Inherited coding variants at the CDKN2A locus influence susceptibility to acute lymphoblastic leukaemia in children

Elaine Mardis  
*Washington University School of Medicine in St. Louis*

Robert Fulton  
*Washington University School of Medicine in St. Louis*

et al.

Follow this and additional works at: https://digitalcommons.wustl.edu/open_access_pubs

Please let us know how this document benefits you.

**Recommended Citation**

https://digitalcommons.wustl.edu/open_access_pubs/4076

This Open Access Publication is brought to you for free and open access by Digital Commons@Becker. It has been accepted for inclusion in Open Access Publications by an authorized administrator of Digital Commons@Becker. For more information, please contact vanam@wustl.edu.
Inherited coding variants at the CDKN2A locus influence susceptibility to acute lymphoblastic leukaemia in children

Heng Xu1,2,*, Hui Zhang1,3,*, Wenjian Yang1, Rachita Yadav4, Alanna C. Morrison5, Maoxiang Qian1, Meenakshi Devidas6, Yu Liu7, Virginia Perez-Andreu1, Xujie Zhao1, Julie M. Gastier-Foster8, Philip J. Lupo9, Geoff Neale10, Elizabeth Raetz11, Eric Larsen12, W. Paul Bowman13, William L. Carroll14, Naomi Winick15, Richard Williams16, Torben Hansen17, Jens-Christian Holm18, Elaine Mardis19, Robert Fulton19, Ching-Hon Pu20,21, Jinghui Zhang7, Charles G. Mullighan20,22, William E. Evans1,20, Stephen P. Hunger23, Ramneeck Gupta4, Kjeld Schmiegelow24, Mignon L. Loh25, Mary V. Relling1,20 & Jun J. Yang1,20

There is increasing evidence from genome-wide association studies for a strong inherited genetic basis of susceptibility to acute lymphoblastic leukaemia (ALL) in children, yet the effects of protein-coding variants on ALL risk have not been systematically evaluated. Here we show a missense variant in CDKN2A associated with the development of ALL at genome-wide significance (rs3731249, \( P = 9.4 \times 10^{-23} \), odds ratio = 2.23). Functional studies indicate that this hypomorphic variant results in reduced tumour suppressor function of p16INK4A, increases the susceptibility to leukaemic transformation of haematopoietic progenitor cells, and is preferentially retained in ALL tumour cells. Resequencing the CDKN2A–CDKN2B locus in 2,407 childhood ALL cases reveals 19 additional putative functional germline variants. These results provide direct functional evidence for the influence of inherited genetic variation on ALL risk, highlighting the important and complex roles of CDKN2A–CDKN2B tumour suppressors in leukaemogenesis.
The risk of developing acute lymphoblastic leukaemia (ALL) is highest between 2 and 5 years after birth\(^1\), with initiating sentinel somatic genomic lesions (for example, chromosomal translocations) detectable at the time of birth in many cases\(^2\). This early disease onset suggests a strong inherited genetic basis for ALL susceptibility, and recent genome-wide association studies (GWAS) have discovered at least six risk loci: \(\text{ARID5B, IKZF1, CEBPE, PIP4K2A-BMI1, GATA3 and CDKN2A-CDKN2B}\)\(^3\)\(^-\)\(^10\). These ALL risk genes are directly involved in haematopoietic stem cell function, lymphocyte differentiation and development, and cell cycle regulation\(^11\)\(^-\)\(^13\), several of which are also commonly targeted by somatic genomic lesions. In particular, the \(\text{CDKN2A-CDKN2B}\) locus is one of the most frequently deleted genomic regions in childhood ALL with focal copy number loss in both B- and T-cell ALL\(^14\)\(^,\)\(^16\).

The vast majority of variants examined in previous ALL GWAS are intronic or intergenic. Although it is now evident that non-coding variants related to disease traits are significantly over-represented in regulatory DNA and often function as modulators of local or distal gene transcription\(^17\)\(^,\)\(^18\), questions also arise regarding the functional consequences of the genome-wide significant variant in the \(\text{CDKN2A}\) gene, and whether coding variants within ALL susceptibility genes might confer even greater effects on disease development. Moreover, a large number of low-frequency and rare-coding germline variants have been discovered by exome-sequencing efforts\(^19\), but their contributions to ALL pathogenesis have yet to be examined systematically.

In the present study, we perform an exome-focused GWAS to systematically examine the impact of germline-coding variants on the development of ALL in children of European descent, experimentally explore the functional consequences of the genome-wide significant variant in the \(\text{CDKN2A}\) gene, and comprehensively characterize coding variation at this locus by targeted resequencing.

**Results**

**Exome-focused GWAS of ALL susceptibility.** In the discovery GWAS, we genotyped 1,773 children with B-ALL and 10,448 non-ALL controls of European descent\(^20\)\(^,\)\(^21\) for 247,505 variants using the Illumina Infinium HumanExome array. Three loci with genome-wide significant association signals were observed: \(\text{ARID5B (10q21.2), IKZF1 (7q12.2) and CDKN2A (9p21.3)}\) (Fig. 1). Non-coding variants rs10821936 in \(\text{ARID5B}\) and rs4132601 in \(\text{IKZF1}\) showed the strongest association \((P = 9.9 \times 10^{-46})\) and \(4.3 \times 10^{-37}\), the logistic regression test, respectively; Fig. 1 and Supplementary Table 1), confirming and associating with ALL susceptibility. The third genome-wide significant hit was a missense SNP at the \(\text{CDKN2A}\) locus \((\text{rs3731249, } P = 9.4 \times 10^{-23})\), the logistic regression test, Fig. 1, Table 1). The T allele at rs3731249 was over-represented in ALL compared with controls \((6.8\%\) versus \(3.0\%,\) Table 1), with every copy of the allele conferring a 2.23-fold increase in disease risk \((95\%\) confidence interval \(1.90-2.61)\). The C-to-T nucleotide substitution at rs3731249 \((\text{c.C442T})\) resulted in an alanine-to-threonine change in amino-acid sequence \((\text{p.A148T}\)) for tumour suppressor \(\text{p16INK4A}\)\(^2\). This variant also locates in the 3′ untranslated region \((3′\text{UTR})\) of the \(\text{p14\text{ARF}}\) transcript, an alternative open reading frame at this locus encoding a different tumour suppressor. Interestingly, previous GWAS had identified an intronic variant in \(\text{CDKN2A}\) (rs3731217) to be strongly associated with susceptibility to ALL in populations of European descent\(^2\). Genotype correlation between the coding variant rs3731249 and the intronic rs3731217 is exceedingly low \((P \leq 0.01\) in Europeans, Supplementary Fig. 1), and multivariate analyses including both SNPs indicated their independent contribution to ALL risk (Supplementary Table 2). In the replication cohort of 409 childhood ALL cases and 1,599 non-ALL controls of European descent in Denmark, the association signal at rs3731249 was validated \((P = 5.2 \times 10^{-4},\) odds ratio \(= 1.73\) \((1.27-2.36)\), the logistic regression test, Table 1) and this variant also remained significant after adjusting for rs3731217.

**Functional characterization of the rs3731249 variant.** To experimentally evaluate the effects of rs3731249 on ALL leukemogenesis, we directly compared the effect of wildtype versus variant allele \(\text{p16INK4A}\) \((\text{p.148A} vs \text{p.148T})\) on \(\text{BCR-ABL1}\)-mediated leukaemic transformation \textit{in vitro}. We chose mouse haematopoietic progenitor Ba/F3 cell line because it is inherently \(\text{p16INK4A}\)-defective due to methylation at the \(\text{Ink4a-Arf}\) locus\(^22\), and ectopic expression of \(\text{BCR-ABL1}\) in Ba/F3 cells efficiently induces exogenous cytokine \((\text{interleukin 3 (IL3)})\)-independent proliferation. Over-expression of wild-type \(\text{p16INK4A}\) \((\text{p.148A})\) significantly inhibited leukaemic transformation by \(\text{BCR-ABL1}\) \((\text{Fig. 2a, Supplementary Fig. 2})\), consistent with its role as a critical tumour suppressor in ALL. In contrast, Ba/F3 cells overexpressing variant \(\text{p16INK4A}\) \((\text{p.148T})\) were significantly more susceptible to \(\text{BCR-ABL1}\) transformation measured by IL3-independent growth, suggesting that the p.148T variant is likely hypomorphistic with reduced tumour suppressor function. In Ba/F3 cells transfected with both variant and wild-type \(\text{p16INK4A}\), the relative ratio of the p.148T (variant) to p.148A

![Figure 1](image-url)
CDKN2A gene (p16INK4A and p14ARF) in children with ALL has a higher burden of rare missense variants relative to controls. The NHLBI GO Exome Sequencing Project (ESP) showed a trend for damaging based on combined annotation dependent depletion (CADD score) in two and five cases, respectively. Five variants were predicted to be damaging based on combined annotation dependent depletion (CADD score), two of which result in amino-acid sequence changes (Fig. 3). These missense variants were all singletons, except for the p.D125H variant in p16INK4A and the p.A121T variant in p14ARF observed in two and five cases, respectively. Five variants were predicted to be damaging based on combined annotation dependent depletion (CADD score) in two and five cases, respectively. Five variants were predicted to be damaging based on combined annotation dependent depletion (CADD score), two of which result in amino-acid sequence changes (Fig. 3).

Table 1 | Genome-wide significant association and replication of novel coding B-ALL susceptibility variant at CDKN2A locus.

<table>
<thead>
<tr>
<th>SNP ID</th>
<th>Chr</th>
<th>Position</th>
<th>Gene</th>
<th>Alleles</th>
<th>Genotype</th>
<th>GWAS series</th>
<th>Replication series</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs3731249</td>
<td>9</td>
<td>21970916</td>
<td>CDKN2A</td>
<td>C/T</td>
<td>Case (N = 4,300)</td>
<td>Control (N = 1,450)</td>
<td>P-value</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8 (0.45%)</td>
<td>7 (0.07%)</td>
<td>9.4 × 10⁻²⁴</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>224 (12.63%)</td>
<td>619 (5.93%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1,541 (86.92%)</td>
<td>9,815 (94%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Case (N = 410)</td>
<td>Control (N = 1,599)</td>
<td>P-value</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7 (1.71%)</td>
<td>2 (0.13%)</td>
<td>5.22 × 10⁻⁴</td>
</tr>
</tbody>
</table>

Chr, chromosome; CI, confidence interval; GWAS, genome-wide association studies; OR, odds ratio; SNP, single nucleotide polymorphism.

*Chromosomal localisations are based on hg19.

†Bold denotes the allele that had a significantly higher frequency in children with B-ALL than in the non-ALL controls (that is, risk allele for B-ALL).

‡OR represents the increase in the risk of developing ALL for each copy of the risk allele compared with subjects who do not carry the risk allele; P-values and ORs were estimated by the logistic regression test.

Targeted resequencing of CDKN2A and CDKN2B in childhood ALL. To comprehensively identify putative functional ALL susceptibility variants at this locus, we resequenced the coding region of the CDKN2A and CDKN2B genes in germline DNA from 2,407 childhood ALL cases (1,450 of which were also included in the discovery GWAS). In addition to rs3731249, we observed another 13 germline exonic variants in tumour suppressors p16INK4A and p14ARF encoded by the CDKN2A gene, 12 of which result in amino-acid sequence changes (Fig. 3, Supplementary Table 3). These missense variants were all singletons, except for the p.D125H variant in p16INK4A and the p.A121T variant in p14ARF observed in two and five cases, respectively. Five variants were predicted to be damaging based on combined annotation dependent depletion (CADD score) (Supplementary Table 3), and we did not observe germline insertions or deletions in CDKN2A in our ALL cohort. Comparing with 4,300 European American individuals from the NHLBI GO Exome Sequencing Project (ESP), there was a trend for a higher burden of rare missense variants in relative to controls the CDKN2A gene (p16INK4A and p14ARF) in children with ALL (0.71% versus 0.23%, P = 0.0045; Fisher’s exact test, Fig. 3). In addition, we identified six germline-coding variants in the adjacent CDKN2B gene in this cohort of children with ALL, although there was no significant over-representation compared with European controls in the ESP cohort (0.83% versus 0.79%, Fig. 3).

Discussion

Encoding three tumour suppressor proteins (p16INK4A, p14ARF and p15INK4B), the CDKN2A–CDKN2B locus at 9p21 is promiscuously associated with tumorigenesis and commonly

Figure 2 | Functional characterization of ALL risk variant at rs3731249.

(a) Mouse haematopoietic progenitor cell Ba/f3 overexpressing wildtype, variant p16INK4A, or transfected with control vector was transduced with leukaemia oncogenic BCR–ABL1 fusion gene. Cell proliferation in the absence of cytokine IL3 was measured daily as an indicator of leukaemogenesis.

(b) Allele-specific expression of p16INK4A in ALL.

© 2015 Macmillan Publishers Limited. All rights reserved.
targeted by somatic mutation, deletion and/or hypermethylation in various cancers. p16INK4A and p15INK4B are highly homologous inhibitors of cyclin-dependent kinase and function mainly as master regulators of cell cycle entry via the Rb-E2F signalling axis. Although also encoded by the CDKN2A gene, p14ARF utilizes a completely different reading frame with distinct tumour suppression functions by inhibiting MDM2 and activating p53. Retained in the nucleus compared with the wild-type p16INK4A, the p.A148T variant is unclear, since this residue is not directly involved in binding to CDK4 or CDK6. The relative contribution of p16 INK4A versus p14ARF to ALL pathogenesis is not unequivocal because somatic deletions at this locus almost always lead to the loss of both genes. In fact, the CDKN2A–CDKN2B locus is either bi- or monoallelicly deleted in 64% of BCR–ABL1-positive ALL cases and in 32–72% of T- or B-ALL cases without the BCR-ABL1 translocation, suggesting positive selection for oncogene-stressed cells. The previously reported ALL susceptibility variant rs3731217 is located in a non-coding region downstream of exon 1β (specific for p14ARF), but distal to exon 1z (specific for p16INK4A) of the CDKN2A gene. The germline genotype at this SNP was not definitively determined to be rs3731249 as a possible functional variant directly contributing to the association with ALL risk. The structural basis of the hypomorphic effects of the p.A148T variant is unclear, since this residue is not directly involved in binding to CDK4 or CDK6. The relative contribution of p16INK4A versus p14ARF to ALL pathogenesis is not unequivocal because somatic deletions at this locus almost always lead to the loss of both genes. Although the rs3731249 variant also results in sequence changes of the 3’-UTR of the p14ARF transcript, bioinformatic prediction did not identify any potential effects on mRNA stability or microRNA binding and no difference was observed in reporter gene transcription under the influence of 3’-UTR containing either the wildtype or variant allele at rs3731249 (Fig. 3). The germline genotype at this SNP was not definitively determined to be rs3731249 in our current GWAS localizes to exon 2 of the CDKN2A. While this exon is shared by both p16INK4A and p14ARF, the C-to-T nucleotide transition causes a missense change for the p16INK4A open reading frame but is in the UTR of the p14ARF, therefore, likely to have a more direct effect on the former. This hypothesis is supported by the fact that haematopoietic progenitor cells (Ba/f3) expressing variant p16INK4A were substantially more susceptible to BCR–ABL1-mediated leukaemic transformation compared with cells with the wild-type protein (Fig. 2a), pointing to rs3731249 as a possible functional variant directly contributing to the association with ALL risk. In this case, the functional consequence of the structural variant rs3731249 is not obvious. The high frequency of the risk allele in African American children with ALL compared with individuals from this racial background in European populations, for example, there was a trend for a higher frequency of the risk allele in African American children with ALL than that in individuals from this racial background in the NHLBI ESP cohort (0.58% in 260 ALL cases versus 0.38% in 2,203 controls), although a much larger sample size is needed to rigorously examine the statistical significance of such differences. It should be noted that we and others previously showed that the
non-coding ALL risk variants (rs17756311 and rs3731217) at this locus had much stronger effects in European Americans than in other race/ethnic groups (Supplementary Table 4), suggesting potential racial differences in genetic susceptibility to ALL.

We subsequently identified additional coding variants in p16INK4A, p14ARF and p15INK4B by resequencing, most of which were low frequency or rare. While there was a modest over-representation of potentially damaging coding variants in ALL cases compared with controls (Fig. 3), our data do not suggest that rare variants contribute substantially to the associations with ALL susceptibility observed at this locus. It should also be noted that the vast majority of coding variants within the CDKN2NA gene affects only one of the two tumour suppressors (either p16INK4A or p14ARF). Interestingly, rs199888003 is the only variant that is located in the coding region of both p16INK4A and p14ARF, resulting in an alanine-to-threonine change in p14ARF (p.A121T) with synonymous effect on p16INK4A. This is also the most frequent germline missense variant in p14ARF in our cohort and was over-represented in ALL compared with non-ALL controls (0.21% versus 0.046%, respectively, Fig. 3). This substitution of threonine in p14ARF adds a possible glycosylation and phosphorylation site and also introduces a phosphoprotein-binding FHA domain implicated in DNA damage response and cell cycling (31). Future studies are warranted to determine the exact consequences of this variant on p14ARF functions. To systemically evaluate the contribution of low frequency and rare-coding variants to ALL risk, we also performed genome-wide gene-level burden test but did not observe any genome-wide significant associations (Supplementary Table 4). Of the six known ALL risk loci, we noted two coding variants in CEBPE (rs141903485 and rs146580935, Supplementary Table 5) nominally associated with ALL susceptibility.

In conclusion, we comprehensively evaluated exonic genetic variations for association with ALL susceptibility and identified novel coding risk variants at the CDKN2A–CDKN2B locus that may directly affect tumour suppressor functions and potentiate leukemic transformation. These results provided functional evidence for the inference of inherited genetic variants on ALL leukemogenesis, further indicating that a continuum of genetic variations in both host and tumour genomes contribute to malignant transformation and cancer risk.

Methods

Subjects and samples. The discovery GWAS consisted of 1,773 childhood B-ALL cases and 10,448 non-ALL controls of European descent (>90% European genetic ancestry as estimated using STRUCTURE(32,33)). ALL cases were from the Children’s Oncology Group (COG) AALL0323 study (N = 1,277) (34), the COG 9906 protocol (N = 115) (35) and St Jude Total Therapy XIIIIB and XV protocols (N = 381) (36). Unrelated individuals of European descent from the Atherosclerosis Risk in Communities (ARIC) study (20,21) were used as non-ALL controls because the prevalence of adult survivors of childhood ALL is less than 1 in 10,000 in the US. The replication series included 409 children with ALL from NOPHO ALL92, ALL2000 and ALL2008 protocols (37) and 1,599 unrelated non-ALL controls from Danish population-based cohorts. Ba/F3 cells with BCR/ABL1 and Ba/F3 cells with BCR/ABL1 and C2H2–Luc2 construct were a gift from Dr Charles Sherr (The Sloan Kettering Cancer Center) were maintained in medium supplemented with 10 ng/ml 1-recombinant mouse IL3. Ba/F3 cells were transfected with lentiviral supernatants containing cl20c–p16INK4A p14A–IRES–GFP or cl20c–p16INK4A p14A–IRES–GFP produced by transient transfection of 293 T cells (American Type Culture Collection) using calcium phosphate. The Mscv (Muscovite MCS)–Bcr/abl–Luc2 construct was a gift from Dr Charles Sherr at St Jude Children’s Research Hospital(38) and retroviral particles were produced using the retroviral packaging and transduction system (39). The (Abcam, ab12187) with 1:100 anti-Tubulin antibody, Sigma-Aldrich, T5618).

Leukemic transformation assay in Ba/F3 cells. The full-length CDKN2A was purchased from GE Healthcare. The p.A14ARF variant (rs3731249) was identified using site-directed mutagenesis (forward primer: 5'-TGGCCAAGATAGTGAGCAGGAGGATGCCTACTGAC-3', reverse primer: 5'-TCTGAGGACCTTCCGTGGCAT CTATGGGGCGCA-3') and cloned into the cl20c–IRES–GFP lentiviral vector, and lentiviral supernatants containing cl20c–p16INK4A p14A–IRES–GFP or cl20c–p16INK4A p14A–T–IRES–GFP were produced by transient transfection of 293 T cells (American Type Culture Collection) using calcium phosphate. The Mscv (Muscovite MCS)–Bcr/abl–Luc2 construct was a gift from Dr Charles Sherr at St Jude Children’s Research Hospital(38) and retroviral particles were produced using the retroviral packaging and transduction system (39). The (Abcam, ab12187) with 1:100 anti-Tubulin antibody, Sigma-Aldrich, T5618).

Genotyping and quality control. SNP genotyping was performed in germline DNA using the Illumina Infinium HumanExome Array v1.0 in the discovery GWAS, and using Illumina HumanCoreExome chip for the replication series. Genotype calls (coded as 0, 1, and 2 for AA, AB and BB genotypes) were determined using the Illumina GenomewideStudio Software. For the ALL cases, samples for which genotyping was not obtained at <98% SNPs on the array were excluded and were not included from the analyses. Quality control procedures were performed for both samples and SNPs on the basis of call rate, minor allele frequency (MAF), and Hardy Weinberg equilibrium (Supplementary Fig. 3). Detailed quality control procedures for the non-ALL controls from the ARIC study was performed at the University of Texas Health Science Center following established protocols (40).

We performed principal component analysis of cases and controls in the discovery GWAS to characterize population substructure (Supplementary Fig. 6).

Genome-wide analyses. The discovery GWAS, the association of each SNP individually with ALL susceptibility was tested by comparing the genotype frequency between ALL cases and non-ALL controls in logistic regression models, after adjusting for top 10 principal components to control for population stratification. A quantitative–quantitative (Q–Q) plot was constructed and the Manhattan plot was only minimal inflation at the upper tail of the distribution (λ = 1.08, Supplementary Fig. 7). In the replication studies, we evaluated the novel genome-wide significant variant rs3731249, using the same logistic regression models. Multivariate logistic regression model including both rs3731217 and rs3731249 were tested to determine independent association signals at the CDKN2A locus in both discovery and replication series.

We also performed gene-level analyses to evaluate the aggregated effects of low-frequency variants on ALL susceptibility, using the SKAT test (41). Missense, stop codon-altering and splice-site variants with MAF < 5% were included. In total, 12,678 genes with at least two variants were tested. R (version 3.0) statistical software was used for all analyses unless indicated otherwise.

CDKN2A–CDKN2B resequencing and rare variant analyses. Germline DNA from 2,407 children with ALL was used to create individual Illumina dual-indexed libraries. These libraries were pooled in sets of 96 and hybridized with a custom version of the Roche NimbleGen SeqCap EZ custom probes to capture the CDKN2A–CDKN2B region on 9p21. Quantitative PCR was used to determine the appropriate capture product titre necessary to efficiently populate an Illumina HiSeq 2000 flowcell for paired-end 2 x 101 bp sequencing. Each sequence pool of 96 samples was demultiplexed, with coverages of >20 x depth across >90% of the targeted regions for nearly all samples. Sequence reads in FASTQ format were formatted and aligned using the BWA and–Samtools pipeline (42). Genotypes were called using the GATK pipeline version 3.1 (ref. 38). We compared the proportion of rare variant-carriers in ALL subjects (either homozygous or heterozygous) versus that in individuals of European descent in the ESP cohort (non-ALL controls), focusing on variants with MAF < 1%. Statistical significance of the difference was estimated using Fisher’s exact test.

CDKN2A sequencing was also performed in matched germline and diagnostic ALL tumour DNA by Complete Genomics for all cases with available data, and in tumour RNA by RNA-seq. Details regarding sequencing, data analysis and coverage are available at ftp://calcipd.ncbi.nih.gov/pub/dcc_target/ALL/Phase_II/sequence/WGS/CGI_TARGET_Pipeline_README.pdf, or as previously described (39) (European Genome Phenome archive: EGAS00001000654).

NATURE COMMUNICATIONS | DOI: 10.1038/ncomms8553 | www.nature.com/naturecommunications

© 2015 Macmillan Publishers Limited. All rights reserved.
For quantitative reverse transcription-PCR (qRT-PCR), total RNA was extracted using the RNeasy Micro kit (Qiagen) according to the manufacturer's protocol. Total RNA (500 ng) was reverse transcribed into cDNA using oligoT primers and the SuperScript III reverse transcriptase kit (Invitrogen). Quantitative real-time PCR was performed by using ABI Prism 7900HT detection system (Applied Biosystems) with Faststart SYBR Green master mix (Roche). Relative expression was calculated as a ratio of BCR-ABL to Hprt. Primer sequences of BCR-ABL1 and Hprt were as follows: BCR-ABL1 (forward: 5'-CTGCGCAACGATTGGCAAC-3'; reverse: 5'-CAGCATCATGCAGCTTCTTC-3'); Hprt (forward: 5'-GAGCAAGATTGTGGTATGTT-3'; reverse: 5'-TTCTTCCCTGGTGATGGAAAT-3').

To co-express rs3731249 variant and wild-type p16INK4A. Ba/F3 cells were transfected with equal molar cl20c-p16ink4a-1.8RA–IRES–GFP and cl20c–p16ink4a-1.8RA–IRES–IPF lentivirus and cells successfully transfected with both were selected by flow cytometry sorting for GFP/YPF double positivity. BCR-ABL1-mediated transformation was performed as described above. Genomic DNA and RNA samples were collected at day 0, 2, 4, and 5 after IL3 removal. p.148A and P.148T cDNA transcript in RNA was quantified using allele-specific Taqman genotyping assay and normalized to allele ratio in matched DNA samples at respective time points. Each experiment was performed three times and each sample was assayed in triplicate.

Luciferase reporter assays. The p14INK4A 3'-UTR vector (3'-UTR for Human NM_058195.2) was placed downstream of luciferase reporter gene on the pEZX-p16INK4Ap.148 T–IRES–iYFP lentivirus and cells successfully transfected with both were selected by flow cytometry sorting for GFP/YPF double positivity. The p16INK4A-3'UTR vectors were constructed using Lipofectamine 2000 (Invitrogen). Firefly luciferase activities were measured 24 h after transfecting the Bcr-Abl luciferase Assay (Promega). The results were normalized against Renilla luciferase. Each reporter construct transfection was replicated at least three times, and each sample was assayed in triplicate.

References
The authors also thank the staff and participants of the ARIC study for their important contributions.

**Author contributions**


**Additional information**

**Accession codes.** The RNA-seq data have been deposited in European Genome Phenome archive under the accession codes EGAS00001000654.

**Supplementary Information** accompanies this paper at http://www.nature.com/naturecommunications

**Competing financial interests:** The authors declare no competing financial interests.

**Reprints and permission** information is available online at http://npg.nature.com/reprintsandpermissions/

**How to cite this article:** Xu, H. *et al.* Inherited coding variants at the CDKN2A locus influence susceptibility to acute lymphoblastic leukaemia in children. Nat. Commun. 6:7553 doi: 10.1038/ncomms8553 (2015).

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/