Risk of Childhood Acute Lymphoblastic Leukemia: Identification of Inherited Susceptibility

11

Amy L. Sherborne and Richard S. Houlston

Contents

Introduction 105

Models of Inherited Susceptibility and Candidate Gene Studies 106

Genome-Wide Association Studies 107

Contribution of GWAS Findings to the Understanding of ALL Development 107

Identifying Functional Variants and the Heritability of ALL 109

Incorporating Non-genetic Risk Factors into Risk Models 109

Conclusions and Future Challenges 110

References 110

Abstract

Acute lymphoblastic leukemia (ALL) is the major paediatric cancer in developed countries. It has long been speculated that common genetic variation influences the development of this haematological malignancy, however until recently evidence for this hypothesis has been lacking. The advent of genome-wide association studies (GWAS) has allowed the search for this class of susceptibility allele to be conducted on a genome-wide basis. Such analyses have identified novel disease genes for ALL and underscore the importance of polymorphic variation in B-cell development genes as determinants of leukemia risk. Furthermore these data indicate that a significant difference in the risk of an individual developing ALL can be attributed to heritable genetic factors.

Introduction

Acute lymphoblastic leukemia (ALL) is the major paediatric cancer in economically developed countries, with precursor B-cell (BCP-ALL; MIM 613065) accounting for approximately 70% of all childhood ALL (Stiller and Parkin 1996). In contrast to many other haematological malignancies, evidence for a familial risk to ALL is weak. Data from the Swedish family-cancer database does, however, lend some support to an excess risk in relatives of patients (Hemminki and Jiang 2002). Although rare (<5% of ALL), direct evi-

A.L. Sherborne (✉) • R.S. Houlston
Division of Genetics and Epidemiology,
Institute of Cancer Research, 15 Cotswold Road,
Sutton, Surrey, SM2 5NG, UK
e-mail: Amy.Sherborne@icr.ac.uk

M.A. Hayat (ed.), *Pediatric Cancer, Volume 4: Diagnosis, Therapy, and Prognosis*, Pediatric Cancer 4,
DOI 10.1007/978-94-007-6591-7_11, © Springer Science+Business Media Dordrecht 2013

dence for inherited genetic susceptibility is provided by the high risk of ALL associated with Bloom's syndrome, neurofibromatosis, ataxia telangiectasia and constitutional trisomy 21. While evidence linking an environmental exposure to risk of childhood ALL has largely been inconsistent (as reviewed in Belson et al. 2007), epidemiological data for an infectious etiology is persuasive, albeit indirect (Greaves 2006). Implicit in a model of ALL having an infectious etiological basis is that ALL is likely to represent a rare sequelae of infection with germline variation influencing host response.

Models of Inherited Susceptibility and Candidate Gene Studies

Modest familial relative risks are compatible with a wide range of genetic models of inheritance. However, the absence of families segregating ALL argues against the role of high penetrance susceptibility to the disease being the norm. Hence it is probable that the risk of ALL attributable to genetic variation is due to the co-inheritance of multiple low-risk variants, as outlined in the polygenic model (Fig. 11.1). Under this model variants conferring relative risks of 1.1–1.5 could make an important contribution to the overall inherited risk. Although such alleles have small effects individually, they could contribute significantly to disease susceptibility in the general population. Furthermore, by acting in concert they have the capacity to generate a high risk of ALL in a subset of the population.

While it has long been speculated that common polymorphic variation contributes to the susceptibility of ALL, evidence for common low risk alleles has only just emerged. The search for risk loci for ALL has until recently centred on association studies of candidate genes, where the frequencies of polymorphic variants, single nucleotide polymorphisms (SNPs), are compared in cases and controls. Most of these studies have

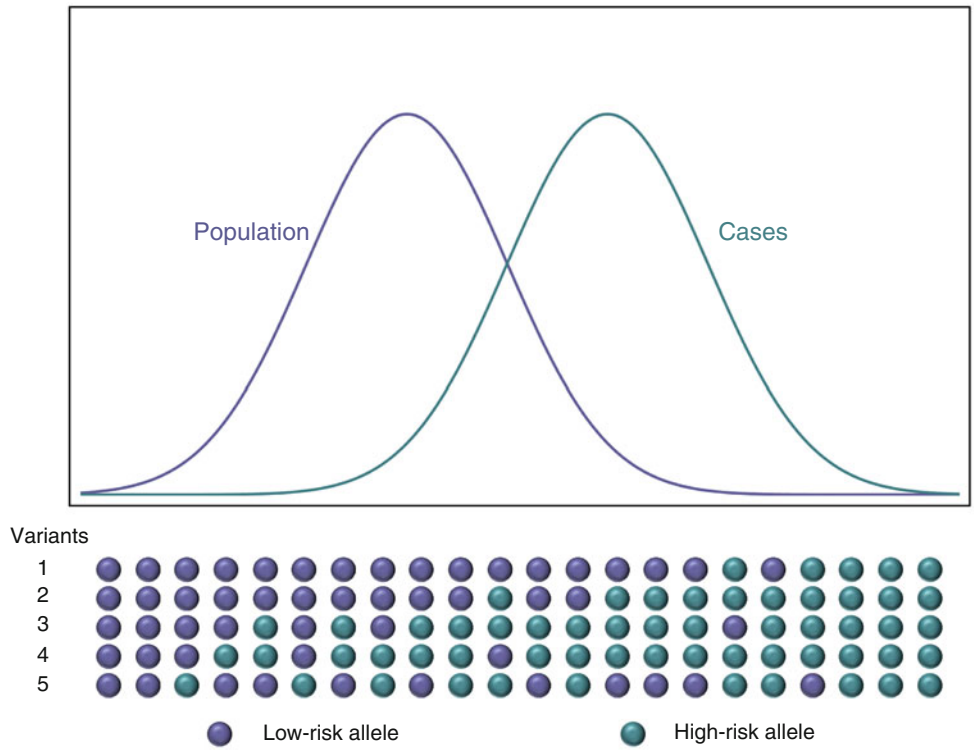

Fig. 11.1 The polygenic model of susceptibility to cancer

evaluated only a restricted number of polymorphisms, such as those influencing methylation or carcinogen metabolism. Although numerous associations have been proposed from such candidate gene analyses conducted over the past 20 years, no definitive susceptibility alleles have been unequivocally identified (Sinnett et al. 2000; Ye and Song 2005; Bolufer et al. 2006; Pereira et al. 2006; Guha et al. 2008). As with many other diseases, positive associations have been reported for various polymorphisms but few of the initial positive results have been replicated in subsequent studies (Vijayakrishnan and Houlston 2010). The inherent statistical uncertainty of studies involving just a few hundred cases and controls seriously limits the power available to reliably identify genetic determinants conferring modest but potentially important risks. Furthermore, without a clear understanding of the biology of predisposition the definition of what truly represents a candidate gene is inherently problematic, making an unbiased approach to loci selection highly desirable.

Genome-Wide Association Studies

Following completion of the Human Genome Project, more than 20 million SNPs have been catalogued in addition to smaller numbers of insertion/deletion and copy number variations. The high resolution LD maps and comprehensive sets of tagging SNPs (tagSNPs) available through the HapMap, coupled with the development of highly efficient analytical platforms, have allowed genome wide association studies (GWAS) to be conducted efficiently and cost effectively. This approach is unbiased and does not depend on prior knowledge of the function or involvement of any gene in disease causation. Furthermore, the strategy offers the prospect of identifying important variants in previously unstudied genes and non-coding regions of the genome.

Recent GWASs of ALL have vindicated the hypothesis of common susceptibility to ALL, identifying SNPs at four novel risk loci7p12.2 (*IKZF1*, rs4132601), 9p21.3 (*CDKN2A*, rs3731217) 10q21.2 (*ARID5B*, rs7089424) and 14q11.2 (*CEBPE*, rs2239633) (Table 11.1) (Papaemmanuil et al. 2009; Trevino et al. 2009; Sherborne et al. 2010; Vijayakrishnan and Houlston 2010). Intriguingly, excluding *CDKN2A*, none of the genes implicated by these GWAS scans have previously been evaluated in targeted association studies, emphasizing that the candidate gene approach was severely limited by inadequate knowledge of tumor biology.

Contribution of GWAS Findings to the Understanding of ALL Development

The SNP genotyped in GWAS are not generally candidates for causality, and enumeration of the causal variant at a specific locus can pose a significant challenge. While fine-mapping and resequencing is required to identify functional variant(s) the associations identified for ALL implicate a number of genes in tumor aetiology.

The strongest association signal for ALL was attained at 7p12.2 with rs4132601, which maps to the 3′ region of the Ikaros family zinc finger 1 (*IKZF1*) gene. Ikaros proteins are master regulators of lymphocyte development (Fig. 11.2) and differentiation, and play a pivotal role in CD4 versus CD8 T-cell lineage commitment decisions (Harker et al. 2002). In homozygous mutant mice deleted for the N-terminal zinc finger DNA binding domain of *IKZF1*, loss of expression leads to arrest of lymphocyte development at its earliest

Table 11.1 ALL susceptibility loci identified through genome-wide association studies

SNP	Chr	MAF	Gene	P_{trend}	OR (95% CI)
rs4132601	7p12.2	0.28	*IKZF1*	1.20×10^{-19}	1.65 (1.58–1.81)
rs3731217	9p21.3	0.14	*CDKN2A*	3.01×10^{-11}	0.71 (0.64–0.78)
rs7089424	10q21.2	0.32	*ARIDB5*	6.69×10^{-19}	1.65 (1.54–1.76)
rs2236933	14q11.2	0.48	*CEBPE*	2.88×10^{-7}	1.34 (1.22–1.45)

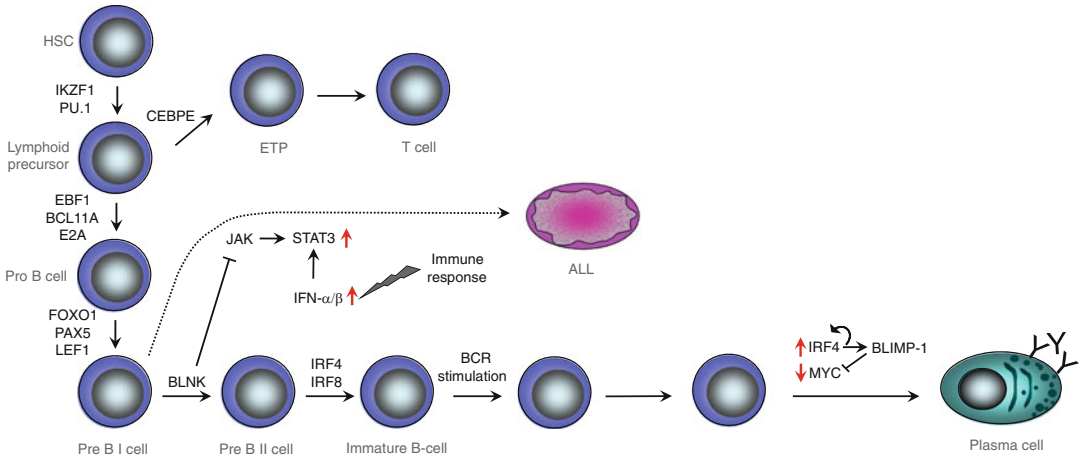

Fig. 11.2 Genes regulating lymphoid development

recognizable stage followed by rapid development of leukemia (Georgopoulos et al. 1994).

The region of association defining the 9p21.3 association encompasses the *CDKN2A* and *CDKN2B* tumor-suppressor genes and the non-coding antisense RNA encoded by *CDKN2BAS*. *CDKN2A* encodes both p16 (INK4A), a negative regulator of cyclin-dependant kinases, and p14 (ARF1), an activator of p53. *CDKN2A* and *CDKN2B* are frequently inactivated in multiple hematological malignancies. Moreover, mono- or biallelic deletion of *CDKN2A* is one of the most frequent genetic events in both childhood BCP and T-ALL (Mullighan and Downing 2009). Perhaps not surprisingly the association between 9p21.3 risk genotype and ALL is generic and not confined to a specific form of ALL.

The association at 10q21.2 implicates the AT rich interactive domain 5B (*ARID5B*) gene in the etiology of ALL. While *ARID5B* has not been extensively studied, evidence for *ARID5B* having a role in defining B-cell lineage is supported by data from homozygous knockout mice, which display decreased bone marrow cellularity and reduced numbers of B-cell progenitors (Lahoud et al. 2001).

The 14q11.2 association with ALL annotates the gene encoding CCAAT/enhancer-binding protein, epsilon (*CEBPE*). CEBP is a suppressor of myeloid leukemogenesis. *CEBPE*, along with other CEBP family members, is occasionally targeted by recurrent IGH translocations in BCP-ALL suggesting opposing functions of CEBP dysregulation in myeloid and lymphoid leukemogenesis and a role in susceptibility to ALL (Akasaka et al. 2007).

Given the biological heterogeneity of ALL, variants are likely to have differential effects on ALL risk depending on cell lineage and phenotype. This is well illustrated by the primary impact of variation defined by the 7p12.2, 9p21.3, 10q21.2 and 14q11.2 risk variants for B-lineage leukemia. Furthermore, subtype analysis of B-precursor ALL provides strong evidence that variation at 10q21.2-*ARID5B* is highly associated with the risk of developing hyperdiploid ALL (Sherborne et al. 2010). Given that the frequency of many ALL subgroups is small, identifying differential effects will only be realistically possible through multi-center pooled analyses.

A role for specific human leukocyte antigen (HLA) variants in the etiology of ALL has been extensively studied over the last 30 years, but no unambiguous association has been identified. It has recently been demonstrated that it is possible to use SNP variation within the 6p21 region to accurately predict alleles at key class I (*HLA-A*, *HLA-B*, and *HLA-C*) and class II (*HLA-DRB1*, *HLA-DQA1*, and *HLA-DQB1*) loci with better than 90% accuracy (Leslie et al.

2008). Information from GWAS has also allowed the role of variation in specific loci such as the MHC to be comprehensively examined. Through such analyses it has been possible to demonstrate that major histocompatibility complex-defined variation in immune-mediated response is unlikely to be a major risk factor for B-cell precursor-ALL (Hosking et al. 2010).

Identifying Functional Variants and the Heritability of ALL

Validated tagSNPs are highly unlikely to directly impact on ALL risk. Identifying a functional variant from a tagSNP that is statistically associated with disease is challenging. Although blocks of LD allow the efficient survey of the genome, they hamper fine mapping of the disease-associated region. Different ethnic groups are likely to have different LD block patterns and they can, therefore, be used to refine the location of a disease susceptibility locus prior to fine mapping genotyping and functional analyses.

While the risks of ALL associated with the identified SNPs are modest, with relative risks of 1.2–1.7 per allele as predicted by the polygenic model, their contribution to ALL incidence is high as the alleles are common within the population. Moreover, the risk of ALL increases with increasing numbers of variant alleles carried by an individual. It is also likely that known loci may carry additional, as of yet unidentified, risk variants, potentially including low-frequency variants with larger influences on disease risk.

The testing of SNPs individually for an association in GWAS necessitates the imposition of a very stringent *P*-value to address the issue of multiple testing. While this reduces false positives, real associations may be missed and therefore any estimate of the total heritability will be negatively biased. By considering all typed SNPs simultaneously it has been calculated that 24% of the total variation in ALL risk can be ascribed to common genetic variation (Enciso-Mora et al. 2012). These findings suggest common variation rather than a restricted number of associations, influence ALL and provide further support for a polygenic basis for susceptibility to the disease. It is, therefore, likely that additional common low risk variants remain to be discovered and should be eminently harvestable in new larger GWAS or through further pooling of additional existing datasets. How much of the unaccounted heritable risk is truly embodied in a long tail of association is currently unknown but will impact on our ability to fully understand the genetic and ultimately biological basis of ALL predisposition.

Incorporating Non-genetic Risk Factors into Risk Models

The risk of developing ALL, like many other cancers, will undoubtedly be determined by complex interactions between genetic, environmental factors and chance. Epidemiological studies have so far provided indirect evidence that ALL may have an infective basis although no specific infectious agent has been implicated (MacMahon 1992; Greaves and Alexander 1993; Kinlen 1995). There is also consistent data supporting birth weight as a risk factor for ALL possibly operating through association with high IGF2 levels and the latter's impact on stem/progenitor cells (Robison et al. 1987). There is little robust evidence linking either pre- or post-natal environmental exposures to risk of childhood ALL, with ionizing radiation being the one notable exception (Mahoney et al. 2004).

Ethnic differences in the risk of ALL are well recognized. Thus, in assessing the interplay between inherited and non-genetic risk factors, analyses using different population cohorts with different incidence rates are likely to be highly informative. This is supported by recent studies of ALL in a Thai population (Vijayakrishnan et al. 2010) and in a black population (Yang et al. 2010) suggesting that 7p12.2 and 10q21.2 variation may contribute to racial differences in ALL risk.

Type 1 diabetes (T1D) is an autoimmune disease for which infectious triggers of disease onset have been sought with increasing evidence pointing

to enteroviruses. Co-morbidity between ALL and type 1 diabetes has been reported from cohort analyses performed in Sweden (Shu et al. 2010), which is especially interesting as variation in *IKZF1* appears to be a determinant of risk for both diseases, albeit reciprocally. Although such observations are intriguing, the robust identification of interactions between genetic variants and environmental risk factors will be contingent on very large datasets, realistically something which can only be achieved through multi-center collaboration.

Conclusions and Future Challenges

Recent studies have provided the first unambiguous evidence that common genetic variation contributes to the risk of developing ALL and implicate genes involved in transcriptional regulation and differentiation of B-cell progenitors as the biological basis of predisposition to B-cell malignancy. Furthermore, their identification provides novel insight into disease causation of these two major haematological malignancies.

The power of recent GWAS to identify common alleles conferring risks of 1.5 or greater (such as the 7p12.2 variant) is high. Hence, there are unlikely to be many additional SNPs with similar effects for alleles with frequencies greater than 0.3 in populations of European ancestry. Tagging SNPs employed for GWAS capture on average approximately 80% of common SNPs in the European population, but only approximately 10% of SNPs with minor allele frequencies of 5–10% are tagged at this level, limiting power to detect this class of susceptibility allele. While coverage of the genome offered by current arrays is generally high, some chromosomal regions cannot be readily typed due to inadequate tagging or technological constraints. GWAS-based strategies are not configured optimally to identify low frequency variants with potentially stronger effects or identify recessively acting alleles. It is, therefore, highly likely that a large number of disease-causing variants remain to be discovered through next generation arrays and through throughput sequencing projects.

References

Akasaka T, Balasas T, Russell LJ, Sugimoto KJ, Majid A, Walewska R, Karran EL, Brown DG, Cain K, Harder L, Gesk S, Martin-Subero JI, Atherton MG, Bruggemann M, Calasanz MJ, Davies T, Haas OA, Hagemeijer A, Kempski H, Lessard M, Lillington DM, Moore S, Nguyen-Khac F, Radford-Weiss I, Schoch C, Struski S, Talley P, Welham MJ, Worley H, Strefford JC, Harrison CJ, Siebert R, Dyer MJ (2007) Five members of the CEBP transcription factor family are targeted by recurrent IGH translocations in B-cell precursor acute lymphoblastic leukemia (BCP-ALL). Blood 109:3451–3461

Belson M, Kingsley B, Holmes A (2007) Risk factors for acute leukemia in children: a review. Environ Health Perspect 115:138–145

Bolufer P, Barragan E, Collado M, Cervera J, Lopez JA, Sanz MA (2006) Influence of genetic polymorphisms on the risk of developing leukemia and on disease progression. Leuk Res 30:1471–1491

Enciso-Mora V, Hosking FJ, Sheridan E, Kinsey SE, Lightfoot T, Roman E, Irving JA, Tomlinson IP, Allan JM, Taylor M, Greaves M, Houlston RS (2012) Common genetic variation contributes significantly to the risk of childhood B-cell precursor acute lymphoblastic leukemia. Leukemia 26:2212–2215

Georgopoulos K, Bigby M, Wang JH, Molnar A, Wu P, Winandy S, Sharpe A (1994) The Ikaros gene is required for the development of all lymphoid lineages. Cell 79:143–156

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

Greaves MF, Alexander FE (1993) An infectious etiology for common acute lymphoblastic leukemia in childhood? Leukemia 7:349–360

Guha N, Chang JS, Chokkalingam AP, Wiemels JL, Smith MT, Buffler PA (2008) NQO1 polymorphisms and de novo childhood leukemia: a HuGE review and meta-analysis. Am J Epidemiol 168:1221–1232

Harker N, Naito T, Cortes M, Hostert A, Hirschberg S, Tolaini M, Roderick K, Georgopoulos K, Kioussis D (2002) The CD8alpha gene locus is regulated by the Ikaros family of proteins. Mol Cell 10:1403–1415

Hemminki K, Jiang Y (2002) Risks among siblings and twins for childhood acute lymphoid leukaemia: results from the Swedish Family-Cancer Database. Leukemia 16:297–298

Hosking FJ, Leslie S, Dilthey A, Moutsianas L, Wang Y, Dobbins SE, Papaemmanuil E, Sheridan E, Kinsey SE, Lightfoot T, Roman E, Irving JA, Allan JM, Taylor M, Greaves M, McVean G, Houlston RS (2010) MHC variation and risk of childhood B-cell precursor acute lymphoblastic leukaemia. Blood 117:1633–1640

Kinlen LJ (1995) Epidemiological evidence for an infective basis in childhood leukaemia. Br J Cancer 71:1–5

Lahoud MH, Ristevski S, Venter DJ, Jermiin LS, Bertoncello I, Zavarsek S, Hasthorpe S, Drago J, de

Kretser D, Hertzog PJ, Kola I (2001) Gene targeting of Desrt, a novel ARID class DNA-binding protein, causes growth retardation and abnormal development of reproductive organs. Genome Res 11:1327–1334

Leslie S, Donnelly P, McVean G (2008) A statistical method for predicting classical HLA alleles from SNP data. Am J Hum Genet 82:48–56

MacMahon B (1992) Is acute lymphoblastic leukemia in children virus-related? Am J Epidemiol 136:916–924

Mahoney MC, Moysich KB, McCarthy PL Jr, McDonald RC, Stepanenko VF, Day RW, Michalek AM (2004) The chernobyl childhood leukemia study: background & lessons learned. Environ Health 3:12

Mullighan CG, Downing JR (2009) Genome-wide profiling of genetic alterations in acute lymphoblastic leukemia: recent insights and future directions. Leukemia 23(7):1209–1218

Papaemmanuil E, Hosking FJ, Vijayakrishnan J, Price A, Olver B, Sheridan E, Kinsey SE, Lightfoot T, Roman E, Irving JAE, Allan JM, Tomlinson IP, Taylor M, Greaves M, Houlston RS (2009) Loci on 7p12.2, 10q21.2 and 14q11.2 are associated with risk of childhood acute lymphoblastic leukemia. Nat Genet 41:1006–1010

Pereira TV, Rudnicki M, Pereira AC, Pombo-De-Oliveira MS, Franco RF (2006) 5,10-methylenetetrahydrofolate reductase polymorphisms and acute lymphoblastic leukemia risk: a meta-analysis. Cancer Epidemiol Biomarkers Prev 15:1956–1963

Robison LL, Codd M, Gunderson P, Neglia JP, Smithson WA, King FL (1987) Birth weight as a risk factor for childhood acute lymphoblastic leukemia. Pediatr Hematol Oncol 4:63–72

Sherborne AL, Hosking FJ, Prasad RB, Kumar R, Koehler R, Vijayakrishnan J, Papaemmanuil E, Bartram CR, Stanulla M, Schrappe M, Gast A, Dobbins SE, Ma Y, Sheridan E, Taylor M, Kinsey SE, Lightfoot T, Roman E, Irving JA, Allan JM, Moorman AV, Harrison CJ, Tomlinson IP, Richards S, Zimmermann M, Szalai C, Semsei AF, Erdelyi DJ, Krajinovic M, Sinnett D, Healy J, Neira AG, Kawamata N, Ogawa S, Koeffler HP, Hemminki K, Greaves M, Houlston RS (2010) Variation in CDKN2A at 9p21.3 influences childhood acute lymphoblastic leukemia risk. Nat Genet 42:492–494

Shu X, Ji J, Li X, Sundquist J, Sundquist K, Hemminki K (2010) Cancer risk among patients hospitalized for Type 1 diabetes mellitus: a population-based cohort study in Sweden. Diabet Med 27:791–797

Sinnett D, Krajinovic M, Labuda D (2000) Genetic susceptibility to childhood acute lymphoblastic leukemia. Leuk Lymphoma 38:447–462

Stiller CA, Parkin DM (1996) Geographic and ethnic variations in the incidence of childhood cancer. Br Med Bull 52:682–703

Trevino LR, Yang W, French D, Hunger SP, Carroll WL, Devidas M, Willman C, Neale G, Downing J, Raimondi SC, Pui CH, Evans WE, Relling MV (2009) Germline genomic variants associated with childhood acute lymphoblastic leukemia. Nat Genet 41:1001–1005

Vijayakrishnan J, Houlston RS (2010) Candidate gene association studies and risk of childhood acute lymphoblastic leukemia: a systematic review and meta-analysis. Haematologica 95:1405–1414

Vijayakrishnan J, Sherborne AL, Sawangpanich R, Hongeng S, Houlston RS, Pakakasama S (2010) 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 51:1870–1874

Yang W, Trevino LR, Yang JJ, Scheet P, Pui CH, Evans WE, Relling MV (2010) ARID5B SNP rs10821936 is associated with risk of childhood acute lymphoblastic leukemia in blacks and contributes to racial differences in leukemia incidence. Leukemia 24:894–896

Ye Z, Song H (2005) Glutathione s-transferase polymorphisms (GSTM1, GSTP1 and GSTT1) and the risk of acute leukaemia: a systematic review and meta-analysis. Eur J Cancer 41:980–989