Current evidence for an inherited genetic basis of childhood acute lymphoblastic leukemia

Kevin Y. Urayama · Anand P. Chokkalingam ·
Atsushi Manabe · Shuki Mizutani

Abstract Acute lymphoblastic leukemia (ALL) is the most common cancer in children, and efforts to understand its etiology has followed a paradigm that common genetic variation in the presence of modifiable environmental factors contribute to disease risk. To date, there are numerous reports of candidate gene association studies suggesting an involvement of genetic loci in childhood ALL risk, but the general lack of consistency in results has underscored the need for careful interpretation and confirmation in additional well-designed studies. Complementary efforts using the genome-wide association study approach have shown indisputable evidence that common low penetrance genetic polymorphisms contribute to childhood ALL risk. However, current calculations show that these established disease loci only explain a portion of the total estimated contribution of common genetic variation on childhood ALL risk. Certain candidate gene loci previously examined likely contribute to this unexplained variation in risk, but the challenge moving forward will be to establish which ones based on the accumulating evidence. In this review, we describe the results of the most recent gene association studies in childhood ALL and discuss options for future efforts to advance this area of research.

Keywords Childhood leukemia · Genetic susceptibility · Genome-wide association study · Predisposition · Single nucleotide polymorphism

K. Y. Urayama (✉)
Center for Clinical Epidemiology,
St. Luke's Life Science Institute,
10-1 Akashi-cho, Chuo-ku, Tokyo 104-0044, Japan
e-mail: kurayama@luke.or.jp

A. P. Chokkalingam
School of Public Health, University of California,
Berkeley, CA, USA

A. Manabe
Department of Pediatrics,
St. Luke's International Hospital, Tokyo, Japan

S. Mizutani
Department of Pediatrics and Developmental Biology,
Graduate School of Medicine,
Tokyo Medical and Dental University, Tokyo, Japan

Introduction

Leukemia, the most common cancer in children, is characterized by dysregulated proliferation of clonally expanded immature lymphoid or myeloid progenitor cells that have encountered a series of catastrophic alternations within key regulatory genes [1]. Within the major subtypes of childhood leukemia defined by cell lineage, acute lymphoblastic leukemia (ALL) is most common comprising nearly 80 % of diagnoses in developed countries [2]. In most childhood leukemia cases, characteristic genetic alterations are observed, including numerical and structural chromosomal changes such as hyperdiploidy (>46 chromosomes) or translocations, as well as the more subtle changes in the form of point mutations and gene deletions [3]. Translocations are considered a hallmark genetic event in leukemia, for example, the t(12;21) translocation (*TEL-AML1*) commonly observed in B-cell lineage ALL and the 11q23/*MLL* gene rearrangements in infant leukemia [4].

Confirmed clinical and epidemiologic associations including sex, age, race, exposure to ionizing radiation in utero, postnatal high-dose radiation, chemotherapeutic agents, and several genetic syndromes, explain only a small proportion of childhood ALL cases diagnosed [2]. The

early age of onset of childhood ALL suggests that inherited genetic traits may play a role, and recent evidence indicates that these contribute to a substantial proportion of the variation in childhood ALL risk [5, 6]. Such genetic traits can range from predisposing rare highly penetrant mutations to more common low penetrance genetic polymorphisms. While a considerable excess risk of childhood ALL has been observed among monozygotic twins compared to dizygotic twins of ALL patients suggesting heritability, recent studies indicate that this excess risk may be due more to intraplacental metastasis rather than highly penetrant risk alleles [7]. Thus, for childhood ALL, as for other multi-factorial diseases, inherited risk alleles are likely to be low penetrance susceptibility alleles that interact with environmental factors to modulate disease risk.

In this review, we focused on evaluating the current published epidemiological literature that has evaluated the influence of inherited (germline) genetic variation on childhood ALL risk. This comprises two types of studies broadly classified as either candidate gene association studies or genome-wide association studies (GWAS).

Studies utilizing the traditional candidate gene approach are initiated with specific a priori hypotheses based on known biological functions of the genes that have relevance to proposed disease pathology. This type of study more commonly focuses on specific genetic variants, such as single nucleotide polymorphisms (SNPs) and/or insertion/deletions that are known to correlate with expression and/or function of the resultant protein product. In addition, some candidate gene studies have adopted a haplotype tagging

approach built on the principle that segments of the genome are arranged into distinct haplotype blocks defined by the level of linkage disequilibrium (LD) exhibited between neighboring genetic markers [8]. Thus, a candidate region can be interrogated without necessarily having previously identified a functional genetic variant, but any positive results from the study would, at most, only localize a region of potential association, requiring replication and additional fine-mapping efforts to identify the causal variant.

The genome-wide approach is similarly based around this concept, but hypothesis testing is completely agnostic in that a large number of variants that represent the genetic diversity of the entire genome, regardless of candidate gene status, are assembled and evaluated in one study [9].

To date, a large number of studies have been published covering a broad range of genes. In this review, we focus our discussion on the most recent publications while referencing previous reviews and meta-analyses [5, 10–12]. Our objective is to provide the reader with an updated account of the current state of evidence regarding the genetic susceptibility to childhood ALL.

Candidate gene association studies

The numbers of candidate gene studies continue to increase. As reported in a previous review [5], a search of the literature published prior to April 2008 identified 59 articles examining the association between variants of 36 different candidate gene loci and childhood ALL risk (blue bars in Fig. 1). As of

Fig. 1 Plot summarizing the number of reports published for each gene arranged by candidate pathway. Genes having greater than three previous publications were reviewed in this article and are represented by *bars* exceeding the *gray horizontal line*. The numbers of reports counted by a previous review as of April 2008 [5] are represented in *blue*, and the numbers of reports published subsequently as of October 2012 are represented in *red*. The *asterisks* (*) are genes with variants that have been evaluated previously with meta-analysis. The "Other" (**) category refers to genes that have been reported by fewer than three previous publications
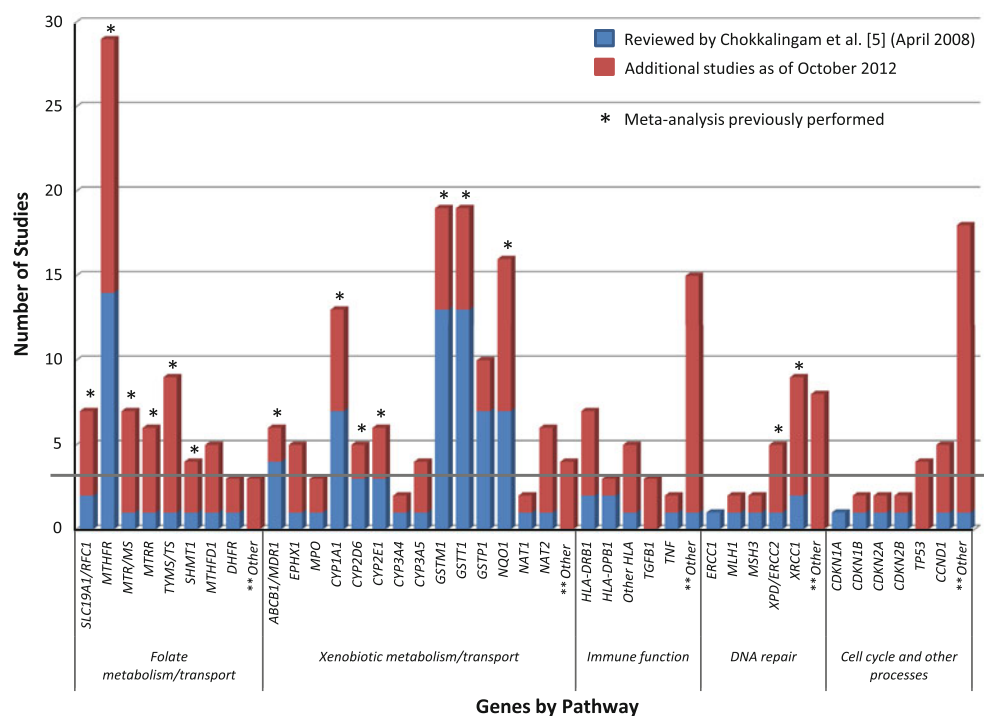
October 2012, a systematic search of the PubMed database for all ALL risk candidate gene study articles published since April 2008 identified an additional 76 articles reporting on studies conducted within populations from 24 countries. These studies examined about 450 additional previously unevaluated genes in childhood leukemia, but this seemingly dramatic increase is influenced by select studies [13–19] that have made use of high-throughput genotyping platforms based around a haplotype tagging SNP approach. With the exception of a few studies [20–23], all utilized the case–control study design to evaluate the associations. Consistency in results across multiple studies is essential to the process of establishing associations, particularly in observational studies of the type reviewed here. Thus, we focused our discussion around genes for which more than three studies have been published (Tables 1, 2; Fig. 1).

Folate metabolism

Folate and its bioactive metabolic substrates are essential to numerous bodily functions, particularly for their role in DNA methylation and synthesis that aid the rapid cell division and growth requirements associated with pregnancy and early infancy [24, 25]. Folate deficiency may contribute to carcinogenesis via hypomethylation of important regulatory genes as well as induction of DNA damage through uracil misincorporation during DNA replication [24]. Variants in more than a dozen genes, including those encoding the methylenetetrahydrofolate reductase (MTHFR), may alter folate metabolism and contribute to the risk of childhood leukemia.

Among candidate gene studies in childhood ALL, MTHFR is the most frequently studied gene and has largely focused on two functional SNPs associated with reduced enzymatic activities, C677T (rs1801133) and A1298C (rs1801131) (Table 1). Reports from three meta-analyses published during this past year showed consistent results, including no evidence of association for A1298C [26–28], but marginally significant reduced risks of childhood ALL for C677T (allele contrast, OR = 0.90, 95 % CI 0.82–1.00) [26, 28]. Using meta-regression, Wang et al. [26] found a potential gender effect associated with C677T and suggested this as a source of between-study heterogeneity in results. Zintzaras et al. [28], despite their marginally significant results particularly in white children, interpreted their meta-analysis with caution due to weak cumulative meta-analysis findings. This conclusion is in line with studies that were published in 2012 subsequent to these meta-analyses in which all five reports showed no association with either C677T or A1298C [20, 29–32]. Several reasons for this discrepancy are acknowledged, including a potential sex and race/ethnicity specific association for C677T, and important interactions with dietary folate

intake that may influence the associations. Current evidence indicates that the effects of the reduced activity of MTHFR conferred by C677T may have greater impact in situations of serum folate deficiency [33, 34]. However, this has not been confirmed in more recent studies of MTHFR-folate gene–environment interaction [29, 31, 35, 36] and continues to be in a relationship that is difficult to evaluate due to widespread dietary folate fortification and education efforts beginning in the mid-1990s.

Other key genes in the folate metabolic pathway examined in previous studies include the solute carrier family 19 member 1 (SLC19A1) gene responsible for membrane transport of folate into the cell, methionine synthase (MTR) and methionine synthase reductase (MTRR) genes involved in methionine biosynthesis that influence DNA methylation, and thymidylate synthetase (TYMS), serine hydroxymethyltransferase 1 (SHMT1), and methylenetetrahydrofolate dehydrogenase 1 (MTHFD1) that contribute to DNA synthesis and replication (Table 1). Of these genes, a recent meta-analysis [12] reported statistically significant ($P < 0.05$) ALL associations with functional SNPs of SLC19A1 (G80A/rs1051266, dominant: OR = 1.37, 95 % CI 1.11–1.69), MTRR (A66G/rs1801394, dominant: OR = 0.73, 95 % CI 0.59–0.91), and SHMT1 (C1420T/rs1051266, CT vs. CC: OR = 0.79, 95 % CI 0.65–0.98) showing minimal evidence of heterogeneity in effect across studies. Subsequent case–control studies of MTRR A66G and SHMT1 C1420T are partly supportive, but the four additional reports on SLC19A1 G80A are inconsistent with both increased [37, 38] and reduced [39] risk estimates reported. Meta-analyses for MTR A2756G (rs1805087) and the TYMS 28 bp repeat [12] including four studies each, do not support an association. However, a recent UK study [40] comprising a larger number of cases (939 cases and 824 controls) than the meta-analysis reported a strong association between MTR A2756G and ALL risk (GG vs. AA, OR = 1.88, 95 % CI 1.16–3.07), particularly pronounced for MLL-positive leukemia (OR = 4.90, 95 % CI 1.30–18.45). For TYMS, the UK study also found a statistically significant increased risk associated with homozygous carriers of the 1494del6 deletion (rs16430, OR = 1.46, 95 % CI 1.02–2.08), but this is not supported by finding from three other smaller studies [37, 39, 41]. Finally, the five studies reporting results for MTHFD1 G401A (rs1950902) [31, 41] and/or G1958A (rs2236225) [37, 39, 41] provide little evidence of an association with childhood ALL.

Xenobiotic metabolism and transport

In order to exert their effects, potentially harmful chemicals (xenobiotics) must gain entry into target cells via

Table 1 Evidence from candidate gene association studies for gene variants in the folate pathway that have been evaluated with meta-analysis or have greater than three previous publications

Gene, variant	Most recent meta-analysis	Other studies subsequent to meta-analysis[a]
MTHFR C677T (rs1801133)	Zintzaras et al. [28]: allele contrast, OR = 0.91 (0.82–1.00) (23 studies, *P* heterogeneity <0.01)	Metayer et al. [31]: additive, 0.96 (0.78–1.19), no GxE with maternal folate
	Wang et al. [26]: allele contrast, OR = 0.92 (0.87–0.98) (24 studies, *P* heterogeneity <0.01)	Lupo et al. [20]: additive, 0.86 (0.53–1.37)
	Yan et al. [27]: recessive, OR = 0.83 (0.72–0.95) (21 studies, *P* heterogeneity = 0.502)	Amigou [29]: dominant, 1.0 (0.8–1.4)
		Nikbakht et al. [32]: dominant, 0.88 (0.50–1.55)
		Azhar et al. [30]: no association
MTHFR A1298C (rs1801131)	Zintzaras et al. [28]: allele contrast, OR = 1.04 (0.93–1.16) (19 studies, *P* heterogeneity <0.01)	Metayer et al. [31]: additive, 1.14 (0.91–1.43), no GxE with maternal folate
	Wang et al. [26]: allele contrast, OR = 1.04 (0.93–1.16) (20 studies, *P* heterogeneity <0.01)	Lupo et al. [20]: additive, 0.95 (0.56–1.62)
	Yan et al. [27]: recessive, OR = 1.02 (0.89–1.17) (17 studies, *P* heterogeneity = 0.138)	Amigou et al. [29]: dominant, 1.0 (0.7–1.4)
		Nikbakht et al. [32]: dominant, 0.94 (0.53–1.66)
		Azhar et al. [30]: no association
MTR/MS A2756G (rs1805087)	Vijayakrishnan and Houlston [12]: dominant, OR = 1.12 (0.93–1.34) (4 studies, *P* heterogeneity = 0.09)	Lightfoot et al. [40]: dominant, 1.31 (1.07–1.60)
		Metayer et al. [31]: additive, 1.10 (0.86–1.40)
		Nikbakht et al. [32]: dominant, 0.95 (0.54–1.68)
		Rahimi et al. [113]: dominant, 1.05 (0.58–1.87)
		Lupo et al. [20]: SNP not evaluated/reported, but rs2275565 was associated (OR=1.68, 1.01–2.86)
MTRR A66G (rs1801394)	Vijayakrishnan and Houlston [12]: dominant, OR = 0.73 (0.59–0.91) (3 studies, *P* heterogeneity = 0.82)[b]	Metayer et al. [31]: additive, 0.93 (0.76–1.15), association with a different SNP rs2287779 noted
		Amigou et al. [29]: dominant, 0.7 (0.5–1.0)
		Lupo et al. [20]: additive, 0.99 (0.61–1.61)
SHMT1 C1420T (rs1979277)	Vijayakrishnan and Houlston [12]: dominant, OR = 0.86 (0.71–1.04); CT vs. CC, 0.79 (0.65–0.98) (2 studies, *P* heterogeneity = 0.37)[b]	Lightfoot et al. [40]: dominant, 0.89 (0.73–1.09); hyperdiploidy, dominant, 0.77 (0.58–1.01)
		Metayer et al. [31]: SNP not evaluated/reported; no association with other SNPs in region noted
SLC19A1/RFC1 G80A (rs1051266)	Vijayakrishnan and Houlston [12]: dominant, OR = 1.37 (1.11–1.69) (2 studies, *P* heterogeneity = 0.49)[b]	Yeoh et al. [39]: Malays-homozygous, 0.65 (0.39–1.10); Chinese-homozygous, 0.86 (0.55–3.33)
		Chan et al. [37]: heterozygous, 1.86 (1.14–3.05)
		Metayer et al. [31]: additive, 0.91 (0.74–1.12)
		Zhao et al. [38]: case/control allele freq, 0.52/0.42 (*P* < 0.05)
TYMS/TS 28 bp repeat (2R > 3R)	Vijayakrishnan and Houlston [12]: dominant, OR = 1.08 (0.87–1.35) (4 studies, *P* heterogeneity = 0.79)	Lightfoot et al. [40]: others vs. 2R/2R, 0.93 (0.74–1.17)
		Yeoh et al. [39]: Chinese-2R/2R vs. 3R/3R, 0.56 (0.22–1.44); Malays-2R/2R vs. 3R/3R, 3.92 (0.74–20.8)
		Chan et al. [37]: 3R/2R vs. 3R/3R, 1.02 (0.56–1.86)
		Canalle et al. [66]: others vs. 2R/2R, 0.60 (0.37–0.99)

Table 1 continued

Gene, variant	Most recent meta-analysis	Other studies subsequent to meta-analysis[a]
		Rahimi et al. [113]: others vs. 3R/3R, 1.40 (0.76–2.60)
		Metayer et al. [31]: variant not evaluated/reported; associations noted in ENOFS1, a regulator of TYMS;
		Lupo et al. [20]: not evaluated/reported; other variants in gene not significantly associated
TYMS/TS 1494 del6 (rs16430)	Not available	Gast et al. [41]: homozygous, 1.0 (0.6–1.5)
		Lightfoot et al. [40]: homozygous, 1.46 (1.02–2.08)
		Yeoh et al. [39]: Malays-homozygous, 1.70 (0.91–3.19); Chinese-homozygous, 0.82 (0.46–1.46)
		Chan et al. [37]: homozygous, 0.79 (0.34–1.84)
MTHFD1 G401A (rs1950902, R134K)	Not available	Gast et al. [41]: homozygous, 0.7 (0.3–1.4)
		Metayer et al. [31]: additive, 0.91 (0.71–1.18), haplotype association in different MTHFD1 region noted
MTHFD1 G1958A (rs2236225, R653Q)	Not available	Gast et al. [41]: homozygous, 1.1 (0.8–1.5)
		Yeoh et al. [39]: Malays-homozygous, 1.44 (0.64–3.24); Chinese-homozygous, 1.15 (0.60–2.19)
		Chan et al. [37]: homozygous, 0.77 (0.29–2.05)

[a] Study results are presented if reported for the specific variant in question. A study may have evaluated the gene, but association with the specific SNP may not have been tested

[b] Results of the meta-analysis by Vijayakrishnan and Houlston [12] were not significant after adjustment using the false-positive report probabilities method

membrane transporters and undergo cellular metabolic processes that alter activity. The complete metabolism of xenobiotic compounds is divided into two phases, each utilizing different sets of metabolic enzymes. The metabolic activation of the xenobiotic performed by the phase I (bioactivation) enzymes are usually necessary in order for the phase II (detoxification) enzymes to convert this activated intermediate into a detoxified water-soluble compound that is eliminated from the cell. Genetic polymorphisms that disrupt the equilibrium between these two phases may compromise the hosts' ability to respond sufficiently to xenobiotics and may potentially increase the hosts' susceptibility to developing cancer.

Of the xenobiotic membrane transporters, the ATP-binding cassette sub-family B member 1 (*ABCB1* or *MDR1*) gene has received the most attention in childhood leukemia, particularly the C3435T (rs1045642) and G2677T/A (rs2032582) variants, both of which have been linked to gene function in numerous studies [42, 43] (Table 2). Results of two meta-analyses [12, 44] of the five previous reports are both suggestive of an increased risk of childhood ALL associated with homozygous carriers of C3435T (OR = 1.27, 95 % CI 0.99–1.63), but when considered with two subsequent studies showing non-significant associations [16, 39], interpretation remains

inconclusive. Meta-analysis of two studies for G2677T/A showed no association [12].

Several variants within the phase I metabolism cytochrome P450 (*CYP*) family of genes have been evaluated in childhood ALL. In 2010, a statistically significant increased risk of childhood ALL associated with *CYP1A1* T6235C (rs4646903, dominant: OR = 1.36, 95 % CI 1.11–1.66) was reported based on a meta-analysis of 7 studies [12], and more recently has been confirmed in another meta-analysis comprising a total of 12 studies (dominant: OR = 1.31, 95 % CI 1.08–1.59) (Table 2) [45]. A second well-characterized variant in this gene, A4889G (rs1048943) evaluated by a fewer number of studies showed an elevated, but non-significant, risk estimate (homozygous: OR = 1.96, 95 % CI 0.92–4.19) in a meta-analysis of three studies [12]. A subsequent study conducted in a multi-ethnic US population of 258 B-lineage ALL cases and 646 matched controls reported a strong increased risk associated with A4889G observed predominantly in Hispanics (dominant: OR = 2.47, 95 % CI 1.13–5.38) [46]. Elevated risk estimates were reported in recent studies conducted in Brazil [47] and a multi-ethnic California population [16], but not in a Korean study [48]. Evidence for an association with *CYP2E1*5B*, a restriction fragment length polymorphism at G-1293C/C-1053T

Table 2 Evidence from candidate gene association studies for gene variants in the xenobiotic transport/metabolism and DNA repair pathways that have been evaluated with meta-analysis or have greater than three previous publications

Gene, variant	Most recent meta-analysis	Other studies subsequent to meta-analysis[a]
Xenobiotic transport *ABCB1/MDR1* C3435T (rs1045642)	Vijayakrishnan and Houlston [12]: dominant, OR = 1.12 (0.92–1.38); homozygous, OR = 1.26 (0.98–1.62) (5 studies, P heterogeneity = 0.04)	Yeoh et al. [39]: Chinese-homozygous, 1.11 (0.67–1.84); Malays-homozygous, 1.38 (0.80–2.37)
	Wang et al. [44]: homozygous, OR = 1.27 (0.99–1.63) (5 studies, P heterogeneity = 0.06)	Chokkalingam et al. [16]: homozygous, 1.11 (0.76, 1.62), haplotype association in different region of gene and GxE with insecticide exposure
ABCB1/MDR1 G2677T/A (rs2032582)	Vijayakrishnan and Houlston [12]: dominant, OR = 1.01 (0.78–1.32) (2 studies, P heterogeneity = 0.88)	Not available
Xenobiotic metabolism *EPHX1* T339C (rs1051740, Y113H)	Not available	Clavel et al. [54]: homozygous, 0.5 (0.2–1.0)
		Silveira et al. [50]: dominant, 0.26 (0.16–0.42)
		Chokkalingam et al. [16]: homozygous, 0.97 (0.63–1.51)
		Bonaventure et al. [49]: homozygous, 1.4 (0.8–2.4)
		Tumer et al. [55]: homozygous, 2.3 (1.2–4.4)
EPHX1 A418G (rs2234922, H139R)	Not available	Clavel et al. [54]: no association
		Silveira et al. [50]: dominant, 1.40 (0.90–2.18)
		Bonaventure et al. [49]: homozygous, 1.1 (0.6–2.8)
		Tumer et al. [55]: dominant, 0.9 (0.6–1.4)
CYP1A1 T6235C (rs4646903, Msp1, CYP1A1*2A)	Vijayakrishnan and Houlston [12]: dominant, OR = 1.36 (1.11–1.66) (7 studies, P heterogeneity = 0.04)[b]	Not available
	Zhuo et al. [45]: dominant, OR = 1.31 (1.08, 1.59) (12 studies, P heterogeneity = 0.10)	
CYP1A1 A4889G (rs1048943, I462 V, CYP1A1*2C)	Vijayakrishnan and Houlston [12]: dominant, OR = 1.22 (0.88–1.70); homozygous, OR = 1.96 (0.92, 4.19) (3 studies, P heterogeneity = 0.180)	Lee et al. [48]: dominant, 0.9 (0.5–1.5), haplotype interaction with paternal smoking
		Yamaguti et al. [47]: dominant, 1.36 (0.76–2.44)
		Swinney et al. [46]: dominant, 2.51 (1.18–5.33)
		Chokkalingam et al. [16]: heterozygous, 1.29 (0.93, 1.79)
CYP2D6 G1934A (rs3892097, CYP2D6*4)	Vijayakrishnan and Houlston [12]: dominant, OR = 1.07 (0.79–1.44) (3 studies, P heterogeneity = 0.19)	Silveira et al. [50]: heterozygous, 1.13 (0.40–3.17), Chokkalingam et al. [16]: SNP not evaluated/reported; other SNPs in gene not associated
CYP2D6 del2637 (rs35742686, CYP2D6*3)	Vijayakrishnan and Houlston [12]: dominant, OR = 0.86 (0.41–1.81) (3 studies, P heterogeneity = 0.33)	Silveira et al. [50]: dominant, 1.57 (0.99–2.48)
CYP2E1 5B (rs3813867/rs2031920, G-1293C/C-1053T)	Vijayakrishnan and Houlston [12]: dominant, OR = 1.99 (1.32–3.00) (4 studies, P heterogeneity = 0.24)[b]	Chokkalingam et al. [16]: homozygous, 2.08 (0.49–8.86)
		Bonaventure et al. [49]: dominant-0.9 (0.5–1.5)
CYP3A5 G6986A (rs776746, CYP3A5*3, splice variant)	Not available	Pakakasama et al. [52]: homozygous, 0.7 (0.3–1.5)
		Borst et al. [51]: A allele, 1.64 (1.01–2.66)
		Chokkalingam et al. [16]: no association
		Silveira et al. [53]: dominant, 0.35 (0.17–0.72)

Table 2 continued

Gene, variant	Most recent meta-analysis	Other studies subsequent to meta-analysis[a]
GSTM1 deletion	Vijayakrishnan and Houlston [12]: null-allele, OR = 1.16 (1.04–1.30) (15 studies, P heterogeneity = 0.20)[b]	Suneetha et al. [57]: 1.96 (1.08–3.57)
		Rimando et al. [56]: 2.37 (1.11–5.04)
		Yeoh et al. [39]: Chinese, 1.11 (0.81–1.52); Malays, 0.87 (0.60–1.24)
		Chan et al. [37]: 1.49 (0.93–2.37)
		Chokkalingam et al. [16]: Hispanics, 1.85 (1.19–2.88); non-Hispanics, 0.62 (0.43–0.89)
GSTT1 deletion	Vijayakrishnan and Houlston [12]: null-allele, OR = 0.96 (0.83–1.11) (15 studies, P heterogeneity <0.01)	Rimando et al. [56]: 1.08 (0.51–2.29)
		Yeoh et al. [39]: Chinese, 1.20 (0.88–1.65); Malays, 1.09 (0.75–1.59)
		Chan et al. [37]: 1.38 (0.88–2.16)
		Chokkalingam et al. [16]: 0.89 (0.63, 1.27)
		Yang et al. [23]: Evidence of case-only GxE with proximity to industrial plants
GSTP1 A1578G (rs1695, Ile105Val)	Vijayakrishnan and Houlston [12]: dominant, OR = 1.08 (0.93–1.26) (7 studies, P heterogeneity = 0.95)	Suneetha et al. [57]: homozygous, 1.41 (0.50–3.96)
		Chan et al. [37]: homozygous, 1.33 (0.56–3.17)
		Chokkalingam et al. [16]: SNP not evaluated/reported; another SNP in gene not associated
GSTP1 C2293T (rs1138272, Ala114Val)	Vijayakrishnan and Houlston [12]: dominant, OR = 0.73 (0.48–1.13) (2 studies, P heterogeneity = 0.49)	Not available
NQO1 C609T (rs1800566, NQO1*2)	Vijayakrishnan and Houlston [12]: dominant, OR = 1.24 (1.02–1.50) (6 studies, P heterogeneity = 0.49)[b]	Rimando et al. [56]: C/C vs. C/T, 4.82 (2.18–10.6)
		Yeoh et al. [39]: Chinese-homozygous, 0.74 (0.46–1.18); Malays-heterozygous, 0.54 (0.36–0.79)
		Chan et al. [37]: homozygous, 1.25 (0.65–2.43)
		Yamaguti et al. [47]: dominant, 2.64 (1.46–4.80)
		Silveira et al. [50]: dominant, 0.63 (0.16–2.51)
		Bonaventure et al. [49]: dominant, 1.1 (0.8–1.4)
		Goncalves et al. [60]: dominant, ALL2-10 years, 0.98 (0.68–1.40)
		Chokkalingam et al. [16]: homozygous, 1.06 (0.60, 1.87)
NQO1 C465T (rs4986998, NQO1*3, Arg139Trp)	Vijayakrishnan and Houlston [12]: dominant, OR = 1.59 (0.95–2.65) (2 studies, P heterogeneity = 0.25)	Not available
NAT2 (rs1801280, NAT2*5)	Not available	Krajinovic et al. [58]: slow (*5A-C, *6A, *7B), 1.5 (1.0–2.2)
		Chokkalingam et al. [16]: no association
		Zanrosso et al. [59]: *5 allele, 2.3 (1.5–3.5)
		Bonaventure et al. [49]: homozygous, 1.7 (1.2–2.6)
		Silveira et al. [53]: slow (homozygous for *5A-C, *6A or *7B), 1.10 (0.40–3.03)

Table 2 continued

Gene, variant	Most recent meta-analysis	Other studies subsequent to meta-analysis[a]
DNA repair		
ERCC2/XPD	Vijayakrishnan and Houlston [12]:	Canalle et al. [66]: no association
G23591A (rs1799793, Asp312Asn)	dominant, OR = 0.78 (0.48–1.27) (2 studies, *P* heterogeneity = 0.59)	Chokkalingam et al. [15]: SNP not evaluated/ reported; haplotype of other SNPs in gene associated
ERCC2/XPD	Vijayakrishnan and Houlston [12]:	Chokkalingam et al. [15]: additive, 1.08 (0.87–1.35)
A35931C (rs13181, Lys751Gln)	dominant, OR = 1.04 (0.67–1.60) (2 studies, *P* heterogeneity = 0.75)	
		Canalle et al. [66]: no association
XRCC1	Wang et al. [64]: dominant, OR = 1.06 (0.85–1.31) (6 studies, *P* heterogeneity = 0.059)	Chokkalingam et al. [16]: additive, 0.94 (0.67–1.31)
C26304T (rs1799782, Arg194Trp)	Wang et al. [65]: dominant, OR = 1.16 (0.87–1.56) (7 studies, *P* heterogeneity = 0.094)	
XRCC1	Wang et al. [64]: dominant, OR = 1.32 (1.10–1.57) (7 studies, *P* heterogeneity = 0.016)	Not available
G28152A (rs25487, Arg399Gln)	Wang et al. [65]: dominant, OR = 1.27 (0.98–1.66); homozygous, OR = 1.44 (1.07–1.93) (8 studies, *P* heterogeneity = 0.448)	
XRCC1	Wang et al. [64]: dominant, OR = 1.13 (0.81–1.57) (3 studies, *P* heterogeneity = 0.534)	Not available
G27466A (rs25489, Arg280His)		

[a] Study results are presented if reported for the specific variant in question. A study may have evaluated the gene, but association with the specific SNP may not have been tested

[b] Results of the meta-analysis by Vijayakrishnan and Houlston [12] were not significant after adjustment using the false-positive report probabilities method

(rs3813867/rs2031920), is provided by a meta-analysis of four studies showing a strong increased risk (dominant: OR = 1.99, 95 % CI 1.32–3.00), and one of two additional studies reported results in a consistent direction [16, 49].

Regarding other phase I metabolism genes previously examined, available evidence has been less consistent for the *CYP2D6*, *CYP3A5*, and epoxide hydrolase 1 (*EPHX1*) genes. Meta-analyses of two variants in *CYP2D6*, G1934A (rs3892097) and del2637 (rs35742686) [12], each based on results from three studies, show weak evidence of an association, which is reinforced by inconclusive results in a recent study [50] (Table 2). Statistically significant risk estimates have been reported in both directions for *CYP3A5* G6986A (rs776746) [16, 51–53] and *EPHX1* T339C (rs1051740) [16, 49, 50, 54, 55], and four studies evaluating *EPHX1* A418G (rs2234922) have shown largely non-significant results [49, 50, 54, 55].

Assessment of the accumulating evidence for genetic variation within several phase II metabolic genes suggests an association with the widely studied glutathione S-transferase class mu (*GSTM1*) gene deletion, and *N*-acetyltransferase 2 (*NAT2*) slow acetylator genotype (Table 2). A meta-analysis of 15 studies showed a statistically significant increased risk

associated with the *GSTM1* deletion (OR = 1.16, 95 % CI 1.04–1.30) [12], and an additional 3 of the 5 subsequent studies reported a statistically significant elevated risk [16, 37, 39, 56, 57]. In contrast, the risk associations for the *GSTT1* deletion and two *GSTP1* variants, A1578G (rs1695) and C2293T (1138272), are not supported by recent meta-analysis results [12] and by studies published subsequently [16, 37, 39, 56, 57].

Recently, an increasing number of studies have evaluated the effect of the slow *NAT2* acetylator genotype on childhood leukemia risk (Table 2). The class of slow acetylator alleles (*NAT2*5A-5C*, *6A*, and *7B*) is represented by combinations of polymorphic sites (C282T, rs1041983; T341C, rs1801280; C481T, rs1799929; G590A, rs1799930; A803G, rs1208; and G857A, rs1799931). While previous studies have used slightly varying classifications of acetylation status for NAT2, three studies have reported statistically significant increased ALL risks associated with slow acetylator alleles [49, 58, 59], and two studies observed elevated but non-significant risk estimates [16, 53].

Evidence from a meta-analysis of 6 studies indicated a marginally significant association for NAD(P)H dehydrogenase quinone type 1 (*NQO1*) C609T (rs1800566,

dominant: OR = 1.24, 95 % CI 1.02–1.50), but results from a large number of studies published more recently appear inconsistent, with statistically significant result reported in both directions [16, 37, 39, 47, 49, 50, 56, 60] (Table 2). One earlier meta-analysis examining the C609T effect by leukemia subtype reported no association in childhood ALL overall, but an increased risk of *MLL* translocation positive leukemia [61]. Heterogeneity in results across studies for C609T may be influenced by the specificity/distribution of leukemia subtype characteristic of each study population. Association with a second *NQO1* variant, C465T (rs4986998), appear suggestive of an increased risk based on two studies [12], but more studies are needed for confirmation.

DNA repair

Childhood ALL results from chromosomal alterations and somatic mutations that disrupt the normal process by which lymphoid progenitor cells differentiate and senesce [1]. These are the result of unrepaired DNA damage such as double-strand breaks (DSB) [62]. Among the few established risk factors for childhood ALL are exposure to ionizing radiation and certain chemotherapeutic agents [10], which are well-known genotoxic exposures. Repair of DNA damage is critical [63], thus, alterations in innate DNA repair pathways including nucleotide excision repair (NER), mismatch repair (MMR), and DSB repair may play a role in leukemia development.

Two independently conducted meta-analyses have recently reported statistically significant increased risks associated with X-ray repair cross-complementing group 1 (*XRCC1*) G28152A (rs25487) [64, 65] (Table 2). Subgroup analyses indicated an effect predominantly in Asians in both reports. No associations with childhood ALL were found for the two other widely studied *XRCC1* variants: C26304T (rs1799782), which showed heterogeneity in results across six studies [64, 65], and G27466A (rs25489) examined in fewer studies [64]. Meta-analysis of results from two studies [12], together with two additional reports [15, 66], shows weak evidence of an association with variants of the excision repair cross-complementing group 2 (*ERCC2* or *XPD*) gene (G23591A, rs1799793 and A35931C, rs13181).

Immune response

Exposure to common infections and the role of immune-related processes have emerged as strong candidate risk factors for childhood ALL [67]. One prevailing hypothesis (the delayed infection hypothesis) suggests that a delay in exposure to immune-modulating factors (e.g., exposure to infections) early in life leaves the immune network under-modulated, and subsequent exposure to infections may result in an adverse immune response that gives proliferative advantage to pre-leukemia cells [68]. An immune response to a foreign antigen involves a complex cascade of events beginning with the activation of T lymphocytes and accompanied by vast production/secretion of cytokines and recruitment of other immune cells. Genetic variation influencing immunological pathways including innate and adaptive immunity may affect ALL susceptibility.

Due to its highly polymorphic nature and central role in immune response, the human leukocyte antigen (*HLA*) genes were one of the first loci within the immune response pathway to be examined in childhood leukemia. *HLA* is polygenic and broadly classified into class I (*HLA-A*, *B*, and *C*) and II (*HLA-DR*, *DQ*, and *DP*) genes which encode cell surface glycoproteins that bind and present processed antigens to T lymphocytes crucial to both cellular and humoral immune response [69]. Adding to the complexity, *HLA* alleles are defined based on combinations of genetic variants (haplotypes) that reside within the region encoding the antigen binding groove of the HLA molecule. Thus, *HLA* genes are multi-allelic with up to many hundreds of alleles segregating at a single locus (http://www.ebi.ac.uk/imgt/hla/).

The HLA-DR53 antigen encoded by the *HLA-DRB4* locus which exists only on haplotypes possessing *HLA-DRB1*04*, *07* and *09* has been associated with increased risks for the major types of leukemia in adults and children [70], including evidence from two studies of childhood ALL indicating a male-specific increase in *HLA-DRB4* alleles in cases compared to controls [70, 71]. More recently, two additional studies showed increased risks associated with *HLA-DRB1*04* in an Iranian [72] and European American population [73]. The *HLA-DRB1*15* allele was also identified as a risk locus in two studies [74, 75]. A second *HLA* gene, *HLA-DPB1*, has also emerged as a potential ALL-associated locus based on results from three studies [76–78].

Summary

Based on single-marker studies, a number of genes appear interesting with the most consistency across studies observed for *MTHFR* C677T, *CYP1A1* T6235C, the *GSTM1* deletion, *NAT2*5*, *XRCC1* G28152A and *HLA-DRB4* encoding the HLA-DR53 antigen. Potentially emerging associations that would benefit from additional supportive data include *MTR* A2756G, *MTRR* A66G, *SHMT1* C1420T, *SLC19A1* G80A of the folate pathway, and *ABCB1* C3435T, *CYP1A1* A4889G, *CYP2E1*5B*, and *NQO1* C609T of the xenobiotic transport/metabolism pathway. Also, a number of studies showing interesting gene–environment and gene–gene interaction have surfaced, but sufficient numbers of comparable reports are not yet available. Since we focus this review on genes with more than three reports, we acknowledge that there are

many other isolated reports of relevant gene associations, including haplotype-based associations, not mentioned in this review that show promise as a potential disease locus.

Genome-wide association studies

Results from the first GWAS in childhood ALL were reported in 2009 by two studies conducted independently in populations of European ancestry [79, 80] (Table 3). Taking a meta-analytic approach and combining the results for two separate GWAS analyses that comprised a total of 907 ALL cases and 2,398 controls, Papaemmanuil et al. [79] identified genome-wide significant (P threshold $<5 \times 10^{-7}$) associations within and/or near the Ikaros family zinc finger 1 (*IKZF1*, chromosomal region 7p12.21) gene, the AT-rich interactive domain 5B (*ARID5B*, chromosomal region 10q21.2) gene, and CCAAT/enhancer-binding protein epsilon (*CEBPE*, chromosomal region 14q11.2). Described in a concurrent report, Trevino et al. [80] also identified associated regions within *ARID5B* and *IKZF1* in their discovery population of 317 ALL cases and 17,958 controls. They performed a validation in a case-only series of 124 ALL patients and showed that the *ARID5B* association clearly distinguishes B-cell hyperdiploidy from other subtypes. The *IKZF1* association appeared to predominantly affect B-cell ALL in the discovery study, but this was not clearly distinguished in the validation series. These subtype specific effects were also reported in the Papaemmanuil et al. study. Together, the two studies offered the first unequivocal evidence of a role for inherited genetic susceptibility in childhood ALL risk conferred by inter-individual variation within specific genomic regions and indicating a likely role of *IKZF1*, *ARID5B*, and *CEBPE*. Subsequently, independent replication studies have shown consistent associations in populations from Poland [81], Germany [82], the UK [82], Thailand [83] and among the US infants [84] for *IKZF1* and *ARID5B*, and additionally in Canada [85] and a US population of African ancestry [86] for *ARID5B*. Replication of the *CEPBE* risk variant has been less consistent with one successful attempt in a German population [82], but as a variant associated with a more modest risk estimate, lack of statistical power cannot be ruled out [81, 83].

Based on the initial GWAS from the Papaemmanuil et al. study conducted in the UK, Sherborne et al. [87] pursued a replication attempt for 34 of the top P value-ranked SNPs using multiple independent populations and identified a fourth associated region that localized to the cyclin-dependent kinase inhibitor 2A (*CDKN2A*, chromosomal region 9p21.3) gene (Table 3). This association remained highly significant in both B- and T-cell lineage ALL which is in contrast to the findings for *IKZF1* and *ARID5B* found to be mostly B-cell subtype specific.

Focused on *TEL-AML1*-positive childhood ALL, Ellinghaus et al. [88] conducted a GWAS comprising a discovery series of 419 German cases and 474 controls, followed by 2 independent replication series comprising 951 cases and 3,061 controls from Germany, Austria, and Italy (Table 3). In addition to confirming the associations at the three previously reported loci (*IKZF1*, *ARID5B*, and *CEBPE*), they identified four additional associations that localize to tumor protein p63 (*TP63*, chromosomal region 3q28), protein tyrosine phosphatase receptor type J (*PTPRJ*, chromosomal region 11p11.2), olfactory receptor family 8 subfamily U member 8 (*OR8U8*, chromosomal region 11q11), and integrator complex subunit 10 (*INTS10*, chromosomal region 8p21.3) genes, the latter 2 of which showed heterogeneity between the German/Austrian and Italian replication series.

In a discovery series of 441 French ALL cases and 1,542 controls, Orsi et al. [89] performed a GWAS and included a replication set comprising 390 Australian ALL cases and 1,202 controls (Table 3). Their strongest findings were consistent with the previously identified *IKZF1* and *ARID5B* regions, followed by suggestive results localizing two potentially novel regions near ankyrin repeat domain 44 (*ANKRD44*, chromosomal region 2q33.1) and solute carrier family 16, member 14 (*SLC16A14*, chromosomal region 2q36.3). However, these latter two loci did not successfully replicate. Their study also provided additional support for the *CDKN2A* and *CEBPE* ALL-associated loci, but *TEL-AML1*-positive ALL-specific results did not confirm the associations for *OR8U8*, *INTS10*, and *PTPRJ* previously identified [88].

The first Asian childhood ALL GWAS was conducted in a Korean population by Han et al. [90] comprising 50 cases and 50 controls (Table 3). Even in this relatively small study, associations for *ARID5B* and *CEBPE*, but not *IKZF1*, were detected and indicated a role in ALL risk in this population. Using the false discovery rate method to adjust for multiple testing, they report potential ALL associations with loci that map to mannosidase alpha class 2A member 1 (*MAN2A1*, chromosomal region 5q21), hydroxyacid oxidase 1 (*HAO1*, chromosomal region 20p12), chromosome 2 open reading frame 3 (*C2orf3*, chromosomal region 2p12), and erythrocyte membrane protein ban 4.1-like 2 (*EPB41L2*, chromosomal region 6q23) genes. Given the potential heterogeneity in associations across race/ethnicities, these novel loci warrant further consideration for replication in comparable populations.

Summary

Collectively based on these five previous GWAS and various follow-up analyses, strong evidence of an

Table 3 Summary of results for previous genome-wide association studies of childhood ALL risk

Author (Reference) and study features	Gene (region)	SNP	OR (95 % CI)	P	Comment
Papaemmanuil et al. [79], UK 907 ALL (GWA1-503; GWA2-404) 2,398 controls (GWA1-1,438; GWA2-960)	*IKZF1* (7p12.2)	rs4132601	1.69 (1.58–1.81)	1.20×10^{-19}	Strongest of 6 significant SNPs in LD; primary impact is for B-lineage ALL
	ARID5B (10q21.2)	rs7089424	1.65 (1.54–1.76)	6.69×10^{-19}	Strongest of 3 significant SNPs in LD; primary impact is for B-lineage ALL; predominately affects hyperdiploid ALL
	CEBPE (14q11.2)	rs2239633	1.34 (1.22–1.45)	2.88×10^{-7}	Primary impact is for B-lineage ALL
Trevino et al. [80], USA Discovery: 317 ALL, 17,958 controls (from 3 sources)	*IKZF1* (7p12.2)	rs11978267	1.69 (1.4–1.9)	8.8×10^{-11}	Distinguishes T-cell from B-lineage ALL; case–case analysis did not replicate, but suggestive
Replication: 124 ALL	*ARID5B* (10q21.2)	rs10821936	1.91 (1.6–2.2)	1.4×10^{-15}	Stronger of 2 significant SNPs in LD; predominately affects hyperdiploid ALL
Han et al. [90], South Korea 50 ALL cases	*MAN2A1* (5q21)	rs3776932	0.11 (0.04–0.34)	1.0×10^{-4}	
50 controls	*HAO1* (20p12)	rs6140264	8.84 (2.92–26.8)	1.0×10^{-4}	Dominant model
	C2orf3 (2p12)	rs12105972	0.13 (0.04–0.36)	1.0×10^{-4}	Dominant model
	EPB41L2 (6q23)	rs9388856	8.97 (2.93–27.5)	1.0×10^{-4}	Strongest of 3 SNPs in LD; dominant model
	ARID5B (10q21.2)	rs10821936	3.21 (1.45–7.07)	0.0039	Reported in Trevino et al. [80]
	ARID5B (10q21.2)	rs7896246	2.90 (1.34–6.27)	0.0067	Strong LD with rs7089424, reported in Papaemmanuil et al. [79]
	IKZF1 (7p12.2)	rs4132601	Not associated	NA	Reported in Papaemmanuil et al. [79]
	CEBPE (14q11.2)	rs4982731	3.29 (1.33–8.13)	0.01	In LD with rs2239633, reported in Papaemmanuil et al. [79]
Sherborne et al. [87], UK, Germany, Austria, Canada, Hungary, Spain Discovery: 907 ALL, 2,398 controls (UK-GWA1, UK-GWA2) Replication (Germany, Canada, Hungary, Spain): 2,386 ALL, 2,419 controls	*CDKN2A* (9p21.3)	rs3731217	0.71 (0.64–0.78)	3.01×10^{-11}	Associated with both B- and T-cell ALL
Ellinghaus et al. [88], Austria, Germany, Italy	*TP63* (3q28)	rs17505102	0.65 (NA)	8.94×10^{-9}	
Discovery: 419 *TEL-AML1* + ALL, 474 controls	*PTPRJ* (11p11.2)	rs3942852	0.72 (NA)	4.95×10^{-7}	
Replication: (Panel B ext) 664 cases, 2,482 controls; (Panel C) 287 cases, 579 controls; (Panel D) 326 cases	*OR8U8* (11q11)	rs1945213	0.69 (NA)	9.14×10^{-11}	More specific to Northern Europeans
	INTS10 (8p21.3)	rs920590	1.36 (NA)	6.12×10^{-9}	More specific to Northern Europeans
	IKZF1 (7p12.2)	rs4132601	1.43 (1.30–1.58)	8.26×10^{-13}	Reported in Papaemmanuil et al. [79]
	IKZF1 (7p12.2)	rs11978267	1.44 (1.30–1.59)	1.07×10^{-12}	Reported in Trevino et al. [80]
	ARID5B (10q21.2)	rs7089424	1.42 (1.30–1.56)	1.98×10^{-13}	Reported in Papaemmanuil et al. [79]
	ARID5B (10q21.2)	rs10821936	1.46 (1.33–1.60)	4.06×10^{-15}	Reported in Trevino et al. [80]
	CEBPE (14q11.2)	rs2239633	0.74 (0.68–0.82)	4.00×10^{-10}	Reported in Papaemmanuil et al. [79]

Table 3 continued

Author (Reference) and study features	Gene (region)	SNP	OR (95 % CI)	P	Comment
Orsi et al. [89], France, Australia Discovery: 441 ALL, 1,542 controls Replication: 390 ALL, 1,202 controls	*IKZF1* (7p12.2)	rs4132601	1.59 (1.34–1.89)	1.2×10^{-7}	Reported in Papaemmanuil et al. [79]
	IKZF1 (7p12.2)	rs11978267	1.6 (1.4–1.9)	$<5 \times 10^{-7}$	Reported in Trevino et al. [80]
	ARID5B (10q21.2)	rs7089424	1.83 (1.55–2.15)	6.1×10^{-13}	Strongest of 3 significant SNPs in region; reported in Papaemmanuil et al. [79]
	ARID5B (10q21.2)	rs10821936	1.9 (1.6–2.2)	$<5 \times 10^{-7}$	Reported in Trevino et al. [80]
	CEBPE (14q11.2)	rs2239633	0.8 (0.7–1.0)	$10^{-3} \leq P < 0.05$	Reported in Papaemmanuil et al. [79]
	CDKN2A (9p21.3)	rs3731217	0.8 (0.6–1.0)	$10^{-3} \leq P < 0.05$	Reported in Sherborne et al. [87]
	ANKRD44 (2q33.1)	rs930372	2.37 (1.65–3.41)	3.1×10^{-6}	Not replicated
	SLC16A14 (2q36.6)	rs12162384	0.56 (0.44–0.73)	9.9×10^{-6}	Not replicated
	PTPRJ (11p11.2)	rs3942852	1.2 (0.8–1.8)	NS	Reported in Ellinghaus et al. [88]
	OR8U8 (11q11)	rs1945213	1.0 (0.6–1.4)	NS	Reported in Ellinghaus et al. [88]
	INTS10 (8p21.3)	rs920590	1.0 (0.7–1.1)	NS	Reported in Ellinghaus et al. [88]

association with ALL overall and/or specific ALL subtypes is available for four independent loci implicating a role for the *IKZF1*, *ARID5B*, *CEBPE*, and *CDKN2A* genes. Strong evidence is also available for another four loci around the *TP63*, *PTPRJ*, *OR8U8*, and *INTS10* genes found associated with *TEL-AML1*-positive ALL, but additional confirmation appears necessary. The IKAROS transcription factor encoded by the *IKZF1* gene is a regulator of lymphocyte differentiation. In germline mutant mice exhibiting loss of *IKZF1* expression, lymphocyte development is inhibited and leads to an aggressive form of lymphoblastic leukemia [79, 91]. In humans, chromosomal deletions involving *IKZF1* are observed in about a third of high-risk B-cell precursor ALL and a large proportion of *BCR-ABL1*-positive ALL patients [92, 93]. As a gene found to be involved in embryogenesis and growth retardation, the putative contributions of *ARID5B* to leukemogenesis is less understood. However, animal studies have shown defects in the B lymphoid compartment in *Arid5b* knockout mice [94], and *ARID5B* expression has been shown to be up-regulated in acute promyelocytic leukemia [95]. A third gene involved in lymphoid development and/or leukemo-genesis, *CEBPE*, is a known target of translocations in B-cell precursor ALL [96]. *CDKN2A* is a tumor suppressor gene that encodes p16, a negative regulator of cyclin-dependent kinases, and p14, an activator of p53. Deletion of *CDKN2A* is among the most common genetic events in

childhood B- and T-lineage ALL [87, 97]. The potential role for *OR8U8* and *INTS10* annotated to SNPs found to be associated with *TEL-AML1*-positive ALL is less clear. Similarly, the direct mechanisms for an involvement of *TP63* and *PTPRJ* in leukemogenesis are not well understood, but *TP63* has been shown to possess features of a tumor suppressor, and *PTPRJ* involvement has been described in a number of cancers [88]. In addition, mice with impaired protein tyrosine phosphatase encoded by *PTPRJ* have been shown to display a partial peripheral B-cell developmental block [98].

Discussion

The strengths and limitations of each of the candidate gene and genome-wide approaches suggest that they should be viewed as complementary strategies to an integrated effort in identifying disease-associated loci. While the candidate gene approach allows the testing of specific hypothesis-driven questions without the stringent P value requirements as those for genome-wide studies, success is dependent on selecting the correct gene (among thousands) and the markers that accurately capture the genetic variation of the locus of interest. In this respect, the agnostic genome-wide strategy has a major advantage, and any finding that is statistically significant at the genome-wide level will have

a very high probability of being a true association in that study population. One limitation of the genome-wide approach that can be addressed by the candidate gene studies is the lack of adequate coverage of certain regions of the genome by the currently available genome-wide arrays, including for multi-allelic and non-SNP markers and certain rare variants. An example is in the evaluation of the *HLA* [73] and killer-cell immunoglobulin-like receptor (*KIR*) [99] genes. While the hope is that any association with these loci would be captured by tag SNPs within the region [100], their highly multi-allelic nature and a potential involvement of rare alleles may require a focused examination using specialized methods.

Although GWAS are widely recognized as rigorous in their approach to identifying disease loci [101], the *P* value requirements for defining a significant association, in return, increases the probability of missing a true association as recently demonstrated specifically in childhood leukemia. Using previously derived methods [102, 103] applied to an analysis of nearly 250,000 SNP genotypes from the UK GWAS [79], Enciso-Mora et al. [6] estimated that about 24 % of the total variation in childhood B-cell precursor ALL risk is accounted for by common genetic variation, and the previously identified loci (*IKZF1*, *ARID5B*, *CEBPE*, and *CDKN2A*) explain only 8 % of this total. This study provides evidence for a polygenic mechanism of susceptibility and rationale for continued investigation of additional susceptibility loci that were likely missed by previous GWAS.

Areas of ongoing and future work

Two main areas to emphasize for ongoing and future efforts are: (1) attempt to explain more of the remaining total variation of childhood ALL risk that is estimated to be due common genetic variants by identifying additional associated loci, and (2) to characterize the identified regions through fine-mapping and functional studies.

Undetected associations due to insufficient sample sizes, which pertain to both genome-wide and candidate gene studies, is a central issue underscored by both the Enciso-Mora et al. report [6] and previous reviews [5, 104]. Power calculations have shown that even the largest GWAS conducted to date (907 cases and 2,398 controls) had limited ability to detect variants conferring relative risks of 1.4 or below and/or those with allele frequencies of <0.10 [6, 104]. Although a candidate gene study hoping to detect an SNP association with an allele frequency of 0.10 conferring a relative risk of 1.4 (assuming power = 0.80 and alpha = 0.05) would require fewer subjects, the sample size would still be substantial: about 600 cases and 600 controls. Furthermore, successful detection depends on whether the correct region and variants were selected a

priori [5]. It is important to note, as well, that studies utilizing small sample sizes are sensitive to even small amounts of bias in genotyping or study design, potentially leading to erroneous results. This is likely a common phenomenon among the previous candidate genes studies reviewed here; about 40 % of studies published within the last 4 years included fewer than 150 cases. These requirements, together with the rarity of childhood leukemia and importance of independent validation [105], amount to advocating coordinated studies and collaborations. The Childhood Leukemia International Consortium (CLIC, https://ccls.berkeley.edu/clic), which currently includes 22 epidemiological studies of childhood leukemia representing 12 countries, was formed to facilitate collaborative efforts such as those needed for genetic studies.

Identification of additional genetic loci will undoubtedly come from rigorous scrutiny of gene–environment and gene–gene interactions. Therefore, it is very important that epidemiological studies continue to collect high-quality environmental, lifestyle, and other exposure data along with the biological samples as a source of DNA. Studies based solely on genetic data, with no collection of environmental data, will be unable to shed light on this critical area. However, it should be noted that widespread testing for combined effects of genes and environmental factors, as well as gene–gene interactions, dramatically increases the number of tests performed, thus compounding the multiple testing problem. Statistical methods for performing such large-scale investigations are being developed [106]. In addition, investigators should develop plans for testing and replicating a limited number of interactions a priori.

While the initiation of new GWAS, particularly in populations of non-European ancestry, still has tremendous potential to yield discoveries in childhood leukemia, the field of genetic epidemiology for the more common cancers and non-cancer phenotypes is now in the "post-GWAS" era [107]. Unlike candidate gene studies where the scientific inquiry was commonly initiated by first having some evidence of a functional consequence of the candidate locus, the genome-wide approach works in reverse order requiring fine-mapping of the localized region to identify the causal variant and subsequent functional characterization to identify the causal genes. The various activities associated with the follow-up of GWAS-identified susceptibility loci are broadly referred to as "post-GWAS" investigations. Data generated from the ongoing 1000 Genomes Project [108], an international collaborative effort with publicly available resources that aims to provide a deep characterization of human genome sequence variation, can aid in fine-mapping efforts.

Complementary to this and for interrogating certain intergenic regions and gene deserts potentially not addressed by the 1000 Genomes Project, targeted re-

sequencing can be undertaken using next generation sequencing technology [109]. The next challenge of identifying/confirming the genes that are regulated by the variants can be addressed using various controlled laboratory experiments, but one initial approach can be to evaluate its association with gene expression. This has already been initiated by previous GWAS investigations for some of the associated SNPs [79, 80, 88]. SNP rs4132601 mapping to the *IKZF1* gene region showed significant correlation with *IKZF1* mRNA expression levels in Epstein–Barr virus-transformed lymphocytes, strongly implicating *IKZF1* as the causal locus tagged by the SNP [79]. Fine-mapping effort for the *ARID5B* SNPs (rs10821936 and rs10994982) did not reveal coding variants in LD, but gene expression studies showed a correlation with global gene expression pattern specifically in B-hyperdiploid ALL blast cells [80]. In addition, existing data from large-scale efforts are available that identify expression quantitative trait loci (eQTLs) [110], genetic variants that tag regions of the genome controlling mRNA expression. GENEVAR is one example of a resource that provides a catalog of information of known eQTLs in select tissue types [111]. Using this resource to examine the four *TEL-AML1*-positive ALL-associated SNPs, significant correlations with expression levels of genes within the localized regions were not observed [88]. Necessary considerations in the interpretation of these data include that eQTLs may be tissue-specific, can vary in time, and are likely to differ between tumor and normal tissue.

Concluding remarks

Just within the past few years, there have been a large number of reports on the genetic susceptibility to childhood ALL, attempting both to confirm candidate gene associations, as well as to make new discoveries by testing novel candidate loci and conducting large-scale GWAS. Such efforts have proven tremendously meaningful, to the point that we now have confirmation of an inherited genetic basis for childhood ALL and have come closer to establishing some candidate genes as risk loci. While most studies have focused on SNPs, which has been the topic of this review, it should be acknowledged that other types of inherited genetic variation may contribute to the overall heritability of childhood ALL, such as epigenetic alterations and copy number variations [112]. Understanding of the influence of these factors in childhood leukemia is growing and it will undoubtedly contribute to major advances in the near future.

Findings described in this review still lie within the realm of research and have not yet yielded tangible clinical results. However, the expectations are that these studies will advance our understanding of the detailed causal pathways leading to disease, and eventually contribute to the development of novel therapeutics, identification of biomarkers for refined disease prediction, and monitoring of disease progression and treatment response [9]. Indeed, this improved knowledge of causal pathways may also reveal clues about modifiable environmental and lifestyle risk factors that can be used to develop public health-based prevention measures. The last few years has seen significant progress in the genetic epidemiology of childhood leukemia, largely fueled by technological advances in genetic analysis and global collaborative efforts in genomics such as the International HapMap and 1000 Genomes Projects. We suspect that with decreasing costs in the use of recent genomic platforms, an increased recognition of the importance of collaboration for validation and to achieve maximal sample sizes, and increased public access to large genomic datasets, the next few years will see even more productivity.

Conflict of interest The authors declare no conflict of interest.

References

1. Pui CH. Childhood leukemias. N Engl J Med. 1995;332: 1618–30.
2. Smith M, Ries L, Gurney J, Ross J. Leukemia. In: Ries L, Smith M, Gurney J, et al., editors. Cancer incidence and survival among children and adolescents: United States SEER Program 1975–1995. Pub. No. 99-4649 ed. Bethesda, MD: National Cancer Institute, SEER Program; 1999. p 17–34.
3. Wiemels J. Perspectives on the causes of childhood leukemia. Chem Biol Interact. 2012;196:59–67.
4. Greaves MF, Wiemels J. Origins of chromosome translocations in childhood leukaemia. Nat Rev Cancer. 2003;3:639–49.
5. Chokkalingam AP, Buffler PA. Genetic susceptibility to childhood leukaemia. Radiat Prot Dosimetry. 2008;132:119–29.
6. Enciso-Mora V, Hosking FJ, Sheridan E, Kinsey SE, Lightfoot T, Roman E, et al. Common genetic variation contributes significantly to the risk of childhood B-cell precursor acute lymphoblastic leukemia. Leukemia. 2012;26:2212–5.
7. Greaves MF, Maia AT, Wiemels JL, Ford AM. Leukemia in twins: lessons in natural history. Blood. 2003;102:2321–33.
8. Wall JD, Pritchard JK. Haplotype blocks and linkage disequilibrium in the human genome. Nat Rev Genet. 2003;4:587–97.
9. McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP, et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet. 2008;9:356–69.
10. Buffler PA, Kwan ML, Reynolds P, Urayama KY. Environmental and genetic risk factors for childhood leukemia: appraising the evidence. Cancer Invest. 2005;23:60–75.
11. Karathanasis NV, Choumerianou DM, Kalmanti M. Gene polymorphisms in childhood ALL. Pediatr Blood Cancer. 2009;52:318–23.
12. Vijayakrishnan J, Houlston RS. Candidate gene association studies and risk of childhood acute lymphoblastic leukemia: a systematic review and meta-analysis. Haematologica. 2010;95: 1405–14.
13. Beuten J, Gelfond JA, Piwkham D, Pollock BH, Winick NJ, Collier AB 3rd, et al. Candidate gene association analysis of

acute lymphoblastic leukemia identifies new susceptibility locus at 11p15 (LMO1). Carcinogenesis. 2011;32:1349–53.

14. Chang JS, Wiemels JL, Chokkalingam AP, Metayer C, Barcellos LF, Hansen HM, et al. Genetic polymorphisms in adaptive immunity genes and childhood acute lymphoblastic leukemia. Cancer Epidemiol Biomarkers Prev. 2010;19:2152–63.

15. Chokkalingam AP, Bartley K, Wiemels JL, Metayer C, Barcellos LF, Hansen HM, et al. Haplotypes of DNA repair and cell cycle control genes, X-ray exposure, and risk of childhood acute lymphoblastic leukemia. Cancer Causes Control. 2011;22:1721–30.

16. Chokkalingam AP, Metayer C, Scelo GA, Chang JS, Urayama KY, Aldrich MC, et al. Variation in xenobiotic transport and metabolism genes, household chemical exposures, and risk of childhood acute lymphoblastic leukemia. Cancer Causes Control. 2012;23:1367–75.

17. Han S, Koo HH, Lan Q, Lee KM, Park AK, Park SK, et al. Common variation in genes related to immune response and risk of childhood leukemia. Hum Immunol. 2012;73:316–9.

18. Han S, Lan Q, Park AK, Lee KM, Park SK, Ahn HS, et al. Polymorphisms in innate immunity genes and risk of childhood leukemia. Hum Immunol. 2010;71:727–30.

19. Park C, Han S, Lee KM, Choi JY, Song N, Jeon S, et al. Association between CASP7 and CASP14 genetic polymorphisms and the risk of childhood leukemia. Hum Immunol. 2012;73:736–9.

20. Lupo PJ, Nousome D, Kamdar KY, Okcu MF, Scheurer ME. A case–parent triad assessment of folate metabolic genes and the risk of childhood acute lymphoblastic leukemia. Cancer Causes Control. 2012;23:1797–803.

21. Taylor M, Bergemann TL, Hussain A, Thompson PD, Spector L. Transmission of HLA-DP variants from parents to children with B-cell precursor acute lymphoblastic leukemia: log-linear analysis using the case–parent design. Hum Immunol. 2011;72:897–903.

22. Yang Y, Jin X, Yan C, Tian Y, Tang J, Shen X. Case-only study of interactions between DNA repair genes (hMLH1, APEX1, MGMT, XRCC1 and XPD) and low-frequency electromagnetic fields in childhood acute leukemia. Leuk Lymphoma. 2008;49:2344–50.

23. Yang Y, Tian Y, Jin X, Yan C, Jiang F, Zhang Y, et al. A case-only study of interactions between metabolic enzyme polymorphisms and industrial pollution in childhood acute leukemia. Environ Toxicol Pharmacol. 2009;28:161–6.

24. Blount BC, Mack MM, Wehr CM, MacGregor JT, Hiatt RA, Wang G, et al. Folate deficiency causes uracil misincorporation into human DNA and chromosome breakage: implications for cancer and neuronal damage. Proc Natl Acad Sci USA. 1997;94:3290–5.

25. Das PM, Singal R. DNA methylation and cancer. J Clin Oncol. 2004;22:4632–42.

26. Wang H, Wang J, Zhao L, Liu X, Mi W. Methylenetetrahydrofolate reductase polymorphisms and risk of acute lymphoblastic leukemia—evidence from an updated meta-analysis including 35 studies. BMC Med Genet. 2012;13:77.

27. Yan J, Yin M, Dreyer ZE, Scheurer ME, Kamdar K, Wei Q, et al. A meta-analysis of MTHFR C677T and A1298C polymorphisms and risk of acute lymphoblastic leukemia in children. Pediatr Blood Cancer. 2012;58:513–8.

28. Zintzaras E, Doxani C, Rodopoulou P, Bakalos G, Ziogas DC, Ziakas P, et al. Variants of the MTHFR gene and susceptibility to acute lymphoblastic leukemia in children: a synthesis of genetic association studies. Cancer Epidemiol. 2012;36:169–76.

29. Amigou A, Rudant J, Orsi L, Goujon-Bellec S, Leverger G, Baruchel A, et al. Folic acid supplementation, MTHFR and MTRR polymorphisms, and the risk of childhood leukemia: the ESCALE study (SFCE). Cancer Causes Control. 2012;23:1265–77.

30. Azhar MR, Rahimi Z, Vaisi-Raygani A, Akramipour R, Madani H, Parsian A. Lack of association between MTHFR C677T and A1298C polymorphisms and risk of childhood acute lymphoblastic leukemia in the Kurdish population from Western Iran. Genet Test Mol Biomarkers. 2012;16:198–202.

31. Metayer C, Scelo G, Chokkalingam AP, Barcellos LF, Aldrich MC, Chang JS, et al. Genetic variants in the folate pathway and risk of childhood acute lymphoblastic leukemia. Cancer Causes Control. 2011;22:1243–58.

32. Nikbakht M, MalekZadeh K, Kumar Jha A, Askari M, Marwaha RK, Kaul D, et al. Polymorphisms of MTHFR and MTR genes are not related to susceptibility to childhood ALL in North India. Exp Oncol. 2012;34:43–8.

33. Krajinovic M, Lamothe S, Labuda D, Lemieux-Blanchard E, Theoret Y, Moghrabi A, et al. Role of MTHFR genetic polymorphisms in the susceptibility to childhood acute lymphoblastic leukemia. Blood. 2004;103:252–7.

34. Pereira AC, Schettert IT, Morandini Filho AA, Guerra-Shinohara EM, Krieger JE. Methylenetetrahydrofolate reductase (MTHFR) c677t gene variant modulates the homocysteine folate correlation in a mild folate-deficient population. Clin Chim Acta. 2004;340:99–105.

35. Milne E, de Klerk NH, van Bockxmeer F, Kees UR, Thompson JR, Baker D, et al. Is there a folate-related gene–environment interaction in the etiology of childhood acute lymphoblastic leukemia? Int J Cancer. 2006;119:229–32.

36. Thirumaran RK, Gast A, Flohr T, Burwinkel B, Bartram C, Hemminki K, et al. MTHFR genetic polymorphisms and susceptibility to childhood acute lymphoblastic leukemia. Blood. 2005;106:2590–1; author reply 1-2.

37. Chan JY, Ugrasena DG, Lum DW, Lu Y, Yeoh AE. Xenobiotic and folate pathway gene polymorphisms and risk of childhood acute lymphoblastic leukaemia in Javanese children. Hematol Oncol. 2011;29:116–23.

38. Zhao W, Yue LJ, Chen XW. Association of single nucleotide polymorphism of reduced folate carrier gene with susceptibility to acute leukemia. Zhonghua Yi Xue Yi Chuan Xue Za Zhi. 2011;28:446–9.

39. Yeoh AE, Lu Y, Chan JY, Chan YH, Ariffin H, Kham SK, et al. Genetic susceptibility to childhood acute lymphoblastic leukemia shows protection in Malay boys: results from the Malaysia-Singapore ALL Study Group. Leuk Res. 2010;34:276–83.

40. Lightfoot TJ, Johnston WT, Painter D, Simpson J, Roman E, Skibola CF, et al. Genetic variation in the folate metabolic pathway and risk of childhood leukemia. Blood. 2010;115:3923–9.

41. Gast A, Bermejo JL, Flohr T, Stanulla M, Burwinkel B, Schrappe M, et al. Folate metabolic gene polymorphisms and childhood acute lymphoblastic leukemia: a case–control study. Leukemia. 2007;21:320–5.

42. Jamroziak K, Robak T. Do polymorphisms in ABC transporter genes influence risk of childhood acute lymphoblastic leukemia? Leuk Res. 2008;32:1173–5.

43. Kimchi-Sarfaty C, Oh JM, Kim IW, Sauna ZE, Calcagno AM, Ambudkar SV, et al. A "silent" polymorphism in the MDR1 gene changes substrate specificity. Science. 2007;315:525–8.

44. Wang J, Wang B, Bi J, Li K, Di J. MDR1 gene C3435T polymorphism and cancer risk: a meta-analysis of 34 case–control studies. J Cancer Res Clin Oncol. 2012;138:979–89.

45. Zhuo W, Zhang L, Qiu Z, Zhu B, Chen Z. Does cytochrome P450 1A1 MspI polymorphism increase acute lymphoblastic leukemia risk? Evidence from 2013 cases and 2903 controls. Gene. 2012;510:14–21.

46. Swinney RM, Beuten J, Collier AB 3rd, Chen TT, Winick NJ, Pollock BH, et al. Polymorphisms in CYP1A1 and ethnic-specific susceptibility to acute lymphoblastic leukemia in children. Cancer Epidemiol Biomarkers Prev. 2011;20:1537–42.

47. Yamaguti GG, Lourenco GJ, Silveira VS, Tone LG, Lopes LF, Lima CS. Increased risk for acute lymphoblastic leukemia in children with cytochrome P450A1 (CYP1A1)- and NAD(P)H:quinone oxidoreductase 1 (NQO1)-inherited gene variants. Acta Haematol. 2010;124:182–4.

48. Lee KM, Ward MH, Han S, Ahn HS, Kang HJ, Choi HS, et al. Paternal smoking, genetic polymorphisms in CYP1A1 and childhood leukemia risk. Leuk Res. 2009;33:250–8.

49. Bonaventure A, Goujon-Bellec S, Rudant J, Orsi L, Leverger G, Baruchel A, et al. Maternal smoking during pregnancy, genetic polymorphisms of metabolic enzymes, and childhood acute leukemia: the ESCALE study (SFCE). Cancer Causes Control. 2012;23:329–45.

50. Silveira Vda S, Canalle R, Scrideli CA, Queiroz RG, Tone LG. Role of the CYP2D6, EPHX1, MPO, and NQO1 genes in the susceptibility to acute lymphoblastic leukemia in Brazilian children. Environ Mol Mutagen. 2010;51:48–56.

51. Borst L, Wallerek S, Dalhoff K, Rasmussen KK, Wesenberg F, Wehner PS, et al. The impact of CYP3A5*3 on risk and prognosis in childhood acute lymphoblastic leukemia. Eur J Haematol. 2011;86:477–83.

52. Pakakasama S, Mukda E, Sasanakul W, Kadegasem P, Udomsubpayakul U, Thithapandha A, et al. Polymorphisms of drug-metabolizing enzymes and risk of childhood acute lymphoblastic leukemia. Am J Hematol. 2005;79:202–5.

53. Silveira VS, Canalle R, Scrideli CA, Queiroz RG, Lopes LF, Tone LG. CYP3A5 and NAT2 gene polymorphisms: role in childhood acute lymphoblastic leukemia risk and treatment outcome. Mol Cell Biochem. 2012;364:217–23.

54. Clavel J, Bellec S, Rebouissou S, Menegaux F, Feunteun J, Bonaiti-Pellie C, et al. Childhood leukaemia, polymorphisms of metabolism enzyme genes, and interactions with maternal tobacco, coffee and alcohol consumption during pregnancy. Eur J Cancer Prev. 2005;14:531–40.

55. Tumer TB, Sahin G, Arinc E. Association between polymorphisms of EPHX1 and XRCC1 genes and the risk of childhood acute lymphoblastic leukemia. Arch Toxicol. 2012;86:431–9.

56. Rimando MG, Chua MN, Yuson E, de Castro-Bernas G, Okamoto T. Prevalence of GSTT1, GSTM1 and NQO1 (609C>T) in Filipino children with ALL (acute lymphoblastic leukaemia). Biosci Rep. 2008;28:117–24.

57. Suneetha KJ, Nancy KN, Rajalekshmy KR, Sagar TG, Rajkumar T. Role of GSTM1 (Present/Null) and GSTP1 (Ile105Val) polymorphisms in susceptibility to acute lymphoblastic leukemia among the South Indian population. Asian Pac J Cancer Prev. 2008;9:733–6.

58. Krajinovic M, Richer C, Sinnett H, Labuda D, Sinnett D. Genetic polymorphisms of N-acetyltransferases 1 and 2 and gene–gene interaction in the susceptibility to childhood acute lymphoblastic leukemia. Cancer Epidemiol Biomarkers Prev. 2000;9:557–62.

59. Zanrosso CW, Emerenciano M, Faro A, Goncalves BA, Mansur MB, Pombo-de-Oliveira MS. Genetic variability in N-acetyltransferase 2 gene determines susceptibility to childhood lymphoid or myeloid leukemia in Brazil. Leuk Lymphoma. 2012;53:323–7.

60. de Aguiar Goncalves BA, Vasconcelos GM, Thuler LC, Andrade C, Faro A, Pombo-de-Oliveira MS. NQO1 rs1800566 (C609T), PON1 rs662 (Q192R), and PON1 rs854560 (L55M) polymorphisms segregate the risk of childhood acute leukemias according to age range distribution. Cancer Causes Control. 2012;23:1811–9.

61. Guha N, Chang JS, Chokkalingam AP, Wiemels JL, Smith MT, Buffler PA. NQO1 polymorphisms and de novo childhood leukemia: a HuGE review and meta-analysis. Am J Epidemiol. 2008;168:1221–32.

62. Gillert E, Leis T, Repp R, Reichel M, Hosch A, Breitenlohner I, et al. A DNA damage repair mechanism is involved in the origin of chromosomal translocations t(4;11) in primary leukemic cells. Oncogene. 1999;18:4663–71.

63. Lindahl T, Wood RD. Quality control by DNA repair. Science. 1999;286:1897–905.

64. Wang L, Yin F, Xu X, Hu X, Zhao D. X-ray repair cross-complementing group 1 (XRCC1) genetic polymorphisms and risk of childhood acute lymphoblastic leukemia: a meta-analysis. PLoS One. 2012;7:e34897.

65. Wang R, Hu X, Zhou Y, Feng Q, Su L, Long J, et al. XRCC1 Arg399Gln and Arg194Trp polymorphisms in childhood acute lymphoblastic leukemia risk: a meta-analysis. Leuk Lymphoma. 2012;PMID: 22712837.

66. Canalle R, Silveira VS, Scrideli CA, Queiroz RG, Lopes LF, Tone LG. Impact of thymidylate synthase promoter and DNA repair gene polymorphisms on susceptibility to childhood acute lymphoblastic leukemia. Leuk Lymphoma. 2011;52:1118–26.

67. McNally RJ, Eden TO. An infectious aetiology for childhood acute leukaemia: a review of the evidence. Br J Haematol. 2004;127:243–63.

68. Greaves M. Infection, immune responses and the aetiology of childhood leukaemia. Nat Rev Cancer. 2006;6(3):193–203.

69. Shiina T, Inoko H, Kulski JK. An update of the HLA genomic region, locus information and disease associations: 2004. Tissue Antigens. 2004;64:631–49.

70. Dorak MT, Oguz FS, Yalman N, Diler AS, Kalayoglu S, Anak S, et al. A male-specific increase in the HLA-DRB4 (DR53) frequency in high-risk and relapsed childhood ALL. Leuk Res. 2002;26:651–6.

71. Dorak MT, Lawson T, Machulla HK, Darke C, Mills KI, Burnett AK. Unravelling an HLA-DR association in childhood acute lymphoblastic leukemia. Blood. 1999;94:694–700.

72. Yari F, Sobhani M, Sabaghi F, Zaman-Vaziri M, Bagheri N, Talebian A. Frequencies of HLA-DRB1 in Iranian normal population and in patients with acute lymphoblastic leukemia. Arch Med Res. 2008;39:205–8.

73. Klitz W, Gragert L, Trachtenberg E. Spectrum of HLA associations: the case of medically refractory pediatric acute lymphoblastic leukemia. Immunogenetics. 2012;64:409–19.

74. Morrison BA, Ucisik-Akkaya E, Flores H, Alaez C, Gorodezky C, Dorak MT. Multiple sclerosis risk markers in HLA-DRA, HLA-C, and IFNG genes are associated with sex-specific childhood leukemia risk. Autoimmunity. 2010;43:690–7.

75. Wang XJ, Ai XF, Sun HY, Wang F, Xu SC, Gui FM, et al. Relation of HLA-DRB1*15 with pathogenesis in 162 childhood cases of acute lymphoblastic leukemia. Zhongguo Shi Yan Xue Ye Xue Za Zhi. 2009;17:1507–10.

76. Taylor GM, Dearden S, Ravetto P, Ayres M, Watson P, Hussain A, et al. Genetic susceptibility to childhood common acute lymphoblastic leukaemia is associated with polymorphic peptide-binding pocket profiles in HLA-DPB1*0201. Hum Mol Genet. 2002;11:1585–97.

77. Taylor GM, Robinson MD, Binchy A. Preliminary evidence of an association between HLA-DPB1*0201 and childhood acute lymphobastic leukemia supports an infectious aetiology. Leukemia. 1995;9:440–3.

78. Urayama KY, Chokkalingam AP, Metayer C, Ma X, Selvin S, Barcellos LF, et al. HLA-DP genetic variation, proxies for early life immune modulation and childhood acute lymphoblastic leukemia risk. Blood. 2012;120:3039–47.

79. Papaemmanuil E, Hosking FJ, Vijayakrishnan J, Price A, Olver B, Sheridan E, et al. Loci on 7p12.2, 10q21.2 and 14q11.2 are associated with risk of childhood acute lymphoblastic leukemia. Nat Genet. 2009;41:1006–10.

80. Trevino LR, Yang W, French D, Hunger SP, Carroll WL, Devidas M, et al. Germline genomic variants associated with childhood acute lymphoblastic leukemia. Nat Genet. 2009;41:1001–5.

81. Pastorczak A, Gorniak P, Sherborne A, Hosking F, Trelinska J, Lejman M, et al. Role of 657del5 NBN mutation and 7p12.2 (IKZF1), 9p21 (CDKN2A), 10q21.2 (ARID5B) and 14q11.2 (CEBPE) variation and risk of childhood ALL in the Polish population. Leuk Res. 2011;35:1534–6.

82. Prasad RB, Hosking FJ, Vijayakrishnan J, Papaemmanuil E, Koehler R, Greaves M, 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. 2010;115:1765–7.

83. Vijayakrishnan J, Sherborne AL, Sawangpanich R, Hongeng S, Houlston RS, Pakasama S. Variation at 7p12.2 and 10q21.2 influences childhood acute lymphoblastic leukemia risk in the Thai population and may contribute to racial differences in leukemia incidence. Leuk Lymphoma. 2010;51:1870–4.

84. Ross JA, Linabery AM, Blommer CN, Langer EK, Spector LG, Hilden JM, et al. Genetic variants modify susceptibility to leukemia in infants: a Children's Oncology Group report. Pediatr Blood Cancer. 2012 doi:10.1002/pbc.24131.

85. Healy J, Richer C, Bourgey M, Kritikou EA, Sinnett D. Replication analysis confirms the association of ARID5B with childhood B-cell acute lymphoblastic leukemia. Haematologica. 2010;95:1608–11.

86. Yang W, Trevino LR, Yang JJ, Scheet P, Pui CH, Evans WE, et al. ARID5B SNP rs10821936 is associated with risk of childhood acute lymphoblastic leukemia in blacks and contributes to racial differences in leukemia incidence. Leukemia. 2010;24:894–6.

87. Sherborne AL, Hosking FJ, Prasad RB, Kumar R, Koehler R, Vijayakrishnan J, et al. Variation in CDKN2A at 9p21.3 influences childhood acute lymphoblastic leukemia risk. Nat Genet. 2010;42:492–4.

88. Ellinghaus E, Stanulla M, Richter G, Ellinghaus D, te Kronnie G, Cario G, et al. Identification of germline susceptibility loci in ETV6-RUNX1-rearranged childhood acute lymphoblastic leukemia. Leukemia. 2012;26:902–9.

89. Orsi L, Rudant J, Bonaventure A, Goujon-Bellec S, Corda E, Evans TJ, et al. Genetic polymorphisms and childhood acute lymphoblastic leukemia: GWAS of the ESCALE study (SFCE). Leukemia. 2012 doi:10.1038/leu.2012.148.

90. Han S, Lee KM, Park SK, Lee JE, Ahn HS, Shin HY, et al. Genome-wide association study of childhood acute lymphoblastic leukemia in Korea. Leuk Res. 2010;34:1271–4.

91. Georgopoulos K, Bigby M, Wang JH, Molnar A, Wu P, Winandy S, et al. The Ikaros gene is required for the development of all lymphoid lineages. Cell. 1994;79:143–56.

92. Mullighan CG, Miller CB, Radtke I, Phillips LA, Dalton J, Ma J, et al. BCR-ABL1 lymphoblastic leukaemia is characterized by the deletion of Ikaros. Nature. 2008;453:110–4.

93. Mullighan CG, Su X, Zhang J, Radtke I, Phillips LA, Miller CB, et al. Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia. N Engl J Med. 2009;360:470–80.

94. Lahoud MH, Ristevski S, Venter DJ, Jermiin LS, Bertoncello I, Zavarsek S, et al. Gene targeting of Desrt, a novel ARID class DNA-binding protein, causes growth retardation and abnormal development of reproductive organs. Genome Res. 2001;11:1327–34.

95. Chang LW, Payton JE, Yuan W, Ley TJ, Nagarajan R, Stormo GD. Computational identification of the normal and perturbed genetic networks involved in myeloid differentiation and acute promyelocytic leukemia. Genome Biol. 2008;9:R38.

96. Akasaka T, Balasas T, Russell LJ, Sugimoto KJ, Majid A, Walewska R, et al. Five members of the CEBP transcription factor family are targeted by recurrent IGH translocations in B-cell precursor acute lymphoblastic leukemia (BCP-ALL). Blood. 2007;109:3451–61.

97. Mullighan CG, Downing JR. Genome-wide profiling of genetic alterations in acute lymphoblastic leukemia: recent insights and future directions. Leukemia. 2009;23:1209–18.

98. Zhu JW, Brdicka T, Katsumoto TR, Lin J, Weiss A. Structurally distinct phosphatases CD45 and CD148 both regulate B cell and macrophage immunoreceptor signaling. Immunity. 2008;28:183–96.

99. Almalte Z, Samarani S, Iannello A, Debbeche O, Duval M, Infante-Rivard C, et al. Novel associations between activating killer-cell immunoglobulin-like receptor genes and childhood leukemia. Blood. 2011;118:1323–8.

100. Hosking FJ, Leslie S, Dilthey A, Moutsianas L, Wang Y, Dobbins SE, et al. MHC variation and risk of childhood B-cell precursor acute lymphoblastic leukemia. Blood. 2011;117:1633–40.

101. The Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007;447:661–78.

102. Lee SH, Wray NR, Goddard ME, Visscher PM. Estimating missing heritability for disease from genome-wide association studies. Am J Hum Genet. 2011;88:294–305.

103. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, et al. Common SNPs explain a large proportion of the heritability for human height. Nat Genet. 2010;42:565–9.

104. Sherborne AL, Hemminki K, Kumar R, Bartram CR, Stanulla M, Schrappe M, et al. Rationale for an international consortium to study inherited genetic susceptibility to childhood acute lymphoblastic leukemia. Haematologica. 2011;96:1049–54.

105. Chanock SJ, Manolio T, Boehnke M, Boerwinkle E, Hunter DJ, Thomas G, et al. Replicating genotype–phenotype associations. Nature. 2007;447:655–60.

106. Mukherjee B, Ahn J, Gruber SB, Chatterjee N. Testing gene–environment interaction in large-scale case–control association studies: possible choices and comparisons. Am J Epidemiol. 2012;175:177–90.

107. Freedman ML, Monteiro AN, Gayther SA, Coetzee GA, Risch A, Plass C, et al. Principles for the post-GWAS functional characterization of cancer risk loci. Nat Genet. 2011;43:513–8.

108. 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012;491:56–65.

109. Su Z, Ning B, Fang H, Hong H, Perkins R, Tong W, et al. Next-generation sequencing and its applications in molecular diagnostics. Expert Rev Mol Diagn. 2011;11:333–43.

110. Cheung VG, Spielman RS. Genetics of human gene expression: mapping DNA variants that influence gene expression. Nat Rev Genet. 2009;10:595–604.

111. Yang TP, Beazley C, Montgomery SB, Dimas AS, Gutierrez-Arcelus M, Stranger BE, et al. Genevar: a database and Java application for the analysis and visualization of SNP-gene associations in eQTL studies. Bioinformatics. 2010;26:2474–6.

112. Hogan LE, Meyer JA, Yang J, Wang J, Wong N, Yang W, et al. Integrated genomic analysis of relapsed childhood acute lymphoblastic leukemia reveals therapeutic strategies. Blood. 2011;118:5218–26.

113. Rahimi Z, Ahmadian Z, Akramipour R, Vaisi-Raygani A, Parsian A (2012) Thymidylate synthase and methionine synthase polymorphisms are not associated with susceptibility to childhood acute lymphoblastic leukemia in Kurdish population from Western Iran. Mol Biol Rep 39(3):2195–2200