The putative role of transforming viruses in childhood acute lymphoblastic leukemia

Jane MacKenzie
Mel F. Greaves
Tim O.B. Eden
Rob A. Clayton
Jacqueline Perry
Katherine S. Wilson
Ruth F. Jarrett

Epidemiological evidence suggests that infection is involved in the etiology of common acute lymphoblastic leukemia, either by stimulating an inappropriate immune response or in the form of a classical transforming agent. In an attempt to elucidate the role that infection is playing in this disease, we used representational difference analysis (RDA) to examine tumor samples for the presence of exogenous genomes. Twenty RDA experiments were carried out, using four different restriction enzymes, but no exogenous sequences were identified within leukemic cells. These results suggest that it is unlikely that a single, direct transforming agent is involved in the pathogenesis of common acute lymphoblastic leukemia.

Key words: etiology, genomes, subtraction, pathogen, leukemia.

Haematologica 2006; 91:240-243

From the LRF Virus Centre, Institute of Comparative Medicine, University of Glasgow, Glasgow, (JMK, RAC, JP, KSW, RFJ); Section of Haemato-Oncology, Institute of Cancer Research, Chester Beatty Laboratories, London, UK (MFG); Academic Unit of Pediatric Oncology, Central Manchester and Manchester Children's Hospitals and Christie Trusts, Manchester, UK (TOBE).

Correspondence:
Ruth F. Jarrett, LRF Virus Center, Institute of Comparative Medicine University of Glasgow, Glasgow, G61 1QH, UK.
E-mail: r.f.jarrett@vet.gla.ac.uk

Childhood acute lymphoblastic leukemia (ALL) comprises several entities with different cancer cell phenotypes and age-specific incidence curves, which are likely to be etiologically distinct. Common acute lymphoblastic leukemia (cALL), in which the leukemic cell is a B-cell precursor, accounts for the peak in incidence rate seen between the ages of 2 and 5 years in areas of high socio-economic status.[1] For cALL, there is now unambiguous molecular evidence that, in most cases, a pre-leukemic state is initiated *in utero*.[2-5] The *MLL-AF4* and *TEL-AML1* fusion genes or a hyperdiploid karyotype can be detected at birth in the blood of children who subsequently go on to develop ALL.[3-5] Identical *TEL-AML1* fusions have also been detected in monozygotic twins with concordant leukemia;[4] however, concordance of cALL in monozygotic twins is modest (~10%) suggesting that post-natal events, i.e., exposures and secondary events, are required for the development of overt leukemia.

The idea that infectious agents may play some role in the development of childhood leukemia is long-standing and plausible. Two distinct, but not mutually exclusive, models have been proposed based on time trends or geographic variation in incidence or clustering of cases.[6-8] The broad association between cALL and socio-economic development[1] has led to the suggestion that cALL may occur as a consequence of delayed exposure in infancy to an infectious agent or agents.[6,7] An abnormal or dysregulated immunological response to post-natal infection, promoting secondary genetic events and leukemia development, is the key feature of this model. The putative agent(s) may be viral or bacterial, but the mechanism of action is indirect or non-transforming.

The second model predicts that childhood leukemia occurs as a rare response to a common transforming virus under certain demographic conditions.[8] It has been shown that influxes of new populations into previously isolated communities, such as occurred during the development of British New Towns during the 1950s, are associated with transiently increased incidences of childhood leukemia. Kinlen postulated that this was due to unrecognized epidemics of infection by an unidentified virus within such communities, and that these epidemics led to the transiently increased incidence of leukemia.[8,9] Similar studies outside the UK support this hypothesis,[10] as does space-time clustering.[11,12] We[13,14] and others[15] have previously investigated the idea that herpesviruses or the polyomaviruses JC and BK might be involved but found no evidence for the presence of viral genomes in ALL cells. Other investigators have retrospectively analyzed patients' neonatal blood spots (Guthrie cards) for these same viral sequences[16,17] or parvovirus B19[18] and reported negative results. To investigate more comprehensively whether DNA genomes of infectious agents are present in leukemic cells, we analyzed a series of cALL cases using the genomic subtractive hybridization technique, representational difference analysis (RDA).[19]

Design and Methods

Considerations on the use of RDA in the investigation of CALL. RDA is designed to identify differences between two complex genomes; in the context of this study the difference is predicted to be a DNA genome present in the leukemic cells but not in normal cells from the same individual. The leukemic sample, containing the putative genome or *target*, is termed the *tester* and the partner sample is

termed the *driver*. In these experiments we used leukemic cells from the patient as the tester and, when possible, granulocytes from a remission blood sample as the driver. Pooled DNA extracted from whole blood samples from both parents of the patient was used in cases in which a satisfactory remission sample was not available.

The probability of detecting a genomic fragment of a given pathogen is dependent on the size of the pathogen's genome and the restriction enzyme used to generate representations. We calculated that a genome of 150 kb or more, such as a herpesvirus genome, would be detected with a probability of 99% following analysis of a single representation defined by three of the four enzymes used in this study. To give a >95% probability of detecting a viral genome of 30 kb, one or two RDAs would have to be performed and for a 9 kb genome three or four RDA would be required, depending on the restriction enzyme used. We, therefore, carried out RDA using representations derived with four different enzymes. According to our original protocol, we intended to perform four RDA on each of five cases (20 RDA); however, there was insufficient DNA from each case to prepare four representations and we therefore required samples from 11 cases to complete these 20 RDA (Table 1).

Case and sample selection. Eleven patients with cALL, aged between 2 and 15 years, were included in this analysis (Table 1). In all cases the leukemic cells fulfilled the phenotypic criteria for the diagnosis of cALL ($CD10^+$, $CD19^+$, TdT^+, FAB L1 or L2) and >90% of the mononuclear cells were blast cells. Following informed consent, pretreatment blood or bone marrow samples and, when possible, remission blood samples were collected. DNA was extracted from mononuclear and granulocyte fractions using standard procedures. Cases and sample details and the RDA performed are shown in Table 1.

Representational difference analysis protocol. In early experiments the original RDA protocol[19] was followed. After evaluation of the results of these experiments, a number of modifications were introduced: the control driver: driver hybridization was replaced with a tester: driver hybridization; small scale pilot experiments were used to determine the optimal number of polymerase chain reaction (PCR) cycles required at each stage to ensure the maintenance of a wide size range of amplified fragments; more efficient and reproducible post-hybridization amplification was achieved by inactivating the mung bean nuclease by phenol chloroform extraction and ethanol precipitation, rather than heat inactivation; and four separate adaptor-primer sets were used rather than the three (R, J and N adaptor-primers) originally described, to reduce the number of illegitimate products generated by annealing of primers to genomic DNA (*supplementary Table, online version*).[19] The original protocol using three adaptor-primers was used for all analyses with BglII, both original and optimized protocols were used for *Bam*HI analyses and the optimized adaptor-primer protocol was used in all RDA performed using *Eco*RI and *Xba*I.

In those experiments in which discrete fragments were identified following three rounds of RDA, the entire third round products, or gel-purified fragments in later experiments, were cloned into pCR®2.1 or pCR®4-TOPO® using the appropriate TA cloning kit (Invitrogen, Paisley, UK), according to the manufacturer's instructions. Nucleotide sequencing of cloned fragments was carried out using primers derived from standard vector sequences and Big DyeTM terminator sequencing kits (Applied Biosystems, Warrington, UK). Thermal cycling was performed using a GeneAmp PCR System 2400 or 9600 (Applied Biosystems) and, following removal of unincorporated fluorescent nucleotides, products were analyzed on an ABI PRISM™ 310 or 3100 Genetic Analyzer (Applied Biosystems). The sequences were analyzed using the NIX sequence analysis package (UK Human Genome Mapping Project Resource Centre, Cambridge, UK), which simultaneously performs Blast searches against vector, protein, mRNA, expressed sequence tag and high throughput genome sequence (HTG) databases. When the origin of third round fragments could not be determined by sequence and Nix analysis, fragments were labeled with α-^{32}P dCTP by random priming using Redivue™ kits (Amersham Biosciences UK Ltd., Little Chalfont, UK) and hybridized to driver and tester representations of the case under investigation, using standard procedures.

Table 1. Patients' details and representations analyzed.

Patient	*Age*	*Sex*	*Enzyme*	*Driver sample**	*Control experiment*	*Difference products†*
1	9	F	*Bam*HI	remission	driver: driver	no
2	3	M	*Bam*HI	remission	driver: driver	no
			*Bgl*II	remission	driver: driver	yes
3	8	M	*Bam*HI	remission	driver: driver	yes
4	15	F	*Bam*HI	remission	tester: driver	no
5	9	M	*Bgl*II	remission	tester: driver	yes
6	2	M	*Bam*HI	remission	tester: driver	yes
			*Bgl*II	remission	tester: driver	yes
7	2	F	*Eco*RI	parental	tester: driver	smear‡
			*Xba*I	parental	tester: driver	no
8	4	F	*Bam*HI	parental	tester: driver	yes
			*Bgl*II	parental	tester: driver	yes
			*Eco*RI	parental	tester: driver	smear‡
			*Xba*I	parental	tester: driver	no
9	3	M	*Eco*RI	parental	tester: driver	no
			*Xba*I	parental	tester: driver	no
10	2	F	*Eco*RI	parental	tester: driver	no
			*Xba*I	parental	tester: driver	no
11	5	F	*Eco*RI	parental	tester: driver	no
			*Xba*I	parental	tester: driver	yes

**Driver samples were either DNA extracted from granulocytes from a remission blood sample from the patient or pooled DNA extracted from the whole blood of the parents. †Difference products detectable in driver: tester experiment following three rounds of representational difference analysis. ‡In two analyses using EcoRI representations, a smear was observed on ethidium bromide-stained agarose gels following electrophoresis of products obtained following 35 cycles of PCR; all other post-hybridization products were detectable after 25 cycles of PCR and we therefore presume that there were no genuine difference products in these two analyses.*

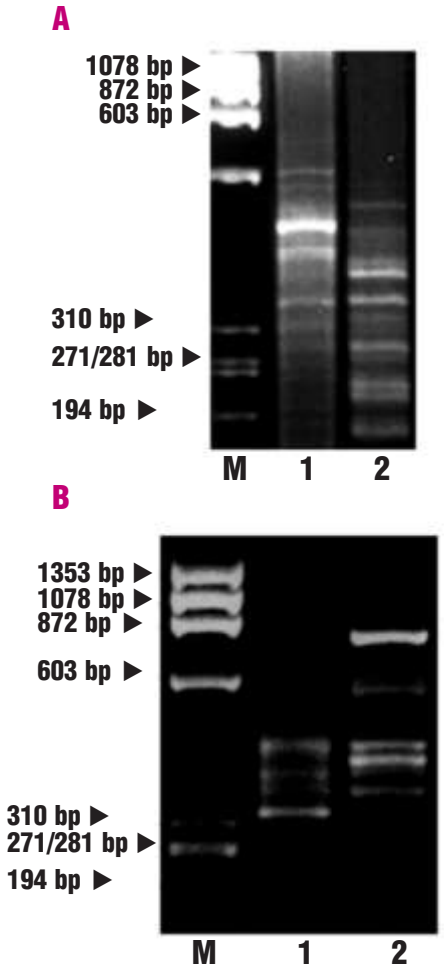

Figure 1. Analysis of third round products of representational difference analysis. Ethidium bromide-stained 3% Nusieve agarose gel showing the third round products of two representational difference analysis (RDA) experiments. M, molecular weight marker; 1, third round products using parental DNA as the tester (control experiment); 2, third round products using tumor DNA as the tester. A, analysis of case #6 using *BamHI* representations: following three rounds of RDA using *BamHI* representations, multiple fragments were detected in the difference products from both the driver: tester and control experiments. The entire products of both driver: tester and tester: driver hybridizations were cloned; 76 clones from the driver: tester experiment and 67 from the reciprocal experiment were subjected to sequence analysis. Thirty-three independent DNA sequences were obtained more than once (range 2–10), and accounted for 72 of the 143 clones analyzed. The most frequently isolated fragment from the driver: tester experiment contained part of a long terminal repeat of the human endogenous retrovirus, HERV9. Nix analysis failed to identify a match for 16 of the 143 sequences; these fragments were used as probes against the tester and driver representations from case #6 but none was found to be tester-specific. Recognized repeat elements were present in 40 of the 143 cloned products. B, analysis of case #11 using *XbaI* representations: following three rounds of RDA using *XbaI* representations, distinct fragments were detected in both the driver: tester and tester: driver experiments. Fragments present in the driver: tester hybridization were excised from agarose gels and subjected to cloning and sequencing and clones corresponding to the bands visualized on the agarose gels were identified. All of the 122 sequences obtained matched known human sequences; 27 were repetitive sequences and 95 had complete homology to named genes.

Results and Discussion

In order to investigate the hypothesis that a single transforming agent is involved in the pathogenesis of cALL, we investigated samples from children with cALL using RDA. This technique facilitates the identification of differences between two complex DNA genomes and can, therefore, be used to identify exogenous DNA sequences, such as viral genomes.[19,20] Following validation of the experimental protocol,[20] RDA experiments, utilizing representations generated following digestion with four different restriction enzymes, were performed using cALL samples as testers (Table 1). In 12 of these experiments no difference products were detected after three rounds of RDA. In the remaining eight experiments, distinct fragments were detected in the third round difference products and these were subjected to cloning and sequencing. Sequences of, or related to, viral genomes were not detected and there was no evidence of non-human DNA in these difference products. Further details are presented in Table 1 and Figure 1.

RDA has a proven track record for viral discovery[20] and has the distinct advantage over other molecular methods of virus discovery that no *a priori* knowledge of the potential infectious agent is required. The method does, however, have certain limitations. First, exogenous sequences must be present at close to single copy level in the tester sample to ensure detection. In the present study the tester samples contained >90% leukemic cells and pilot experiments confirmed that we were able to detect single copy differences between tester and driver samples. Thus, it is unlikely that we have missed an exogenous sequence present in all the tumor cells but we cannot exclude the presence of a pathogen in a small proportion of the tumor cells, or in another cell population, which is contributing to leukemogenesis by an indirect mechanism. Secondly, infectious agents will only be detected if the starting tester representation includes a fragment containing part of the genome of the pathogen. In order to minimize the risk of missing small genomes, we performed RDA with four different enzymes. We predicted that this would give us a >95% chance of detecting a viral genome of >9 kb, such as a retroviral genome. Lastly, RDA relies on differences between two samples, in this case leukemic cells and either remission granulocytes or pooled parental samples. Genomes present in both samples at similar levels would escape detection by RDA. TT viruses have small genomes (~3.7 kb), infect granulocytes and lymphocytes, and show extensive genomic variation and, therefore, could have been missed in this study; however, there is currently no evidence that these viruses are involved in the pathogenesis of childhood ALL.[21]

Epidemiological and genetic data support a role for common infections in childhood ALL.[7-12] Our failure to identify exogenous viral sequences in ALL cells via a sensitive, generic screen makes it unlikely that direct transformation is involved in most cases. This leaves an indirect, immunological mechanism as a more plausible explanation.[7]

RJ, MG and JMK conceived the study. The study was performed in RJ's laboratory under the direction of JMK, who also performed all the pilot experiments and most of the analysis. RC, JP and KW performed the benchwork and KSW helped with some of the later analyses. MG and TE provided the samples. JMK, MG and RJ wrote the manuscript with contributions from other authors. The authors declare that they have no potential conflicts of interest.

Manuscript received August 24, 2005. Accepted November 21, 2005.

References

1. Greaves MF, Colman SM, Beard ME, Bradstock K, Cabrera ME, Chen PM, et al. Geographical distribution of acute lymphoblastic leukaemia subtypes: second report of the collaborative group study. Leukemia 1993;7:27-34.
2. Mori H, Colman SM, Xiao Z, Ford AM, Healy LE, Donaldson C, et al. Chromosome translocations and covert leukemic clones are generated during normal fetal development. Proc Natl Acad Sci USA 2002;99:8242-7.
3. Greaves MF, Wiemels J. Origins of chromosome translocations in childhood leukaemia. Nat Rev Cancer 2003;3:639-49.
4. Greaves MF, Maia AT, Wiemels JL, Ford AM. Leukemia in twins: lessons in natural history. Blood 2003; 102: 2321-33.
5. Maia AT, Tussiwand R, Cazzaniga G, Rebulla P, Colman S, Biondi A, et al. Identification of preleukemic precursors of hyperdiploid acute lymphoblastic leukemia in cord blood. Genes Chromosomes Cancer 2004; 40:38-43.
6. Bhatia S, Ross JA, Greaves MF, Robison LL. In: Pui CH ed. Childhood Leukemias. Cambridge: Cambridge University Press; 1999. p. 38-49.
7. Greaves MF. Speculations on the cause of childhood acute lymphoblastic leukemia. Leukemia 1988;2:120-5.
8. Kinlen L. Evidence for an infective cause of childhood leukaemia: comparison of a Scottish new town with nuclear reprocessing sites in Britain. Lancet 1988;2:1323-7.
9. Kinlen LJ. Epidemiological evidence for an infective basis in childhood leukaemia. Br J Cancer 1995;71:1-5.
10. Alexander FE, Chan LC, Lam TH, Yuen P, Leung NK, Ha SY, et al. Clustering of childhood leukaemia in Hong Kong: association with the childhood peak and common acute lymphoblastic leukaemia and with population mixing. Br J Cancer 1997; 75:457-63.
11. Birch JM, Alexander FE, Blair V, Eden OB, Taylor GM, McNally RJ. Space-time clustering patterns in childhood leukaemia support a role for infection. Br J Cancer 2000;82:1571-6.
12. Gustafsson B, Carstensen J. Evidence of space-time clustering of childhood acute lymphoblastic leukaemia in Sweden. Br J Cancer 1999;79:655-7.
13. MacKenzie J, Perry J, Ford AM, Jarrett RF, Greaves M. JC and BK virus sequences are not detectable in leukaemic samples from children with common acute lymphoblastic leukaemia. Br J Cancer 1999;81:898-9.
14. MacKenzie J, Gallagher A, Clayton RA, Perry J, Eden OB, Ford AM, et al. Screening for herpesvirus genomes in common acute lymphoblastic leukemia. Leukemia 2001;15:415-21.
15. Smith MA, Strickler HD, Granovsky M, Reaman G, Linet M, Daniel R, et al. Investigation of leukemia cells from children with common acute lymphoblastic leukemia for genomic sequences of the primate polyomaviruses JC virus, BK virus, and simian virus 40. Med Pediatr Oncol 1999; 33: 441-3.
16. Priftakis P, Dalianis T, Carstensen J, Samuelsson U, Lewensohn-Fuchs I, Bogdanovic G, et al. Human polyomavirus DNA is not detected in Guthrie cards (dried blood spots) from children who developed acute lymphoblastic leukemia. Med Pediatr Oncol 2003;40:219-23.
17. Bogdanovic G, Jernberg AG, Priftakis P, Grillner L, Gustafsson B. Human herpes virus 6 or Epstein-Barr virus were not detected in Guthrie cards from children who later developed leukaemia. Br J Cancer 2004;91:913-5.
18. Isa A, Priftakis P, Broliden K, Gustafsson B. Human parvovirus B19 DNA is not detected in Guthrie cards from children who have developed acute lymphoblastic leukemia. Pediatr Blood Cancer 2004;42:357-60.
19. Lisitsyn N, Lisitsyn N, Wigler M. Cloning the differences between two complex genomes. Science 1993; 259: 946-51.
20. Chang Y, Cesarman E, Pessin MS, Lee F, Culpepper J, Knowles DM, et al. Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma. Science 1994;266:1865-9.
21. Shiramizu B, Yu Q, Hu N, Yanagihara R, Nerurkar VR. Investigation of TT virus in the etiology of pediatric acute lymphoblastic leukemia. Pediatr Hematol Oncol 2002;19:543-51.