Acquisition of genome-wide copy number alterations in monozygotic twins with acute lymphoblastic leukemia

Caroline M. Bateman,1 Susan M. Colman,1 Tracy Chaplin,2 Bryan D. Young,2 Tim O. Eden,3 Manoo Bhakta,4 Eric J. Gratias,4 Elisabeth R. van Wering,5 Giovanni Cazzaniga,6 Christine J. Harrison,7 Richard Hain,8 Philip Ancliff,8 Anthony M. Ford,1 Lyndal Kearney,1 and Mel Greaves1

1Section of Haemato-Oncology, The Institute of Cancer Research, Surrey, United Kingdom; 2Medical Oncology Unit, Cancer Research UK, London, United Kingdom; 3Academic Unit of Paediatric Oncology, University of Manchester, Teenage Cancer Trust Oncology Unit, Christie Hospital National Health Service Trust, Manchester, United Kingdom; 4University of Tennessee at Chattanooga and T. C. Thompson Children’s Hospital, Chattanooga; 5Dutch Childhood Oncology Group, The Hague, The Netherlands; 6Centro Ricercro Tettamanti, Clinica Pediatrica University di Milano Bicocca, Monza, Italy; 7Northern Institute for Cancer Research, Newcastle University, Newcastle upon Tyne, United Kingdom; 8Children’s Hospital for Wales, Cardiff, United Kingdom; and 9Camelia Botnar Laboratories, Great Ormond Street Hospital, London, United Kingdom

Chimeric fusion genes are highly prevalent in childhood acute lymphoblastic leukemia (ALL) and are mostly prenatal, early genetic events in the evolutionary trajectory of this cancer. ETV6-RUNX1-positive ALL also has multiple (−6 per case) copy number alterations (CNAs) as revealed by genome-wide single-nucleotide polymorphism arrays. Recurrent CNAs are probably “driver” events contributing critically to clonal diversification and selection, but at diagnosis, their developmental timing is “buried” in the leukemia’s covert natural history. This conundrum can be resolved with twin pairs. We identified and compared CNAs in 5 pairs of monozygotic twins with concordant ETV6-RUNX1-positive ALL and 1 pair discordant for ETV6-RUNX1 positive ALL. We compared, within each pair, CNAs classified as potential “driver” or “passenger” mutations based upon recurrency and, where known, gene function. An average of 5.1 (range 3-11) CNAs (excluding immunoglobulin/T-cell receptor alterations) were identified per case. All “driver” CNAs (total of 32) were distinct within each of the 5 twin pairs with concordant ALL. “Driver” CNAs in another twin with ALL were all absent in the shared ETV6-RUNX1-positive preleukemic clone of her healthy co-twin. These data place all “driver” CNAs secondary to the prenatal gene fusion event and most probably postnatal in the sequential, molecular pathogenesis of ALL. (Blood. 2010;115(17):3553-3558)

Introduction

The common chromosome abnormalities observed in childhood acute lymphoblastic leukemia (ALL), chimeric fusion genes generated by chromosome translocation and hyperdiploidy, are predominantly prenatal in origin but, in themselves, insufficient for leukemia to develop, suggesting a minimal 2 “hit,” pre/postnatal model for molecular pathogenesis and etiology.1 This view finds support in modeling experiments with mice2,3 and human cells.4 Recent observations suggest, however, substantially more genetic complexity in ALL than previously suspected. Data derived from high-resolution single-nucleotide polymorphism (SNP) arrays have indicated that cases of B-cell precursor ALL with the common chimeric fusion gene ETV6-RUNX1 have, in addition to the fusion gene, an average of 6 additional DNA aberrations or copy number alterations (CNAs; range 1-21), several being recurrent deletions in genes with functions impacting on B-cell lineage differentiation or cell-cycle control5-8 and therefore presumed “driver” mutations.9 This provides a challenge to the presumption that the fusion gene itself is necessary and sufficient as a first prenatal hit generating the preleukemic phase of disease.1,4 A key question then becomes the issue of when, in relation to fusion genes, these multiple CNAs arise in multistep molecular pathogenesis and clonal evolution. This is difficult to ascertain in diagnostic samples as the most recent subclonal expansions can disguise prior historical sequences of genetic events. Identical twin pairs with concordant ETV6-RUNX1-positive ALL provide a unique and tractable approach to this problem.

ETV6-RUNX1 genes have highly variable but clone-specific breakpoints and fusion region genomic sequences and are therefore patient specific.1 The only exception to this is with monozygotic twins with concordant ALL who share the same acquired (ie, nonconstitutive) but unique fusion gene sequence reflecting monoclonal origin in 1 fetus in utero.10 Concordance of ALL then arises as a consequence of intraplacental spread of an initiated preleukemic clone from one twin to the other. The ALL concordance rate of 10% to 15% in such a twin context10 suggests the necessity of additional genetic hits for which there is significant discordance (ie, ~85%-90%). The most common additional genetic abnormality in ETV6-RUNX1-positive ALL is deletion of the normal ETV6 allele,11 which can be subclonal (by fluorescent in situ hybridization [FISH]) and therefore secondary to ETV6-RUNX1.11,12 ETV6 deletions can also be discordant in twin pairs, indicating they are probably postnatal in origin.12-15 If ETV6 deletions and all other recurrent and presumed “driver” CNAs identified more recently in ALL5-8 are consistently secondary to ETV6-RUNX1 and postnatal,
then a straightforward and testable prediction is that they should be distinct or different within twin pairs with concordant ALL. Contrariwise, if one or all of such genetic alterations precede or occur simultaneously with gene fusion, then they should be identical in twin pairs with ALL.

**Methods**

**Patients**

Five pairs of identical (monozygotic, monochorionic) twins with concordant ETV6-RUNX1 fusion gene positive ALL and a further twin pair, who were discordant for ETV6-RUNX1 fusion gene positive ALL, were accrued from several centers for the study. The clinical, immunophenotypic, cytogenetic, and FISH data are given in supplemental Table 1 (available on the Blood Web site; see the Supplemental Materials link at the top of the online article). Ethical review committee approval was obtained for the study from the Royal Marsden NHS Trust. We used high-resolution (500 K) SNP mapping arrays, and by comparing with matched normal or remission blood DNA, we explored the clonal relationships of their respective “within pair” genetic abnormalities.

**Genome-wide copy number analyses**

DNA was extracted according to standard methods from leukemic bone marrow acquired at diagnosis and peripheral blood taken during clinical remission. The diagnostic leukemic DNA was derived from bone marrow samples with a median leukemic blast cell count of 90% (supplemental Table 1) SNP array analysis was carried out using GeneChip Human Mapping 250K Nsp and 250K Sty arrays according to the manufacturer’s instructions (Affymetrix). This accounts for 500 000 SNPs with a mean physical distance between SNPs of 2.5 kb. GTYPE was used for analysis of signal intensity and for genotype calling (Affymetrix; full SNP genotype data are available at www.icr.ac.uk/array/array.html). The mean call rate was 95% (range 81.25%-98.84%; supplemental Table 5). For copy number estimates, the Genome Orientated Laboratory File (GOLF Version 2.2.9) software package was used.16,17 Hybridization values were normalized to the median value of each array. Copy number was determined based on the log2 ratio of the signal intensity from the leukemia sample versus matched peripheral blood DNA taken during clinical remission, in all but 1 twin pair, and was done by visual inspection; potential copy number changes included at least 3 SNPs. In this case (twin pair 3), no remission sample was available, and the pooled signal intensity of 15 unrelated ALL remission samples was used.

**FISH**

Interphase FISH was performed on methanol-acetic acid fixed cells, which were only available for twin pairs 5 and 6, using standard methods. Bacterial artificial chromosome (BAC) or fosmid probes for the region of interest were obtained from the BACPAC Resource Center (Children’s Hospital, Oakland Research Institute, Oakland, CA; http://bACPac.chori.org), labeled with biotin-16-dUTP or digoxigenin-11-dUTP, hybridized, and detected with streptavidin Cy5 (for biotin-labeled probes) and (1) monoclonal anti-digoxigenin (Sigma-Aldrich), (2) horse anti–mouse
immunoglobulin (Ig)–Texas Red (Vector Laboratories), and (3) goat anti–horse Ig–Texas Red (Jackson Immunochemicals; for digoxigenin-labeled probes). Fluorescent signals were viewed using a Zeiss Axioskop fluorescence microscope equipped with epifluorescence and filters for DAPI (4′,6-diamidino-2-phenylindole), FITC (fluorescein isothiocyanate), Texas Red, and Cy5, and images were captured and analyzed using a charge-coupled device (Photometrics) and SmartCapture software (Digital Scientific). The fosmid and BAC probes used were tested on normal metaphase smears to confirm correct chromosome localization. No metaphase cells were available for twin pairs 5 and 6.

Results

Monozygotic twins concordant for ETV6-RUNX1–positive ALL

Table 1 summarizes the data, and Figure 1A through C provide illustrative examples of some key “paired” SNP array data in the monozygotic twins concordant for ETV6-RUNX1 ALL. The total number of CNAs (including TCR [T-cell receptor] and Ig rearrangements) in the 5 twin pair samples averaged 7.9 aberrations, with a range of 5 to 15. To facilitate evaluation of the complex datasets generated, we elected, before data acquisition, to divide CNAs into potential “drivers,” potential neutral or nonfunctional “passengers,” and “physiological” CNAs, while recognizing that these distinctions are not entirely unambiguous. Potentially functional or driver CNAs (“drivers”) were defined as those that were observed to be recurrent in prior series of SNP or comparative genomic hybridization arrays screens of ETV6-RUNX1 ALL5-8 and, when deletions were small, then occurring in or close to genes with functions known to be relevant to leukemogenesis or B-cell regulation5 (see supplemental Table 2 for list of recurrent CNAs). It is accepted that driver status ultimately requires confirmation by functional studies.9 CNAs that were nonrecurrent (this study and prior series5-8) in gene poor regions or in genes of no known functional relevance to lymphocyte biology or leukemia were considered as likely “passengers.” We recognize that anonymous and sporadic CNAs could occasionally transpire to be true “drivers.” As in prior studies,5 we classify gene deletions in IGH or TCR loci as most probably “physiological” and associated with the developmental rearrangements of these loci associated with diversification of antigen receptors in the lymphoid lineages (supplemental Table 3).

Table 1. CNAs identified by SNP array

<table>
<thead>
<tr>
<th>Twin sets</th>
<th>Number of CNAs (identical)</th>
<th>Twin set 1</th>
<th>Twin set 2</th>
<th>Twin set 3</th>
<th>Twin set 4</th>
<th>Twin set 5</th>
<th>Total (identical)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1a</td>
<td>1b</td>
<td>2a</td>
<td>2b</td>
<td>3a</td>
<td>3b</td>
</tr>
<tr>
<td>“Drivers”</td>
<td></td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>“Passengers”</td>
<td></td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>“Physiological”</td>
<td></td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>1</td>
</tr>
</tbody>
</table>

CNA indicates copy number alteration.

Table 2. Presence and size of “driver” CNAs and associated genes in twin pairs

<table>
<thead>
<tr>
<th>Cytoband</th>
<th>Size of CNA (Mb)*</th>
<th>Twin set 1</th>
<th>Twin set 2</th>
<th>Twin set 3</th>
<th>Twin set 4</th>
<th>Twin set 5</th>
<th>Gene(s) in region</th>
<th>Reported in karyotype†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deletions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1p32.3–pter</td>
<td>No CNA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3q13.2 No CNA</td>
<td>0.147</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3q26.32 No CNA</td>
<td>0.28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5q21.1–q35.3 No CNA</td>
<td>79.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6p22.1 0.147 No CNA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6q15–q27 No CNA</td>
<td>79.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7q34–q35 No CNA</td>
<td>6.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9p All 9p† No CNA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Whole arm</td>
</tr>
<tr>
<td>9p13.2 0.248 0.436 No CNA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9p21.3 1.39% No CNA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12p 8.98‡ 20.7‡ No CNA</td>
<td>22.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Many, including ETV6</td>
</tr>
<tr>
<td>12p13 0.866‡ 4.44‡ 9.96 6.5 0.259 No CNA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Many including ETV6</td>
</tr>
<tr>
<td>13q12.3–q34 83.5 81.2 No CNA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Many, including RB1</td>
</tr>
<tr>
<td>21q22.3–qter 4.9 No CNA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Many (not including RUNX1)</td>
</tr>
<tr>
<td>Gains</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5pter–q21.1 no CNA</td>
<td>101.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6pter–q15 no CNA</td>
<td>91.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All 10 All 10 No CNA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15q14–q15.1 No CNA</td>
<td>4.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21q All 21q No CNA</td>
<td>19.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Many, including LTK</td>
</tr>
</tbody>
</table>
| CNA indicates copy number alteration.  
*No CNA indicates no copy number alteration detected.  
†Abnormally reported by conventional karyotyping and/or FISH.  
‡Identical sized focal deletion presumed to occur at the same time as the ETV6-RUNX1 rearrangement.  
§Deletion appeared subclonal.
Physiological IGH or TCR CNAs were either identical (4/28) or different (24/28) in twin pairs (Table 1 and supplemental Table 3). Where such deletions are shared, we take this to indicate that the ETV6-RUNX1 fusion arose in a progenitor B cell that had either already undergone these rearrangements or simultaneously acquired them.

Sporadic CNAs classified as “passengers” (Table 1) were either identical (4/19) or distinct (15/19) in twin pairs. The most likely explanation for sharing of “passenger” CNAs is that either these preceded the ETV6-RUNX1 fusion event in the same cell or its predecessor, or that they represent collateral damage from whatever caused a simultaneous ETV6-RUNX1 fusion in the same “target” cell. Formally, it is also possible that shared passenger CNAs arose after the initiating ETV6-RUNX1 lesion. However, this would require fixation in the same subclone of both twins by subsequent (and distinct) “driver” CNAs. Passenger CNAs were all in regions with either no annotated gene or involved genes with no known or predictable role in leukemogenesis (supplemental Table 3), thus endorsing their likely passenger or neutral status.

A total of 32 “driver” mutations were identified (average 3.2 per case, range 2-7). These were all within the same chromosome regions or genes as observed in previous studies of childhood ALL, including genes encoding predominantly B-lineage transcription factors or cell-cycle regulators5-8 (Table 2). Significantly, all 5 twin pairs were discordant for all 32 CNAs (Table 1, Figure 1A-B). Discordance was reflected in either both twins sharing the same gene deletion (eg, in ETV6) but with different genomic boundaries of deletion (Figure 1A) or particular genes (or regions) being deleted in one twin and present in the other (Figure 1B). We have confirmed differences in “driver” CNAs in twin pair 5 by multicolor FISH (Figure 2). Insufficient cells were available from the other 4 twin pairs.

Some prior data have indicated that deletion of the untranslocated ETV6 can be clonal or secondary to ETV6-RUNX1 fusion.11,12 The present study suggests that this is both a consistent feature of ETV6 deletions (9/9 deletions detected here) and, more significantly, a feature of all other CNAs in ALL (23/23 deletions or gains detected here in the 5 pairs of twins). Some discordance of genetics in leukemias can be deduced from chromosome karyotype alone.13-15 However, in the present series, 18 of the 23 non-ETV6 “driver” CNAs were invisible by karyotype (Table 2 and supplemental Table 1).

Monozygotic twins discordant for ETV6-RUNX1-positive ALL

We sought to further confirm these findings with a twin pair, previously reported,2 who were discordant for ALL (Figure 3). Twin 6a had ALL with ETV6-RUNX1 and, as identified by SNP arrays (supplemental Table 4), multiple CNAs with potential “drivers” including ETV6, PAX5 deletion, and 10p gain (Figure 3B). Her twin sister was healthy but had a persistent (over 3 years) but clinically covert preleukemic population with ETV6-RUNX1, reflecting their common or single-cell origin in utero.1,10 When these preleukemic cells (defined by positivity for ETV6-RUNX1 using FISH) were interrogated by multicolor FISH for the status of the normal (nontranslocated) ETV6 allele, PAX5 genes, and 10p, no CNA was detected in any of the cells assessed (Figure 3C). We cannot exclude that these ETV6-RUNX1 preleukemic cells harbor these CNAs at low frequencies or indeed have other genetic abnormalities, but these data further support the contention that recurrent and presumed functional gene deletions in ALL arise secondarily to gene fusion.

Discussion

ETV6-RUNX1 fusion is usually a prenatal, acquired genetic event in ALL. This is formally demonstrated by the sharing of identical ETV6-RUNX1 genomic sequences in all 5 monozygotic twin pairs previously assessed10,13 (2 of these pairs, numbers 1 and 3 being included in the current series), and the presence of these clonotypic sequences in the archived, neonatal blood spots of patients with ALL.1,18 This conclusion is further endorsed by the identification of fusion gene positive cells in unselected cord bloods.19 The fact that newborn cord bloods score positive for such putative preleukemic populations (with clinically undetectable frequencies in the lymphocyte fraction of 10−3 to 10−4) at rates substantially higher (~100×) than the cumulative risk of childhood ALL with ETV6-RUNX1 indicates the obligatory requirement for further, postnatal genetic changes or mutations. The concordance rate of ALL in monozygotic twin children, at 10% to 15%, also supports this contention, as does modeling with murine and human cells.2-4

ALL occurring in the twin context is no different in terms of cellular phenotype, molecular genetics and clinical features to ALL in non-twinned singletons.10 This includes the common prenatal origin of the ETV6-RUNX1 fusion1,18 and patterns of CNAs detected by SNP arrays (this study vs Mullighan et al,2 Strefford et al,4 Lilliebjo¨r ne et al,7 and Tsuzuki et al18). It is therefore highly likely that the sequence of genetic events we observe in twins applies more generally to childhood ALL.

These data suggest the following sequence of mutational events in the covert, evolutionary history of childhood B-cell precursor ALL. First, and as previously shown,1 an early and likely initiating (or first “hit”) gene fusion (ETV6-RUNX1) occurring prenatally,
being sufficient to spawn clonal progeny, including preleukemic stem cells and as endorsed by modeling experiments with human cord blood progenitor cells. Second, the acquisition of multiple genetic alterations including (but possibly not restricted to) the CNAs described here and in prior studies. The latter might be expected to drive the evolution of overt leukemic stem cells and to culminate in a clinical diagnosis of ALL. The precise timing of these secondary CNAs has not been determined. They could occur proximal to a diagnosis, perhaps linked to environmental, exposure-based promotion of disease. The lack of concordance of all “driver” CNAs identified in the leukemic cells of 6 pairs of twins is most compatible with these genetic events occurring postnatally. Formally, there is a possibility that one or more of them could have arisen prenatally albeit secondary to $ETV6$-$RUNX1$ fusion and therefore in a subclone. ALL clearly does have significant subclonal diversity as indicated by long-standing karyotypic observations and more recent SNP array derived data. However, twin blood chimaerism should result in all subclones generated before birth being shared also. Subsequent discordance of all CNAs in twin pairs would therefore require differential subclone selection by further, postnatal “driver” events. Retrospective screening of serial blood samples of a healthy co-twin that eventually evolves from preleukemia to overt ALL might shed light on the CNAs timing issue. That they can arise significantly before a diagnosis of ALL as evidenced by a case of $ETV6$-$RUNX1$–positive ALL with multiple CNAs present in an aplastic phase preceding overt ALL, by some 7 months. The conclusion remains that all potentially or functional CNAs in ALL appear to be secondary to the likely initiating lesion, $ETV6$-$RUNX1$.

Complete genomic sequencing will be required to determine whether other mutational events, such as balanced translocations or small mutations not detectable by SNP array, contribute to this pattern of clonal evolution in childhood ALL.

Acknowledgments

We would like to thank the patients and families involved in this study.

This leukemia research was funded by a Leukaemia Lymphoma Research (London) clinical fellowship for C.M.B., a program grant for M.F.G., and Cancer Research UK for funding T.C. and B.D.Y.

Authorship

Contribution: C.M.B. carried out the majority of the experimental work, compiled data, and contributed to writing the paper; S.M.C. and T.C. carried out experimental work; B.D.Y. provided expertise in SNP array data analysis; T.O.E., M.B., E.J.G., C.J.H., R.H., and PA. provided patient samples (cell and/or DNA) and patient diagnostic information; A.M.F. carried out experimental work and provided expertise in molecular techniques; L.K. helped manage the study, supervised FISH studies, and assisted in writing the paper; and M.G. conceived the study, acquired twin leukemia cell DNA samples, helped analyze the data, and drafted the paper.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

Correspondence: Prof Mel Greaves, FRSt, Section of Haematological Oncology, The Institute of Cancer Research, Brooks Lawley Bldg, 15 Cotswold Rd, Belmont, Sutton, Surrey SM2 5NG, United Kingdom; e-mail: mel.greaves@icr.ac.uk.

References

duplications and identify several putative target genes within the commonly gained region. Leukemia. 2007;21(10):2137-2144.


Acquisition of genome-wide copy number alterations in monozygotic twins with acute lymphoblastic leukemia

Caroline M. Bateman, Susan M. Colman, Tracy Chaplin, Bryan D. Young, Tim O. Eden, Manoo Bhakta, Eric J. Gratias, Elisabeth R. van Wering, Giovanni Cazzaniga, Christine J. Harrison, Richard Hain, Philip Ancliff, Anthony M. Ford, Lyndal Kearney and Mel Greaves