Clinical & Experimental Immunology
The Journal of Translational Immunology
British Society for immunology
Clinical and Experimental Immunology ORIGINAL ARTICLE doi:10.1111/cei.12812

Illegitimate RAG-mediated recombination events are involved in *IKZF1* Δ3–6 deletion in *BCR-ABL1* lymphoblastic leukaemia

Y. Dong,*† F. Liu,‡ C. Wu,*† S. Li,*† X. Zhao,*† P. Zhang,*† J. Jiao,*† X. Yu,*† Y. Ji*† and M. Zhang§
*Department of Pathogenic Biology and Immunology, School of Basic Medical Sciences, Xi'an Jiaotong University Health Science Center, †Ministry of Education of China, Key Laboratory of Environment and Genes Related to Diseases (Xi'an Jiaotong University), ‡Department of Hematology, Xi'an Central Hospital, and §Department of Hematology, the First Affiliated Hospital of Xi'an Jiaotong University, Shaanxi, China

Accepted for publication 17 March 2016
Correspondence: *Y. Ji, School of Basic Medical Sciences, Xi'an Jiaotong University Health Science Center, No. 76 Yanta West Road, Xi'an, Shaanxi 710061, China.
E-mail: jiyanhong@xjtu.edu.cn
*M. Zhang, Department of Hematology, the First Affiliated Hospital of Xi'an Jiaotong University, No. 277 Yanta West Road, Xi'an, Shaanxi 710061, China.
E-mail: zhangmei@medmail.com.cn

Summary

Breakpoint cluster region-Abelson murine leukaemia viral oncogene homologue 1 (BCR–ABL1), encoded by the Philadelphia (Ph) chromosome, is the characteristic of chronic myeloid leukaemia (CML) and a subset of acute lymphoblastic leukaemia (ALL). We demonstrated that expression of the *Ik6* transcript, which lacked exons 3–6, was observed exclusively in *BCR–ABL1*$^+$B ALL and lymphoid blast crisis CML (BC–CML) patients harbouring the *IKZF1* Δ3–6 deletion. To confirm the hypothesis that illegitimate recombination activating gene protein (RAG)-mediated recombination events are involved in *IKZF1* Δ3–6 deletion in *BCR–ABL1* lymphoblastic leukaemia, we first demonstrated that the expression rates of *RAG1* and *RAG2*, collectively called *RAG*, were higher in ALL and BC–CML (lymphoid). Notably, analysis of relationships among *RAG*, *BCR–ABL1* and Ikaros 6 (*Ik6*) showed that *Ik6* can be generated only if *RAG* and *BCR–ABL1* are co-existing. The sequencing data showed that the deleted segments of introns 2 and 6 contained cryptic recombination signal sequences (cRSSs) and frequently had non-template nucleotides inserted between breakpoints. Furthermore, we used chromatin immunoprecipitation (ChIP) technology and demonstrated that the sequences directly flanking *IKZF1* Δ3–6 deletion breakpoints have significantly higher levels of histone H3 lysine 4 trimethylation (H3K4me3) modifications. Overall, RAG expression, good-quality cRSS and a specific chromatin modification, H3K4me3, satisfy the conditions of RAG's off-target effects on *IKZF1*. Our work provides evidence for RAG-mediated *IKZF1* Δ3–6 deletion. Our results raise the prospect that *RAG* is a valuable biomarker in disease surveillance. Dissecting the contribution of RAG should not only provide valuable mechanistic insights, but will also lead to a new therapeutic direction.

Keywords: ALL, BCR–ABL1, CML, Ik6, RAG

Introduction

Acute lymphoblastic leukaemia (ALL) is comprised of multiple subtypes with constellations of chromosomal rearrangements, deletions, trisomies and mutations [1]. A subset of ALL with a high risk of treatment failure and disease relapse is associated with the fusion protein breakpoint cluster region–Abelson murine leukaemia viral oncogene homologue 1 (BCR–ABL1) [2], a constitutively activated tyrosine kinase, which is encoded by the Philadelphia (Ph) chromosome resulting from the t(9;22)(q34;q11.2) translocation [3]. *BCR–ABL1*-positive ALL occurs in 2–5% of children and 25% of adult ALL [4–6]. Expression of BCR–ABL1 is also the defining characteristic of chronic myeloid leukaemia (CML). Most CML has a natural course of progression from an initial chronic phase (CP) to accelerated phase (AP) and blast crisis (BC). The final transformation phase causes an acute leukaemic-like illness, including lymphoblastic (25%) and myeloblastic (50%) subtypes or other undifferentiated phenotypes [7]. The characteristics of BC–CML (lymphoid) are always similar to those in B ALL [8]. Previous mouse studies demonstrated that transgenic expression of BCR–ABL1 in haematopoietic stem cells alone can induce a CML-like myeloproliferative disease [9]. However, additional cytogenetic aberrations and mutations in tumour suppressor genes, beyond *BCR–*

ABL1, are required for the generation of a blastic leukaemia [10].

IKZF1 encodes the DNA-binding transcription factor Ikaros (Ik), which is required for the development of all lymphoid lineages [11]. *IKZF1* comprises eight exons (0–7), of which exons 1–7 are coding. Exons 3–5 encode four N-terminal zinc fingers (ZnFs) required for DNA binding, and exon 7 encodes two C-terminal ZnFs that mediate dimerization. Ikaros has approximately eight isoforms (*Ik1–8*). The longer isoforms *Ik1–3*, which have at least three ZnFs, can bind DNA normally. However, the shorter isoforms *Ik4–8*, which lack two or more ZnFs, are unable to bind DNA and regarded as dominant-negative (DN) isoforms [12–14]. Notably, *Ik6* exerts a DN effect resulting into a loss of the tumour suppressor function attributed to wild-type *IKZF1* [15]. Illegitimate RAG-mediated *IKZF1* Δ3–6 deletion may be the underlying mechanism of *Ik6* generation [16]. However, this hypothesis is not proved, due to lack of direct evidence of RAG participation.

The diverse antigen receptor repertoire of B and T lymphocytes is generated by V(D)J recombination, which assembles the variable regions of immunoglobulin (Ig) or T cell receptor (TCR) genes from discontinuous variable (V), diversity (D) and joining (J) gene segments. V(D)J recombination is initiated by the RAG recombinase – a protein complex consisting primarily of the proteins encoded by recombination activating gene 1 (*RAG1*) and *RAG2*. RAG introduces DNA double-strand breaks (DSBs) specifically between antigen receptor gene segments and their flanking recombination signal sequences (RSSs). RSSs are comprised of highly conserved heptamer (consensus 5′-CACAGTG-3′) and nonamer (consensus 5′-ACAAAAACC-3′) elements separated by a relatively non-conserved 12 or 23 base pairs (bp) spacer (referred to as 12RSS or 23RSS, respectively) [17–19]. The human and mouse genomes contain millions of cryptic RSSs (cRSSs), which might be off-targets of RAG. Illegitimate RAG-mediated recombination can occur between a RSS flanking gene segment in *Tcr* or *Ig* loci and a cRSS flanking non-antigen receptor loci or between the two cRSS flanking non-antigen receptor loci, which are associated usually with genome alterations in haematological malignancies [20–23].

Here we demonstrated that *Ik6* was expressed exclusively in BCR-ABL1$^+$ B-ALL and BC–CML (lymphoid) harbouring *IKZF1* Δ3–6 deletion patients. To confirm the hypothesis that RAG-mediated recombination events are involved in *IKZF1* Δ3–6 deletion in *BCR–ABL1* lymphoblastic leukaemia, we investigated the mRNA expression of *RAG1* and *RAG2* by reverse transcription–polymerase chain reaction (RT–PCR) and analysed the relationships among *RAG*, *BCR–ABL1* and *Ik6*. Moreover, we searched for the presence of cRSS in the deleted segments of *IKZF1* introns 2 and 6 and then estimated their recombination potential. Furthermore, we used chromatin immunoprecipitation (ChIP) technology to demonstrate that the modification levels of histone H3 trimethylated at lysine 4 (H3K4me3) in the sequences directly flanking *IKZF1* Δ3–6 deletion breakpoints. Our discoveries provide valuable mechanistic insights for *IKZF1* Δ3–6 deletion in *BCR–ABL1* lymphoblastic leukaemia.

Table 1. Clinical and laboratory characteristics of patients with ALL

Patient no.	Sex	Age	WBC (10^9/l)	Blasts (BM, %)	Blasts (periphery, %)	Immunological markers	Disease status
BCR–ABL1$^+$B ALL-1	F	52	76·51	46	35	CD10(+)CD19(+)	Relapse
BCR–ABL1$^+$B ALL-2	F	63	35	45	37	CD10(+)CD19(+)CD34(+)	Initial
BCR–ABL1$^+$B ALL-3	F	20	21·3	59	31	CD10(+)CD19(+)CD34(+)	Initial
BCR–ABL1$^+$B ALL-4	F	49	19·34	71	49	CD10(+)CD19(+)CD34(+)	Initial
BCR–ABL1$^+$B ALL-5	M	46	93·72	33	30	CD10(+)CD19(+)CD34(+)	Initial
BCR–ABL1$^+$B ALL-6	F	60	24·9	73	41	CD10(+)CD19(+)CD34(+)	Initial
BCR–ABL1$^+$B ALL-7	F	44	77	52	59	CD10(+)CD19(+)CD20(+)	Initial
BCR–ABL1$^-$B ALL-8	F	23	27·67	55	31	CD19(+)CD34(+)	Initial
BCR–ABL1$^-$B ALL-9	F	53	23·4	54	29	CD10(+)CD19(+)	Initial
BCR–ABL1$^-$B ALL-10	M	45	38·89	38	42	CD10(+)CD19(+)	Initial
BCR–ABL1$^-$B ALL-11	F	15	66	39	31	CD19(+)CD20(+)CD34(+)	Initial
BCR–ABL1$^-$B ALL-12	M	40	49	71	44	CD19(+)CD20(+)CD34(+)	Relapse
BCR–ABL1$^-$B ALL-13	M	23	30	44	57	CD10(+)CD19(+)CD34(+)	Relapse
BCR–ABL1$^-$T ALL-14	F	22	117	28	34	CD2 (+)CD5(+)	Initial
BCR–ABL1$^-$T ALL-15	F	24	275	66	77	CD7(+)CD34(+)	Initial
BCR–ABL1$^-$T ALL-16	M	49	174	78	59	CD7(+)CD34(+)	Initial
BCR–ABL1$^-$T ALL-17	F	15	21	49	58	CD7(+)CD34(+)	Initial
BCR–ABL1$^-$T ALL-18	M	26	26·95	28	32	CD2(+)CD5(+)CD7(+)	Initial

ALL = acute lymphoblastic leukaemia; M = male; F = female; WBC = white blood cell; BM = bone marrow; blasts (BM, %) = the percentage of blasts in bone marrow; blasts (periphery, %) = the percentage of blasts in peripheral blood.

Table 2. Clinical and laboratory characteristics of patients with CML

Patient no.	Sex	Age	WBC (10^9/l)	*BCR-ABL1*	Immunological markers
CP–CML-1	M	56	104·3	+	n.a.
CP–CML-2	M	73	116·1	+	n.a.
CP–CML-3	M	24	168	+	n.a.
CP–CML-4	M	34	215	+	n.a.
CP–CML-5	M	65	260	+	n.a.
CP–CML-6	M	34	449·08	+	n.a.
CP–CML-7	M	58	29	+	n.a.
CP–CML-8	F	42	117·01	+	CD13(+)CD15(+)CD33(+)
CP–CML-9	F	70	457·77	+	CD13(+)CD33(+)
CP–CML-10	F	35	88·22	+	n.a.
CP–CML-11	M	42	140·09	+	n.a.
CP–CML-12	F	44	337	+	n.a.
CP–CML-13	F	58	75·25	+	n.a.
CP–CML-14	M	27	8·86	+	CD13(+)CD33(+)
CP–CML-15	F	55	61·13	+	n.a.
CP–CML-16	M	27	228·38	+	n.a.
CP-CML-17	M	30	588·53	+	CD15(+)CD33(+)CD64(+)
CP–CML-18	M	40	324·26	+	N.A.
CP–CML-19	M	69	108·93	+	CD33(+)CD34(+)CD64(+)
CP–CML-20	M	46	302·99	+	n.a.
CP–CML-21	M	40	225·78	+	n.a.
CP–CML-22	F	38	69·97	+	CD15(+)CD33(+)CD64(+)
CP–CML-23	F	40	39·14	+	CD33(+)CD34(+)
CP–CML-24	F	39	103·8	+	n.a.
CP–CML-25	F	43	254·07	+	CD15(+)CD16(+)CD33(+)
CP–CML-26	M	37	580·49	+	n.a.
AP–CML-27	F	36	61·63	+	CD33(+)CD34(+)CD117(+)
BC–CML-28 (myeloid)	F	60	13·45	+	CD33(+)CD13(+)
BC–CML-29 (myeloid)	F	56	175·55	+	n.a.
BC–CML-30 (myeloid)	M	35	20·11	+	CD13(+)CD33(+)CD34(+)
BC–CML-31 (myeloid)	F	43	273	+	CD13(+)CD34(+)CD117(+)
BC–CML-32 (myeloid)	M	57	53	+	CD33(+)CD34(+)CD64(+)
BC–CML-33 (myeloid)	F	73	56·13	+	CD33(+)CD34(+)
BC–CML-34 (lymphoid)	M	32	40	+	CD10(+)CD19(+)CD34(+)
BC–CML-35 (lymphoid)	F	35	19	+	CD10(+)CD19(+)CD34(+)
BC–CML-36 (lymphoid)	M	18	21	+	CD10(+)CD19(+)CD34(+)
BC–CML-37 (lymphoid)	F	76	68	+	CD10(+)CD19(+)CD34(+)
BC–CML-38 (lymphoid)	F	49	68	+	CD10(+)CD19(+)CD34(+)

CML = chronic myeloid leukaemia; CP–CML = chronic myeloid leukaemia in chronic phase; AP-CML = chronic myeloid leukaemia in acceleration phase; BC–CML (lymphoid) = lymphoid blast crisis chronic myeloid leukaemia; BC–CML (myeloid) = myeloid blast chronic myeloid leukaemia; M = male; F = female; WBC = white blood cell; blasts (BM, %) = the percentage of blasts in bone marrow; blasts (periphery, %) = the percentage of blasts in peripheral blood; n.a. = not available.

Patients and methods

Patients

Fifty-six peripheral blood samples were obtained from 18 patients with ALL and 38 patients with CML between 2012 and 2014 at the First Affiliated Hospital of Xi'an Jiaotong University. The ALL cohort included five T ALL (5 *BCR–ABL1*$^-$) and 13 B ALL (7 *BCR–ABL1*$^+$ and 6 *BCR–ABL1*$^-$). The CML cohort included 26 CP, one AP and 11 BC (five lymphoid BC and six myeloid BC). The clinical and laboratory features of ALL and CML patients were summarized, respectively, in Tables 1 and 2. Eight peripheral blood samples from healthy donors were selected as normal controls at the First Affiliated Hospital of Xi'an Jiaotong University. The study protocol was approved by the ethics committee of our institution and has therefore been conducted according to the Declaration of Helsinki. Informed consent from all patients and healthy controls was obtained before enrolment.

Table 3. Primers used for reverse transcription–polymerase chain reaction (RT–PCR), genomic PCR and quantitative (q)RT–PCR analysis

Gene	NCBI Accession no.		Sequence (5′–3′)	Annealing temperature (°c)	Size (bp)
RT–PCR					
IKZF1	NM_006060	Forward	CGAGGATCAGTCTTGGCCCCA	59	*Ik1*:1107
		Reverse	GCAGCTGGTACATCGGGCTGAT		*Ik2/3*:846
					Ik4:720
					Ik6:417
					Ik8:552
					Ik9:297
					Ik10: 242
RAG1	NM_000448	Forward	GGAGAGAGCAGAGAACACACT	56	225
		Reverse	GATCTCACCCGGAACAGCTT		
RAG2	NM_000536	Forward	AGCCCCTCTGGCCTTCAG	58	261
		Reverse	AAGAGGAGGGAGGTAGCAGG		
BCR–ABL1	AF113911	Forward	ACTCGCAACAGTCCTTCGAC	60	276
(P190)		Reverse	GGTTGGGGTCATTTTCACTG		
BCR–ABL1	AJ131466	Forward	ACTCCAGACTGTCCACAGCA	60	233
(P210)	AJ131467	Reverse	GGTTGGGGTCATTTTCACTG		158
β-actin	NM_001101	Forward	AGTGTGACGTGGACATCCGCAAAGAC	62	228
		Reverse	GCTTGCTGATCCACATCTGCTGGAAG		
Genomic PCR					
IKZF1	NC_000007	Forward	TATGAAGTCAATCAGCAGTGTTTCTAA	60	1600
		Reverse	AAGAAAGGAAGTAAAACACAGACCATA		
γ-actin	NC_000017	Forward	TGCTGCATGGGTTAATTGAG	60	307
		Reverse	CAGACTCACCAAGCCACAGA		
ChIP–qPCR					
IKZF1 intron 2	NC_000007	Forward	TTAAAAGGGAAAGGAGCAGTG	60	104
		Reverse	CGCAGTTACTTTTGCACCAA		
IKZF1 intron 6	NC_000007	Forward	TGAGGGATACCAGTTGTCCA	60	239
		Reverse	TTCGGGTCTCAGGAGACAAA		
γ-actin	NC_000017	Forward	GCTCACCGGCAGAGAAAC	60	114
		Reverse	GCCGCTTCCGCTTAAATA		
β-globin	NC_000011	Forward	GGTGAGTCTATGGGACGCTT	60	144
		Reverse	CGATCCTGAGACTTCCACAC		

ChIP = chromatin immunoprecipitation; bp = base pairs.

Sample collection and isolation of nucleated cells

Three ml of fresh peripheral blood was collected into aseptic tubes containing ethylenediamine tetraacetic acid (EDTA) anti-coagulant from the veins of fasting patients in the morning. Red blood cells were lysed by red blood cell lysis buffer. After the supernatant was discarded, the nucleated cell pellets at the bottom of the tube were cleaned twice with phosphate-buffered saline (PBS) buffer and separated into DNase/RNase-free Eppendorf tubes.

RT–PCR

Total cellular RNA was extracted with TRIzol Regent (Invitrogen, Carlsbad, CA, USA), according to the manufacturer's protocol. Extracted RNA was treated with RNase-free DNase I (Roche, Basel, Switzerland) in order to eliminate residual genomic DNA. Five hundred ng of total RNA was reverse-transcribed using random hexamer primers with PrimeScript^{TM}RT reagent Kit (TaKaRa, Shiga, Japan), according to the manufacturer's protocol. PCR was performed in a total volume of 50 μl containing 50 ng of cDNA, 0·2 mM of deoxynucleotides (dNTPs), 1·5 mM of $MgCl_2$, 0·2 μM of forward and reverse primer, 1 U of KOD Plus Neopolymerase and 1× KOD Plus Neo buffer (Toyobo, Osaka, Japan) with the following conditions: 94°C for 5 min; 35 cycles of 30 s at 94°C; 30 s at annealing temperature and 1 min at 72°C; and 1 min at 72°C for 5 min (BioRad-s1000; BioRad, Hercules, CA, USA). The primer sequences used for PCR are listed in Table 3.

Genomic PCR

Genomic DNA was isolated using conventional Proteinase K digestion and phenol–chloroform extraction according to the Molecular Cloning protocols. Genomic PCR was performed with 50 ng of genomic DNA in a 50-μl reaction by the use of KOD-Plus-Neo (Toyobo). Amplification conditions were as follows: 94°C for 5 min; 35 cycles of 30 s at 94°C, 30 s at 60°C and 1 min at 72°C; and 72°C for 5 min

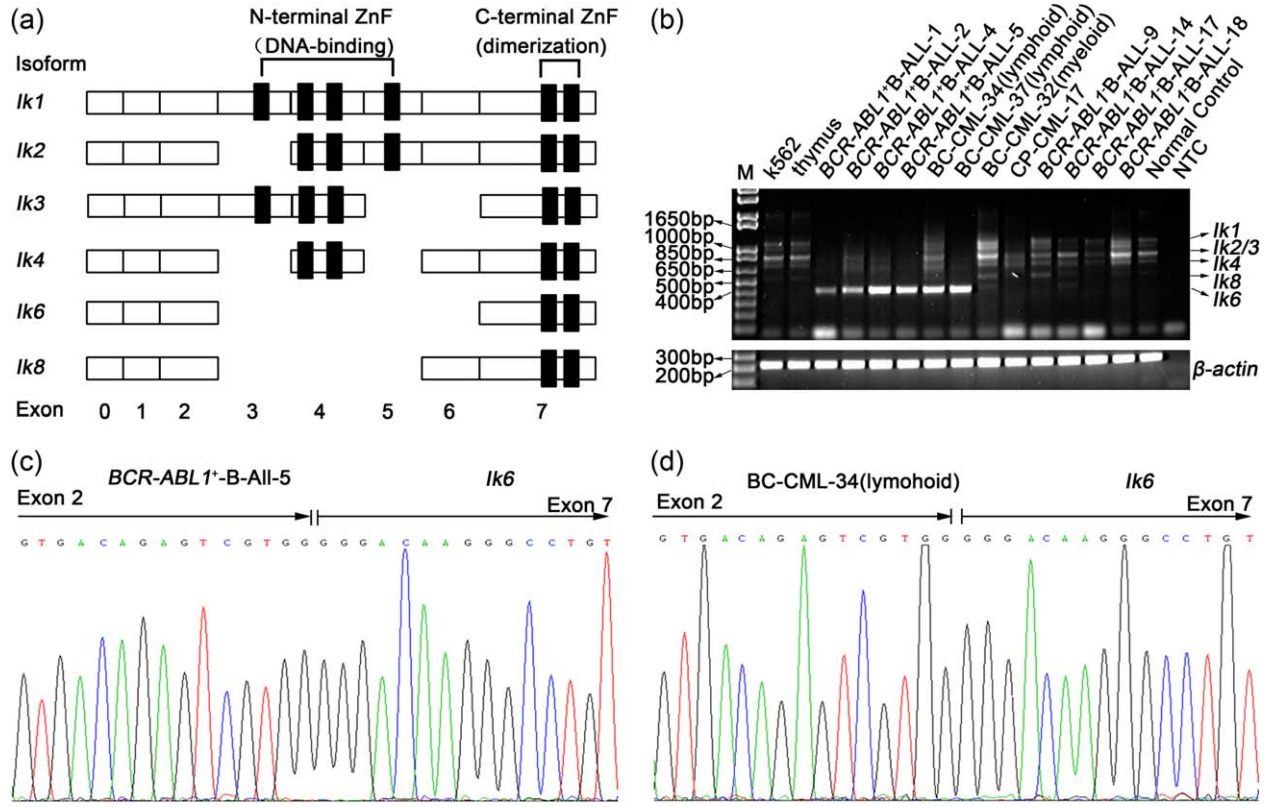

Fig. 1. Expression of different *IKZF1* transcripts in acute lymphoblastic leukaemia (ALL) and chronic lymphoblastic leukaemia (CML). (a) Domain structure of different *IKZF1* transcripts are detected by reverse transcription–polymerase chain reaction (RT–PCR), examples of which are shown in (b). (b) RT–PCR for *IKZF1* transcripts in representative cases of ALL and CML. (c,d) Sequencing of RT–PCR products confirms the characteristic of *Ik6* in representative cases of *BCR–ABL1*$^+$ B ALL and BC–CML (lymphoid), respectively. Regions matching the reference *IKZF1* cDNA sequences are shown by arrows.

(BioRad-s1000). The primer sequences used for PCR are listed in Table 3.

Cloning and sequencing analysis

RT–PCR and genomic PCR products were separated by agarose gels electrophoresis. Corrected DNA bands were excised and gel was extracted with E.Z.N.A.^{TM} Gel Extraction Kit (Omega Biotek, Inc., Norcross, GA, USA). Gel-purified DNA was ligated into the pMD19-T vector (TaKaRa) with T4 DNA ligase (New England Biolabs, Ipswich, MA, USA). Plasmid DNA was extracted with E.Z.N.A.^{TM} Plasmid Mini Kit I (Omega Biotek, Inc.) and sequenced (ABI, 3730XL) by Shanghai Sunny Biotechnology Co., Ltd (Shanghai, China). Reference genome sequences were obtained from the NCBI's RefSeq, and sequence comparisons were performed with the BLAST software tool (www.ncbi.nlm.nih.gov/BLAST/).

Computational detection of putative RAG RSSs

The deleted segments of human *IKZF1* gene were screened for the presence of cryptic recombination signal sequences (cRSSs) using the IMGT software algorithms (http://www.itb.cnr.it/rss/analyze.html). A significant correlation was observed between RSS information content (RIC) scores and RSS functionality. In the current version, pass/fail RIC thresholds were from Cowell *et al.* [24], as follows: 12 RSS: pass with RIC $> -38{\cdot}81$ and 23 RSS: pass with RIC $> -58{\cdot}45$.

Chromatin immunoprecipitation

The ChIP procedure has been described in detail previously [25]. Briefly, 30 million cells were cross-linked with 1% HCHO (Sigma, St Louis, MO, USA) for 13 min at room temperature (RT), and the reaction was terminated with 0·125 M glycine. After centrifugation at 900 g for 3 min, the precipitation was washed, resuspended in radioimmunoprecipitation assay (RIPA) buffer (10 mM Tris pH 7·4, 1 mM EDTA pH 8·0, 0·1% sodium deoxycholate, 0·1% sodium dodecyl sulphate (SDS), 0·8 M NaCl, 1% Triton X-100 and sonicated using Bioruptor^{TM}UCD-200 (Diagenode, Seraing, Belgium) at high power (30" on, 30" off) to shear DNA to lengths of between 200 and 300 bp three times every 15 min. After preclearing the chromatin with Dynabeads Protein G beads (×2) (Invitrogen), an aliquot

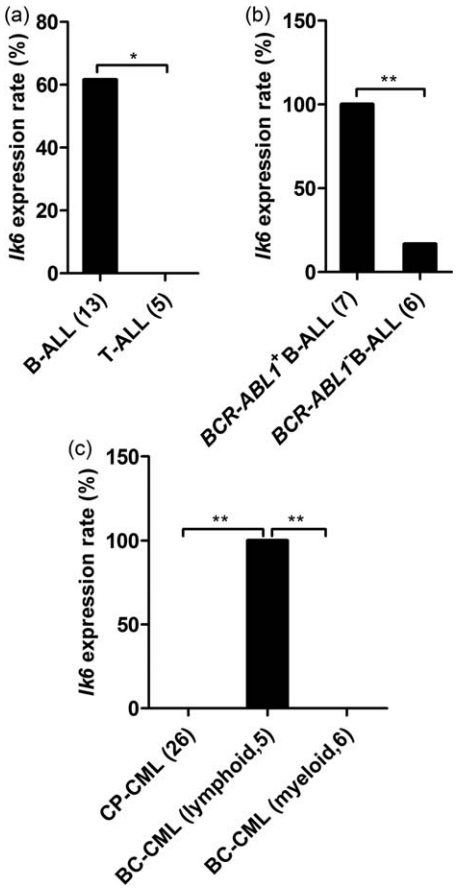

Fig. 2. *Ik6* expression rates in acute lymphoblastic leukaemia (ALL) and chronic lymphoblastic leukaemia (CML). (a) *Ik6* expression rates in B ALL and T ALL. (b) *Ik6* expression rates in *BCR–ABL1*$^+$ B ALL and *BCR–ABL1*$^-$ B ALL. (c) *Ik6* expression rates in the different phases of CML. CP = chronic phase; BC = blast crisis. Numbers of cases in each group are included in the brackets. Asterisks indicate statistically significant difference between different groups. (a,b) $^*P < 0{\cdot}05$; $^{**}P < 0{\cdot}01$; (c) $^{**}P < 0{\cdot}0167$ [corrected testing level = $2\alpha/(K(K-1))$; α = uncorrected testing level = 0·05; K = the number of experimental groups = 3.

(5×10^5 cell equivalents) was set aside as the input sample. Chromatin from 5×10^6 cells was then incubated with specific antibody or normal rabbit immunoglobulin (Ig)G (Millipore, Billerica, MA, USA) overnight at 4°C. Immune complexes were pulled down with Dynabeads Protein G beads (×2) (Invitrogen). After reverse cross-link, DNA was isolated using conventional phenol–chloroform extraction. Quantitative real-time PCR was performed using SYBR Premix Ex Taq II (TaKaRa) with an Mx3000 thermocycler (Agilent Technologies, Santa Clara, CA, USA). The primer sequences used for quantitative PCR (qPCR) are listed in Table 3. Input samples were diluted so that each IP and input samples would provide approximately equal qPCR signals. Using standard curves generated for each region analysed in each experiment, the amount of DNA recovered in immunoprecipitates and the input chromatin was calculated. ChIP–qPCR signals were expressed as the following equation: $\text{IP/input}_{\text{corr}} = [(\text{IP}_{\text{specific antibody}} - \text{IP}_{\text{IgG}})/\text{input}] \times 1000$. ChIP experiments were performed with antibodies for H3K4me3 (Millipore).

Statistical analysis

Statistics were calculated using the SPSS version 18·0 statistical software package. For continuous variables, descriptive results were presented as mean ± standard deviation (s.d.). Intergroup comparisons were made using one-way analysis of variance (ANOVA) for data that followed normal distribution. Likelihood-ratio χ^2 or Fisher's exact tests were applied for 2 × 2 tables or R × C contingency tables in case the number of patients in individual groups was lower than 40. If there were significant differences in multiple comparisons, we performed partition of χ^2 statistics {corrected testing level $\alpha' = 2\alpha/[K(K-1)]$ α' uncorrected testing level = 0·05; K = the number of experimental groups}. All tests with *P*-values less than 0·05 or corrected testing level α' was considered statistically significant.

Results

Ik6 was a dominant *IKZF1* transcript in *BCR–ABL1*$^+$ B ALL and BC–CML (lymphoid)

To determine which *IKZF1* transcripts appear in ALL and CML, we performed RT–PCR using a pair of specific primers located in exons 0 and 7 of *IKZF1*, respectively, which could detect all *IKZF1* transcripts. We found that there were *Ik1*, *Ik2/3*, *Ik4*, *Ik6* and *Ik8* in our cases. However, *Ik6* was the most common transcript in *BCR–ABL1*$^+$ B ALL and BC–CML (lymphoid) (Fig. 1a,b). Sequencing of RT–PCR products from *Ik6* transcripts demonstrated that *Ik6* lacked exons 3–6, and contained only exons 0, 1, 2 and 7 (Fig. 1c,d). The expression rates of *Ik6* in B ALL (eight of 13, 61·54%) was significantly higher than that in T ALL (none of five, 0%) ($P < 0{\cdot}05$) (Fig. 2a). Moreover, the expression rates of *Ik6* in *BCR–ABL1*$^+$ B ALL (seven of seven, 100%) was significantly higher than that in *BCR–ABL1*$^-$ B ALL (one of six, 16·7%) ($P < 0{\cdot}01$) (Fig. 2b). Furthermore, *Ik6* was expressed in BC–CML (lymphoid; five of five, 100%), but not in 26 cases of CML in CP, one case of CML in AP (data not shown) and six cases of BC–CML (myeloid) (Fig. 2c).

Ik6 was caused by *IKZF1* genomic Δ3–6 deletion

To explore whether the genomic alteration was responsible for *Ik6* generation, we performed genomic PCR on DNA from patients with the *Ik6* transcript using a pair of specific primers located in *IKZF1* introns 2 and 6, respectively (Fig. 3a). The PCR products were then cloned and sequenced. BLAST results showed that the sequences were matched in

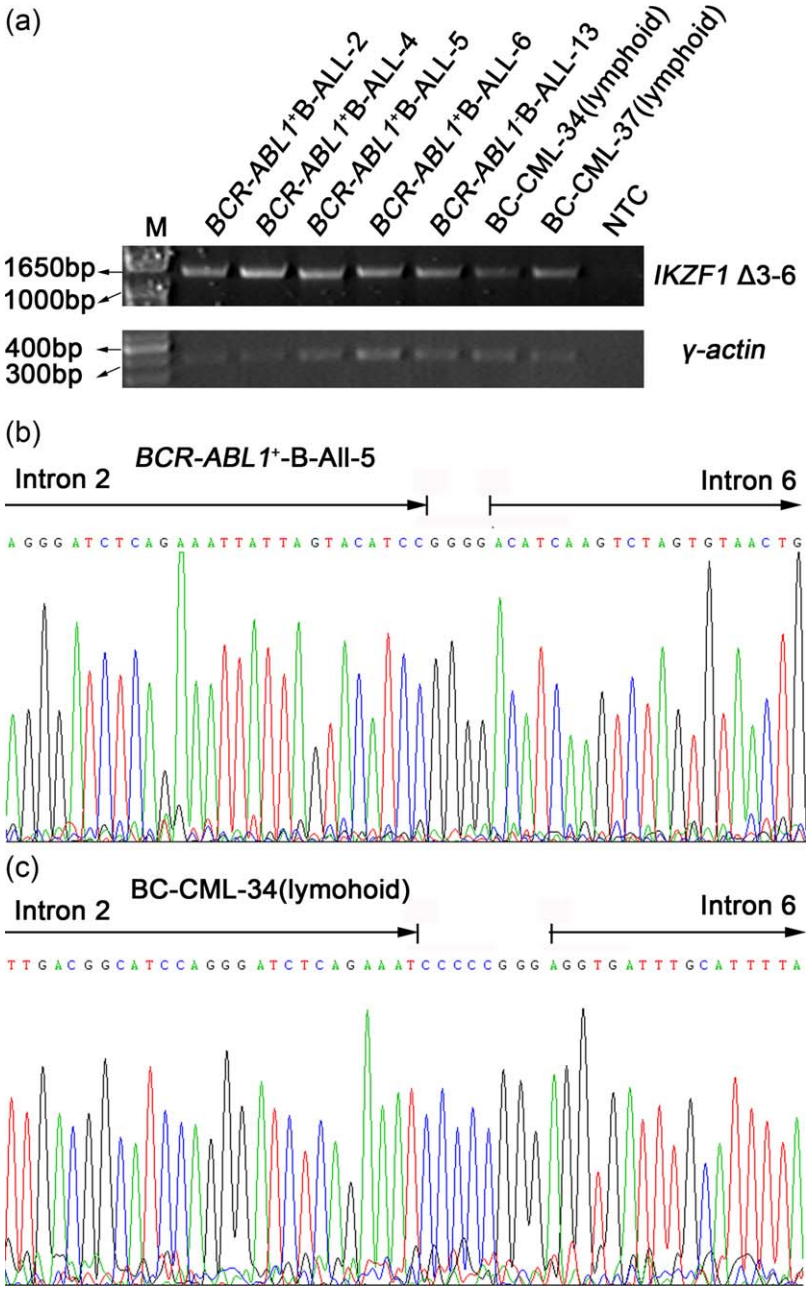

Fig. 3. Genomic polymerase chain reaction (PCR) and sequencing of *IKZF1* Δ3–6. (a) Genomic PCR of *IKZF1* Δ3–6 in representative cases with *Ik6* transcript; γ-actin is used as a control for DNA loading. (b) Sequencing of genomic PCR products of *IKZF1* Δ3–6. Regions matching the reference genomic *IKZF1* sequence are shown by arrows, separated by additional nucleotides not matching the consensus sequence.

introns 2 and 6 of *IKZF1*, which were separated by additional nucleotides not belonging to the consensus *IKZF1* sequence (Fig. 3b,c), indicating strongly that *Ik6* was caused by *IKZF1* genomic Δ3–6 deletion.

High *RAG1* and *RAG2* expression rates in ALL and BC–CML (lymphoid)

Due to the fact that RAG is always expressed in the early stage of lymphocyte development, we analysed the characteristics of leukaemia cells from ALL and BC–CML (lymphoid). According to the immunological markers [26–29], 13 B–ALL cases expressed CD19 or CD20, which were all B cell-specific antigens, and also expressed CD10 or CD34 that marked B cells consistently as progenitors or precursors. Five T ALL cases constituted three pro-T ALL ($CD7^+CD34^+$) and two pre-T ALL ($CD2^+CD5^+$) (Table 1). Five BC–CML (lymphoid) cases were characterized by expression of CD10, CD19 and CD34, which were consistent with the markers of early B ALL (Table 2). These cytological characteristics suggested that most leukaemia cells from ALL and BC–CML (lymphoid) might be blocked in the early stage of lymphocyte development.

To confirm further the expression of *RAG1* and *RAG2* in ALL and CML, we performed reverse transcription on total

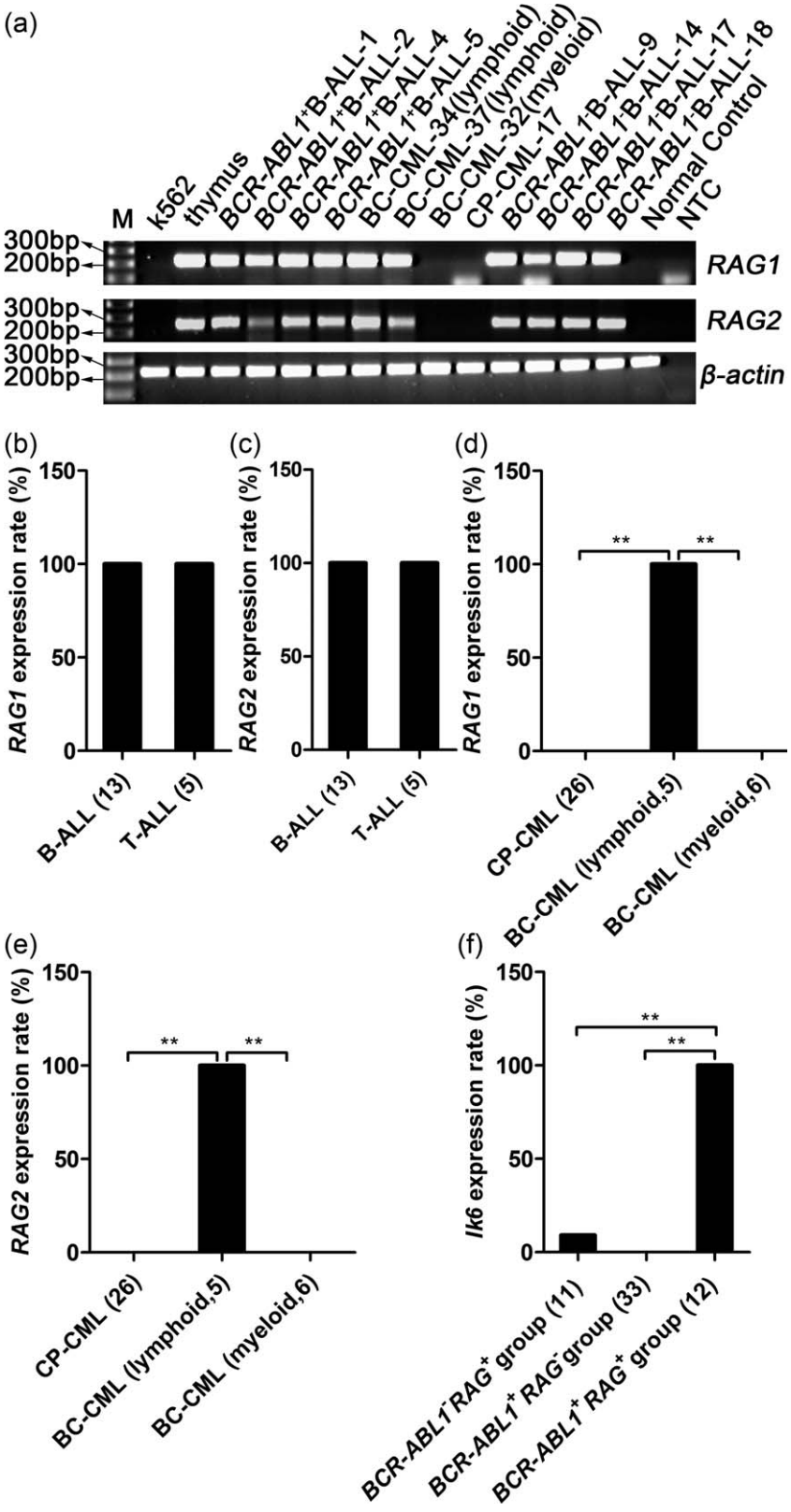

Fig. 4. *RAG1* and *RAG2* expression rates in acute lymphoblastic leukaemia (ALL) and chronic lymphoblastic leukaemia (CML). (a) Reverse transcription–polymerase chain reaction (RT–PCR) for *RAG1* and *RAG2* transcripts in representative cases of ALL and CML. β-actin is used as a control for DNA loading. (b) *RAG1* expression rates in B ALL and T ALL. (c) *RAG2* expression rates in B ALL and T ALL. (d) *RAG1* expression rates in the different phases of CML. CP = chronic phase; BC = blast crisis. (e) *RAG2* expression rates in the different phases of CML. (f) *Ik6* expression rates in the different situations of *BCR-ABL1* and *RAG*. Numbers of cases in each group are included in the brackets. Asterisks indicate statistically significant difference between different groups. (d–f) $^{**}P < 0{\cdot}0167$ [corrected testing level = $2\alpha/(K(K-1)] = 0{\cdot}0167$; α = uncorrected testing level = 0·05; K = the number of experimental groups = 3.

RNA from the leukaemia cells of 56 patients using random hexamers, followed by PCR using *RAG1*- and *RAG2*-specific primers for transcription analysis (Fig. 4). The data showed that the expression rates of *RAG1* and *RAG2* were 100% in leukaemia cells from B ALL and T ALL, respectively (Fig. 4b,c). Moreover, the expression rates of *RAG1* and *RAG2* in BC–CML (lymphoid; five of five, 100%) were significantly higher than those in 26 cases of CML in

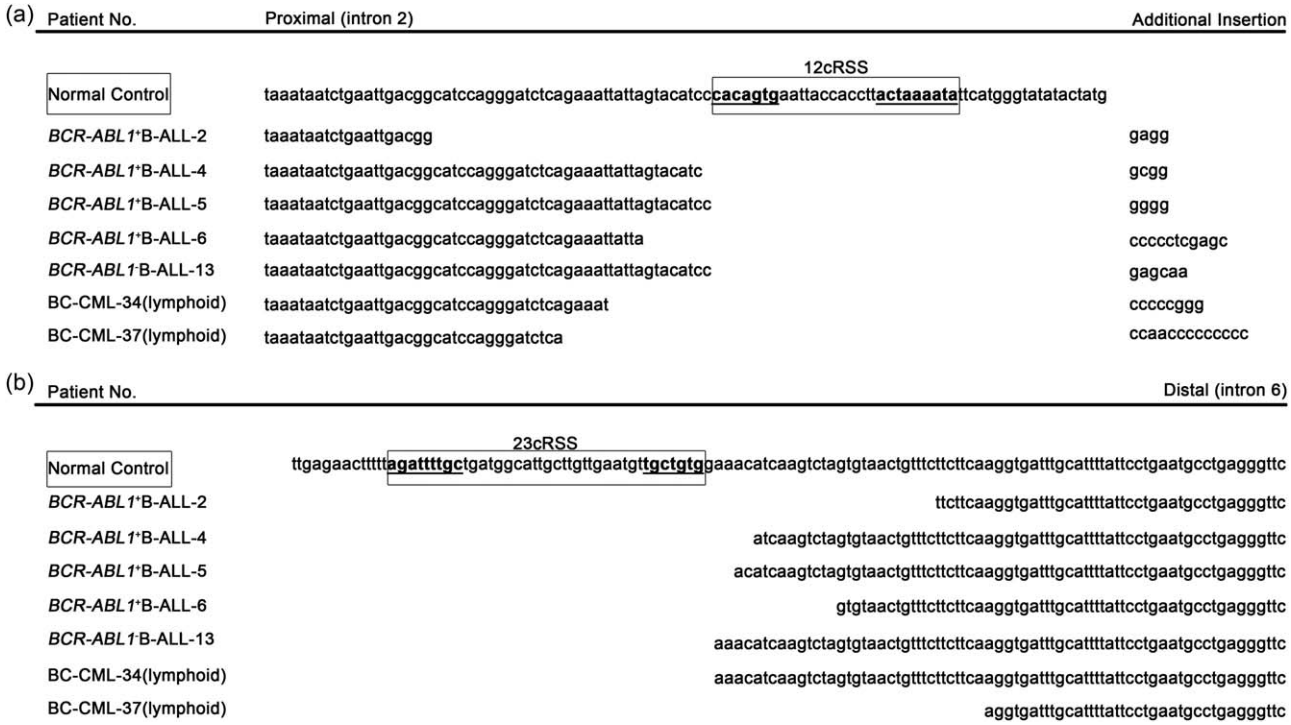

Fig. 5. (Cryptic) recombination signal sequences (cRSSs) in the deleted segments of *IKZF1* Δ3–6. (a) Sequences of intron 2 breakpoints for representative cases with *Ik6* transcript. 12cRSS is indicated by a box with the canonic heptamer sequence and the nonamer sequence indicated in bold type and underlined. Introns 2 and 6 are separated by additional nucleotides. (b) Sequences of intron 6 breakpoints for representative cases with *Ik6* transcript. 23cRSS is indicated by a box with the canonic heptamer sequence and the nonamer sequence indicated in bold type and underlined.

chronic phase (CP) and six cases of BC–CML (myeloid) ($P < 0{\cdot}0167$) (Fig. 4d,e).

RAG and *BCR–ABL1* were indispensable for *Ik6* generation

To explore further the relationships among *RAG*, *BCR–ABL1* and *Ik6*, we analysed *Ik6* expression rates in the different situations of *BCR–ABL1* and *RAG*. *Ik6* was expressed in only one patient of the *BCR–ABL1*$^-$ *RAG*$^+$ group, and not expressed in the *BCR–ABL1*$^+$ *RAG*$^-$ group. Instead, *Ik6* expression rates were 100% in the *BCR–ABL1*$^+$ *RAG*$^+$ group (Fig. 4f). These data suggested that *RAG* and *BCR–ABL1* were all indispensable for *Ik6* generation.

The deleted segments of *IKZF1* Δ3–6 contained cRSSs with similar recombination potential to that of authentic RSSs flanking immunoglobulin V(D)J gene segments

Analysis of sequences directly flanking the breakpoints revealed a canonic CAC/GTG sequence, which is a hallmark of the heptamer sequences of RSSs. The nonamer sequence with A/T-nucleotide enrichment nearby heptamer was separated by a spacer of 12 or 23 nucleotides. A 12cRSS was located within the deleted segment of intron 2 with a 5′–3′ orientation, and a 23cRSS was located in the deleted segment of intron 6 with a 3′–5′ orientation (Fig. 5a,b). A variable number of additional nucleotides were inserted in between breakpoints (Fig. 5a).

To assess further the quality of cRSSs in the deleted segments of *IKZF1* Δ3–6, we used a publically available program (http://www.itb.cnr.it/rss/analyze.html) that can predict RSSs and calculate RSS information content (RIC) scores. This search in the *IKZF1* Δ3–6 deletion segments predicted a 12cRSS with a RIC score of −30·62 as well as 23cRSS with a RIC score of −54.97. The two cRSSs obtained were assigned a passing score, indicating that they had recombination potential resembling that of authentic RSSs flanking antigen receptor gene segments.

The sequences directly flanking *IKZF1* Δ3–6 deletion breakpoints had significantly higher levels of H3K4me3

We then performed ChIP technology to analyse the levels of H3K4me3 in the sequences directly flanking *IKZF1* Δ3–6 deletion breakpoints in ALL and CML patients, respectively. The primer sequences used are summarized in Table 3. A pair of *IKZF1* intron 2 primers were located upstream of 12cRSS and two *IKZF1* intron 6 primers were located

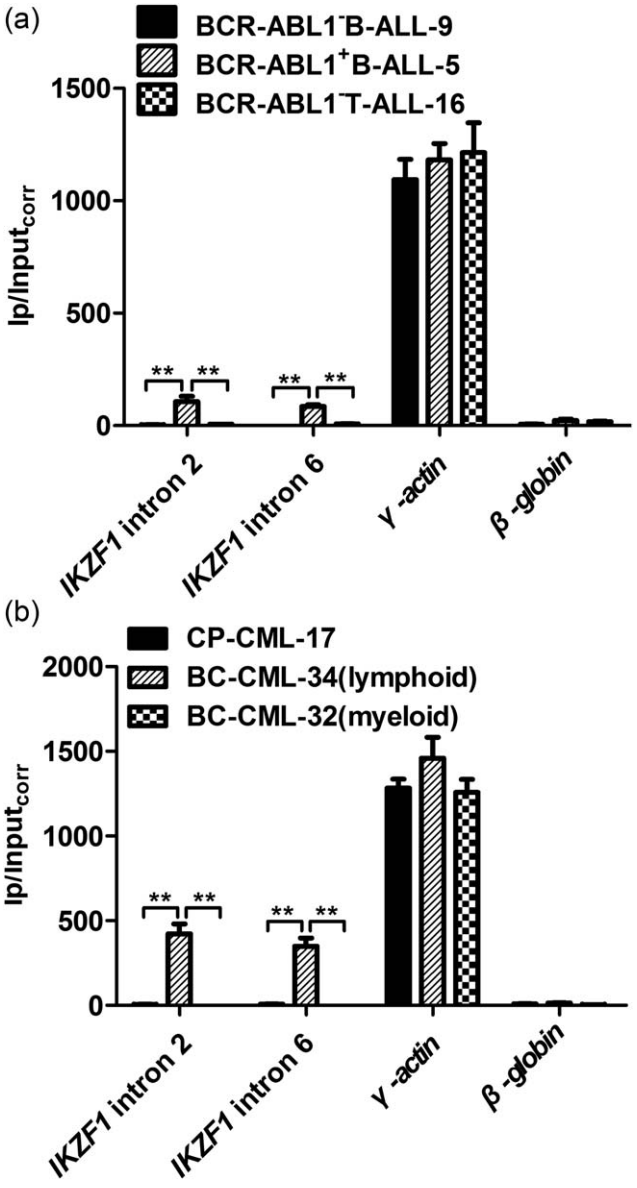

Fig. 6. The levels of histone H3 lysine 4 trimethylation (H3K4me3) in the sequences directly flanking *IKZF1* Δ*3–6* deletion breakpoints in acute lymphoblastic leukaemia (ALL) and chronic lymphoblastic leukaemia (CML) patients. Chromatin immunoprecipitation–quantitative polymerase chain reaction (ChIP–qPCR) signals were expressed in the following equation: IP/Input$_{corr}$ = [(IP$_{specific\ antibody}$–IP$_{IgG}$)/input] × 1000; γ-actin was used as positive control and β-globin was used as negative control. All measurements were conducted in triplicate.

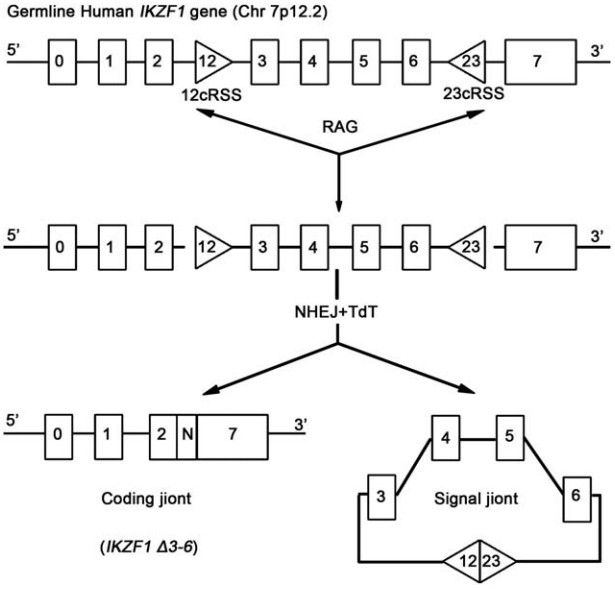

Fig. 7. Schematic model of illegitimate RAG-mediated *IKZF1* Δ3–6 deletion. The 12cRSS and 23cRSS are represented as 5′–3′ orientation and 3′–5′ orientation triangles, respectively. *IKZF1* exons are indicated by box; *IKZF1* introns are indicated by line between two boxes or between a box and a triangle; NHEJ = non-homologous end joining; TdT = terminal deoxynucleotidyltransferase. *IKZF1* Δ3–6 deletion can be divided into two stages. The first stage is that recombination activating gene (RAG) introduces DSB in *IKZF1*. The second stage is the formation of coding joint and signal joint.

downstream of 23cRSS, which were adjacent to breakpoints of *IKZF1* Δ3–6 deletions. The data showed that the sequences directly flanking the deletion breakpoints of *IKZF1* introns 2 and 6 had significantly higher levels of H3K4me3 in *BCR–ABL1*$^+$ B ALL than *BCR–ABL1*$^-$ B ALL and BC–CML (lymphoid) than CP–CML patients, respectively (Fig. 6).

Discussion

In this study, we demonstrated that expression of the *Ik6* transcript, which lacked exons 3–6, was observed exclusively in *BCR–ABL1*$^+$ B ALL and BC–CML (lymphoid) harbouring the *IKZF1* Δ3–6 deletion, but not in BCR–ABL1$^-$ ALL and CP–CML. This observation suggests that *Ik6* is associated with the pathogenesis of BCR–ABL1$^+$ ALL and blast transformation of CML. The results are consistent with previous findings [16,30]. However, the *Ik6* generation mechanism remains controversial. Alternative splicing has been reported to be the underlying mechanism [13]. This hypothesis has been challenged by our study, stating that expression of *Ik6* transcript is determined by the presence of *IKZF1* deletions.

RAG-mediated recombination is an intricate and tightly regulated process. Errors in this process always produce aberrant genomic rearrangements. The mechanisms of illegitimate recombination may include regulation of *RAG* expression, the RSS specificity of RAG endonuclease itself and chromatin features that define possible recombination sites (transcription-associated accessibility and histone H3 methylation, etc.) [31]. First, our study found that most leukaemia cells from ALL and BC–CML (lymphoid) were blocked in the early stage of lymphocyte development and

expressed *RAG*, which may serve as an initiator enzyme for V(D)J recombination. Secondly, the quality of cRSSs determines binding the specificity of RAG itself. Our results suggested that 12 of 23 cRSS in the deleted segments of *IKZF1* introns 2 and 6 had a similar recombination potential to that of authentic RSSs flanking V(D)J gene segments of antigen receptor loci. Thirdly, the integrity of *IKZF1* is not damaged, although the developing lymphocytes in the early stage express RAG and contain good-quality cRSSs. The ability of RAG to initiate V(D)J recombination also depends upon the accessibility of RSSs, which includes high-transcription, active histone modifications such as H3K4me3. Our data demonstrated that the sequences directly flanking the *IKZF1* Δ3–6 deletion breakpoints had significantly higher levels of H3K4me3 in *BCR–ABL1*$^+$ B ALL and BC–CML (lymphoid) patients. Moreover, RAG2 contains a plant homeodomain (PHD) finger that recognizes H3K4me3, and this interaction is important for V(D)J recombination [19]. However, the origin of H3K4me3-rich in the *IKZF1* loci is under investigation. The relationships among *RAG*, *BCR–ABL1* and *Ik6* emphasize the importance of *RAG* and *BCR–ABL1* in *Ik6* generation. *Ik6* can be generated only if RAG and *BCR–ABL1* co-exist. Previous studies have demonstrated that BCR–ABL1 expression can promote genomic instability and acquisition of chromosomal aberrations [32]. We speculate that *BCR–ABL1* may increase H3K4me3 modification levels on *IKZF1* breakpoints sequences, which further predispose RAG to exert off-target effects.

Based on the above findings, we propose that *IKZF1* cleavage is a two-step process similar to the steps of canonical V(D)J recombination [19]. In the first phase, RAG recognizes a pair of 12cRSS (located in intron 2) and 23cRSS (located in intron 6) resulting in DNA double-strand breaks (DSBs) adjacent to each heptamer. In the second phase, the DNA ends are rejoined by non-homologous end-joining (NHEJ) DNA repair factors. Coding joints and signal joints are generated typically during normal T or B cell receptor gene segment rearrangements. In the formation of *IKZF1* Δ3–6, introns 2 and 6 breakpoint ends not containing cRSS undergo non-templated nucleotide addition by terminal deoxynucleotidyltransferase (TdT) before being joined to form the coding joint. Signal joints are the consequence of heptamer–heptamer sequence fusions (Fig. 7).

Several advantages advocate the potential clinical application of *RAG* mRNA detection in *BCR–ABL1* lymphoblastic leukaemia. First, *RAG* mRNA are detected in leukaemia cells from peripheral blood samples, which are common specimens in clinical investigation. Secondly, detection can be accomplished with high reproducibility using commercially available RT–PCR detection kits. Thirdly, *RAG* mRNA expression can be detected accurately and interpreted easily. Finally, *RAG* is the necessary factor for *Ik6* generation in *BCR–ABL1* lymphoblastic leukaemia. The prognostic information provided by *RAG* appears superior to *Ik6*.

In conclusion, our data confirm that expression of *Ik6* transcript arises from *IKZF1* Δ3–6 deletion. We further provide evidence that illegitimate RAG-mediated recombination events are involved in *IKZF1* Δ3–6 deletion in *BCR–ABL1* lymphoblastic leukaemia. Currently, our group are purifying rabbit anti-human RAG ChIP grade antibodies; further research on the binding of RAG to the *IKZF1* cRSS loci is anticipated. The emergence of RAG may provoke disease progression and relapse, adding a new level of complexity to be addressed in the development of anti-leukaemia strategies. Our results raise the prospect that *RAG* is a valuable biomarker in disease surveillance.

Acknowledgements

This work was supported by grants from the National Natural Scientific Foundation of China (no. 31170821).

Author contributions

Y. D., Y. J. and M. Z. designed the study. Y. D., F. L., S. L., X. Z. and P. Z. participated in sample collection, processing, and storage. Y. D. and F. L. performed experiments. Y. D., C. W., J. J. and X. Y. analysed and interpreted the data. Y. J. and Y. D. wrote the paper.

Disclosure

The authors declare that they have no disclosures.

References

1 Mullighan CG. The molecular genetic makeup of acute lymphoblastic leukemia. Hematology Am Soc Hematol Educ Program 2012; **2012**:389–96.
2 Pui C-H, Robison LL, Look AT. Acute lymphoblastic leukaemia. Lancet 2008; **371**:1030–43.
3 Goldman JM, Melo JV. Chronic myeloid leukemia – advances in biology and new approaches to treatment. N Engl J Med 2003; **349**:1451–64.
4 Gleißner B, Gökbuget N, Bartram CR *et al.* Leading prognostic relevance of the BCR–ABL translocation in adult acute B-lineage lymphoblastic leukemia: a prospective study of the German Multicenter Trial Group and confirmed polymerase chain reaction analysis. Blood 2002; **99**:1536–43.
5 Bernt KM, Hunger SP. Current concepts in pediatric Philadelphia chromosome-positive acute lymphoblastic leukemia. Front Oncol 2014; **4**:54.
6 Moorman AV, Harrison CJ, Buck GA *et al.* Karyotype is an independent prognostic factor in adult acute lymphoblastic leukemia (ALL): analysis of cytogenetic data from patients treated on the Medical Research Council (MRC) UKALLXII/Eastern Cooperative Oncology Group (ECOG) 2993 trial. Blood 2007; **109**:3189–97.
7 Apperley JF. Chronic myeloid leukaemia. Lancet 2015; **385**: 1447–59.

8 Skorski T. Genetic mechanisms of chronic myeloid leukemia blastic transformation. Curr Hematol Malig Rep 2012; **7**:87–93.

9 Daley GQ, Van Etten RA, Baltimore D. Blast crisis in a murine model of myelogenous leukemia. Proc Natl Acad Sci USA 1991; **88**:11335–8.

10 Mullighan CG, Downing JR. Genome-wide profiling of genetic alterations in acute lymphoblastic leukemia: recent insights and future directions. Leukemia 2009; **23**:1209–18.

11 Georgopoulos K, Bigby M, Wang JH *et al.* The Ikaros gene is required for the development of all lymphoid lineages. Cell 1994; **79**:143–56.

12 Molnar A, Georgopoulos K. The Ikaros gene encodes a family of functionally diverse zinc finger DNA-binding proteins. Mol Cell Bio 1994; **14**:8292–303.

13 Molnar A, Wu P, Largespada DA *et al.* The Ikaros gene encodes a family of lymphocyte-restricted zinc finger DNA binding proteins, highly conserved in human and mouse. J Immunol 1996; **156**:585–92.

14 Mullighan C, Downing J. Ikaros and acute leukemia. Leuk Lymphoma 2008; **49**:847–9.

15 Van der Veer A, Zaliova M, Mottadelli F *et al.* IKZF1 status as a prognostic feature in BCR–ABL1-positive childhood ALL. Blood 2014; **123**:1691–8.

16 Mullighan CG, Miller CB, Radtke I *et al.* BCR–ABL1 lymphoblastic leukaemia is characterized by the deletion of Ikaros. Nature 2008; **453**:110–4.

17 Gellert M. V(D)J recombination: RAG proteins, repair factors, and regulation. Annu Rev Biochem 2002; **71**:101–32.

18 Schatz DG, Swanson PC. V(D)J recombination: mechanisms of initiation. Annu Rev Genet 2011; **45**:167–202.

19 Schatz DG, Ji Y. Recombination centres and the orchestration of V(D)J recombination. Nat Rev Immunol 2011; **11**:251–63.

20 Larmonie NS, Dik WA, Meijerink JP, Homminga I, van Dongen JJ, Langerak AW. Breakpoint sites disclose the role of the V (D) J recombination machinery in the formation of T-cell receptor (TCR) and non-TCR associated aberrations in T-cell acute lymphoblastic leukemia. Haematologica 2013; **98**:1173–84.

21 Clappier E, Auclerc M, Rapion J *et al.* An intragenic ERG deletion is a marker of an oncogenic subtype of B-cell precursor acute lymphoblastic leukemia with a favorable outcome despite frequent IKZF1 deletions. Leukemia 2014; **28**:70–7.

22 Mendes RD, Sarmento LM, Canté-Barrett K *et al.* PTEN microdeletions in T-cell acute lymphoblastic leukemia are caused by illegitimate RAG-mediated recombination events. Blood 2014; **124**:567–78.

23 Onozawa M, Aplan PD. Illegitimate V (D) J recombination involving nonantigen receptor loci in lymphoid malignancy. Genes Chromosomes Cancer 2012; **51**:525–35.

24 Cowell LG, Davila M, Kepler TB, Kelsoe G. Identification and utilization of arbitrary correlations in models of recombination signal sequences. Genome Biol 2002; **3**:126.

25 Ji Y, Resch W, Corbett E, Yamane A, Casellas R, Schatz DG. The *in vivo* pattern of binding of RAG1 and RAG2 to antigen receptor loci. Cell 2010; **141**:419–31.

26 Onciu M. Acute lymphoblastic leukemia. Hematol Oncol Clin North Am 2009; **23**:655–74.

27 Li S, Lew G. Is B-lineage acute lymphoblastic leukemia with a mature phenotype and l1 morphology a precursor B-lymphoblastic leukemia/lymphoma or Burkitt leukemia/lymphoma? Arch Pathol Lab Med 2003; **127**:1340–4.

28 Nelson BP, Treaba D, Goolsby C *et al.* Surface immunoglobulin positive lymphoblastic leukemia in adults; a genetic spectrum. Leuk Lymphoma 2006; **47**:1352–9.

29 Bene MC, Castoldi G, Knapp W *et al.* Proposals for the immunological classification of acute leukemias. European Group for the Immunological Characterization of Leukemias (EGIL). Leukemia 1995; **9**:1783–6.

30 Wang L, Howarth A, Clark RE. Ikaros transcripts Ik6/10 and levels of full-length transcript are critical for chronic myeloid leukaemia blast crisis transformation. Leukemia 2014; **28**: 1745–7.

31 Mijušković M, Chou YF, Gigi V *et al.* Off-target V(D)J recombination drives lymphomagenesis and is escalated by loss of the Rag2 C terminus. Cell Rep 2015; **12**:1842–52.

32 Dierov J, Sanchez PV, Burke BA *et al.* BCR/ABL induces chromosomal instability after genotoxic stress and alters the cell death threshold. Leukemia 2009; **23**:279–86.