 15;70(6):2328–38.

[11] Scholtysik R, **Kreuz M**, Klapper W, Burkhardt B, Feller AC, Hummel M, Loeffler M, Rosolowski M, Schwaenen C, Spang R, Stein H, Thorns C, Trümper L, Vater I, Wessendorf S, Zenz T, Siebert R, Küppers R; Molecular Mechanisms in Malignant Lymphomas Network Project of Deutsche Krebshilfe; Detection of genomic aberrations in molecularly defined Burkitt's lymphoma by array-based, high resolution, single nucleotide polymorphism analysis. Haematologica. 2010 Dec;95(12):2047–55.

[12] Sellmann L, Scholtysik R, **Kreuz M**, Cyrull S, Tiacci E, Stanelle J, Carpinteiro A, Nückel H, Boes T, Gesk S, Siebert R, Klein-Hitpass L, Dührsen U, Dürig J, Küppers R; Gene dosage effects in chronic lymphocytic leukemia. Cancer Genet Cytogenet. 2010 Dec;203(2):149-60.

[13] Salaverria I, Philipp C, Oschlies I, Kohler CW, **Kreuz M***, Szczepanowski M, Burkhardt B, Trautmann H, Gesk S, Andrusiewicz M, Berger H, Fey M, Harder L, Hasenclever D, Hummel M, Loeffler M, Mahn F, Martin-Guerrero I, Pellissery S, Pott C, Pfreundschuh M, Reiter A, Richter J, Rosolowski M, Schwaenen C, Stein H, Trümper L, Wessendorf S, Spang R, Küppers R, Klapper W, Siebert R for the Molecular Mechanisms in Malignant Lymphomas Network Project of the Deutsche Krebshilfe; the German High-Grade Lymphoma Study Group; the Berlin-Frankfurt-Münster-NHL trial group: Translocations activating IRF4 identify a subtype of germinal center-derived B-cell lymphoma affecting predominantly children and young adults. Blood. 2011 Jul 7;118(1):139–147.

[14] Ahlgrimm M, Pfreundschuh M, **Kreuz M**, Regitz E, Preuss KD, Bittenbring J; The impact of Fc-γ receptor polymorphisms in elderly patients with diffuse large B-cell lymphoma treated with CHOP with or without rituximab. Blood. 2011 Oct 27;118(17):4657–62.

[15] Sanjmyatav J, Junker K, Matthes S, Muehr M, Sava D, Sternal M, Wessendorf S, **Kreuz M**, Gajda M, Wunderlich H, Schwaenen C; Identification of genomic alterations associated with metastasis and cancer specific survival in clear cell renal cell carcinoma. J Urol. 2011 Nov;186(5):2078–83.

[16] Scholtysik R, Nagel I, **Kreuz M**, Vater I, Giefing M, Schwaenen C, Wessendorf S, Trümper L, Loeffler M, Siebert R, Küppers R; Recurrent deletions of the TNFSF7 and TNFSF9 genes in 19p13.3 in diffuse large B-cell and Burkitt lymphomas. Int J Cancer. 2011 Dec 31.

[17] Klapper W, **Kreuz M***, Kohler CW, Burkhardt B, Szczepanowski M, Salaverria I, Hummel M, Loeffler M, Pellissery S, Woessmann W, Schwänen C, Trümper L, Wessendorf S, Spang R, Hasenclever D, Siebert R; Molecular Mechanisms in Malignant Lymphomas Network Project of the Deutsche Krebshilfe; Patient age at diagnosis is associated with the molecular characteristics of diffuse large B-cell lymphoma. Blood. 2012 Feb 23; 119(8):1882–7.

[18] Heemann C, **Kreuz M**, Stoller I, Schoof N, von Bonin F, Ziepert M, Löffler M, Jung W, Pfreundschuh M, Trümper L, Kube D; Circulating levels of TNF- receptor II are prognostic for patients with peripheral T-cell Non-Hodgkin lymphoma. Clin Cancer Res. 2012 May 9

[19] Salaverria I, Martin-Guerrero I, Burkhardt B, **Kreuz M**, Zenz T, Oschlies I, Arnold N, Baudis M, Bens S, García-Orad A, Lisfeld J, Schwaenen C, Szczepanowski M, Wessendorf S, Pfreundschuh M, Trümper L, Klapper W, Siebert R; High resolution copy number analysis of IRF4 translocation-positive diffuse large B-cell and follicular lymphomas. Genes Chromosomes Cancer. 2013 Feb;52(2):150-5.

[20] Richter J, Schlesner M, Hoffmann S, **Kreuz M***, Leich E, Burkhardt B, Rosolowski M, Ammerpohl O, Wagener R, Bernhart SH, Lenze D, Szczepanowski M, Paulsen M, Lipinski S, Russell RB, Adam-Klages S, Apic G, Claviez A, Hasenclever D, Hovestadt V, Hornig N, Korbel JO, Kube D, Langenberger D, Lawerenz C, Lisfeld J, Meyer K, Picelli S, Pischimarov J, Radlwimmer B, Rausch T, Rohde M, Schilhabel M, Scholtysik R, Spang R, Trautmann H, Zenz T, Borkhardt A, Drexler HG, Moller P, Macleod RA, Pott C, Schreiber S, Trumper L, Loeffler M, Stadler PF, Lichter P, Eils R, Kuppers R, Hummel M, Klapper W, Rosenstiel P, Rosenwald A, Brors B, and Siebert R; Recurrent mutation of the ID3 gene in Burkitt lymphoma identified by integrated genome, exome and transcriptome sequencing. Nat Genet, 2012.

[21] Schif B, Lennerz JK, Kohler CW, Bentink S, **Kreuz M**, Melzner I, Ritz O, Trümper L, Loeffler M, Spang R, Möller P; SOCS1 mutation subtypes predict divergent outcomes in diffuse large B-Cell lymphoma (DLBCL) patients. Oncotarget. 2013 Jan;4(1):35-47.

[22] Aukema SM, **Kreuz M***, Kohler CW, Rosolowski M, Hasenclever D, Hummel M, Küppers R, Lenze D, Ott G, Pott C, Richter J, Rosenwald A, Szczepanowski M, Schwaenen C, Stein H, Trautmann H, Wessendorf S, Trümper L, Loeffler M, Spang R, Kluin PM, Klapper W, Siebert R; Biologic characterization of adult MYC-translocation positive mature B-cell lymphomas other than molecular Burkitt lymphoma. Haematologica. 2013 Oct 31. [Epub ahead of print]

[23] Salaverria I, Martin-Guerrero I, Wagener R, **Kreuz M***, Kohler CW, Richter J, Pienkowska-Grela B, Adam P, Burkhardt B, Claviez A, Damm-Welk C, Drexler HG, Hummel M, Jaffe ES, Küppers R, Lefebvre C, Lisfeld J, Löffler M, Macleod RA, Nagel I, Oschlies I, Rosolowski M, Russell RB, Rymkiewicz G, Schindler D, Schlesner M, Scholtysik R, Schwaenen C, Spang R, Szczepanowski M, Trümper L, Vater I, Wessendorf S, Klapper W, Siebert R; Molecular Mechanisms in Malignant Lymphoma Network Project and Berlin-Frankfurt-Münster Non-Hodgkin Lymphoma Group; A recurrent 11q aberration pattern characterizes a subset of MYC-negative high-grade B-cell lymphomas resembling Burkitt lymphoma. Blood. 2014 Feb 20;123(8):1187-98.

 (***** Geteilte Erstautorschaft)

7 Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig und ohne unzulässige Hilfe oder Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Ich versichere, dass Dritte von mir weder unmittelbar noch mittelbar geldwerte Leistungen für Arbeiten erhalten haben, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen, und dass die vorgelegte Arbeit weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde zum Zweck einer Promotion oder eines anderen Prüfungsverfahrens vorgelegt wurde. Alles aus anderen Quellen und von anderen Personen übernommene Material, das in der Arbeit verwendet wurde oder auf das direkt Bezug genommen wird, wurde als solches kenntlich gemacht. Insbesondere wurden alle Personen genannt, die direkt an der Entstehung der vorliegenden Arbeit beteiligt waren.

Leipzig, den 24. April 2014

....
Datum

....
Unterschrift

8 Danksagung

9 Curriculum vitae

PERSÖNLICHE DATEN

Name	Markus Tim Kreuz
Akademischer Grad	Diplom-Informatiker
Anschrift	Weißestraße 34
	04299 Leipzig
Geburtstag	02.05.1980
Geburtsort	Leipzig
Familienstand	ledig

SCHULBILDUNG

1986–1992	Polytechnische Oberschule „Walter Heise" Holzhausen
1992–1998	Gymnasium „Neue Nikolaischule" Leipzig Abschluss: Allgemeine Hochschulreife

STUDIUM UND BERUF

1998–1999	Grundwehrdienst
1999–2006	Studium der medizinischen Informatik an der Universität Leipzig Abschluss: Diplominformatiker
2006	Wissenschaftlicher Mitarbeiter am Institut für Medizinische Informatik, Statistik und Epidemiologie (IMISE) der Universität Leipzig
2007–2009	Stipendiat im Graduiertenkolleg 1034: „Die Bedeutung genetischer Polymorphismen in der Onkologie – Von den Grundlagen zur individualisierten Therapie"
seit 2010	Wissenschaftlicher Mitarbeiter am Institut für Medizinische Informatik, Statistik und Epidemiologie (IMISE) der Universität Leipzig