

Haplotypes of DNA repair and cell cycle control genes, X-ray exposure, and risk of childhood acute lymphoblastic leukemia

Anand P. Chokkalingam · Karen Bartley · Joseph L. Wiemels · Catherine Metayer ·
Lisa F. Barcellos · Helen M. Hansen · Melinda C. Aldrich · Neela Guha ·
Kevin Y. Urayama · Ghislaine Scélo · Jeffrey S. Chang · Stacy R. Month ·
John K. Wiencke · Patricia A. Buffler

Received: 8 July 2011 / Accepted: 24 September 2011 / Published online: 11 October 2011
© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract

Background Acute leukemias of childhood are a heterogeneous group of malignancies characterized by cytogenetic abnormalities, such as translocations and changes in ploidy. These abnormalities may be influenced by altered DNA repair and cell cycle control processes.

Methods We examined the association between childhood acute lymphoblastic leukemia (ALL) and 32 genes in DNA repair and cell cycle pathways using a haplotype-based approach, among 377 childhood ALL cases and 448 controls enrolled during 1995–2002.

Results We found that haplotypes in *APEX1*, *BRCA2*, *ERCC2*, and *RAD51* were significantly associated with total ALL, while haplotypes in *NBN* and *XRCC4*, and *CDKN2A* were associated with structural and numerical change subtypes, respectively. In addition, we observed statistically significant interaction between exposure to 3 or

more diagnostic X-rays and haplotypes of *XRCC4* on risk of structural abnormality-positive childhood ALL.

Conclusions These results support a role of altered DNA repair and cell cycle processes in the risk of childhood ALL, and show that this genetic susceptibility can differ by cytogenetic subtype and may be modified by exposure to ionizing radiation. To our knowledge, our study is the first to broadly examine the DNA repair and cell cycle pathways using a haplotype approach in conjunction with X-ray exposures in childhood ALL risk. If confirmed, future studies are needed to identify specific functional SNPs in the regions of interest identified in this analysis.

Keywords Leukemia · Childhood cancer · DNA repair · Genetic susceptibility

Anand P. Chokkalingam and Karen Bartley contributed equally.

Electronic supplementary material The online version of this article (doi:10.1007/s10552-011-9848-y) contains supplementary material, which is available to authorized users.

A. P. Chokkalingam (✉) · K. Bartley · C. Metayer ·
L. F. Barcellos · P. A. Buffler
Division of Epidemiology, School of Public Health, University
of California Berkeley, 1995 University Avenue, Suite 460,
Berkeley, CA 94704, USA
e-mail: anandc@berkeley.edu

J. L. Wiemels
Department of Epidemiology and Biostatistics, University
of California San Francisco, San Francisco, CA, USA

H. M. Hansen · J. K. Wiencke
Department of Neurological Surgery, University of California
San Francisco, San Francisco, CA, USA

Introduction

Leukemia, a heterogeneous group of hematological malignancies, is the most common cancer in children. Nearly all pediatric cases of leukemia are acute in nature,

M. C. Aldrich
Vanderbilt University Medical Center, Vanderbilt University,
Nashville, TN, USA

N. Guha · K. Y. Urayama · G. Scélo
International Agency for Research on Cancer, Lyon, France

J. S. Chang
National Institute of Cancer Research, National Health Research
Institutes, Tainan, Taiwan

S. R. Month
Kaiser Permanente Medical Group, Oakland, CA, USA

and the majority, ~80%, are acute lymphoblastic leukemia (ALL) [1]. ALL is characterized by specific chromosomal abnormalities, such as translocations and changes in ploidy, that may result from unrepaired DNA damage such as double-strand breaks (DSBs) [2, 3]. The few known risk factors for childhood leukemia include exposure to ionizing radiation and some chemotherapeutic agents [4], which are known to cause cellular DNA damage [5]. Repair of damage resulting from exposure to such agents is a critical cellular maintenance function. Therefore, alterations in innate DNA repair, cell cycle, or genomic maintenance processes may play a role in leukemia development [6–8].

In addition to altered DNA repair processes, DNA DSBs can also arise via the inhibition of DNA topoisomerases [9], or immune-cell specific processes such as somatic hypermutation and class-switch recombination [10, 11]. Germline single nucleotide polymorphisms (SNPs) in genes that control these processes may alter their function, resulting in the accumulation of DNA damage and chromosomal instability.

Cancers at numerous sites have been linked to SNPs in specific DNA repair pathways, including nucleotide excision repair (NER), mismatch repair (MMR), and DSB repair and breast cancer [12, 13]. In addition, SNPs in genes that control cell cycling and checkpoints may determine whether cells with substantial DNA damage and instability progress to replication or undergo apoptosis [5, 14].

In order to explore the association between germline genetic variants and ALL, we examined 21 genes in base excision (BER), nucleotide excision (NER), and DSB repair pathways, and 11 genes in cell cycle and genomic maintenance pathways, among 377 childhood ALL cases and 448 controls from a population-based study of childhood leukemia in Northern and Central California. In addition, while the association between high doses of ionizing radiation and leukemia is well established [15, 16], the evidence for low dose sources such as diagnostic X-rays is evolving. As we previously found exposure to diagnostic X-rays postnatally to be a risk factor for childhood leukemia [17], we examined whether the risk of childhood ALL associated with diagnostic irradiation is modified by variants in DNA repair genes.

Materials and methods

Study subjects

The study was conducted among children participating in the Northern California Childhood Leukemia Study (NCCLS), an ongoing population-based case–control study. The study enrollment and recruitment procedures have been described in detail previously [18]. Briefly, case

children under 15 years of age with incident leukemia were ascertained within 72 h of diagnosis via a rapid reporting system established with the diagnosing hospitals. Control children selected from California birth certificates were matched to case children on date of birth, sex, maternal race, and child's Hispanic ethnicity (having at least one parent reporting Hispanic ethnicity). Participation rates among eligible cases and controls were 87 and 86%, respectively. Data on potential risk factors were elicited from an English- or Spanish-speaking parent (usually the mother) by trained interviewers using a structured questionnaire.

This study was reviewed and approved by institutional review committees at the University of California Berkeley, the California Department of Public Health, and the participating hospitals. Written informed consent was obtained from all parent respondents.

Biospecimen collection and DNA processing

Buccal cytobrush specimens successfully collected at interview from 95% of participating children were processed by heating in the presence of 0.5 N NaOH. Isolated DNA was later repurified using an automated DNA extraction system (AutoGen, Holliston, MA), whole-genome amplified using GenomePlex reagents (Rubicon Genomics, Ann Arbor, MI). When buccal cytobrush DNA was inadequate (26.6% of subjects), DNA was isolated from dried bloodspots (DBS) collected at birth and archived by the Genetic Diseases Screening Program of the California Department of Public Health. After extraction (QIAamp 96 DNA Blood Kit, QIAGEN, Germany), these DNA samples were whole-genome amplified using REPLI-g reagents (QIAGEN). Regardless of source, DNA specimens were quantitated using human-specific Alu-PCR to confirm a minimum level of amplifiable human DNA [19]. We have previously shown that highly multiplexed GoldenGate genotyping (Illumina, San Diego, CA) of whole-genome amplified buccal cell DNA yields genotypes that are highly concordant with those from genomic DNA from peripheral blood [20]. We genotyped DNA specimens from both buccal cells and DBS for 9 subjects; genotype concordance between paired samples was 98.9%.

Selection of variants

This analysis included SNPs in 32 genes involved in DNA repair, cell cycle and genomic stability processes. These genes were selected after consensus review by our investigative team after review of the literature. These include base excision repair (*APEX1*, *MUTYH*, *UNG2*, *XRCC1*), nucleotide excision repair (*ERCC2*), double-strand break, nonhomologous end-joining (*LIG4*, *PRKDC*, *XRCC4*,

XRCC5, *XRCC6*), double-strand break, homologous recombination (*BRCA1*, *BRCA2*, *MRE11*, *NBN*, *RAD50*, *RAD51*, *RAD54B*, *RAD54L*, *XRCC2*, *XRCC3*), direct reversal of damage (*MGMT*), cell cycle (*TP53*, *TP53BP1*, *CCND1*, *CDKN2A* (*p16*), *CDKN2B* (*p15*)), topoisomerases (*TOP1*, *TOP2A*, *TOP2B*), and recombination of immunoglobulin and T-cell receptor genes (*AICDA*, *RAG1*, *RAG2*).

Using HaploView [21] in conjunction with SNP data from the 30 Caucasian trios in the HapMap project (Release 19, Build 34, www.hapmap.org) and the 23 Hispanics in the SNP500Cancer project (www.snp500cancer.nci.nih.gov), we selected haplotype-tagging SNPs (htSNPs) that captured at least 80% of the haplotype diversity for common haplotypes (>5% frequency) in either the Caucasian or Hispanic populations [22]. As Hispanics are a recently admixed ethnic group and comprise 42% of our study population, we placed special emphasis on capturing haplotype structures in this population. To maximize capture of potential regulatory regions, we included 10 kb stretches both up- and downstream from the gene boundaries reported in the UCSC Genome Browser. In addition, we identified non-synonymous SNPs within these genes. Finally, a set of ancestry informative markers (AIMs) was included for SNP genotyping; these had been previously identified to distinguish Amerindian, African, and European populations [23], three populations that make up the genetic ancestry of US Hispanics.

Genotyping

Although the study recruitment was originally of an individually matched design, we performed a frequency-matched analysis by including all subjects with available biospecimens from 1995 through 2002. Accordingly, we genotyped whole-genome amplified DNAs of 385 available ALL cases and 456 available controls enrolled from 1995 to 2002 using a custom Illumina GoldenGate assay panel with a GenCall threshold of 0.25. Fifty-nine duplicate samples that were processed in the same manner and genotyped on the same plate showed >99% concordance in genotype. For these duplicates, the sample with the higher SNP call rate within each pair was included in the final data. SNPs were excluded if the call rate was less than 90% (29 SNPs), had minor allele frequencies less than 5% in both Hispanics and non-Hispanics (6 SNPs), or failed Hardy–Weinberg equilibrium ($p < 0.01$) in both Hispanic and non-Hispanic controls (1 SNP). After applying these data quality thresholds, data for 238 SNPs in the 32 selected genes were available for 377 ALL cases and 448 controls. A total of 80 AIMs were successfully typed in these subjects.

Cytogenetic characterization

The cytogenetic classification methods used in this analysis have been described in detail elsewhere [24]. Briefly, pretreatment diagnostic karyotype and fluorescence in situ hybridization (FISH) data were abstracted from leukemia patient records shortly after diagnosis. Additional FISH analyses were conducted at the University of California, Berkeley, to identify *ETV6-RUNX1* [t(12; 21)] translocations and hyperdiploidy when not done at hospitals. As shown in Table 1, the presence of any structural or numerical change was common among the NCCLS case population. The most frequent structural abnormality, t(12;21), was more common among non-Hispanics than Hispanics. Among cases with a numerical chromosomal change of any kind, high hyperdiploidy (>51 chromosomes) was most commonly observed. Because of the low prevalence for most other individual structural and numeric changes, cytogenetic subgroup analyses were limited to t(12;21), high hyperdiploidy, and the broader categories “any structural change” and “any numerical change”.

Diagnostic X-ray exposure assessment

Information on child’s X-rays received prior to the date of diagnosis for cases or corresponding reference date for matched controls (hereafter called “postnatal X-rays”), was collected during the in-person interview as described previously [17]. All postnatal diagnostic X-ray exposures, with the exception of dental X-rays, were reported by the following broadly defined regions of the body: chest, skull, broken bone, and “other”. Respondents also reported the number of postnatal X-rays received and the age at first X-ray. Postnatal X-ray exposure information was available for 746 (90.4%) of the genotyped study participants. In our genotyped study sample, having 3+ x-diagnostic X-rays postnatally was associated with a significantly increased risk of childhood ALL (OR = 2.49, 95% CI 1.55–3.98), similar to our previous report on a larger sample size [25].

Statistical analysis

Based on the AIMs, individual estimates of genetic ancestry, i.e., percent contribution of each of the three ancestral populations per person, were obtained from maximum likelihood estimation as described previously [26]. Using the confounding relative risk (CRR) [27], we found no evidence of major confounding by estimated genetic ancestry (>10%) over and above adjustment for self-identified race and ethnicity. As a result, our analysis proceeded with stratification by or adjustment for self-identified race and ethnicity.

Table 1 Demographic and cytogenetic characteristics of Hispanic and Non-Hispanic children, the Northern California Childhood Leukemia Study, 1995–2002

| | Non-Hispanic | | | Hispanic | | |
|--|--|---|----------------|--|---|----------------|
| | Case (<i>n</i> = 221) <i>n</i> (%) | Control (<i>n</i> = 269) <i>n</i> (%) | <i>p</i> Value | Case (<i>n</i> = 156) <i>n</i> (%) | Control (<i>n</i> = 179) <i>n</i> (%) | <i>p</i> Value |
| Male gender, <i>n</i> (%) | 123 (55.7) | 150 (55.8) | 0.9 | 77 (49.4) | 87 (48.6) | 0.6 |
| Age, years (SE) | 5.6 (3.4) | 5.5 (3.6) | 0.7 | 5.5 (3.4) | 5.4 (3.5) | 0.8 |
| Maternal race, <i>n</i> (%) | | | | | | |
| White | 162 (73.3) | 198 (73.6) | 0.5 | 52 (33.3) | 57 (31.8) | 0.4 |
| Black | 13 (5.9) | 14 (5.2) | | 2 (1.3) | 2 (1.1) | |
| Native american | 1 (0.4) | 1 (0.4) | | 6 (3.8) | 7 (3.9) | |
| Asian/pacific islander | 25 (11.3) | 33 (12.2) | | 0 (0) | 2 (1.1) | |
| Mixed/other | 20 (9.1) | 23 (8.6) | | 96 (61.5) | 111 (62.0) | |
| Post-natal X-rays, <i>n</i> (%) ^a | | | | | | |
| 0–2 | 173 (87.4) | 233 (93.2) | 0.04 | 112 (81.8) | 154 (95.7) | <0.01 |
| 3+ | 25 (12.6) | 17 (6.8) | | 25 (18.2) | 7 (4.3) | |
| ALL Cytogenetic characteristics | | | | | | |
| Any structural change | 106 (48.0) | N/A | | 63 (40.4) | N/A | |
| <i>ETV6-RUNX1</i> [t(12;21)] | 43 (19.5) | N/A | | 19 (12.2) | N/A | |
| Any numerical ploidy change | 119 (53.8) | N/A | | 90 (57.7) | N/A | |
| High hyperdiploidy (>51 chromosomes) | 95 (40) | N/A | | 51 (56) | N/A | |

^a x-ray exposure information available for 746 (90.4%) of study subjects

As a preliminary step prior to haplotype analysis, we tested for potential interactions of individual SNPs with Hispanic ethnicity using the likelihood ratio test at the 0.05 significance level. We used unconditional logistic regression to estimate odds ratios (ORs) for the log-additive associations of individual SNPs, both overall and by cytogenetic subtype, after adjusting for age at diagnosis, sex, and child's race, plus child's Hispanic ethnicity in analyses combining Hispanics and non-Hispanics.

For haplotype analyses, we applied a sliding window approach for each gene, as implemented in the haplo.stats package for R [28], using window sizes of 2–5 SNPs. This approach examines sub-haplotypes using the full set of SNP data, with differently sized “windows” of adjacent alleles. This is an effective means of combining multi-locus data for Hispanics and non-Hispanics, as it is agnostic to differences in haplotype structure, provided no individual SNPs for a given gene show significant effect heterogeneity by Hispanic ethnicity ($p_{\text{interaction}} \leq 0.05$). If none of the individual SNPs in a given gene showed such heterogeneity, data for both ethnic groups were combined for sliding window analyses; for genes in which significant single-SNP heterogeneity by ethnicity existed, the sliding window analysis for that gene was conducted separately for Hispanics and non-Hispanics. We utilized GrASP, a graphical tool [29], to display and visualize sliding window results. We used haplotype trend regression [30] to estimate the magnitude of effect associated with risk

haplotypes of the windows with the smallest global *p* values.

Assessment of potential interactions with exposure to ionizing radiation was limited to single SNPs or haplotypes with significant main effects; the significance of these was assessed in logistic regression models using the likelihood ratio test. The number of postnatal X-rays received by the child was modeled as a dichotomous variable (0–2 vs. 3+ X-rays). Postnatal X-rays that occurred less than a year prior to the date of diagnosis for cases or corresponding reference date for controls were excluded from analysis. Maternal X-rays received during the year prior to conception and during pregnancy were excluded from analysis due to low prevalence of exposure in the study population.

Results

Characteristics of the study population are presented in Table 1 stratified by Hispanic ethnicity. Cases and controls were similar with respect to age, sex, and maternal race.

Of the 32 genes examined, 6 included one or more individual SNPs with effects that differed significantly between Hispanics and non-Hispanics (*APEX1*, *BRCA1*, *BRCA2*, *MGMT*, *RAD51*, and *RAG2*). Accordingly, haplotype sliding window analyses for these 6 genes were stratified by Hispanic ethnicity, while those for the other 26 genes were conducted with both ethnic groups combined.

Total ALL and subtype-specific ALL results for genes with significant ($p \leq 0.05$) haplotype effects that persisted through increasingly larger windows are presented in the Supplemental Material. Haplotype trend regression results estimating the magnitudes of effect for haplotypes with the lowest multi-SNP p value in sliding window analyses are shown in Table 2 for total ALL, and in Table 3 for specific ALL subtypes.

For total ALL, among both ethnicities combined, *ERCC2* showed a significant haplotype association that persisted through progressively larger SNP windows, with haplotype G-A-A showing a significantly reduced risk (OR = 0.59, $p = 0.018$). Among Hispanics, haplotypes A–A–A and A-G-A of *RAD51* were significantly associated with increased risks of total ALL (OR = 1.55 and $p = 0.05$, OR = 1.51 and $p = 0.04$, respectively). Among non-Hispanics, *APEX1* and *BRCA2* showed significant haplotype associations with total ALL: haplotype A–A of *APEX1* was significantly associated with an increased risk (OR = 1.90, $p = 0.003$), and haplotype G-A of *BRCA2* was significantly associated with an increased risk (OR = 1.77 $p = 0.02$).

We performed haplotype analyses for specific ALL subtypes for genes with nominally significant associations among both ethnicities combined (Table 3). Two genes (*NBN* and *XRCC4*) showed significant associations with t(12;21) translocation-positive childhood ALL (global $p \leq 0.05$). For *NBN*, haplotype A–A-G-A-G was borderline significantly associated with a reduced risk (OR = 0.55, $p = 0.057$). For *XRCC4*, haplotypes C-G-G-G-A and C-G-A-G-A were significantly associated with reduced risk (OR = 0.39 and $p = 0.039$, OR = 0.56 and $p = 0.05$, respectively). *XRCC4* was also significantly associated with childhood ALL with any structural abnormalities: haplotypes G-A-G and G–G–G were both significantly associated with a reduced risk (OR = 0.60 and $p = 0.006$, and OR = 0.55 and $p = 0.012$, respectively). Of particular interest, the two 3-SNP risk haplotypes of *XRCC4* are part of the two significant 5-SNP risk haplotypes associated with t(12;21) translocation-positive childhood ALL, and have a similar effect.

For high-hyperdiploid-positive childhood ALL, and more broadly for childhood ALL with any numerical ploidy changes, there were significant haplotype associations with

Table 2 Significant haplotype trend regression results (global $p \leq 0.05$), for total childhood ALL, by ethnicity, NCCLS, 1995–2002

| Genes/haplotypes | Haplotype frequency ^a (%) | OR (95% CI) ^b | p Value | Global p Value ^c |
|---------------------------------|--------------------------------------|--------------------------|-----------|-------------------------------|
| <i>ERCC2</i> —both ethnicities | | | | |
| rs3916874, rs238416, rs171140 | | | | |
| GGA | 37.0 | Reference | N/A | 0.02 |
| CGA | 12.0 | 0.96 (0.60, 1.52) | 0.853 | |
| CGC | 9.0 | 1.11 (0.80, 1.53) | 0.537 | |
| GAA | 4.0 | 0.59 (0.38, 0.91) | 0.018 | |
| GAC | 36.0 | 1.65 (0.92, 2.95) | 0.094 | |
| GGC | 2.0 | 0.91 (0.72, 1.16) | 0.461 | |
| <i>APEX1</i> -non-Hispanics | | | | |
| rs11160711, rs3120073 | | | | |
| GA | 46.0 | Reference | N/A | 0.004 |
| AA | 29.0 | 1.90 (1.25, 2.89) | 0.003 | |
| GG | 25.0 | 1.58 (0.94, 2.67) | 0.090 | |
| <i>BRCA2</i> —non-Hispanics | | | | |
| rs4942448, rs9943876 | | | | |
| GG | 71.0 | Reference | N/A | 0.006 |
| GA | 8.0 | 1.77 (1.10, 2.85) | 0.020 | |
| AA | 21.0 | 0.73 (0.52, 1.05) | 0.090 | |
| <i>RAD51</i> —Hispanics | | | | |
| rs2304579, rs1717265, rs2304580 | | | | |
| AAG | 46.0 | Reference | N/A | 0.03 |
| AAA | 21.0 | 1.55 (1.01, 2.42) | 0.050 | |
| AGA | 24.0 | 1.51 (1.01, 2.26) | 0.040 | |
| GAA | 8.0 | 0.58 (0.27, 1.25) | 0.170 | |
| Rare haplotypes ^d | 1.0 | 1.83 (0.40, 8.27) | 0.430 | |

^a Haplotype frequency among the 448 controls. All case subtypes were compared to this control group

^b ORs were adjusted for age, sex, and race (plus Hispanic ethnicity in analyses of both ethnicities combined) using unconditional logistic regression

^c Global p values for haplotype trend regression in other ethnic group: *APEX1* haplotype in Hispanics, $p = 0.09$; *BRCA2* haplotype in Hispanics, $p = 0.95$; *RAD51* haplotype in non-Hispanics, $p = 0.32$

^d Haplotypes with frequencies less than 1% were grouped together for analysis

Table 3 Significant haplotype trend regression results (global $p \leq 0.05$), by major childhood ALL subtype, both ethnicities combined, NCCLS, 1995–2002

| Genes/haplotypes | Haplotype frequency ^a (%) | OR (95% CI) ^b | p Value | Global p Value |
|---|--------------------------------------|--------------------------|-----------|------------------|
| T12;21 ($n = 62$ cases) | | | | |
| <i>NBN</i> | | | | |
| rs12680687, rs6470522, rs7840099, rs1805812, rs709816 | | | | |
| GCGGA | 40 | Reference | N/A | 0.04 |
| AAGAA | 17 | 0.55 (0.30, 1.01) | 0.057 | |
| AAGAG | 6 | 1.23 (0.63, 2.42) | 0.539 | |
| GCAGA | 16 | 0.64 (0.35, 1.17) | 0.147 | |
| GCGAA | 8 | 0.52 (0.22, 1.22) | 0.134 | |
| Rare haplotypes ^c | 13 | 0.38 (0.16, 0.88) | 0.025 | |
| <i>XRCC4</i> | | | | |
| rs7711825, rs1193695, rs301276, rs301287, rs3777018 | | | | |
| CAGGA | 34 | Reference | N/A | 0.006 |
| AAGGA | 4 | 1.85 (0.87, 3.94) | 0.109 | |
| AAGGG | 6 | 0.62 (0.24, 1.64) | 0.339 | |
| CGAGA | 21 | 0.56 (0.31, 1.00) | 0.050 | |
| CGGAA | 6 | 0.91 (0.40, 2.09) | 0.821 | |
| CGGGA | 11 | 0.39 (0.16, 0.95) | 0.039 | |
| Rare haplotypes ^c | 15 | 0.35 (0.16, 0.74) | 0.007 | |
| Any structural change ($n = 169$ cases) | | | | |
| <i>XRCC4</i> | | | | |
| rs1193695, rs301276, rs301287 | | | | |
| AGG | 47 | Reference | N/A | 0.01 |
| AGA | 6 | 0.65 (0.35, 1.20) | 0.170 | |
| GAG | 23 | 0.60 (0.42, 0.86) | 0.006 | |
| GGA | 12 | 0.83 (0.56, 1.24) | 0.365 | |
| GGG | 13 | 0.55 (0.35, 0.88) | 0.012 | |
| Hyperdiploidy ($n = 163$ cases) | | | | |
| <i>CDKN2A</i> | | | | |
| rs3731257, rs2518719 | | | | |
| GA | 60 | Reference | N/A | <0.001 |
| AA | 29 | 0.73 (0.51, 1.05) | 0.090 | |
| GG | 11 | 0.30 (0.14, 0.64) | 0.002 | |
| Any numerical ploidy change ($n = 209$ cases) | | | | |
| <i>CDKN2A</i> | | | | |
| rs3731257, rs2518719 | | | | |
| GA | 63 | Reference | N/A | <0.001 |
| AA | 28 | 0.67 (0.50, 0.90) | 0.008 | |
| GG | 8 | 0.44 (0.26, 0.73) | 0.001 | |

^a Haplotype frequency among the 448 controls. All case subtypes were compared to this control group

^b ORs were adjusted for age, sex, Hispanic ethnicity and race using unconditional logistic regression

^c Haplotypes less with frequencies less than 5% were grouped together for analysis

just one gene: *CDKN2A*. The G–G haplotype was significantly associated with a reduced risk of hyperdiploid-positive disease (OR = 0.30, $p = 0.002$), while risk reduction associated with the A–A haplotype was marginally significant (OR = 0.73, $p = 0.09$). Looking more broadly at childhood ALL with any numerical changes, these same haplotypes showed significant associations (OR = 0.44 and $p = 0.01$, OR = 0.67 and $p = 0.008$, respectively).

In testing gene-environment interactions, we investigated interactions of postnatal X-ray exposures (0–2 vs. 3 or more) with those haplotypes that showed nominally significant main effects in both ethnicities combined, both for total ALL and by major ALL subtype. At the $p \leq 0.05$ level, we found the main effect of the G–G–G haplotype of *XRCC4* with ALL with any structural abnormalities was significantly modified by number of postnatal X-rays ($p = 0.027$). No such interactions were observed for *NBN* or *ERCC2*.

Discussion

In this study, we utilized a haplotype-tagging approach to examine the role of variants in genes involved in DNA repair and genomic maintenance in risk of childhood ALL. We found significant haplotype associations with total ALL for *ERCC2*, *RAD51*, *APEX1*, and *BRCA2*; and with specific ALL subtypes for *NBN*, *XRCC4*, and *CDKN2A*. In addition, we observed significant gene-environment interactions between *XRCC4* variants and exposure to diagnostic X-rays modulating the risk of structural abnormality-positive childhood ALL. Our results provide strong support for a role of the DNA repair and cell cycle control pathways in risk of childhood ALL.

APEX1 encodes the major apurinic/aprimidinic (AP) endonuclease in human cells. AP sites occur frequently in DNA as a result of spontaneous hydrolysis, damaging agents, or glycosylases that remove specific abnormal bases; therefore, *APEX1* plays an important role in base excision repair. Gene expression analyses of *APEX1* have shown high levels of expression in many different types of tumors, including osteosarcoma, ovarian, and digestive cancers ([31] Endonuclease, [32]). The two SNPs involved in the observed *APEX1* risk haplotype were borderline significantly associated with total childhood ALL after adjustment for the false discovery rate ($p_{FDR} = 0.06$ for both) [33]. Of these, rs3120073, whose variant allele was associated with decreased risk ALL among both ethnicities combined, is located 6.5 kb upstream of *APEX1*, in intron 6 of *OSGEP*, a probable endopeptidase. Analysis of HapMap CEU population data shows that this SNP is not in strong linkage disequilibrium ($r^2 > 0.80$) with any nearby SNPs (<10 kbp), though HapMap genotyping data for this variant was available for less than 50% of the CEU population. The other involved SNP, rs11160711, whose variant allele was associated with an increased risk of total childhood ALL among non-Hispanics, is 10kb upstream of *APEX1*. Neither SNP has known function. Additional studies are warranted to replicate these findings and identify functional variants that may be in strong linkage disequilibrium with these SNPs.

The *ERCC2* gene product is involved in nucleotide excision repair. Defects in this gene can cause rare genetic syndromes, including the cancer-prone syndrome xeroderma pigmentosum complementation group D, which necessitates protection from ultraviolet light [34]. We found a significant haplotype association for this gene with total childhood ALL risk. Separately, the *BRCA2* gene product, a tumor suppressor, interacts with the *RAD51* gene product, a recombinase, to effect homologous repair of double-strand breaks [35]. We observed significant haplotype associations for *BRCA2* and *RAD51*, albeit in different ethnic groups. These results must be confirmed, but support a role for DSB repair pathways in risk of childhood ALL.

The X-ray repair cross-complementing protein encoded by *XRCC4* is involved in nonhomologous end-joining repair of double-strand DNA breaks and the completion of V(D)J recombination events [36, 37]. In addition to the significant haplotype associations with this gene for ALL with any structural changes, we observed a significant interaction between the *XRCC4* risk haplotype and exposure to postnatal X-rays on risk of the same disease subtype. These effects appear to be driven largely by intronic SNP rs1193695, which showed nominally significant main effects and interactions with postnatal X-ray exposure for both total ALL as well as ALL subtypes defined by any structural changes and any numerical changes. Our results provide compelling support for a role of the nonhomologous end-joining repair pathway in risk of ALL, particularly ALL with structural abnormalities, both alone and in conjunction with exposure to ionizing radiation.

Other genes showed significant subtype-specific haplotype associations as well. Haplotypes of *NBN* were associated with t(12;21) positive childhood ALL. *NBN* is part of the *MRE11/RAD50* DSB repair complex, involved in homologous DSB repair. Mutations in this gene can lead to Nijmegen breakage syndrome, a chromosomal instability syndrome that predisposes to cancer, among other diseases [38]. These results for *NBN* did not extend more broadly to ALL with any structural changes.

In contrast, haplotypes of *CDKN2A* were associated with both hyperdiploid ALL and ALL with any numerical ploidy changes. *CDKN2A* is a cell cycle control gene recognized as a tumor suppressor for its role in stabilizing p53. A study of secondary hits from a genome-wide association study found *CDKN2A* SNP rs3731217, which was not included for genotyping in our study, to be associated with childhood ALL, specifically B-cell precursor disease [39]. Our results support these previous findings for a role of *CDKN2A* in childhood ALL risk, and suggest that further studies should examine effects by ploidy.

Few previous candidate gene studies have investigated the main effects of DNA repair pathway gene variants in the etiology of childhood ALL, and those that have were typically of smaller sample size (less than 200 cases) and focused on analyses of individual SNPs with putative function (i.e., inducing amino acid changes or located in potential regulatory regions). One study reported significant association of a promoter variant of *CDKN2A* with childhood pre-B-cell ALL [40], while another reported an association of a functional *ERCC1* variant with total childhood ALL ([41] 2006). Other studies have found significant main effects of functional SNPs in *XRCC1* [42, 43], and a significant haplotype combining three of these *XRCC1* functional variants of *XRCC1* has also been observed [42]. In our study, we did not observe a significant haplotype association for *XRCC1* or *ERCC1* though

we did observe significant effects for *CDKN2A*. Inconsistencies between our study and previous candidate gene studies may be attributable to a number of factors, including the differences in approach (haplotype vs. individual SNP), population and/or allele frequency differences and sample size differences that impact power to detect effects, as well as chance.

To our knowledge, this study is the first to examine the joint effects of genes in the DNA repair and cell cycle control pathways with ionizing radiation on risk of childhood ALL using a haplotype approach. We focused on interactions involving haplotypes with significant main effects, as the biological basis is unclear for interactions involving significant subgroup variation in exposure–disease associations with no main effects when subgroups are combined [44]. We also limited our investigation of potential interactions to those haplotypes with significant main effects in both Hispanics and non-Hispanics combined, as the sizes of the individual ethnic groups were considered too small to permit adequately powered examinations of interactions. A previous study focusing on single SNPs using a case-only interaction study design reported a suggestive interaction between an *APEXI* SNP and postnatal X-ray exposure [45]. Per our a priori analysis plan, we did not examine *APEXI* haplotype interactions since this gene showed significant main effects in only non-Hispanics. In this regard, we recognize that our total sample size (343 cases, 406 controls with both genetic and exposure data) may be insufficient to observe modest interaction effects with adequate statistical power. In addition, although the role of post-natal diagnostic X-rays in childhood ALL risk is not entirely clear [16], we previously found exposure to postnatal X-rays to be significantly associated with childhood ALL (OR = 1.85, 95%CI: 1.22–2.79 for 3+ vs. 0–2 X-rays) in a larger sample size ($n = 711$ cases), of which the current study population is a subset [17]. However, these measures are derived from maternal reports of the child's exposures and exposure periods and are potentially subject to reporting errors. Further studies with improved measures of ionizing radiation exposure, perhaps via review of medical records, as well as data on other potential DNA damaging exposures and larger sample sizes, are needed to confirm the associations observed.

One of the strengths of our study is the inclusion of U.S. Hispanics, an understudied population whose childhood leukemia incidence rates are the highest reported in California [46]. We selected SNPs in a manner to maximize capture of genetic variation in Hispanics, and examined Hispanics separately from non-Hispanics where there was significant heterogeneity in between-group effects of individual SNPs. Although this approach may have limited our ability to detect associations in the population as a whole, we believe it was necessary given that genetic susceptibility and/or patterns of linkage disequilibrium may be

different in Hispanics versus non-Hispanics due to the Hispanic population's relatively recent genetic admixture [23]. Results that differ between Hispanics and non-Hispanics may be due to differences in allele frequency and/or haplotype structure, or may reflect underlying differences in exposures that modulate the effects of genes. Regardless, if the results are not spurious, they represent potential risk loci, and we present them in either or both ethnic groups for replication and further followup. A final point is that the limited size of certain racial/ethnic sub-populations among non-Hispanics precluded further stratification of this group; therefore, heterogeneity among non-Hispanics might have obscured results.

In gauging these results, consideration must be given to several factors. First, despite this study's relatively large sample size compared to those of most previous candidate gene studies, the presence of genetic heterogeneity due to the ethnic and racial diversity of the California population may have influenced our ability to detect associations. However, in this study population, we found no evidence of strong confounding due to estimated genetic ancestry, minimizing concerns about the impact of population stratification on the results. In addition, as noted above, our findings for *CDKN2A* confirm those from analyses of secondary hits from a recent genome-wide association study [39]. However, associations identified in our study for other genes, including *APEXI*, *ERCC2*, and others, were not observed in the primary or secondary hits of the large genome-wide studies conducted to date [39, 47, 48]. This may be due to stringent multiple testing adjustment (at the $p \leq 1 \times 10^{-7}$ level) to account for the large number of individual variants investigated in the genome-wide studies. In contrast to the agnostic approach to discovery used in genome-wide studies, our study focused on relatively few genes representing key elements of the DNA repair and cell cycle control pathways. Other DNA repair genes not included in our study may also be associated with disease; we are unable to comment on these. We concede that results of our study may be due to chance, and therefore must be replicated. However, the haplotype-tagging approach we adopted maximizes capture of total variation within each candidate gene and the haplotype analysis increases statistical power to detect associations over analyses of individual variants. Finally, differences in genetic risk factors by cytogenetic subtypes may be obscured in studies that do not stratify analyses by these subtypes.

In summary, our results indicate that elements of the DNA repair and cell cycle pathways are likely to be associated with childhood ALL, and that some of these elements may interact with ionizing radiation exposures to modulate risk. The associations and interactions identified should be considered targets for further analysis in studies with larger sample sizes, high quality environmental

exposure data, and finer coverage of SNPs in the identified associated regions.

Acknowledgments We thank our clinical collaborators and NCCLS participating hospitals: University of California Davis (Dr. J. Ducore), University of California San Francisco (Drs. M. Loh and K. Matthay), Children's Hospital of Central California (Dr. V. Crouse), Lucile Packard Children's Hospital (Dr. G. Dahl), Children's Hospital Oakland (Dr. J. Feusner), Kaiser Permanente Roseville (Drs. K. Jolly and V. Kiley), Kaiser Permanente Santa Clara (Drs. A. Wong and C. Russo), Kaiser Permanente San Francisco (Dr. K. Leung), and Kaiser Permanente Oakland (Drs. D. Kronish and S. Month). We thank the entire NCCLS staff and the UC Berkeley Survey Research Center for their effort and dedication. Finally, we thank the families who participated in the NCCLS for their strong support and selflessness, without which this research could not have been conducted. We acknowledge funding support from the National Institute of Environmental Health Sciences (PS42ES04705 and R01ES09137) and the Children with Cancer UK Foundation (2005/027, 2005/028, 2006/053). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIEHS, NIH or the Children with Cancer UK Foundation.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Bhatia S, Ross JA, Greaves MF, Robinson LL (1999) Epidemiology and etiology. In: Pui CH (ed) *Childhood leukemias*. Cambridge University Press, New York
- Greaves MF, Wiemels J (2003) Origins of chromosome translocations in childhood leukaemia. *Nat Rev Cancer* 3:639–649
- Gillert E, Leis T, Repp R et al (1999) A DNA damage repair mechanism is involved in the origin of chromosomal translocations t(4;11) in primary leukemic cells. *Oncogene* 18:4663–4671
- Buffler PA, Kwan ML, Reynolds P, Urayama KY (2005) Environmental and genetic risk factors for childhood leukemia: appraising the evidence. *Cancer Invest* 23:60–75
- Kastan MB, Bartek J (2004) Cell-cycle checkpoints and cancer. *Nature* 432:316–323
- Hoeijmakers JH (2001) Genome maintenance mechanisms for preventing cancer. *Nature* 411:366–374
- Papaefthymiou MA, Giaginis CT, Theocharis SE (2008) DNA repair alterations in common pediatric malignancies. *Med Sci Monit* 14:RA8–RA15
- Friedenson B (2007) The BRCA1/2 pathway prevents hematologic cancers in addition to breast and ovarian cancers. *BMC Cancer* 7:152
- Froelich-Ammon SJ, Osheroff N (1995) Topoisomerase poisons: harnessing the dark side of enzyme mechanism. *J Biol Chem* 270:21429–21432
- Schlissel MS, Kaffer CR, Curry JD (2006) Leukemia and lymphoma: a cost of doing business for adaptive immunity. *Genes Dev* 20:1539–1544
- Novara F, Beri S, Bernardo ME et al (2009) Different molecular mechanisms causing 9p21 deletions in acute lymphoblastic leukemia of childhood. *Hum Genet* 126:511–520
- Lindahl T, Wood RD (1999) Quality control by DNA repair. *Science* 286:1897–1905
- Goode EL, Ulrich CM, Potter JD (2002) Polymorphisms in DNA repair genes and associations with cancer risk. *Cancer Epidemiol Biomarkers Prev* 11:1513–1530
- Zhou BB, Elledge SJ (2000) The DNA damage response: putting checkpoints in perspective. *Nature* 408:433–439
- Monson RR, MacMahon B (1984) Prenatal X-ray exposure and cancer in children. In: Boice JD, Fraumeni JF Jr (eds) *Radiation carcinogenesis: epidemiology and biological significance*. Raven Press, New York, pp 97–105
- Wakeford R (2008) Childhood leukaemia following medical diagnostic exposure to ionizing radiation in utero or after birth. *Radiat Prot Dosim* 132:166–174
- Bartley K, Metayer C, Selvin S, Ducore J, Buffler P (2010) Diagnostic X-rays and risk of childhood leukaemia. *Int J Epidemiol* 39:1628–1637
- Ma X, Buffler PA, Layefsky M, Does MB, Reynolds P (2004) Control selection strategies in case-control studies of childhood diseases. *Am J Epidemiol* 159:915–921
- Hansen HM, Wiemels JL, Wrensch M, Wiencke JK (2007) DNA quantification of whole genome amplified samples for genotyping on a multiplexed bead array platform. *Cancer Epidemiol Biomarkers Prev* 16:1686–1690
- Paynter RA, Skibola DR, Skibola CF, Buffler PA, Wiemels JL, Smith MT (2006) Accuracy of multiplexed Illumina platform-based single-nucleotide polymorphism genotyping compared between genomic and whole genome amplified DNA collected from multiple sources. *Cancer Epidemiol Biomarkers Prev* 15:2533–2536
- Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21:263–265
- Gabriel SB, Schaffner SF, Nguyen H et al (2002) The structure of haplotype blocks in the human genome. *Science* 296:2225–2229
- Collins-Schramm HE, Chima B, Morii T et al (2004) Mexican American ancestry-informative markers: examination of population structure and marker characteristics in European Americans, Mexican Americans, Amerindians and Asians. *Hum Genet* 114:263–271
- Aldrich MC, Zhang L, Wiemels JL et al (2006) Cytogenetics of Hispanic and White children with acute lymphoblastic leukemia in California. *Cancer Epidemiol Biomarkers Prev* 15:578–581
- Bartley K, Metayer C, Selvin S, Ducore J, Buffler P (2010) Diagnostic X-rays and risk of childhood leukaemia. *Int J Epidemiol* 39:1628–1637
- Aldrich MC, Selvin S, Hansen HM et al (2009) CYP1A1/2 haplotypes and lung cancer and assessment of confounding by population stratification. *Cancer Res* 69:2340–2348
- Wacholder S, Rothman N, Caporaso N (2000) Population stratification in epidemiologic studies of common genetic variants and cancer: quantification of bias. *J Natl Cancer Inst* 92:1151–1158
- Schaid DJ, Rowland CM, Tines DE, Jacobson RM, Poland GA (2002) Score tests for association between traits and haplotypes when linkage phase is ambiguous. *Am J Hum Genet* 70:425–434
- Mathias RA, Gao P, Goldstein JL et al (2006) A graphical assessment of p-values from sliding window haplotype tests of association to identify asthma susceptibility loci on chromosome 11q. *BMC Genet* 7:38
- Zaykin DV, Westfall PH, Young SS, Karnoub MA, Wagner MJ, Ehm MG (2002) Testing association of statistically inferred haplotypes with discrete and continuous traits in samples of unrelated individuals. *Hum Hered* 53:79–91
- Al-Attar A, Gossage L, Fareed KR et al (2010) Human apurinic/apyrimidinic endonuclease (APE1) is a prognostic factor in ovarian, gastro-oesophageal and pancreatico-biliary cancers. *Br J Cancer* 102:704–709
- Yang J, Yang D, Cogdell D et al (2010) APEX1 gene amplification and its protein overexpression in osteosarcoma: correlation

- with recurrence, metastasis, and survival. *Technol Cancer Res Treat* 9:161–169
33. Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: a practical and powerful approach to multiple testing. *J Royal Stat Soc Ser B* 57:289–300
 34. Benhamou S, Sarasin A (2002) ERCC2/XPD gene polymorphisms and cancer risk. *Mutagenesis* 17:463–469
 35. Sharan SK, Morimatsu M, Albrecht U et al (1997) Embryonic lethality and radiation hypersensitivity mediated by Rad51 in mice lacking Brca2. *Nature* 386:804–810
 36. Mahaney BL, Meek K, Lees-Miller SP (2009) Repair of ionizing radiation-induced DNA double-strand breaks by non-homologous end-joining. *Biochem J* 417:639–650
 37. van Gent DC, van der Burg M (2007) Non-homologous end-joining, a sticky affair. *Oncogene* 26:7731–7740
 38. Tauchi H, Matsuura S, Kobayashi J, Sakamoto S, Komatsu K (2002) Nijmegen breakage syndrome gene, NBS1, and molecular links to factors for genome stability. *Oncogene* 21:8967–8980
 39. Sherborne AL, Hosking FJ, Prasad RB et al (2010) Variation in CDKN2A at 9p21.3 influences childhood acute lymphoblastic leukemia risk. *Nat Genet* 42:492–494
 40. Healy J, Belanger H, Beaulieu P, Lariviere M, Labuda D, Sinnett D (2007) Promoter SNPs in G1/S checkpoint regulators and their impact on the susceptibility to childhood leukemia. *Blood* 109:683–692
 41. Wang SL, Zhao H, Zhou B et al (2006) Polymorphisms in ERCC1 and susceptibility to childhood acute lymphoblastic leukemia in a Chinese population. *Leukemia Res* 30:1341–1345
 42. Pakakasama S, Sirirat T, Kanchanachumpol S et al (2007) Genetic polymorphisms and haplotypes of DNA repair genes in childhood acute lymphoblastic leukemia. *Pediatr Blood Cancer* 48:16–20
 43. Joseph T, Kusumakumary P, Chacko P, Abraham A, Pillai MR (2005) DNA repair gene XRCC1 polymorphisms in childhood acute lymphoblastic leukemia. *Cancer Lett* 217:17–24
 44. Weiss NS (2008) Subgroup-specific associations in the face of overall null results: should we rush in or fear to tread? *Cancer Epidemiol Biomarkers Prev* 17:1297–1299
 45. Infante-Rivard C (2003) Diagnostic x rays, DNA repair genes and childhood acute lymphoblastic leukemia. *Health Phys* 85:60–64
 46. Campleman SL, Wright WE. (2004) Childhood cancer in California 1988 to 1999 vol I: birth to age 14. California Department of Health Services, Cancer Surveillance Section, Sacramento, CA, pp 16–17
 47. Papaemmanuil E, Hosking FJ, Vijayakrishnan J et al (2009) Loci on 7p12.2, 10q21.2 and 14q11.2 are associated with risk of childhood acute lymphoblastic leukemia. *Nat Genet* 41:1006–1010
 48. Trevino LR, Yang W, French D et al (2009) Germline genomic variants associated with childhood acute lymphoblastic leukemia. *Nat Genet* 41:1001–1005