Rationale for an international consortium to study inherited genetic susceptibility to childhood acute lymphoblastic leukemia

Amy L. Sherborne,[1] Kari Hemminki,[2] Rajiv Kumar,[2] Claus R. Bartram,[3] Martin Stanulla,[4] Martin Schrappe,[4] Eleni Petridou,[5] Ágnes F. Semsei,[6] Csaba Szalai,[7,8] Daniel Sinnett,[9,10] Maja Krajinovic,[9,10] Jasmine Healy,[9] Marina Lanciotti,[11] Carlo Dufour,[11] Stefania Indaco,[11] Eman A El-Ghouroury,[12] Ruchchadol Sawangpanich,[13] Suradej Hongeng,[13] Samart Pakakasama,[13] Anna Gonzalez-Neira,[14] Evelia L. Ugarte,[15] Valeria P. Leal, [15] Juan P.M. Espinoza,[15] Azza M. Kamel,[16] Gamal T.A. Ebid,[16] Eman R. Radwan,[17] Serap Yalin,[18] Erdinc Yalin,[18] Mehmet Berkoz,[19] Jill Simpson,[20] Eve Roman,[20] Tracy Lightfoot,[20] Fay J. Hosking,[1] Jayaram Vijayakrishnan,[1] Mel Greaves,[21] and Richard S. Houlston[1]

[1]Section of Cancer Genetics, Institute of Cancer Research, Sutton, Surrey, UK; [2]Division of Molecular Genetic Epidemiology, German Cancer Research Center, Heidelberg, Germany; [3]Institute of Human Genetics, University of Heidelberg, Heidelberg, Germany; [4]Department of Pediatrics, ALL-BFM Clinical Trial Center, University of Kiel, Kiel, Germany; [5]Center for Research and Prevention of Injuries (CEREPI), Department of Hygiene and Epidemiology, School of Medicine, Athens University, Athens, Greece; [6]Department of Genetics, Cell and Immunobiology, Semmelweis University, Budapest, Hungary; [7]Inflammation Biology and Immunogenomics Research Group, Hungarian Academy of Sciences, Semmelweis University, Budapest, Hungary; [8]Heim Pal Pediatric Hospital, Budapest, Hungary; [9]Hematology-Oncology Service, Sainte-Justine UHC Research Center, University of Montreal, Montreal, Quebec, Canada; [10]Department of Pediatrics, Faculty of Medicine, University of Montreal, Quebec, Canada; [11]Haematology Unit, Department of Paediatric Haemato-Oncology, G. Gaslini Institute, Genova, Italy; [12]National Research Center, Cairo, Egypt; [13]Department of Pediatrics, Faculty of Medicine, Ramathibodi Hospital, Bangkok, Thailand; [14]Spanish National Cancer Research Center (CNIO), Madrid, Spain; [15]Unidad Academica de Ciencias de la Salud y Tecnologia, Universidad Autonoma de Tamaulipas, Tamaulipas, Mexico; [16]BMT Lab Unit, Clinical Pathology Department, NCI, Cairo University, Cairo, Egypt; [17]Clinical Pathology Department, Faculty of Medicine, Cairo University, Kasr El-Aini Hospital, Manial, Cairo, Egypt; [18]Department of Biochemistry, Faculty of Pharmacy, Mersin University, Turkey; [19]Department of Biochemistry, Faculty of Medicine, Cukurova University, Turkey; [20]Epidemiology and Genetics Unit, Department of Health Sciences, University of York, UK; and [21]Section of Haemato-Oncology, Institute of Cancer Research, Sutton, Surrey, UK

ABSTRACT

Acute lymphoblastic leukemia is the major pediatric cancer in developed countries. To date most association studies of acute lymphoblastic leukemia have been based on the candidate gene approach and have evaluated a restricted number of polymorphisms. Such studies have served to highlight difficulties in conducting statistically and methodologically rigorous investigations into acute lymphoblastic leukemia risk. Recent genome-wide association studies of childhood acute lymphoblastic leukemia have provided robust evidence that common variation at four genetic loci confers a modest increase in risk. The accumulated experience to date and relative lack of success of initial efforts to identify novel acute lymphoblastic leukemia predisposition loci emphasize the need for alternative study designs and methods. The International Childhood Acute Lymphoblastic Leukaemia Genetics Consortium includes 12 research groups in Europe, Asia, the Middle East and the Americas engaged in studying the genetics of acute lymphoblastic leukemia. The initial goal of this consortium is to identify and characterize low-penetrance susceptibility variants for acute lymphoblastic leukemia through association-based analyses. Efforts to develop genome-wide association studies of acute lymphoblastic leukemia, in terms of both sample size and single nucleotide polymorphism coverage, and to increase the number of single nucleotide polymorphisms taken forward to large-scale replication should lead to the identification of additional novel risk variants for acute lymphoblastic leukemia. Ethnic differences in the risk of acute lymphoblastic leukemia are well recognized and 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. Given that the frequency of many acute lymphoblastic leukemia subgroups is small, identifying differential effects will realistically only be possible through multi-center pooled analyses. Here, we review the rationale for identifying genetic risk variants for acute lymphoblastic leukemia and our proposed strategy for establishing the International Childhood Acute Lymphoblastic Leukaemia Genetics Consortium.

Key words: acute lymphoblastic leukemia, genetics, consortium, etiology.

Citation: Sherborne AL, Hemminki K, Kumar R, Bartram CR, Stanulla M, Schrappe M, Petridou E, Semsei AF, C Szalai, Sinnett D, Krajinovic M, Healy J, Lanciotti M, Dufour C, Indaco S, El-Ghouroury EA, Sawangpanich R, Hongeng S, Pakakasama S, Gonzalez-Neira A, Ugarte EL, Leal VP, Espinoza JMP, Kamel AM, Ebid GTA, Radwan ER, Yalin S, Yalin E, Berkoz M, Simpson J, Roman E, Lightfoot T, Hosking FJ, Vijayakrishnan J, Greaves M, and Houlston RS. Rationale for an international consortium to study inherited genetic susceptibility to childhood acute lymphoblastic leukemia. Haematologica 2011;96(7):1049-1054. doi:10.3324/haematol.2011.040121

Funding: in the UK work is supported by grants from the Leukaemia Research Fund and the Kay Kendall Leukaemia Fund. In Canada work is supported by grants from the Canadian Institutes of Health Research (CIHR), the Leukemia and Lymphoma Society of Canada, the Charles Bruneau Foundation, and the François-Karl Viau Research Chair in Pediatric Oncogenomics. In Turkey work is supported by the Research Foundation of Mersin University, Mersin, Turkey, grant N. BAP-SBE BB (MB) 2005-3 YL. In Italy the work is funded by the Compagnia di San Paolo, ERG s.p.a, SAAR Depositi Oleari Portuali, Genova.
Manuscript received on January 4, 2011. Revised version arrived on March 1, 2011. Manuscript accepted on March 18, 2011.
Correspondence: Richard Houlston, Section of Cancer Genetics, Institute of Cancer Research, Sutton, Surrey, UK. E-mail:richard.houlston@icr.ac.uk

Background

Acute lymphoblastic leukemia (ALL) is the major pediatric cancer in developed countries. B-cell precursor (BCP) ALL accounts for approximately 70% of childhood ALL and characteristically affects children between three and five years of age. While evidence linking most environmental exposures to risk of childhood ALL has largely been inconsistent, epidemiological data for an infectious etiology is persuasive, albeit indirect.[1] Nevertheless, causation pathways are likely to be multifactorial and it is probable that the risk of ALL from environmental exposure is influenced by genetic variation through the co-inheritance of multiple low-risk variants.

The "common-disease common-variant" model of cancer susceptibility implies that association analyses based on scans of polymorphic variants should be a powerful strategy for identifying common, low-penetrance susceptibility alleles. This assertion has been vindicated by recently conducted genome-wide association (GWA) studies of childhood ALL, which have provided robust evidence that common variation at four genetic loci confers a modest increase in risk.[2-4] As well as establishing a role for genetic susceptibility in the development of ALL, these data provide novel insights into disease causation; notably risk variants annotate genes involved in transcriptional regulation and differentiation of B-cell progenitors. While the risk of ALL associated with each of the variants individually is modest, they make a significant contribution to disease burden by virtue of their high frequencies in the population.

To facilitate the study of inherited predisposition to ALL we have established the IALLGC (International Childhood Acute Lymphoblastic Leukaemia Genetics Consortium) with the goal of identifying and characterizing additional genetic variants influencing ALL risk. Here, we review the rationale for studying genetic susceptibility to ALL and our proposed strategy for establishing the IALLGC.

Problems in conducting methodologically rigorous association studies

To date, most association studies of ALL have been based on the candidate gene approach and have evaluated a restricted number of polymorphisms, primarily in genes implicated in the metabolism of carcinogens (e.g. *CYP1A1, GSTM1, GSTT1, NQO1*), folate metabolism (e.g. *MTHFR, MTRR, SHMT1, RFC1*), protection of DNA from carcinogen-induced damage (e.g. *XRCC1, ERCC2, ATM*) and cell-cycle regulation (e.g. *CDKN2A, CDKN2B, CDKN1A, CDKN1B, TGFB1*).[5]

Reports from most candidate gene studies have been disappointing, with many positive associations initially being reported which subsequent studies fail to replicate. A number of studies have, however, reported case-control data on the same variant thereby allowing data to be pooled (Table 1). While *P* values from meta-analyses of studies provide support for the role of variants in *GSTM1, MTRR, SHMT1, RFC1, CYP1A1, CYP2E1, NQO1* and *XRCC1*, analyses should be interpreted with caution even if the issue of publication bias is ignored.[5] Use of false positive report probability value (FPRP)[6] which integrates the prior probability for the association and statistical power provides a method of assessing robustness of summary estimates derived from pooled analyses. While prior probabilities are partly subjective, the prior probability for variants in candidate genes is unlikely to be better than 1 in 1,000 (or 0.001).[7] Imposing a 'best case' prior probability of 0.001, and stipulating an odds ratio of 1.5 for associations, it is noteworthy that the likelihood that any of the variants so far reported from candidate gene analyses is associated with ALL risk is low (i.e. FPRP >0.2 is the value suggested to be appropriate for summary analyses [6]). Hence, despite much research prior to GWA studies, few if any definitive susceptibility alleles for ALL have been unequivocally identified through candidate gene association studies.[5]

The *in utero* origins of childhood ALL are well established.[1, 8-10] The age-incidence pattern of childhood ALL[11,12] and genetic studies of monozygotic twins[13] support the prenatal origin of the disease. Parental exposures to environmental hazards or use of parental medication have been reported to contribute to ALL risk.[14-18] Given the possible role of parental factors in disease susceptibility, focusing solely on the genotype of the child in ALL association studies may be insufficient to fully address disease etiology.[19]

The accumulated experience to date and relative lack of success of initial efforts to identify novel ALL predisposition loci emphasize the need for alternative study designs and methods. The current difficulties in conducting statistically and methodologically rigorous ALL association studies are summarized below.

1. The increase in ALL risk conferred by any common variant is almost certainly small (i.e. typically relative risk <1.5). The inherent statistical uncertainty of case-control studies involving a small number of cases and controls severely constrains study power to reliably identify genet-

Table 1. Polymorphisms reported to be statistically significant in meta-analyses. Adapted from Vijayakrishnan and Houlston.[5]

Polymorphism	Risk group	MAF/at risk frequency	OR (95% CI)	*P* value	N. studies	Power for OR=1.5	FPRP @ prior probability of 0.001
GSTM1-Null	Null *vs.* present	-	1.16 (1.04-1.30)	0.008	15	100 %	0.91
MTRR A66G	AG+GG *vs.* AA	0.23-0.45	0.73 (0.59-0.91)	0.005	3	96%	0.87
SHMT1-C1420T	CT *vs* CC	0.05-0.43	0.79 (0.65-0.98)	0.028	2	97%	0.97
RFC1-G80A	GA+AA *vs.* GG	-	1.37 (1.11-1.69)	0.003	2	98%	0.80
CYP1A1-T6235C	CT+TT *vs.* CC	0.19-0.54	1.36 (1.11-1.66)	0.003	7	99%	0.75
CYP2E1-5B	GC+CC *vs.* GG	0.01-0.29	1.99 (1.32-3.00)	0.001	4	41%	0.92
NQ01-C609T	CT+TT *vs.* CC	0.19-0.52	1.24 (1.02-1.50)	0.030	2	97%	0.97
XRCC1-G28152A	GA+AA *vs.* GG	0.10-0.28	1.78 (1.32-2.42)	0.001	3	77%	0.63

CI, confidence interval; FPRP false-positive report probability; MAF,minor allele frequency; OR, odds ratio.

ic determinants conferring modest, but potentially important, risks.

2. Because of the large number of polymorphisms in the genome, false-positive associations are inevitably more frequent than true-positive associations when testing large numbers of markers even if studies are rigorously conducted. Hence associations need to attain a high level of statistical significance to be established beyond reasonable doubt. For this reason, in GWA studies, a *P* value threshold of 5.0×10^{-8} has been advocated as being appropriate for genome-wide significance.[20,21]

3. Positive associations need to be replicated in independent case-control series to limit type 1 error rate. However, to increase the power of the replication studies, the allelic architecture of the population from which these additional case-control series are ascertained need to have similar ancestry and, ideally, the same linkage disequilibrium (LD) structure.

4. It should be recognized that ALL is heterogeneous both in terms of cellular origins, molecular biology and clinical response to therapy. Hence a given variant may not affect the risk of all subtypes. The power of any analysis stratified by cell lineage/genotype will be limited because of the small numbers of cases in each subgroup.

5. Careful attention must be paid to population stratification as a source of confounding findings, because disease rates and allele frequencies vary with race and ethnicity. This is one possible explanation for some of the false-positive associations reported in the literature. Use of genome-wide single nucleotide polymorphism (SNP) data allows the issue of population stratification to be addressed and adjustment of association statistics through statistical procedures such as principal component analysis.[22]

6. Rare germline variants may confer more profound effects on risk and hence have greater significance for individuals, though the population-attributable risk may be low. Only through genotyping and sequencing of large numbers of affected individuals can such rare variants be identified. Advances in sequencing technology will make this a feasible proposition in the very near future.

7. Given the unique nature of this early-onset disease, the investigation of parental genetics, particularly maternally-contributed effects is a prerequisite not only for understanding the etiology of ALL but also to pave the way toward new opportunities in preventive medicine. In addition to the urgent call for larger ALL case-control cohorts and replication datasets, there is also a need in the childhood leukemia research community for larger ALL family-based cohorts (i.e. case triads).

8. It is advantageous to consider taking epidemiological risk factor data into consideration to allow examination of interactions between the known or suspected etiological factors (e.g. birth weight, infectious exposures) and genetic risk variants in future studies. Higher order gene-gene and gene-environment interactions may contribute to ALL risk. Very large sample sizes will, however, be needed to detect such interactions, hence the power of individual cohorts will be limited.

9. There is increasing interest in the view that germline variation may influence patient outcome from malignancy. Large consortia provide the opportunity to investigate the relationships between genotype and patient phenotypes; however, harmonization of data from diverse patient cohorts is challenging.

Characteristics of low-penetrance variants

Most studies to date aimed at identifying low-penetrance alleles for ALL have been based on a candidate gene approach formulated on etiological hypotheses and/or limited epidemiological evidence for pathway exposure. However, without a clear understanding of causality, the choice of suitable genes for the disease is inherently problematic, making unbiased but comprehensive screening methodologies attractive.

The availability of comprehensive sets of tagging single nucleotide polymorphisms (SNPs) that capture much of the common sequence variation in the genome and the availability of high-resolution linkage disequilibrium (LD) maps allow GWA studies for disease associations to be conducted efficiently. This approach does not depend upon prior knowledge of function or presumptive involvement of any gene in disease causation. Moreover, it minimizes the probability of failing to identify important common variants in hitherto unstudied loci (i.e. genes and regulatory regions). This tactic has been very useful in identifying risk variants for a number of common tumors including, breast, colorectal, prostate, lung, and ovarian cancers, melanoma, glioma and chronic lymphocytic leukemia.[23]

Two GWA studies of ALL have so far been reported and four independent loci shown conclusively to be associated with ALL risk: 7p12.2 (*IKZF1*), 9p12 (*CDKN2A/CDKN2B*), 10q21.2 (*ARID5B*) and 14q11.2 (*CEBPE*)[2-4] (Table 2). Risks associated with common variants at each of these loci are modest (Odds ratios 1.5-1.6) and there is little evidence of interactive effects. With homozygous risk variants conferring approximately twice the heterozygote risk, the distribution of risk alleles follows a normal distribution in both case and controls, with a shift towards a higher number of risk alleles in affected individuals consistent with a polygenic model of disease predisposition (Figure 1).

Data from gene discovery efforts in other cancers are proving highly informative regarding the allelic architecture of cancer susceptibility. The number of common variants accounting for more than 1% of inherited risk appears low. Furthermore, only a small proportion of the total heritability of any cancer can be explained by the currently identified loci even though the contribution of the identified loci to disease risk may be conservative because of imperfect tagging surrogates for the true etiological loci. Multiple causal variants may also exist at each locus, including low frequency variants with significantly larger cumulative effects on risk. Few of the observed disease-associated variants are coding variants with many of the loci mapping to regions bereft of genes or protein-encod-

Table 2. Risk of ALL associated with 7p12.2, 9p21.3, 10q22.1, 14q11.2 variants. Data taken from Papaemmanuil[3] and Sherborne[2].

Tagging SNP	Gene	Chromosome/ position	Risk allele	Risk allele frequency in controls	Odds ratio	95% CI
rs4132601	*IKZF1*	7p12.2	C	0.27	1.69	1.58-1.81
rs3731217	*CDKN2A*	9p21.3	G	0.15	0.71	0.64-0.78
rs7089424	*ARIDB5*	10q22.1	C	0.34	1.65	1.54-1.76
rs2239633	*CEBPE*	14q11.2	G	0.52	1.34	1.22-1.45

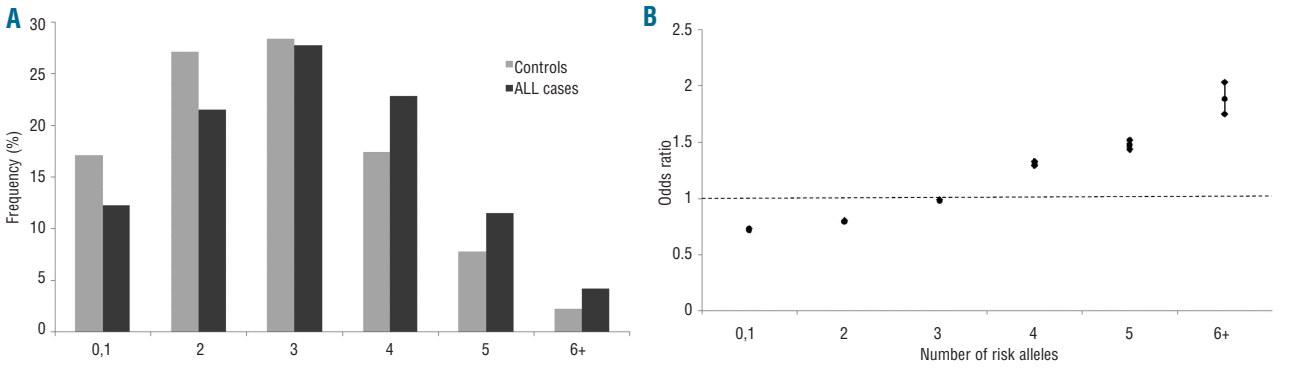

Figure 1. Cumulative impact of 7p12.2, 9p12, 10q21.2 and 14q11.2 variants on ALL risk.[2,3] (A) Distribution of risk alleles in controls (blue bars) and ALL cases (red bars) for the 4 loci (rs4132601, rs3731217, rs7089424 and rs2239633). (B) Plot of the increasing ORs for ALL with increasing number of risk alleles. The ORs are relative to the median number of 3 risk alleles. Vertical bars correspond to 95% confidence intervals. The distribution of risk alleles follows a normal distribution in both case and controls, with a shift towards a higher number of risk alleles in cases. Horizontal line denotes the null value (OR=1.0).

ing transcripts. It is likely that much of the common variation in ALL risk is mediated through sequence changes influencing gene expression, perhaps in a subtle fashion, or through effects on pathway components mitigated by functional redundancy. It is also possible that risk genotype may facilitate preferential somatic mutations as exemplified by an association between JAK2 mutation and haplotype in myeloproliferative neoplasms.[24]

Future directions

Prospects for identifying additional common variants

The power of the two reported GWA studies of ALL over a range of allele frequencies and relative risks is shown in Figure 2. The power of these studies to identify common alleles conferring risks of 1.5 or greater (such as the 7p12.2 variant) is high.[3,4] 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. In contrast, the two GWA studies have had low power to detect alleles with smaller effects and/or minor allele frequency (MAF) less than 0.1. Tagging SNPs employed for GWA studies capture on average approximately 80% of common SNPs in the European population, but only approximately 10% of SNPs with MAFs 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. GWA-based strategies are not configured optimally to identify low frequency variants with potentially stronger effects or identify recessively acting alleles. Format of many of the current commercial arrays is not ideal and may not appropriately capture copy number variants or other structural variants, which may also impact on ALL risk. A systematic search of these classes of polymorphic variant should be seen as complementary to classical GWA studies based on analysis of tagging SNPs. It is, therefore, highly likely that a large number of low-penetrance variants remain to be discovered. Efforts to develop GWA studies of ALL, in terms of both sample size and SNP coverage, and to increase the number of SNPs taken forward to large-scale replication should lead to the identification of additional novel risk variants for ALL using sample sets in excess of 3,000 cases and a similar number of controls that are likely to be forthcoming through multicenter collaborations (Figure 2). Ongoing GWA studies of ALL being conducted by different research groups will inevitably generate SNP data from different array platforms with different SNP representation and coverage. Using statistical methodology whereby imputation of untyped SNPs can be generated in datasets allows for harmonization and pooled analyses to be conducted.[26]

Subgroup analyses

Given the biological heterogeneity of ALL, risk 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, 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.[3,4] Given that the frequency of many ALL subgroups is small, identifying differential effects will only be realistically possible through multi-center pooled analyses.

Identifying causal variants

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.

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.[1,27]

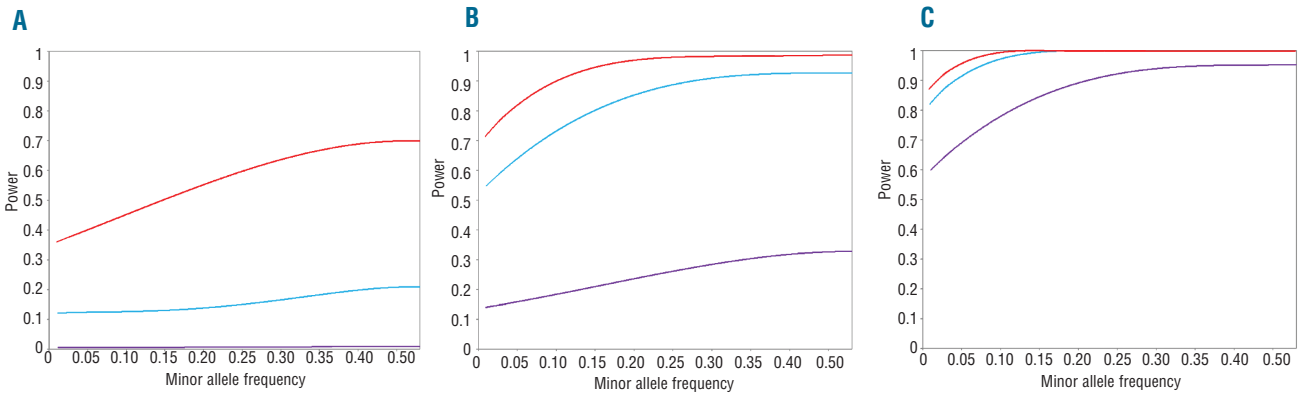

Figure 2. Power to identify risk loci for acute lymphoblastic leukemia over a range of minor allele frequencies and relative risks for $P=5x10^{-7}$ calculated using GWApower.[25] (A) Study of 441 cases and 17,958 controls reported by Trevino *et al.*[4] using Affymetrix 500K arrays. (B) Study of 907 cases and 2,398 controls reported by Papaemmanuil *et al.*[3] using Illimina 370K arrays: (C). Analysis of 3,000 cases and 3,000 controls using Illumina Omni 1M arrays.

Epidemiological studies have so far provided indirect evidence that ALL may have an infective basis although no specific infectious agent has been implicated.[1] 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.[28]

Ethnic differences in the risk of ALL are well recognized.[29] 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 and in a black population suggesting that 7p12.2 and 10q21.2 variation may contribute to racial differences in ALL risk.[30,31]

Identification of interaction between genetic variants and environmental risk factors is contingent on very large datasets, realistically something which can only be achieved through multi-center collaboration.

Inherited prognostic and predictive variants

In addition to influencing the risk of developing ALL, inherited genetic factors may play a role in determining the natural course of the disease and its response to therapies. As a prognostic factor, the concept of germline variation imparting inter-individual variability in tumor development and progression is currently receiving increasing attention. Prognostic studies of ALL have generally examined the same candidate genes as those hypothesized to play a role in susceptibility. It is, however, more probable that a genetic variant affecting inter-individual disease expression will impact on later stages of clonal development rather than early events associated with an inherited susceptibility. For example, variants in growth factors or immune surveillance signaling pathways might not impact on risk of initiation but could have a substantial effect on progression or outcome of established disease. Chemotherapy response and toxicity may also be related to germline genotype. As with conventional association studies, it is essential to impose appropriate statistical thresholds and conduct replication analyses to avoid the reporting of false positives. Linking GWA data to patient outcome provides an attractive strategy for identifying prognostic markers of outcome from ALL. Using this strategy it has recently been shown that germline variation in *IL15* influences the risk of minimal residual disease.[32]

Rationale for consortia

The recognition that common genetic variants contribute to ALL risk represents a major advance in our understanding of this disease. In view of the issues discussed above, collaboration has been developing between research groups that are engaged in searching for low-penetrance variants for ALL through association-based studies. These relatively loose affiliations centered around work on specific projects are beginning to crystallize into a more formal collaborative network. At present, 12 groups that are performing case-control genetic association studies have joined the IALLGC and the consortium is keen to involve other interested researchers. The IALLGC can be contacted at *www.iallgc.co.uk*.

This international collaborative initiative is seeking to comprehensively understand genetic susceptibility to ALL and to describe the genetic landscape of the disease. The immediate goal is to work together to study polymorphisms that have been shown to be associated with risk and to plan future high quality studies. This article provides a framework for these future studies, with biological analyses of risk variants and epidemiological studies as longer-term aims.

Authorship and Disclosures

The information provided by the authors about contributions from persons listed as authors and in acknowledgments is available with the full text of this paper at www.haematologica.org.

Financial and other disclosures provided by the authors using the ICMJE (www.icmje.org) Uniform Format for Disclosure of Competing Interests are also available at www.haematologica.org.

References

1. Greaves M. Infection, immune responses and the aetiology of childhood leukaemia. Nat Rev Cancer. 2006;6(3):193-203.
2. 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(6):492-4.
3. 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(9):1006-10.
4. 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(9):1001-5.
5. Vijayakrishnan J, Houlston R. Candidate gene association studies and risk of childhood acute lymphoblastic leukemia: a systematic review and meta-analysis. Haematologica. 2010;95(8):1405-14.
6. Wacholder S, Chanock S, Garcia-Closas M, El Ghormli L, Rothman N. Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. J Natl Cancer Inst. 2004;96(6):434-42.
7. Thomas DC, Clayton DG. Betting odds and genetic associations. J Natl Cancer Inst. 2004;96(6):421-3.
8. McHale CM, Wiemels JL, Zhang L, Ma X, Buffler PA, Feusner J, et al. Prenatal origin of childhood acute myeloid leukemias harboring chromosomal rearrangements t(15;17) and inv(16). Blood. 2003;101(11):4640-1.
9. Wiemels JL, Cazzaniga G, Daniotti M, Eden OB, Addison GM, Masera G, et al. Prenatal origin of acute lymphoblastic leukaemia in children. Lancet. 1999;354 (9189):1499-503.
10. Greaves MF. Aetiology of acute leukaemia. Lancet. 1997;349(9048):344-9.
11. Linet MS, Devesa SS. Descriptive epidemiology of childhood leukaemia. Br J Cancer. 1991;63(3):424-9.
12. Sandler DP, Ross JA. Epidemiology of acute leukemia in children and adults. Semin Oncol. 1997;24(1):3-16.
13. Greaves MF, Maia AT, Wiemels JL, Ford AM. Leukemia in twins: lessons in natural history. Blood. 2003;102(7):2321-33.
14. Chang JS, Selvin S, Metayer C, Crouse V, Golembesky A, Buffler PA. Parental smoking and the risk of childhood leukemia. Am J Epidemiol. 2006;163(12):1091-100.
15. Wen W, Shu XO, Potter JD, Severson RK, Buckley JD, Reaman GH, et al. Parental medication use and risk of childhood acute lymphoblastic leukemia. Cancer. 2002;95 (8):1786-94.
16. Infante-Rivard C, Sinnett D. Preconceptional paternal exposure to pesticides and increased risk of childhood leukaemia. Lancet. 1999;354(9192):1819.
17. Kwan ML, Metayer C, Crouse V, Buffler PA. Maternal illness and drug/medication use during the period surrounding pregnancy and risk of childhood leukemia among offspring. Am J Epidemiol. 2007;165(1):27-35.
18. Pombo-de-Oliveira MS, Koifman S. Infant acute leukemia and maternal exposures during pregnancy. Cancer Epidemiol Biomarkers Prev. 2006;15(12):2336-41.
19. Labuda D, Krajinovic M, Sabbagh A, Infante-Rivard C, Sinnett D. Parental genotypes in the risk of a complex disease. Am J Hum Genet. 2002;71(1):193-7.
20. Dudbridge F, Gusnanto A. Estimation of significance thresholds for genomewide association scans. Genet Epidemiol. 2008; 32(3):227-34.
21. The International HapMap C. A haplotype map of the human genome. Nature. 2005; 437(7063):1299-320.
22. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006;38(8):904-9.
23. Fletcher O, Houlston RS. Architecture of inherited susceptibility to common cancer. Nat Rev Cancer. 10(5):353-61.
24. Jones AV, Chase A, Silver RT, Oscier D, Zoi K, Wang YL, et al. JAK2 haplotype is a major risk factor for the development of myeloproliferative neoplasms. Nat Genet. 2009;41(4):446-9.
25. Spencer CCA, Su Z, Donnelly P, Marchini J. Designing Genome-Wide Association Studies: Sample Size, Power, Imputation, and the Choice of Genotyping Chip. PLoS Genet. 2009;5(5):e1000477.
26. Infante-Rivard C. Diagnostic x rays, DNA repair genes and childhood acute lymphoblastic leukemia. Health Phys. Health Phys. 2003;85(1):60-4.
27. Roman E, Simpson J, Ansell P, Kinsey S, Mitchell CD, McKinney PA, et al. Childhood acute lymphoblastic leukemia and infections in the first year of life: a report from the United Kingdom Childhood Cancer Study. Am J Epidemiol. 2007;165(5):496-504.
28. Callan AC, Milne E. Involvement of the IGF system in fetal growth and childhood cancer: an overview of potential mechanisms. Cancer Causes Control. 2009;20 (10):1783-98.
29. Stiller CA, Parkin DM. Geographic and ethnic variations in the incidence of childhood cancer. Br Med Bull. 1996;52(4):682-703.
30. Vijayakrishnan J, Sherborne AL, Sawangpanich R, Hongeng S, Houlston RS, Pakakasama 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(10):1870-4.
31. 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 (4):894-6.
32. Yang JJ, Cheng C, Yang W, Pei D, Cao X, Fan Y, et al. Genome-wide interrogation of germline genetic variation associated with treatment response in childhood acute lymphoblastic leukemia. JAMA. 2009;301 (4):393-403.