Infectious triggers and novel therapeutic opportunities in childhood B cell leukemia

Cesar Cobaleda[1#], Carolina Vicente-Dueñas[2], and Isidro Sanchez-Garcia[2,3#]

[1]*Immune system development and function Unit, Centro de Biología Molecular Severo Ochoa (CSIC -Universidad Autónoma de Madrid), Madrid, Spain*

[2]*Institute for Biomedical Research of Salamanca (IBSAL), Salamanca, Spain*

[3]*Experimental Therapeutics and Translational Oncology Program, Instituto de Biología Molecular y Celular del Cáncer, CSIC/Universidad de Salamanca, Salamanca, Spain.*

(#) corresponding authors
Email to: cesar.cobaleda@csic.es or isg@usal.es

Abstract

B cell acute lymphoblastic leukemia (B-ALL) is the most common form of childhood cancer. Although treatment has advanced remarkably in the last 50 years, it still fails in ~20% of patients. Recent studies revealed that a large number (>5%) of healthy newborns carry pre-leukemic clones that originate *in utero*, but only a small percentage of these carriers will progress to overt B-ALL. The drivers of progression are unclear, but B-ALL incidence seems to be increasing in parallel with the adoption of modern lifestyles. Emerging evidence shows that a major driver for the conversion from the pre-leukemic to the B-ALL state is exposure to immune stressors, like infection. Here, we discuss our current understanding of the environmental triggers and genetic predispositions that may lead to B-ALL, highlighting lessons from epidemiology, the clinic, and from animal models, and identifying priority areas for future research.

Introduction [H1]

Leukemia is the most frequent type of cancer in children and accounts for up to 35% of all malignancies under 15 years of age[1]. Although its incidence is reduced in comparison with adult tumors, in countries where infectious diseases are not endemic, cancer is the leading cause of pathology-related childhood deaths, and acute lymphoblastic leukemia (ALL) is the most frequent cause of cancer-associated death before 20 years of age[2]. ALL is one of the four main categories of human leukemias, and most of them are of the B cell type (B-ALL). In this Review, we will focus on childhood B-ALL (also known as B cell precursor ALL, BCP-ALL), a clonal malignant disease characterized by the accumulation of blast cells that are phenotypically similar to the normal earliest stages of B-cell differentiation[3].

In spite of its molecular heterogeneity (Table 1), a large percentage of childhood B-ALLs share some common characteristics in their biology; first, childhood B-ALL has a distinctive age distribution characterized by a sharp peak between 2 and 5 years of age, before declining to a much more reduced rate[4,5]. Also, B-ALLs are in general highly responsive to chemotherapy, leading to large improvements (up to almost 90%) in the survival rate in affected children over the last 50 years[6], even though most current treatments are associated with substantial toxicity and morbidity[7]. But perhaps the most relevant biological characteristic of many types of childhood B-ALLs is the existence of a latent silent pre-leukemic phase in which the initiating leukemogenic hit is present but leukemia does not develop (see Box 1, Table 1 and Figure 1). This clinically silent pre-leukemic condition, present in up to 5% of the healthy childhood population, will however not evolve to disease in the majority of the cases[8,9]. However, most likely due to the presence of a given external/environmental stimulus, a small percentage (<1%) of those predisposed children will develop B-ALL (see Box 1 and Figure 1). The incidence of B-ALL has increased in the past few decades, a rise that seems to be associated with modern life-style[10-12]. Since it is believed that the incidence of pre-leukemic clones is largely constant, this increase is considered to be

related to new exposures or circumstances affecting the children in westernized countries[10-12]. Although infections have been proposed to be one of the possible causes of childhood leukemias for more than a century[13], there was no relevant biological evidence supporting this possibility. This was mainly due to the facts that i) many of these proofs cannot be obtained from studies in human patients, because of the low incidence of the disease and the genetic heterogeneity of human populations and ii) that the empirical requirements necessary to obtain unequivocal answers to the postulated hypotheses can only be applied in experimental animal models. Now, several recent studies have provided strong evidence confirming the hypothesis that exposure to infections can indeed be a driver of clonal evolution of preleukemic clones towards overt leukemia[14-18].

The trend of increasing childhood leukemia and cancer incidence associated with modern lifestyles[12,19] underscores the urgency in addressing cancer disease prevention. Deciphering the properties of the complex interplay between genetic leukemia predisposition, environmental triggers and immune evasion will result in a better understanding of the host mechanisms of defence against progression to leukemia, and of the pathogenesis of childhood B-ALL. The conceptual and mechanistic advances in this field will open new possibilities for clinical applications as well as for disease prevention and treatment. In this review, we discuss the latest discoveries in the field of childhood B-ALL and highlight possible directions for future research.

[H1] Genetic defects and predisposition to B-ALL

In the past decades, we have gained an enormous amount of knowledge about the molecular and cellular biology of tumors; unfortunately, with some exceptions like tyrosine kinase inhibitors-based therapies in Philadelphia[+] (Ph[+]) B-ALL[20,21], this knowledge has not correlated with our capacity to successfully develop personalized therapies[22,23]. From a developmental perspective, the key events in cancer

development are its origin (the first moment in which a fully normal cell evolves towards cancer) and the posterior evolutionary steps that lead to the acquisition of additional malignant capacities. However, we still don't know how to stop a pre-cancerous cell from becoming a cancerous one, mainly because we don't understand the nature of the early events that determine this conversion. In childhood B-ALL, first, a predisposing mutation (either germline or somatic, see below) leads to the generation of a preleukemic clone in which normal B cell development is perturbed, but not enough to give rise to a malignant clone by itself (Box 1 and Figure 1). Therefore, from this perspective, childhood B-ALL is an ideal model of study to gain definitive insights into how to prevent this transition to a full-blown malignant cell.

There are different categories of genetic events linked to B-ALL (Table 1), originated by various types of molecular aberrations, which can mediate B cell leukemogenesis by diverse mechanisms, and which are associated with different disease subtypes, specific age onsets and variable prognoses[24-26]. For example, high hyperdiploid or ETV6-RUNX1[+] childhood B-ALLs (described hereunder) peak at 2-5 years of age, whereas mixed lineage leukemia (MLL)-rearranged or NUT family member 1 (NUTM1)-rearranged infant ALLs have a very early onset, and ZNF384-based, double homeobox protein 4 (DUX4), Ph[+], or Ph-like leukemias have a later onset[27-31] (Table 1).

At the molecular level, a large subgroup of B-ALLs is originated by aneuploidies whose mechanism of action, due to their complexity, is still not fully understood but that, at least in the case of hyperdiploid B-ALLs, is normally associated with a defective condensin complex, altered Aurora B kinase activity, and an impaired spindle assembly checkpoint[32]. Another large and more homogenous group is triggered by alterations affecting the activity of signalling kinases, and this subtype is much more prevalent in adults and it is associated with a poor prognosis[31].

The third important subgroup in terms of molecular origin is composed by B-ALLs in which different genetic alterations lead to the dysregulation of transcription

factors involved in early hematopoietic and early B-cell development[24,26]; such dysregulation can be due to either somatic or germline alterations. In turn, the latter can be classified into 2 groups: low-penetrance susceptibility to B-ALL (a 1.5-fold to 2-fold increase in relative risk) conferred by common germline polymorphisms, and high-penetrance (10-fold increase in relative risk) predisposition conferred by more infrequent germline variants[33-35]. Genome wide association studies (GWASs) have identified common polymorphisms in several genomic loci associated with hematopoietic transcription factors (for example, *ARID5B*[36,37], *IKZF1*[36,37], *GATA3*[38,39], *CEBPE*[36], *ERG*[40], or *IKZF3*[41]), associated with ALL susceptibility. Individually, each one of these risk alleles only has a modest effect and is most likely of limited clinical significance; however, combined they can lead to even a nine-fold increase in leukemia risk[42]. On the other hand, there are infrequent germline variations that are associated with strong leukemia predisposition, often presenting in familial clusters. Thanks to the modern next-generation sequencing technologies, numerous genes involved in normal hematopoietic or lymphoid development have now been implicated in familial B-ALL predisposition, including *IKZF1*[43], *PAX5*[44-46], *ETV6*[47-51], *SH2B3*[52], *RUNX1*[53,54], or *TYK2*[55]. Each of these genes is now considered to predispose to B-ALL in ~1% of sporadic cases each[43,47]; therefore, their cumulative effect on predisposition can be higher than 1%, although the full extent of this effect is not clear. In any case, these mutations offer an invaluable glimpse into the intimate mechanisms of B-ALL origins.

Regarding somatic alterations affecting transcription factor genes, they can also be classified into two groups (Table 1): i) chromosomal rearrangements resulting either in chimeric transcription factors or in copy-number alterations, and ii) sequence mutations that directly alter the factor's activity[24]. More than half of the gene fusion events in B-ALL involve one or more transcription factor genes[56,57]; indeed, the most characteristic and most frequent somatic chromosomal translocations[3,6,58] that occur in childhood B-ALL are those leading to the generation of the chimeric fusion proteins ETV6-RUNX1 (which is found in 25% of childhood ALL cases), TCF3-PBX1 or TCF3-

HLF, whose contribution to the founding of the preleukemic state have been previously reviewed thoroughly[3,6,58]. Sequence mutations are the other type of somatic alteration affecting transcription factor activity in B-ALLs, as is the case with PAX5[56,59], EBF1[59], IKZF1[60], or BTG1[59]. Broadly speaking, gene fusions tend to act as initiating events during early leukemogenesis, while small genomic aberrations more often occur as late secondary events potentiating and promoting leukemogenic effects.

In general, it is accepted that all these genomic defects act by interfering with correct B cell development and therefore cause a block in differentiation[61,62]. The consequent accumulation of an immature B cell progenitor cell pool increases the chance of additional oncogenic mutations, with ensuing leukemic transformation. However, one must not forget that most transcription factors fulfill complex and varied functions (very often not limited to haematopoiesis) and, most likely, the haematopoietic consequences of their mutations are not only limited to a block in B cell differentiation.

Early B cell development and leukemia are both lineage-commitment processes in which developmental potential becomes restricted[24]. Normally, the first oncogenic hit restricts the target cells to a single cell lineage (B cells, in the context of B-ALL). However, a single hit is typically insufficient to trigger full B-ALL development (with some notable exceptions, for example *MLL* translocations in infant ALLs), as demonstrated by the existence of twins with concordant childhood ALL and identical preleukemic translocations in their blood cells[63] (see Box 1). Since the first hit seems to act by specifying the leukemic B cell lineage identity, what is then the role of the second hit (that is, *PAX5* or *IKZF1* deletions)? It has been shown that these genes are important DNA damage hotspots during leukemic transformation of B cell precursors[64]; therefore, the consequent downregulation of their activity would not have an instructive role in the process of B-ALL leukemogenesis, but rather a permissive one. Indeed, functional factors would prevent cells with the first oncogenic hit from being further reprogrammed into leukemic B cells. In this regard, it has been proposed that

functional B cell transcription factors act as metabolic gatekeepers, limiting the amount of cellular ATP to levels that are insufficient for the malignant transformation of precursor B cells[65,66], and therefore allowing preleukemic clones to remain in a latent state[65]. In p53-deficient cancers, the p53-mediated DNA damage response that restrains the reprogramming capacity of the first hit is lost[67], and it is conceivable that B cell transcription factors might have a similar role in limiting the malignant reprogramming function of the first hit. In this regard, *TP53* and *PAX5* alterations seem to be mutually exclusive in human B-ALL development[68]. It seems contradictory that the presence of PAX5 mediates B cell commitment in normal development but that its absence is required for full-blown B-ALL development. Nevertheless, although the data suggest that PAX5 downregulation restricts cell lineage options to a leukemic B cell lineage fate (lineage infidelity), such activity still needs to be clearly demonstrated.

[H1] Risk factors and preleukemic conversion

Although the genomic landscape of patients with B-ALL has been extensively characterized[35,56] (Table 1), the external factors that promote the conversion of the preleukemic clone into a leukemia are not yet understood. Many factors have been suggested to be the presumed cause of childhood leukemia[69-71]. However, up to now, the only risk factors clearly associated with childhood B-ALL are i) Down syndrome, ii) sex (boys are ~1.2 more affected than girls), iii) chemotherapeutic drugs and iv) severe exposure to ionizing radiations above 100 mSv, although some studies suggested that this increased risk may also be associated with lower doses of radiation[72-75]. Consequently, after reports of increased leukemia incidence in the vicinity of nuclear fuel reprocessing plants[76] or nuclear power plants[77], living near a nuclear facility became a suspected risk factor for childhood leukemia. However, leukemia clusters also have been reported elsewhere and are not specific to nuclear installations (see below), suggesting the existence of radiation-independent mechanisms. Exposure to low-frequency electromagnetic fields (ELF-EMF) has also been suggested by

epidemiological studies to be a risk factor for childhood leukemia[78], but so far there is no solid biological or mechanistic evidence supporting these findings[79,80]. Lastly, as mentioned above, there is also consistent evidence from epidemiological studies showing that high weight at birth is a determinant of disease risk[81-83]. In any case, it must be taken into account that epidemiological associations do not imply a direct causality.

Given the clinical characteristics of the disease, one of the aspects that have attracted more attention in relation to the risk of developing childhood B-ALL is the exposure to infectious agents and the role of immune function[84,85]. In all countries that keep accurate records, the age peak of childhood B-ALL is 2-5 years, corresponding with the interval when children are first exposed to more common infections in day care, kindergarten and school (reviewed in [86]). Accordingly, the idea that infection might be a trigger for childhood B-ALL was proposed more than a century ago[13]. This is epidemiologically supported by the existence of several reported B-ALL space-time clusters associated with different specific pathogens like group A Streptococcus[87], adenovirus[88], or influenza A H1N1 swine flu virus[89]. Furthermore, peaks of B-ALL that occur approximately 6 months after seasonal influenza epidemics have also been described[90]. These results also suggest that attributing a single, specific, infectious agent to the second step in leukemia development is unlikely. Interestingly, the incidence of B-ALL has increased in parallel with the increasing levels of cleanliness associated with modern lifestyles[10-12]. On the other hand, epidemiological studies have shown that exposure to infectious agents in early life can provide a protective effect against B-ALL development, including associations with day care attendance[91-98], birth order[91,98,99], mode of delivery[100-102], breastfeeding[103], and BCG vaccination[104,105] (which is known to confer improved immune responses against other non-mycobacterial pathogens[106]). Kinlen suggested that childhood B-ALL could arise as an infrequent consequence of exposure to an unidentified common infection, but this would only become noticeable at times of population mixing, when a relatively large number of

susceptible children face a large number of infected individuals, leading to a (mainly subclinical) widespread epidemic[107-109]. One example of such a situation would be the construction of a nuclear facility in a distant rural area, where a large influx of non-local workers suddenly moves into a previously isolated area, potentially introducing new pathogens to the resident population[88,108,110]. In parallel, Mel Greaves hypothesized that a lack of exposure to infectious agents in the first year of life, together with a posterior 'delayed' infectious challenge may lead to the development of B-ALL in the peak ages of 2 to 5 years[84,111]. Studies that have investigated self-reported day-care attendance as a surrogate marker for social contacts support this hypothesis[112], but other studies have observed that children developing B-ALL in the age range of 2–5 years had significantly higher reports of diagnosed infectious episodes in their first year of life when compared to controls[113]. This suggests that the immune stress in children who will end up developing B-ALL may be taking place several years before diagnosis[114-118], a possibility supported by the finding of abnormal profiles of inflammatory markers in neonatal blood spot samples of children who later developed B-ALL[119] and also by *in vitro* data suggesting that signalling by the pro-inflammatory cytokines IL-6, IL-1β and TNF[120] or TGFβ-dependent signalling[121] can predispose pre-leukemic B cells to malignant transformation. As can be seen from the evidence presented, the balance between the 'delayed infection' challenge and 'early exposure to pathogen' challenge is very hard to quantify. Epidemiologic data support a 'delayed infection' model in which common infections promote the posterior secondary genetic events, but only in the context of previous insufficient infectious exposure of the children (revised in [84]). This model implies that the absence of timely exposure to infections in postnatal life in the clean environments of modern societies might prompt the immune system towards aberrant responses following a posterior or 'delayed' exposure to common pathogens. However, as we have previously mentioned, epidemiologic studies have also found that infections in the first year of life were not protective[114,117]. Finally, recent epidemiologic studies have also shown that infections could be involved in early B-ALL

initiating events[115,116,122]. In summary, there is a large number of epidemiologic data supporting that the development of B-ALL in susceptible children could be the result of an immune stress after infection exposure in postnatal life.

[H1] Genetic vulnerability of preleukemic pre-B cells

Early B cell development is a precisely orchestrated process. Therefore, it is not surprising that B-ALL predisposing mutations involving genes important for haematopoietic development might introduce a specific vulnerability to transformation in those preleukemic cells once exposed to stress (Figure 1). This susceptibility to transformation is sometimes associated with altered B cell function (for example, immunodeficiencies) or deregulated B cell development. Indeed, as we have discussed, B cell vulnerability can be triggered by the malfunctioning of key developmental genes that lead to a partial block in normal B cell development in the preleukemic stage. This block is well illustrated both in humans and in animal models by the diverse alterations affecting genes like *PAX5*[44,45,123], *ERG*[40,124,125] or *ARID5B*[36,37,126], or by the appearance of chimeric proteins like ETV6–RUNX1 [17,127,128]. By way of an example, the transcription factor PAX5 is one of the key regulators of B cell development[123,129], and heterozygous *Pax5*[+/-] animals tend to accumulate a larger population of B cell progenitors[16,130]; this also happens in mice bearing B-ALL-associated oncogenes like the *ETV6–RUNX1* fusion[17,127,128]. Other studies integrating GWAS with transcriptomic, epigenomic, and 3D chromatin interaction data for these leukemia risk loci suggest that the key mechanisms controlling genetic susceptibility to B-ALL are indeed the deregulation of B cell development, apoptosis and cell cycle signalling[131].

The accumulation of an enlarged, vulnerable, progenitor population would increase the possibility of malignant transformation through the accumulation of secondary mutations in the presence of a selective pressure triggered, for example, by an otherwise 'normal' infection. The activity of the recombination-activating genes

11

RAG1 and RAG2 is essential for immunoglobulin gene rearrangement at the early stages of normal lymphocyte development, but it also poses a threat to the lymphocyte genome[132] as, at the time of rapid expansion of the B cell compartment during development (which in humans overlaps with the 2-5 years-old peak of B-ALL incidence), RAG activity can mediate undesired mutations. Indeed, RAG-mediated deletions at this stage seem to be the dominant mutational process, characterized by illegitimate cryptic RAG-mediated recombination events of pseudo-random nature that lead to the appearance of second hits in key B cell developmental genes, likely triggering B-ALL progression[133-135]. However, although the majority of the carriers of all these susceptibility mutations (for example, in *PAX5*, *IKZF1*) will never develop leukemia[26,35], there are also examples in mice showing that *IKZF1* alterations promoted the development of lymphoid leukemia from preleukemic BCR-ABL1 cells in a pathogen-free environment[136]. This suggests that the presence of preleukemic cells in itself may not be sufficient to trigger full-blown leukemia, and that something else is required, perhaps an environmental trigger or an additional susceptibility gene that accelerates a process that could also arise naturally (Figure 2). A key emerging question in this context is the biological, evolutionary reason behind the existence of stable genetic variants predisposing the carrier to the development of B-ALL.

[H1] Infection as a modulator of B-ALL development

The concept of 'gene–environment interaction' refers to the concept of both genetic and environmental factors combining to predispose individuals to developing a particular disease or to acquire a particular characteristic[24,67]. As we have described, childhood B-ALL is a good example of this situation[86]. The interaction between infection and B-ALL development has been a focus of epidemiological studies for over 60 years. As already mentioned above, the peak prevalence of B-ALL occurence at 2-5 years of age corresponds to the time of B cell maturation and rapid expansion, associated with the activation of RAG enzymes, which are responsible for many of the second hits in B-

ALL (*PAX5, IKZF1, ERG, CDKN2A/B*, etc.)[133,134]. This conceptually fits with the inital hypotheses of both Leo Kinlen and Mel Greaves[86,107,108,111] and suggests that the immune modulation could influence subsequent disease progression. Of course, B cell population expansion can clearly be influenced to different degrees by infection exposure (and by factors such as child care and living conditions). However, until recently, there was no direct evidence supporting exposure to infection as a second hit in the progression of the disease from its prenatal initiation (Figure 2). In 2015, the group of Markus Müschen showed that *ex vivo* exposure to LPS triggers transformation of ETV6–RUNX1$^+$ precursor B cells[135]. Also, early innate immune response induction by Toll-like receptor (TLR) ligation could reduce leukemia penetrance in the Eμ-ret and E2A–PBX1 transgenic mouse models[14]. These studies provide support for the hypothesis that immune modulation during the preleukemic phase can significantly alter progression to B-ALL. In line with this, a heterozygous *Pax5* mouse model has recently been used to mimic the first step leading to the B-ALL predisposition found in some human patient families[16]. In this setting, *Pax5*$^{+/-}$ mice housed in a specific pathogen-free (SPF) environment never developed B-ALL; however, when these predisposed mice were moved to a conventional facility in the presence of infectious agents, 22% of them developed B-ALL[16]. These findings suggest that exposure to infections might be a triggering factor for full-blown B-ALL development in genetically predisposed patients. Furthermore, the specific genetic alterations associated with progression to leukemia in these mice models matched those seen in human patients[44,45], further supporting the idea that this model closely recapitulates the human pathology. Another observation sustaining this finding is the fact that this relationship between infection and B-ALL is not restricted to the *Pax5*-based cases. Indeed, mice expressing *ETV6–RUNX1* at the HSC level behaved in a similar way, with absence of disease in SPF conditions and development of B-ALL after exposure to natural infections[17,127,128]. In fact, this mechanism might also be at work in B-ALLs with germline *ETV6* variants[26], a possibility also supported by the fact that *ETV6* risk alleles

are mutually exclusive with the ETV6-RUNX1 fusion in B-ALL, suggesting that they may be involved in a common leukemogenic pathway[47].

A second layer of this gene–infection interaction is defined by the fact that these leukemia-predisposing mutations, either germline or somatic, progress to B-ALL through different mechanisms, even if they are exposed to the same type of pro-oncogenic environment (such as infections). For example, *Pax5*[+/-] heterozygous mice exposed to pathogens develop leukemia with mutated *Jak3* kinase, while mice expressing the *ETV6–RUNX1* fusion gene in the haematopoietic stem/progenitor (HSPC) compartment suffer recurrent alterations affecting genes of the lysine demethylase family[17]. However, when *ETV6–RUNX1* mice are triggered by a single dose of the mutagen 1-ethyl-1-nitrosourea (ENU), they developed T cell neoplasias[128]. This suggests that the type of vulnerability to B-ALL generated by the initiating mutations may be specific for a given mutation, even though progressing through phenotypically similar stages and being triggered by apparently similar secondary exposures. The behaviour of these models closely mimics what we know about the incomplete penetrance of B-ALL-predisposing mutations in human patients, therefore enforcing the credibility of the 'infectious hypothesis'[13]. However, this mechanism does not necessarily apply to all B-ALL genetic predisposition conditions, as illustrated by the case of the Ph1[+] B-ALLs with the *BCR-ABL*[p190] oncogene, where preleukemic clones carrying *BCR-ABL*[p190] oncogenic lesions can also be found in human neonatal cord blood[137,138] while, in mouse models, *BCR-ABL*[p190+] B-ALL development is infection-independent[66], in agreement with the fact that, in humans, this B-ALL subtype rarely occurs in children[58,133]. Also, chromosomal translocations associated with the distinct infant leukemias (for example, mixed lineage leukemia (MLL) see above) seem to be capable of triggering full-blown leukemic development without the need of secondary events, therefore using other molecular mechanisms to achieve these aims[139-141].

Therefore, the varied aetiology of childhood B-ALLs implies that, most likely, infection exposure is not the only immune stress capable of triggering preleukemia-to-leukemia conversion. In order to ascertain this possibility, one would need to keep predisposed (pre-leukemic) carriers in a pathogen-free environment, to rule out a role of infection exposure. When *Pax5*[+/-] mice were exposed early in life to antibiotic treatment as an immune stressor while being maintained under SPF conditions throughout their life span[142], in spite of the SPF conditions, the antibiotic-treated animals developed B-ALL with an even higher frequency than when exposed to natural infections in the 'delayed exposure' setting (see above), and presented the same molecular alterations and disease latency. This *in vivo* result is in line with experimental data from *in vitro* studies of *ETV6–RUNX1*[+] B-ALL[135] and *E2A–PBX*[+] B-ALL[14], which collectively show that several types of immune stress can trigger clonal evolution of preleukemic clones. Further, these recent data from both *Pax5*[+/-] and *ETV6–RUNX1*[+] mice show that the genetic predisposition is systematically associated with changes in the microbiota regardless of whether the animals develop B-ALL or not[142,143]. This suggests an important role for the gut microbiota in cancer predisposition and aetiology[144,145], a possibility that had been previously proposed for childhood B-ALL based on the similarity of its development with other early-life immune disorders, like autoimmune and allergic conditions[86]. Indeed, a similar function of the gut microbiota has been suggested for other multifactorial diseases such as Crohn's disease[146], Parkinson's disease[147] and other cancers[145]. These findings have important implications for the use of antibiotics in infants and children, especially those carrying a B-ALL predisposition[143].

In summary, although the preclinical findings have exciting epidemiological and public health implications[143], clearly further research is necessary to identify the precise mechanisms regulating the interaction of the microbiota with B cell development in both healthy individuals and B-ALL-predisposed ones. Given the low frequency of predisposing *PAX5* germline mutations in children with B-ALL, other preleukemic

conditions caused by the more frequent ETV6–RUNX1 fusion, *ETV6* germline mutations or single nucleotide polymorphisms (SNPs) associated with the recently discovered high polygenic risk score for childhood ALL[131] should be studied in more detail with respect to microbiome dysbiosis, ideally by screening of newborns for preleukemic alterations together with microbiome and immune analyses. However, since many variables may be involved (for example, delivery mode, nutrition and others.) the cohorts would need to be large, population-based cohorts, and such samples are not easily collected, since a large series of longitudinal sampling would be required from birth onwards[9,148]. Population-based studies are currently being initiated and their results may have broad implications for clinical practice and public health.

[H1] An integrative model for B-ALL

All of the information from the epidemiological and clinical studies and mouse models that we have summarized so far points to some important conclusions.

One, the crucial step in the transition from preleukemia to full-blown B-ALL seems to be the interaction between the preleukemic cell and an immune stressor (since in mouse models at least, B-ALL never develops without an immune stress) (Figure 2).

Two, the data from the animal models reconcile apparently discordant epidemiological findings, and the time at which this interaction takes place seems to be less critical than originally thought, since the disease may appear in adult mice when exposed to a natural infection environment (although, in the case of children, they will nevertheless be exposed to immune stress sooner or later, something that is also suggested by the 2-5 years-old peak). Children encounter many different infectious agents, but all attempts to identify a single consistent infectious agent have failed in humans with B-ALL[149]. This supports the idea that what is important for leukemia development is a sequence of infections and immune responses in an unprimed immune system, rather than exposure to a single specific pathogen. Furthermore, months prior to B-ALL presentation, children show a marked deterioration of immune response and profound

susceptibility to a variety of infections[113]. Three, the immune stress can be induced by infections (natural infections or through specific TLR stimulation) but also by other factors like antibiotics or microbiome alterations (Figure 2). Four,subclinical genetic predisposition to leukemia can shape a gut microbiota different from that of healthy individuals. Five, an intact gut microbiome protects these predisposed mice against leukemia, and its alteration is linked to the onset of the disease. Six, the immune stress does not seem to act by expanding a previously existing silent transformed clone or to induce a random mutagenic process[130], a fact that is in agreement with the discordant genetics of the twin data, the peak incidence of childhood B-ALL[4,63,84,150,151], and the low rate of mutations normally present in childhood B-ALL[152]. On the contrary, there seems to be a specific interaction between the first hit and the immune stress, and this interaction is capable of defining the nature of the second hit.

We need to emphasize that further research is required to see if this mechanism can apply to all genetic preleukemic conditions or only to some of them; but, in summary, we postulate (Figure 2) that an immune evasion takes place in a reduced number of occasions where a cell takes advantage of the immune stress to adquire a second hit and to escape immune control, progressing to leukemic development, therefore implying that childhood B-ALL may be a preventable cancer.

If childhood B-ALL could indeed be a preventable cancer triggered by immune stress, it is therefore critical to identify the immune mechanisms involved in this evolution to leukemia. There are several potential immune mechanisms that could be involved in this immune evasion (Box 2), although the knowledge of some of these mechanisms themselves is still in its infancy, so further research into their biological foundations is still necessary, and of course also about their potential role in leukemia. Besides, their participation in leukemogenesis doesn't need to be mutually exclusive and, furthermore they could be specific for or restricted to different preleukemia-initiating alterations. It would be advisable to carry out proof-of-principle experiments in mouse models before the screening of newborns for preleukemic alterations, alongside

17

with microbiome and immune analyses. As previously mentioned, since many variables can play a role in the human setting, the posterior patient cohorts would need to be large population-based cohorts. In any case, it is important to emphasize that any attempt of enhancing immune responses in order to prevent leukemia development should be balanced to avoid an unwelcome immune stress that may result in unintended and harmful consequences.

Concluding remarks

Despite remarkable advances made in the field, the understanding and potential targeting of the interaction between the preleukemic cell and the immune stress poses a unique challenge, warranting multidisciplinary investigation[71]. Concerted international collaboration is an unavoidable requisite for a successful translational research aimed at preventing infection-driven leukemogenesis. Key collaborative goals should include, among others: the generation of biobanks of fresh cord blood samples; the standardization of experimental and analytical protocols; the standardization of the phenotypical markers used for the discrimination of preleukemic cells. Also, since chemotherapy or radiotherapy, or even malignant progression, may alter the patterns of expression of surface antigens on preleukemic cells, like in the rest of the haematopoietic system[153], the stability of these markers should be first determined.

The use of unified cell culture methods whenever *ex vivo* expansion of preleukemic cells will be required to obtain enough cells for these studies. Also, a complete understanding of how environmental factors affect B-ALL development will require intact, unmanipulated animals carrying a genetic susceptibility similar to that of predisposed children. Finally, the use of next generation sequencing technologies —single-cell RNA sequencing, together with proteomic and advanced cytometries like multicolor quantitative confocal cytometry or mass cytometry — should allow the identification of the fine molecular mechanisms governing preleukemic cell behavior at the single-cell level.

Hopefully, the elucidation of the mechanisms regulating the interaction between preleukemic cells and immune stress will provide us with the strategies that will allow us to prevent childhood B-ALL development.

Acknowledgements

Author contributions

The authors contributed equally to all aspects of the article.

Peer review information

Nature Reviews Immunology thanks C. Mullighan, M. Witkowski and the other,

anonymous, reviewer(s) for their contribution to the peer review of this work.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps

and institutional affiliations.

BOX1. The discovery of the pre-leukemic clone

In 1998, a seminal study[150] showed that identical twins had developed *ETV6-RUNX1* leukemias with identical t(12;21) chromosomal breakpoints, something that could only be explained by the existence of a preleukemic clone already present *in utero* and passed through the shared placenta. This was confirmed by studies tracing back childhood leukemias to blood samples taken at birth (Guthrie cards)[148,154,155]. Furthermore, studying pairs of B-ALL-sick and healthy twins showed that the healthy co-twin also presented a population of preleukemic cells carrying the same primary oncogenic hit, although they did not develop leukemia[151]. Additionally, in the cases of two leukemic twins, it was also found that a number of mutations differed between the twins[150], therefore indicating the postnatal secondary origin of those alterations. This evidence demonstrated the existence of a preleukemic clone originated in embryonic life, but unable by itself to give rise to a full B-ALL, a process for which a secondary alteration was required after birth[86] (Figure 1). The kinetics of this process was further clarified when it was found that up to 5% of healthy children may carry the preleukemic clone, the majority of them without any consequence, since only a small percentage of them will develop a full-blown leukemia[8,9]. For most aberrations linked to childhood pre-B-ALL, there is evidence showing that the occurrence of the aberration is more frequent than the incidence of the corresponding leukemia[156]. Therefore, the existence of a preleukemic hit is a necessary, but not sufficient, condition to develop childhood B-ALL, and the rate-limiting factor for the development of full-blown leukemia is the second hit[86].

BOX 2 Immune evasion and B-ALL prevention: challenges

A therapeutic enhancement of the immune control might allow childhood B-ALL prevention. It is therefore critical to identify the immune mechanisms involved in this progression, which might include:

Microbiota as a modifiable therapeutic target. A diet-dependent approach might be used to reduce or abolish the impact of immune stress in predisposed carriers[86].

T cell immunosurveillance. *ETV6-RUNX1*[+] B-ALLs have autoreactive T cells directed against leukemic cells[157], and could therefore be responsive to checkpoint blockade or other T cell–targeted therapies.

Trained immunity.[158,159] Innate immune training of granulopoiesis suppresses tumor growth and potentiates checkpoint inhibition[160,161]. The identification of an approach to train innate immunity, avoiding the immune stress, and therefore preventing B-ALL from developing would be a major advance in the prevention of the disease.

A transient immune amnesia state. Immune amnesia has already been observed in measles virus infections where the infection diminishes preexisting antibodies that offer protection against other pathogens[162,163]. A potential anti-ALL immune response might be lost (erased) by an immune stress, therefore triggering leukemia development.

Variation and selective pressure within preleukemic cells. Population variability could likely provide a means for preleukemic cells to adapt their behavior to the environmental conditions. Measuring and understanding the dynamics of preleukemic cells under immune stress is key for leukemia prevention.

Remodeling of the immune microenvironment. The extent to which preleukemic cells might shape the bone marrow immune microenvironment remains unknown, and its understanding may shed light on mechanisms of extrinsic regulation triggering B-ALL conversion.

Immune evasion as a 'bad luck' effect. Immune evasion as a bad luck effect refers to the fact that the nature and number of potential immune-based causes can be very variable, while being each one of them of low penetrance. Under these circumstances, the immune evasion mechanisms would be virtually indistinguishable from chance.

FIGURE 1. Genetic vulnerability of B-ALL preleukemic clones. A) Transcriptional regulation determines normal B cell development. A network of transcriptional and epigenetic regulatory circuits drives the normal differentiation of HSCs to mature B cells in a stepwise, tightly regulated, process. **B)** A first oncogenic hit, usually acquired *in utero*, such as a chromosomal translocation that generated a chimeric protein (for example, ETV6-RUNX1 or TCF3-PBX1) or a mutation in a key developmental gene (such as *PAX5*, *IKAROS* and *ERG*), arises in early haematopoietic or B cell development. This will occur in a large number of healthy newborns (>5%, without taking into account inherited predisposition), and will introduce a vulnerability through the generation and expansion of a preleukemic B cell clone which nevertheless allows normal B cell development to take place; carriers are therefore clinically silent. **C)** Most children carrying this alteration will never develop B-ALL but, in a small percentage of cases, given certain conditions (such as environmental exposures), a preleukemic clone will acquire a second hit, therefore giving rise to a full-blown B-ALL. Only after these additional secondary genetic alterations have occurred (usually also affecting genes involved in normal B cell development), will a full-blown B-ALL develop. HSC: Hematopoietic Stem Cell; LMPP: Lymphoid-primed MultiPotential Progenitor; CLP: Common Lymphoid Progenitor.

FIGURE 2. A unified model for childhood B-ALL development. The first event is the presence of a genetic lesion (either congenital or generated *de novo*) leading to the development of a preleukemic clone. In the absence of such lesion **(A)**, development takes place normally, and in the context of a normal microbiome. The preleukemic clone can, however **(B)** shape the gut microbiome and lead to the appearance of a modified microbiome whose characteristics are determined by the nature of the driver genetic lesion. Still, in the absence of an immune stress, this preleukemic clone doesn't progress to give rise to leukemia. However, in the presence of an immune stress (caused for example by delayed-exposure to common infections, antibiotic treatment,

exposure to low-dose ionizing radiation, etc.) it might happen that, in a reduced number of occasions, an immune evasion takes place where a cell with a second hit takes advantage of the effects of this immune stress to escape to immune control and lead to leukemic development. The percentage of conversion to full-blown leukemia is highly dependent on the specific first genetic hit, both in human patients and in mouse models, but is always much less than the number of cases where preleukemic cells remain in this state and do not progress to 2nd hit and B-ALL. Finally, since we still don't know the mechanistic basis of the influence of the microbiome on the acquisition of the 2nd hit, or in the pre-cancerous clonal expansion, we must bear in mind that the alteration in the gut microbiome associated with the genetic predisposition seen in mice (*Pax5$^{+/-}$* and *ETV6-RUNX1^{+}*) might however not play a role in the transformation of preleukemic clones, and then there might be cases **(C)** where such progression takes place *via* microbiome-independent mechanisms.

REFERENCES

1.	Parkin, D. M., Stiller, C. A., Draper, G. J. & Bieber, C. A. The international incidence of childhood cancer. *Int J Cancer* **42**, 511-520 (1988).
2.	Linabery, A. M. & Ross, J. A. Trends in childhood cancer incidence in the U.S. (1992-2004). *Cancer* **112**, 416-432 (2008).
3.	Mullighan, C. G. The genomic landscape of acute lymphoblastic leukemia in children and young adults. *Hematology Am Soc Hematol Educ Program* **2014**, 174-180 (2014).
4.	The United Kingdom Childhood Cancer Study: objectives, materials and methods. UK Childhood Cancer Study Investigators. *Br J Cancer* **82**, 1073-1102 (2000).
5.	Greaves, M. F., Pegram, S. M. & Chan, L. C. Collaborative group study of the epidemiology of acute lymphoblastic leukaemia subtypes: background and first report. *Leuk Res* **9**, 715-733 (1985).
6.	Inaba, H., Greaves, M. & Mullighan, C. G. Acute lymphoblastic leukaemia. *Lancet* **381**, 1943-1955 (2013).
7.	Yen, H. J. et al. Patient-Reported Outcomes in Survivors of Childhood Hematologic Malignancies with Hematopoietic Stem Cell Transplant. *Blood* (2020).
8.	Mori, H. et al. Chromosome translocations and covert leukemic clones are generated during normal fetal development. *Proc Natl Acad Sci U S A* **99**, 8242-8247 (2002).
9.	Schafer, D. et al. Five percent of healthy newborns have an ETV6-RUNX1 fusion as revealed by DNA-based GIPFEL screening. *Blood* **131**, 821-826 (2018).

This study, together with Mori et al, *PNAS USA* 2002, demonstrates that the presence of the first hit in the general children population is much higher than previously anticipated.

10. Linet, M. S. et al. International long-term trends and recent patterns in the incidence of leukemias and lymphomas among children and adolescents ages 0-19 years. *Int J Cancer* **138**, 1862-1874 (2016).
11. Steliarova-Foucher, E. et al. International incidence of childhood cancer, 2001-10: a population-based registry study. *Lancet Oncol* **18**, 719-731 (2017).
12. Steliarova-Foucher, E. et al. Changing geographical patterns and trends in cancer incidence in children and adolescents in Europe, 1991-2010 (Automated Childhood Cancer Information System): a population-based study. *Lancet Oncol* **19**, 1159-1169 (2018).
13. Ward, G. The infective theory of acute leukemia. *Br J Child Dis* **14**, 11 (1917). **Classical paper, first linking infections to B-ALL development.**

14. Fidanza, M. et al. Inhibition of precursor B-cell malignancy progression by toll-like receptor ligand-induced immune responses. *Leukemia* **30**, 2116-2119 (2016). **This work shows that early innate immune response induction by Toll-like receptor (TLR) ligation could reduce leukemia penetrance in mouse models.**

15. Fidanza, M. et al. IFN-gamma directly inhibits murine B-cell precursor leukemia-initiating cell proliferation early in life. *Eur J Immunol* **47**, 892-899 (2017).
16. Martin-Lorenzo, A. et al. Infection Exposure is a Causal Factor in B-cell Precursor Acute Lymphoblastic Leukemia as a Result of Pax5-Inherited Susceptibility. *Cancer Discov* **5**, 1328-1343 (2015).
17. Rodriguez-Hernandez, G. et al. Infection Exposure Promotes ETV6-RUNX1 Precursor B-cell Leukemia via Impaired H3K4 Demethylases. *Cancer Res* **77**, 4365-4377 (2017). **Together with Martin-Lorenzo et al., *Cancer Discov.* 2015, this study provides evidence that delayed exposure to common infections trigger preleukemic-to-leukemic conversion in different mouse models, recapitulating human B-ALL genetic predispositions.**

18. Swaminathan, S. & Muschen, M. Infectious origins of childhood leukemia. *Oncotarget* **6**, 16798-16799 (2015).
19. Grabas, M. R. et al. Incidence and time trends of childhood cancer in Denmark, 1943-2014. *Acta Oncol* **59**, 588-595 (2020).
20. Biondi, A. et al. Imatinib after induction for treatment of children and adolescents with Philadelphia-chromosome-positive acute lymphoblastic leukaemia (EsPhALL): a randomised, open-label, intergroup study. *Lancet Oncol* **13**, 936-945 (2012).
21. Schultz, K. R. et al. Long-term follow-up of imatinib in pediatric Philadelphia chromosome-positive acute lymphoblastic leukemia: Children's Oncology Group study AALL0031. *Leukemia* **28**, 1467-1471 (2014).
22. Prasad, V. Perspective: The precision-oncology illusion. *Nature* **537**, S63 (2016).
23. Tannock, I. F. & Hickman, J. A. Limits to Personalized Cancer Medicine. *N Engl J Med* **375**, 1289-1294 (2016).
24. Fischer, U. et al. Cell Fate Decisions: The Role of Transcription Factors in Early B-cell Development and Leukemia. *Blood Cancer Discovery* (2020).
25. Kimura, S. & Mullighan, C. G. Molecular markers in ALL: Clinical implications. *Best Pract Res Clin Haematol* **33**, 101193 (2020).

26.   Gocho, Y. & Yang, J. J. Genetic defects in hematopoietic transcription factors and predisposition to acute lymphoblastic leukemia. *Blood* **134**, 793-797 (2019).
27.   Roberts, K. G. Genetics and prognosis of ALL in children vs adults. *Hematology Am Soc Hematol Educ Program* **2018**, 137-145 (2018).
28.   Paulsson, K. et al. The genomic landscape of high hyperdiploid childhood acute lymphoblastic leukemia. *Nat Genet* **47**, 672-676 (2015).
29.   Andersson, A. K. et al. The landscape of somatic mutations in infant MLL-rearranged acute lymphoblastic leukemias. *Nat Genet* **47**, 330-337 (2015).
30.   Yasuda, T. et al. Corrigendum: Recurrent DUX4 fusions in B cell acute lymphoblastic leukemia of adolescents and young adults. *Nat Genet* **48**, 1591 (2016).
31.   Roberts, K. G. et al. High Frequency and Poor Outcome of Philadelphia Chromosome-Like Acute Lymphoblastic Leukemia in Adults. *J Clin Oncol* **35**, 394-401 (2017).
32.   Molina, O. et al. Impaired condensin complex and Aurora B kinase underlie mitotic and chromosomal defects in hyperdiploid B-cell ALL. *Blood* **136**, 313-327 (2020).
33.   Kratz, C. P., Stanulla, M. & Cave, H. Genetic predisposition to acute lymphoblastic leukemia: Overview on behalf of the I-BFM ALL Host Genetic Variation Working Group. *Eur J Med Genet* **59**, 111-115 (2016).
34.   Moriyama, T., Relling, M. V. & Yang, J. J. Inherited genetic variation in childhood acute lymphoblastic leukemia. *Blood* **125**, 3988-3995 (2015).
35.   Pui, C. H., Nichols, K. E. & Yang, J. J. Somatic and germline genomics in paediatric acute lymphoblastic leukaemia. *Nat Rev Clin Oncol* **16**, 227-240 (2019).
36.   Papaemmanuil, E. et al. Loci on 7p12.2, 10q21.2 and 14q11.2 are associated with risk of childhood acute lymphoblastic leukemia. *Nat Genet* **41**, 1006-1010 (2009).
37.   Trevino, L. R. et al. Germline genomic variants associated with childhood acute lymphoblastic leukemia. *Nat Genet* **41**, 1001-1005 (2009).
       This article, together with Papaemmanuil et al., ***Nat. Genet.*** **2009, first showed that common low-penetrance susceptibility alleles contribute to the risk of developing childhood B-ALL, mainly by affecting transcriptional regulators of early B cell development.**

38.   Perez-Andreu, V. et al. Inherited GATA3 variants are associated with Ph-like childhood acute lymphoblastic leukemia and risk of relapse. *Nat Genet* **45**, 1494-1498 (2013).
39.   Migliorini, G. et al. Variation at 10p12.2 and 10p14 influences risk of childhood B-cell acute lymphoblastic leukemia and phenotype. *Blood* **122**, 3298-3307 (2013).
40.   Qian, M. et al. Novel susceptibility variants at the ERG locus for childhood acute lymphoblastic leukemia in Hispanics. *Blood* **133**, 724-729 (2019).
41.   Wiemels, J. L. et al. GWAS in childhood acute lymphoblastic leukemia reveals novel genetic associations at chromosomes 17q12 and 8q24.21. *Nat Commun* **9**, 286 (2018).
42.   Xu, H. et al. Novel susceptibility variants at 10p12.31-12.2 for childhood acute lymphoblastic leukemia in ethnically diverse populations. *J Natl Cancer Inst* **105**, 733-742 (2013).
43.   Churchman, M. L. et al. Germline Genetic IKZF1 Variation and Predisposition to Childhood Acute Lymphoblastic Leukemia. *Cancer Cell* **33**, 937-948 e938 (2018).
44.   Shah, S. et al. A recurrent germline PAX5 mutation confers susceptibility to pre-B cell acute lymphoblastic leukemia. *Nat Genet* **45**, 1226-1231 (2013).

45. Auer, F. et al. Inherited susceptibility to pre B-ALL caused by germline transmission of PAX5 c.547G>A. *Leukemia* **28**, 1136-1138 (2014).

46. Duployez, N. et al. Germline PAX5 mutation predisposes to familial B acute lymphoblastic leukemia. *Blood* (2020).

47. Moriyama, T. et al. Germline genetic variation in ETV6 and risk of childhood acute lymphoblastic leukaemia: a systematic genetic study. *Lancet Oncol* **16**, 1659-1666 (2015).

48. Zhang, M. Y. et al. Germline ETV6 mutations in familial thrombocytopenia and hematologic malignancy. *Nat Genet* **47**, 180-185 (2015).

49. Noetzli, L. et al. Germline mutations in ETV6 are associated with thrombocytopenia, red cell macrocytosis and predisposition to lymphoblastic leukemia. *Nat Genet* **47**, 535-538 (2015).

50. Topka, S. et al. Germline ETV6 Mutations Confer Susceptibility to Acute Lymphoblastic Leukemia and Thrombocytopenia. *PLoS Genet* **11**, e1005262 (2015).

51. Melazzini, F. et al. Clinical and pathogenic features of ETV6-related thrombocytopenia with predisposition to acute lymphoblastic leukemia. *Haematologica* **101**, 1333-1342 (2016).

52. Perez-Garcia, A. et al. Genetic loss of SH2B3 in acute lymphoblastic leukemia. *Blood* **122**, 2425-2432 (2013).

53. Sood, R., Kamikubo, Y. & Liu, P. Role of RUNX1 in hematological malignancies. *Blood* **129**, 2070-2082 (2017).

54. Nishimoto, N. et al. T cell acute lymphoblastic leukemia arising from familial platelet disorder. *Int J Hematol* **92**, 194-197 (2010).

55. Waanders, E. et al. Germline activating TYK2 mutations in pediatric patients with two primary acute lymphoblastic leukemia occurrences. *Leukemia* **31**, 821-828 (2017).

56. Gu, Z. et al. PAX5-driven subtypes of B-progenitor acute lymphoblastic leukemia. *Nat Genet* **51**, 296-307 (2019).

57. Qian, M. et al. Whole-transcriptome sequencing identifies a distinct subtype of acute lymphoblastic leukemia with predominant genomic abnormalities of EP300 and CREBBP. *Genome Res* **27**, 185-195 (2017).

58. Pui, C. H., Roberts, K. G., Yang, J. J. & Mulligan, C. G. Philadelphia Chromosome-like Acute Lymphoblastic Leukemia. *Clin Lymphoma Myeloma Leuk* **17**, 464-470 (2017).

59. Mulligan, C. G. et al. Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature* **446**, 758-764 (2007).
    This study found that mutations in genes encoding principal regulators of B lymphocyte development and differentiation were present in 40% of B-ALL cases.

60. Mulligan, C. G. et al. Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia. *N Engl J Med* **360**, 470-480 (2009).

61. Pang, S. H., Carotta, S. & Nutt, S. L. Transcriptional control of pre-B cell development and leukemia prevention. *Curr Top Microbiol Immunol* **381**, 189-213 (2014).

62. Hu, Y., Yoshida, T. & Georgopoulos, K. Transcriptional circuits in B cell transformation. *Curr Opin Hematol* **24**, 345-352 (2017).

63. Maia, A. T. et al. Prenatal origin of hyperdiploid acute lymphoblastic leukemia in identical twins. *Leukemia* **17**, 2202-2206 (2003).

64. Boulianne, B. et al. Lineage-Specific Genes Are Prominent DNA Damage Hotspots during Leukemic Transformation of B Cell Precursors. *Cell Rep* **18**, 1687-1698 (2017).

65. Chan, L. N. & Muschen, M. B-cell identity as a metabolic barrier against malignant transformation. *Exp Hematol* **53**, 1-6 (2017).

66. Martin-Lorenzo, A. et al. Loss of Pax5 Exploits Sca1-BCR-ABL(p190) Susceptibility to Confer the Metabolic Shift Essential for pB-ALL. *Cancer Res* **78**, 2669-2679 (2018).

67. Vicente-Duenas, C., Hauer, J., Cobaleda, C., Borkhardt, A. & Sanchez-Garcia, I. Epigenetic Priming in Cancer Initiation. *Trends Cancer* **4**, 408-417 (2018).

68. Ma, X. et al. Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. *Nature* **555**, 371-376 (2018).

69. Buffler, P. A., Kwan, M. L., Reynolds, P. & Urayama, K. Y. Environmental and genetic risk factors for childhood leukemia: appraising the evidence. *Cancer Invest* **23**, 60-75 (2005).

70. Kaatsch, P. & Mergenthaler, A. Incidence, time trends and regional variation of childhood leukaemia in Germany and Europe. *Radiation Protection Dosimetry* **132**, 107-113 (2008).

71. Laurier, D. et al. Childhood leukaemia risks: from unexplained findings near nuclear installations to recommendations for future research. *J Radiol Prot* **34**, R53-68 (2014).

72. Doll, R. & Wakeford, R. Risk of childhood cancer from fetal irradiation. *Br J Radiol* **70**, 130-139 (1997).

73. Kendall, G. M. et al. A record-based case-control study of natural background radiation and the incidence of childhood leukaemia and other cancers in Great Britain during 1980-2006. *Leukemia* **27**, 3-9 (2013).

74. Pearce, M. S. et al. Radiation exposure from CT scans in childhood and subsequent risk of leukaemia and brain tumours: a retrospective cohort study. *Lancet* **380**, 499-505 (2012).

75. Wakeford, R. The risk of childhood leukaemia following exposure to ionising radiation--a review. *J Radiol Prot* **33**, 1-25 (2013).

76. Rubery, E. D. Investigation of the possible increased incidence of cancer in West Cumbria. *Journal of the Society for Radiological Protection* **5**, 187-191 (1985).

77. Laurier, D. et al. Epidemiological studies of leukaemia in children and young adults around nuclear facilities: a critical review. *Radiat Prot Dosimetry* **132**, 182-190 (2008).

78. IARC. Non-ionizing radiation, Part 1: Static and extremely low-frequency (ELF) electric and magnetic fields. IARC Monographs on the evaluation of carcinogenic risks to humans. *Lyon, France: IARC Press* (2002).

79. Campos-Sanchez, E. et al. Novel ETV6-RUNX1 Mouse Model to Study the Role of ELF-MF in Childhood B-Acute Lymphoblastic Leukemia: a Pilot Study. *Bioelectromagnetics* **40**, 343-353 (2019).

80. Schuz, J. et al. Extremely low-frequency magnetic fields and risk of childhood leukemia: A risk assessment by the ARIMMORA consortium. *Bioelectromagnetics* **37**, 183-189 (2016).

81. Caughey, R. W. & Michels, K. B. Birth weight and childhood leukemia: a meta-analysis and review of the current evidence. *Int J Cancer* **124**, 2658-2670 (2009).

82. Comstock, G. W. Snippets from the past: 70 years ago in the journal. *Am J Epidemiol* **157**, 183-184 (2003).

83. O'Neill, K. A., Bunch, K. J. & Murphy, M. F. Intrauterine growth and childhood leukemia and lymphoma risk. *Expert Rev Hematol* **5**, 559-576 (2012).

84. Greaves, M. Infection, immune responses and the aetiology of childhood leukaemia. *Nat Rev Cancer* **6**, 193-203 (2006).

85. Kinlen, L. Childhood leukaemia, nuclear sites, and population mixing. *Br J Cancer* **104**, 12-18 (2011).

86. Greaves, M. A causal mechanism for childhood acute lymphoblastic leukaemia. *Nat Rev Cancer* **18**, 471-484 (2018).

87.   Hagan, T. et al. Antibiotics-Driven Gut Microbiome Perturbation Alters Immunity to Vaccines in Humans. *Cell* **178**, 1313-1328 e1313 (2019).

88.   Francis, S. S., Selvin, S., Yang, W., Buffler, P. A. & Wiemels, J. L. Unusual space-time patterning of the Fallon, Nevada leukemia cluster: Evidence of an infectious etiology. *Chem Biol Interact* **196**, 102-109 (2012).

89.   Cazzaniga, G. et al. Possible role of pandemic AH1N1 swine flu virus in a childhood leukemia cluster. *Leukemia* **31**, 1819-1821 (2017).
       This is the first report connecting a space–time cluster of childhood B-ALL with a specific viral infection.

90.   Kroll, M. E., Stiller, C. A., Murphy, M. F. & Carpenter, L. M. Childhood leukaemia and socioeconomic status in England and Wales 1976-2005: evidence of higher incidence in relatively affluent communities persists over time. *Br J Cancer* **105**, 1783-1787 (2011).

91.   Ajrouche, R. et al. Childhood acute lymphoblastic leukaemia and indicators of early immune stimulation: the Estelle study (SFCE). *Br J Cancer* **112**, 1017-1026 (2015).

92.   Gilham, C. et al. Day care in infancy and risk of childhood acute lymphoblastic leukaemia: findings from UK case-control study. *BMJ* **330**, 1294 (2005).

93.   Goodman, R. A., Osterholm, M. T., Granoff, D. M. & Pickering, L. K. Infectious diseases and child day care. *Pediatrics* **74**, 134-139 (1984).

94.   Kamper-Jorgensen, M. et al. Childcare in the first 2 years of life reduces the risk of childhood acute lymphoblastic leukemia. *Leukemia* **22**, 189-193 (2008).

95.   Ma, X. et al. Daycare attendance and risk of childhood acute lymphoblastic leukaemia. *Br J Cancer* **86**, 1419-1424 (2002).

96.   Neglia, J. P. et al. Patterns of infection and day care utilization and risk of childhood acute lymphoblastic leukaemia. *Br J Cancer* **82**, 234-240 (2000).

97.   Rudant, J. et al. Childhood acute lymphoblastic leukemia and indicators of early immune stimulation: a Childhood Leukemia International Consortium study. *Am J Epidemiol* **181**, 549-562 (2015).

98.   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).

99.   Dockerty, J. D., Draper, G., Vincent, T., Rowan, S. D. & Bunch, K. J. Case-control study of parental age, parity and socioeconomic level in relation to childhood cancers. *Int J Epidemiol* **30**, 1428-1437 (2001).

100.   Greenbaum, S. et al. Cesarean Delivery and Childhood Malignancies: A Single-Center, Population-Based Cohort Study. *J Pediatr* **197**, 292-296 e293 (2018).

101.   Marcotte, E. L. et al. Caesarean delivery and risk of childhood leukaemia: a pooled analysis from the Childhood Leukemia International Consortium (CLIC). *Lancet Haematol* **3**, e176-185 (2016).

102.   Wang, R. et al. Cesarean Section and Risk of Childhood Acute Lymphoblastic Leukemia in a Population-Based, Record-Linkage Study in California. *Am J Epidemiol* **185**, 96-105 (2017).

103.   Amitay, E. L. & Keinan-Boker, L. Breastfeeding and Childhood Leukemia Incidence: A Meta-analysis and Systematic Review. *JAMA Pediatr* **169**, e151025 (2015).

104.   Morra, M. E. et al. Early vaccination protects against childhood leukemia: A systematic review and meta-analysis. *Sci Rep* **7**, 15986 (2017).

105.   Shimkln, M. B. BCG and cancer prevention. *West J Med* **122**, 185 (1975).

106.   Miller, A. et al. Correlation between universal BCG vaccination policy and reduced morbidity and mortality for COVID-19: an epidemiological study. *medRxiv* (2020).

107. Kinlen, L. Evidence for an infective cause of childhood leukaemia: comparison of a Scottish new town with nuclear reprocessing sites in Britain. *Lancet* **2**, 1323-1327 (1988).

108. Kinlen, L. J. Epidemiological evidence for an infective basis in childhood leukaemia. *Br J Cancer* **71**, 1-5 (1995).

109. Kinlen, L. J. An examination, with a meta-analysis, of studies of childhood leukaemia in relation to population mixing. *Br J Cancer* **107**, 1163-1168 (2012).

110. Steinmaus, C., Lu, M., Todd, R. L. & Smith, A. H. Probability estimates for the unique childhood leukemia cluster in Fallon, Nevada, and risks near other U.S. Military aviation facilities. *Environ Health Perspect* **112**, 766-771 (2004).

111. Greaves, M. F. Speculations on the cause of childhood acute lymphoblastic leukemia. *Leukemia* **2**, 120-125 (1988).

112. Urayama, K. Y., Ma, X. & Buffler, P. A. Exposure to infections through day-care attendance and risk of childhood leukaemia. *Radiat Prot Dosimetry* **132**, 259-266 (2008).

113. Crouch, S. et al. Infectious illness in children subsequently diagnosed with acute lymphoblastic leukemia: modeling the trends from birth to diagnosis. *Am J Epidemiol* **176**, 402-408 (2012).

114. Cardwell, C. R., McKinney, P. A., Patterson, C. C. & Murray, L. J. Infections in early life and childhood leukaemia risk: a UK case-control study of general practitioner records. *Br J Cancer* **99**, 1529-1533 (2008).

115. Francis, S. S. et al. In utero cytomegalovirus infection and development of childhood acute lymphoblastic leukemia. *Blood* **129**, 1680-1684 (2017).

116. Kreis, C., Grotzer, M., Hengartner, H. & Spycher, B. D. Space-time clustering of childhood cancers in Switzerland: A nationwide study. *Int J Cancer* **138**, 2127-2135 (2016).

117. Roman, E. et al. Childhood acute lymphoblastic leukemia and infections in the first year of life: a report from the United Kingdom Childhood Cancer Study. *Am J Epidemiol* **165**, 496-504 (2007).

118. Roman, E., Simpson, J., Ansell, P., Lightfoot, T. & Smith, A. Infectious proxies and childhood leukaemia: findings from the United Kingdom Childhood Cancer Study (UKCCS). *Blood Cells Mol Dis* **42**, 126-128 (2009).

119. Soegaard, S. H. et al. Neonatal Inflammatory Markers Are Associated with Childhood B-cell Precursor Acute Lymphoblastic Leukemia. *Cancer Res* **78**, 5458-5463 (2018).

120. Beneforti, L. et al. Pro-inflammatory cytokines favor the emergence of ETV6-RUNX1-positive pre-leukemic cells in a model of mesenchymal niche. *Br J Haematol* (2020).

121. Portale, F. et al. Activin A contributes to the definition of a pro-oncogenic bone marrow microenvironment in t(12;21) preleukemia. *Exp Hematol* **73**, 7-12 e14 (2019).

122. Kreis, C. et al. Space-Time Clustering of Childhood Leukemia: Evidence of an Association with ETV6-RUNX1 (TEL-AML1) Fusion. *PLoS One* **12**, e0170020 (2017).

123. Cobaleda, C., Schebesta, A., Delogu, A. & Busslinger, M. Pax5: the guardian of B cell identity and function. *Nat Immunol* **8**, 463-470 (2007).

124. Ng, A. P. et al. An Erg-driven transcriptional program controls B cell lymphopoiesis. *Nat Commun* **11**, 3013 (2020).

125. Sondergaard, E. et al. ERG Controls B Cell Development by Promoting Igh V-to-DJ Recombination. *Cell Rep* **29**, 2756-2769 e2756 (2019).

126. Lahoud, M. H. et al. Gene targeting of Desrt, a novel ARID class DNA-binding protein, causes growth retardation and abnormal development of reproductive organs. *Genome Res* **11**, 1327-1334 (2001).
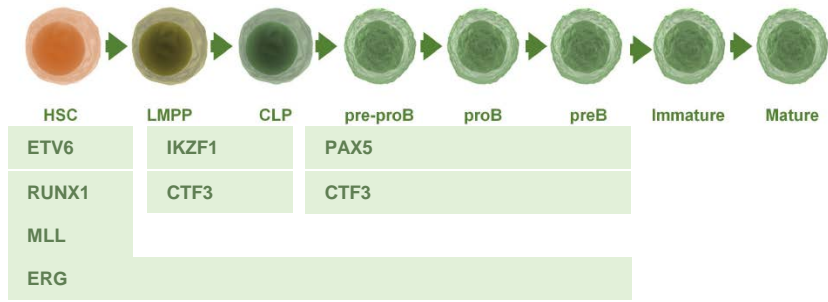
127. Boiers, C. et al. A Human IPS Model Implicates Embryonic B-Myeloid Fate Restriction as Developmental Susceptibility to B Acute Lymphoblastic Leukemia-Associated ETV6-RUNX1. *Dev Cell* **44**, 362-377 e367 (2018).

128. Schindler, J. W. et al. TEL-AML1 corrupts hematopoietic stem cells to persist in the bone marrow and initiate leukemia. *Cell Stem Cell* **5**, 43-53 (2009).

129. Cobaleda, C., Jochum, W. & Busslinger, M. Conversion of mature B cells into T cells by dedifferentiation to uncommitted progenitors. *Nature* **449**, 473-477 (2007).

130. Rodriguez-Hernandez, G. et al. Infectious stimuli promote malignant B-cell acute lymphoblastic leukemia in the absence of AID. *Nat Commun* **10**, 5563 (2019).

131. Vijayakrishnan, J. et al. Identification of four novel associations for B-cell acute lymphoblastic leukaemia risk. *Nat Commun* **10**, 5348 (2019).

132. Teng, G. et al. RAG Represents a Widespread Threat to the Lymphocyte Genome. *Cell* **162**, 751-765 (2015).

133. Mullighan, C. G. et al. BCR-ABL1 lymphoblastic leukaemia is characterized by the deletion of Ikaros. *Nature* **453**, 110-114 (2008).

134. 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).
This study identifies RAG-mediated deletions as the dominant mutational process leading to the acquisition of additional somatic mutations in the progress from preleukemic cells to overt B-ALL.

135. Swaminathan, S. et al. Mechanisms of clonal evolution in childhood acute lymphoblastic leukemia. *Nat Immunol* **16**, 766-774 (2015).
This study shows that ex vivo exposure to lipopolysaccharide (LPS) triggers transformation of ETV6-RUNX1$^+$ precursor B-cells.

136. Churchman, M. L. et al. Efficacy of Retinoids in IKZF1-Mutated BCR-ABL1 Acute Lymphoblastic Leukemia. *Cancer Cell* **28**, 343-356 (2015).

137. Cazzaniga, G. et al. Developmental origins and impact of BCR-ABL1 fusion and IKZF1 deletions in monozygotic twins with Ph+ acute lymphoblastic leukemia. *Blood* **118**, 5559-5564 (2011).

138. Kosik, P. et al. Low numbers of pre-leukemic fusion genes are frequently present in umbilical cord blood without affecting DNA damage response. *Oncotarget* **8**, 35824-35834 (2017).

139. Sanjuan-Pla, A. et al. Revisiting the biology of infant t(4;11)/MLL-AF4+ B-cell acute lymphoblastic leukemia. *Blood* **126**, 2676-2685 (2015).

140. Marschalek, R. Mechanisms of leukemogenesis by MLL fusion proteins. *Br J Haematol* **152**, 141-154 (2011).

141. Hyrenius-Wittsten, A. et al. De novo activating mutations drive clonal evolution and enhance clonal fitness in KMT2A-rearranged leukemia. *Nat Commun* **9**, 1770 (2018).

142. Vicente-Duenas, C. et al. An intact gut microbiome protects genetically predisposed mice against leukemia. *Blood* **136**, 2003-2017 (2020).
This study shows that gut microbiome alterations caused by antibiotic treatment were sufficient to induce leukemia in genetically predisposed mice in the absence of infectious stimuli.

143. Roy, A. A "gut feeling" about precursor B-ALL. *Blood* **136**, 1995-1996 (2020).

144. Gilbert, J. A. et al. Current understanding of the human microbiome. *Nat Med* **24**, 392-400 (2018).

145. Poore, G. D. et al. Microbiome analyses of blood and tissues suggest cancer diagnostic approach. *Nature* **579**, 567-574 (2020).

146. Jostins, L. et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* **491**, 119-124 (2012).
147. Sampson, T. R. et al. Gut Microbiota Regulate Motor Deficits and Neuroinflammation in a Model of Parkinson's Disease. *Cell* **167**, 1469-1480 e1412 (2016).
148. Hein, D. et al. The preleukemic TCF3-PBX1 gene fusion can be generated in utero and is present in approximately 0.6% of healthy newborns. *Blood* **134**, 1355-1358 (2019).
149. McNally, R. J. & Eden, T. O. An infectious aetiology for childhood acute leukaemia: a review of the evidence. *Br J Haematol* **127**, 243-263 (2004).
150. Ford, A. M. et al. Fetal origins of the TEL-AML1 fusion gene in identical twins with leukemia. *Proc Natl Acad Sci U S A* **95**, 4584-4588 (1998).
This study found that leukemia in identical twins had its origins in utero, in the fetus, therefore showing the existence of a prenatal predisposition to B-ALL.

151. Hong, D. et al. Initiating and cancer-propagating cells in TEL-AML1-associated childhood leukemia. *Science* **319**, 336-339 (2008).
152. Grobner, S. N. et al. The landscape of genomic alterations across childhood cancers. *Nature* **555**, 321-327 (2018).
153. Mendez-Ferrer, S., Scadden, D. T. & Sanchez-Aguilera, A. Bone marrow stem cells: current and emerging concepts. *Ann N Y Acad Sci* **1335**, 32-44 (2015).
154. Taub, J. W. et al. High frequency of leukemic clones in newborn screening blood samples of children with B-precursor acute lymphoblastic leukemia. *Blood* **99**, 2992-2996 (2002).
155. Wiemels, J. L. et al. Prenatal origin of acute lymphoblastic leukaemia in children. *Lancet* **354**, 1499-1503 (1999).
156. Hein, D., Borkhardt, A. & Fischer, U. Insights into the prenatal origin of childhood acute lymphoblastic leukemia. *Cancer Metastasis Rev* **39**, 161-171 (2020).
157. Zamora, A. E. et al. Pediatric patients with acute lymphoblastic leukemia generate abundant and functional neoantigen-specific CD8(+) T cell responses. *Sci Transl Med* **11** (2019).
158. Netea, M. G. et al. Defining trained immunity and its role in health and disease. *Nat Rev Immunol* (2020).
159. Netea, M. G. et al. Trained immunity: A program of innate immune memory in health and disease. *Science* **352**, aaf1098 (2016).
160. Kalafati, L. et al. Innate Immune Training of Granulopoiesis Promotes Anti-tumor Activity. *Cell* **183**, 771-785.e712 (2020).
161. Priem, B. et al. Trained Immunity-Promoting Nanobiologic Therapy Suppresses Tumor Growth and Potentiates Checkpoint Inhibition. *Cell* **183**, 786-801.e719 (2020).
162. Mina, M. J. et al. Measles virus infection diminishes preexisting antibodies that offer protection from other pathogens. *Science* **366**, 599-606 (2019).
163. Petrova, V. N. et al. Incomplete genetic reconstitution of B cell pools contributes to prolonged immunosuppression after measles. *Sci Immunol* **4** (2019).
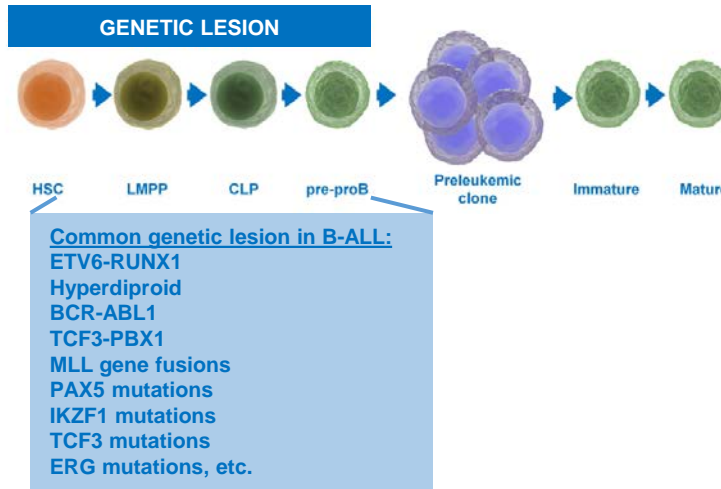
A. Normal B cell development

HSC	LMPP	CLP	pre-proB	proB	preB	Immature	Mature
ETV6	IKZF1		PAX5				
RUNX1	CTF3		CTF3				
MLL							
ERG							

B. Preleukemic B cell development

GENETIC LESION

HSC — LMPP — CLP — pre-proB — Preleukemic clone — Immature — Mature

Common genetic lesion in B-ALL:
ETV6-RUNX1
Hyperdiproid
BCR-ABL1
TCF3-PBX1
MLL gene fusions
PAX5 mutations
IKZF1 mutations
TCF3 mutations
ERG mutations, etc.

C. B-ALL development

2nd hit

Environmental exposures

GENETIC LESION

HSC — LMPP — CLP — pre-proB — Preleukemic clone — 2nd hit — B-ALL

PAX5 loss/alteration
ETV6 loss/alteration
IKZF1 loss
ERG loss, etc.

Cobaleda et al., Fig. 1

A.

Healthy child

WT cell

WT Microbiome

Normal B cell Development

B.

Healthy carrier

Preleukemic clone

Modified Microbiome

Immune Stress

2nd hit

B-ALL

C.

Healthy carrier

Preleukemic clone

?

Immune Stress

2nd hit

B-ALL

Genetic predisposition

2nd hit

Cobaleda et al., Fig. 2

	Subtype	Childhood	Adolescent and young adult	Adult	Prognosis
Aneuploid	High Hyperdiploid	21,10%	6.7%	2.9%	Very good
	Low Hyperdiploid	3,10%	3.6%	0.3%	Poor
	Low Hypodiploid	0,80%	4.5%	13.0%	Poor
	Near-haploid	2,10%	0.5%	0.8%	Intermediate
	iAMP21	2,50%	2.1%	0.3%	Good
Transcription factor rearrangement	ETV6-RUNX1	15,85%	1.4%	0.3%	Very good
	ETV6-RUNX1-like	3,35%	0.7%	0.3%	Good
	DUX4-rearranged	5,10%	7.9%	3.2%	Very good
	KMT2A	4,40%	4.1%	16.1%	Poor
	KMT2A-like	0,30%	0.2%	0.0%	Unknown
	TCF3-PBX1	5,20%	2.9%	1.1%	Good
	ZNF384/ZNF362	2,15%	3.8%	1 .3%	Intermediate
	MEF2Dr	2,10%	2.9%	1.1%	Intermediate
	NUTM1	0,90%	0.0%	0.0%	Very good
	TCF3-HLF	0,55%	0.0%	0.8%	Very poor
Transcription factor - Other	PAX5alt	6,30%	9.3%	7.7%	Intermediate
	PAX5 P80R	1,15%	3.1%	4.2%	Intermediate
	IKZF1 N159Y	0,30%	0.5%	0.5%	Intermediate
	BCL2/MYC	0,15%	1.4%	2.6%	Very poor
Kinase Signaling	Ph-like	12,80%	29.4%	20.4%	Poor
	BCR-ABL1	3,15%	5.5%	15.9%	Poor
	CRLF2 (no Ph-like)	1,05%	1 .0%	0.0%	Poor
Other	ZNF384-like	0,00%	0.7%	0.3%	Unknown
	Other	5,55%	7.9%	7.1%	Intermediate

Table I. Current molecular classification of B-ALL based on transcriptome sequencing.

(*Data extracted from* [56])