Title
Mechanisms of clonal evolution in childhood acute lymphoblastic leukemia

Journal
Nature Immunology, 16(7)

ISSN
1529-2908

Authors
Swaminathan, S
Klemm, L
Park, E
et al.

Publication Date
2015-07-20

DOI
10.1038/ni.3160

nature immunology

Mechanisms of clonal evolution in childhood acute lymphoblastic leukemia

Srividya Swaminathan[1,11], Lars Klemm[1,2,11], Eugene Park[1,3], Elli Papaemmanuil[4], Anthony Ford[5], Soo-Mi Kweon[6], Daniel Trageser[6], Brian Hasselfeld[6], Nadine Henke[7], Jana Mooster[7], Huimin Geng[1], Klaus Schwarz[8], Scott C Kogan[1], Rafael Casellas[9], David G Schatz[10], Michael R Lieber[6], Mel F Greaves[5] & Markus Müschen[1]

Childhood acute lymphoblastic leukemia (ALL) can often be traced to a pre-leukemic clone carrying a prenatal genetic lesion. Postnatally acquired mutations then drive clonal evolution toward overt leukemia. The enzymes RAG1-RAG2 and AID, which diversify immunoglobulin-encoding genes, are strictly segregated in developing cells during B lymphopoiesis and peripheral mature B cells, respectively. Here we identified small pre-BII cells as a natural subset with increased genetic vulnerability owing to concurrent activation of these enzymes. Consistent with epidemiological findings on childhood ALL etiology, susceptibility to genetic lesions during B lymphopoiesis at the transition from the large pre-BII cell stage to the small pre-BII cell stage was exacerbated by abnormal cytokine signaling and repetitive inflammatory stimuli. We demonstrated that AID and RAG1-RAG2 drove leukemic clonal evolution with repeated exposure to inflammatory stimuli, paralleling chronic infections in childhood.

Childhood pre-B cell acute lymphoblastic leukemia (pre-B ALL) can frequently be traced to a pre-leukemic clone carrying a genetic lesion that was acquired *in utero* (for example, *ETV6-RUNX1*)[1,2]. During early childhood, pre-leukemic clones can acquire secondary mutations and evolve toward overt leukemia[3]. Leukemia cells follow Darwinian trajectories of clonal evolution, in which sequentially acquired genetic lesions increase or decrease the adaptive fitness of pre-leukemic clones, which results in their positive or negative selection[4]. While this concept is well established, the mechanism(s) that drive(s) clonal evolution is (are) not known. Epidemiological studies have suggested a model of clonal evolution that is driven by recurrent infections during childhood[4–8]. It has been hypothesized that delayed exposure of infants to infections in the highly protective environments of developed societies causes protracted and excessive immune responses with substantial collateral damage[6,8]. Early exposure to common infections (such as via attendance at a daycare facility) mitigates this risk[5,8]. The systematic introduction of vaccination programs during early childhood has reduced the incidence of ALL, presumably by reducing chronic infections and the immune responses that they instigate[7]. For example, the administration of vaccine against *Haemophilus influenzae* type b in early childhood significantly reduced the risk of childhood ALL in three different case-control studies of American, Finnish and Canadian cohorts[9–11].

Altered cytokine environments in the context of inflammation eliminate multiple normal pre-B cell clones from the repertoire and favor the selective outgrowth of pre-B cell clones that already harbor a pre-leukemic genetic lesion, such as *ETV6-RUNX1* (ref. 12). *ETV6-RUNX1* was identified as a prenatally acquired genetic rearrangement[13]. Additional postnatally acquired lesions over the course of up to 15 years can give rise to leukemia in the subgroup of patients carrying the *ETV6-RUNX1* lesion. While *ETV6-RUNX1* is frequently found in umbilical cord B cells and neonatal Guthrie blood spots[13], less than 1% of newborn children carrying *ETV6-RUNX1* eventually develop ALL[14]. Such studies have led to the idea that persistent infection and delayed exposure to strong inflammatory stimuli in childhood may increase the risk of acquiring the post-natal genetic lesions. However, the mechanistic basis of how such pre-leukemic clones evolve has remained elusive.

Two classes of enzymes are required for the somatic recombination and mutation of immunoglobulin-encoding genes in B cells: the proteins encoded by recombination-activating genes (RAG1 and RAG2) introduce DNA double-strand breaks and recombine segments encoding the variable (V), diversity (D) and joining (J) regions[15]; and AID deaminates cytosine residues in immunoglobulin V and switch regions and thereby enables somatic hypermutation (SHM) and class-switch recombination (CSR) of immunoglobulin-encoding genes[16]. We note that genetic diversification of antibodies

[1]Department of Laboratory Medicine, University of California San Francisco, San Francisco, California, USA. [2]University of Freiburg, Faculty of Biology, Freiburg, Germany. [3]Department of Haematology, University of Cambridge, Cambridge, UK. [4]Cancer Genome Project, Wellcome Trust Sanger Institute, Hinxton, UK. [5]Centre for Evolution and Cancer, The Institute of Cancer Research, London, UK. [6]University of Southern California, Los Angeles, California, USA. [7]Heinrich-Heine-Universität Düsseldorf, Düsseldorf, Germany. [8]Institute for Transfusion Medicine, University of Ulm, Ulm, Germany. [9]Genomics & Immunity, NIAMS, NIH, Bethesda, Maryland, USA. [10]Yale University, New Haven, Connecticut, USA. [11]These authors contributed equally to this work. Correspondence should be addressed to M.M. (markus.muschen@ucsf.edu).

Received 20 January; accepted 26 March; published online 18 May 2015; doi:10.1038/ni.3160

by AID and RAG1-RAG2 is activated in response to infectious and inflammatory stimuli, both in early B cells[17] and in mature B cells[18]. A comprehensive breakpoint analysis[19] has suggested that some chromosomal rearrangements in human B cell malignancies may result from cooperation between RAG1-RAG2 and AID. During normal B cell development, however, the enzymatic activities of these proteins are normally strictly segregated[20].

Here we report that the enzymes noted above not only genetically diversified the B lymphocyte repertoire[21] but also contributed to the acquisition of genetic lesions and promoted clonal evolution toward leukemia. We also delineate the molecular mechanism by which the normal physiological process of antibody diversification is subverted in pre-leukemic B cell clones (*ETV6-RUNX1*), leading to overt disease.

RESULTS

Genes frequently deleted in childhood ALL are targets of AID

RAG1 and RAG2 are constitutively expressed in pro- and pre-B cells and target recombination signal sequences in both genes encoding immunoglobulins and those that do not encode immunoglobulins. Motif analyses of deletion breakpoints in the tumor suppressor–encoding genes *IKZF1* (ref. 22), *CRLF2*, *TBL1XR1*, *CDKN2A*, *NR3C1*, *NR3C2* and *BTG1* (ref. 23) has revealed that the recombination signal sequence motifs in these genes are indeed targeted by RAG1-RAG2.

To test the plausibility of the contribution of AID to RAG1-RAG2-mediated deletions and rearrangements, we assessed whether lesions in childhood ALL (data from the Children's Oncology Group (COG) clinical trial P9906) occur 'preferentially' at genes bound by AID in normal B cells. We identified 40 genes that were recurrently mutated, deleted or rearranged in childhood ALL (207 patients in trial P9906) and plotted the frequency of these lesions relative to the targeting of these genes by mouse AID (mAID) in mouse B cells. 4,167 mAID-target genes were identified in a published study by chromatin immunoprecipitation followed by deep sequencing (ChIP-seq), on the basis of a greater number of mAID ChIP-seq tags in *Aicda*$^{+/+}$ B cells than in AID-deficient (*Aicda*$^{-/-}$) B cells[24]. 18,248 genes that were not targets of mAID did not have significantly higher mAID ChIP-seq tag counts in *Aicda*$^{+/+}$ B cells than in than in *Aicda*$^{-/-}$ B cells[24]. Notably, the frequency of recurrent genetic lesions correlated significantly with mAID ChIP-seq counts (**Fig. 1a**). In this analysis, 34 of 40 genes with recurrent lesions in childhood ALL were targets of mAID, which indicated a possible contribution of AID to frequent genetic lesions in childhood ALL. Of the remaining 6 genes that were not targets of mAID, 5 (*CDKN2A*, *TBL1XR1*, *NR3C1*, *NR3C2* and *PBX1*)[23,25] have breakpoints at known recombination signal sequence motifs with characteristic N-nucleotide addition at the break junctions, which suggests that the lesions at these loci were caused by RAG1-RAG2 activity alone. We note that not all targets of AID are syntenic between humans and mice. While this represents a limitation to our approach, disparity between targets of AID in humans and mice will only lead to an underestimation of the correlation we observed. Moreover, a published study identifying targets of human AID in B cell lymphoma[26], although not quantitative, includes various genes from our analysis (*PAX5*, *BTG1*, *BTG2*, *BACH2*, *RHOH*, *EBF1* and *TCF3*). These studies confirmed that AID might be responsible for mutating a substantial percentage of the genes frequently found to be altered in pre-B ALL.

Abundant *AICDA* and *RAG1* mRNA indicates poor ALL outcome

If AID allows clonal evolution toward childhood ALL, its expression and activity might correlate with clinical outcomes of patients. To test this possibility, we correlated *AICDA* mRNA expression in patients at diagnosis with overall and relapse-free survival. We segregated patients with ALL in two clinical trials (Eastern Cooperative Oncology Group (ECOG) trial E2993 and COG trial P9906) into two groups on the basis of *AICDA* mRNA abundance at the time of diagnosis that was higher or lower than the median. The ECOG E2993 trial included 215 patients (106 bone marrow samples and 109 peripheral blood samples), and the COG P9906 trial had 207 pediatric patients (131 bone marrow samples and 76 peripheral blood samples).

Notably, higher-than-median *AICDA* expression was strongly indicative of poor overall patient survival ($P = 0.026$; **Supplementary Fig. 1a**, left). Likewise, higher-than-median *RAG1* mRNA abundance was indicative of shorter relapse-free and overall survival of patients

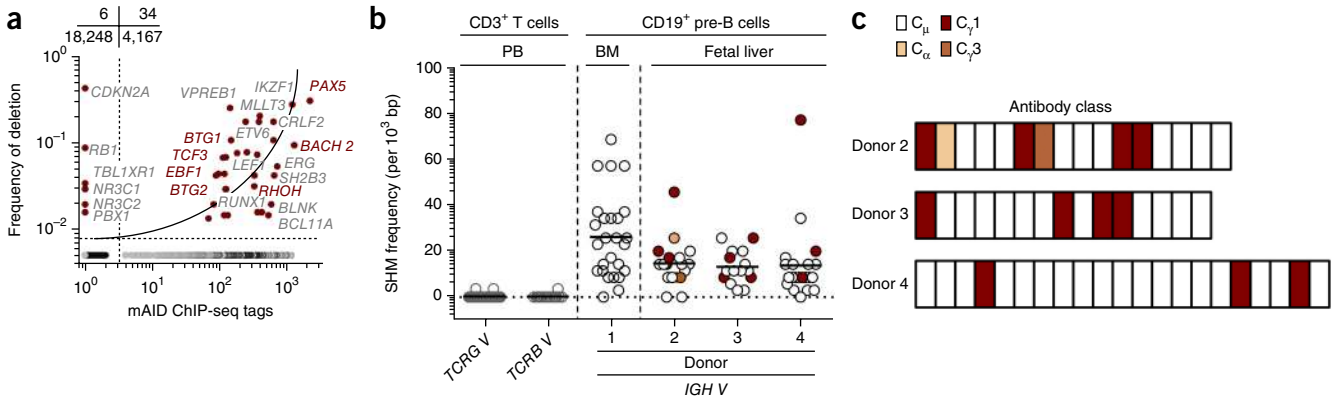

Figure 1 Expression and activity of AID in human B cell precursors and B lineage ALL. (**a**) Correlation between genes with somatic hypermutation that are frequently bound and targeted by mAID[24] (horizontal axis; normalized to *Aicda*$^{-/-}$ B cells) versus the frequency of deletion of genes commonly found in childhood ALL (vertical axis; data from clinical trial P9906; $n = 207$ subjects), by a two-by-two contingency table analysis ($P = \chi^2$ test with Yates' correction). Red symbols indicate known targets of AID in human B cells; gray circles (above horizontal axis) indicate genes without correlation to genes altered in childhood ALL. Inset (top left), numbers indicate total genes in each area marked in the plot below. (**b**) SHM frequency of *IGH* V_H regions (*IGH* V) in single pre-B cell clones isolated from human bone marrow (BM) ($n = 1$ donor) and fetal liver ($n = 3$ donors), and in V regions of *TCRG* (*TCRG* V) and *TCRB* (*TCRB* V) in CD3$^+$ T cells from peripheral blood (PB) (negative control). $P = 3.3 \times 10^{-18}$ (Student's *t*-test). (**c**) CSR in early human B cells from fetal liver ($n = 3$ donors), presented as the absolute number of clones with C_μ, C_α, $C_\gamma 3$ and $C_\gamma 1$ regions (each rectangle indicates one clone). Data are representative of one experiment with 207 patients (**a**), one experiment with one bone marrow donor and three fetal liver donors (**b**) or one experiment with three donors (**c**).

with ALL (**Supplementary Fig. 1a**, center and right). In the COG P9906 trial, 49 patients relapsed after successful initial therapy. In a comparison of *AICDA* mRNA expression in matched sample pairs at diagnosis and relapse, for most patients, *AICDA* mRNA abundance was increased at relapse (**Supplementary Fig. 1b**). These findings suggested that *AICDA* expression at diagnosis can be used to predict the outcome of patients with ALL.

RAG and AID proteins are active in precursors of human B cells

Pre-B ALL arises from sites of early B lymphopoiesis (fetal liver and bone marrow) during prenatal development and postnatal development, respectively. RAG enzymes are active in precursors of B cells in both sites. Studies of identical twins with concordant *ETV6-RUNX1* lesions pre-B ALL have indicated clonal origin of cells undergoing RAG-dependent rearrangement of the immunoglobulin heavy-chain locus (*IGH*) V region and continued recombinant activity during subsequent clonal evolution[13,27]. To measure AID activity in precursors of human fetal liver and bone marrow B cells, we sorted $CD19^+$ pre-B cells lacking κ- and λ-chains (immunoglobulin light chains) from liver tissues of three fetal donors and bone marrow of one healthy adult donor. We then cloned and sequenced the *IGH* V regions (V_H) to assess SHM and cloned and sequenced the *IGH* constant regions (C_H) to assess CSR (**Supplementary Tables 1** and **2**). Notably, most bone marrow pre-B cell clones expressed *IGH* with mutated V_H regions (with a frequency of 26 mutations per 10^3 base pairs (bp); **Fig. 1b** and **Supplementary Table 1**). Likewise, liver pre-B cells from the three fetal donors carried mutated V_H regions (14 mutations per 10^3 bp; **Fig. 1b** and **Supplementary Table 2**). To control for errors induced by reverse transcriptase and Pfu DNA polymerase, we sorted $CD3^+$ T cells lacking *AICDA* expression from peripheral blood, amplified rearranged V regions from genes encoding the T cell antigen receptor (TCR) β-chain (*TCRB*) and γ-chain (*TCRG*) and analyzed these in parallel with *IGH* V_H regions of bone marrow and fetal liver pre-B cells. Overall, *IGH* V_H regions in sorted bone marrow and fetal liver pre-B cells carried a significantly greater frequency of somatic mutations than did the *TCRB* and *TCRG* V regions in peripheral blood T lymphocytes (**Fig. 1b**). We also found that a fraction of the sorted $CD19^+$ pre-B cells sorted from fetal liver tissues expressed class-switched *IGH* and contained $C_\gamma3$, $C_\gamma1$ and C_α regions (**Fig. 1b**,**c** and **Supplementary Table 2**). We note that our method of *IGH* cloning and sequencing was not strictly quantitative. Therefore, we were not able to predict the exact fraction of fetal liver pre-B cells that expressed class-switched *IGH* C regions. These findings suggested that AID might be prematurely activated in human bone marrow and fetal liver pre-B cells. While we found genetic traces of AID activity in RAG1-RAG2–expressing pre-B cells, these findings did not establish concurrent activity of AID and RAG1-RAG2 in the same cells.

Small pre-BII cells have increased genetic vulnerability

To elucidate whether and how premature activity of mAID is regulated during early B lymphopoiesis, we sorted B cell precursors from normal mouse bone marrow as described[20] by flow cytometry. We isolated pro-B cells ($c\text{-}Kit^+B220^+$), large pre-BII cells ($CD25^+B220^+FSC^{hi}SSC^{lo}$) and small pre-BII ($CD25^+B220^+$ $FSC^{lo}SSC^{lo}$). As large and small pre-BII cells carry similar surface markers, we used their size differences (forward scatter (FSC)) to separate them. We then measured

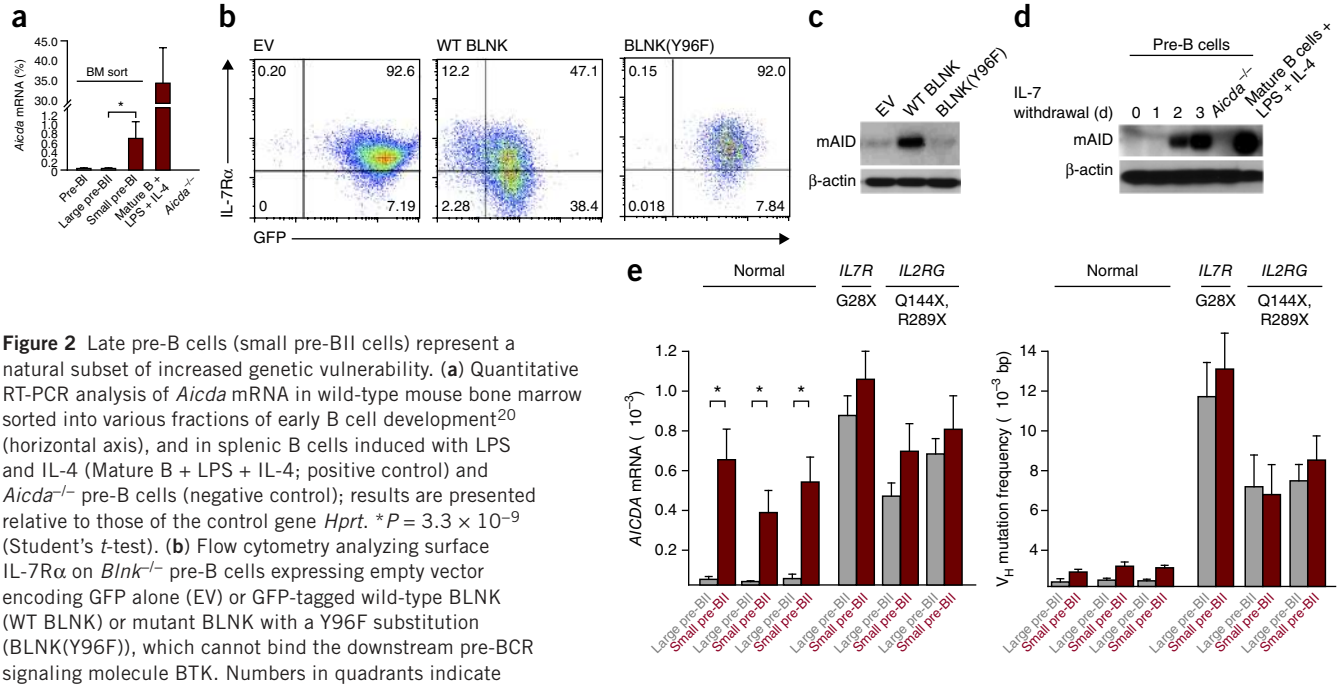

Figure 2 Late pre-B cells (small pre-BII cells) represent a natural subset of increased genetic vulnerability. (**a**) Quantitative RT-PCR analysis of *Aicda* mRNA in wild-type mouse bone marrow sorted into various fractions of early B cell development[20] (horizontal axis), and in splenic B cells induced with LPS and IL-4 (Mature B + LPS + IL-4; positive control) and $Aicda^{-/-}$ pre-B cells (negative control); results are presented relative to those of the control gene *Hprt*. $^*P = 3.3 \times 10^{-9}$ (Student's *t*-test). (**b**) Flow cytometry analyzing surface IL-7Rα on $Blnk^{-/-}$ pre-B cells expressing empty vector encoding GFP alone (EV) or GFP-tagged wild-type BLNK (WT BLNK) or mutant BLNK with a Y96F substitution (BLNK(Y96F)), which cannot bind the downstream pre-BCR signaling molecule BTK. Numbers in quadrants indicate percent living cells in each. (**c**) Immunoblot analysis of mAID and β-actin (loading control throughout) in $Blnk^{-/-}$ pre-B cells transduced as in **b**. (**d**) Immunoblot analysis of mAID and β-actin in mouse pre-B cell cultures before (0) and after withdrawal of IL-7 (above lanes), as well as in $Aicda^{-/-}$ cells and in mature B cells treated with LPS and IL-4. (**e**) Abundance of *AICDA* mRNA (left) and frequency of V_H mutations (right) in large pre-BII cells and small pre-BII pre-B cells in bone marrow from children with mutation in one of the chains encoding IL-7R (above plots) and from human donors without such mutation (Normal), as a negative control; mRNA results are presented relative to those of the control gene *COX6B*. Mean V_H mutation frequency, 1.51 ($\times10^{-3}$ bp; Normal) and 8.16 ($\times10^{-3}$ bp; *IL7R* or *IL2RG*), $P = 0.0008$ (not shown). $^*P = 0.001$ (Student's *t*-test). Data are representative of two experiments (**a**; mean and s.d. of $n = 3$ technical replicates) or one experiment (**e**; mean and s.d. of $n = 3$ technical replicates from three healthy donors and three donors with *IL7R* or *IL2RG* mutation) or are from one experiment representative of at least two experiments (**b**–**d**).

the abundance of *Aicda* mRNA in each fraction of early B cell development (**Fig. 2a**). *Aicda* mRNA was generally less abundant in (or absent from) bone marrow precursors of B cells than in activated mature splenic B cells (**Fig. 2a**). However, *Aicda* mRNA was ~21-fold more abundant at the small pre-BII cell than at the large pre-BII cell stage of early B lymphopoiesis (**Fig. 2a**). This transition is characterized by signals from the pre-B cell antigen receptor (pre-BCR) and its intracellular signaling components that are assembled by the pre-BCR linker BLNK[28,29]. One consequence of pre-BCR signaling at the transition from large pre-BII cell to small pre-BII cell was downregulation of surface expression of the interleukin 7 receptor α-chain (IL-7Rα) by activation of BLNK (**Fig. 2b**). Attenuation of IL-7R expression dephosphorylates and inactivates the signal transducer STAT5 in small pre-BII cells[30]. Therefore, to investigate whether the abundance of *Aicda* mRNA was increased in small pre-BII cells, we abrogated IL-7R–STAT5 signaling via two methods. First, we reconstituted BLNK in *Blnk*$^{-/-}$ pre-B cells to induce downregulation of IL-7R expression and measured the abundance of *Aicda* mRNA (**Supplementary Fig. 2a**) and mAID protein (**Fig. 2c**). Second, we removed IL-7 from pre-B cell cultures (**Fig. 2d** and **Supplementary Fig. 2b**) to force transition from the large pre-BII cell stage to the small pre-BII cell stage. *In vitro* differentiation by each method induced upregulation of mAID protein expression, which demonstrated that active IL-7R signaling in mouse large pre-BII cells served as a safeguard against premature mAID expression.

IL-7R safeguards human pre-B cells against AID activation

We next investigated whether IL-7R signaling protects human pre-B cells from premature AID expression. While IL-7R signaling is essential for B cell development in mice, humans with inherited *IL7R* deficiency have almost normal B cell counts[31], which allows experimental analysis of human B lymphopoiesis in the absence of IL-7R signaling. We isolated large and small pre-BII cells from bone marrow samples of three patients with biallelic germline mutations in either gene encoding IL-7R (*IL7R* and *IL2RG*) that result in a lack of functional IL-7R signaling. One patient harbored a mutation in *IL7R* resulting in a G28X substitution in IL-7R; the two other patients each carried a mutation in *IL2RG* resulting in a Q144X or R289X substitution in IL-7R (**Fig. 2e**). As a control, we studied large and small pre-BII cells sorted from the bone marrow of three healthy donors. As in mouse B lymphopoiesis, we observed that human pre-B cells upregulated *AICDA* mRNA by 10- to 25-fold at the transition from the large pre-BII cell stage to the small pre-BII cell stage (**Fig. 2e**). The pre-B cells from the three patients lacking functional IL-7R signaling expressed more *AICDA* mRNA than did those from the healthy control subjects (**Fig. 2e**). We observed no significant difference between large pre-BII cell subset and small pre-BII cell subset from IL-7R-deficient patients in their abundance of *AICDA* mRNA (**Fig. 2e**). Amplification and sequencing of *IGH* V_H regions revealed a low frequency of SHM in small pre-BII cells from healthy control subjects (**Fig. 2e**). The frequency of somatic mutations in the *IGH* V_H region was about fivefold higher in both large pre-BII cells and small pre-BII cells of patients lacking functional IL-7R than in those from healthy control subjects (**Fig. 2e**). Thus, IL-7R served as a safeguard against premature activation of AID in human pre-B cells.

Safeguard mechanism of IL-7R against AID activation

IL-7R–STAT5 signaling in early pre-B cells[32] prevents upregulation of the expression of *Rag1-Rag2* mRNA and recombination of V and J segments in the locus encoding the immunoglobulin κ-chain[32,33]. Downregulation of IL-7R expression attenuated the activity of STAT5 and the kinase Akt and induced transcription of *Rag1* mRNA (**Supplementary Fig. 2c**). STAT5 recruits the polycomb repressor EZH2 to *Rag1-Rag2*, while Akt phosphorylates and inactivates Foxo1, a potent transcriptional activator of *Rag1* and *Rag2* (ref. 34). We found that inducible, Cre recombinase–mediated deletion of

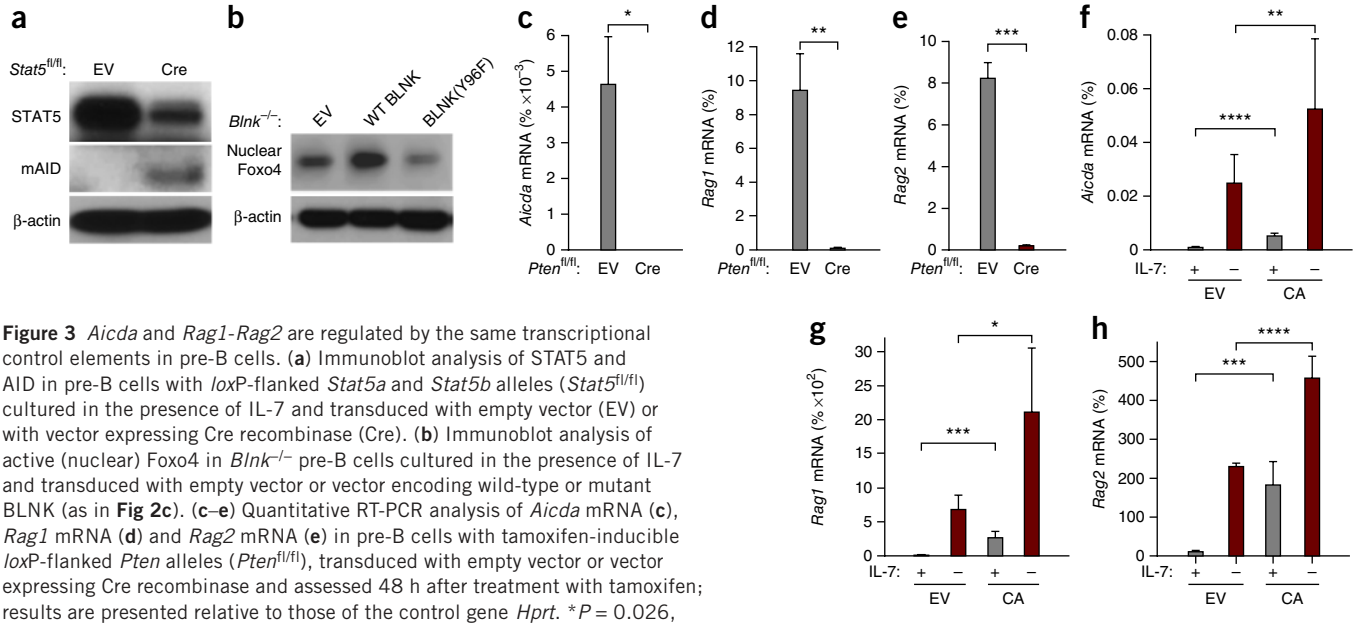

Figure 3 *Aicda* and *Rag1-Rag2* are regulated by the same transcriptional control elements in pre-B cells. (**a**) Immunoblot analysis of STAT5 and AID in pre-B cells with *lox*P-flanked *Stat5a* and *Stat5b* alleles (*Stat5*$^{fl/fl}$) cultured in the presence of IL-7 and transduced with empty vector (EV) or with vector expressing Cre recombinase (Cre). (**b**) Immunoblot analysis of active (nuclear) Foxo4 in *Blnk*$^{-/-}$ pre-B cells cultured in the presence of IL-7 and transduced with empty vector or vector encoding wild-type or mutant BLNK (as in **Fig 2c**). (**c**–**e**) Quantitative RT-PCR analysis of *Aicda* mRNA (**c**), *Rag1* mRNA (**d**) and *Rag2* mRNA (**e**) in pre-B cells with tamoxifen-inducible *lox*P-flanked *Pten* alleles (*Pten*$^{fl/fl}$), transduced with empty vector or vector expressing Cre recombinase and assessed 48 h after treatment with tamoxifen; results are presented relative to those of the control gene *Hprt*. **P* = 0.026, **0.018 and ****P* = 0.003 (Student's *t*-test). (**f**–**h**) Quantitative RT-PCR analysis of *Aicda* mRNA (**f**), *Rag1* mRNA (**g**) and *Rag2* mRNA (**h**) in mouse pre-B cells cultured in the presence of IL-7 and transduced with retrovirus expressing empty vector (EV) or a constitutively active form of Foxo1 (CA), then cultured for 24 h in the presence of IL-7 (+) or after withdrawal of IL-7 (–); results presented as in **c**–**e**. **P* = 0.06, ***P* = 0.04, ****P* = 0.039 and *****P* = 0.002 (Student's *t*-test). Data are representative of at least two independent experiments (**a**,**b**), two experiments (**c**–**e**; mean and s.d. of *n* = 3 technical replicates) or at least two experiments (**f**–**h**; mean and s.d. of *n* = 3 technical replicates).

Stat5a and *Stat5b* in large pre-BII cells resulted in the upregulation of AID expression (**Fig. 3a**). We assessed the mechanistic role of Akt-Foxo1 signaling in regulating the abundance of *Aicda* mRNA through the use of genetic loss of function of PTEN, a negative regulator of Akt, and gain-of-function studies of Foxo1. Inducible deletion of *Pten* results in hyperactivation of Akt and, hence, phosphorylation and inactivation of Foxo transcription factors[35]. Pre-BCR signaling via BLNK not only decreased the surface expression of IL-7R (**Fig. 2b**) but also caused activation of Foxo factors[35] such as Foxo4 (**Fig. 3b**). Cre-mediated inducible ablation of *Pten* caused nearly complete loss of the expression of *Rag1*, *Rag2* and *Aicda* mRNA (**Fig. 3c–e**). In agreement with those findings, a constitutively active form of Foxo1 that is protected from Akt-mediated phosphorylation substantially activated the transcription of *Rag1*, *Rag2* and *Aicda* (**Fig. 3f–h**). Notably, these effects were further enhanced by removal of IL-7 from the cell culture medium, to elicit differentiation into small pre-BII cells (**Fig. 3f–h** and **Supplementary Fig. 2d**). In summary, IL-7R signaling protected pre-B cells from premature expression of *Aicda* through STAT5- and Akt-dependent pathways.

Concurrent RAG and AID activity in pre-B cell clones

AID expression is upregulated in mature B cells upon antigen encounter in germinal center, where it introduces SHM in *Igh* V_H regions and promotes CSR of C_H regions[16,36]. Antigen encounter in germinal centers can be mimicked by the treatment of mature splenic B cells with lipopolysaccharide (LPS) and IL-4. We investigated whether small pre-BII cells respond to inflammatory agents such as LPS in a manner similar to that of mature B cells, by upregulating mAID expression. To measure such expression of mAID in pre-B cells at the single-cell level, we studied pre-B cells from an *Aicda*-GFP reporter mouse strain in which the gene encoding green fluorescent protein (GFP) is fused in-frame to exon 5 of *Aicda* (*Aicda*-GFP mice, which produce mAID-GFP)[37]. We propagated *Aicda*-GFP pre-B cells in IL-7 and then left the cells untreated or treated them with LPS, in the continued presence or withdrawal of IL-7. Treatment with LPS induced a minimal increase in mAID expression (**Fig. 4a** and **Supplementary Fig. 3a,b**). However, combined treatment with LPS plus withdrawal of IL-7 resulted in 20-fold greater upregulation of mAID expression than did treatment with LPS in the presence of IL-7 (**Fig. 4a,b**). To measure the expression of mAID and activity of RAG1-RAG2 in parallel at the single-cell level, we monitored mAID-GFP and *de novo* surface expression of immunoglobulin κ-light chains after RAG-mediated rearrangement of $V_κ$-$J_κ$ gene segments. Treatment of *Aicda*-GFP pre-B cells with LPS in the presence of IL-7 induced activation of the *Aicda*-GFP reporter in 2.8% of cells, compared with its activation in 0.01% in the condition without LPS treatment (**Fig. 4a** and **Supplementary Fig. 3b**). However, the combination of treatment with LPS plus with removal of IL-7 from the cell culture medium increased the fraction of mAID-expressing cells to 42%, from 2.8% in the presence of IL-7 (**Fig. 4a** and **Supplementary Fig. 3b**). Notably, the majority of *Aicda*-GFP cells also had *de novo* expression of κ-light chains (**Fig. 4a** and **Supplementary Fig. 3b**), which provided direct evidence that mAID expression and RAG-mediated recombination of $V_κ$ and $J_κ$ gene segments occurred in the same cells. We confirmed those findings in a second *Aicda* reporter mouse model, in which mAID drives expression of Cre for excision of a *lox*P-flanked 'stop' cassette encoding an enhanced yellow fluorescent protein marker, located in the ubiquitous *Rosa26* locus[37] (**Supplementary Fig. 3c,d**).

Evidence of ongoing RAG and AID activity in human pre-B ALL

Studying *IGH* V_H regions amplified from 72 patients with childhood ALL (13 with the *ETV6-RUNX1* lesion, 2 with the *TCF3-PBX1* lesion, 3 with rearrangement of *MLL*, 1 with the *MYC-IGH* lesion, 17 with hyperdiploidy, 26 with a normal or undetermined karyotype, and 10 with sporadic chromosomal translocations), we found evidence of ongoing RAG1-RAG2 and AID activity in all subgroups (**Supplementary Tables 3–5**). Pre-B ALL clones consistently carried somatically mutated V_H and J_H gene segments (**Supplementary Tables 3–5**), indicative of AID activity. Notably, in 10 of 72 patients, we found subclones in each that carried the same defining D-J_H junction (**Supplementary Tables 3–5**). Multiple distinct V_H segments diversified these subclones (**Supplementary Tables 3–5**) through

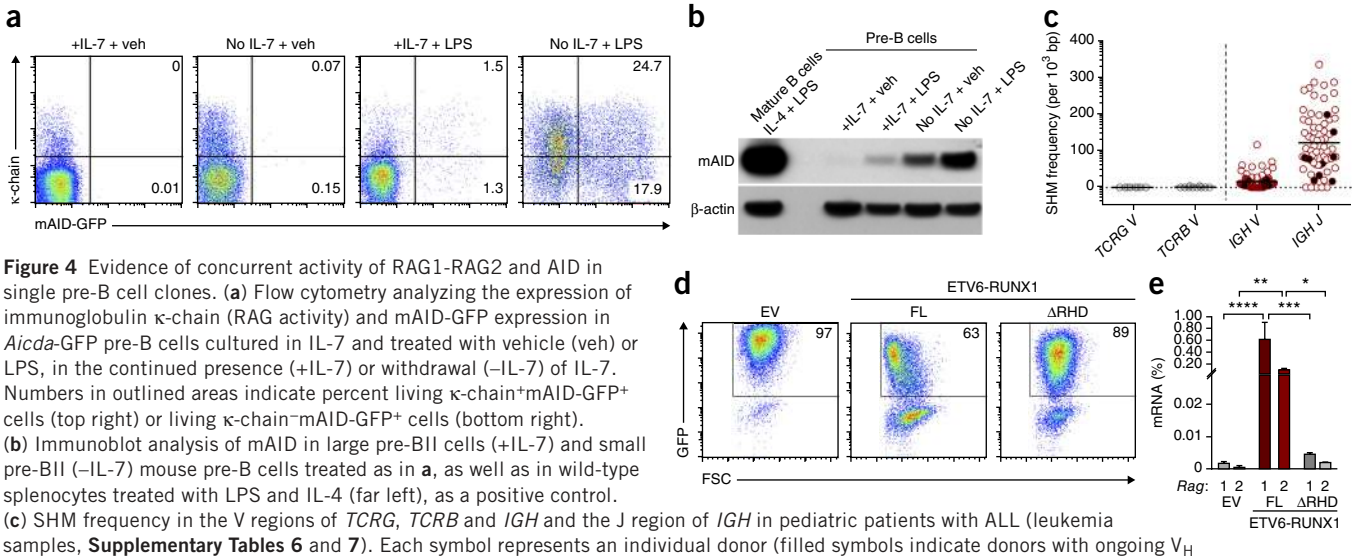

Figure 4 Evidence of concurrent activity of RAG1-RAG2 and AID in single pre-B cell clones. (**a**) Flow cytometry analyzing the expression of immunoglobulin κ-chain (RAG activity) and mAID-GFP expression in *Aicda*-GFP pre-B cells cultured in IL-7 and treated with vehicle (veh) or LPS, in the continued presence (+IL-7) or withdrawal (–IL-7) of IL-7. Numbers in outlined areas indicate percent living κ-chain$^+$mAID-GFP$^+$ cells (top right) or living κ-chain$^-$mAID-GFP$^+$ cells (bottom right). (**b**) Immunoblot analysis of mAID in large pre-BII cells (+IL-7) and small pre-BII (–IL-7) mouse pre-B cells treated as in **a**, as well as in wild-type splenocytes treated with LPS and IL-4 (far left), as a positive control. (**c**) SHM frequency in the V regions of *TCRG*, *TCRB* and *IGH* and the J region of *IGH* in pediatric patients with ALL (leukemia samples, **Supplementary Tables 6** and **7**). Each symbol represents an individual donor (filled symbols indicate donors with ongoing V_H replacement); small horizontal lines indicate the mean. (**d**) Flow cytometry of mouse pre-B cells on day 2 after transduction (via retroviral vectors) to overexpress full-length ETV6-RUNX1 (FL) or mutant ETV6-RUNX1 lacking the DNA-binding runt homology domain (ΔRHD), or empty vector (EV), as a negative control. (**d**) Numbers in outlined areas percent GFP$^+$ (transduced) living cells. FSC, forward scatter. (**e**) Quantitative RT-PCR analysis of *Rag1* mRNA (*1*) and *Rag2* mRNA (*2*) in the GFP$^+$ cells in **d**; results are presented relative to those of *Hprt*. $^*P = 0.010$, $^{**}P = 0.009$, $^{***}P = 0.003$ and $^{****}P = 0.002$ (Student's *t*-test). Data are representative of three experiments (**a**,**b**), one experiment (**c**), two experiments with one of three technical replicates (**d**) or two experiments (**e**; mean and s.d. of $n = 3$ technical replicates).

sequential RAG-mediated V_H replacement[38]. We present here the process of leukemic clonal evolution by cooperative activities of AID (SHM) and the RAG recombinase (V_H replacement) in a pediatric patient with pre-B ALL (**Supplementary Fig. 4** and **Supplementary Table 5**). Comparing mutation frequencies in V_H and J_H gene segments, we found that J_H gene segments carried a significantly higher load of somatic mutations (102 per 10^3 bp (J_H) versus 16 per 10^3 bp (V_H); $P = 0.0007$; **Fig. 4c** and **Supplementary Tables 3–5**). Multiple rounds of V_H replacement, during which acquired mutations in V_H segments are erased but D-J_H junctions remain constant, could explain this unexpected difference. In 7 of 13 patients with pre-B ALL with the *ETV6-RUNX1* lesion, we found multiple distinct subclones that had undergone V_H replacement (**Supplementary Table 3**), which suggested that RAG enzymes were particularly active in this subset. This finding was in agreement with a study highlighting aberrant activation of RAG-mediated recombination as the primary driver of *ETV6-RUNX1* leukemogenesis[23]. To determine if ETV6-RUNX1 activity is linked to deregulated RAG expression, we transduced mouse pre-B cells with vector encoding full-length ETV6-RUNX1 or mutant ETV6-RUNX1 lacking the DNA-binding runt homology domain, or with empty vector, as a negative control. Notably, full-length ETV6-RUNX1 induced expression of *Rag1-Rag2* in mouse pre-B cells, but the mutant ETV6-RUNX1 lacking the DNA-binding domain did not (**Fig. 4d,e**); this suggested that the abnormally high recombinase activity in this ALL subset was caused by the ETV6-RUNX1 fusion. Hence, cooperative AID and RAG activities might explain the mechanism of clonal evolution in *ETV6-RUNX1* leukemias.

Enrichment for cytosine deamination of CpG in patients with ALL

As the targets of AID are frequently deleted or mutated in childhood ALL, we analyzed mutational 'hot-spot' targeting in samples derived from 24 patients with ALL: 10 with the *ETV6-RUNX1* lesion, 4 with the *MLL-AF4* lesion, 1 with the *TCF3-PBX1* lesion, and 9 lacking a recurrent chromosomal translocation (**Supplementary Table 6**; mutation data not shown). In agreement with two published studies[23,39], whole-exome sequencing revealed substantial enrichment for C-to-T transitions at CpG dinucleotides, which accounted for 44% of all substitutions. Cytosine deamination at methylated CpG dinucleotides indicates the involvement of AID-mediated deamination of methylated CpG[19]. However, in agreement with published studies[23,38], somatic mutations in samples from human subjects with ALL did not reveal the typical hot-spot targeting described for SHM of *IGH* V_H gene segments (such as the DNA sequences RGYW or WRCY, where 'R' indicates A or G, 'Y' indicates C or T, and 'W' indicates A or T)[40]. The targeting of hot spots by AID typically emerges from a large number of somatic mutations in *IGH* V_H gene segments that were non-productively rearranged; hence, the mutations do not affect the outcome of clonal selection among B cells. The mutations in genes not encoding immunoglobulins in the cells from subjects with ALL studied here occurred at much lower frequencies and in coding regions, which left open the possibility that hot-spot targeting had been obscured by low mutation frequencies and effects of clonal selection.

Coordinate AID-RAG activity subverts genetic integrity in B cells

On the basis of our findings reported above, we propose that premature AID expression in small pre-BII cells represents a vulnerability that exposes early B lymphocyte development to genetic lesions in the context of repeated exposure to inflammatory stimuli (**Supplementary Fig. 5**). We therefore assessed the relevance of the cooperation of RAG and AID in clonal evolution in normal cord blood–derived human B cells through a genetic gain-of-function experiment. We sorted $CD19^+$ B cells from human cord blood and immortalized the cells through treatment with supernatants of B95-8 marmoset B cell cultures containing Epstein-Barr virus (EBV). We then transduced the proliferating cord blood B cells with vector encoding AID tagged with the red fluorescent protein iRFP670, RAG1 tagged with enhanced greed fluorescent protein (eGFP), or RAG2 tagged with the red fluorescent protein dsRedE2, either alone or in combination, or empty-vector controls for the AID, RAG1 and RAG2 vectors (**Fig. 5a,b** and **Supplementary Figs. 6** and **7**). We sorted transduced cord blood B cells in each condition and confirmed by immunoblot analysis the expression of AID, RAG1 and RAG2 (or lack thereof) (**Fig. 5a**). Our finding that AID expression was absent in EBV-infected B cells was consistent with published literature. EBV suppresses AID expression[41],

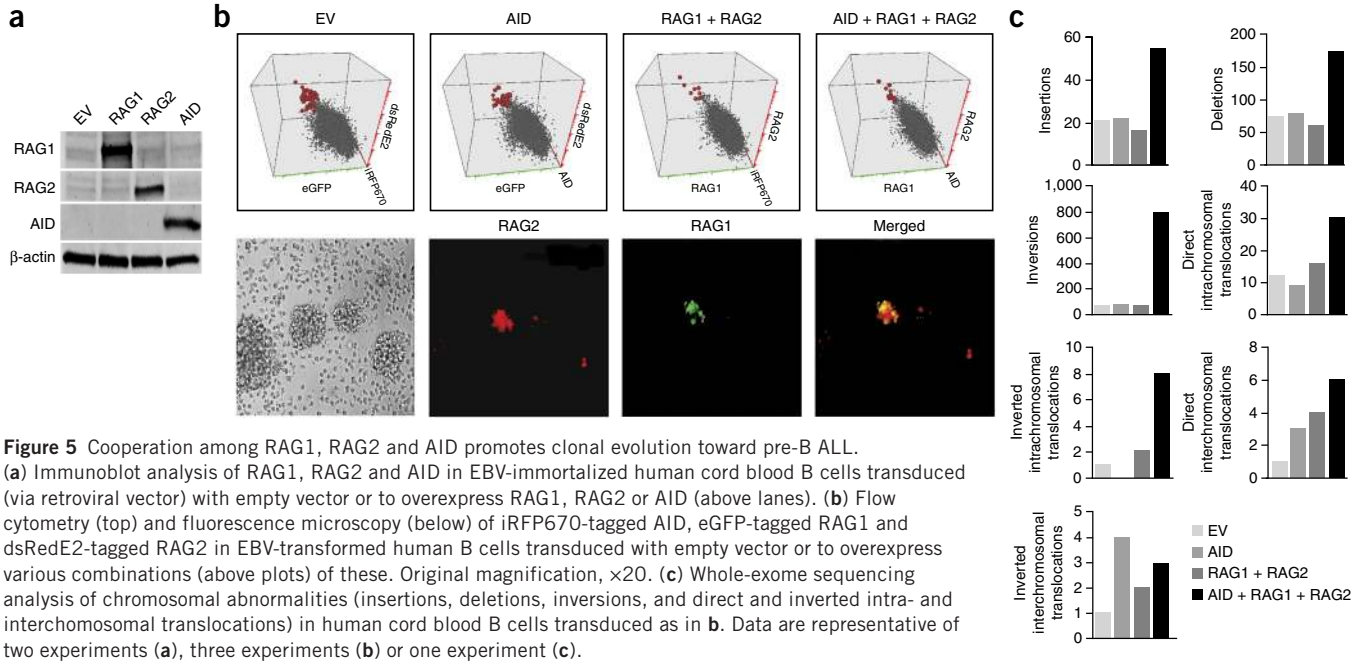

Figure 5 Cooperation among RAG1, RAG2 and AID promotes clonal evolution toward pre-B ALL. (**a**) Immunoblot analysis of RAG1, RAG2 and AID in EBV-immortalized human cord blood B cells transduced (via retroviral vector) with empty vector or to overexpress RAG1, RAG2 or AID (above lanes). (**b**) Flow cytometry (top) and fluorescence microscopy (below) of iRFP670-tagged AID, eGFP-tagged RAG1 and dsRedE2-tagged RAG2 in EBV-transformed human B cells transduced with empty vector or to overexpress various combinations (above plots) of these. Original magnification, ×20. (**c**) Whole-exome sequencing analysis of chromosomal abnormalities (insertions, deletions, inversions, and direct and inverted intra- and interchomosomal translocations) in human cord blood B cells transduced as in **b**. Data are representative of two experiments (**a**), three experiments (**b**) or one experiment (**c**).

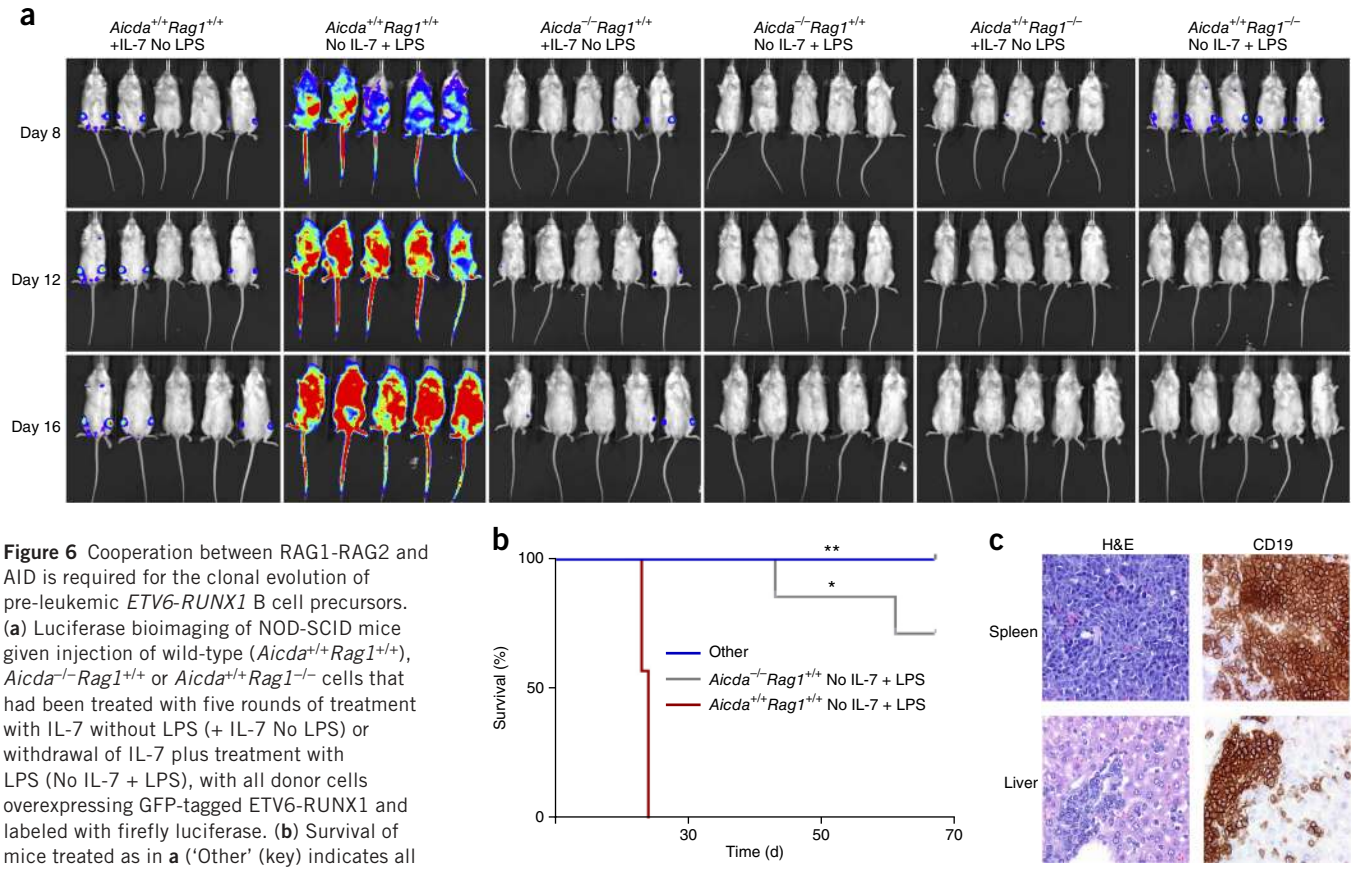

Figure 6 Cooperation between RAG1-RAG2 and AID is required for the clonal evolution of pre-leukemic *ETV6-RUNX1* B cell precursors. (**a**) Luciferase bioimaging of NOD-SCID mice given injection of wild-type (*Aicda*$^{+/+}$*Rag1*$^{+/+}$), *Aicda*$^{-/-}$*Rag1*$^{+/+}$ or *Aicda*$^{+/+}$*Rag1*$^{-/-}$ cells that had been treated with five rounds of treatment with IL-7 without LPS (+ IL-7 No LPS) or withdrawal of IL-7 plus treatment with LPS (No IL-7 + LPS), with all donor cells overexpressing GFP-tagged ETV6-RUNX1 and labeled with firefly luciferase. (**b**) Survival of mice treated as in **a** ('Other' (key) indicates all conditions not in red and gray lines) $*P = 0.00044$ and $**P = 0.00006$ (log-rank (Mantel-Cox) test). (**c**) Immunohistochemistry of sections of the spleen (top) and liver (bottom) from a host mouse given wild-type (*Aicda*$^{+/+}$ *Rag1*$^{+/+}$) cells treated by withdrawal of IL-7 in the presence of LPS (as in **a**), stained with hematoxylin and eosin (H&E) (left) or antibody to CD19 (right). Original magnification, ×40. Data are from one experiment with five mice per group (among seven total mice per group) (**a**), one experiment with seven mice in each group (**b**) or one experiment with one mouse representative of seven mice (**c**).

which is the reason that EBV$^+$ B cell lymphomas do not exhibit ongoing SHM and CSR. EBV-mediated suppression of AID is induced by the EBV oncoprotein EBNA2 (ref. 42).

To be able to study genetic lesions at the level of individual clones, we plated single transduced human B cell clones into individual wells of 96-well plates. We harvested cord blood B cell clones and subjected them to whole-exome sequencing. We analyzed EBV-immortalized cord blood B cells carrying three empty-vector controls (iRFP670, eGFP and dsRedE2); vector encoding AID and two empty vector controls; vectors encoding RAG1 and RAG2 and one empty vector control; or vectors encoding AID, RAG1 and RAG2. This analysis focused on the following structural genetic lesions: insertions, deletions, inversions and translocations. Cells with overexpression of either AID alone or RAG1-RAG2 had only a greater frequency of interchromosomal translocations than did cord blood B cells transduced with empty-vector controls (**Fig. 5c**). However, cord blood B cells expressing AID in combination with RAG1-RAG2 had a greater absolute number of insertions, deletions, inversions and intrachromosomal translocations than did cord blood B cells transduced with empty-vector controls (**Fig. 5c**).

RAG1-RAG2 and AID act together in ETV6-RUNX1 leukemia

We hypothesized that AID and RAG1-RAG2 would act together and induce secondary genetic lesions to accelerate the leukemogenesis of an *ETV6-RUNX1* small pre-BII clone in the context of inflammation. To test this, we transduced pre-B cells from wild-type (*Aicda*$^{+/+}$*Rag1*$^{+/+}$), *Aicda*$^{-/-}$*Rag1*$^{+/+}$ and *Aicda*$^{+/+}$*Rag1*$^{-/-}$ mice with *ETV6-RUNX1* and subjected them to five rounds of withdrawal of IL-7 and treatment with LPS. During each cycle, a large fraction of stimulated pre-B cells died (data not shown). Before entering the next round of stimulation, the pre-B cells among a small fraction of surviving pre-B cells recovered over the course of ~2 weeks (data not shown). The cell death after each cycle occurred independently of the *Aicda* or *Rag1* genotype (data not shown). This was not unexpected, because withdrawal of IL-7 induces many AID- and RAG1-RAG2–independent signaling pathways that promote apoptosis at the small pre-BII cell stage of B lymphopoiesis.

After five cycles of treatment as described above, we labeled the treated pre-B cells, as well as their untreated counterparts that did not undergo IL-7 withdrawal or treatment with LPS, with firefly luciferase and injected the cells intravenously into mice of the non-obese diabetic (NOD) severe combined immunodeficiency (SCID) strain (**Fig. 6a**). All mice that received wild-type (*Aicda*$^{+/+}$*Rag1*$^{+/+}$) pre-B cells died of leukemia within 3 weeks (**Fig. 6b**). We killed three to four mice in that same group that received wild-type (*Aicda*$^{+/+}$*Rag1*$^{+/+}$) pre-B cells at the same time because they were terminally ill with leukemia. This phenomenon was reflected in the rapid decrease in the Kaplan-Meier survival curve for this group (**Fig. 6b**). In contrast, the absence of *Aicda* or *Rag1* delayed or abrogated, respectively, the development of leukemia (**Fig. 6a**,**b** and **Supplementary Figs. 8** and **9**), which demonstrated that both AID and RAG1 were required for the clonal evolution of *ETV6-RUNX1* pre-B cells toward leukemia.

In the absence of *Aicda*, two mice developed leukemia, albeit after a substantially prolonged latency period (**Fig. 6b**). We were able to verify that the sick mice given injection of wild-type (*Aicda*$^{+/+}$*Rag1*$^{+/+}$) pre-B cells that had been stimulated by withdrawal of IL-7 plus treatment with LPS indeed died of pre-B ALL (**Fig. 6c** and **Supplementary Figs. 8** and **9**). These findings provided genetic evidence that the clonal evolution of pre-leukemic *ETV6-RUNX1* pre-B cells in the context of inflammatory and/or repetitive infectious stimulation required the activity of both AID and the RAG recombinase.

DISCUSSION

A published study has demonstrated that aberrant RAG-mediated V(D)J recombination targeting genes not encoding immunoglobulins causes genetic lesions that drive the clonal evolution of pre-B ALL[23]. Likewise, AID targets such genes, promotes drug resistance in B lineage ALL[43,44] and gives rise to AID-dependent chromosomal rearrangements. A comprehensive sequence analysis of 1,700 breakpoints of chromosomal rearrangements in human B cell malignancies has suggested that these lesions are caused by cooperation between the RAG recombinase and AID[19]. That conclusion was based on the unexpected finding that ~70% of chromosomal rearrangements in pro- and pre-B cells involve a break at CpG dinucleotides. Cytosine residues in CpG dinucleotides are frequently methylated. While AID-mediated deamination of cytosine typically results in conversion to uracil and triggers highly efficient base-excision repair, AID-mediated deamination of methylated cytosine (for example, in CpG) results in the conversion of cytosine to thymine. The resulting T:G mismatch is more stable than a U:G mismatch, and repair of T:G mismatches is ~2,500 times less efficient than that of U:G mismatches[45]. These T:G mismatches generate a one-base 'bubble' that is recognized by the RAG recombinase, which results in the introduction of a single-strand nick[19]. AID thus generates multiple one-base 'bubbles' (T:G mismatches) at methylated CpG sequences that form substrates for RAG-mediated cleavage. An important premise for this scenario is that under some circumstances, B cells express RAG proteins and AID concomitantly.

Our findings have highlighted the observation that small pre-BII cells represented a subset of increased genetic vulnerability because they concurrently expressed AID and RAG1-RAG2 in the presence of strong inflammatory stimuli. The concurrent activation of these enzymes would provide an explanation for the thus-far-elusive origin of childhood leukemia. We have formally demonstrated that combined activities of AID and the RAG recombinase were required for clonal evolution from pre-leukemic lesions and that such cooperative activities potentiated genetic instability in these clones. We elucidated how IL-7R signaling served as a safeguard protecting the early B cell genome from premature expression of AID. Altered cytokine environments have been previously attributed to the clonal evolution of ETV6-RUNX1–driven pre-B ALLs[12]. Here we have provided one more example of this phenomenon by which cytokine signaling in pre-B cells can contribute to genomic stability and, hence, leukemic transformation.

Our work provides a mechanistic underpinning for the epidemiological findings that altered patterns of infection during early childhood provide a promotional drive to leukemogenesis, especially for subtypes of childhood leukemia initiated *in utero* via *ETV6-RUNX1* fusion or hyperdiploidy[6]. In the future, it may be useful to investigate whether leukemic children with known infectious origins have high expression of AID and RAG enzymes. It might also be important to identify the classes of pathogens responsible for overt leukemogenesis in such children. Delineation of the molecular bases of the infectious and inflammatory origins of pre-B ALL could inform the development of public health policies on the benefits of vaccination programs during early childhood for the prevention of leukemia.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank J.L. Wiemels and C. Gawad for encouragement and critical discussions; D.B. Kohn (University of California, Los Angeles) for the envelope and packaging vectors for lentivirus and retrovirus productions; and H. Hanenberg (Indiana University) for the plasmid pCL6-IRES-eGFP-wo. Supported by the National Cancer Institute of the US National Institutes of Health (R01CA137060, R01CA139032, R01CA157644, R01CA169458 and R01CA172558 to M.M.), the Leukemia and Lymphoma Society (1479-11, 6132-09, 6097-10 and 6221-12 to M.M.), the William Lawrence and Blanche Hughes Foundation, the California Institute for Regenerative Medicine (TR2-01816 to M.M.), Leukaemia, Lymphoma Research UK (M.F.G. and A.F.), Cancer Research UK (CRUK 18131 to M.M.) and The Wellcome Trust (WT105104/Z/14/Z to M.F.G. and WT101880AIA to M.M.).

AUTHOR CONTRIBUTIONS

M.M. conceived of the study; S.S., L.K. and M.M. designed experiments and interpreted the data; S.S. and L.K. performed most of the experiments; E. Park, A.F., S.-M.K., D.T., B.H., N.H. and J.M. performed experiments; H.G. generated survival analyses for samples from patients; E. Papaemmanuil, A.F., K.S., S.C.K., R.C., D.G.S., M.R.L. and M.F.G. provided reagents, mouse samples and patient data; and S.S., L.K., M.F.G. and M.M. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Wiemels, J.L. *et al.* Prenatal origin of acute lymphoblastic leukaemia in children. *Lancet* **354**, 1499–1503 (1999).
2. Greaves, M.F. & Wiemels, J. Origins of chromosome translocations in childhood leukaemia. *Nat. Rev. Cancer* **3**, 639–649 (2003).
3. Bateman, C.M. *et al.* Acquisition of genome-wide copy number alterations in monozygotic twins with acute lymphoblastic leukemia. *Blood* **115**, 3553–3558 (2010).
4. Greaves, M. & Maley, C.C. Clonal evolution in cancer. *Nature* **481**, 306–313 (2012).
5. Gilham, C. *et al.* Day care in infancy and risk of childhood acute lymphoblastic leukaemia: findings from UK case-control study. *Br. Med. J.* **330**, 1294 (2005).
6. Greaves, M. Infection, immune responses and the aetiology of childhood leukaemia. *Nat. Rev. Cancer* **6**, 193–203 (2006).
7. Greaves, M. in *The Hygiene Hypothesis and Darwinian Medicine* (ed. Rook, G. A. W.) 239–255 (Birkhäuser Basel, 2009).
8. Urayama, K.Y., Buffler, P.A., Gallagher, E.R., Ayoob, J.M. & Ma, X. A meta-analysis of the association between day-care attendance and childhood acute lymphoblastic leukaemia. *Int. J. Epidemiol.* **39**, 718–732 (2010).
9. Auvinen, A., Hakulinen, T. & Groves, F. Haemophilus influenzae type B vaccination and risk of childhood leukaemia in a vaccine trial in Finland. *Br. J. Cancer* **83**, 956–958 (2000).
10. Groves, F.D., Sinha, D., Kayhty, H., Goedert, J.J. & Levine, P.H. Haemophilus influenzae type b serology in childhood leukaemia: a case-control study. *Br. J. Cancer* **85**, 337–340 (2001).
11. Ma, X. *et al.* Vaccination history and risk of childhood leukaemia. *Int. J. Epidemiol.* **34**, 1100–1109 (2005).
12. Ford, A.M. *et al.* The TEL-AML1 leukemia fusion gene dysregulates the TGF-beta pathway in early B lineage progenitor cells. *J. Clin. Invest.* **119**, 826–836 (2009).
13. Ford, A.M. *et al.* Fetal origins of the TEL-AML1 fusion gene in identical twins with leukemia. *Proc. Natl. Acad. Sci. USA* **95**, 4584–4588 (1998).
14. Mori, H. *et al.* Chromosome translocations and covert leukemic clones are generated during normal fetal development. *Proc. Natl. Acad. Sci. USA* **99**, 8242–8247 (2002).
15. Oettinger, M.A., Schatz, D.G., Gorka, C. & Baltimore, D. RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. *Science* **248**, 1517–1523 (1990).
16. Muramatsu, M. *et al.* Class switch recombination and hypermutation require activation-induced cytidine deaminase (AICDA), a potential RNA editing enzyme. *Cell* **102**, 553–563 (2000).

17. Kumar, S. *et al.* Flexible ordering of antibody class switch and V(D)J joining during B-cell ontogeny. *Genes Dev.* **27**, 2439–2444 (2013).
18. Yu, W. *et al.* Continued RAG expression in late stages of B cell development and no apparent re-induction after immunization. *Nature* **400**, 682–687 (1999).
19. Tsai, A.G. *et al.* Human chromosomal translocations at CpG sites and a theoretical basis for their lineage and stage specificity. *Cell* **135**, 1130–1142 (2008).
20. Hardy, R.R. & Hayakawa, K. B cell development pathways. *Annu. Rev. Immunol.* **19**, 595–621 (2001).
21. Rajewsky, K. Clonal selection and learning in the antibody system. *Nature* **381**, 751–758 (1996).
22. Mullighan, C.G. *et al.* BCR-ABL1 lymphoblastic leukaemia is characterized by the deletion of Ikaros. *Nature* **453**, 110–114 (2008).
23. Papaemmanuil, E. *et al.* RAG-mediated recombination is the predominant driver of oncogenic rearrangement in ETV6-RUNX1 acute lymphoblastic leukemia. *Nat. Genet.* **46**, 116–125 (2014).
24. Yamane, A. *et al.* Deep-sequencing identification of the genomic targets of the cytidine deaminase AICDA and its cofactor RPA in B lymphocytes. *Nat. Immunol.* **12**, 62–69 (2011).
25. Wiemels, J.L. *et al.* Site-specific translocation and evidence of postnatal origin of the t(1;19) E2A–PBX1 fusion in childhood acute lymphoblastic leukemia. *Proc. Natl. Acad. Sci. USA* **99**, 15101–15106 (2002).
26. Qian, J. *et al.* B cell super-enhancers and regulatory clusters recruit AID tumorigenic activity. *Cell* **159**, 1524–1537 (2014).
27. Alpar, D. *et al.* Clonal origins of ETV6-RUNX1+ acute lymphoblastic leukemia: studies in monozygotic twins. *Leukemia* **29**, 839–846 (2015).
28. Fu, C., Turck, C.W., Kurosaki, T. & Chan, A.C. BLNK: a central linker protein in B cell activation. *Immunity* **9**, 93–103 (1998).
29. Geier, J.K. & Schlissel, M.S. Pre-BCR signals and the control of Ig gene rearrangements. *Semin. Immunol.* **18**, 31–39 (2006).
30. Johnson, K. *et al.* Regulation of immunoglobulin light-chain recombination by the transcription factor IRF-4 and the attenuation of interleukin-7 signaling. *Immunity* **28**, 335–345 (2008).
31. Puel, A., Ziegler, S.F., Buckley, R.H. & Leonard, W.J. Defective IL7R expression in $T^-B^+NK^+$ severe combined immunodeficiency. *Nat. Genet.* **20**, 394–397 (1998).
32. Mandal, M. *et al.* Epigenetic repression of the *Igk* locus by STAT5-mediated recruitment of the histone methyltransferase Ezh2. *Nat. Immunol.* **12**, 1212–1220 (2011).
33. Duy, C. *et al.* BCL6 is critical for the development of a diverse primary B cell repertoire. *J. Exp. Med.* **207**, 1209–1221 (2010).
34. Amin, R.H. & Schlissel, M.S. Foxo1 directly regulates the transcription of recombination-activating genes during B cell development. *Nat. Immunol.* **9**, 613–622 (2008).
35. Herzog, S. *et al.* SLP-65 regulates immunoglobulin light chain gene recombination through the PI(3)K-PKB-Foxo pathway. *Nat. Immunol.* **9**, 623–631 (2008).
36. Muramatsu, M. *et al.* Specific expression of activation-induced cytidine deaminase (AICDA), a novel member of the RNA-editing deaminase family in germinal center B cells. *J. Biol. Chem.* **274**, 18470–18476 (1999).
37. Crouch, E.E. *et al.* Regulation of AICDA expression in the immune response. *J. Exp. Med.* **204**, 1145–1156 (2007).
38. Zhang, Z. *et al.* Contribution of Vh gene replacement to the primary B cell repertoire. *Immunity* **19**, 21–31 (2003).
39. Gawad, C., Koh, W. & Quake, S.R. Dissecting the clonal origins of childhood acute lymphoblastic leukemia by single-cell genomics. *Proc. Natl. Acad. Sci. USA* **111**, 17947–17952 (2014).
40. Dörner, T. *et al.* Analysis of the frequency and pattern of somatic mutations within nonproductively rearranged human variable heavy chain genes. *J. Immunol.* **158**, 2779–2789 (1997).
41. Kurth, J., Hansmann, M.-L., Rajewsky, K. & Küppers, R. Epstein-Barr virus-infected B cells expanding in germinal centers of infectious mononucleosis patients do not participate in the germinal center reaction. *Proc. Natl. Acad. Sci. USA* **100**, 4730–4735 (2003).
42. Tobollik, S. *et al.* Epstein-Barr virus nuclear antigen 2 inhibits AID expression during EBV-driven B-cell growth. *Blood* **108**, 3859–3864 (2006).
43. Feldhahn, N. *et al.* Activation-induced cytidine deaminase acts as a mutator in BCR-ABL1-transformed acute lymphoblastic leukemia cells. *J. Exp. Med.* **204**, 1157–1166 (2007).
44. Gruber, T.A., Chang, M.S., Sposto, R. & Müschen, M. Activation-induced cytidine deaminase accelerates clonal evolution in BCR-ABL1-driven B-cell lineage acute lymphoblastic leukemia. *Cancer Res.* **70**, 7411–7420 (2010).
45. Schmutte, C., Yang, A.S., Beart, R.W. & Jones, P.A. Base excision repair of U:G mismatches at a mutational hotspot in the p53 gene is more efficient than base excision repair of T:G mismatches in extracts of human colon tumors. *Cancer Res.* **55**, 3742–3746 (1995).

ONLINE METHODS

Mouse strains and cell culture. Bone marrow cells were obtained from young female and male mice (less than 6 weeks of age). The following strains of mice were used: *Blnk*$^{-/-}$ mice[46], mice with *lox*P-flanked *Pten* alleles (*Pten*$^{fl/fl}$ mice)[47], *Aicda-EGFP* mice[37], *Aicda*-Cre × *Rosa26*-EYFP mice[37], mice with *lox*P-flanked *Stat5a* and *Stat5b* alleles (*Stat5*$^{fl/fl}$ mice)[48], *Rag1*$^{-/-}$ mice[49], *Aicda*$^{-/-}$ mice[16] and NOD-SCID mice (Jackson Laboratories). Bone marrow cells were isolated by flushing of the cavities of the femur and tibia with PBS. After filtration through a 40-µm filter and depletion of erythrocytes with RBC lysis buffer (BD PharmLyse; BD Biosciences), washed cells were either cryopreserved or used for further experiments. Bone marrow cells were cultured with 10 ng/ml IL-7 on non–tissue culture dishes coated with RetroNectin (Takara) or on irradiated OP9 stroma layers. Precursor B cell populations were sorted from mouse bone marrow as pro-B cells (c-Kit^{+}B220^{+}), large pre-BII cells (CD25^{+}B220^{+}FSC^{hi}SSC^{lo}) and small pre-BII cells (CD25^{+}B220^{+}FSC^{lo}SSC^{lo}) and were separated by flow cytometry with antibodies (**Supplementary Table 8**) on a FACSAriaII (BD). All pre-B cells were maintained at 37 °C in a humidified incubator with 5% CO_2 in Iscove's modified Dulbecco's medium (Invitrogen) with GlutaMAX containing 20% FBS, 100 IU/ml penicillin, 100 µg/ml streptomycin, 50 µM 2-mercaptoethanol and 10 ng/ml recombinant mouse IL-7 (Peprotech). All mouse experiments were approved by the Institutional Animal Care and Use Committee of the University of California, San Francisco.

Patient samples, cord blood and human cell lines. We obtained diagnostic bone marrow samples of subjects with primary pre-B ALL[23,43,50,51] from our collaborators, and human pre-B ALL cell lines from Deutsche Sammlung von Mikroorganismen und Zellkulturen, Germany (**Supplementary Tables 6** and **7**). These primary pre-B ALL samples, along with some pre-B ALL patient xenografts, were used for *IGH* sequencing studies. Our primary patient xenograft repository is approved by the University of California, San Francisco, and the Institutional Review Board of the Cancer Therapy Evaluation Program of the National Cancer Institute and ECOG (protocol E2993T5). Human CD19^{+} cord blood B cells were obtained from AllCells. AllCells procures cord blood subsequent to receiving informed consent from its donors. We have approval to use cord blood cells in our experiments (protocol BU085087; University of California, San Francisco).

***In vivo* model for inflammatory origins of *ETV6-RUNX1*–driven leukemia.** IL-7-dependent pre-B cells from wild-type, *Aicda*$^{-/-}$ *Rag1*$^{+/+}$, *Aicda*$^{+/+}$ *Rag1*$^{-/-}$ mice were transformed with the retroviral vector ETV6-RUNX1–internal ribosomal entry site (IRES)-GFP. Next these cells were either cultured with 10 ng/ml IL-7 on OP9 plates or were subjected to five cycles of withdrawal of IL-7 and treatment with 2.5 µg/ml LPS. For the latter, each cycle was interrupted by a phase in which cells were given 10 ng/ml IL-7 for 1–2 d to aid their recovery before the subsequent cycle was started. We then transduced cells with a retroviral vector encoding luciferase enzyme and injected cells intravenously into sublethally irradiated (250 cGy) female NOD-SCID mice. Recipient mice were monitored for weight loss, hunched posture and inability to move, as indicators of leukemia progression. Recipient mice were imaged for bioluminescence on days 7, 10 and 12 after injection. Mice were killed after they became terminally ill. For the group that displayed signs of leukemia; immunostaining for CD19 was performed on the bone marrow and spleen of mouse recipients for confirmation of leukemic disease. We killed mice from all the healthy groups at day 70 after injection.

Immunoblot analysis. Cells were lysed in CelLytic buffer (Sigma) supplemented with 1% protease inhibitor 'cocktail' (Pierce). Protein samples were subsequently separated by electrophoresis through NuPAGE (Invitrogen) 4–12% Bis-Tris gradient gels and were transferred to PVDF membranes (Immobilion; Millipore). For the detection of mouse and human proteins by immunoblot analysis, we used primary antibodies (all antibodies used for immunoblotting and their corresponding dilutions, **Supplementary Table 8**.) together with the WesternBreeze immunodetection system (Invitrogen).

Flow cytometry. All antibodies for flow cytometry measurements, as well as their respective isotype-matched control antibodies are in **Supplementary Table 8**. Surface-staining antibodies that were phycoerythrin conjugated were used at a dilution of 1:10, and those conjugated to fluorescein isothiocyanate were used at a dilution of 1:5.

Quantitative RT-PCR. Quantitative real-time PCR carried out with the SYBRGreenER mix from Invitrogen according to standard PCR conditions and an ABI7900HT real-time PCR system (Applied Biosystems). Primers for quantitative RT-PCR are in **Supplementary Table 9**.

Sequencing for somatic hypermutation of the VH region. Genomic DNA was isolated from the patient samples described in **Supplementary Tables 6** and **7**. V_H-DJ_H segments were amplified with a mixture of V_H forward primers (V_H mix) and J_H reverse primers (J_H mix) (**Supplementary Table 9**). PCR was carried out with a Phusion Hot Start Kit (NEB). PCR fragments were cloned with a Topo TA cloning kit (Invitrogen) and were subsequently sequenced unidirectionally in a 96-well format at Eurofins MWG Operon. Comprehensive sequence analysis was used for the identification of somatic mutations in the V_H-DJ_H junction sequences with IMGT-V Quest software (The International Immunogenetics Information System).

Retrovirus production and transduction. Transfection of retroviral constructs and their corresponding empty vector controls[52,53] (**Supplementary Table 10a**) was performed using Lipofectamine 2000 (Invitrogen) with Opti-MEM media (Invitrogen). Retroviral supernatants were used for infection of cells by co-transfection of HEK 293FT cells with the plasmids pHIT60 (gag-pol) and pHIT123 (ecotropic envelope) (provided by D.B. Kohn). To infect human cells, we replaced pHIT123 with pHIT456 (amphotropic envelope) (provided by D.B. Kohn). Cells were cultured in high-glucose DMEM (Invitrogen) with GlutaMAX containing 10% FBS, 100 IU/ml penicillin, 100 µg/ml streptomycin, 25 mM HEPES, pH 7.2, 1 mM sodium pyruvate and 0.1 mM nonessential amino acids. Serum-free medium was replaced after 16 h with growth medium containing 10 mM sodium butyrate. After 8 h of incubation, the medium was changed back to growth medium without sodium butyrate. 24 h later, we harvested the viral supernatants, passed them through a 0.45-µm filter, and centrifuged them twice at 2,000*g* for 90 min at 32 °C on 50 µg/ml RetroNectin-coated non–tissue-culture-treated six-well plates. Pre-B cells (2×10^6 to 3×10^6 cells per well) were transduced by centrifugation at 600*g* for 30 min and were maintained overnight at 37 °C with 5% CO_2 before transfer into culture flasks.

Lentivirus production and transduction. Transfection of lentivirus (**Supplementary Table 10b**) was performed with Lipofectamine 2000 (Invitrogen) in Opti-MEM medium (Invitrogen). We produced lentiviral supernatants to infect mouse cells by cotransfecting HEK 293FT cells with the plasmids pCD/NL-BH*DDD[54] (gag-pol; 17531; Addgene) and pMD2.G (VSV-G envelope; 12259; Addgene). Cells were cultivated in high-glucose DMEM (Invitrogen) with GlutaMAX as described above. Culture supernatants were concentrated with Lenti-X according to the manufacturer's recommendation (Clontech). The concentrate of three viruses was incubated for 30 min on 50 µg/ml RetroNectin-coated non-tissue-culture-treated six-well plates. EBV-transformed cells were transduced (2×10^6 cells per well) by centrifugation at 600*g* for 30 min and were maintained overnight at 37 °C with 5% CO_2 before transfer into culture flasks.

Immunohistochemistry. Spleen and liver obtained from experimental mice were mounted in OCT embedding compound and were frozen at −80 °C. Sections 4 µm in thickness were cut with a cryostat and were stained with antibody to CD19 (ab25232; Abcam). Antigen-antibody interactions were detected and visualized with a secondary antibody to rat IgG (P0450; Dako). The slides were counterstained with hematoxylin and eosin and then were dehydrated in graded series of ethanol solutons. Immunohistochemistry sections were mounted with antifade reagent (Pro-Long Gold; Invitrogen) and coverslips were sealed before acquisition of images on a Zeiss Axiovert 200M inverted confocal microscope with a 40× objective. Immunohistochemistry was performed by the Cellular Imaging Core at Children's Hospital Los Angeles.

Cloning of lentiviral vectors encoding AID and RAG1-RAG2. The plasmids pCL6-IRES-dsRedExpress2-wo and pCL6-IRES-iRFP670-wo were generated

by replacement of the eGFP-encoding region of pCL6-IRES-eGFP-wo (obtained from H. Hanenberg) with sequence encoding dsRedExpress2 from pDsRed-Express2 (Clontech) or with sequence encoding iRFP670 from pNLS-iRFP670 (45466; Addgene). *Rag1* and *Rag2* were amplified with Phusion Hot Start II (Thermo Fisher Scientific) from C57BL/6J cDNA and were inserted via In-Fusion HD cloning (Clontech) into the vector pCL6-IRES-eGFP-wo or pCL6IEdsRedExpress2-wo, respectively. *Aicda* was amplified with Phusion Hot Start II (Thermo Fisher Scientific) from the plasmid MIG-AICDA and was inserted into the vector pCL6-IRES-iRFP670-wo.

EBV transformation of cord blood CD19$^+$ B cells. 2×10^6 fresh CD19$^+$ B cells obtained from human cord blood were infected with 1 ml supernatant from B95-8 cells containing EBV (VR-1492; American Type Culture Collection) at a dilution of 1:4 with RPMI medium (Invitrogen) containing GlutaMAX with 20% FBS, 100 IU/ml penicillin and 100 µg/ml streptomycin.

Introduction of *Aicda* and *Rag1-Rag2* into human CD19$^+$ B cells. Lentiviral vectors containing mouse *Aicda* (pCL6-Aicda-IRES-iRFP670-wo), *Rag1* (pCL6-Rag1-IRES-eGFP-wo), *Rag2* (pCL6-Rag2-IRES-dsRedExpress2-wo) and the corresponding empty vector controls were introduced into EBV-transformed human CD19$^+$ cord blood B cells by the transduction protocol described above. Cells were transduced with empty vector or vector containing *Aicda* alone, the *Rag1-Rag2* combination or both *Aicda* and *Rag1-Rag2*. After 4 d, live EBV-transformed cord blood B cells, stained with DAPI (4,6-diamidino-2-phenylindole), that were triply positive for eGFP, iRFP670 and dsRedExpress2 were sorted as single cells into 96-well plates with a configuration of 488 nm (525/50), 640 nm (670/30), 561 nm (582/15) and 355 nm (450/50) on an BD FACSAriaII Sorter. The DNA of the sorted triple-positive single cells was amplified with a REPLI-g Single Cell Kit according to the manufacturer's protocol (Qiagen). The resulting DNA underwent enrichment for targets by SureSelect Human All Exon V5 plus UTR target enrichment (Agilent), then paired-end libraries were prepared and sequenced by Illumina with two 100-nucleotide paired-end primers (MOgene). The resulting files were analyzed with GeneSpring software (Agilent). The 'mate' status was fixed to exclude reads with an inconsistent 'mate' status. A minimum insertion-deletion detection size of 100 bp was applied with the PEMer Indel detection algorithm. The minimum local deviant read coverage used was 11 and the minimum local deviant read fraction was 0.2.

Statistical analysis. The comparisons for the mean values between two sample groups were made by two-tailed Student's *t*-test or Wilcoxon's rank-sum test with Graphpad Prism software or R software (R Development Core Team 2009). For experiments involving transplantation of cells into mice, the minimal number of mice in each group was calculated through use of the 'cpower' function in the R/Hmisc package. The Kaplan-Meier method was used for estimation of overall survival and relapse-free survival. The log-rank test was used for comparison of the difference between patient groups in their survival. The R package 'survival' version 2.35-8 was used for the survival analysis. The level of significance was set at $P < 0.05$.

46. Jumaa, H. *et al.* Abnormal development and function of B lymphocytes in mice deficient for the signaling adaptor protein SLP-65. *Immunity* **11**, 547–554 (1999).
47. Lesche, R. *et al.* Cre/loxP-mediated inactivation of the murine Pten tumor suppressor gene. *Genesis* **32**, 148–149 (2002).
48. Liu, X. *et al.* Stat5a is mandatory for adult mammary gland development and lactogenesis. *Genes Dev.* **11**, 179–186 (1997).
49. Mombaerts, P. *et al.* RAG-1-deficient mice have no mature B and T lymphocytes. *Cell* **68**, 869–877 (1992).
50. Rosenfeld, C. *et al.* Phenotypic characterisation of a unique non-T, non-B acute lymphoblastic leukaemia cell line. *Nature* **267**, 841–843 (1977).
51. Height, S.E. *et al.* Analysis of clonal rearrangements of the Ig heavy chain locus in acute leukemia. *Blood* **87**, 5242–5250 (1996).
52. Kumar, M.S. *et al.* Dicer1 functions as a haploinsufficient tumor suppressor. *Genes Dev.* **23**, 2700–2704 (2009).
53. Yusuf, I., Zhu, X., Kharas, M.G., Chen, J. & Fruman, D.A. Optimal B-cell proliferation requires phosphoinositide 3-kinase-dependent inactivation of FOXO transcription factors. *Blood* **104**, 784–787 (2004).
54. Zhang, X.-Y., La Russa, V.F. & Reiser, J. Transduction of bone-marrow-derived mesenchymal stem cells by using lentivirus vectors pseudotyped with modified RD114 envelope glycoproteins. *J. Virol.* **78**, 1219–1229 (2004).

doi:10.1038/ni.3160