

Ressortforschungsberichte zur kerntechnischen Sicherheit und zum Strahlenschutz

**Untersuchung umweltbedingter genetischer Veränderungen
im Zusammenhang mit der Entstehung von Leukämien
im Kindesalter - Vorhaben 3609S30013**

**Auftragnehmer:
Christian-Albrechts-Universität zu Kiel
Klinik für Allgemeine Pädiatrie**

**M. Stanulla
M. Schrappe
S. Schreiber
A. Franke**

Das Vorhaben wurde mit Mitteln des Bundesministeriums für Umwelt, Naturschutz und Reaktorsicherheit (BMU) und im Auftrag des Bundesamtes für Strahlenschutz (BfS) durchgeführt.

Dieser Band enthält einen Ergebnisbericht eines vom Bundesamt für Strahlenschutz im Rahmen der Ressortforschung des BMU (UFOPLAN) in Auftrag gegebenen Untersuchungsvorhabens. Verantwortlich für den Inhalt sind allein die Autoren. Das BfS übernimmt keine Gewähr für die Richtigkeit, die Genauigkeit und Vollständigkeit der Angaben sowie die Beachtung privater Rechte Dritter. Der Auftraggeber behält sich alle Rechte vor. Insbesondere darf dieser Bericht nur mit seiner Zustimmung ganz oder teilweise vervielfältigt werden.

Der Bericht gibt die Auffassung und Meinung des Auftragnehmers wieder und muss nicht mit der des BfS übereinstimmen.

BfS-RESFOR-85/13

Bitte beziehen Sie sich beim Zitieren dieses Dokumentes immer auf folgende URN:
urn:nbn:de:0221-2013101011077

Salzgitter, Oktober 2013

UNTERSUCHUNG UMWELTBEDINGTER GENETISCHER VERÄNDERUNGEN IM ZUSAMMENHANG MIT DER ENTSTEHUNG VON LEUKÄMIEN IM KINDESALTER

Martin Stanulla

Martin Schrappe

Stefan Schreiber

Andre Franke

Klinik für Allgemeine Pädiatrie

Institut für Klinische Molekularbiologie

Universitätsklinikum Schleswig-Holstein, Campus Kiel

Christian Albrechts-Universität, Kiel

UNTERSUCHUNG UMWELTBEDINGTER GENETISCHER VERÄNDERUNGEN IM ZUSAMMENHANG MIT DER ENTSTEHUNG VON LEUKÄMIEN IM KINDESALTER

A	ZUSAMMENFASSUNG	3
B	SUMMARY	3
C	PROJEKTHINTERGRUND	5
D	METHODEN	7
1.	Rekrutierung von Patienten und gesunden Kontrollpersonen	7
2.	Diagnose und Charakterisierung der Leukämien	8
3.	Zellen und Nukleinsäuren	8
4.	Genomweite SNP-Genotypisierung	9
5.	SNP-Auswahl für die Replikation, SNPlex und TaqMan Genotypisierung	9
6.	Sequenzierung zur Mutationsdetektion	9
7.	Statistische Analysen	10
8.	Imputation	10
9.	Plotten	10
E	ERGEBNISSE	10
F	DISKUSSION	18
G	LITERATUR	20

A ZUSAMMENFASSUNG

Die akute lymphoblastische Leukämie (ALL) ist eine bösartige Erkrankung der weißen Blutkörperchen, bei der sich frühe lymphoide Vorläuferzellen unkontrolliert vermehren und das normale blutbildende Gewebe des Knochenmarks ersetzen. Wir haben im Rahmen einer genomweiten Assoziationsstudie 355.750 SNPs in 419 Pa-tienten mit einem der häufigsten Subtypen einer ALL - gekennzeichnet durch die chromosomale Translokation $t(12;21)(p13;q22)$, die zu einer ETV6-RUNX1-Genfusion führt - sowie 474 gesunden Kontrollen analysiert. Die 100 am stärksten mit dem Leukämierisiko assoziierten SNPs wurden in 693 ETV6-RUNX1-positiven Leukämiefällen und 2261 Kontrollen, bestehend aus zwei unabhängigen Gruppen aus Deutschland/Österreich und Italien, im Rahmen einer Replikationsanalyse typisiert. In diesen Untersuchungen identifizierten wir zwei neue Risikoloci auf den Chromosomenbanden 3q28 (*TP63*, rs17505102, $P = 1,18 \times 10^{-7}$, OR=0.62) und 14q24.3 (benachbart mit *C14orf118*, rs7156960, $P = 1,10 \times 10^{-7}$, OR=0,78). Die separate Analyse der kombinierten deutsch/österreichischen Proben, offenbarte weitere genomweite signifikante Assoziationen in den Bereichen 11q11 (*OR8U8*, rs1945213, $P = 8,54 \times 10^{-10}$, OR=0.69), 8p21.3 (in der Nähe *INTS10*, rs920590, $P = 4,76 \times 10^{-8}$, OR=1,36) und 11p11.2 (*PTPRJ*, rs3942852, $P = 2,04 \times 10^{-7}$, OR=0,72). Die Ergebnisse blieben in den deutsch/österreichischen Replikationsproben auch nach Bonferroni-Panel-Korrektur für multiples Testen signifikant. Die erzielten Ergebnisse zeigen erstmalig, dass es neben allgemeinen genetischen Risikoassoziationen für die ALL auch für ALL-Subgruppen spezifische Risikoloci gibt. Die Identifikation von *TP63* und *PTPRJ* als Suszeptibilitätsgene verdeutlicht die Rolle der TP53 Genfamilie und die Bedeutung von Proteinen, die zelluläre Prozesse regulieren, bei der Krebsentstehung.

B SUMMARY

Acute lymphoblastic leukemia is a malignant disease of the white blood cells. The etiology of ALL is believed to be multifactorial and likely to involve interplay of environmental and genetic variables. We performed a genome-wide association study of 355,750 SNPs in 474 controls and 419 childhood ALL cases characterized by a $t(12;21)(p13;q22)$ - the most common chromosomal translocation observed in child-hood ALL - which leads to an ETV6-RUNX1 gene fusion. The one hundred most strongly associated SNPs were followed-up in 693 cases and 2,261 controls from Germany/Austria and Italy, respectively. We identified two novel, genome-wide significant risk loci at 3q28 (*TP63*, rs17505102, $P_{CMH}=1.18 \times 10^{-7}$, OR=0.62) and at 14q24.3 (close to *C14orf118*, rs7156960, $P_{CMH} = 1.10 \times 10^{-7}$, OR=0,78). The separate analysis of the combined German/Austrian sample only, revealed additional genomewide significant associations at 11q11

(OR8U8, rs1945213, $P = 8.54 \times 10^{-10}$, OR=0.69) and 8p21.3 (near *INTS10*, rs920590, $P = 4.76 \times 10^{-8}$, OR=1.36). These associations and another association at 11p11.2 (*PTPRJ*, rs3942852, $P = 2.04 \times 10^{-7}$, OR=0.72) remained significant in the German/Austrian replication panel after correction for multiple testing. Our findings demonstrate that germline genetic variation can specifically contribute to the risk of ETV6-RUNX1-positive childhood ALL. The identification of *TP63* and *PTPRJ* as susceptibility genes emphasizes the role of the TP53 gene family and the importance of proteins regulating cellular processes in connection with tumorigenesis.

C PROJEKTHINTERGRUND

Eine Studie zur Häufung von Leukämie bei Kindern im Umkreis von Kernkraftwerken (KiKK-Studie) hat in der jüngeren Vergangenheit die Diskussion um die Ursachen der Leukämie im Kindesalter intensiviert. Nationale und internationale Forschungsbemühungen, die sich vor allem auf Assoziationen zwischen potentiell leukämogenen Faktoren und der Inzidenz der akuten Leukämie bezogen haben, konnten andererseits bislang keinen gut belegbaren ursächlichen Zusammenhang zwischen exogenen Faktoren (wie z.B. ionisierender Strahlung) und den in Deutschland und andernorts beschriebenen Leukämie-Clustern im Umkreis von Kernkraftwerken liefern.

Im Vordergrund der Diskussion steht im oben angesprochenen Sachverhalt die akute lymphoblastische Leukämie (ALL). Die ALL ist mit einem Anteil von 30% und ca. 550 Neuerkrankungen pro Jahr die häufigste bösartige Erkrankung des Kindesalters in Deutschland. Im Vergleich hierzu wird die akute myeloische Leukämie (AML) bei ca. 120 Kindern (bis 15 Jahre) pro Jahr in Deutschland diagnostiziert. Basierend auf der Identifizierung prognostischer Marker und der Entwicklung risikoadaptierter Behandlungsstrategien wurden bezüglich der Behandlung der ALL in den letzten Jahrzehnten bedeutende Fortschritte erzielt. So konnte durch eine systematische und kontinuierliche Evaluation der Therapie im Rahmen bundesweiter kooperativer Therapiestudien, die erfolgreiche Behandelbarkeit der ALL in den letzten vier Jahrzehnten von etwa 10% in den 1960er Jahren auf nahezu 80% in der heutigen Zeit verbessert werden. Im Gegensatz zum Wissenszuwachs in der klinischen Forschung sind die Ursachen der ALL bislang leider nahezu unbekannt.

Die ALL ist durch eine maligne klonale Proliferation von lymphoiden Zellen gekennzeichnet, die in einer frühen Differenzierungsphase blockiert sind. Mehr als 60% der Patienten, die mit einer ALL diagnostiziert werden, sind Kinder und Jugendliche im Alter bis zu 15 Jahren. Die jährliche Inzidenz der ALL im Kindesalter variiert weltweit zwischen einem und vier neuen Fällen pro 100.000 Kindern und Jugendlichen jünger als 15 Jahre und findet ihren Höhepunkt zwischen dem zweiten und fünften Lebensjahr. Die Ursachen der ALL im Kindesalter sind vermutlich multifaktoriell und wahrscheinlich durch ein Zusammenspiel von genetischen und umweltassoziierten Variablen bestimmt.

Die international geführte Diskussion zum Zusammenhang von Kernkraftwerken mit dem Erkrankungsrisiko für Leukämien hat aufgezeigt, dass die bisherigen Studienansätze höchstwahrscheinlich nicht den Durchbruch bezüglich eines besseren Verständnisses der Leukämieentstehung liefern können. Es ist davon auszugehen, dass es wahrscheinlich erfolgversprechender ist, über ein besseres Verständnis der Leukämieentstehung im allgemeinen, in nachgeordneten Untersuchungen auch die diskutierte Beziehung von Kernkraftwerken mit dem

Auftreten von Leukämien zielführender zu beleuchten. Aktuell werden v.a. in den USA, in Großbritannien - und jetzt auch in Deutschland - Ansätze gesucht, die molekulargenetische, populationsgenetische und epidemiologische Informationen zu verknüpfen versuchen, um dem genannten Ziel näher zu kommen.

Unter dem Begriff ALL lassen sich verschiedene Subtypen der Erkrankung unterscheiden, die entweder immunphänotypisch (z.B. B-Zell-Vorläufer- versus T-Zell-ALL) oder durch eine insbesondere in den letzten zwei Jahrzehnten ständig wachsende Zahl von bekannten genetischen Veränderungen in den Leukämiezellen charakterisiert werden können. Die beiden häufigsten genetischen Subtypen der ALL im Kindesalter sind entweder durch Hyperdiploidie (>50 Chromosomen pro Leukämiezelle) oder die chromosomale Translokation t(12;21)(p13;q22), die zu einer ETV6-RUNX1 (früher TEL-AML1 genannt)-Genfusion führt, charakterisiert. Auf die hyperdiploide und die ETV6-RUNX1-positive ALL entfallen rund 25% bzw. 20% der gesamten ALLs des Kindesalters. Neben den häufigen erworbenen genetischen Veränderungen, die sich in den ALL-Zellen nachweisen lassen, wird eine Minderheit von ALLs (<5%) im Zusammenhang mit spezifischen genetischen Vorerkrankungen wie dem Down-Syndrom oder Erkrankungen mit beeinträchtigter DNA-Reparatur beobachtet (z.B. Ataxia teleangiectasia, Nijmegen-Breakage-Syndrom oder Bloom-Syndrom). Für den Großteil der ALLs im Kindesalter waren möglicherweise prädisponierende genetische Faktoren bislang nicht bekannt. Vor kurzem haben jedoch zwei genomweite Assoziationsstudien (GWAS) zur ALL im Kindesalter, die in Großbritannien und den USA durchgeführt wurden, erstmals zeigen können, dass sich bestimmte Keimbahnloci mit dem Risiko einer ALL im Kindesalter assoziieren lassen. Die britische Studie verwendete hierbei rund 291.000 Einzelnukleotid-Polymorphismen (Single Nucleotide Polymorphism, SNP), die in 907 ALL-Fällen in einer Meta-Analyse ohne unabhängige Replikation analysiert wurden. Die US-Studie untersuchte fast 308.000 SNPs in 317 Fällen. Die beiden signifikantesten Loci wurden übereinstimmend in beiden Studien gefunden und in der Folge mittlerweile durch unabhängige Studien bestätigt. Es handelt sich hierbei um *IKZF1*, kodierend für einen Transkriptionsregulator der lymphatischen Zelldifferenzierung, und *ARID5B*, ein Mitglied der "AT-rich interaction domain"-Transkriptionsfaktorfamilie.

In einem an die Bestrebungen zur Verknüpfung von molekulargenetischen, populationsgenetischen und epidemiologischen Informationen angelehnten Projekt wird derzeit auch am Universitätsklinikum Schleswig-Holstein in Kiel versucht, genetische Ursachen für die ALL bei Kindern und Jugendlichen aufzudecken, um hierüber zu-künftig u.a. gezielt Gen-Umwelt-Interaktionen im Zusammenhang leukämogener genetischer Aberrationen beleuchten zu können. So ist in einem gemeinsamen Forschungsprojekt des Instituts für Klinische Molekularbiologie (IKMB) und der Klinik für Allgemeine Pädiatrie am Universitätsklinikum Schleswig-Holstein, Campus Kiel, im Vorfeld des hier genannten Projekts eine genomweite Assoziationsstudie (GWAS) durchgeführt worden. Bei dieser Untersuchung wurde nach krankheitsassoziierten

Genpolymorphismen (SNPs) bei der mit 25% am häufigsten vorkommenden spezifischen Untergruppe der ALL des Kindes- und Jugendalters gesucht, der durch die chromosomale Translokation t(12;21) charakterisierten sogenannten ETV6-RUNX1 (TEL-AML1)-rearrangierten Subgruppe. Dabei wurden bei 419 Patienten – dem so genannten Screening Panel A - gegenüber einer großen Anzahl von gesunden Kontrollen Hinweise auf eine mögliche genetische Assoziation mit funktionell plausiblen Bezug als Risikofaktor gefunden. Das Fehlen von häufigen, hochpenetranten Mutationen deckt sich hierbei mit den Erfahrungen aus GWAS für andere komplexe Erkrankungen. In einer ersten Validierung konnten die initial beobachteten Ergebnisse bestätigt werden. Das Patientenkollektiv sollte dann in der Folge in einem vom Bundesamt für Strahlenschutz geförderten Vorhaben (Kenn-Nr. 3609S30013) durch die Rekrutierung weiterer Patienten – sowohl auf nationaler als auch auf internationaler Ebene – vergrößert werden, um zusätzliche Replikationsstudien mit neuen Patienten durchzuführen. Weiterhin sollten die Untersuchungen auf TEL-AML1-negative Leukämien des Kindes- und Jugendalters ausgedehnt werden, um allgemeine Leukämie-Suszeptibilitätsgene identifizieren zu können. Die erzielten Ergebnisse sollten hinsichtlich ihrer Bedeutung für den Strahlenschutz diskutiert werden.

D METHODEN

1. Rekrutierung von Patienten und gesunden Kontrollpersonen

Alle Patienten der Fallgruppe wurden aus der österreichisch-deutsch-italienisch-schweizerischen klinischen Multicenterstudie AIEOP ALL-BFM 2000 zur Behandlung der ALL im Kindesalter rekrutiert. Die Patienten wurden in einem von 121 teilnehmenden Studienzentren in Österreich, Deutschland und Italien zwischen 1999 und 2008 diagnostiziert und behandelt. Zweihundert ALL-Patienten wurden in zwei weiteren klinischen Studien der deutschen Studiengruppe COALL behandelt, COALL-06-97 und COALL-07, die zwischen 1997 und 2008 in 19 deutschen Behandlungszentren durchgeführt wurden. Die deutschen gesunden Kontrollpersonen der Panels A und B wurden über die Popgen-Biobank am IKMB im Universitätsklinikum Schleswig-Holstein, Campus Kiel, erhalten (Tabelle 1). Die gesunden italienischen Kontrollindividuen wurden durch die Blutbank des Krankenhauses San Gerardo in Monza, Italien, rekrutiert. Von allen Studienteilnehmern oder ihren Erziehungsberechtigten wurden im Rahmen der Therapieprotokolle Einverständniserklärungen eingeholt, verbleibendes diagnostisches Untersuchungsmaterial für Forschungszwecke verwenden zu dürfen. Die Protokolle wurden von den jeweiligen institutionellen Ethik-Ausschüssen geprüft und genehmigt.

Tabelle 1 Übersicht zu den in der Studie verwendeten Panels

Panels	Kontrollen	Fälle
A – ALL GWAS Panel (Deutschland)	474	419
B – ALL Replikationspanel (Deutschland/Österreich)	1682	406
C – ALL Replikationspanel (Italien)	579	287
D – ALL Replikationspanel (Deutschland)	1682	326

2. Diagnose und Charakterisierung der Leukämien

Die Diagnose wurde auf der Grundlage von Zytomorphologie (FAB-Kriterien) und Zytochemie bei $\geq 25\%$ Leukämieblasten im Knochenmark gestellt. Die durchflusszytometrische Immunphänotypisierung wurde nach Konsensprotokollen gemäß den Leitlinien der Europäischen Gruppe für die immunologische Charakterisierung von Leukämien (EGIL) durchgeführt. Die Diagnostik bezüglich ETV6-RUNX1-, BCR-ABL- und MLL-AF4-Fusionsgentranskripten wurde mittels eines Multiplex-Polymerase-Kettenreaktions (PCR)-Assays in einer einzigen PCR-Reaktion durchgeführt. In der Mehrzahl der Patienten wurden positive Ergebnisse durch Interphase-Fluoreszenz in situ Hybridisierung (FISH) bestätigt.

3. Zellen und Nukleinsäuren

Bei Erstdiagnose und den Follow-up-Zeitpunkten nach Induktions- und Konsolidierungstherapie wurden Knochenmarkproben für die Bestimmung der minimalen Resterkrankung gewonnen. Hierfür wurden mononukleäre Zellen durch Gradientenzentrifugation aus dem Knochenmark isoliert und in der Folge aus den so gewonnenen Zellen hochmolekulare genomische DNA nach standardisierten Protokollen unter Zuhilfenahme des Puregene- (Gentra Systems, Minneapolis, MN, USA) oder Qiagen-DNA-Blood-Isolierungssystems (Qiagen, Hilden, Deutschland) isoliert. Qualität und Quantität der genomischen DNA wurden spektrophotometrisch bestimmt. In der hier vorliegenden Studie wurde nur auf Remissions-DNA, zurückgegriffen, die an den Follow-up-Zeitpunkten nach Induktions- oder Konsolidierungstherapie gewonnen wurde. Die DNA der deutschen Kontrollindividuen wurde mit dem Blut-NucleoSpin-Kit (Macherey-Nagel, Deutschland) aus peripherem Blut isoliert, die der italienischen Kontrollen aus peripherem Blut unter Verwendung des Qiagen-DNA-Blood-Isolierungssystems. Da es sich in der vorliegenden Studie um heterogene Quellen bezüglich der Herkunft des Untersuchungsmaterials handelte, wurden zur Qualitätskontrolle alle DNA-Proben mittels Gelelektrophorese zunächst auf ihre Integrität kontrolliert. Degradierete Proben wurden von den folgenden Experimenten ausgeschlossen.

4. Genomweite SNP-Genotypisierung

Die Genotypisierung für die GWAS Kontrollen, die als Teil der deutschen NGFN-GWAS-Initiative vom NGFN gefördert wurden, wurde als Serviceprojekt mit dem Genome-Wide Human SNP Array 5.0 (500K, 443.816 Marker) von Affymetrix® (South San Francisco, CA, USA) durchgeführt. Die genomweiten Genotypisierungen der Fälle wurden nach dem gleichen Protokoll am Cologne Center for Genomics der Universität Köln in Köln durchgeführt.

5. SNP-Auswahl für die Replikation, SNPlex und TaqMan Genotypisierung

Alle 355.750 SNPs des genomweiten Scans, die die Qualitätskontrolle bestanden haben, wurden mittels gPLINK v2.050 in Kombination mit PLINK v1.07. analysiert. Korrelierte SNPs (basierend auf Kopplungsungleichgewicht) wurden unter Verwendung des --clump- Befehls in PLINK identifiziert. Zu Kontrollzwecken wurden bei allen Index-SNPs mit einem P-Wert von $<3 \times 10^{-4}$ die entsprechenden Cluster-Plots visuell inspiziert. SNPs, die diese visuelle Inspektion nicht bestanden, wurden von der weiteren Analyse ausgeschlossen. Die 100 am stärksten assoziierten SNPs ($P < 2 \times 10^{-4}$) in diesen Analysen wurden dann in zusätzlichen Experimenten weiterverfolgt. Die nachgeschalteten Genotypisierungen (Panels B bis D) wurden unter Verwendung von SNPlex™ und TaqMan® Technologie (Applied Biosystems, Foster City, CA, USA) in einem automatisierten Laborablauf durchgeführt und die ermittelten Informationen mittels eines datenbankgestützten Laborinformationsmanagementsystems (LIMS) verwaltet. Von den 100 ausgewählten SNPs überstanden 87 SNPs die Qualitätskontrollen in beiden Replikationspanels. Diese SNPs hatten eine hohe Call-Rate (Rate der erfolgreichen Genotypisierungen; $>90\%$ in Fällen und Kontrollen), waren weder monomorph noch selten (Frequenz des selteneren Allels $>1\%$ in Fällen und Kontrollen) und wichen nicht vom Hardy-Weinberg-Gleichgewicht (HWE) in der Kontrollgruppe ab ($PHWE > 0,001$).

6. Sequenzierung zur Mutationsdetektion

Amplifikate wurden mit Hilfe des folgenden Touchdown-PCR-Programms generiert:

95°C für 12 min, (95°C für 30 sec, T_m für 30 sec, 72°C für 30 sec) × 16 Zyklen [td -0,5°C], (95°C für 30 sec, T_m für 30 sec, 72°C für 30 sec) × 19 Zyklen, 72°C für 10 min, 10°C bis zum Ende des Experiments. Die Sequenzierung der DNA erfolgte dann mittels BigDye® Terminator Technologie (v3.1, Applied Biosystems) und Kapillarsequenzierung (ABI3730 Sequenzer, Applied Biosystems) nach Angaben des Herstellers. Die ermittelten Daten wurden nach SNPs, Insertionen und Deletionen unter Verwendung von novoSNP-Software (v2.03) inspiziert.

7. Statistische Analysen

Zur Berechnung der statistischen Kraft unserer Studie wurden die Software-Programme PS Power und Sample Size (v3.0.12) eingesetzt. Genomweite Assoziationsanalysen wurden mit gPLINK (v2.050) in Kombination mit PLINK (v1.07) durchgeführt. GWAS Daten wurden auch mit R Statistical Environment (v2.10.0) analysiert. Korrelierte SNPs wurden nach den unten aufgeführten Schwellenwerten mittels des `--clump`-Befehls in PLINK entfernt.

Schwellenwerte:

<code>--clump-p1 0.001</code>	Schwellenwert für Index-SNPs
<code>--clump-p2 0.01</code>	Sekundärer Schwellenwert für <i>clumped</i> SNPs
<code>--clump-r2 0.50</code>	Schwellenwert Kopplungsungleichgewicht für <i>clumping</i>
<code>--clump-kb 250</code>	Distanzschwellenwert für <i>clumping</i>

8. Imputation

Das Software-Programm BEAGLE (v3.1.1) wurde verwendet, um die Genotypen autosomaler SNPs basierend auf den 1000 Genomes-Daten zu imputieren. Zur Imputation wurden nur SNPs herangezogen, die initial auf der Affymetrix-Plattform die Qualitätskontrolle erfolgreich durchlaufen hatten.

9. Plotten

Regionale Plots wurden mit einer modifizierten Version von deBakker's R-Skript generiert. Auf der vertikalen Achse finden sich dabei SNPs mit den entsprechenden P-Werten (log10-Skala) dargestellt, auf der horizontalen lässt sich ihre chromosomale Position ablesen. Die Einzelmarker-Assoziationsanalyse erfolgte mittels des allellischen Tests für Fall-Kontroll-Gruppen in PLINK unter Verwendung sowohl genotypisierter als auch imputierter SNPs als Input.

E ERGEBNISSE

Ausgehend von einer Häufigkeit eines erkrankungsassoziierten Allels von mindestens 30% in der Kontrollgruppe, hatte unser initiales Screening-Panel A eine statistische Power von 80%, eine Variante mit einer relativen Quote von 1,48 oder höher auf einem Signifikanzniveau von 5% zu erkennen (Abbildung 1).

Für weiterführende statistische Analysen wurden nur Daten verwendet, die stringente Qualitätsanforderungen erfüllten (SNP Call-Rate über 95% in Fällen und Kontrollen, Häufigkeit des seltenen Allels über 1% in Fällen und Kontrollen sowie keine Abweichung vom Hardy-Weinberg Gleichgewicht in den Kontrollen ($P > 0.001$). Aus diesen Anforderungen ergaben sich 355.750 SNPs, die in 419 Fällen und 474 gesunden Kontrollen in die initialen Analysen einbezogen wurden. Die Gesamtgenotypisierungsrate bei diesen Proben lag bei 99,8%. Weitere Qualitätskontrollen zeigten eine europäische Abstammung der typisierten Individuen und ergaben keinen Hinweis auf versteckte Individuen mit nicht europäischem Hintergrund oder unbekannte Verwandtschaftsverhältnisse innerhalb der analysierten Gruppe.

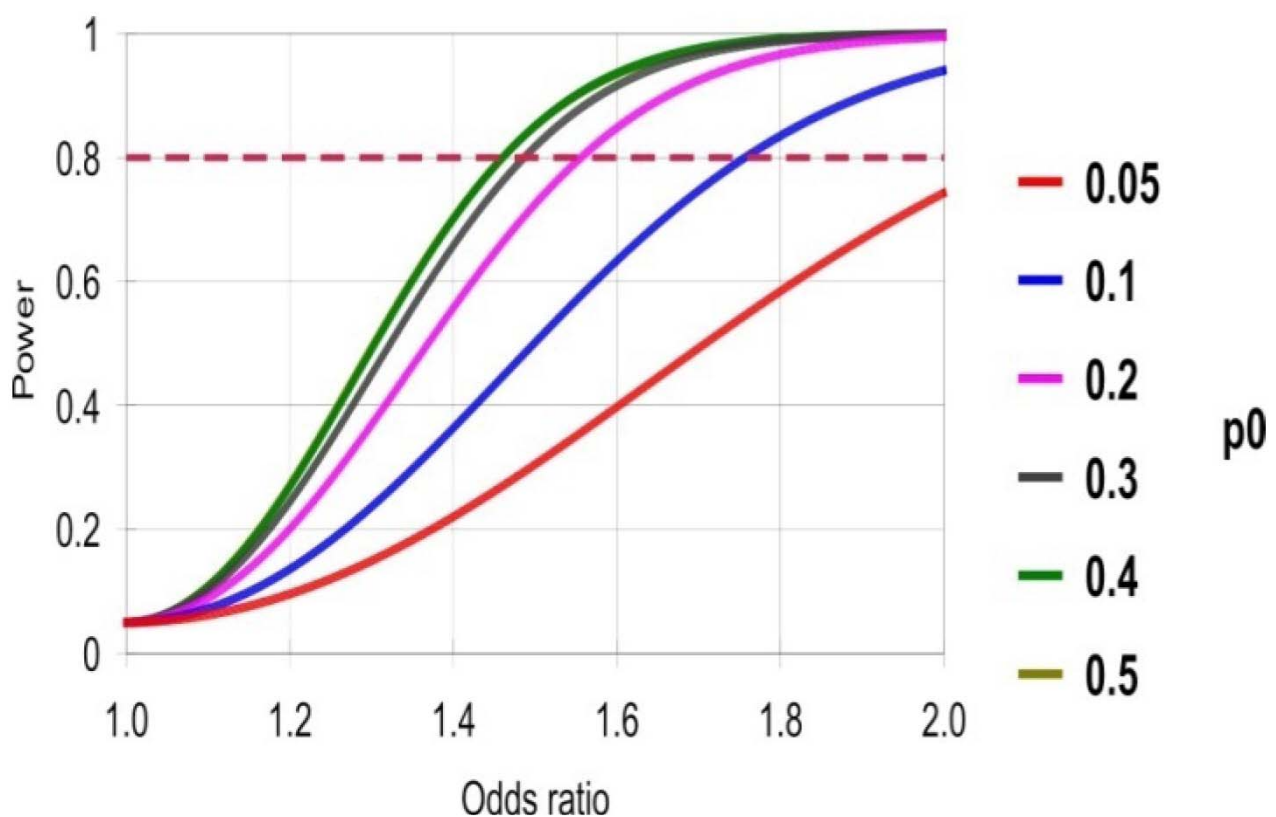


Abbildung 1. Statistische Power im Screening Panel A

Statistische Power im Screening Panel A (474 Kontrollen, 419 Fälle) eine gegebene allelische Erkrankungsassoziation für Träger des selteneren Allels zu detektieren. Die Berechnungen wurden mittels der Programme PS Power und Sample Size v3.012 durchgeführt und schlossen Annahmen für unterschiedliche Allelfrequenzen ein (bezeichnet als p_0). Die statistische Power findet sich in Abhängigkeit von der Odds ratio dargestellt und die gestrichelte rote Linie zeigt den Schwellenwert für eine statistische Power von 80% an.

Initial wurden die SNP-Häufigkeiten in Fällen und Kontrollen verglichen, die SNP-Liste bei redundanten Daten in Folge von Kopplungsungleichgewicht entsprechend gekürzt (unter

Verwendung des `--clump`-Kommandos in PLINK) und Cluster-Plots der Index-SNPs mit einem P-Wert von weniger als $1,5 \times 10^{-4}$ visuell inspiziert. Dann wurden die 100 SNPs mit den stärksten Assoziationen für Replikationsanalysen ausgewählt und im deutsch/österreichischen Replikationspanel B (406 Fälle und 1682 Kontrollen) und dem italienischen Replikationspanel C (287 Fälle und 579 Kontrollen) typisiert (Tabelle 1). Die Genotypisierungsanalysen wurden mittels SNPlex®- oder TaqMan®-Assays (Applied Biosystems, Foster City, CA, USA) durchgeführt. Nach Qualitätskontrolle wurden 13 Follow-up-SNPs wegen einer SNP Call-Rate in Fällen und Kontrollen von 95% oder weniger, einer Häufigkeit des seltenen Allels von weniger als 1% in Fällen und Kontrollen oder einer Abweichung vom Hardy-Weinberg Gleichgewicht in den Kontrollen ($P > 0.001$) ausgeschlossen.

Unter Berücksichtigung des unterschiedlichen geographischen Ursprungs der Replikationspanels, wurden in kombinierten Analysen der Replikationspanels B und C sowie in der Gesamtpopulation (Panels A bis C) der Cochran-Mantel-Haenszel-Test (P_{CMH}) und der Breslow-Day-Test auf Heterogenität der relativen Quote (P_{BD}) angewendet. Neue signifikante Erkrankungsassoziationen, die einer Bonferroni-Korrektur widerstanden (adjustiertes Signifikanzniveau $\alpha = 5,75 \times 10^{-4}$), wurden in der kombinierten Analyse der Replikationspanels B und C für rs17505102, einen intronischen SNP in der kodierenden Region des *TP63*-Gens (tumor protein p63 isoform 3) auf Chromosomenbande 3q28 ($P_{CMH} = 1,41 \times 10^{-5}$, odds ratio (OR)=0,62, 95% Konfidenzintervall (CI) = 0,50-0,77), für rs7156960 - 34 kb abwärts von *C14orf118* liegend ($P_{CMH} = 1,08 \times 10^{-4}$, OR=0,78, 95% CI=0,69-0,89) - und für den intronischen SNP rs3942852 im *PTPRJ*-Gen (protein tyrosine phosphatase, receptor type, J) auf Chromosomenbande 11p11.2 ($P_{CMH} = 3,11 \times 10^{-4}$, OR=0,74, 95% CI=0,63-0,87) gefunden (Tabelle 2 und Abbildung 2).

Ogleich die Assoziationen dieser Loci mit ETV6-RUNX1-positiver ALL nicht genomweite Signifikanz im initialen Screen zeigten (konventionelles Level $P = < 5,0 \times 10^{-7}$), wurde für diese in den anschließenden Analysen unter Berücksichtigung der Panels A bis C mit 1112 ETV6-RUNX1-positiven Fällen und 2735 gesunde Kontrollen genomweite Signifikanz für rs17505102 ($P_{CMH} = 1,18 \times 10^{-7}$) im *TP63*-Gen und rs7156960 ($P_{CMH} = 1,10 \times 10^{-7}$) erzielt (Tabelle 2). Die kombinierte Analyse für rs3942852 ergab ein P_{CMH} von $7,27 \times 10^{-7}$.

Da das initiale Screening (Panel A) nur unter Verwendung deutscher Proben durchgeführt wurde, wurden zusätzlich zu den oben genannten Untersuchungen Analysen der nachverfolgten SNPs exklusiv für das deutsch/österreichische Replikationspanel B durchgeführt, um potentielle Loci zu identifizieren, die - im Vergleich zu einem eher allgemeinen Risiko in der gesamteuropäischen Population - spezifisch mit dem ALL-Risiko in Nordeuropäern verbunden waren. Bei diesen Analysen konnten neben den drei oben genannten neuen gemeinsamen Suszeptibilitätsloci (repräsentiert durch rs17505102, rs7156960 und rs3942852) drei weitere mit ETV6-RUNX1-positiver ALL assoziierte SNPs identifiziert werden, die einer Korrektur für multiples Testen standhielten. Dieses waren der auf Chromosomenbande 11q11 liegende SNP rs1945213

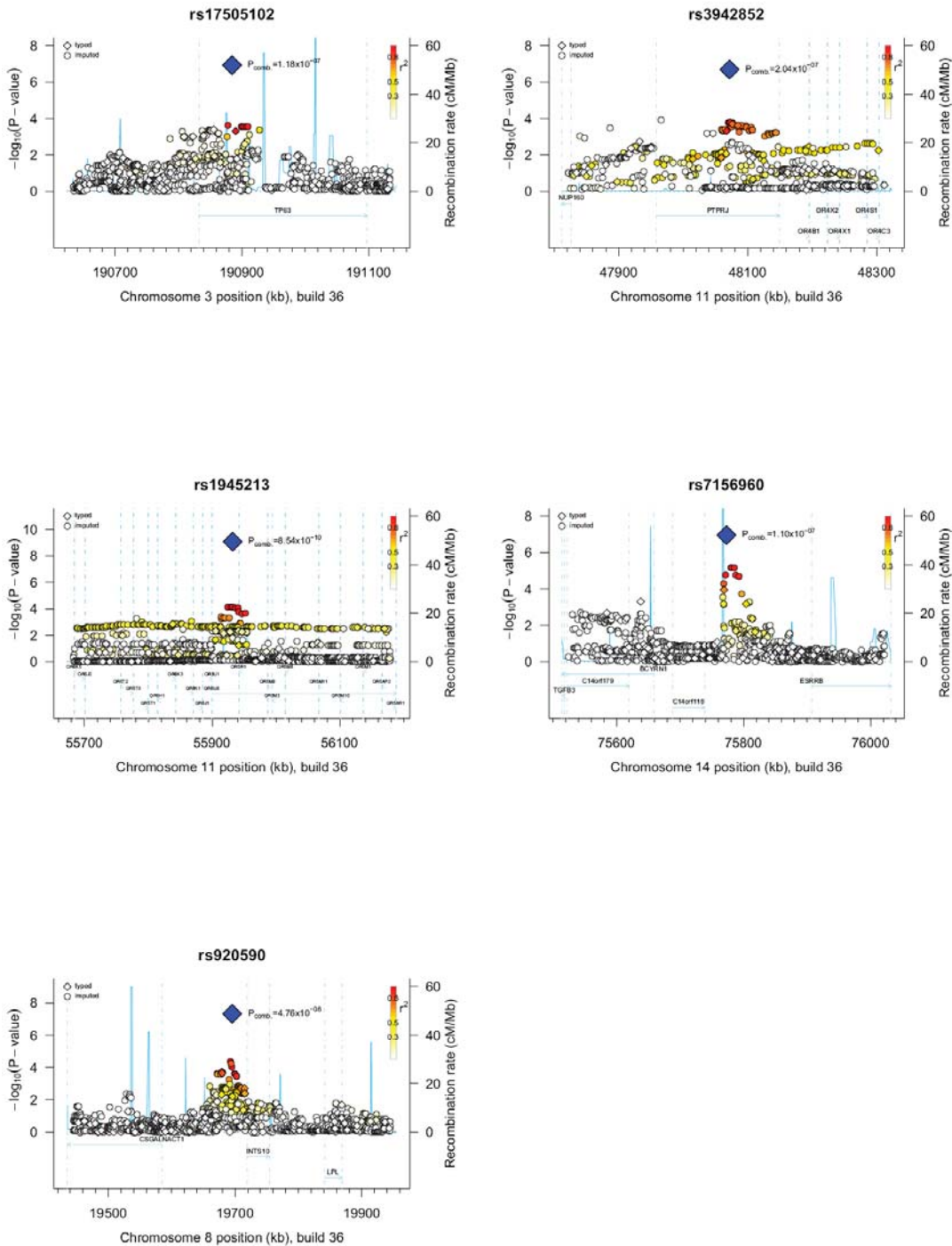


Abbildung 2. Regionale Plots

Die regionalen Plots der assoziierten Loci *TP63*, *PTPRJ*, *OR8U8*, 14q24.3, 8p21.3 und 6q14.1. stellen die für Panel A errechneten P-Werte ($\log_{10}$ -Skala) auf der vertikalen Achse und die chromosomale Position auf der horizontalen Achse dar. Für die sogenannten Lead-SNPs rs3942852, rs1945213, rs920590 und rs7738636 findet sich der P-Wert der Panels A und B, bzw. für rs17505102, rs7156960 der Panels A bis C gegeben. Die Stärke des Kopplungsungleichgewichts (LD) für SNPs gemessen als r^2 ist durch die Farbe der SNP-Symbole gekennzeichnet (für Farbkodierung, siehe rechte obere Ecke der Plots). Die Rekombinationsaktivität (in Centimorgan [cM] per Mb) ist als blaue Linie dargestellt. Positionen basieren auf NCBI build 36-Koordinaten. Weitere Details finden sich in Tabelle 2.

innerhalb des *OR8U8*-Gens ($P=1,61 \times 10^{-6}$), rs920590, 24 kb aufwärts des *INTS10*-Gens auf Chromosomenbande 8p21.3 liegend ($P=1,86 \times 10^{-4}$) und rs7738636 auf Chromosomenbande 6q14.1 ($P=5,20 \times 10^{-4}$) (Tabelle 2). Die sechs gerade genannten SNPs zeigten auch genomweite signifikante Assoziationen in der kombinierten Analyse der deutsch/österreichischen Gruppen A und B. Weiterhin zeigten zwei weitere SNPs - rs17423910 innerhalb des *PDE4B*-Gens auf Chromosom 1 ($P=1,45 \times 10^{-7}$) und rs6445754 innerhalb des *ERC2*-Gens auf Chromosom 3 ($P=1,71 \times 10^{-7}$) - genomweite Signifikanz in der kombinierten Analyse der Gruppenpanels A und B. Diese Ergebnisse hielten aber in einer alleinigen Analyse des Replikationspanels B einer statistischen Korrektur für multiples Testen nicht stand (Tabelle 2 und Abbildung 2).

Tabelle 2. Zusammenfassung der Assoziationsanalyseergebnisse

Chr. Pos. (bp)	dbSNP ID	Gen (relative Position)	A1 A2	Panel A - GWAS Deutschland ETV6-RUNX1-pos. 474 Kontrollen 419 Fälle		Panel B Deutschland/Österreich ETV6-RUNX1-pos. 1682 Kontrollen 406 Fälle		Panel A + B Deutschland/Österreich ETV6-RUNX1-pos. 2156 Kontrollen 825 Fälle		Panel C Italien ETV6-RUNX1-pos. 579 Kontrollen 287 Fälle		Panel B + C Replikation ETV6-RUNX1-pos. 2261 Kontrollen 693 Fälle			Panel A + B + C Kombinierte Analyse ETV6-RUNX1-pos. 2735 Kontrollen 1112 Fälle	
				AF _{A1} Kontr. AF _{A1} Fälle	P _{CCA}	AF _{A1} Kontr. AF _{A1} Fälle	P _{CCA}	AF _{A1} Kontr. AF _{A1} Fälle	P _{CCA}	AF _{A1} Kontr. AF _{A1} Fälle	P _{CCA}	P _{CMH}	P _{BD}	OR (95% CI)	P _{CMH}	P _{BD}
14 75,773,104	rs17156960	<i>C14orf118</i> (± 50 kb)	G C	0.52 0.41	9.16×10 ⁻⁰⁶	0.47 0.40	2.97×10⁻⁰⁴	0.48 0.41	2.35×10 ⁻⁰⁷	0.41 0.37	0.10	1.08×10⁻⁰⁴	0.38	0.78 (0.69-0.89)	1.10×10 ⁻⁰⁷	0.27
3 190,884,470	rs17505102	<i>TP63</i> (within gene)	C G	0.18 0.11	1.15×10 ⁻⁰⁴	0.15 0.09	4.14×10⁻⁰⁵	0.15 0.10	3.77×10 ⁻⁰⁷	0.08 0.06	0.11	1.41×10⁻⁰⁵	0.38	0.62 (0.50-0.77)	1.18×10 ⁻⁰⁷	0.53
11 55,932,247	rs1945213	<i>OR8U8</i> (within gene)	C G	0.33 0.24	5.82×10 ⁻⁰⁵	0.31 0.22	1.61×10⁻⁰⁶	0.31 0.23	8.54×10 ⁻¹⁰	0.21 0.24	0.32	7.84×10 ⁻⁰⁴	2.26×10 ⁻⁰⁴	0.78 (0.68-0.90)	6.16×10 ⁻⁰⁷	1.28×10 ⁻⁰⁴
11 48,071,665	rs3942852	<i>PTPRJ</i> (within gene)	C T	0.24 0.17	1.58×10 ⁻⁰⁴	0.22 0.16	1.06×10⁻⁰⁴	0.22 0.16	2.04×10 ⁻⁰⁷	0.18 0.17	0.43	3.11×10⁻⁰⁴	0.08	0.74 (0.63-0.87)	7.27×10 ⁻⁰⁷	0.06
6 77,846,527	rs7738636	—	C A	0.26 0.18	9.84×10 ⁻⁰⁵	0.24 0.18	5.20×10⁻⁰⁴	0.25 0.18	3.87×10 ⁻⁰⁷	0.16 0.15	0.46	9.76×10⁻⁰⁴	0.17	0.76 (0.65-0.90)	1.10×10 ⁻⁰⁶	0.09
15 90,458,377	rs207954	<i>SLCO3A1</i> (within gene)	T C	0.29 0.21	1.24×10 ⁻⁰⁴	0.28 0.25	0.09	0.28 0.23	6.58×10 ⁻⁰⁵	0.26 0.20	5.60×10⁻⁰³	2.45×10⁻⁰³	0.22	0.80 (0.70-0.93)	1.35×10 ⁻⁰⁶	0.61
4 59,198,483	rs282708	—	A G	0.39 0.49	3.91×10 ⁻⁰⁵	0.41 0.45	0.03	0.40 0.47	5.41×10 ⁻⁰⁶	0.46 0.48	0.42	0.03	0.51	1.15 (1.01-1.30)	1.38×10 ⁻⁰⁵	0.12
8 19,695,441	rs920590	<i>INTS10</i> (± 50 kb)	C T	0.29 0.38	5.19×10 ⁻⁰⁵	0.31 0.37	1.86×10⁻⁰⁴	0.30 0.38	4.76×10 ⁻⁰⁸	0.45 0.43	0.36	0.02	2.45×10 ⁻⁰³	1.16 (1.02-1.32)	2.57×10 ⁻⁰⁵	3.83×10 ⁻⁰⁴
19 13,727,957	rs2910225	<i>CCDC130</i> (within gene)	C T	0.42 0.32	1.02×10 ⁻⁰⁵	0.36 0.33	0.16	0.37 0.32	6.97×10 ⁻⁰⁴	0.30 0.24	0.01	9.83×10⁻⁰³	0.23	0.84 (0.74-0.96)	3.23×10 ⁻⁰⁵	0.54
15 97,081,324	rs11857366	<i>IGF1R</i> (within gene)	A G	0.37 0.46	5.98×10 ⁻⁰⁵	0.40 0.43	0.08	0.39 0.45	7.01×10 ⁻⁰⁵	0.42 0.45	0.16	0.02	0.98	1.15 (1.02-1.30)	3.41×10 ⁻⁰⁵	0.45
1 66,461,697	rs17423910	<i>PDE4B</i> (within gene)	G A	0.14 0.23	2.18×10 ⁻⁰⁶	0.17 0.22	1.11×10⁻⁰³	0.17 0.22	1.45×10 ⁻⁰⁷	0.23 0.21	0.43	0.04	8.82×10 ⁻⁰³	1.16 (1.00-1.35)	4.67×10 ⁻⁰⁵	8.97×10 ⁻⁰⁴
5 10,085,335	rs7734914	—	A G	0.40 0.50	1.32×10 ⁻⁰⁵	0.44 0.45	0.42	0.43 0.48	6.09×10 ⁻⁰⁴	0.46 0.52	0.03	4.71×10⁻⁰³	0.21	1.13 (1.00-1.28)	4.81×10 ⁻⁰⁵	0.82
1 111,060,293	rs343604	<i>KCNA3</i> (± 50 kb)	T C	0.07 0.03	8.39×10 ⁻⁰⁵	0.09 0.06	0.03	0.08 0.05	1.19×10 ⁻⁰⁶	0.08 0.09	0.67	0.14	0.08	0.84 (0.67-1.06)	9.25×10 ⁻⁰⁵	1.46×10 ⁻⁰³
3 55,782,295	rs6445754	<i>ERC2</i> (within gene)	C T	0.17 0.28	1.82×10 ⁻⁰⁷	0.23 0.28	1.07×10⁻⁰³	0.21 0.28	1.71×10 ⁻⁰⁷	0.32 0.29	0.24	0.08	3.19×10 ⁻⁰³	1.13 (0.99-1.30)	1.31×10 ⁻⁰⁴	2.36×10 ⁻⁰⁴
8 109,041,474	rs630662	<i>RSP02</i> (within gene)	A G	0.30 0.22	3.61×10 ⁻⁰⁵	0.28 0.23	4.46×10⁻⁰³	0.29 0.22	2.02×10 ⁻⁰⁶	0.23 0.25	0.45	0.06	0.02	0.87 (0.76-1.01)	1.36×10 ⁻⁰⁴	2.65×10 ⁻⁰³
13 38,845,574	rs7336133	<i>LHFP</i> (within gene)	A G	0.38 0.29	3.53×10 ⁻⁰⁵	0.32 0.31	0.42	0.34 0.30	6.01×10 ⁻⁰³	0.34 0.28	0.01	0.03	0.13	0.86 (0.75-0.98)	2.71×10 ⁻⁰⁴	0.39
6 143,700,705	rs6901152	<i>AIG1</i> (within gene)	T C	0.46 0.36	3.19×10 ⁻⁰⁵	0.42 0.36	4.63×10⁻⁰³	0.43 0.36	4.71×10 ⁻⁰⁶	0.44 0.47	0.37	0.09	0.01	0.90 (0.79-1.02)	4.19×10 ⁻⁰⁴	2.12×10 ⁻⁰³
4 159,663,764	rs936094	<i>RXFP1</i> (within gene)	C T	0.15 0.23	9.42×10 ⁻⁰⁶	0.17 0.21	0.01	0.17 0.22	1.43×10 ⁻⁰⁶	0.25 0.23	0.28	0.24	0.02	1.09 (0.94-1.27)	4.74×10 ⁻⁰⁴	6.48×10 ⁻⁰⁴
14 58,908,882	rs4901921	<i>DAAM1</i> (± 10 kb)	T C	0.52 0.43	1.48×10 ⁻⁰⁴	0.47 0.45	0.29	0.48 0.44	5.41×10 ⁻⁰³	0.40 0.35	0.04	0.04	0.32	0.88 (0.78-0.99)	6.45×10 ⁻⁰⁴	0.66
6 73,500,100	rs6910780	<i>KCNQ5</i> (within gene)	T C	0.16 0.23	1.47×10 ⁻⁰⁴	0.19 0.23	0.01	0.18 0.23	3.89×10 ⁻⁰⁵	0.22 0.21	0.64	0.09	0.06	1.13 (0.98-1.31)	9.29×10 ⁻⁰⁴	0.02

Legende zu Tabelle 2. Zusammenfassung der Assoziationsanalyseergebnisse

Die 100 am stärksten assoziierten SNPs ($P < 2 \times 10^{-4}$) der initialen Analysen von Panel A wurden in zusätzlichen Experimenten weiterverfolgt. Die nachgeschalteten Genotypisierungen (Panels B bis D) wurden unter Verwendung von SNPlex™ und TaqMan® Technologie (Applied Biosystems) durchgeführt. Von den 100 ausgewählten SNPs überstanden 87 SNPs die Qualitätskontrollen in beiden Replikationspanels. Diese SNPs hatten eine hohe Call-Rate ($>90\%$ in Fällen und Kontrollen), waren weder monomorph noch selten (Frequenz des selteneren Allels $>1\%$ in Fällen und Kontrollen) und wichen nicht vom Hardy-Weinberg-Gleichgewicht (HWE) in der Kontrollgruppe ab ($PHWE > 0,001$). Die Tabelle zeigt Daten der 20 SNPs, die entweder signifikant im Replikationspanel B waren ($P_{CCA} < 0.05$) oder in der kombinierten Analyse der Panels B und C ($P_{CMH} < 0.05$ und $P_{BD} > 0.05$). SNPs finden sich geordnet nach ihrem P -Wert aus den kombinierten Analysen der Panels A, B und C. Nukleotidpositionen (**Pos. (bp)**) beziehen sich auf NCBI build 36. Chromosom: **Chr. A1** bezeichnet das seltenere Allel, **A2** das häufigere. Allelfrequenzen finden sich in Fällen und Kontrollen für A1 angezeigt (**AF_{A1} Kontr./ AF_{A1} Fälle**). Gelistete P -Werte aus der Fall-Kontroll-Analyse wurden mittels allelbasierter χ^2 -Tests errechnet (P_{CCA}). Die Odds Ratios (**OR**) und 95% Konfidenzintervalle (**95% CI**) für das Tragen des Allels A1 gründen auf den Replikationsanalysen (Panels B und C). Die Spalte P_{BD} zeigt P -Werte aus den Breslow-Day-Tests für Heterogenität, bei denen ein signifikanter P -Wert signifikante Heterogenität zwischen den Replikationspanels bezüglich OR und Erkrankungsassoziation nahe legt. Die Spalte P_{CMH} zeigt kombinierte P -Werte (P_{CMH}) des Cochran-Mantel-Haenszel-Tests. Signifikante P -Werte ($P_{CCA} < 0.05$ oder $P_{CMH} < 0.05$ [nur wenn $P_{BD} > 0.05$]) der Replikation sind **fett** und *kursiv* angegeben. SNP-Assoziationen, die nach Korrektur für multiples Testen Signifikanz bewahrten (P_{CMH} der Panels B und C) und genomweit signifikant assoziierte SNPs der kombinierten Analyse (kombinierter P -Wert der Panels A und B oder C oder Panels A bis C) sind mit grauer Schattierung unterlegt.

Um zu beurteilen, ob die von uns erhobenen Befunde speziell dem ETV6-RUNX1-positiven ALL-Subtyp zugeordnet werden können oder auch für andere Subgruppen relevant sind, haben wir nachfolgend die 20 am stärksten assoziierten SNPs in einer weiteren deutschen Replikationskohorte (Panel D) (Tabelle 1) typisiert. Diese Gruppe setzte sich aus 326 ETV6-RUNX1-negativen ALL des Kindesalters zusammen und wurde mit den 1.682 Kontrollen des Panels B analysiert. Von 19 erfolgreich die Qualitätskontrollen passierenden SNPs wurde jedoch nur einer - rs7738636 auf Chromosomenbande 6q14.1 - als signifikant mit dem ALL-Risiko assoziiert gefunden ($P = 2,55 \times 10^{-3}$). Die kombinierte Analyse der Replikationspanels B bis D für rs7738636 ergab ein P_{CMH} von $3,71 \times 10^{-5}$. Diese Ergebnisse implizieren, dass rs7738636 ein allgemeiner ALL-Suszeptibilitätslocus sein könnte.

In weiteren Analysen typisierten wir die von Papaemmanuil et al. und Treviño et al. in zwei publizierten nicht subgruppenspezifischen GWAS zur ALL im Kindesalter identifizierten SNPs in einer Teilmenge unserer Panels A bis D, um sowohl die zuvor berichteten Loci zu bestätigen, als auch eine Positivkontrolle für die von uns durchgeführten Analysen vorzunehmen. Diese Teilmenge bestand aus Fällen des Panels A und den kompletten Panels B bis D.

Tabelle 3. Assoziationsanalyseergebnisse bereits bekannter Risikoloci

	Chr Pos. (bp)	dbSNP ID Genname	A1 A2	Panel A* (nur Fälle) Deutschland ETV6-RUNX1-pos. — 419 Fälle	Panel B Deutschland/Österreich ETV6-RUNX1-pos. 1682 Kontrollen 406 Fälle	Panel C Italien ETV6-RUNX1-pos. 579 Kontrollen 287 Fälle	Panel A* + B + C Kombinierte Analyse ETV6-RUNX1-pos. 2261 Kontrollen 1112 Fälle			Panel D Deutschland ETV6-RUNX1-neg. — 326 Fälle	Panel A* + B + C + D Kombinierte Analyse ETV6-RUNX1-pos. + -neg. 2261 Kontrollen 1438 Fälle		
				AF _{A1} Fälle	AF _{A1} Kontr. AF _{A1} Fälle	AF _{A1} Kontr. AF _{A1} Fälle	P _{CMH}	P _{Bo}	OR (95% CI)	AF _{A1} Fälle	P _{CMH}	P _{Bo}	OR (95% CI)
Papaemmanuil et al., 2009	7	rs6964823 <i>IKZF1</i>	A	—	0.49	0.51	3.03×10⁻⁰³	0.95	0.83	0.44	8.34×10⁻⁰⁴	0.99	0.83
	50,427,590	—	G	0.43	0.44	0.47			(0.73-0.94)				(0.75-0.93)
	7	rs4132601 <i>IKZF1</i>	G	—	0.27	0.26	1.77×10⁻⁰⁸	0.77	1.46	0.35	2.66×10⁻¹¹	0.84	1.47
	50,438,098	—	T	0.34	0.35	0.34			(1.28-1.67)				(1.31-1.65)
	7	rs6944602 <i>IKZF1</i>	A	—	0.19	0.22	2.99×10⁻⁰⁷	0.29	1.45	0.26	2.91×10⁻⁰⁹	0.31	1.45
	50,441,245	—	G	0.26	0.27	0.27			(1.26-1.67)				(1.28-1.64)
	7	rs3779084 <i>DDC</i>	C	—	0.20	0.16	8.06×10⁻⁰⁴	0.12	1.29	0.25	3.61×10⁻⁰⁵	0.19	1.31
	50,536,229	—	T	0.25	0.23	0.22			(1.11-1.5)				(1.15-1.49)
	7	rs880028 <i>DDC</i>	C	—	0.20	0.16	9.16×10⁻⁰⁴	0.06	1.29	0.24	1.35×10⁻⁰⁴	0.08	1.28
	50,537,630	—	T	0.25	0.23	0.22			(1.11-1.5)				(1.13-1.46)
	7	rs7809758 <i>DDC</i>	G	—	0.34	0.32	3.57×10⁻⁰⁵	0.30	1.31	0.41	1.83×10⁻⁰⁷	0.48	1.33
	50,540,827	—	A	0.40	0.39	0.40			(1.15-1.48)				(1.20-1.49)
	10	rs7073837 <i>ARID5B</i>	A	—	0.37	0.43	1.89×10⁻⁰⁵	0.68	1.31	0.46	8.95×10⁻⁰⁸	0.53	1.34
	63,369,901	—	C	0.44	0.44	0.49			(1.16-1.48)				(1.20-1.49)
Trevino et al., 2009	10	rs10740055 <i>ARID5B</i>	C	—	0.47	0.51	6.41×10⁻⁰⁸	0.99	1.33	0.56	1.92×10⁻⁰⁸	0.81	1.36
	63,388,485	—	A	0.53	0.54	0.58			(1.18-1.51)				(1.22-1.51)
	10	rs7089424 <i>ARID5B</i>	G	—	0.32	0.38	7.96×10⁻⁰⁹	0.08	1.45	0.43	4.40×10⁻¹³	0.05	1.49
	63,422,165	—	T	0.38	0.42	0.44			(1.28-1.64)				(1.34-1.66)
	14	rs2239633 <i>CEBPE</i>	T	—	0.50	0.43	9.51×10⁻⁰⁷	0.34	0.73	0.41	1.09×10⁻⁰⁸	0.36	0.73
	22,658,897	—	C	0.42	0.41	0.38			(0.64-0.83)				(0.65-0.81)
	1	rs10873876 <i>SIAT7C</i>	T	—	0.14	0.18	2.43×10⁻⁰⁴	0.40	1.34	0.17	2.01×10⁻⁰⁴	0.64	1.30
	76,544,916	—	C	0.15	0.19	0.22			(1.15-1.58)				(1.13-1.50)
	1	rs6428370	G	—	0.31	0.36	0.10	0.61	1.12	0.35	0.01	0.89	1.15
	195,111,216	—	A	0.31	0.32	0.40			(0.98-1.27)				(1.03-1.29)
	1	rs7554607 <i>RYS2</i>	G	—	0.43	0.46	0.91	0.05	1.01	0.41	0.82	0.04	0.99
	235,333,226	—	A	0.48	0.41	0.51			(0.89-1.14)				(0.89-1.10)
	1	rs1881797 <i>OR2C3</i>	C	—	0.19	0.17	0.40	0.64	1.07	0.16	0.82	0.24	0.98
	245,756,155	—	T	0.17	0.19	0.19			(0.91-1.25)				(0.86-1.13)
	2	rs12621643 <i>KCNE4</i>	T	—	0.28	0.42	0.23	7.54×10 ⁻⁰⁴	1.08	0.32	0.09	8.97×10 ⁻⁰⁴	1.10
	223,626,227	—	G	0.31	0.34	0.37			(0.95-1.23)				(0.98-1.23)
	3	rs9290663 <i>KCNMB2</i>	T	—	0.14	0.23	0.01	0.97	1.22	0.17	5.35×10⁻⁰³	0.96	1.21
	179,912,633	—	A	0.16	0.17	0.27			(1.04-1.42)				(1.06-1.39)
	6	rs11155133	G	—	0.01	0.01	0.65	0.52	1.15	0.01	0.65	0.59	1.13
	141,211,518	—	A	0.01	0.01	0.01			(0.63-2.09)				(0.67-1.92)
	7	rs11978267 <i>IKZF1</i>	G	—	0.27	0.25	5.41×10⁻⁰⁸	0.80	1.45	0.35	1.26×10⁻¹⁰	0.88	1.47
	50,433,798	—	A	0.34	0.34	0.33			(1.27-1.66)				(1.30-1.65)
	7	rs2242041 <i>DDC</i>	G	—	0.09	0.05	3.24×10⁻⁰³	0.64	1.37	0.12	3.69×10⁻⁰⁴	0.69	1.38
	50,496,943	—	C	0.10	0.12	0.08			(1.11-1.69)				(1.16-1.65)
	7	rs2167364 <i>DDC</i>	G	—	0.31	0.29	5.38×10⁻⁰⁶	0.45	1.35	0.37	1.52×10⁻⁰⁷	0.47	1.35
	50,533,321	—	A	0.37	0.37	0.37			(1.19-1.53)				(1.20-1.50)
	10	rs563507 <i>PAR3D3</i>	A	—	0.04	0.07	0.59	0.59	1.08	0.05	0.39	0.77	1.11
	34,857,994	—	G	0.06	0.04	0.08			(0.82-1.42)				(0.87-1.41)
	10	rs10994982 <i>ARID5B</i>	A	—	0.46	0.51	2.40×10⁻⁰⁴	0.99	1.26	0.53	5.64×10⁻⁰⁸	0.88	1.28
	63,380,110	—	G	0.55	0.52	0.57			(1.11-1.42)				(1.15-1.42)
	10	rs10821936 <i>ARID5B</i>	C	—	0.31	0.37	3.78×10⁻¹⁰	0.23	1.49	0.42	2.86×10⁻¹⁴	0.18	1.52
	63,393,583	—	T	0.39	0.42	0.44			(1.32-1.69)				(1.37-1.70)
	12	rs10849033 <i>C12orf5</i>	G	—	0.01	0.02	5.08×10⁻⁰³	0.49	1.74	0.01	0.03	0.95	1.50
	4,295,383	—	A	0.01	0.03	0.03			(1.17-2.59)				(1.04-2.16)
	12	rs2089222 <i>KRTHB5</i>	A	—	0.04	0.03	0.91	0.36	1.02	0.03	0.68	0.23	0.94
	115,487,041	—	G	0.04	0.04	0.03			(0.73-1.43)				(0.70-1.26)
	18	rs1879352	G	—	0.16	0.17	0.93	0.11	1.01	0.15	0.90	0.14	1.01
	2,488,054	—	A	0.14	0.14	0.19			(0.85-1.19)				(0.87-1.17)
	19	rs2191566 <i>ZNF230</i>	C	—	0.29	0.34	0.34	0.03	0.94	0.33	1.00	7.47×10 ⁻⁰³	1.00
	49,203,229	—	A	0.29	0.30	0.29			(0.82-1.07)				(0.89-1.12)
	19	rs6509133 <i>ZNF230</i>	C	—	0.30	0.35	0.51	0.03	0.96	0.33	0.82	8.09×10 ⁻⁰³	1.01
	49,214,753	—	T	0.29	0.32	0.30			(0.84-1.09)				(0.90-1.14)

Legende zu Tabelle 3. Assoziationsanalyseergebnisse bereits bekannter Risikoloci

Die 28 von Papaemmanuil *et al.* and Treviño *et al.* als mit dem ALL-Risiko assoziierten publizierten SNPs wurden in den Fällen des Panels A (Panel A*) sowie den Panels B bis D genotypisiert. SNPs finden sich geordnet nach ihrer chromosomalen Position und getrennt nach Studie. Nukleotidpositionen (**Pos. (bp)**) beziehen sich auf NCBI build 36. Chromosom: **Chr.** **A1** bezeichnet das seltenere Allel, **A2** das häufigere. Allelfrequenzen finden sich in Fällen und Kontrollen für A1 angezeigt (**AF_{A1} Kontr./ AF_{A1} Fälle**). Kombinierte *P*-Werte des Cochran-Mantel-Haenszel-Tests (**P_{CMH}**) und kombinierte Odds Ratios (**OR**) sind gezeigt für Panel A* (nur Fälle), Panels B und C sowie Panels A* bis D. Die Spalte **P_{BD}** zeigt *P*-Werte aus den Breslow-Day-Tests für Heterogenität, bei denen ein signifikanter *P*-Wert signifikante Heterogenität zwischen den Replikationspanels bezüglich OR und Erkrankungsassoziation nahe legt. Signifikante *P*-Werte ($P_{CMH} < 0.05$ [nur wenn $P_{BD} > 0.05$]) sind **fett** und *kursiv* angegeben. SNP-Assoziationen, die auch genomweit in der kombinierten Analyse der Panels A* bis C oder A* bis D Signifikanz bewahrten, sind mit grauer Schattierung unterlegt.

In diesen Analysen wurden in der kombinierten Stichprobe, die 1438 Fälle und 2261 Kontrollen umfasste, genomweit signifikante Assoziationen ($P < 5 \times 10^{-7}$) für *IKZF1*, *DDC*, *ARID5B* und *CEBPE* gefunden (Tabelle 3). Die Assoziation für *SIAT7C* war nicht genomweit signifikant, widerstand aber der Bonferroni-Korrektur für multiples Testen ($P = 1,79 \times 10^{-3}$). Suggestive Hinweise auf eine Assoziation mit dem Risiko für eine ALL in unserer analysierten Gruppe ($P < 0,05$) wurden für 1q31.3 (rs6428370), *KCNMB2* und *C12orf5* gefunden.

F DISKUSSION

Trotz der nur begrenzten Ausgangsfallzahl von nur 419 Patienten, konnten wir in der hier beschriebenen GWAS neue Kandidatenloci für die ALL im Kindesalter identifizieren. Von diesen neuen Suszeptibilitätsloci sind insbesondere *PTPRJ* und *TP63* als interessante Kandidaten hervorzuheben. Da sich der SNP rs17505102 in Intron 1 des *TP63*-Gens befindet, haben wir in weiterführenden Untersuchungen in 47 Patienten die *TP63*-Exons 1 bis 3 sequenziert, die sich im gleichen Kopplungsungleichgewichtsblock wie der Marker-SNP befinden, um festzustellen, ob dieser mit einem Polymorphismus in der kodierenden Region assoziiert sein könnte. Allerdings haben wir keinen solchen SNP der kodierenden Region ermitteln können und werden nun zusätzliche auch funktionelle Analysen durchführen, um den Mechanismus von rs17505102 bezüglich der Modulation des ALL-Risikos besser verstehen zu lernen. Weiterhin wurden zwei validierte nicht-synonyme SNPs im *TP63*-Gen nach "NCBI dbSNP built 130" zur Genotypisierung in den Panels B und C ausgewählt (rs79939938 und rs77231541). Dabei stellte sich heraus, dass sie in unserer Kohorte monomorph vorlagen und somit leider nicht informativ waren.

Das *TP63*-Gen - ein Mitglied der *TP53*-Gen-Familie - kodiert für p63, das eine bedeutende Rolle in der Embryonalentwicklung spielt. *TP63* besitzt zwei Transkriptionsstartstellen. Die hierdurch

bedingten zwei Isoformen tragen entweder die Trans-aktivierungsdomäne (TAp63) oder nicht (Δ Np63). Mäuse, die kein *TP63*-Gen tragen (*TP63*-null), sterben innerhalb weniger Stunden nach der Geburt. Die Tiere zeigen schwerwiegende Verluste von Gliedmaßen sowie Anlagestörungen einer breiten Palette von epithelialen Strukturen in verschiedenen Organsystemen, einschließlich Haut, Brust und Urogenitalsystem. Die genannten Beobachtungen im Tiermodell deuten insgesamt darauf hin, dass p63 als ein wichtiger Regulator von Stammzeleigenschaften – wie z.B. Selbsterneuerung und Proliferation – in sich entwickelnden Epithelien fungiert. Beim Menschen finden sich *TP63*-Mutationen mit ektodermaler Dysplasie, Lippen-Kiefer-Gaumenspalten und anderen körperlichen Fehlbildungen, aber nicht mit einer höheren Tumorzinzidenz verknüpft. Im Einklang mit diesen Beobachtungen sind somatische Mutationen des *TP63*-Gens gegenwärtig nur selten – etwa bei nicht-kleinzelligen Lungenkarzinomen und chronisch-myeloischer Leukämie (CML) – in humanen Krebszellen berichtet worden. Ähnlich verhält es sich mit *TP63*-Keimbahnvarianten – hier ist eine Assoziation (rs710521) bislang nur mit Blasenkrebs beschrieben worden. Trotz der nur vereinzelt berichteten Verbindungen von *TP63* mit neoplastischen Erkrankungen beim Menschen, gibt es im Tiermodell Daten, die eine potentielle Rolle von *TP63* eindrücklich unterstreichen. So entwickeln für *TP63*-Mutationen heterozygote Mäuse maligne Läsionen, was darauf hindeutet, dass *TP63* als Tumorsuppressor fungieren kann. Weiterhin sind in Mäusen, die heterozygot für Mutationen sowohl in *TP53* als auch in *TP63* sind, eine höhere Tumoralast und stärkere Metastasierung im Vergleich zu Mäusen feststellbar, die nur heterozygot für *TP53*-Mutationen sind. Von besonderem Interesse für unsere Studie ist ebenso die Beschreibung von TAp63 in Verbindung mit einem anti-apoptotischen Signalweg, über den in normalen B-Zellen sowie in chronisch lymphatischen Leukämiezellen (CLL) das Überleben der Zellen abhängig vom Zelloberflächenmolekül CD74 reguliert wird. Bei der CLL führt die Hochregulation von TAp63 zu verstärkter *VLA-4*-Integrin-Expression, was letztlich die Migration und die Affinität der Zellen zum Knochenmark erhöht.

PTPRJ ist eine Protein-Tyrosin-Phosphatase vom Rezeptortyp. Sie findet sich in Prozesse wie die normale Regulation von Zellwachstum, Zelldifferenzierung und Zellteilung eingebunden, wurde aber auch im Zusammenhang onkogener Transformation beschrieben. Somatische Mutationen im *PTPRJ*-Gen sind bei Darm-, Brust- und Lungenkrebs beschrieben worden. Auf Ebene der Keimbahnvariation wurde ein spezifischer *PTPRJ*-Haplotyp mit dem Brustkrebsrisiko assoziiert. Letzteres legt nahe, dass *PTPRJ*-Keimbahnvarianten als Suszeptibilitätsallele für Brustkrebs fungieren können. Von besonderem Interesse für die hier von uns berichteten Ergebnisse, zeigen Mäuse mit konstitutiv deletierter transmembranöser Domäne des *PTPRJ*-Gens einen partiellen peripheren Block der B-Zell-Entwicklung. Heterozygote Mäuse zeigen dagegen keinen bemerkenswerten Phänotyp.

Die oben aufgeführte aktuelle Datenlage erlaubt zwar deutlich erkennbar einen möglichen Zusammenhang von *TP63* und *PTPRJ* sowohl mit der normalen B-Zell-Entwicklung als auch mit

der Entwicklung bösartiger Tumoren herzustellen, lässt aber nicht zu, genauere Mechanismen zu postulieren, wie die in unserer Studie gefundene Keimbahnvariation von *TP63* und *PTPRJ* die Anfälligkeit für die *ETV-RUNX1*-positive ALL im Kindesalter moduliert. Dieses bleibt vorerst spekulativ und unterstreicht die Notwendigkeit für nun durchzuführende funktionelle Untersuchungen. Letztere wären auch notwendig, um eine potentielle Modulation eines bislang ungeklärten möglichen strahlungsassoziierten Risikos für die ALL im Kindesalter durch die neuen Suszeptibilitätsloci zu untersuchen.

Über die gerade beschriebenen neuartigen Erkenntnisse hinaus konnten wir in unserer Studie auch die fünf bislang veröffentlichten Suszeptibilitätsloci für die ALL im Kindesalter bestätigen: *IKZF1*, *DDC*, *ARID5B*, *CEBPE* und *SIAT7C*. Nominal signifikante P-Werte ($P < 0,05$) wurden für 1q31.3, *KCNMB2* und *C12orf5* errechnet. Obwohl unsere Studie eine statistische Kraft von mehr als 80% besaß, um alle SNPs/Loci zu identifizieren, die durch Papaemmanuil et al. und Trevino et al. berichtet wurden, konnten einige der bislang publizierten Loci von uns nicht repliziert werden (*RYR2*, *OR2C3*, *KCNE4*, 6q24.1, *PARD3*, *KRTHB5*, 18p11.32, *ZNF230*).

Insgesamt zeigen unsere Ergebnisse erstmalig, dass Keimbahnvariation spezifische Risiken für Subgruppen der ALL im Kindesalter vermitteln kann - in unserem Fall für die *ETV6-RUNX1*-positive ALL. Unsere Ergebnisse legen somit nahe, dass genetische Variation als Risikomodulator einerseits übergreifend für die Gesamtheit der ALL im Kindesalter, andererseits aber auch spezifisch beschränkt auf bestimmte Subtypen von Bedeutung ist.

Die Ergebnisse des Vorhabens wurden im Jahr 2012 in der Zeitschrift *Leukemia* veröffentlicht (Ellinghaus, E. et al. 2012).

G LITERATUR

1. Pui, C.H. & Evans, W.E. Treatment of acute lymphoblastic leukemia. *N Engl J Med* 354, 166-178 (2006).
2. Stanulla, M. & Schrappe, M. Treatment of childhood acute lymphoblastic leukemia. *Semin Hematol* 46, 52-63 (2009).
3. Redaelli, A., Laskin, B.L., Stephens, J.M., Botteman, M.F. & Pashos, C.L. A systematic literature review of the clinical and epidemiological burden of acute lymphoblastic leukaemia (ALL). *Eur J Cancer Care (Engl)* 14, 53-62 (2005).
4. Linabery, A.M. & Ross, J.A. Trends in childhood cancer incidence in the U.S. (1992-2004). *Cancer* 112, 416-432 (2008).
5. Howard, S.C., et al. Childhood cancer epidemiology in low-income countries. *Cancer* 112, 461-472 (2008).
6. Belson, M., Kingsley, B. & Holmes, A. Risk factors for acute leukemia in children: a review. *Environ Health Perspect* 115, 138-145 (2007).

7. Eden, T. Aetiology of childhood leukaemia. *Cancer Treat Rev* 36, 286-297 (2010).
8. Teitell, M.A. & Pandolfi, P.P. Molecular genetics of acute lymphoblastic leukemia. *Annu Rev Pathol* 4, 175-198 (2009).
9. Mullighan, C.G. & Downing, J.R. Genome-wide profiling of genetic alterations in acute lymphoblastic leukemia: recent insights and future directions. *Leukemia* 23, 1209-1218 (2009).
10. Greaves, M. Darwin and evolutionary tales in leukemia. The Ham-Wasserman Lecture. *Hematology Am Soc Hematol Educ Program*, 3-12 (2009).
11. Papaemmanuil, E., et al. Loci on 7p12.2, 10q21.2 and 14q11.2 are associated with risk of childhood acute lymphoblastic leukemia. *Nat Genet* 41, 1006-1010 (2009).
12. Trevino, L.R., et al. Germline genomic variants associated with childhood acute lymphoblastic leukemia. *Nat Genet* 41, 1001-1005 (2009).
13. Prasad, R.B., et al. Verification of the susceptibility loci on 7p12.2, 10q21.2, and 14q11.2 in precursor B-cell acute lymphoblastic leukemia of childhood. *Blood* 115, 1765-1767 (2010).
14. WTCCC. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447, 661-678 (2007).
15. Sherry, S.T., et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* 29, 308-311 (2001).
16. Collavin, L., Lunardi, A. & Del Sal, G. p53-family proteins and their regulators: hubs and spokes in tumor suppression. *Cell Death Differ* 17, 901-911 (2010).
17. Mills, A.A., et al. p63 is a p53 homologue required for limb and epidermal morphogenesis. *Nature* 398, 708-713 (1999).
18. Yang, A., et al. p63 is essential for regenerative proliferation in limb, craniofacial and epithelial development. *Nature* 398, 714-718 (1999).
19. Senoo, M., Pinto, F., Crum, C.P. & McKeon, F. p63 is essential for the proliferative potential of stem cells in stratified epithelia. *Cell* 129, 523-536 (2007).
20. Su, X., et al. TAp63 prevents premature aging by promoting adult stem cell maintenance. *Cell Stem Cell* 5, 64-75 (2009).
21. Celli, J., et al. Heterozygous germline mutations in the p53 homolog p63 are the cause of EEC syndrome. *Cell* 99, 143-153 (1999).
22. Hagiwara, K., McMenamin, M.G., Miura, K. & Harris, C.C. Mutational analysis of the p63/p73L/p51/p40/CUSP/KET gene in human cancer cell lines using intronic primers. *Cancer Res* 59, 4165-4169 (1999).
23. Sunahara, M., et al. Mutational analysis of p51A/TAp63gamma, a p53 homolog, in non-small cell lung cancer and breast cancer. *Oncogene* 18, 3761-3765 (1999).
24. Yamaguchi, H., Inokuchi, K., Sakuma, Y. & Dan, K. Mutation of the p51/p63 gene is associated with blastic crisis in chronic myelogenous leukemia. *Leukemia* 15, 1729-1734 (2001).
25. Kiemeny, L.A., et al. A sequence variant at 4p16.3 confers susceptibility to urinary bladder cancer. *Nat Genet* 42, 415-419 (2010).
26. Flores, E.R., et al. Tumor predisposition in mice mutant for p63 and p73: evidence for broader tumor suppressor functions for the p53 family. *Cancer Cell* 7, 363-373 (2005).

27. Lantner, F., et al. CD74 induces TAp63 expression leading to B-cell survival. *Blood* 110, 4303-4311 (2007).
28. Binsky, I., et al. TAp63 regulates VLA-4 expression and chronic lymphocytic leukemia cell migration to the bone marrow in a CD74-dependent manner. *J Immunol* 184, 4761-4769 (2010).
29. Ruivenkamp, C.A., et al. Ptprij is a candidate for the mouse colon-cancer susceptibility locus Scc1 and is frequently deleted in human cancers. *Nat Genet* 31, 295-300 (2002).
30. Ruivenkamp, C., et al. LOH of PTPRJ occurs early in colorectal cancer and is associated with chromosomal loss of 18q12-21. *Oncogene* 22, 3472-3474 (2003).
31. Lesueur, F., et al. Allelic association of the human homologue of the mouse modifier Ptprij with breast cancer. *Hum Mol Genet* 14, 2349-2356 (2005).
32. Zhu, J.W., Brdicka, T., Katsumoto, T.R., Lin, J. & Weiss, A. Structurally distinct phosphatases CD45 and CD148 both regulate B cell and macrophage immunoreceptor signaling. *Immunity* 28, 183-196 (2008).
33. Krawczak, M., et al. PopGen: population-based recruitment of patients and controls for the analysis of complex genotype-phenotype relationships. *Community Genet* 9, 55-61 (2006).
35. Purcell, S., et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81, 559-575 (2007).
36. Dupont, W.D. & Plummer, W.D. PS power and sample size program available for free on the Internet. *Controlled Clin Trials* 18, (1997).
37. Browning, B.L. & Browning, S.R. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am J Hum Genet* 84, 210-223 (2009).
38. Ellinghaus, E., et al. Identification of germline susceptibility loci in ETV6-RUNX1-rearranged childhood acute lymphoblastic leukemia. *Leukemia* 26, 902-909 (2012). Hier Textbeitrag einfügen

| Verantwortung für Mensch und Umwelt |

Kontakt:

Bundesamt für Strahlenschutz

Postfach 10 01 49

38201 Salzgitter

Telefon: + 49 30 18333 - 0

Telefax: + 49 30 18333 - 1885

Internet: www.bfs.de

E-Mail: ePost@bfs.de

Gedruckt auf Recyclingpapier aus 100 % Altpapier.



Bundesamt für Strahlenschutz