npg

ORIGINAL ARTICLE

Clonal origins of *ETV6-RUNX1*⁺ acute lymphoblastic leukemia: studies in monozygotic twins

D Alpar[1,2], D Wren[3], L Ermini[4], MB Mansur[1,5], FW van Delft[1], CM Bateman[1], I Titley[1], L Kearney[1], T Szczepanski[6], D Gonzalez[3], AM Ford[1,7], NE Potter[1,7] and M Greaves[1,7]

Studies on twins with concordant acute lymphoblastic leukemia (ALL) have revealed that *ETV6-RUNX1* gene fusion is a common, prenatal genetic event with other driver aberrations occurring subclonally and probably postnatally. The fetal cell type that is transformed by *ETV6-RUNX1* is not identified by such studies or by the analysis of early B-cell lineage phenotype of derived progeny. Ongoing, clonal immunoglobulin (*IG*) and cross-lineage T-cell receptor (*TCR*) gene rearrangements are features of B-cell precursor leukemia and commence at the pro-B-cell stage of normal B-cell lineage development. We reasoned that shared clonal rearrangements of *IG* or *TCR* genes by concordant ALL in twins would be informative about the fetal cell type in which clonal advantage is elicited by *ETV6-RUNX1*. Five pairs of twins were analyzed for all varieties of *IG* and *TCR* gene rearrangements. All pairs showed identical incomplete or complete variable-diversity-joining junctions coupled with substantial, subclonal and divergent rearrangements. This pattern was endorsed by single-cell genetic scrutiny in one twin pair. Our data suggest that the pre-leukemic initiating function of *ETV6-RUNX1* fusion is associated with clonal expansion early in the fetal B-cell lineage.

INTRODUCTION

It is assumed that most cancers evolve from a single cell that has acquired a mutation, but in few clinical cases there is definitive identification of the founder mutation or the precise cell of origin. Childhood acute lymphoblastic leukemia (ALL) has multiple genetically and clinically distinct subtypes but the two most common variants have either a t(12;21) chromosomal translocation, which results in a chimeric *ETV6-RUNX1* fusion gene, or chromosomal hyperdiploidy.[1] There is strong evidence that the former is a recurrent founder mutation. Monozygotic twins concordant for ALL were found to share the fusion gene break point sequence, suggesting that a common pre-leukemic clone was spawned *in utero* and shared between the twins.[2] The blood chimeras that developed between these twins are a consequence of a shared single placenta and intraplacental vascular anastomoses.[3] Archived neonatal blood spots of patients with ALL were also found to be positive for the gene fusion confirming *ETV6-RUNX1* as an early event in the pathogenesis of this disease.[4] More recent single-cell analyses showed the presence of the gene fusion in every leukemic cell.[5,6] The fusion gene is the only acquired, 'driver' (recurrent) mutation in the coding region in *ETV6-RUNX1*⁺ ALLs developed in monozygotic twins[7,8] and at relapse, the *ETV6-RUNX1*⁺ genomic fusion sequence is always preserved, whereas copy number abnormalities change.[9] Murine models with human precursor cells harboring the *ETV6-RUNX1* fusion gene generated a pre-leukemic state that only resulted in an overt leukemic phenotype upon the acquisition of additional genetic abnormalities.[10–14]

ETV6-RUNX1⁺ ALLs can be characterized by a CD19⁺ B-cell lineage precursor immunophenotype with ongoing recombinase-associated *IGH* rearrangements.[15] These features reflect the predominant level of differentiation arrest in the B-cell lineage but not necessarily the cell of origin in which the initial functional impact of the fusion gene occurs. This could lie anywhere antecedent to B precursor cells in the lineage hierarchy. Experiments with mouse and human stem and progenitor cells suggest that *ETV6-RUNX1* may arise in a multi-lineage stem cell as a weak oncogene but transforms or impacts more profoundly on early pro-B cells. Models are, however, by their nature artifactual and may not definitively define the cell of origin in a clinical context.

Sequential rearrangements at the multi-gene segment loci *IGH*, *IGK*, *IGL* and *TCRA/D*, *G* and *B* are unique to the B-cell and T-cell lineages, respectively, and are initiated by DH to JH segment rearrangements in pro-B-cells.[16,17] These markers are present in >90% of ALL samples[18,19] but cannot in themselves define the cell of origin as they are primarily products of extensive and continuing RAG-driven clonal diversification in downstream progeny of the cell that was originally transformed.[20,21] However, we reasoned that the clonal status of immunoglobulin/T-cell receptor (*IG/TCR*) rearrangements in the context of monozygotic twins concordant for *ETV6-RUNX1*⁺ ALL could be indicative of the fetal cell of origin that initially expanded. Although, *IG/TCR* rearrangements are ongoing in ALL, including replacement variable-diversity-joining (V(D)J) switches, any diversity-joining (DJ) or V(D)J rearrangements shared by a twin pair would be indicative of the *IG/TCR* status of the cell giving rise to pre-leukemic clones. Limited screening of *IG/TCR* rearrangements in

[1]Centre for Evolution and Cancer, The Institute of Cancer Research—London, London, UK; [2]Department of Pathology, University of Pecs, Pecs, Hungary; [3]Haemato-Oncology Research Unit, Division of Molecular Pathology, The Institute of Cancer Research—London, London, UK; [4]Centre for Geogenetics, Natural History Museum of Denmark, University of Copenhagen, Copenhagen K, Denmark; [5]Pediatric Hematology-Oncology Program, Instituto Nacional de Câncer, Rio de Janeiro, Brazil and [6]Department of Pediatric Hematology and Oncology, Zabrze, Medical University of Silesia, Katowice, Poland. Correspondence: Professor M Greaves or Dr D Alpar, Centre for Evolution and Cancer, Division of Molecular Pathology, The Institute of Cancer Research—London, Brookes Lawley Bldg, 15 Cotswold Rd, Belmont, Sutton SM2 5NG, Surrey, UK.
E-mail: mel.greaves@icr.ac.uk or donat.alpar@icr.ac.uk
[7]These authors share senior authorship.
Received 7 October 2014; accepted 10 October 2014; accepted article preview online 12 November 2014; advance online publication, 16 January 2015

twin pairs has been previously reported.[2,22,23] We have carried out a detailed and comprehensive screen of *IG/TCR* rearrangements in five twin pairs with concordant *ETV6-RUNX1*⁺ ALL. For one twin pair, we expanded the analysis incorporating single leukemic cells and included genomic mutations and copy number alterations (CNAs) common to the leukemias of each twin. Phylogenetic analysis combined all data describing disease evolution in this twin pair.

MATERIALS AND METHODS

Patient samples

Diagnostic bone marrow material was available from five monozygotic, monochorionic twin pairs concordant for pediatric *ETV6-RUNX1*⁺ precursor B-cell ALL. DNA was available from all patients with additional methanol/acetic acid-fixed cells available from one twin pair. Clinical and pathological information along with SNP-array data for all patients has previously been described.[7] For one twin pair (3A/B), whole-genome sequencing data were also published.[8] Ethical Committee approval from the Royal Marsden NHS Trust and written informed consent from the patients' parents have been obtained for the study, which was conducted in accordance with the principles expressed in the Declaration of Helsinki.

IG/TCR gene rearrangement screening

Standard screening. Our standard *IG/TCR* screening approach has recently been described in detail.[24] In brief, for all cases, polymerase chain reaction (PCR) amplifications of *IG* heavy chain variable-diversity-joining (*IGH* V(D)J; complete and incomplete), *IG* kappa variable-joining (IGK VJ), *IGK* V-kappa-deleting element (Kde), intron recombination signal sequence-Kde, *IG* lambda (*IGL*), *TCR* delta (*TCRD*), *TCR* gamma (*TCRG*) and *TCR* beta (*TCRB*) gene rearrangements were performed using primer mixes and conditions recommended by the BIOMED-2 Consortium.[25,26] Clonality was assessed by GeneScan profiling (Applied Biosystems, Paisley, UK)[27] and in the case of positivity, monoplex PCR reactions were performed followed by cloning of the products (pCR2.1 TA Cloning kit; Life Technologies, Paisley, UK or pMOS*Blue* Blunt-Ended Cloning Kit; GE Healthcare Life Sciences, Amersham, UK) and Sanger sequencing of randomly picked colonies using BigDye chemistry. Sequences were aligned by basic local alignment search tool search to the human germ-line sequences deposited in GenBank and junction analyses were performed using the Ig basic local alignment search tool (www.ncbi.nlm.nih.gov/igblast/) and/or the ImMunoGeneTics tools (www.imgt.org).

Targeted next-generation sequencing. We developed a targeted capture approach using the NimbleGen system (Roche, Burgess Hill, UK) by tiling DNA baits across the V, D and J segments of the *IGH*, *IGK*, *IGL*, *TCRD*, *TCRG* and *TCRB* loci. This method avoids potential bias that can be caused by the initial PCR and cloning steps in standard screening methods. Genomic regions of interest were assayed using 100 ng DNA from six patients (twin pairs 2, 3 and 4) and a previously described capture protocol.[28] Samples were barcoded using Illumina indexes, multiplexed and sequenced on a MiSeq platform generating 150-base pair paired-end reads. After base calling and quality control metrics, fastq reads were aligned to the reference human genome (build GRCh37) using BWA 0.6.2 and Stampy 1.0.20. On average, a median depth of $325 \times$ was achieved in the genomic regions of interest across the six diagnostic samples analyzed. In addition to manual inspection of the whole data set in IGV, identification of clonal complete and incomplete *IG/TCR* rearrangements was also performed with a combination of Samtools and ImMunoGeneTics tools. Rare variants were considered to be true if the rearrangement was present at least three times independently in both forward and reverse high-quality reads.

Cross-test of IG/TCR gene rearrangements with allele-specific Taqman PCR

Each clonotypic rearrangement detected in only one twin (with either screening method) was screened in his/her co-twin by Taqman PCR using custom allele-specific assays, covering the unique junctional regions; assays were designed using the Primer Express v3.0 software (Applied Biosystems). Custom assay probes were labeled with FAM reporter dye and a commercial VIC-labeled Taqman RNase P copy number assay (Applied Biosystems) was used as endogenous reference. Optimal thermocycling conditions were established and reactions were performed in duplicates or

triplicates depending on the amount of sample available. All experiments were performed using an ABI 7900HT Fast Real-Time Q-PCR System equipped with SDS 2.2.2 software (Applied Biosystems). Positivity was interpreted according to the ESG-MRD-ALL guidelines[29] and the specificity of each Taqman PCR reaction was confirmed by cloning and sequencing the products using the methods described above.

Single-cell genotyping

Single-cell genetic analysis was performed for twin pair 3 using available methanol/acetic acid-fixed cells and our previously established multiplex quantitative-PCR (Q-PCR) approach.[6] with minor modifications.[6] In brief, fixed propidium-iodide-labeled single cells were sorted into individual wells of 96-well plates by fluorescently activated cell sorting (BDFACSAria I SORP). Eleven control cord blood cells were included in each plate. After cell lysis (G-Storm GS4 Multi Block Thermal Cycler) multiplex target-specific DNA amplification was completed using custom Taqman assays (designed with Primer Express v3.0) or commercially available DNA copy number assays (Life Technologies) to amplify regions of interest. *IG/TCR* rearrangements, the *ETV6-RUNX1* gene fusion, *ETV6* and *RUNX1* gene CNAs and nine coding-region somatic nucleotide variants (SNVs) (previously identified by whole-genome sequencing[8]) were analyzed, simultaneously. Amplified samples (diluted 1:6) and assays were then loaded into a 96.96 dynamic array (Fluidigm, San Francisco, CA, USA) and the final detection Q-PCR reaction was completed using the BioMark HD System (Fluidigm). Assays aiming to detect CNAs were completed in quadruplicates, whereas all remaining assays were used in duplicates. The gene fusion, CNA, SNV and *IG/TCR* status for each single cell was determined as described previously[6] and the clonal phylogeny from these single-cell data was inferred using the Maximum Parsimony approach.[6,30] Data were assembled using an in-house Perl script and all parsimony analyses were performed using the computer software PAUP* version 4.0b10 for Linux.[6] Trees were visualized using Dendroscope Software version 3. Owing to limitations arising from the junctional nucleotide sequence, it was impossible to include an assay for detection of the *IGH*D3-10/J4 junction into the Fluidigm microfluidic multiplex Q-PCR experiment. This rearrangement was found to be present in both patients of twin pair 3; thus, we performed an additional single-cell experiment using an ABI 7900HT fast real-time Q-PCR System and analyzed this incomplete *IGH* junction together with the *ETV6-RUNX1* gene fusion and other previously identified *IG* rearrangements in 60–60 cells of twins 3A and 3B, respectively.

RESULTS

IG/TCR gene rearrangement profiling

Standard screening for clonal *IG* and *TCR* gene rearrangements was performed on the diagnostic DNA of all 10 individual patients by multiplex PCR reactions and ABI GeneScan profiling. Thirty-seven reactions detected clonalities with the following distribution: *IGH* DJ (#0), *IGH* V(D)J (#10), IGK VJ (#3), *IGK* Kde (#8), *IGL* (#4), *TCRD* (#4), *TCRB* (#2) and *TCRG* (#6) (Supplementary Table 1). In each patient, at least two affected genes were found. We next performed monoplex PCR reactions to scrutinize all loci that showed clonal rearrangements. Amplified products were cloned, random colonies picked and directly sequenced followed by junction analyses using the Ig basic local alignment search tool and ImMunoGeneTics tools. In total, > 670 bidirectional sequencing reactions were carried out and 65 rearrangements were identified in the five twin pairs.

Targeted next-generation sequencing was successfully performed for twin pairs 2, 3 and 4. On average, a median depth of $325 \times$ was achieved in the genomic regions of interest (Supplementary Table 2). In total, 42 genuine *IG/TCR* gene rearrangements present in at least three independent reads were identified in twin pairs 2A/B, 3A/B and 4A/B.

All *IG/TCR* rearrangements detected in the five twin pairs using our screening approaches are summarized in Supplementary Table 3. Identical *IGH* V(D)J and *IGK* Kde rearrangements were observed in the siblings of twin pairs 2A/B and 3A/B, whereas twins 4A/B shared an identical *TCRD* Vδ2-Dδ3 junction.

In order to further refine the immunogenotypes of twins in the context of their sibling's profile, clonal rearrangements detected in

only one patient were screened in his/her sibling by allele-specific Taqman PCR. For *IGH* gene rearrangements, complete V(D)J and incomplete DJ junctions were investigated separately. In a subset of rearrangements, it was not possible to design sensitive allele-specific assays due to the sub-optimal sequence at the junctions; a technical limitation of this approach that is well-known from previous studies.[19] With these sibling cross-tests, we were able to identify three additional rearrangements with identical junctions common to twins 1A/B (*TCRD* Vδ2-Dδ3), 3A/B (*IGH* DJ) and 5A/B (*IGH* DJ) (Supplementary Table 4). The number and pattern of all rearrangements detected in our twin cohort are summarized in Figure 1.

Previous bulk DNA SNP-array analysis of these twin pairs confirmed that all samples harbor diploid genomes[7] allowing oligoclonality to be interpreted on a background of two rearranged allelic variants of the *IG* or *TCR* genes. Oligoclonality was observed by analyzing the *IGH* gene in patients 1A, 2A, 2B, 3A, 3B, 4B and 5B. In patients 2A and 2B, V-replacement on one allele produced a third rearrangement, whereas in patients 1A, 3A, 3B and 5B identical intra-sample DJ junctions were not observed suggesting a post-oncogeneic DJ recombination event, generating multiple rearrangements. In case 4B, both an oligoclonal *IGH* DJ constitution and V(D)J junctions suggesting V-replacement activity could be detected. Oligoclonal *IGK* repertoires including three different Kde junctions were observed in cases 3B and 5B. In patients 1A and 5A, more than two *IGK* rearrangements were identified; however, monoclonality cannot be excluded in these cases as the rearrangements include an IGK VJ and an intron recombination signal sequence-Kde that can be located on the same allele.[31]

Among the cross-lineage rearrangements, *TCRD* showed oligoclonality in patients 1B, 4A and 4B. In case 4A, the two Vδ2-Jα29 junctions comprised two independently detected Vδ2-Dδ3 joints, indicating a direct clonal relationship. Identical Vδ2-Dδ3 joints were not found in case 1B, suggesting that the various rearrangements are located on different alleles or possibly

found in different cells. In patient 4B, aside from clonally related Vδ2-Dδ3 and Vδ2-Jα29 junctions, three unrelated *TCRD* rearrangements were detected. In both twin pairs 1A/B and 4A/B, identical Vδ2-Dδ3 rearrangements were shared by the siblings coupled with *TCRD* oligoclonalities mentioned above. These data demonstrate that both the B-cell lineage-specific *IG* genes and the cross-lineage *TCRD* locus may be accessible for recombination pre- and postnatally during leukemia development.

All clonal rearrangements detected in our twin pairs are presented in the context of pre- and postnatal leukemogenesis in Figures 2a–e. In each pair, we observed shared identical clonal *IG/TCR* rearrangement(s) in addition to the founder lesion *ETV6-RUNX1*, and therefore these rearrangements must have occurred *in utero*. The vast majority of clonal markers proved to be patient-specific and are likely to have arisen postnatally in patient-specific subclones.
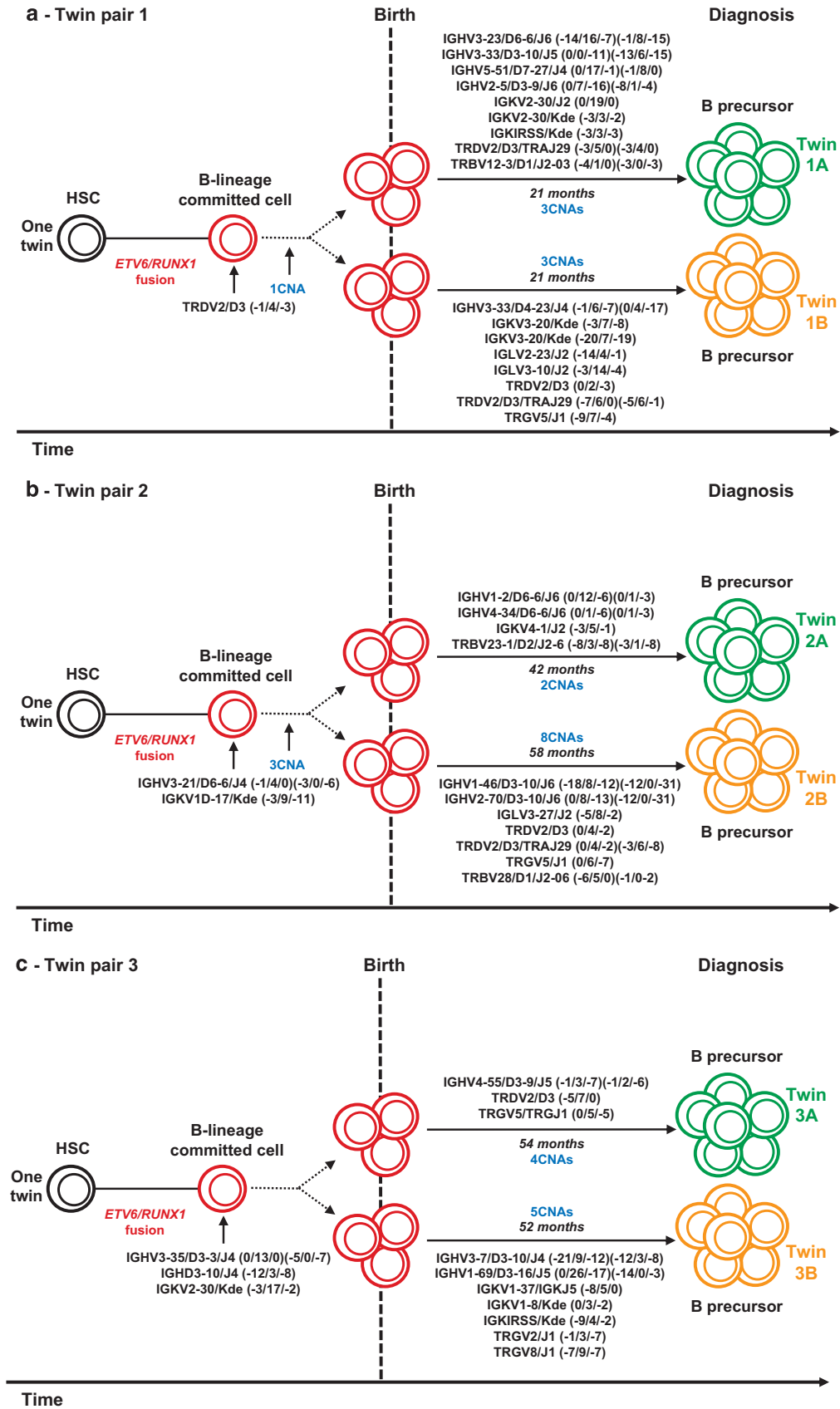
Single-cell genotyping

Archived diagnostic leukemia cells from twin pair 3 were subjected to unbiased high-throughput single-cell mulitplex Q-PCR analysis that allowed the simultaneous investigation of previously defined gene fusions, CNAs, SNVs and *IG/TCR* gene rearrangements for each leukemia. The gene targets included the *ETV6-RUNX1* gene fusion, *ETV6* and *RUNX1* copy numbers, nine coding-region SNVs and eight *IG/TCR* rearrangements. A complete list of genetic targets analyzed and assays applied can be found in Supplementary Table 5.

After quality-control assessments and analysis as previously described[6] 111 and 110 leukemic cells provided reliable data for patients 3A and 3B, respectively. Five subclones with 9–10 genetic markers and eight subclones with 7–15 markers were identified in patients 3A and 3B, respectively. The most likely phylogenetic history and relationships between the subclones found in each twin is shown in Supplementary Figures 1 and 2. The phylogenetic trees are branching and complex in both leukemias but share a common ancestry.
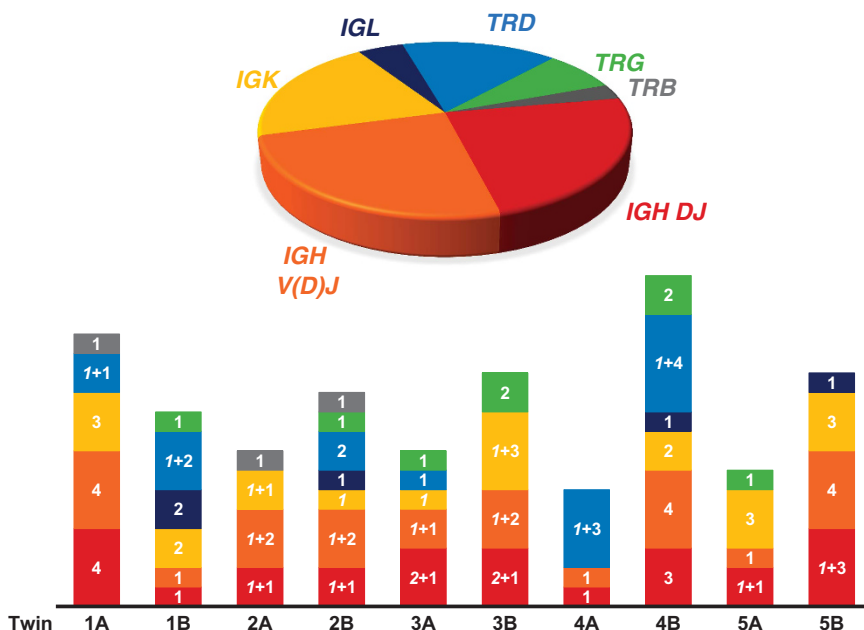
Figure 1. Pattern of immunoglobulin (*IG*) and T-cell receptor (*TCR*) gene rearrangements detected in the five monozygotic twin pairs with concordant *ETV6-RUNX1*[+] ALL. The figure includes all rearrangements identified by standard screening (Sanger sequencing), targeted next-generation sequencing or allele-specific Taqman PCR cross-tests between siblings. The number of identical rearrangements shared by the siblings of each twin pair is highlighted in italics. Additional alterations found in each individual are indicated with normal numbers. The incomplete *IGH* DJ rearrangements shown here belong to the complete *IGH* V(D)J rearrangements indicated in the same twin pair.

a - Twin pair 1

Birth

Diagnosis

IGHV3-23/D6-6/J6 (-14/16/-7)(-1/8/-15)
IGHV3-33/D3-10/J5 (0/0/-11)(-13/6/-15)
IGHV5-51/D7-27/J4 (0/17/-1)(-1/8/0)
IGHV2-5/D3-9/J6 (0/7/-16)(-8/1/-4)
IGKV2-30/J2 (0/19/0)
IGKV2-30/Kde (-3/3/-2)
IGKIRSS/Kde (-3/3/-3)
TRDV2/D3/TRAJ29 (-3/5/0)(-3/4/0)
TRBV12-3/D1/J2-03 (-4/1/0)(-3/0/-3)

B precursor

Twin
1A

HSC

B-lineage
committed cell

One
twin

ETV6/RUNX1
fusion

1CNA

TRDV2/D3 (-1/4/-3)

21 months
3CNAs

3CNAs
21 months

IGHV3-33/D4-23/J4 (-1/6/-7)(0/4/-17)
IGKV3-20/Kde (-3/7/-8)
IGKV3-20/Kde (-20/7/-19)
IGLV2-23/J2 (-14/4/-1)
IGLV3-10/J2 (-3/14/-4)
TRDV2/D3 (0/2/-3)
TRDV2/D3/TRAJ29 (-7/6/0)(-5/6/-1)
TRGV5/J1 (-9/7/-4)

B precursor

Twin
1B

Time

b - Twin pair 2

Birth

Diagnosis

IGHV1-2/D6-6/J6 (0/12/-6)(0/1/-3)
IGHV4-34/D6-6/J6 (0/1/-6)(0/1/-3)
IGKV4-1/J2 (-3/5/-1)
TRBV23-1/D2/J2-6 (-8/3/-8)(-3/1/-8)

B precursor

Twin
2A

HSC

B-lineage
committed cell

One
twin

ETV6/RUNX1
fusion

3CNA

IGHV3-21/D6-6/J4 (-1/4/0)(-3/0/-6)
IGKV1D-17/Kde (-3/9/-11)

42 months
2CNAs

8CNAs
58 months

IGHV1-46/D3-10/J6 (-18/8/-12)(-12/0/-31)
IGHV2-70/D3-10/J6 (0/8/-13)(-12/0/-31)
IGLV3-27/J2 (-5/8/-2)
TRDV2/D3 (0/4/-2)
TRDV2/D3/TRAJ29 (0/4/-2)(-3/6/-8)
TRGV5/J1 (0/6/-7)
TRBV28/D1/J2-06 (-6/5/0)(-1/0-2)

B precursor

Twin
2B

Time

c - Twin pair 3

Birth

Diagnosis

IGHV4-55/D3-9/J5 (-1/3/-7)(-1/2/-6)
TRDV2/D3 (-5/7/0)
TRGV5/TRGJ1 (0/5/-5)

B precursor

Twin
3A

HSC

B-lineage
committed cell

One
twin

ETV6/RUNX1
fusion

IGHV3-35/D3-3/J4 (0/13/0)(-5/0/-7)
IGHD3-10/J4 (-12/3/-8)
IGKV2-30/Kde (-3/17/-2)

54 months
4CNAs

5CNAs
52 months

IGHV3-7/D3-10/J4 (-21/9/-12)(-12/3/-8)
IGHV1-69/D3-16/J5 (0/26/-17)(-14/0/-3)
IGKV1-37/IGKJ5 (-8/5/0)
IGKV1-8/Kde (0/3/-2)
IGKIRSS/Kde (-9/4/-2)
TRGV2/J1 (-1/3/-7)
TRGV8/J1 (-7/9/-7)

B precursor

Twin
3B

Time

Figure 2. Postulated evolutionary sequence of clonal immunogenotypic markers identified in each twin pair (**a**–**e**: twin pairs 1–5). Rearrangements are depicted in the context of *in utero* origin and diverging postnatal evolution of the leukemia in twin siblings. Copy number abnormality (CNA) data have been adopted from Bateman *et al*.[7] in which the leukemia from same twin pairs was examined. HSC: hematopoietic stem cell.
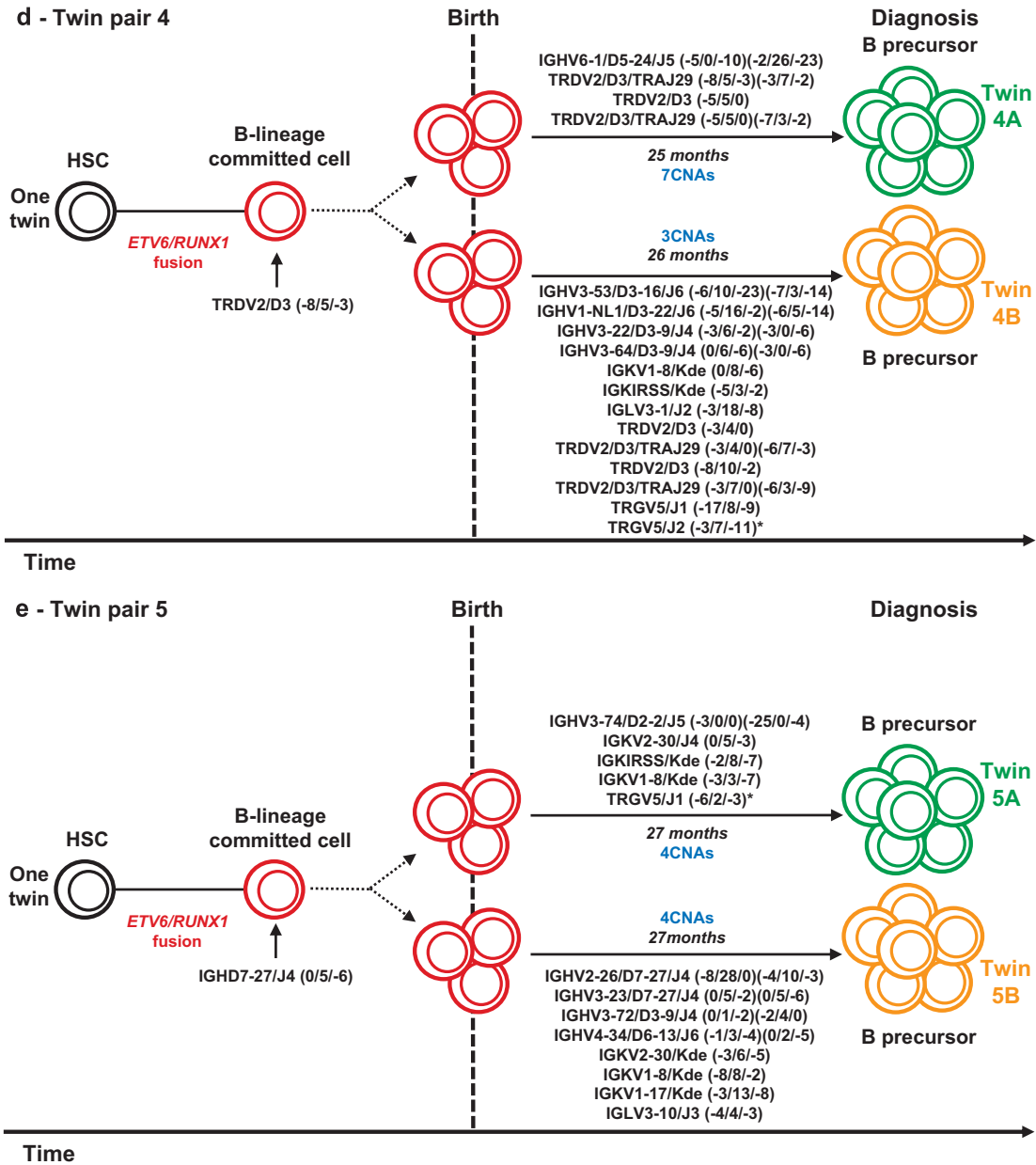
d - Twin pair 4

Birth

Diagnosis
B precursor

HSC

B-lineage committed cell

One twin

ETV6/RUNX1 fusion

TRDV2/D3 (-8/5/-3)

IGHV6-1/D5-24/J5 (-5/0/-10)(-2/26/-23)
TRDV2/D3/TRAJ29 (-8/5/-3)(-3/7/-2)
TRDV2/D3 (-5/5/0)
TRDV2/D3/TRAJ29 -5/5/0)(-7/3/-2)

25 months
7CNAs

Twin 4A

3CNAs
26 months

IGHV3-53/D3-16/J6 (-6/10/-23)(-7/3/-14)
IGHV1-NL1/D3-22/J6 (-5/16/-2)(-6/5/-14)
IGHV3-22/D3-9/J4 (-3/6/-2)(-3/0/-6)
IGHV3-64/D3-9/J4 (0/6/-6)(-3/0/-6)
IGKV1-8/Kde (0/8/-6)
IGKIRSS/Kde (-5/3/-2)
IGLV3-1/J2 (-3/18/-8)
TRDV2/D3 (-3/4/0)
TRDV2/D3/TRAJ29 (-3/4/0)(-6/7/-3)
TRDV2/D3 (-8/10/-2)
TRDV2/D3/TRAJ29 (-3/7/0)(-6/3/-9)
TRGV5/J1 (-17/8/-9)
TRGV5/J2 (-3/7/-11)*

Twin 4B

B precursor

Time

e - Twin pair 5

Birth

Diagnosis

IGHV3-74/D2-2/J5 (-3/0/0)(-25/0/-4)
IGKV2-30/J4 (0/5/-3)
IGKIRSS/Kde (-2/8/-7)
IGKV1-8/Kde (-3/3/-7)
TRGV5/J1 (-6/2/-3)*

B precursor

HSC

B-lineage committed cell

One twin

ETV6/RUNX1 fusion

IGHD7-27/J4 (0/5/-6)

Twin 5A

27 months
4CNAs

4CNAs
27months

IGHV2-26/D7-27/J4 (-8/28/0)(-4/10/-3)
IGHV3-23/D7-27/J4 (0/5/-2)(0/5/-6)
IGHV3-72/D3-9/J4 (0/1/-2)(-2/4/0)
IGHV4-34/D6-13/J6 (-1/3/-4)(0/2/-5)
IGKV2-30/Kde (-3/6/-5)
IGKV1-8/Kde (-8/8/-2)
IGKV1-17/Kde (-3/13/-8)
IGLV3-10/J3 (-4/4/-3)

Twin 5B

B precursor

Time

Figure 2. (Continued)

For patient 3A, a subclonal architecture with relatively modest complexity was found (Supplementary Figure 1). In addition to the *ETV6-RUNX1* gene fusion, four somatic SNVs (three shared with the co-twin and one patient-specific) and an *IGK* Kde rearrangement were observed in every leukemic cell, suggesting that these alterations occurred early in leukemogenesis. Two complete *IGH* V(D)J junctions were also acquired early in the development of this leukemia; the disappearance of one of these rearrangements in 8% of the leukemic cells (subclones A4 and A5) is most likely due to the ongoing RAG enzyme-driven recombination activity replacing parts of the V(D)J junction with upstream V segments. The subclonal position of the *TCRG* rearrangement and *ETV6* gene loss suggests that these events occurred later. *ETV6* gene loss is considered to be a potential secondary 'driver' alteration in pediatric *ETV6-RUNX1*+ ALL. The re-iterative loss of *ETV6*, twice in independent branches of the tree confirms convergent evolution and strengthens the notion that this gene is a secondary 'driver'.[5] Maximum parsimony algorithm predicted the most recent common ancestor from which all subclones of the phylogenetic tree descended.

Phylogenetic analysis of patient 3B defined a more complex tree than in the co-twin (Supplementary Figure 2). *ETV6-RUNX1* and the three identical somatic SNVs shared with patient 3A were observed in every leukemic cell. Patient-specific SNVs (five) and *IGK* VJ and *TCRG* rearrangements were detected in 97% of leukemic cells with an early acquired *IGH* V(D)J rearrangement (*IGH*V3-7/D3-10/J4) showing loss in minor subclones (B5 and B7—total 9%); presumably a later event due to ongoing V-replacement. Re-iterative loss of *ETV6* and gain of *RUNX1* were also common, supporting previous findings that CNAs in *ETV6-RUNX1*+ ALL can occur multiple times during leukemogenesis.[5] Subclone B8 harboring the highest number of genetic changes, including both *ETV6* loss and *RUNX1* gain, was not numerically dominant, albeit this result is based on analysis
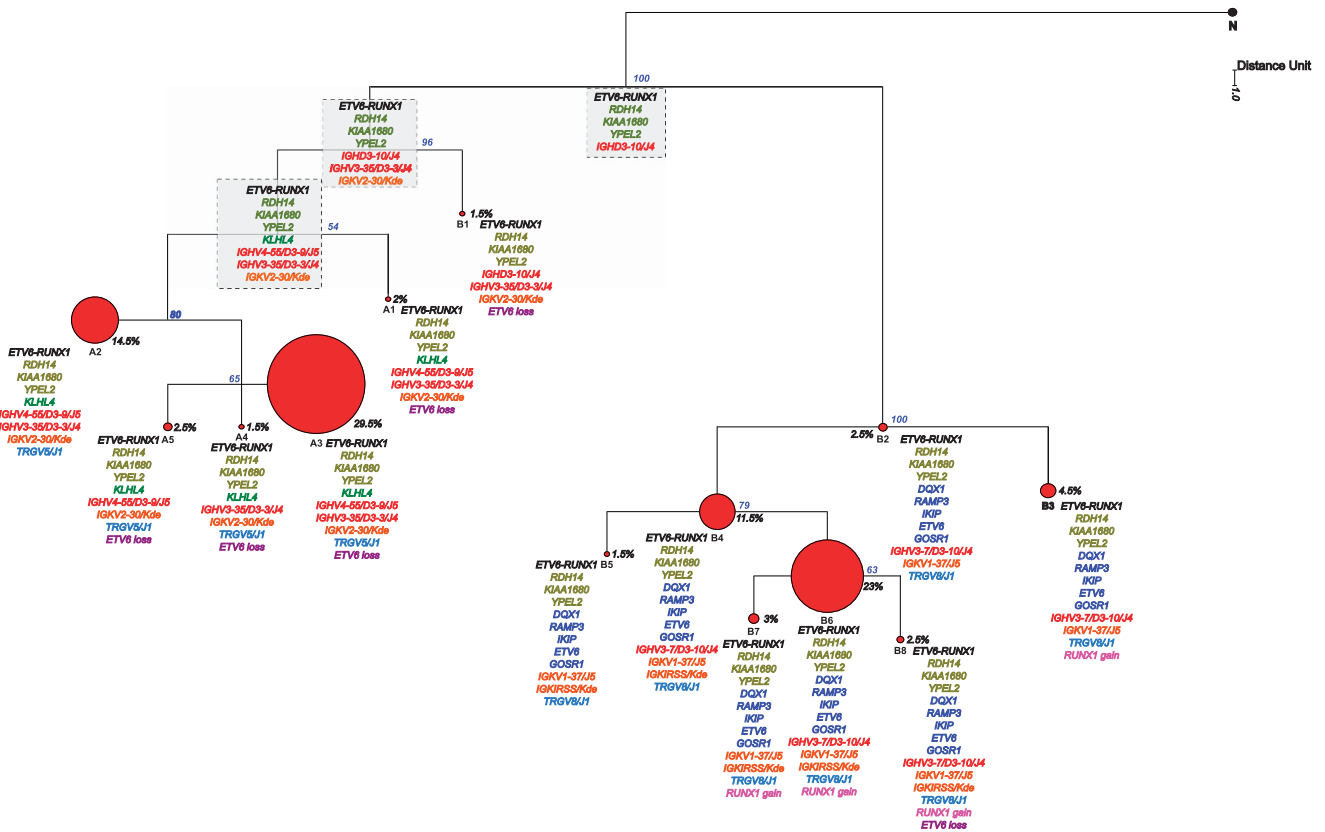
Figure 3. Combined phylogenetic history of *ETV6-RUNX1*[+] ALLs revealed in monozygotic twins by single-cell genotyping. The phylogenetic tree was generated based on subclones detected in twins 3A/B using maximum parsimonious analysis.[6] Leukemic subclones are represented by red circles and the normal state is indicated by a black circle. The size of the circle is proportional to the number of cells in each subclone and the detected genetic markers are listed below each circle. Inferred subclones are represented by gray boxes; these are groups of cells, which have died out, been outcompeted or if still present, exist at low frequencies below the level of reliable detection using our approach. Tree branch lengths are directly proportional to the number of evolutionary changes inferred and the points at which the branches diverge (nodes) represent the ancestor state of a clonal clade; a monophyletic group, which includes all descendants of the ancestor. The number in italics at each node indicates the jackknifing value. The phylogeny shows how the clonal expansion has evolved from a common ancestor toward the observed states. Branched subclonal architectures were observed in both twins 3A and 3B. Maximum parsimonious analysis has shown that subclone B1, detected in twin 3B, is the least evolved, earliest occurring subclone with a closer phylogenetic relationship to the subclones detected in co-twin 3A than its own subclonal counterparts. *IGH*D3-10/J4 was added to the phylogenetic tree based on a separate single-cell experiment and bulk DNA Taqman PCR analysis.

at a single time point. Again, phylogenetic analysis predicted an undetectable common ancestor from which all subclones were potentially derived.

A striking finding in this analysis was that patient 3B harbored a minor subclone (B1) with *IG* rearrangements that dominated the leukemia of co-twin 3A but none of the patient-specific SNVs or *IG/TCR* rearrangements common to the other subclones in twin 3B. Whereas this unique genetic pattern was found only in 3% of cells, the 11 markers that distinguished subclone B1 from the closest clonally related subclone (B2) provide strong evidence that this is a true population of cells.

In the additional single-cell experiment performed using an ABI 7900HT fast real-time Q-PCR System, all analyzed *ETV6-RUNX1*[+] leukemic cells of patient 3A were found to be positive for the *IGK*V2-30/Kde and negative for the *IGH*D3-10/J4 rearrangement. Ninety percent of these cells harbored both complete *IGH* rearrangements analyzed (*IGH*V4-55/D3-9/J5 and *IGH*V3-35/D3-3/J4), whereas 10% were positive only for one of those; this proportion is comparable to the combined size of subclones A4 and A5 (8% in total) detected in the Fluidigm microfluidic experiment. In twin 3B, all *ETV6-RUNX1*[+] leukemic cells were positive for the incomplete *IGH*D3-10/J4 rearrangement and 96% of them were negative for all three complete or end-stage *IG*

rearrangements mentioned above. Four percent of cells harbored the *IGK*V2-30/Kde rearrangement previously detected in subclone B1 during the Fluidigm experiment.

The leukemias that arose in twins 3A/B have a common *in utero* origin, supported by the identical *ETV6-RUNX1* genomic fusion gene sequence and three somatic SNVs detected in both siblings. Thus, the phylogenetic trees derived for each sibling should be considered as diverging subclonal architectures of a single leukemia. We combined and analyzed all subclones detected in the twins to acquire a more comprehensive picture of leukemogenesis (Figure 3). This integrated parsimonious tree revealed a very close phylogenetic relationship between subclone B1 detected in patient 3B and the subclones observed in co-twin 3A. The inferred most recent common ancestor of this combined tree harbored the *ETV6-RUNX1* gene fusion, the three shared SNVs and an incomplete *IGH* DJ rearrangement; a profile characteristic of a cell at the pro-B-cell stage of development. Two of the three *IG* gene rearrangements found to be shared by the twins by bulk DNA analysis were present only in the minor subclone B1 in sibling 3B and not in all cells as initially presumed. The third *IG* rearrangement (*IGH* DJ) found in all leukemic 3B cells was detected in patient 3A only by bulk DNA analysis, suggesting its presence in a very small subclone. These results demonstrate that

immunogenotype diversification began before birth in this twin pair, with the formation of multiple *in utero* subclones, originating from the same pre-leukemic clone. Before birth, these subclones could pass from one twin to the other and contribute to the development of clinically overt leukemia.

DISCUSSION

ETV6-RUNX1⁺ ALLs can be characterized by phenotypes (CD19⁺, CD10⁺, TdT⁺ and RAG1/2⁺) and genotypes (ongoing *IG* rearrangements) indicative of differentiation arrest in the B-cell precursor developmental compartments.[32] This cancer usually originates *in utero* during fetal hemopoiesis with the *ETV6-RUNX1* gene fusion as the likely initiating genetic event.[33] The sequencing of the *ETV6-RUNX1* fusion region suggests that it arises as a consequence of non-homologous end-joining as there are no binding motifs indicative of RAG1/2 or TdT activity.[34–36] In contrast, recurrent secondary postnatal CNAs do have consistent signatures of RAG1/2 and TdT activity.[36] These data are compatible with the notion that *ETV6-RUNX1* fusion arises *in utero*, prior to the pro-B-cell stage in which RAGs are active, whereas secondary genetic hits impart significant clonal advantage and possibly self-renewal capacity to downstream B-cell precursors. Modeling data involving murine[12,37] and human[11] cells have also suggested a possible lympho-myeloid precursor or stem cell origin of the fusion gene. Yet, *ETV6-RUNX1*⁺ leukemias never present or relapse with a myeloid phenotype. This is in contrast to some *MLL-AF4* pro-B-cell ALLs[38] or pre-B-cell ALL blast crisis of prior *BCR-ABL1*⁺ CML[39] and *BCR-ABL1*⁺ ALL[15] that may derive from lympho-myeloid stem cells and can 'switch' lineages.

The investigation of bulk DNA from monozygotic, monochorionic twin pairs suggests that the pre-leukemic clone spawned *in utero* has *IG/TCR* rearrangements indicative of a pro- or pre-B-cell in all cases. In those cases, where only identical *TCR* gene rearrangements were shared by the siblings, it could theoretically be possible that the proliferation associated with pre-leukemic transformation started before B-lineage commitment. However, cross-lineage rearrangements in B-cell ALL are thought to occur owing to continuing recombinase activity after the clonal expansion in the B-lineage commenced[40] and our single-cell data, demonstrating the later occurrence of *TCRG* compared with early *IG* rearrangements, also support this concept. Scrutiny of the subclonal architecture at single-cell level in one of the twin pairs also provided evidence that the pre-leukemic transformation may occur at the pro-B-cell stage; even in cases where complete and end-stage *IG* rearrangements are shared by the twins.

The most parsimonious interpretation of these results is that the *ETV6-RUNX1* fusion arises in a fetal progenitor or stem cell that lies upstream of B-cell lineage-restricted RAG1/2 active precursors, but is either permissive only for B-cell lineage differentiation or only has a proliferative fitness impact on early B-cells. The pre-leukemic clone therefore arises and expands preferentially in the pro- or pre-B-lineage compartment, initially in one twin, possibly in the fetal liver and undergoes DJH and V(D)JH rearrangements. All or most clonally descendent cells with self-renewal stem cell activity are shared and sustained in both twins after birth as potential targets for secondary genetic hits essential for clinical development of ALL. This interpretation is compatible with our prior observation that identical twins discordant for ALL share a pre-leukemic stem cell population with a common DJH rearrangement.[11] The observation that all detectable *ETV6-RUNX1*⁺ cells in both pre-leukemic[11,41] and overtly leukemic populations[42,43] are CD19⁺ also indicates that if the fusion gene does indeed arise in a lympho-myeloid progenitor, then it imparts no clonal advantage to these cells—and may even totally block myeloid differentiation as opposed to partial differentiation arrest in the B-cell lineage.

It is striking that *ETV6-RUNX1*⁺ ALL is extremely rare after the age of 15 years and most of these cases diagnosed and assessed at ages up to 14 years, have a prenatal origin of the fusion gene.[33] One interpretation of this finding is that there is a progenitor target cell for *ETV6-RUNX1* transformation that is unique to fetal development—or a progenitor-tissue microenvironmental relationship that is restricted to this narrow prenatal time frame. A precedent for this exists with acute megakaryoblastic leukemia in patients with Down's syndrome, which may arise in a megakaryocytic/erythroid progenitor in fetal liver uniquely (compared with adult) dependent upon *GATA1* signals.[44]

In summary, we have performed the first comprehensive screen of *IG/TCR* rearrangements in a cohort of twins concordant for ALL, using an approach including massive parallel sequencing. We have also carried out the first in-depth scrutiny of subclonal immunogenotypic architecture in twins with ALL by a highly multiplexed combined analysis of *IG/TCR* rearrangements and three different types of somatic aberrations. Our results suggest that the pre-leukemic transformation conferring the *in utero* clonal expansion of *ETV6-RUNX1*⁺ cells occurs in an early B-cell lineage committed progenitor, most likely at the pro-B or possibly at the pre-B-cell stage. Further functional studies are needed to unveil the key mechanisms channeling the evolution of *ETV6-RUNX1*⁺ lympho-myeloid progenitor or stem cells into the B-lineage in this type of ALL.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

DA and MG designed the study. DA, DW, LE, MBM, FWvD, CMB, IT, LK, TS, DG, AMF, NEP and MG generated and/or analyzed the experimental data. DA and MG wrote the paper. AMF, NEP and MG supervised the study. All authors critically reviewed and approved the final draft of the manuscript.

REFERENCES

1 Inaba H, Greaves M, Mullighan CG. Acute lymphoblastic leukaemia. *Lancet* 2013; **381**: 1943–1955.
2 Ford AM, Bennett CA, Price CM, Bruin MC, Van Wering ER, Greaves M. Fetal origins of the TEL-AML1 fusion gene in identical twins with leukemia. *Proc Natl Acad Sci USA* 1998; **95**: 4584–4588.

3 Greaves MF, Maia AT, Wiemels JL, Ford AM. Leukemia in twins: lessons in natural history. *Blood* 2003; **102**: 2321–2333.

4 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**: 1499–1503.

5 Anderson K, Lutz C, van Delft FW, Bateman CM, Guo Y, Colman SM *et al.* Genetic variegation of clonal architecture and propagating cells in leukaemia. *Nature* 2011; **469**: 356–361.

6 Potter NE, Ermini L, Papaemmanuil E, Cazzaniga G, Vijayaraghavan G, Titley I *et al.* Single-cell mutational profiling and clonal phylogeny in cancer. *Genome Res* 2013; **23**: 2115–2125.

7 Bateman CM, Colman SM, Chaplin T, Young BD, Eden TO, Bhakta M *et al.* Acquisition of genome-wide copy number alterations in monozygotic twins with acute lymphoblastic leukemia. *Blood* 2010; **115**: 3553–3558.

8 Ma Y, Dobbins SE, Sherborne AL, Chubb D, Galbiati M, Cazzaniga G *et al.* Developmental timing of mutations revealed by whole-genome sequencing of twins with acute lymphoblastic leukemia. *Proc Natl Acad Sci USA* 2013; **110**: 7429–7433.

9 van Delft FW, Horsley S, Colman S, Anderson K, Bateman C, Kempski H *et al.* Clonal origins of relapse in ETV6-RUNX1 acute lymphoblastic leukemia. *Blood* 2011; **117**: 6247–6254.

10 Tsuzuki S, Seto M, Greaves M, Enver T. Modeling first-hit functions of the t(12;21) TEL-AML1 translocation in mice. *Proc Natl Acad Sci USA* 2004; **101**: 8443–8448.

11 Hong D, Gupta R, Ancliff P, Atzberger A, Brown J, Soneji S *et al.* Initiating and cancer-propagating cells in TEL-AML1-associated childhood leukemia. *Science* 2008; **319**: 336–339.

12 Schindler JW, Van Buren D, Foudi A, Krejci O, Qin J, Orkin SH *et al.* TEL-AML1 corrupts hematopoietic stem cells to persist in the bone marrow and initiate leukemia. *Cell Stem Cell* 2009; **5**: 43–53.

13 van der Weyden L, Giotopoulos G, Rust AG, Matheson LS, van Delft FW, Kong J *et al.* Modeling the evolution of ETV6-RUNX1-induced B-cell precursor acute lymphoblastic leukemia in mice. *Blood* 2011; **118**: 1041–1051.

14 Li M, Jones L, Gaillard C, Binnewies M, Ochoa R, Garcia E *et al.* Initially disadvantaged, TEL-AML1 cells expand and initiate leukemia in response to irradiation and cooperating mutations. *Leukemia* 2013; **27**: 1570–1573.

15 Hotfilder M, Röttgers S, Rosemann A, Schrauder A, Schrappe M, Pieters R *et al.* Leukemic stem cells in childhood high-risk ALL/t(9;22) and t(4;11) are present in primitive lymphoid-restricted CD34+CD19- cells. *Cancer Res* 2005; **65**: 1442–1449.

16 Hoffmann R, Melchers F. A genomic view of lymphocyte development. *Curr Opin Immunol* 2003; **15**: 239–245.

17 van Zelm MC, van der Burg M, de Ridder D, Barendregt BH, de Haas EF, Reinders MJ *et al.* Ig gene rearrangement steps are initiated in early human precursor B cell subsets and correlate with specific transcription factor expression. *J Immunol* 2005; **175**: 5912–5922.

18 van der Velden VH, Szczepański T, Wijkhuijs AJM, Hart PG, Hoogeveen PG, Hop WC *et al.* Age-related patterns of immunoglobulin and T cell receptor gene rearrangements in precursor-B-ALL: implications for detection of minimal residual disease. *Leukemia* 2003; **17**: 1834–1844.

19 Flohr T, Schrauder A, Cazzaniga G, Panzer-Grümayer R, van der Velden V, Fischer S *et al.* Minimal residual disease-directed risk stratification using real-time quantitative PCR analysis of immunoglobulin and T-cell receptor gene rearrangements in the international multicenter trial AIEOP-BFM ALL 2000 for childhood acute lymphoblastic leukemia. *Leukemia* 2008; **22**: 771–782.

20 Hübner S, Cazzaniga G, Flohr T, van der Velden VH, Konrad M, Pötschger U *et al.* High incidence and unique features of antigen receptor gene rearrangements in TEL-AML1-positive leukemias. *Leukemia* 2004; **18**: 84–91.

21 Gawad C, Pepin F, Carlton VE, Klinger M, Logan AC, Miklos DB *et al.* Massive evolution of the immunoglobulin heavy chain locus in children with B precursor acute lymphoblastic leukemia. *Blood* 2012; **120**: 4407–4417.

22 Teuffel O, Betts DR, Dettling M, Schaub R, Schäfer BW, Niggli FK. Prenatal origin of separate evolution of leukemia in identical twins. *Leukemia* 2004; **18**: 1624–1629.

23 Bungaro S, Irving J, Tussiwand R, Mura R, Minto L, Molteni C *et al.* Genomic analysis of different clonal evolution in a twin pair with t(12;21) positive acute lymphoblastic leukemia sharing the same prenatal clone. *Leukemia* 2008; **22**: 208–211.

24 Bateman CM, Alpar D, Ford AM, Colman SM, Wren D, Morgan M *et al.* Evolutionary trajectories of hyperdiploid ALL in monozygotic twins. *Leukemia* 2015; **29**: 58–65.

25 van Dongen JJ, Langerak AW, Brüggemann M, Evans PA, Hummel M, Lavender FL *et al.* Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 Concerted Action BMH4-CT98-3936. *Leukemia* 2003; **17**: 2257–2317.

26 Szczepański T, Van der Velden VH, Hoogeveen P, de Bie M, Jacobs DC, van Wering ER *et al.* Vδ2-Jα rearrangements are frequent in precursor-B-acute lymphoblastic leukemia but rare in normal lymphoid cells. *Blood* 2004; **103**: 3798–3804.

27 Langerak AW, Groenen PJ, Brüggemann M, Beldjord K, Bellan C, Bonello L *et al.* EuroClonality/BIOMED-2 guidelines for interpretation and reporting of Ig/TCR clonality testing in suspected lymphoproliferations. *Leukemia* 2012; **26**: 2159–2171.

28 Kozarewa I, Rosa-Rosa JM, Wardell CP, Walker BA, Fenwick K, Assiotis I *et al.* A modified method for whole exome resequencing from minimal amounts of starting DNA. *PLoS One* 2012; **7**: e32617.

29 van der Velden VH, Cazzaniga G, Schrauder A, Hancock J, Bader P, Panzer-Grumayer ER *et al.* Analysis of minimal residual disease by Ig/TCR gene rearrangements: guidelines for interpretation of real-time quantitative PCR data. *Leukemia* 2007; **21**: 604–611.

30 Page RMD, Holmes EC. *Molecular evolution: a phylogenetic approach,* 1st edn. Wiley-Blackwell: UK, 1998.

31 Langerak AW, Nadel B, De Torbal A, Wolvers-Tettero IL, van Gastel-Mol EJ, Verhaaf B *et al.* Unraveling the consecutive recombination events in the human IGK locus. *J Immunol* 2004; **173**: 3878–3888.

32 Romana SP, Poirel H, Leconiat M, Flexor MA, Mauchauffé M, Jonveaux P *et al.* High frequency of t(12;21) in childhood B-lineage acute lymphoblastic leukemia. *Blood* 1995; **86**: 4263–4269.

33 Greaves MF, Wiemels J. Origins of chromosome translocations in childhood leukaemia. *Nat Rev Cancer* 2003; **3**: 639–649.

34 Wiemels JL, Greaves M. Structure and possible mechanisms of TEL-AML1 gene fusions in childhood acute lymphoblastic leukemia. *Cancer Res* 1999; **59**: 4075–4082.

35 Tsai AG, Lu H, Raghavan SC, Muschen M, Hsieh CL, Lieber MR *et al.* Human chromosomal translocations at CpG sites and a theoretical basis for their lineage and stage specificity. *Cell* 2008; **135**: 1130–1142.

36 Papaemmanuil E, Rapado I, Li Y, Potter NE, Wedge DC, Tubio J *et al.* RAG-mediated recombination is the predominant driver of oncogenic rearrangement in ETV6-RUNX1 acute lymphoblastic leukemia. *Nat Genet* 2014; **46**: 116–125.

37 Fischer M, Schwieger M, Horn S, Niebuhr B, Ford A, Roscher S *et al.* Defining the oncogenic function of the TEL/AML1 (ETV6/RUNX1) fusion protein in a mouse model. *Oncogene* 2005; **24**: 7579–7591.

38 Ridge SA, Cabrera ME, Ford AM, Tapia S, Risueño C, Labra S *et al.* Rapid intraclonal switch of lineage dominance in congenital leukemia with a MLL gene rearrangement. *Leukemia* 1995; **9**: 2023–2026.

39 Janossy G, Woodruff RK, Pippard MJ, Prentice G, Hoffbrand AV, Paxton A *et al.* Relation of 'lymphoid' phenotype and response to chemotherapy incorporating vincristine-prednisolone in the acute phase of Ph1 positive leukaemia. *Cancer* 1979; **43**: 426–434.

40 Szczepański T, Beishuizen A, Pongers-Willemse MJ, Hählen K, Van Wering ER, Wijkhuijs AJ *et al.* Cross-lineage T cell receptor gene rearrangements occur in more than ninety percent of childhood precursor-B acute lymphoblastic leukemias: alternative PCR targets for detection of minimal residual disease. *Leukemia* 1999; **13**: 196–205.

41 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**: 8242–8247.

42 Castor A, Nilsson L, Astrand-Grundström I, Buitenhuis M, Ramirez C, Anderson K *et al.* Distinct patterns of hematopoietic stem cell involvement in acute lymphoblastic leukemia. *Nat Med* 2005; **11**: 630–637.

43 Hotfilder M, Röttgers S, Rosemann A, Jürgens H, Harbott J, Vormoor J. Immature CD34+CD19- progenitor/stem cells in TEL/AML1-positive acute lymphoblastic leukemia are genetically and functionally normal. *Blood* 2002; **100**: 640–646.

44 Li Z, Godinho FJ, Klusmann JH, Garriga-Canut M, Yu C, Orkin SH. Developmental stage-selective effect of somatically mutated leukemogenic transcription factor GATA1. *Nat Genet* 2005; **37**: 613–619.

Supplementary Information accompanies this paper on the Leukemia website (http://www.nature.com/leu)