Corrected: Author Correction

A causal mechanism for childhood acute lymphoblastic leukaemia

Mel Greaves

Abstract | In this Review, I present evidence supporting a multifactorial causation of childhood acute lymphoblastic leukaemia (ALL), a major subtype of paediatric cancer. ALL evolves in two discrete steps. First, in utero initiation by fusion gene formation or hyperdiploidy generates a covert, pre-leukaemic clone. Second, in a small fraction of these cases, the postnatal acquisition of secondary genetic changes (primarily V(D)J recombination-activating protein (RAG) and activation-induced cytidine deaminase (AID)-driven copy number alterations in the case of ETS translocation variant 6 (*ETV6*)–runt-related transcription factor 1 (*RUNX1*)$^{+}$ ALL) drives conversion to overt leukaemia. Epidemiological and modelling studies endorse a dual role for common infections. Microbial exposures earlier in life are protective but, in their absence, later infections trigger the critical secondary mutations. Risk is further modified by inherited genetics, chance and, probably, diet. Childhood ALL can be viewed as a paradoxical consequence of progress in modern societies, where behavioural changes have restrained early microbial exposure. This engenders an evolutionary mismatch between historical adaptations of the immune system and contemporary lifestyles. Childhood ALL may be a preventable cancer.

Childhood acute leukaemia is the most common paediatric cancer in developed societies, accounting for one-third of all cases, with a variable incidence rate of 10–45 per 10^6 children per year and a cumulative risk of ~1 in 2,000 up to the age of 15 years[1]. The most common paediatric leukaemia, acute lymphoblastic leukaemia (ALL), is an intrinsically lethal cancer, as evidenced by a universally adverse clinical outcome before effective therapy was developed[2]. Currently, however, cure rates for ALL using combination chemotherapy are around 90%[3], making this one of the real success stories of oncology.

While this is a cause for celebration, the current treatment remains toxic, traumatic for young patients and their families, and carries some long-term health consequences[4,5]. It is unfortunate that we have remained ignorant as to the cause of ALL. The open question as to whether this cancer is potentially preventable is therefore important.

Environmental exposures possibly linked to ALL are numerous[6], but in many cases, these associations are weak, inconsistent or lacking in biological plausibility[7]. Large and multidisciplinary nationwide studies or international consortia[8,9] have provided a more enabling framework for addressing this question, but to date, the only accepted causal agent for ALL, albeit under exceptional circumstances, is ionizing radiation[10–12]. The causes of ALL might be best understood by using biological insights into the cancer itself as the foundation for designing, testing and validating hypotheses.

Childhood ALL includes a number of subtypes defined by cell lineage (B cell or T cell), differentiation status and genetics (FIG. 1a). These differ by age distribution (FIG. 1b) and clinical outcome (see FIG. 1 legend) and could have distinctive aetiologies. In this Review, I focus on the body of evidence — epidemiological, biological and genetic — that has accumulated, particularly over the past decade, and supports a causal mechanism that is selective for the common, or B cell precursor, subtype of childhood ALL (designated here as BCP-ALL)[7]. This is suggested to be a multifactorial mix of infectious exposure, inherited or constitutive genetics and chance, with patterns or timing of common infection in early life identified as the critical component and a potential route for preventive intervention.

Infection hypotheses

The idea that infections might play a causal role in childhood ALL is ~100 years old[13]. When it became clear that leukaemia in a number of animal species — chickens, mice, cattle and cats — was viral in origin[14], there was an expectation that a similar transforming virus might be responsible for childhood ALL, as well as for other blood cell cancers. To date, all attempts to molecularly identify or otherwise incriminate a leukaemogenic virus in ALL have failed[7].

Centre for Evolution and Cancer, The Institute of Cancer Research, London, UK.

e-mail: mel.greaves@icr.ac.uk

https://doi.org/10.1038/s41568-018-0015-6

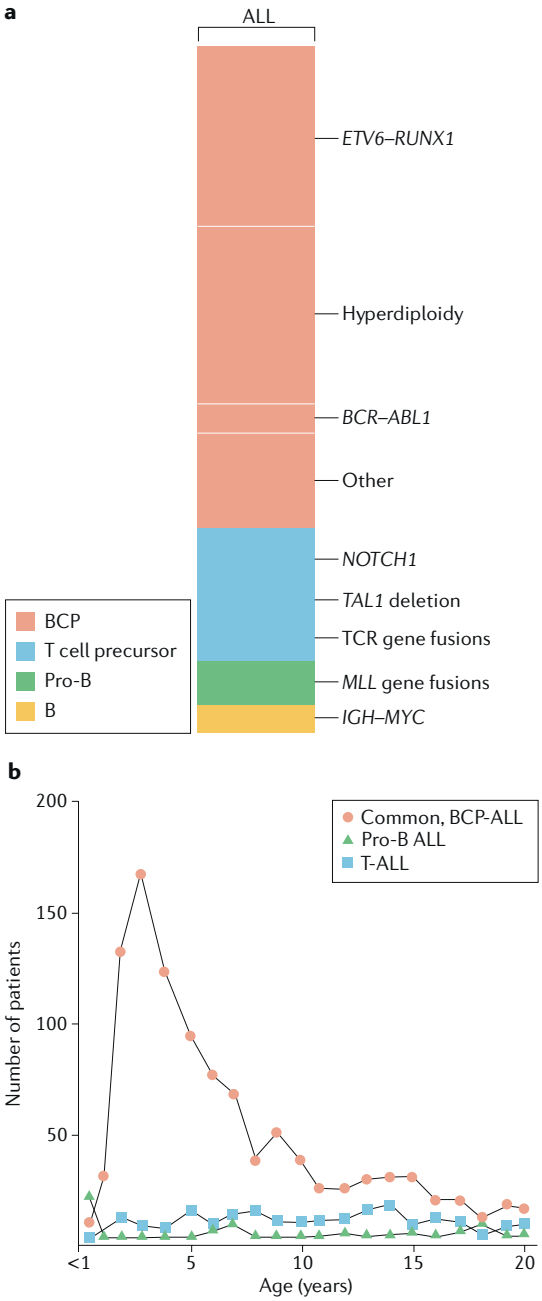

Fig. 1 | **Major subtypes of childhood ALL.** **a** | Immunophenotype screens in the 1970s and early 1980s established that acute lymphoblastic leukaemia (ALL) could be divided into subsets corresponding to early developmental compartments of the B and T cell lineages, as indicated in the key. Common or B cell precursor ALL (BCP-ALL) is genetically diverse (as illustrated), with the two most prevalent alterations being ETS translocation variant 6 (*ETV6*)–runt-related transcription factor 1 (*RUNX1*) fusion and hyperdiploidy. The rare (~2%) subtype with a mature B cell immunophenotype (and frequent *IGH–MYC* rearrangements) was subsequently recognized (and treated) not as ALL but as B cell lymphoblastic lymphoma. For more detailed descriptions of genomic diversity in ALL, see REFS[179–181]. **b** | The age distribution of ALL subtypes from a cohort of 1,184 patients with ALL entered into MRC-UKALL clinical trials[182] (1975–1984) is shown. This pattern of age-associated ALL subtypes was validated in a later cohort of MRC-UKALL trials (1991–1996; 1,088 patients up to 14 years of age)[8]. It had been known that childhood ALL had a very marked incidence peak at 2–5 years of age throughout the developed world. But this peak appeared to be diminished or absent in less developed societies and appeared in particular countries and ethnic groups at different times[182]. Immunophenotypic screens, linked to clinical trials in the UK[183], and an international collaborative group study[182] documented that the peak in incidence was selective for common or BCP-ALL. Recent epidemiological data indicate that the incidence of this subtype of leukaemia in Europe has continued to increase at ~1% per year[184–186]. This suggested that the increase over time was real, rather than ascertainment bias, and that BCP-ALL had a distinctive aetiology. The BCP-ALL subtype was also found to have a much more favourable clinical outcome[183,187,188], emphasizing its distinct biology. *MLL*, mixed-lineage leukaemia; Pro-B, B cell progenitor; TCR, T cell receptor. Part **b** is adapted with permission from REF.[182], Elsevier.

In 1988, two hypotheses were presented that suggested a new perspective on this problem. The two models are sometimes considered as alternative or competing explanations. I believe they portray the same picture through different lenses. Both propose that childhood leukaemia may arise as a consequence of an abnormal immune response to common infection(s). One model advanced by epidemiologist Leo Kinlen was based on transient and localized increases in the incidence of childhood leukaemia that could be ascribed, epidemiologically, to population mixing[15,16] (BOX 1).

The other model that I proposed was dubbed the 'delayed infection' hypothesis, the focus of this article, and was more biological than epidemiological in its origins and was applied specifically to BCP-ALL[7,17]. Central to this were two propositions.

First was the idea that the immune system, in both its innate and adaptive arms, had evolved to both anticipate and require microbial infectious exposure perinatally or in infancy[18]. The dynamics and composition of the microbiome and virome of infants is highly variable[19], and early microbial exposures have lasting impacts on immune function and health[20,21]. Metabolites of commensal bacteria promote regulatory T cells and affect subsequent inflammatory signalling pathways[22]. Deficits of this natural microbial experience, especially in modern societies, result in an unmodulated or distorted immune network[23]. A consequence of an under-exposed immune network in infancy was predicted to be subsequent dysregulated responses to common infections that could promote or trigger BCP-ALL. The increased incidence in childhood ALL in developed societies was therefore considered to be a paradox of progress, and the link to infection was considered to be inverted: the problem might be lack of infection. An equivalent mismatch of evolutionary adaptations and modern lifestyles may underlie the causation of several common adult cancers in the developed world[24].

The second proposition related to the natural history of the disease. The speculation was that ALL most likely developed by two critical steps: first, an initiating event

Box 1 | **The population-mixing hypothesis**

The Kinlen hypothesis[15,16] was prompted by public concerns over apparent increases in childhood leukaemia in the vicinity of nuclear power plants in the UK. The Kinlen model proposed that childhood leukaemia was caused by a rare or abnormal reaction to a common infection of low pathogenicity in a population at risk because of migration and mixing in the context of a lack of herd immunity[150]. By analogy with animal leukaemias, Kinlen favoured a specific virus. The model was also considered to apply across the board to childhood blood cell malignancy, irrespective of subtype, and was not prescriptive with respect to timing of infection in the life of a child.

The population-mixing hypothesis for childhood leukaemia has been explored by Kinlen and others in a variety of geographic and demographic settings. The data suggest that where an influx of adults and families occurs, particularly into somewhat isolated or rural areas, a transient increase, on the order of twofold, occurs in incidence rates, which is compatible with an infectious aetiology[16,106,151–153]. These studies have not been informative with respect to timing of infection in relation to the natural history of disease, the nature of the infection(s) involved or mechanistic aspects. Nevertheless, they provide an important body of evidence supporting a role for common infection(s) in childhood leukaemia.

in utero and second, a postnatal mutational event that promotes clinical leukaemia development. The prediction was then that an abnormal immune response to infection(s) indirectly triggered the requisite secondary mutational events. No specific infection was proposed in relation to either protection in early life or postnatal promotion and the immunological mechanism was considered to be indirect and therefore not akin to a transforming virus.

Common infections were therefore proposed to have two opposing impacts on risk of ALL that depend on timing — antagonistic (early) or promotional (late). A parallel would be with the divergent roles of microbial infections and chronic inflammation in gastrointestinal and other common cancers in adults[25,26].

The two-hit model of childhood ALL

With a few informative exceptions (Supplementary Box 1), ALL has a clinically silent natural history before diagnosis. My colleagues and I developed three different tactics for backtracking the origins of this covert process to before birth. These exploited the fact that common fusion genes in ALL (for example, ETS translocation variant 6 (*ETV6*)–runt-related transcription factor 1 (*RUNX1*) (also known as *TEL–AML1*), mixed-lineage leukaemia (*MLL*)–*AF4* (also known as *KMT2A–AFF1*) and *BCR–ABL1*) have uniquely variable or idiosyncratic breakpoints within the intronic, breakpoint cluster regions of the two partner fusion genes involved. The genomic sequences at the gene fusion junction provide stable, sensitive and clone-specific markers[27,28].

Comparative genomics of concordant ALL in monozygotic twins. Studies on monozygotic twins have been especially informative[29]. The possibility that concordance of leukaemia in identical twins might be attributable not to co-inheritance of genetic susceptibility but to an in utero origin in one twin was first proposed in 1962 (REF.[30]) and elaborated on in 1971 (REF.[31]). This idea was based on previous understanding that monochorionic or single placentae have vascular anastomoses permitting twin–twin blood transfusion with consequent blood cell chimerism[32].

Vascular anastomoses
Interconnected blood vessels (venous or arterial) in a twin, monochorionic placenta that facilitate the transfer of blood cells and fluids between identical twins.

Blood cell chimerism
The sharing of blood cells in monozygotic twins that developed in a single, or monochorionic, placenta.

The prediction was then that ALL in both twins should arise in one twin but be monoclonal[31]. Unambiguous evidence that this was the case derived from the finding of identical, but non-constitutive, clone-specific fusion gene breakpoints and sequences in a series of twin pairs[29,33] (FIG. 2). That evidence is strengthened by the observation of shared, clone-specific immunoglobulin heavy chain (*IGH*) diversity–joining (DJ) or variable–diversity–joining (V(D)J) genomic sequences of concordant BCP-ALL in twins[34].

With further exploration of ALL genomes in twins, it has become clear that while patients with concordant ALL share the identical and singular fusion gene event, other genetic alterations present, including copy number alterations (CNAs) and single nucleotide variants (SNVs), were different in twin pairs[35,36] (FIG. 2). This then suggested that such distinctive mutational events reflected independent and divergent subclonal evolution postnatally. Similarly, the majority of ongoing V(D)J rearrangements in *IGH* are subclonal and distinctive in twin pairs[34]. In one twin pair with concordant *ETV6–RUNX1*$^{+}$ ALL, whole-genome sequencing revealed that the fusion gene was the only shared or clonal genetic lesion[37]. These data endorsed the likelihood that *ETV6–RUNX1* fusion was an initiating event or founder mutation for ALL.

The concordance rate in monozygotic twins varies according to age and ALL subtype. In infants (<18 months) with B cell progenitor ALL (pro-B ALL) and *MLL* fusions, the rate approximates to 100% for those with a monochorionic or single placenta[29]. This suggested that *MLL* fusion-driven leukaemogenesis in such infants was essentially completed in utero and that the fusion gene, or a single mutation, was sufficient for leukaemogenesis. Subsequent genomic sequencing of these patients is compatible with this possibility even though other subclonal mutations, for example, in RAS family genes, do occur[38,39].

Pre-leukaemic clones in healthy co-twins. The concordance rate in older children with BCP-ALL was calculated to be around 10–15%, which is lower than that in infants[29]. A prediction for those pairs of twins with a monochorionic placenta, where only one twin develops ALL, is that the healthy co-twin should have a population of covert pre-leukaemic cells harbouring the same initiating lesion as his or her co-twin with ALL, that is, the twins are discordant for the critical postnatal secondary genetic event. This has been confirmed in three twin pairs with BCP-ALL with hyperdiploidy[40], *BCR–ABL1* fusion[36] or *ETV6–RUNX1* fusion[35,41]. In this context, the healthy co-twin provides a rare 'experiment of nature' and unique access to the pre-leukaemic clone. Putative pre-leukaemic cells from the blood of one healthy co-twin and propagation in vitro and in vivo (in NOD–severe combined immunodeficient (SCID) mice) established that these cells have both self-renewal capacity and intact B cell differentiation capacity[41], features commensurate with a pre-leukaemic status. Equivalent pre-leukaemic stem cells for acute myeloid leukaemia (AML) have now been identified in patients with AML[42] and healthy adults[43].

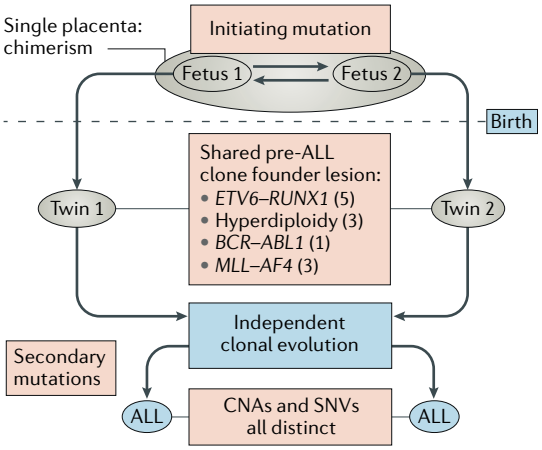

Fig. 2 | **Summary of comparative genomics of ALL in identical twin pairs.** The figure is based on analysis of 12 monozygotic twin pairs (the number of pairs with each founder lesion is noted in parentheses) with concordant acute lymphoblastic leukaemia (ALL)[29,33,37,40,48]. The sharing of a patient-specific and clone-specific fusion gene that is not inherited in the twins indicates that in such cases of concordant ALL, the leukaemia must have been initiated in a single cell, in one twin of the pair in utero, and the clonal progeny of that cell then disseminated to the co-twin via intra-placental anastomoses. In further support of this notion, it was noted that concordance of ALL occurred only when the twins shared a single or monochorionic placenta, providing a route for cellular transmission[29]. CNA, copy number alteration; *ETV6*, ETS translocation variant 6; *MLL*, mixed-lineage leukaemia; *RUNX1*, runt-related transcription factor 1; SNV, single nucleotide variant. Figure adapted with permission from REF.[189], Elsevier.

Backtracking early genetic events in ALL to neonatal blood spots. Less than 1% of childhood BCP-ALL cases occur in twins. However, ALL in twins is no different in its biological and clinical features or age incidence to that in singletons. This suggests that many or most childhood ALL cases in singletons are also initiated in utero.

To validate this proposition, my colleagues and I exploited the fact that neonatal blood spots, also known as Guthrie cards, contain reasonably intact DNA. Archived blood spots from patients with ALL were screened for clone-specific fusion gene sequences identified at diagnosis. This was first carried out in three infants with ALL with *MLL–AF4* fusion, and blood spots from all three patients evaluated were positive[44]. Subsequent studies with samples from children with *ETV6–RUNX1*$^+$ ALL found that ~75% were positive[45]. These results have been independently confirmed[46,47]. Negative blood spot results are uninterpretable, as this could reflect either a postnatal origin or an inadequate number of leukaemic cells in the sample. The conclusion drawn from these screens was therefore that the majority of childhood ALL cases, though possibly not all, were prenatal in origin.

The twin and blood spot studies also provided insight into persistence of pre-leukaemic stem cells and postnatal latency in ALL. The oldest twin with concordant ALL originating in utero was 14 years at diagnosis, and her twin sibling had been diagnosed with ALL some 9 years earlier[48]. The oldest non-twin patient with ALL and a positive neonatal blood spot to date was diagnosed at 9 years and 4 months old[49].

Frequency of ALL initiation in utero. The data on discordant, monozygotic twins suggested that some or possibly most individuals harbouring a prenatally generated, covert pre-leukaemic clone never progress to overt ALL. This begs the important question, relevant to aetiology, of the frequency of initiation of ALL in utero and the frequency of its transition to overt leukaemia.

To address this issue, my colleagues and I screened a large cohort of unselected cord blood samples for *ETV6–RUNX1* fusion mRNA (data summarized in Supplementary Fig. 1). The striking result was that ~1% of newborn babies (6/567) had a covert and modest sized, putative pre-leukaemic population of ~1 in 10^4 B lineage cells[50]. This result, initially challenged[51], has been independently confirmed[52–54]. An ~1% incidence for *ETV6–RUNX1* in relation to incidence of the leukaemia itself reflects a low transition probability of ~1%, with 99% of pre-leukaemic clones initiated during fetal development never progressing to clinical ALL. This low transition probability could reflect either lack of persistence of the pre-leukaemic stem cells after birth or a severe bottleneck in acquisition of the necessary secondary genetic changes.

These data suggest that initiation of leukaemia in utero is far more common than indicated by the incidence of disease and carry implications for causation. The same may apply to some other paediatric cancers. Histological evidence and some genetic data suggest that the frequency of precursor lesions for neuroblastoma[55] and Wilms tumour[56] is also some 100 times the incidence of clinical cancer[50].

Other subtypes of ALL. These lines of investigation were pursued using fusion genes as the predominant clonal markers of early genetic events in ALL. The most frequent subtype of BCP-ALL is, however, characterized by chromosomal hyperdiploidy, which is harder to track than fusion genes. There is evidence, however, that the key findings described above for the *ETV6–RUNX1* subset are likely to apply to hyperdiploid ALL. Monozygotic, monochorionic twin pairs concordant for hyperdiploid ALL are described with identical karyotypes[40], and neonatal blood spots of children with hyperdiploid ALL have clone-specific *IGH* sequences[57–59]. In one case of hyperdiploid ALL, the child's cord blood had been frozen at birth. Retrieval of this sample led to the identification of putative pre-leukaemic cells in the cord blood with the same triploid chromosomes as in the child's subsequent ALL[60]. Hyperdiploidy, generated by a one-off abnormal mitosis resulting in trisomies[61], can therefore occur in utero as an alternative initiating event to gene fusion for BCP-ALL.

Further genomic exploration of ALL

Whole-genome sequencing of a cohort of 57 patients with *ETV6–RUNX1*$^+$ BCP-ALL provided an audit of all mutational changes[62]. This confirmed the previous finding that the most common recurrent events were CNAs,

Neonatal blood spots
Also known as Guthrie cards, these are samples of dried blood collected from a newborn baby shortly after birth via a heel prick that are routinely used to detect genetic conditions.

primarily gene deletions[63]. SNVs were also present but with low or undetectable recurrency[62].

Genomic sequencing in cancer can reveal mutational signatures of relevance to aetiology[64]. Almost 50% of CNAs in *ETV6–RUNX1*$^{+}$ BCP-ALL had partial or complete V(D)J recombination-activating protein (RAG) heptamer-nonamer recognition motifs within 20 bp of the breakpoints[62]. This finding may explain the observation that highly recurrent CNAs in BCP-ALL are reiteratively present in subclones of individual patients[65,66]. A comparison with CNAs in ~14,000 patients with breast, prostate or pancreatic cancer revealed none with RAG motifs[62]. SNVs in *ETV6–RUNX1*$^{+}$ BCP-ALL had two main mutated signatures: one was C>T transitions at CpGs and C>G and C>T mutations at TpCs, and a second was transitions and transversions in a TpC context at NpCpG trinucleotides[62]. This second signature is very common in cancer and generally reflects apolipoprotein B mRNA-editing enzyme catalytic subunit (APOBEC) cytidine deaminase activity[64,67].

These genomic studies indicate that BCP-ALL has very restricted but informative mutational signatures and a low level of background or neutral genetic alterations[62]. This makes it less likely that BCP-ALL is caused by genotoxic exposures, which generally precipitate more widespread genomic instability with multiple distinctive signatures[64].

The other genetic subtype of BCP-ALL — hyperdiploid ALL — also has recurrent CNAs that may be RAG-mediated in genes including *PAX5*, *IKZF1* and *ETV6*. In contrast to *ETV6–RUNX1*+ ALL, however, hyperdiploid ALL has recurrent mutations in receptor tyrosine kinase (RTK)–RAS pathways and histone modifiers[61,63].

Collectively, these data provide a firm cellular, genetic and mechanistic framework for the two-step model for BCP-ALL and highlight both critical time windows, prenatally and postnatally, and mutational mechanisms. Any proposed causative mechanism should accommodate this natural history profile. The initiating role of *ETV6–RUNX1* and the postulated sequence of events in BCP-ALL are endorsed by modelling within both human and animal cells (BOX 2 and below). The timing and tissue site of BCP-ALL initiated by *ETV6–RUNX1* or hyperdiploidy in utero is uncertain but may involve transformation of a unique, fetal liver progenitor cell (BOX 3).

Inherited susceptibility

Childhood ALL only very rarely runs in families, but this observation may underplay inherited genetic risk because the disease itself is rare. Twin concordance is unhelpful in this respect because the risk has a mostly non-genetic basis: blood cell chimerism in utero. The risk of ALL in non-identical twins is unknown, but sibling risk has been estimated to be ~3.0 times higher

Box 2 | **Insights from modelling**

The initiating role of ETS translocation variant 6 (*ETV6*)–runt-related transcription factor 1 (*RUNX1*) in acute lymphoblastic leukaemia (ALL) has been modelled in zebrafish[154] and mice[155–162] and with human cells[41]. A consensus view is that the ETV6– RUNX1 protein functions as a weak oncoprotein, imparting only a modest proliferative or survival advantage to pro-B or pre-B cells. This accords with the small clone size of *ETV6–RUNX1*$^{+}$ pre-leukaemic cells in cord blood and in the blood of co-twins of patients with ALL[50]. *ETV6–RUNX1*$^{+}$ ALL cells ectopically express erythropoietin (EPO) receptors, and modelling with both human and mouse cells suggests that the receptors are functional and that EPO provides a survival signal to *ETV6–RUNX1*-expressing pre-leukaemic cells[163,164].

The observations in identical twins and cord blood suggest that *ETV6–RUNX1* fusion is insufficient by itself to induce overt or clinically evident ALL. This is supported by various models (see the table). However, as anticipated, *ETV6–RUNX1*-expressing clones in mice do progress to B cell precursor ALL (BCP-ALL) if additional, cooperative mutations are introduced by cross-breeding with mice heterozygous or homozygous null for the genes recurrently deleted in *ETV6–RUNX1*$^{+}$ ALL, including *Cdkn2a* and *Pax5* (REFS[161,162]). ALL also develops in some of these models if mice expressing *ETV6–RUNX1* are subjected to insertional mutagenesis[160] or chemical exposure[159].

The impact of these additional genetic lesions that complement *ETV6–RUNX1* is to impede or block differentiation, mirroring the early B cell lineage phenotypes observed in the clinical disease. This is in accord with the normal function of the transcription factors ETV6 and RUNX1, which is primarily to promote differentiation, and with their loss of function or deletion in ALL[63]. This interpretation is further endorsed by modelling studies in which experimental regulation of *Pax5* expression and gene dosage govern both early B cell lineage differentiation and leukaemogenesis[165].

These murine models support the 'two-hit' hypothesis for BCP-ALL and have also provided evidence for the role of common infections in the development of BCP-ALL (see the table)[134–136].

Initiating, transgenic lesion	Manipulation	Outcome	Refs
ETV6–RUNX1	None	Covert pre-leukaemia only	125,157
ETV6–RUNX1	Activation of AID, in vitro, with LPS; transplant to *Rag1*$^{+/+}$ or *Rag1*$^{-/-}$ mice	BCP-ALL in *Rag1*$^{+/+}$ background	126
ETV6–RUNX1	Switch to non-germ-free housing	BCP-ALL	135
Pax5$^{+/-}$	Switch to non-germ-free housing	BCP-ALL	134
Eμ-Ret or *E2A–PBX1*	CpG ODNs (TLR9 ligands) at 4–8 weeks	Delay and reduced penetrance of BCP-ALL	136

AID, activation-induced cytidine deaminase; CpG ODNs, oligodeoxynucleotides containing CpG motifs; LPS, lipopolysaccharide; TLR9, Toll-like receptor 9.

Box 3 | **Is there a unique fetal target cell for childhood B cell precursor ALL?**

Adults do develop B cell precursor acute lymphoblastic leukaemia (BCP-ALL), but it is striking that hyperdiploid ALL and ETS translocation variant 6 (*ETV6*)–runt-related transcription factor 1 (*RUNX1*)$^+$ BCP-ALL decline in incidence in teenage years and are infrequent after 20 years of age. Adult BCP-ALL has a higher frequency of *BCR–ABL1* fusion or of a *BCR–ABL1*-like signalling phenotype, perhaps explaining the poorer prognosis of adult BCP-ALL[166]. Childhood ALL in this respect appears to be a different cancer. Why should *ETV6–RUNX1* or hyperdiploidy initiate BCP-ALL preferentially in early life and especially in utero? *ETV6–RUNX1* expression, driven by the endogenous *Etv6* promoter in mice, increases quiescence and persistence of fetal liver stem cells[159]. Moreover, in this mouse model, fetal haemopoietic stem cells expressing the ETV6– RUNX1 protein had a protracted but still limited lifespan as self-renewing cells and did not contribute to the early B cell progenitor pool of adults. This suggests that the low penetrance of pre-leukaemia in children and the declining incidence of *ETV6–RUNX1*$^+$ (and hyperdiploid) ALL with age is, at least in part, due to natural attrition of the fetal pre-leukaemic clone.

One possibility is that fetal B lymphopoiesis is distinctive, in cell type and/or microenvironment, and provides a selective context in which *ETV6–RUNX1* and hyperdiploidy have effective transforming functions. This appears to be the explanation for *GATA1* mutations in Downs syndrome-associated acute megakaryoblastic leukaemia being restricted to fetal liver[167]. Recently, a strong candidate target cell for childhood BCP-ALL has been identified[168]. This cell type is developmentally restricted to fetal lymphopoiesis, with a mixed myeloid and B cell lineage phenotype. Similar to the putative target cell suggested by our studies in identical twins[34,41], this fetal liver myeloid and B cell progenitor also has DJ rearrangements of *IGH*[168].

than the risk in the general population, which provides evidence of a modest but definite contribution of constitutive, genetic variation to risk[68].

Earlier, targeted gene screening approaches suggested that inherited allelic variants encoding proteins involved in DNA repair, carcinogen metabolism or the folic acid pathway are linked to risk of childhood ALL[69,70]. Unfortunately, most of these studies were under-powered to detect small effects or have not been consistently replicated, so their significance remains uncertain[70].

Genome-wide association studies (GWAS) have provided unambiguous evidence for multiple gene variants that affect the risk of ALL[70,71] (TABLE 1). The individual alleles described to date have a significant but somewhat modest impact (see odds ratio in TABLE 1) and appear to be functionally additive rather than synergistic. The functional logic of these associations is unclear, but as most of the relevant single nucleotide polymorphisms (SNPs) lie outside coding regions, they are likely to be regulatory, affecting levels of proteins[72,73].

It is striking that in ALL, as in many other cancers, most of the candidate risk genes implicated in GWAS (TABLE 1) are the same genes that have acquired (non-constitutive) mutations in the same cancer type. One interpretation of this is that the inherited allelic variants interact functionally (or epistatically) with the mutated alleles to increase vulnerability of cells to transformation. A low functioning inherited allele, for example, would render a deletion in the other allele functionally homozygous, with a potentially increased impact on cellular fitness. A prediction that follows from this is that there should be a preferential loss, by deletion, of the non-risk allele (in heterozygotes for that allele), as only that deletion would increase clonal fitness. Evidence for this has been presented with respect to risk variants of *CDKN2A*[74,75]. However, for *ARID5B*, there is preferential gain of the risk allele (via trisomy 10) in heterozygotes[72].

To date, GWAS have provided no evidence implicating immune response gene variants, as might have been anticipated from infection-based hypotheses for the aetiology of ALL. However, previous studies examining major histocompatibility complex (MHC, also known as HLA) genes[76–78], interferon-γ (*IFNG*)[79], Toll-like receptor 6 (*TLR6*)[80] or the presence and/or absence of specific killer-cell immunoglobulin-like receptor (KIR) family genes[81] did record significant associations with particular allelic variants. Notably, *TLR6* variants and KIR genes were associated with decreased risks of all childhood ALL types. It remains unclear whether the large GWAS multi-cohort studies invalidate these data or whether the SNP screens in GWAS adequately detect the relevant variants. This is an important discrepancy to resolve.

Childhood ALL can also arise in a rare familial syndrome context, with inherited mutations in genes also implicated as acquired mutations in leukaemia, including *PAX5* (REF.[82]) and *ETV6* (REF.[83]). The relatively infrequent low hypodiploid subset of BCP-ALL is strongly associated with inherited *TP53* mutations or Li Fraumeni syndrome[84]. Further rare risk alleles, but with intermediate-to-high penetrance, are likely to be uncovered in ongoing, large-scale studies. Children with Downs syndrome have an approximately 20-fold to 30-fold increased risk of BCP-ALL[85]. Trisomy 21 in Downs syndrome is associated with overexpression of the nucleosome-remodelling protein high mobility group nucleosome-binding domain-containing protein 1 (HMGN1) and enhanced self-renewal of B cell progenitors[86]. In patients with ALL, this is complimented by secondary genetic changes, including those in *CRLF2*, *JAK2, NRAS* and *KRAS*[85]. All told, however, familial syndromes and Downs-associated ALL are likely to account for only a small fraction of patients diagnosed with childhood ALL.

The general conclusion to be drawn from these genetic studies is that inherited susceptibility does contribute to risk of BCP-ALL. The attributable risk or quantitative contribution is, however, unclear. A sibling risk of threefold for ALL, seen against a background risk of 1:2,000, suggests that the genetic component, though real, is minor, at least compared with some other common adults cancers (prostate and breast)[87]. On the other hand, there could be a complex interplay between genes

Table 1 | **Inherited alleles and risk of childhood B cell precursor ALL from genome-wide association studies[a]**

Candidate genes (chromosome)[b]	ALL subset bias	OR[c] (95% CI)	Refs
IKZF1 (7p12.2)	None detected	1.69 (1.58–1.81)	190
		1.69 (1.40–1.90)	191
ARID5B (10q21.2)	Hyperdiploid	1.65 (1.54–1.36)	190
		1.91 (1.60–2.20)	191
CDKN2A and *CDKN2B* (9p21.3)	None detected	1.36 (1.18–1.56)	75
		1.72 (1.50–1.97)	192
		2.41 (1.99–2.92)[b]	193
CEBPE (14q11.2)	Hyperdiploid	1.34 (1.22–1.45)	190
		1.45 (1.30–1.62)	194
GATA3 (10p14)	*BCR–ABL1*-like	1.31 (1.21–1.41)	195
		3.85 (2.70–5.40)	196
PIP4K2A (10p12.2)	None detected	1.40 (1.40–1.53)	75,195
TP63 (3q28)	*ETV6–RUNX1*	0.63 (0.52–0.75)	197
PTPRJ (11p11.2)	*ETV6–RUNX1*	0.72 (0.68–0.89)	197
LHPP (10q26.13)	None detected	1.20 (1.15–1.28)	198
ELK3 (12q23.1)	None detected	1.19 (1.12–1.26)	198
Unknown[d] (8q24.1)	None detected	1.34 (1.21–1.47)	199,200
Unknown (2q22.3)	*ETV6–RUNX1*	1.32 (1.17–1.49)	199
IKZF3 (17q12)	None detected	1.18 (1.11–1.25)	200

ALL, acute lymphoblastic leukaemia; NA, not available. [a]Cohort sizes in these studies were between ~400 and 3,500 patients. All patients were white and of European descent, in which the alleles exist at a relatively high frequency (10–50%). [b]More low risk alleles are likely to be uncovered as larger cooperative group studies are pursued and other populations are considered. Rare allelic variants, missed in previous studies, may also be identified that have a stronger impact on risk of ALL as described for *CDKN2A* (OR 2.4)[193]. [c]*P* values in these studies were between 10^{-5} and 10^{-19}. [d]The locus at 8q24.1 may regulate expression of *MYC*[199]

and environmental exposures[73,88] in which genetic background makes a more substantial difference. This has yet to be fully explored.

Possible causes of initiating events

No epidemiological studies to date have clearly implicated exposures during pregnancy, linked to risk of ALL, that might explain how the initiating mutations for BCP-ALL arise. *ETV6–RUNX1*$^{+}$ BCP-ALL has no mutational signatures that might implicate any particular type of aetiological exposure. There is no evidence for RAG involvement, but in common with *IGH* rearrangements[89], the recombination event appears to involve non-homologous end joining via microhomologies[27].

If ~1% of unselected newborn babies have an in-frame *ETV6–RUNX1* fusion in an expanded clone derived from the appropriate cell type for BCP-ALL (B cell lineage progenitor), then considerably more newborn babies should have acquired this or other fusion genes in the wrong cell types or out of frame for a functional protein. It therefore seems likely that whatever causes this genetic alteration could be very common or possibly ubiquitous.

The original proposition[17] was that BCP-ALL was initiated in utero by a spontaneous mutation or with no external exposure involvement. Developmental, endogenous factors such as proliferative and oxidative stress or the profound apoptotic signalling in early lymphopoiesis could be involved. Spontaneous mutations or mutations caused by endogenous processes are common during fetal development[90]. Endogenously driven double-strand breaks, required for fusion gene recombinants, occur at ~50 per cell cycle in human cells[91]. It has been suggested that most paediatric cancers arise during embryonic or fetal life and can similarly be considered as developmental errors[92]. In the absence of evidence to the contrary, this remains the most plausible explanation for initiation of BCP-ALL and focuses attention on the postnatal triggering of promotional events, which are required for clinical disease. There is clearly scope for more research on the origins and mechanisms involved in fusion gene formation, and hyperdiploidy, in utero.

Epidemiological evidence

Epidemiological evidence suggests that patterns of infection after birth have a causal role in triggering ALL. The delayed infection hypothesis lends itself to epidemiological evaluation in a case–control setting. A prediction of the model was that common infections in infancy should be protective against BCP-ALL. There is no prior reason to implicate any particular infectious agent (for example, bacteria, virus or parasite), and the relevant infections need not be symptomatic or pathological. A longstanding need for microbial, immune network modulation might reflect common, commensal, or 'old friend' organisms, such as gut microbiota, soil mycobacteria or helminth parasites[93]. In this context, a surrogate of overall infectious exposure during infancy could be considered an appropriate variable. Quantifiable surrogates include

Non-homologous end joining via microhomologies
A common (error-prone) recombination method that cells use to repair double-stranded DNA breaks without a sequence template that often involves the use of microhomologies of a few bases in pairing DNA strands.

social exposures of infants in the home, related to the number of siblings and birth order, or in day care centres, and breastfeeding. These variables have been investigated in epidemiological case–control or cohort studies for risk associations with ALL overall and, in some instances, selectively for the major BCP-ALL subset.

Impact of day care attendance in infancy. In the 1990s, the UK Children's Cancer Study Group (UKCCS) was set up to test the delayed infection hypothesis, in a case–control context, in addition to analysing other exposures including ionizing and non-ionizing (such as that from electromagnetic fields) radiation and chemicals[8]. Day care attendance was chosen as one surrogate for infectious exposure because this was well documented as a context for increased social contacts facilitating spread of common infections[94]. The UKCCS involved almost 1,300 patients with ALL (all subtypes) and over 6,000 matched controls. Although only a relatively small number of controls experienced day care in the first 12 months of life, the data showed a significant protective impact on risk of ALL overall and on BCP-ALL[95].

This association has been documented in additional studies in California[96], Scandinavia[97] and France[98] and in an international consortium[99], and has been endorsed by a meta-analysis[100]. The meta-analysis noted significant between study heterogeneity, and one early, large-scale study failed to detect any impact of day care on risk of ALL[101]. No protection is afforded against childhood AML by day care attendance or, to date, any other paediatric cancer, which increases the confidence that the associations seen in ALL were not confounded by social or other variables.

It was anticipated that assessing, in a case–control fashion, actual infections in infancy would be informative. This, however, is fraught with difficulties and has provided mixed results. Parental recall is known to be suspect or inaccurate in this respect[102]. Medical records are more reliable, particularly in the UK, with nationwide registration of children with general practitioners and a free health service. One such analysis found more, rather than fewer, infections reported for children who subsequently developed ALL than for controls[102]. The main difficulty here, other than possible bias in use of general practitioner services, is that we do not know whether the relevant modulating infectious exposures in infancy are necessarily symptomatic; they might well not be. In this sense, the surrogate of day care could be considered preferable. However, several studies have reported, in accord with the hypothesis, an inverse relationship between common infections in early life, including inner ear infections, and risk of ALL[98,101,103–106].

Birth order and risk of ALL. A further surrogate measure of infectious exposures in infancy is the number of siblings cohabiting and, in particular, birth order. The prediction was that firstborns would be more at risk than laterborns, who would, as infants, benefit from protective exposures via older siblings. One large UK-based study (with >3,000 patients with ALL and the same number of matched controls) found a striking association of birth order with risk of ALL, but not ofAML[107]. Other case–control studies in France[98] and California[100] also found a significantly increased risk of ALL for firstborn (versus thirdborn) children, as did a recent international cohort analysis (O. Paltiel, personal communication).

If natural infections early in life reduce risk of ALL, then it might be expected that some vaccinations would have an effect. The data on vaccination histories have produced null or inconsistent results. However, there is one exception: immunization against *Haemophilus influenzae* type B in infancy appears to confer a degree of protection against ALL[108].

If the natural microbiota is part of a longstanding and critical interaction with the developing immune system, then antibiotic use in infancy might increase risk of ALL. This has not been systematically evaluated to date, though an earlier report from China did suggest an increased risk associated with exposure to chloramphenicol[109].

Mode of delivery, breastfeeding and risk of ALL. The mode of delivery at birth influences the early exposure of newborn babies to benign microbiota[110], as caesarean delivery deprives newborn babies of the microbial exposures associated with vaginal passage. Cohort and case–control studies have reported a significantly increased risk of ALL associated with caesarean delivery[111–114]. No such increased risk was observed for brain cancer or lymphoma[111].

Breastfeeding during infancy provides nutritional support, maternal antibodies, anti-inflammatory molecules, some maternal cells, microorganisms (lactobacilli) and oligosaccharides that nourish the infant's intestinal microbiome (*Bifidobacteria* spp.)[115]. It might be anticipated that long-term breastfeeding would have a modulating effect on the immune system of infants and reduce the risk of ALL. Seventeen case–control studies of the impact of breastfeeding on ALL risk have been published[116]. In the largest of these, from the USA[117] and UK[118], there was a reduced risk of ALL (10–20%) associated with breastfeeding of 6 months or more. Five meta-analyses have now been published with concordant conclusions, and the latest of these indicated a reduced ALL risk of ~20% for breastfeeding of 6 months or more[116].

Clusters of ALL. Although no specific microbial agent or a unique transforming virus is suspected in ALL, there might be one circumstance where a single type of infection is involved: the very rare cases of space–time clusters. A prediction of the hypothesis would be that a single cluster of affected patients might be associated with a single infection or microorganism species, but independent, space–time clusters could involve different infectious triggers.

Many putative clusters of childhood leukaemia have been reported, but two stand out. The first was in Niles, a suburb of Chicago, Illinois, USA, in 1957–1960 — where there were eight patients (relative risk (RR) 4.3) diagnosed with ALL or 'stem cell' leukaemia[119]. All patients and/or their older siblings attended the same school. The second cluster involved 13 patients with BCP-ALL over 4 years (2000–2004; RR 12.0), but 10 of those patients were diagnosed within just 10 months in 2001, in the small town of Fallon, Nevada, USA[120].

Space–time clusters
A set of patients diagnosed in a short time frame and resident in the same, small area.

Relative risk
(RR). The calculated increase in diagnoses in a group of patients expressed, proportionally, in relation to that expected in the general, age-matched population.

A neglected aspect of these two clusters is that the patients, though resident in the cluster area at the time of diagnosis, were mostly born outside of that area[7]. Additionally, the clusters involved children diagnosed with ALL at different ages (2–11 years) and a narrow time frame of diagnoses. Given what we now know of the natural history of ALL, these data then indicate that any causal exposure linked to the cluster would have to be proximal in time to diagnosis (rather than prenatal) and, therefore, promotional. The Niles cluster was linked, observationally, to an outbreak of streptococcal fever[119]. The cause(s) of the Fallon cluster of ALL remain unresolved and contentious, though a possible role of adenovirus was hypothesized[120].

A significant space–time cluster of BCP-ALL in Milan, Italy has recently been recorded[121]. Seven patients were diagnosed in a 4-week period; four of these lived within one small residential area, and three of these four attended a single school. The Kulldorff scan method[122] identified this as a significant space–time cluster ($P=0.017$). Given the narrow time window of the diagnoses (4 weeks) and the age range of the patients (2–11 years), any causal exposure, as in Niles and Fallon, would be proximal to diagnosis, promoting overt ALL evolution from a prior and covert pre-leukaemic state. The Milan cases sparked substantial public anxiety, particularly in relation to the school, and a detailed epidemiological investigation was launched. No link was found with ionizing radiation, non-ionizing radiation or chemicals. There was, however, an association with a particular common infection. All seven patients had been infected with endemic influenza A H1N1 swine flu virus during the epidemic that preceded the ALL cluster by 3–6 months. The infection frequency in children in Milan during the same period was relatively high, at around one-third, but this still indicated that the link with patients was significantly different from expected ($P=0.01$)[121]. Six of the seven patients were firstborn children, and none attended day care in the first year of life.

Proof of a causal role for infections in these situations is not possible, and clustering of cases by chance cannot be excluded. But the observations accord with predictions of the infection hypothesis and highlight that influenza viruses are potential promoting agents for ALL. A previous study in the UK observed peaks in the incidence of childhood ALL ~6 months after seasonal influenza epidemics[123]. A final piece of epidemiological evidence indirectly supporting a role of common infection in childhood ALL comes from anecdotal but striking observations of rapid changes in the incidence of ALL that were preceded by major social changes in Germany and Hong Kong (Supplementary Box 2).

There is no compelling reason for postulating an exclusive role for influenza viruses or, indeed, for viruses. A role for cytomegalovirus (CMV) in ALL has been proposed but as an early, in utero modulator of immunity rather than as a proximal trigger[124].

In some respects, it is surprising that the epidemiological data are as consistent as they are for individual factors related to infection in ALL, because many variables will interact to influence patterns of microbial exposures in early life (Supplementary Box 3).

Modelling the missing link in ALL

There are limits to what epidemiological studies can achieve and to the robustness of the findings. Nevertheless, the associations described are compatible with the infection model proposed, and their selectivity for BCP-ALL versus AML is striking. But associations are not necessarily causal. Functional components of the infection hypothesis are best addressed by modelling studies in mice (BOX 2). These have proved very informative. One inflammatory cytokine — transforming growth factor-β (TGFβ) — was found to induce preferential expansion of *ETV6–RUNX1*-driven pre-leukaemic cells of both mouse and human origin[125]. Normal B cell precursor proliferation is inhibited by TGFβ via activation of the cell cycle (cyclin-dependent kinase) inhibitor p27. *ETV6–RUNX1* blocks this activity, giving pre-leukaemic cells a fitness advantage in the presence of TGFβ[125].

The missing link in the chain of events between infection, inflammatory responses and promotion of BCP-ALL may be activation-induced cytidine deaminase (AID; also known as AICDA)[126,127]. As noted above, genomic sequencing in BCP-ALL revealed that the recurrent CNAs have signatures of RAG activity[62]. Physiological RAG activity in germinal centre immunoglobulin class switching or hypermutation requires AID[128], as does illegitimate recombination between the *IGH* locus and oncogenes[129,130]. In B cell precursors, AID is not normally co-expressed with RAGs but is inducible by infection-driven cytokine signals[131]. This suggested that one route by which infection or chronic inflammation triggers RAG-mediated CNAs and ALL is via activation of AID expression in pre-leukaemic stem cells.

A mouse model has tested the requirement for RAGs and AID in the transition from *ETV6–RUNX1* pre-leukaemia to overt ALL[126]. The data revealed that lentiviral transfection of *ETV6–RUNX1* into progenitor cells leads to BCP-ALL when those cells are treated with a surrogate inflammatory signal (bacterial lipopolysaccharide binding to TLR4) that activates AID. Critically, mice did not develop ALL if the same experiment was conducted in a *Rag1*$^{-/-}$ genetic background. More recently, my colleagues and I have screened a series of inflammatory cytokines for their ability to trigger AID expression in human B cell precursors. The most potent was TGFβ (V. Cazzaniga, A. M. Ford and M. Greaves, unpublished observations). TGFβ is known to promote other cancers, often in the context of chronic inflammation[132]. In ALL, its role may include not only selective expansion of pre-leukaemic cells[125] and activation of AID but compromise of natural killer cell-based immune-surveillance[133].

If the aetiological hypothesis is correct, then it should be possible to influence risk or penetrance of ALL in murine models by timed exposure of the immune system to natural infections. Using a model of BCP-ALL, it was shown that ALL developed if *Pax5*$^{+/-}$ mice were switched from a germ-free environment to one providing exposure to common microbial pathogens[134]. Similarly, another study found that ~10% of

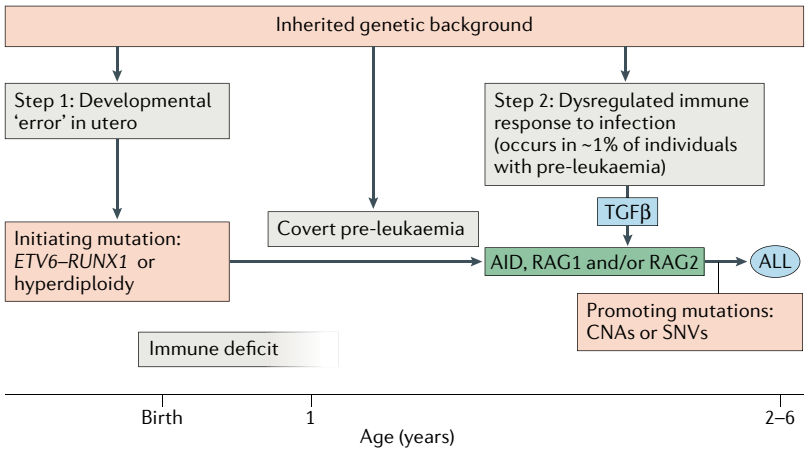

Fig. 3 | **Summary of the two-hit model for role of infections in B cell precursor ALL.** Genetic, inherited risk alleles are depicted (top of figure) as having effects at any or several stages of the stepwise process of acute lymphoblastic leukaemia (ALL) development. Step 1 is the prenatal initiation lesion (ETS translocation variant 6 (*ETV6*)–runt-related transcription factor 1 (*RUNX1*) or hyperdiploidy), which is common (~100 times clinical ALL frequency) and postulated to arise as a spontaneous, developmental error. This generates a clinically silent pre-leukaemic clone that can persist for up to 14 years. Step 2 is that in a small fraction (~1%) of patients with a covert pre-leukaemic clone, an abnormal immune response to one or more common infections triggers (probably via transforming growth factor-β (TGFβ) and possibly other cytokines) activation-induced cytidine deaminase (AID) activation, which, in combination with V(D)J recombination-activating protein 1 (RAG1) and/or RAG2, induces secondary genetic changes (predominantly copy number alterations (CNAs)). This occurs in patients who carry a covert pre-leukaemic clone and have a deficit of infectious exposures in infancy. The postulated immune deficit in infancy may increase the risk of Step 2 either by failure of immune network modulation and/or by affecting the persistence of a pre-leukaemic clone. SNV, single nucleotide variant.

mice with *ETV6–RUNX1* developed BCP-ALL after exposure to common pathogens[135]. These experiments provide evidence, albeit in murine models, that common infections can have, as predicted, a promotional role in ALL.

Another mouse model has provided evidence that early stimulation of the immune system can be protective. Exposure of mice with transgenic *Eμ-Ret* or *E2A* (also known as *TCF3*)–*PBX1* to oligodeoxynucleotides (which bind to TLR9) at 4 weeks depleted both normal and pre-leukaemic precursors and both delayed and diminished the risk of progression to ALL[136]. This effect was dependent upon IFNγ. By contrast, binding of polyinosinic:polycytidylic acid (poly(I:C)), a TLR3 ligand that does not induce IFNγ, resulted in an expansion of the pre-leukaemic cell pool. These data hint that the nature of infectious exposures in infancy and responses of the innate immune system may influence not only subsequent immune responses but also the fate of pre-leukaemic cells.

Conclusions: paradoxes of progress

> *We incline on our evidence to the belief that the solution of the problem of leukaemia lies rather in some peculiar reaction to infection than in the existence of some specific infective agent.*
> F. J. Poynton, H. Thursfield and D. Paterson, Great Ormond Street Hospital for Sick Children, 1922 (REF.[137])

Collectively, the accumulated evidence derived from epidemiological studies, GWAS, genome sequencing, biological scrutiny of the natural history and molecular pathogenesis of BCP-ALL and mechanistic and modelling studies provide us with a more substantive and credible version of the original[7,17] two-hit model for childhood ALL, as summarized in FIG. 3. The model applies selectively to the common, B cell precursor subset of ALL, although the evidence is currently more compelling for the *ETV6–RUNX1*$^{+}$ subset of BCP-ALL than for the hyperdiploid subset. The rarer pro-B ALL in infants appears likely to have a different causation and molecular pathogenesis, as does childhood AML and childhood lymphoma. There are insufficient data for thymic or T cell precursor ALL (T-ALL) in this respect. Other causal associations in leukaemia and cancer in general might be revealed or strengthened by a focus on well-defined subtypes, as suggested for breast cancer[138].

The causal mechanism proposed here is multifactorial, involving patterns of infection, inherited genetics and other modulators of risk including chance and, probably, diet (BOX 4). It has a logical coherence[139] and is grounded in the fundamental biology of leukaemia and evolutionary logic of the immune system network functions. The central thesis posits BCP-ALL as a paradox of progress in developed societies contingent upon a mismatch between the historical or evolutionary programming of the immune system and contemporary lifestyles that restrain opportunities for early-life microbial exposures. Childhood ALL is probably not the only unanticipated, deleterious health consequence of diminished infectious exposure in infancy[93]. Similar epidemiological associations exist for Hodgkin lymphoma in young adults[140] as well as for childhood allergies and autoimmune disease[141] (Supplementary Box 4). In all these clinical situations, the common theme is that acquisition of common microbial infections in early life has an impact on later responses of the immune system to challenge and the subsequent presence or absence of pathology[93,141,142]. Diminished exposure early in life to microorganisms that are pathological has been highly beneficial, reducing infant mortality, but it seems plausible that a suite of illnesses prevalent now in young people in more developed societies, including BCP-ALL, could be due to an unanticipated consequence of this advance[93,141].

The infection hypothesis would benefit from further scrutiny, including validation and extension of the animal modelling, but its public health implication is clear. Most cases of childhood ALL are potentially preventable. But how? Lifestyle changes including day care attendance or protracted breastfeeding in the first year of life can be advocated but would be difficult to achieve. A more realistic prospect might be to design a prophylactic vaccine that mimics the protective impact of natural infections in infancy, correcting the deficit in modern societies. Reconstitution or manipulation of the natural microbiome[143–146] or helminth injections[147,148] are strategies under consideration for early-life immune disorders

Box 4 | **Other modulators of risk in childhood ALL**

In addition to patterns of infectious exposure and inherited genetics, other factors are likely to contribute to multifactorial risk, including diet and chance. For acute lymphoblastic leukaemia (ALL) as well as acute myeloid leukaemia (AML) and most other paediatric cancers, risk is significantly and consistently elevated in association with higher birthweights[169,170] or, possibly, accelerated fetal growth[171]. A plausible interpretation of this link is that higher weight, possibly orchestrated via insulin-like growth factor 1 (IGF1) levels[172,173], may provide a greater number of cells at risk. IGF1 potentiates expansion of B cell lineage progenitors[174]. Recently, evidence has been presented, using mouse models of ALL, that a restricted diet can have a risk-reducing impact[175]. Intermittent fasting was shown to block expansion of leukaemic cell populations and progression of disease. The effect operated via attenuation of leptin receptor expression on leukaemic cells, possibly enforcing differentiation. Diet or calorie intake may, therefore, have a modulating impact on risk of ALL, reinforcing the likely multifunctional nature of causation of ALL, as in cancer in general.

Random events or chance get short shrift in cancer epidemiology, but it has long been recognized that contingency and chance pervades all of biology[176]. Some posit that a substantial number of cancers are due to chance alone[177], but this has been contentious. Chance is likely to be an ingredient in each and every cancer, including childhood ALL. This is because inheritance of risk alleles is a lottery at conception, because exposures including infections, at particular times, may or may not happen and because mutational mechanisms alter genes independently of their function[178].

in modern societies, including autoimmune and allergic conditions. Oral administration of benign synbiotics (bacteria species such as *Lactobacillus* spp. and oligosaccharides) can have profound and beneficial modulating effects on the developing immune system[149]. The results of those endeavours might inform approaches for preventing BCP-ALL. Cross collaboration of scientists working in disparate fields of early-life immune dysfunction — allergy, autoimmune disease and ALL — would be beneficial.

Published online 21 May 2018

1. Parkin, D. M. et al. (eds). *International incidence of childhood cancer* (IARC Scientific Publications, Lyon, 1988).
2. Pinkel, D. in *White Blood. Personal journeys with childhood leukaemia* (ed. Greaves, M.) 13–46 (World Scientific, Singapore, 2008).
3. Inaba, H., Greaves, M. & Mullighan, C. G. Acute lymphoblastic leukaemia. *Lancet* **381**, 1943–1955 (2013).
4. Essig, S. et al. Risk of late effects of treatment in children newly diagnosed with standard-risk acute lymphoblastic leukaemia: a report from the Childhood Cancer Survivor Study cohort. *Lancet Oncol.* **15**, 841–851 (2014).
5. Winther, J. F. & Schmiegelow, K. How safe is a standard-risk child with ALL? *Lancet Oncol.* **15**, 782–783 (2014).
6. Bhatia, S. & Robison, L. L. in *Hematology and Oncology of Infancy and Childhood* (eds Orkin, S. H. et al.) 1239–1256 (Elsevier, Philadelphia, 2015).
7. Greaves, M. Infection, immune responses and the aetiology of childhood leukaemia. *Nat. Rev. Cancer* **6**, 193–203 (2006).
Before the current Review, this was the most recent comprehensive review of the delayed infection hypothesis for BCP-ALL.
8. UK Childhood Cancer Study Investigators. The United Kingdom Childhood Cancer Study: objectives, materials and methods. *Br. J. Cancer* **82**, 1073–1102 (2000).
9. Metayer, C. et al. The Childhood Leukemia International Consortium. *Cancer Epidemiol.* **37**, 336–347 (2013).
10. Preston, D. L. et al. Cancer incidence in atomic bomb survivors. Part III: Leukemia, lymphoma and multiple myeloma, 1950–1987. *Radiat. Res.* **137** (Suppl.), S68–S97 (1994).
11. Doll, R. & Wakeford, R. Risk of childhood cancer from fetal irradiation. *Br. J. Radiol.* **70**, 130–139 (1997).
12. Bartley, K., Metayer, C., Selvin, S., Ducore, J. & Buffler, P. Diagnostic X-rays and risk of childhood leukaemia. *Int. J. Epidemiol.* **39**, 1628–1637 (2010).
13. Ward, G. The infective theory of acute leukaemia. *Br. J. Child Dis.* **14**, 10–20 (1917).
14. Schulz, T. F. & Neil, J. C. in *Leukemia* (eds Henderson, E. S., Lister, T. A. & Greaves, M. F.) 200–225 (Saunders, Philadelphia, 2002).
15. 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).
16. Kinlen, L. Childhood leukaemia, nuclear sites, and population mixing. *Br. J. Cancer* **104**, 12–18 (2011).
This is a review of the Kinlen population-mixing hypothesis for childhood leukaemia.
17. Greaves, M. F. Speculations on the cause of childhood acute lymphoblastic leukemia. *Leukemia* **2**, 120–125 (1988).
18. Torow, N. & Hornef, M. W. The neonatal window of opportunity: setting the stage for life-long host-microbial interaction and immune homeostasis. *J. Immunol.* **198**, 557–563 (2017).
19. Lim, E. S. et al. Early life dynamics of the human gut virome and bacterial microbiome in infants. *Nat. Med.* **21**, 1228–1234 (2015).
20. Olszak, T. et al. Microbial exposure during early life has persistent effects on natural killer T cell function. *Science* **336**, 489–493 (2012).
21. Biesbroek, G. et al. Early respiratory microbiota composition determines bacterial succession patterns and respiratory health in children. *Am. J. Respir. Crit. Care Med.* **190**, 1283–1292 (2014).
22. Arpaia, N. et al. Metabolites produced by commensal bacteria promote peripheral regulatory T cell generation. *Nature* **504**, 451–455 (2013).
23. Wills-Karp, M., Santeliz, J. & Karp, C. L. The germless theory of allergic disease: revisiting the hygiene hypothesis. *Nat. Rev. Immunol.* **1**, 69–75 (2001).
24. Greaves, M. Darwinian medicine: a case for cancer. *Nat. Rev. Cancer* **7**, 213–221 (2007).
25. Elinav, E. et al. Inflammation-induced cancer: crosstalk between tumours, immune cells and microorganisms. *Nat. Rev. Cancer* **13**, 759–771 (2013).
26. de Visser, K. E., Eichten, A. & Coussens, L. M. Paradoxical roles of the immune system during cancer development. *Nat. Rev. Cancer* **6**, 24–37 (2006).
27. Wiemels, J. L. & Greaves, M. Structure and possible mechanisms of *TEL-AML1* gene fusions in childhood acute lymphoblastic leukemia. *Cancer Res.* **59**, 4075–4082 (1999).
28. Rowley, J. D., Le Beau, M. M. & Rabbitts, T. H. (eds) *Chromosomal Translocations and Genome Rearrangements in Cancer* (Springer International Publishing, Switzerland, 2015).
29. Greaves, M. F., Maia, A. T., Wiemels, J. L. & Ford, A. M. Leukemia in twins: lessons in natural history. *Blood* **102**, 2321–2333 (2003).
30. Wolman, I. J. Parallel responses to chemotherapy in identical twin infants with concordant leukemia. *J. Pediatr.* **60**, 91–96 (1962).
31. Clarkson, B. & Boyse, E. A. Possible explanation of the high concordance for acute leukaemia in monozygotic twins. *Lancet* **1**, 699–701 (1971).
32. Strong, S. J. & Corney, G. *The Placenta in Twin Pregnancy* (Pergamon Press, Oxford, 1967).
33. Ford, A. M. et al. In utero rearrangements in the trithorax-related oncogene in infant leukaemias. *Nature* **363**, 358–360 (1993).
This is the first description of comparative genomics of ALL in twins with evidence for in utero origin.
34. Alpar, D. et al. Clonal origins of ETV6-RUNX1$^+$ acute lymphoblastic leukemia: studies in monozygotic twins. *Leukemia* **29**, 839–846 (2015).
35. Bateman, C. M. et al. Acquisition of genome-wide copy number alterations in monozygotic twins with acute lymphoblastic leukemia. *Blood* **115**, 3553–3558 (2010).
36. 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**, 5559–5565 (2011).
37. Ma, Y. et al. Developmental timing of mutations revealed by whole-genome sequencing of twins with acute lymphoblastic leukemia. *Proc. Natl Acad. Sci. USA* **110**, 7429–7433 (2013).
This study presents data from whole-genome sequencing of leukaemia in twin pairs.
38. Andersson, A. K. et al. The landscape of somatic mutations in infant MLL-rearranged acute lymphoblastic leukemias. *Nat. Genet.* **47**, 330–337 (2015).
39. Dobbins, S. E. et al. The silent mutational landscape of infant MLL-AF4 pro-B acute lymphoblastic leukemia. *Genes Chromosomes Cancer* **52**, 954–960 (2013).
40. Bateman, C. M. et al. Evolutionary trajectories of hyperdiploid ALL in monozygotic twins. *Leukemia* **29**, 58–65 (2015).
41. Hong, D. et al. Initiating and cancer-propagating cells in *TEL-AML1*-associated childhood leukemia. *Science* **319**, 336–339 (2008).
This study identifies putative pre-leukaemic stem cells in ALL.
42. Shlush, L. I. et al. Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. *Nature* **506**, 328–333 ()2014).
43. Jaiswal, S. et al. Age-related clonal hematopoiesis associated with adverse outcomes. *N. Engl. J. Med.* **371**, 2488–2498 (2014).
44. Gale, K. B. et al. Backtracking leukemia to birth: identification of clonotypic gene fusion sequences in neonatal blood spots. *Proc. Natl Acad. Sci. USA* **94**, 13950–13954 (1997).
This is the first study to identify leukaemia fusion genes in neonatal blood spots.
45. Wiemels, J. L. et al. Prenatal origin of acute lymphoblastic leukaemia in children. *Lancet* **354**, 1499–1503 (1999).
46. Hjalgrim, L. L. et al. Presence of clone-specific markers at birth in children with acute lymphoblastic leukaemia. *Br. J. Cancer* **87**, 994–999 (2002).
47. McHale, C. M. et al. Prenatal origin of *TEL-AML1*-positive acute lymphoblastic leukemia in children born

in California. *Genes Chromosomes Cancer* **37**, 36–43 (2003).
48. Wiemels, J. L., Ford, A. M., Van Wering, E. R., Postma, A. & Greaves, M. Protracted and variable latency of acute lymphoblastic leukemia after *TEL-AML1* gene fusion in utero. *Blood* **94**, 1057–1062 (1999).
49. Maia, A. T. et al. Protracted postnatal natural histories in childhood leukemia. *Genes Chromosomes Cancer* **39**, 335–340 (2004).
50. 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).
This study identifies a high rate of pre-leukaemic clone initiation before birth.
51. Lausten-Thomsen, U. et al. Prevalence of t(12;21) [ETV6-RUNX1]-positive cells in healthy neonates. *Blood* **117**, 186–189 (2011).
52. Zuna, J. et al. ETV6/RUNX1 (TEL/AML1) is a frequent prenatal first hit in childhood leukemia. *Blood* **117**, 368–369 (2011).
53. Skorvaga, M. et al. Incidence of common preleukemic gene fusions in umbilical cord blood in Slovak population. *PLoS ONE* **9**, e91116 (2014).
54. 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).
55. Beckwith, J. B. & Perrin, E. V. In situ neuroblastomas: a contribution to the natural history of neural crest tumors. *Am. J. Pathol.* **43**, 1089–1104 (1963).
56. Charles, A. K., Brown, K. W. & Berry, P. J. Microdissecting the genetic events in nephrogenic rests and Wilms' tumor development. *Am. J. Pathol.* **153**, 991–1000 (1998).
57. Yagi, T. et al. Detection of clonotypic IGH and TCR rearrangements in the neonatal blood spots of infants and children with B cell precursor acute lymphoblastic leukemia. *Blood* **96**, 264–268 (2000).
58. 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).
59. Fasching, K. et al. Presence of clone-specific antigen receptor gene rearrangements at birth indicates an in utero origin of diverse types of early childhood acute lymphoblastic leukemia. *Blood* **95**, 2722–2724 (2000).
60. Maia, A. T. et al. Identification of preleukemic precursors of hyperdiploid acute lymphoblastic leukemia in cord blood. *Genes Chromosomes Cancer* **40**, 38–43 (2004).
61. Paulsson, K. et al. The genomic landscape of high hyperdiploid childhood acute lymphoblastic leukemia. *Nat. Genet.* **47**, 672–676 (2015).
This study presents data on the genomics of the hyperdiploid subset of BCP-ALL.
62. 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 presents data on the genomics of *ETV6–RUNX1*$^{+}$ BCP-ALL.
63. Mullighan, C. G. et al. Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature* **446**, 758–764 (2007).
This study identifies recurrent CNAs in ALL.
64. Alexandrov, L. B. et al. Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
65. Anderson, K. et al. Genetic variegation of clonal architecture and propagating cells in leukaemia. *Nature* **469**, 356–361 (2011).
66. Waanders, E. et al. The origin and nature of tightly clustered *BTG1* deletions in precursor B cell acute lymphoblastic leukemia support a model of multiclonal evolution. *PLoS Genet.* **8**, e1002533 (2012).
67. Roberts, S. A. et al. An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nat. Genet.* **45**, 970–976 (2013).
68. Kharazmi, E. et al. Familial risks for childhood acute lymphocytic leukaemia in Sweden and Finland: far exceeding the effects of known germline variants. *Br. J. Haematol.* **159**, 585–588 (2012).
69. Sinnett, D., Krajinovic, M. & Labuda, D. Genetic susceptibility to childhood acute lymphoblastic leukemia. *Leuk. Lymphoma* **38**, 447–462 (2000).
70. Vijayakrishnan, J. & Houlston, R. S. Candidate gene association studies and risk of childhood acute lymphoblastic leukemia: a systematic review and meta-analysis. *Haematologica* **95**, 1405–1414 (2010).
71. Moriyama, T., Relling, M. V. & Yang, J. J. Inherited genetic variation in childhood acute lymphoblastic leukemia. *Blood* **125**, 3988–3995 (2015).
This is a review of GWAS studies in ALL.
72. Studd, J. B. et al. Genetic and regulatory mechanism of susceptibility to high-hyperdiploid acute lymphoblastic leukaemia at 10p21.2. *Nat. Commun.* **8**, 14616 (2017).
73. Sud, A., Kinnersley, B. & Houlston, R. S. Genome-wide association studies of cancer: current insights and future perspectives. *Nat. Rev. Cancer* **17**, 692–704 (2017).
74. Walsh, K. M. et al. A heritable missense polymorphism in CDKN2A confers strong risk of childhood acute lymphoblastic leukemia and is preferentially selected during clonal evolution. *Cancer Res.* **75**, 4884–4894 (2015).
75. Xu, H. et al. Inherited coding variants at the CDKN2A locus influence susceptibility to acute lymphoblastic leukaemia in children. *Nat. Commun.* **6**, 7553 (2015).
76. Taylor, G. M. et al. Strong association of the HLA-DP6 supertype with childhood leukaemia is due to a single allele, DPB1*0601. *Leukemia* **23**, 863–869 (2009).
77. Urayama, K. Y. et al. HLA-DP genetic variation, proxies for early life immune modulation and childhood acute lymphoblastic leukemia risk. *Blood* **120**, 3039–3047 (2012).
78. Urayama, K. Y. et al. SNP association mapping across the extended major histocompatibility complex and risk of B cell precursor acute lymphoblastic leukemia in children. *PLoS ONE* **8**, e72557 (2013).
79. Cloppenborg, T. et al. Immunosurveillance of childhood ALL: polymorphic interferon-gamma alleles are associated with age at diagnosis and clinical risk groups. *Leukemia* **19**, 44–48 (2005).
80. Miedema, K. G. et al. Polymorphisms in the TLR6 gene associated with the inverse association between childhood acute lymphoblastic leukemia and atopic disease. *Leukemia* **26**, 1203–1210 (2012).
81. Almalte, Z. et al. Novel associations between activating killer-cell immunoglobulin-like receptor genes and childhood leukemia. *Blood* **118**, 1323–1328 (2011).
82. Shah, S. et al. A recurrent germline PAX5 mutation confers susceptibility to pre-B cell acute lymphoblastic leukemia. *Nat. Genet.* **45**, 1226–1231 (2013).
83. 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).
84. Holmfeldt, L. et al. The genomic landscape of hypodiploid acute lymphoblastic leukemia. *Nat. Genet.* **45**, 242–252 (2013).
85. Lee, P., Bhansali, R., Izraeli, S., Hijiya, N. & Crispino, J. D. The biology, pathogenesis and clinical aspects of acute lymphoblastic leukemia in children with Down syndrome. *Leukemia* **30**, 1816–1823 (2016).
86. Lane, A. A. et al. Triplication of a 21q22 region contributes to B cell transformation through HMGN1 overexpression and loss of histone H3 Lys27 trimethylation. *Nat. Genet.* **46**, 618–623 (2014).
87. Lichtenstein, P. et al. Environmental and heritable factors in the causation of cancer. *N. Engl. J. Med.* **343**, 78–85 (2000).
88. Rudant, J. et al. ARID5B, IKZF1 and non-genetic factors in the etiology of childhood acute lymphoblastic leukemia: the ESCALE study. *PLoS ONE* **10**, e0121348 (2015).
89. Yan, C. T. et al. IgH class switching and translocations use a robust non-classical end-joining pathway. *Nature* **449**, 478–482 (2007).
90. Paashuis-Lew, Y. R. & Heddle, J. A. Spontaneous mutation during fetal development and post-natal growth. *Mutagenesis* **13**, 613–617 (1998).
91. Vilenchik, M. M. & Knudson, A. G. Endogenous DNA double-strand breaks: production, fidelity of repair, and induction of cancer. *Proc. Natl Acad. Sci. USA* **100**, 12871–12876 (2003).
92. Marshall, G. M. et al. The prenatal origins of cancer. *Nat. Rev. Cancer* **14**, 277–289 (2014).
93. Rook, G. A. W. (ed.) *The Hygiene Hypothesis and Darwinian Medicine* (Birkhäuser, Basel, 2009).
94. Goodman, R. A., Osterholm, M. T., Granoff, D. M. & Pickering, L. K. Infectious diseases and child day care. *Pediatrics* **74**, 134–139 (1984).
95. Gilham, C. et al. Day care in infancy and risk of childhood acute lymphoblastic leukaemia: findings from a UK case-control study. *Br. Med. J.* **330**, 1294–1297 (2005).
96. Ma, X. et al. Daycare attendance and risk of childhood acute lymphoblastic leukaemia. *Br. J. Cancer* **86**, 1419–1424 (2002).
97. Kamper-Jørgensen, M. et al. Childcare in the first 2 years of life reduces the risk of childhood acute lymphoblastic leukemia. *Leukemia* **22**, 189–193 (2007).
98. 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).
99. 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).
100. 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).
This is a meta-analysis of studies reporting the protective effect of day care attendance in infancy on total ALL and BCP-ALL occurrence.
101. Neglia, J. P. et al. Patterns of infection and day care utilization and risk factors of childhood acute lymphoblastic leukemia. *Br. J. Cancer* **82**, 234–240 (2000).
102. Simpson, J., Smith, A., Ansell, P. & Roman, E. Childhood leukaemia and infectious exposure: a report from the United Kingdom Childhood Cancer Study (UKCCS). *Eur. J. Cancer* **43**, 2396–2403 (2007).
103. Rudant, J. et al. Childhood acute leukemia, early common infections, and allergy: the ESCALE Study. *Am. J. Epidemiol.* **172**, 1015–1027 (2010).
104. Lin, J. N. et al. Risk of leukaemia in children infected with enterovirus: a nationwide, retrospective, population-based, Taiwanese-registry, cohort study. *Lancet Oncol.* **16**, 1335–1343 (2015).
105. Urayama, K. Y. et al. Early life exposure to infections and risk of childhood acute lymphoblastic leukemia. *Int. J. Cancer* **128**, 1632–1643 (2011).
106. Chan, L. C. et al. Is the timing of exposure to infection a major determinant of acute lymphoblastic leukaemia in Hong Kong? *Paediatr. Perinat. Epidemiol.* **16**, 154–165 (2002).
107. 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).
108. Ma, X. et al. Vaccination history and risk of childhood leukaemia. *Int. J. Epidemiol.* **34**, 1100–1109 (2005).
109. Shu, X. O. et al. Chloramphenicol use and childhood leukaemia in Shanghai. *Lancet* **2**, 934–937 (1987).
110. Dominguez-Bello, M. G. et al. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proc. Natl Acad. Sci. USA* **107**, 11971–11975 (2010).
111. Greenbaum, S. et al. Cesarean delivery and childhood malignancies: a single-center, population-based cohort study. *J. Pediatr*. https://doi.org/10.1016/j.jpeds.2017.12.049 (2018).
112. 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).
113. 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–e185 (2016).
114. Sevelsted, A., Stokholm, J., Bonnelykke, K. & Bisgaard, H. Cesarean section and chronic immune disorders. *Pediatrics* **135**, e92–e98 (2015).
115. Penders, J. et al. Factors influencing the composition of the intestinal microbiota in early infancy. *Pediatrics* **118**, 511–521 (2006).
116. Amitay, E. L. & Keinan-Boker, L. Breastfeeding and childhood leukemia incidence: a meta-analysis and systematic review. *JAMA Pediatr.* **169**, e151025 (2015).
This is a review of epidemiological evidence supporting a protective effect of protracted breastfeeding on ALL occurrence.
117. Shu, X. O. et al. Breast-feeding and risk of childhood acute leukemia. *J. Natl Cancer Inst.* **91**, 1765–1772 (1999).
118. UK Childhood Cancer Study Investigators. Breastfeeding and childhood cancer. *Br. J. Cancer* **85**, 1685–1694 (2001).
119. Heath, C. W. Jr & Hasterlik, R. J. Leukemia among children in a suburban community. *Am. J. Med.* **34**, 796–812 (1963).
120. 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).
This is a report of the most striking space–time cluster of childhood BCP-ALL to date.
121. 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 linking a space–time cluster of childhood BCP-ALL with a specific viral infection.
122. Kulldorff, M. & Nagarwalla, N. Spatial disease clusters: detection and inference. *Stat. Med.* **14**, 799–810 (1995).
123. Kroll, M. E., Draper, G. J., Stiller, C. A., & Murphy, M. F. G. Childhood leukemia incidence in Britain, 1974–2000: time trends and possible relation to influenza epidemics. *J. Natl Cancer Inst.* **98**, 417–420 (2006).
124. Francis, S. S. et al. In utero cytomegalovirus infection and development of childhood acute lymphoblastic leukemia. *Blood* **129**, 1680–1684 (2017).
125. Ford, A. M. et al. The *TEL-AML1* leukemia fusion gene dysregulates the TGFb pathway in early B lineage progenitor cells. *J. Clin. Invest.* **119**, 826–836 (2009).
126. Swaminathan, S. et al. Mechanisms of clonal evolution in childhood acute lymphoblastic leukemia. *Nat. Immunol.* **16**, 766–774 (2015).
This study models the link between infections and activation of secondary genetic changes in ALL via RAG and AID.
127. Olinski, R., Styczynski, J., Olinska, E. & Gackowski, D. Viral infection-oxidative stress/DNA damage-aberrant DNA methylation: separate or interrelated events responsible for genetic instability and childhood ALL development? *Biochim. Biophys. Acta* **1846**, 226–231 (2014).
128. de Yébenes, V. & Ramiro, A. R. Activation-induced deaminase: light and dark sides. *Trends Mol. Med.* **12**, 432–439 (2006).
129. Robbiani, D. F. et al. AID is required for the chromosomal breaks in c-myc that lead to c-myc/IgH translocations. *Cell* **135**, 1028–1038 (2008).
130. Chesi, M. et al. AID-dependent activation of a MYC transgene induces multiple myeloma in a conditional mouse model of post-germinal center malignancies. *Cancer Cell* **13**, 167–180 (2008).
131. Rosenberg, B. R. & Papavasiliou, F. N. Beyond SHM and CSR: AID and related cytidine deaminases in the host response to viral infection. *Adv. Immunol.* **94**, 215–244 (2007).
132. Pickup, M., Novitskiy, S. & Moses, H. L. The roles of TGFbeta in the tumour microenvironment. *Nat. Rev. Cancer* **13**, 788–799 (2013).
133. Rouce, R. H. et al. The TGF-beta/SMAD pathway is an important mechanism for NK cell immune evasion in childhood B-acute lymphoblastic leukemia. *Leukemia* **30**, 800–811 (2016).
134. 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).
This study demonstrates, in a mouse model, that common infections promote BCP-ALL.
135. Rodriguez-Hernandez, G. et al. Infection exposure promotes ETV6-RUNX1 precursor B cell leukemia via impaired H3K4 demethylases. *Cancer Res.* **77**, 4365–4377 (2017).
136. Fidanza, M. et al. Inhibition of precursor B cell malignancy progression by toll-like receptor ligand-induced immune responses. *Leukemia* **30**, 2116–2119 (2016).
137. Poynton, F. J., Thursfield, H. & Paterson, D. The severe blood diseases of childhood: a series of observations from the Hospital for Sick Children, Great Ormond Street. *Br. J. Child Dis.* **XIX**, 128–144 (1922).
138. Holm, J. et al. Assessment of breast cancer risk factors reveals subtype heterogeneity. *Cancer Res.* **77**, 3708–3717 (2017).
139. Thagard, P. *How Scientists Explain Disease* (Princeton Univ. Press, Princeton, 1999).
140. Cozen, W. et al. A protective role for early oral exposures in the etiology of young adult Hodgkin lymphoma. *Blood* **114**, 4014–4020 (2009).
141. Bach, J. F. The effect of infections on susceptibility to autoimmune and allergic diseases. *N. Engl. J. Med.* **347**, 911–920 (2002).
142. Dunne, D. W. & Cooke, A. A worm's eye view of the immune system: consequences for evolution of human autoimmune disease. *Nat. Rev. Immunol.* **5**, 420–426 (2005).
143. Parker, W. & Ollerton, J. Evolutionary biology and anthropology suggest biome reconstitution as a necessary approach toward dealing with immune disorders. *Evol. Med. Public Health* **2013**, 89–103 (2013).
144. Marino, E. et al. Gut microbial metabolites limit the frequency of autoimmune T cells and protect against type 1 diabetes. *Nat. Immunol.* **18**, 552–562 (2017).
145. Durack, J. et al. Delayed gut microbiota development in high-risk for asthma infants is temporarily modifiable by *Lactobacillus* supplementation. *Nat. Commun.* **9**, 707 (2018).
146. Trompette, A. et al. Gut microbiota metabolism of dietary fiber influences allergic airway disease and hematopoiesis. *Nat. Med.* **20**, 159–166 (2014).
147. Elliott, D. E. & Weinstock, J. V. Helminth-host immunological interactions: prevention and control of immune-mediated diseases. *Ann. NY Acad. Sci.* **1247**, 83–96 (2012).
148. Navarro, S. et al. Hookworm recombinant protein promotes regulatory T cell responses that suppress experimental asthma. *Sci. Transl Med.* **8**, 362ra143 (2016).
149. Panigrahi, P. et al. A randomized synbiotic trial to prevent sepsis among infants in rural India. *Nature* **548**, 407–412 (2017).
150. Anderson, R. M. & May, R. M. Immunisation and herd immunity. *Lancet* **335**, 641–645 (1990).
151. Bellec, S. et al. Childhood leukaemia and population movements in France, 1990–2003. *Br. J. Cancer* **98**, 225–231 (2008).
152. Stiller, C. A., Kroll, M. E., Boyle, P. J. & Feng, Z. Population mixing, socioeconomic status and incidence of childhood acute lymphoblastic leukaemia in England and Wales: analysis by census ward. *Br. J. Cancer* **98**, 1006–1011 (2008).
153. Dickinson, H. O., Hammal, D. M., Bithell, J. F. & Parker, L. Population mixing and childhood leukaemia and non-Hodgkin's lymphoma in census wards in England and Wales, 1966–1987. *Br. J. Cancer* **86**, 1411–1413 (2002).
154. Sabaawy, H. E. et al. TEL-AML1 transgenic zebrafish model of precursor B cell acute lymphoblastic leukemia. *Proc. Natl Acad. Sci. USA* **103**, 15166–15171 (2006).
155. Bernardin, F. et al. TEL-AML1, expressed from t(12;21) in human acute lymphocytic leukemia, induces acute leukemia in mice. *Cancer Res.* **62**, 3904–3908 (2002).
156. Morrow, M., Horton, S., Kioussis, D., Brady, H. J. M. & Williams, O. TEL-AML1 promotes development of specific hematopoietic lineages consistent with preleukemic activity. *Blood* **103**, 3890–3896 (2004).
157. Tsuzuki, S., Seto, M., Greaves, M. & Enver, T. Modelling first-hit functions of the t(12;21) *TEL-AML1* translocation in mice. *Proc. Natl Acad. Sci. USA* **101**, 8443–8448 (2004).
158. Fischer, M. et al. Defining the oncogenic function of the TEL/AML1 (*ETV6/RUNX1*) fusion protein in a mouse model. *Oncogene* **24**, 7579–7591 (2005).
159. 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).
160. van der Weyden, L. et al. Modeling the evolution of *ETV6-RUNX1*-induced B cell precursor acute lymphoblastic leukemia in mice. *Blood* **118**, 1041–1051 (2011).
161. van der Weyden, L. et al. Somatic drivers of B-ALL in a model of ETV6-RUNX1; Pax5$^{(+/-)}$ leukemia. *BMC Cancer* **15**, 585 (2015).
162. Li, M. et al. Initially disadvantaged, TEL-AML1 cells expand and initiate leukemia in response to irradiation and cooperating mutations. *Leukemia* **27**, 1570–1573 (2013).
163. Inthal, A. et al. Role of the erythropoietin receptor in ETV6/RUNX1-positive acute lymphoblastic leukemia. *Clin. Cancer Res.* **14**, 7196–7204 (2008).
164. Torrano, V., Procter, J., Cardus, P., Greaves, M. & Ford, A. M. ETV6-RUNX1 promotes survival of early B lineage progenitor cells via a dysregulated erythropoietin receptor. *Blood* **118**, 4910–4918 (2011).
165. Liu, G. J. et al. Pax5 loss imposes a reversible differentiation block in B-progenitor acute lymphoblastic leukemia. *Genes Dev.* **28**, 1337–1350 (2014).
166. Hunger, S. P. & Mullighan, C. G. Redefining ALL classification: toward detecting high-risk ALL and implementing precision medicine. *Blood* **125**, 3977–3987 (2015).
167. Li, Z. et al. Developmental stage-selective effect of somatically mutated leukemogenic transcription factor GATA1. *Nat. Genet.* **37**, 613–619 (2005).
168. 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.e7 (2018).
This study identifies putative fetal liver target cells for *ETV6–RUNX1*$^{+}$ BCP-ALL.
169. Hjalgrim, L. L. et al. Birth weight as a risk factor for childhood leukemia: a meta-analysis of 18 epidemiologic studies. *Am. J. Epidemiol.* **158**, 724–735 (2003).
170. Roman, E. et al. Childhood acute lymphoblastic leukaemia and birthweight: insights from a pooled analysis of case-control data from Germany, the United Kingdom and the United States. *Eur. J. Cancer* **49**, 1437–1447 (2013).
171. Milne, E. et al. Fetal growth and childhood acute lymphoblastic leukemia: findings from the childhood leukemia international consortium. *Int. J. Cancer* **133**, 2968–2979 (2013).
172. Chokkalingam, A. P. et al. Fetal growth and body size genes and risk of childhood acute lymphoblastic leukemia. *Cancer Causes Control* **23**, 1577–1585 (2012).
173. Tower, R. L. & Spector, L. G. The epidemiology of childhood leukemia with a focus on birth weight and diet. *Crit. Rev. Clin. Lab. Sci.* **44**, 203–242 (2007).
174. Gibson, L. F., Piktel, D. & Landreth, K. S. Insulin-like growth factor-1 potentiates expansion of interleukin-7-dependent pro-B cells. *Blood* **82**, 3005–3011 (1993).
175. Lu, Z. et al. Fasting selectively blocks development of acute lymphoblastic leukemia via leptin-receptor upregulation. *Nat. Med.* **23**, 79–90 (2017).
176. Monod, J. *Chance & Necessity* (Alfred A. Knopf, New York, 1970).
177. Tomasetti, C. & Vogelstein, B. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science* **347**, 78–81 (2015).
178. Greaves, M. Evolutionary determinants of cancer. *Cancer Discov.* **5**, 806–820 (2015).
179. Iacobucci, I. & Mullighan, C. G. Genetic basis of acute lymphoblastic leukemia. *J. Clin. Oncol.* **35**, 975–983 (2017).
180. Roberts, K. G. & Mullighan, C. G. Genomics in acute lymphoblastic leukaemia: insights and treatment implications. *Nat. Rev. Clin. Oncol.* **12**, 344–357 (2015).
181. Belver, L. & Ferrando, A. The genetics and mechanisms of T cell acute lymphoblastic leukaemia. *Nat. Rev. Cancer* **16**, 494–507 (2016).
182. 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).
183. Greaves, M. F., Janossy, G., Peto, J. & Kay, H. Immunologically defined subclasses of acute lymphoblastic leukaemia in children: their relationship to presentation features and prognosis. *Br. J. Haematol.* **48**, 179–197 (1981).
184. Shah, A. & Coleman, M. P. Increasing incidence of childhood leukaemia: a controversy re-examined. *Br. J. Cancer* **97**, 1009–1012 (2007).
185. Spix, C., Eletr, D., Blettner, M. & Kaatsch, P. Temporal trends in the incidence rate of childhood cancer in Germany 1987–2004. *Int. J. Cancer* **122**, 1859–1867 (2008).
186. Steliarova-Foucher, E. et al. Trends in childhood cancer incidence in Europe, 1970–1999. *Lancet* **365**, 2088 (2005).
187. Chessells, J. M., Hardisty, R. M., Rapson, N. T. & Greaves, M. F. Acute lymphoblastic leukaemia in children: classification and prognosis. *Lancet* **2**, 1307–1309 (1977).
188. Sallan, S. E. et al. Cell surface antigens: prognostic implications in childhood acute lymphoblastic leukemia. *Blood* **55**, 395–402 (1980).
189. Greaves, M. in *Nathan & Oski's Hematology & Oncology of Infancy & Childhood* (eds Orkin, S. H. et al.) 1229–1238 (Elsevier Saunders, Philadelphia, 2015).
190. 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).
191. Trevino, L. R. et al. Germline genomic variants associated with childhood acute lymphoblastic leukemia. *Nat. Genet.* **41**, 1001–1005 (2009).
192. Hungate, E. A. et al. A variant at 9p21.3 functionally implicates CDKN2B in paediatric B cell precursor acute lymphoblastic leukaemia aetiology. *Nat. Commun.* **7**, 10635 (2016).

193. Vijayakrishnan, J. et al. The 9p21.3 risk of childhood acute lymphoblastic leukaemia is explained by a rare high-impact variant in CDKN2A. *Sci. Rep.* **5**, 15065 (2015).
194. Wiemels, J. L. et al. A functional polymorphism in the CEBPE gene promoter influences acute lymphoblastic leukemia risk through interaction with the hematopoietic transcription factor Ikaros. *Leukemia* **30**, 1194–1197 (2016).
195. 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).
196. 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).
197. Ellinghaus, E. et al. Identification of germline susceptibility loci in ETV6-RUNX1-rearranged childhood acute lymphoblastic leukemia. *Leukemia* **26**, 902–909 (2012).
198. Vijayakrishnan, J. et al. A genome-wide association study identifies risk loci for childhood acute lymphoblastic leukemia at 10q26.13 and 12q23.1. *Leukemia* **31**, 573–579 (2017).
199. Vijayakrishnan, J. et al. Genome-wide association study identifies susceptibility loci for B-cell childhood acute lymphoblastic leukemia. *Nat. Commun.* **9**, 1340 (2018).
200. 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).

Acknowledgements
The author thanks the Leukaemia Research Fund UK (now Bloodwise), The Kay Kendall Leukaemia Fund, the Wellcome Trust [105104/Z/14/Z] and The Institute of Cancer Research, London, for long-term funding support. The author also thanks R. Houlston for constructive comments on the manuscript. The author is indebted to many students and postdoctoral fellows in his laboratory who have researched this topic over many years and to the many scientists and clinicians who have been excellent collaborators. These include A. Ford, J. Wiemels, A. T. Maia, L. C. Chan, C. Bateman, D. Alpar, I. Titley, S. Colman, H. Mori, V. Cazzaniga, K. Anderson, N. Potter, L. Kearney and R. Houlston (all at The Institute of Cancer Research, London), as well as A. Lister, T. Eden, J. Chessells, P. Ancliff, R. Jarrett, A. Borkhardt, P. Buffler, E. Roman, G. Cazzaniga, M. E. Cabrera, M. P. de Oliveira, A. Biondi, J. Kersey, M. Müschen, P. Campbell, E. Papaemmanuil and T. Enver.

Author contributions
M.G. researched data for the article and wrote, reviewed and edited the manuscript.

Competing interests
The author declares no competing interests.

Publisher's note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Reviewer information
Nature Reviews Cancer thanks K. Paulsson, S. Sallan and J. Wiemels for their contribution to the peer review of this work.

Supplementary information
Supplementary information is available for this paper at https://doi.org/10.1038/s41568-018-0015-6.