npg

MEETING REPORT

Research recommendations toward a better understanding of the causes of childhood leukemia

G Ziegelberger[1], C Baum[2], A Borkhardt[3], C Cobaleda[4], C Dasenbrock[5], A Dehos[1], B Grosche[1], J Hauer[3], S Hornhardt[1], T Jung[1], T Kammertoens[6], I Lagroye[7], H Lehrach[8], T Lightfoot[9], MP Little[10], C Rossig[11], I Sanchez-Garcia[12], M Schrappe[13], J Schuez[14], S Shalapour[6], R Slany[15], M Stanulla[13] and W Weiss[1]

[1]Department of Radiation and Health, Federal Office for Radiation Protection (BfS), Oberschleissheim, Germany; [2]Department of Experimental Hematology, Hannover Medical School, Hannover, Germany; [3]Department of Paediatric Oncology, Haematology and Clinical Immunology, Center for Child and Adolescent Health, Heinrich Heine University, Duesseldorf, Germany; [4]CBMSO, Spanish Research Council (CSIC), Universidad Autonoma, Madrid, Spain; [5]Fraunhofer Institute of Toxicology and Experimental Medicine (ITEM), Hannnover, Germany; [6]Institute of Immunology and Department of Paediatric Oncology and Haematology, Charite Universitaetsmedizin, Berlin, Germany; [7]Laboratoire IMS UMR 5218, University of Bordeaux, Bordeaux, France; [8]Max Planck Institute for Molecular Genetics, Berlin, Germany; [9]Department of Health Sciences, Epidemiology & Genetics Unit, University of York, York, UK; [10]Radiation Epidemiology Branch, National Cancer Institute, Rockville, MD, USA; [11]University Children's Hospital Muenster, Pediatric Hematology and Oncology, Muenster, Germany; [12]IBMCC, Spanish Research Council (CSIC), University of Salamanca, Salamanca, Spain; [13]Department of General Paediatrics, University Medical Center Schleswig-Holstein, Kiel, Germany; [14]Section of Environment and Radiation, International Agency for Research on Cancer (IARC), Lyon, France and [15]Department of Genetics, Friedrich Alexander University, Erlangen, Germany

A small expert group met in conclave in July 2010 to define a long-term strategic research agenda toward further clarification of the etiology of childhood leukemia (CL). The motivation and invitation for this project came from the German Office for Radiation Protection (BfS) because radiation experts have been puzzled for some time by epidemiological findings of an increased incidence of CL near German nuclear facilities,[1] as well as by a statistical association with exposure to residential low-frequency magnetic fields.[2] Both findings are difficult to explain given the current knowledge of the biological mechanisms of ionizing or non-ionizing radiation, as both types of exposure deposit far too little energy in cellular DNA and other likely targets to be considered directly causative. A previous workshop on risk factors held in May 2008[3] and a follow-up meeting led to the conclusion that understanding of the causes of CL development requires a broadened, interdisciplinary approach.

The discussion and work on a research agenda was focused on the main leukemia type in childhood, acute lympho-blastic leukemia (ALL), and especially B-cell precursor ALL. Based on short presentations by all participants and the views of experts from different disciplines (epidemiology, clinic, experimental modeling, theoretical modeling, molecular biology and genetics), the following key features and related *key questions* of childhood ALL have been developed.

Incidence rates and epidemiological findings

The reported increasing incidence rates seen for B-cell precursor ALL (but not for T-ALL or acute myelogenous leukemia (AML)) in industrialized countries[4] point toward a role for the modern lifestyle. However, to date, epidemiological studies have failed to confirm any associations with lifestyle exposures such as smoking, alcohol, diet, social status, and so on. Indeed, to date, the only recognized risk factors for ALL are heavy birth weight[5,6] and gender, with boys more often affected than girls (approximate ratio 1.2:1).—*However, the validity of these annual increases in industrialized countries is questionable. Are the observed increases partially due to improvements in cancer registration and disease detection and classification? What is the base line incidence world wide or, in other words, how large is the contribution of potentially avoidable environmental factors to the disease? Why are boys more likely to develop ALL?*

So far, epidemiological studies could not correlate various potential risk factors with the initial versus subsequent genetic events. A translocation event as the first hit has to be the consequence of a double-strand break that is misrepaired, whereas different risk factors reflecting different mechanisms are supposed to be responsible for the additional cooperating genetic changes.—*What does the age at onset of leukemia tell us? B-cell precursor ALL has a distinct age peak between 2–5 years of age, except for the subtype characterized by mixed lineage leukemia (MLL) rearrangements that occurs mainly below 1 year. What are the patterns of incidence of ALL and AML in developing countries—is there evidence of a peak between 2–5 years as seen in developed countries?*

Molecular genetics and inherited susceptibilities

Specific chromosomal translocations and DNA aneuploidies are frequent first hits, but further genetic or epigenetic events must cooperate to cause an outbreak of acute leukemia. Several independent genome-wide association studies (GWASs) of a large number of ALL patients have identified and confirmed

a handful of gene variants, namely IKZF1, ARIDB5, CEBPE and CDKN2A,[7–12] that affect genomic susceptibility to ALL, although they are individually modest in their effects. These gene variants are predominantly related to pathways controlling B-cell development and differentiation. How these germline variations influence the risk of ALL is unresolved.

Besides the already ongoing GWAS, a strong recommendation was to initiate next-generation sequencing (whole-genome or transcriptome sequences, exome capture and sequencing, analyses of the methylome) of ALL cases, as it is already being done for other cancer types (see projects of the International Cancer Genome Consortium). GWASs can reveal population-based variations, but deep sequencing can focus on detailed differences and mutations, for example, of tumor versus normal cells, which could, at least partly, reflect the effects of the environment. This could help to characterize subgroups that might not show up in an overall approach, or detect common patterns/footprints possibly correlated to external risk factors.[13] The previous finding that childhood ALL is the consequence of a limited number of genetic alterations[7] supports the feasibility of identifying the relevant hits.

Studying monozygotic twins, consanguineous siblings and other familial predisposition syndromes will add valuable information on the etiology of ALL, but due to the limited number of familial cases the latter cannot be studied systematically.[14]

The preleukemic clone

Studies of monozygotic twins and retrospective analysis of neonatal blood spots have found that in many CLs the first hit that converts a fetal hematopoietic precursor or stem cell to a preleukemic clone originates in utero.[15,16] One study found that the frequency of newborns carrying the leukemia-specific translocation TEL-AML1 in their blood is about 100 times higher than the overall incidence rate of ALL.[17] This implies that preleukemic clones are frequently and normally extinguished (or at least kept at bay) by natural processes. The preleukemic 'original' clones also seem to be of major importance for ALL relapse.[18] Whether the preleukemic clones originated in utero have also a role in adult leukemia is currently not known. A recent study published after the meeting[19] analyzed 1417 umbilical cord blood samples but did not confirm the findings by Mori et al., indicating that the frequency of TEL-AML1-positive cells may be markedly lower in healthy children than previously suggested. The prevalence of preleukemic clones in newborns needs to be verified, as it forms the base for many hypotheses and study designs. Once the frequency of preleukemic clones is established in an industrialized population, it must be compared with data from developing countries where less is known about the prevalence of these clones and of ALL incidence. This would allow us to test whether or not the observed increasing incidence rates in industrialized countries are correlated with an increase of initial events or rather of secondary events.—In other words: Is the first hit occurring naturally (at random) or is it already an effect of a factor that can be prevented? If the first hit occurs naturally, only the secondary genetic events would be relevant to explain the increasing incidence rates in developed countries. One critical hurdle is still the reliable detection of rare preleukemic clones.

Recent functional studies on fusion gene products suggest that those cells carrying fusion proteins might be more susceptible to secondary events. MLL fusion proteins interfere with the downregulation of HOX genes, leading to a differentiation block and immortalizing a precursor cell as target for subsequent genetic changes.[20] Similarly, the TEL–AML1 fusion protein increases the number of hematopoietic stem cells in the bone marrow as a latent reservoir for the accumulation of further genetic hits.[21]—What is the developmental step at which the preleukemic clone is arrested to accumulate secondary mutations? Answering this question may offer insight into ways to prevent tumorigenic progression.

Epidemiological and pediatric observations on the role of the immune system

The suggestion that the timing and pattern of infectious exposure in early life may be a determinant of childhood leukemia was mooted over 70 years ago.[22,23] More recent, however, has been the hypothesis that ALL could arise as a rare abnormal response to a common infection; two distinct and related hypotheses about how this might occur have been proposed by Kinlen and Greaves. Kinlen suggested that ALL could arise as a consequence of exposure to a specific infection, which is particularly evident at times of unusual population movements and mixing.[24,25] Greaves hypothesized that a deficit of exposure to infectious agents in infancy and subsequent 'delayed' infectious challenge may be causal in the development of B-cell precursor ALL in the peak ages of 2–5 years.[26] While the Greaves hypothesis is supported by studies that have looked at day-care attendance,[27] less support is derived from examination of primary care data revealing that cases who developed B-cell precursor ALL between 2–5 years of age had significantly more (not less) clinically diagnosed infectious illness episodes in the first year of life compared with controls. This suggests that immune dysregulation in children who develop ALL may be detectable several years before diagnosis.[28,29] However, findings on the role of infections depend strongly on the infectious proxies that have been employed (for example, social contact of parents and child, socioeconomic status, maternally reported infections versus medical records, day care attendance), and suggest an important role for the immune system in the pathogenesis of B-cell precursor ALL.—Does the infection-driven inflammation lead to tumor development and promotion? In this respect, a potential protective effect of atopic disease as observed in epidemiological studies[30] needs further considerations and might be a key in our understanding of the role of the immune system in both diseases.

Are preleukemic clones affected by virus-induced (auto-) immune reactions? Evidence for either direct or indirect immune-mediated rejection of a preleukemic clone has been observed in a healthy monozygotic twin of a child with MLL-rearranged ALL. The child lost the preleukemic cell population during transient, likely immune-mediated cytopenia.[31]—Can this be generalized and, if yes, by what mechanisms are virus infections involved in the clearance of the preleukemic clone? Or, alternatively, can infections trigger secondary events or enhance their potency to drive preleukemic cells toward leukemia? What are the conditions that trigger enhanced (re)generation of B cells and their precursors?

Cellular plasticity and the role of the microenvironment

Although cellular differentiation processes had long been considered unidirectional and irreversible, a set of experiments in mice uncovered a high developmental plasticity of

hematopoietic cells. The loss of a single transcription factor, Pax 5, allows mature B cells from peripheral lymphoid organs to de-differentiate *in vivo* back to uncommitted progenitors in the bone marrow. These de-differentiated progenitors were subsequently even able to develop into macrophages.[32] The high plasticity and lineage conversion imply that studying the disease at the onset of clinical leukemia does not allow to retrospectively reveal the individual history, for example, the initial first event. Thus, the nature of the cell-of-origin cannot be identified by studying ALL samples at diagnosis.[33] As a consequence, prospective study designs and animal models are needed.—*Which prospective study design can be performed?*

The stromal microenvironment is known to be important in B-lineage development and differentiation. Studies in mice could demonstrate that alterations in the hematopoietic stem cells niche (for example, mesenchymal cells) can lead to the development of myeloproliferative disease and secondary leukemia.[34] Surprisingly, leukemia-specific genetic aberrations were recently also found in mesenchymal stem cells (MSCs) of some ALL cases, but the frequency of this finding varies significantly and seems to be higher in relapse.[35,36]—*What is the role of the microenvironment in the persistence of preleukemic clones? Are (some of) the genetic aberrations caused by plasticity of leukemia cells? Is the correlation between aberrations in MSCs and prognosis/relapse observed in a limited number of patients also significant for a larger study cohort? (If so, this could be useful as prognostic marker.) Can the highly variable frequency of leukemia-specific translocations in MSC be correlated with any known risk factor? Do other cells in the affected individuals also carry these aberrations?*

Animal models and innovative study designs

A 'multi-hit' model is highly desirable to study the mechanisms of initiating events (genetic, epigenetic or environmental hits) as well as the nature and role of the further hits, their kinetics and age dependence (intrinsic versus extrinsic factors: chemicals, radiation, infections/inflammation, immune control of malignant progression), as this would allow hazard assessment studies of potential risk factors.[37]

So far, the exposure of wild-type and some genetically altered mice strains to magnetic fields that have been classified as 'possibly carcinogenic' (class 2b) by IARC did not confirm the epidemiological findings of low-frequency fields as risk factors. It cannot be excluded, however, that the negative results are at least partially due to the use of inadequate, that is, insensitive, animal models. Data from a new double-knockout mouse model (loss of p19Arf and Rag1) were presented at the meeting[38] and showed that p19Arf is a tumor suppressor in the Rag1-deficient B-cell precursor cells, as the combined loss resulted in B-cell precursor leukemia in mice. This raised the question whether this and other animal models generated so far (for example, MLL-AF9 model, see Stubbs et al.[39]; TEL-AML1 model, see Ford et al.[40]; TEL-AML1 model, see Schindler et al.[21]) are suitable and generally available or still have shortcomings that need to be overcome. There was a strong feeling in the group that, in addition to refinement of existing models, the generation of new, more adequate animal models is also necessary, as during the last years new insights on the hierarchical structure of the disease and on genotype–phenotype correlations have been gained.[41] The risk that these new models might still not manage to faithfully reproduce all aspects of human pathology is still high, but their generation and analysis are considered to add valuable information on mechanisms that cannot be gained otherwise.

A solid mechanistic understanding of leukemia development is also intended in animal studies and *in vitro* assays by using retrovirus-based vectors for insertional mutagenesis.[42,43] The retroviral vector insertion sites are frequently located in or near proto-oncogenes or signaling genes and the resulting insertional dominance database has led to the identification of genes that stimulate or transform hematopoietic stem cells or progenitor cells. Taken together, insertional mutagenesis provides a relatively unbiased tool for screening approaches as well as for functional studies on genes of interest complementing other ongoing study designs.

Modeling

As the latency for childhood ALL is short, it has been suggested that only a small number of rate-limiting steps are responsible for overt leukemia. Little et al.[44] could model most subtypes of ALL by assuming only rate-limiting steps. Recent studies in mice revealing an unanticipated high plasticity of lymphocytes[32,45] and the newly uncovered role of the microenvironment[35,36] will lead to development of existing models, which have been recently reviewed.[46]

In conclusion, the discussion on the above topics has led to the identification of the most important research areas in the near future and the following first-research agenda was developed.

Research recommendations

Human studies
(A) *Investigations on the prevalence of 'first-hit' events.* The focus should be on confirming the prevalence of TEL-AML1, but also other less frequent chromosomal translocations such as AML1-ETO, MLL-AF4, MLL-ENL, MLL-AF9 and E2A-PBX should be considered.

(A1) *Prevalence of the preleukemic clone*: Samples from existing Western Caucasian birth cohorts are preferable, as this would offer the possibility of follow-up for disease and, thus, would allow for the first time to develop and apply a prospective study design. In addition, many data and possible risk factors from the individuals of a birth cohort are available (parental age, birth weight, atopic disease, and so on), including some environmental exposure data. The anticipated number of samples that would need to be screened is approximately 40 000. The sample size needs to be sufficiently large to address the question whether the gender difference in ALL incidence rates can be explained by a gender difference in the prevalence of TEL-AML1.

(A2) *Comparison of regional differences* by focusing on regions with maximally reported ALL incidence rate differences (for example, regional differences reported to be up to threefold between the Caucasian population and developing countries in Central Africa). Data relating to the incidence of ALL in developing countries are sparse and those existing are often of poor quality. Therefore it is important to try and ascertain what the disease incidence and the prevalence of TEL-AML are in these countries.

Methodology and conditions to be developed:

- A network has to be established for defining a study protocol and for developing an infrastructure for standardized testing of ten thousands of samples (one place versus several places, for example, set up a lab in Africa that could also be used for other purposes).

- Development of a set of validated PCR primers as a tool to detect the presence of the most frequent translocations at the genomic level.
- The availability of previously collected cord blood samples (with low or known ethnic diversity) has to be established, as well as possible overlap or synergism with ongoing or planned studies.
- There are several ethical issues with a prospective design that have to be considered. Informed consent of parents is needed. One issue that needs to be carefully addressed is how to deal with those children who have the TEL-AML1 or other pre-disposing translocations.

(B) Deep sequencing. Characterization of the somatic changes by whole-genome sequencing of ALL cases taken from populations with different incidence rates (genome or exome, transcriptome, epigenome).

(B1) Deep sequencing of ALL individuals in a stepwise approach: A pilot study ($n = 10$, whole-genome or transcriptome sequences, exome capture and sequencing) on individuals already analyzed by GWASs would enable the comparison of different methods. In a next step, whole-genome sequencing of unselected ALL incident cases of one diagnostic year (for example, in Germany) is recommended (including data collection on possible risk factors like parental age, birth weight, birth order, diet, and so on). In addition, deep sequencing of predisposed children (from Project A1) is considered to add valuable information. Finally, sequencing of ALL cases from a population with low incidence rates in developing countries (Central Africa) should reveal population differences.

Information on somatic and transcriptome changes will also be a prerequisite for the development of predictive models that could help in detecting environmental effects and allow individually optimized treatment (http://www.treat1000.org).

(B2) Deep sequencing of the preleukemic clone from predisposed children (from Project A1): Critical step: How can the preleukemic clone be detected and isolated? As one single cell might not be enough for 'error-free' whole-genome sequencing, transcript analysis of the preleukemic clone could be a starting point.

Methodology and conditions to be developed:

- The development of a standardized questionnaire is necessary.
- In the prospective study design on predisposed individuals (from Project A1), sample collection, for example, skin biopsy or saliva samples, has to be validated.
- Again, ethical issues are considered to be difficult and need clarification. Informed consent of parents is needed and ascertainment of a 'newly screened' population versus already collected biological material. How to convert a program of genetic screening in, for example, an African country into something beneficial for the population as whole should be determined.

(C) Study the role of the hematopoietic stem cells niche (for example, mesenchymal cells) for the origin and maintenance of ALL. *(C1) Verification of the correlation between leukemia-specific aberrations in MSCs and prognosis/relapse* observed in a limited number of patients in a larger cohort (for example, in the unselected ALL incident cases of one diagnostic year from Project B1), including the correlation with data collected on possible risk factors such as parental age, birth weight, birth order, diet, and so on. It should be analyzed whether other cells in the affected patients also carry these aberrations (for example, endothelial cells, CD133 cells, osteoblasts, myeloid cells).

In a next step, the results should be compared with frequencies of leukemia-specific aberrations in MSCs in populations with different incidence rates.

Methodology and conditions to be developed:

- Usage of the developed set of validated PCR primers of Project A1 for detecting the most frequent translocations.

Animal models

The ultimate goal is to be able to mimic in the mouse the entire molecular, cellular, tissular and organic features of human B-ALL, including its initiation, progression, evolution, response to therapy and eventual cure or relapse.

(D) Generation of appropriate mouse models. *(D1) Check availability of existing animal models and the need for further refinement.*

(D2) Generation of several new mouse strains (at the same time): Among the different technical approaches available (transgenesis, retroviral introduction of mice or human hematopoietic cells, or injection of human leukemic cells into mice), germ-line transgenesis is preferred, as the generated mice strains can be shared. Coupling with imaging techniques/markers and/or gene marking or biotinylation is of help for further analysis. A moderate susceptibility of the strain is desired, as this allows the analysis of further hits. In all steps a comparison with clinical data has to be established.

Subproject: Generate mouse strains to dissect the contribution of non-hematopoietic cells to leukemia development (translocations under an MSC-specific promoter).

Analysis of host genetic background in leukemia development: As shown in a mouse model by Wang et al.,[47] the lineage fate of leukemia cells (B-ALL versus AML) was determined by the host microenvironment.

(D3) After having decided on a given model from Projects D1 or D2

1. *Exposure of animals to possible risk factors* (radiation, chemicals, infections/inflammation, replicative stress and so on) followed by complete phenotyping with standardized protocols.
2. *Generation of a B-cell leukemia model of controlled genetic variability* with a backcross between two syngeneic mouse strains (a resistant one and a susceptible one). It will permit to study, using an integrative systems biology approach and in a common scenario, the (multiple) low-susceptibility genes that determine essential aspects of variability in B-ALL behavior among individuals.
3. *Exposure of a cohort of this B-cell leukemia model of controlled genetic variability to possible risk factors.*

These steps need biometric planning and consideration of the 3Rs (reduce, refine, replace).

Methodology and conditions to be developed:

- Standardization of mouse phenotyping (mouse hospital, pathology, imaging techniques, molecular assays and so on).
- Development of methods for detection of clonal kinetics.

(E) Verification of the contribution of gene variants identified in human studies to the development of B-cell leukemias in mice

(F) Verification that the mechanisms outlined in A–E can quantitatively account for the totality of human data, via construction of novel quasi-mechanistic models

Concluding remarks

It was concluded that the exchange between different disciplines is a valuable and complementary approach to many ongoing studies. Some essential questions can be addressed only by combined efforts in an interdisciplinary network. The research agenda will be used by BfS for fund raising at the national and European level.

Conflict of interest

The authors declare no conflict of interest.

References

1 Kaatsch P, Spix C, Schulze-Rath R, Schmiedel S, Blettner M. Leukaemia in young children living in the vicinity of German nuclear power plants. *Int J Cancer* 2008; **122**: 721–726.
2 Schuez J, Ahlbom A. Exposure to electromagnetic fields and the risk of childhood leukemia: a review. *Rad Prot Dos* 2008; **132**: 202–211.
3 Risk factors for childhood leukemia. In Matthes R, Ziegelberger G (eds). Proceedings of an ICNIRP Workshop, Berlin, May 5–7, 2008. *Rad Prot Dos* 2008: Special Issue **132**: 107–274.
4 Steliarova-Foucher E, Coebergh JW, Kaatsch P, Pritchard-Jones K, Stiller C (eds). *Eur J Cancer* 2006; **42**: 1913–2190.
5 Hjalgrim LL, Westergaard T, Rostgaard K, Schmiegelow K, Melbye M, Hjalgrim H et al. Birth weight as a risk factor for childhood leukemia: a meta-analysis of 18 epidemiologic studies. *Am J Epidemiol* 2003; **158**: 724–735.
6 Caughey RW, Michels KB. Birth weight and childhood leukemia: a meta-analysis and review of the current evidence. *Int J Cancer* 2009; **124**: 2658–2670.
7 Mullighan CG, Goorha S, Radtke I, Miller CB, Coustan-Smith E, Dalton JD et al. Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature* 2007; **446**: 758–764.
8 Treviño LR, Yang W, French D, Hunger SP, Carroll WL, Devidas M et al. Germline genomic variants associated with childhood acute lymphoblastic leukemia. *Nat Genet* 2009; **41**: 1001–1005.
9 Papaemmanuil E, Hosking FJ, Vijayakrishnan J, Price A, Olver B, Sheridan E et al. Loci on 7p12.2, 10q21.2 and 14q11.2 are associated with risk of childhood acute lymphoblastic leukemia. *Nat Genet* 2009; **41**: 1006–1010.
10 Prasad RB, Hosking FJ, Vijayakrishnan J, Papaemmanuil E, Koehler R, Greaves M et al. Verification of the susceptibility loci on 7p12.2, 10q21.2, and 14q11.2 in precursor B-cell acute lymphoblastic leukemia of childhood. *Blood* 2010; **115**: 1765–1767.
11 Hosking FJ, Papaemmanuil E, Sheridon E, Kinsey SE, Lightfoot T, Roman E et al. Genome-wide homozygosity signatures and childhood acute lymphoblastic leukemia risk. *Blood* 2010; **115**: 4472–4477.
12 Sherborne AL, Hosking FJ, Prasad RB, Kumar R, Koehler R, Vijayakrishnan J et al. Variation in CDKN2A at 9p21.3 influences childhood acute lymphoblastic leukemia risk. *Nat Genet* 2010; **42**: 492–494.
13 Pleasance ED, Stephens PJ, O'Meara S, McBride DJ, Meynert A, Jones D et al. A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature* 2010; **463**: 184–190.
14 Linden T, Schnittger S, Groll AH, Juergens H, Rossig C. Childhood B-cell precursor acute lymphoblastic leukemia in a patient with familial thrombocytopenia and RUNX1 mutation. *Br J Haematol* 2010; **151**: 528–530.
15 Ford AM, Ridge SA, Cabrera ME, Mahmoud H, Steel CM, Chan LC. In utero rearrangements in the trothorax-related oncogene in infant leukaemias. *Nature* 1993; **363**: 358–360.
16 Wiemels JL, Cazzaniga G, Daniotti M, Eden OB, Addison GM, Masera G et al. Prenatal origin of acute lymphoblastic leukaemia in children. *Lancet* 1999; **354**: 1499–1503.
17 Mori H, Colman SM, Xiao Z, Ford AM, Healy LE, Donaldson C et al. Chromosome translocations and covert leukemic clones are generated during normal fetal development. *Proc Natl Acad Sci USA* 2002; **99**: 8242–8247.
18 Mullighan CG, Philips LA, Su X, Ma J, Miller CB, Shurtleff SA et al. Genomic alterations of the clonal origins of relapsed acute lymphoblastic leukemia. *Science* 2008; **322**: 1377–1380.
19 Lausten-Thomsen U, Madsen HO, Vestergaard TR, Hjalgrim H, Nersting J, Schmiegelow K. Prevalence of t(12;21) [ETV6-RUNX1]-positive cells in healthy neonates. *Blood* 2010, DOI: 10.1182/blood-2010-05-282764 (e-pub ahead of print 16 August).
20 Slany RK. The molecular biology of mixed lineage leukemia. *Haematologica* 2009; **94**: 984–993.
21 Schindler JW, Van Buren D, Foudi A, Krejci O, Qin J, Orkin SH et al. TEL-AML1 corrupts hematopoietic stem cells to persist in the bone marrow and initiate leukemia. *Cell Stem Cell* 2009; **5**: 43–53.
22 Cooke J. The incidence of acute leukemia in children. *JAMA* 1942; **119**: 547–550.
23 Ward G. Infective theory of acute leukaemia. *Br J Child Dis* 1917; **14**: 10–20.
24 Kinlen LJ. Evidence for an infective cause of childhood leukaemia: comparison of a Scottish new town with nuclear reprocessing sites in Britain. *Lancet* 1988; **2**: 1323–1327.
25 Kinlen LJ. Epidemiological evidence for an infective basis in childhood leukaemia. *Br J Cancer* 1995; **71**: 1–5.
26 Greaves MF. Speculations on the cause of childhood acute lymphoblastic leukemia. *Leukemia* 1988; **2**: 120–125.
27 Urayama KY, Buffler PA, Gallagher ER, Ayoob JM, Ma X. A meta-analysis of the association between day-care attendance and childhood acute lymphoblastic leukaemia. *Int J Epidemiol* 2010; **39**: 718–732.
28 Roman E, Simpson J, Ansell P, Kinsey S, Mitchell CD, Mc Kinney PA et al. Childhood acute lymphoblastic leukemia and infections in the first year of life: a report from the United Kingdom Childhood Cancer Study. *Am J Epidemiol* 2007; **165**: 496–504.
29 Roman E, Simpson J, Ansell P, Lightfoot T, Smith A. Infectious proxies and childhood leukaemia: findings from the United Kingdom Childhood Cancer Study (UKCCS). *Blood Cells Mol Dis* 2009; **42**: 126–128.
30 Dahl S, Schmidt LS, Vestergaard T, Schuez J, Schmiegelow K. Allergy and the risk of childhood leukemia: a meta-analysis. *Leukemia* 2009; **23**: 2300–2304.
31 Chuk MK, McIntyre E, Small D, Brown P. Discordance of MLL-rearranged (MLL-R) infant acute lymphoblastic leukemia in monozygotic twins with spontaneous clearance of preleukemic clone in unaffected twin. *Blood* 2009; **113**: 6691–6694.
32 Cobaleda C, Jochum W, Busslinger M. Conversion of mature B cells into T cells by dedifferentiation to uncommitted progenitors. *Nature* 2007; **449**: 473–477.
33 Cobaleda C, Sánchez-García I. B-cell acute lymphoblastic leukaemia: towards understanding its cellular origin. *Bioessays* 2009; **31**: 600–609.
34 Raaijmakers MH, Mukherjee S, Guo S, Zhang S, Kobayashi T, Schoonmaker JA et al. Bone progenitor dysfunction induces myelodysplasia and secondary leukaemia. *Nature* 2010; **464**: 852–857.
35 Menendez P, Catalina P, Rodriguez R, Melen GJ, Bueno C, Arriero M et al. Bone marrow mesenchymal stem cells from infants with MLL-AF4+ acute leukemia harbor and express the MLL-AF4 fusion gene. *J Exp Med* 2009; **206**: 3131–3141.
36 Shalapour S, Eckert C, Seeger K, Pfau M, Prada J, Henze G et al. Leukemia-associated genetic aberrations in mesenchymal stem cells of children with acute lymphoblastic leukemia. *J Mol Med* 2010; **88**: 249–265.
37 McCormick DL, Kavet R. Animal Models for the study of childhood leukemia: considerations for model identification and optimization to identify potential risk factors. *Int J Toxicol* 2004; **23**: 149–161.
38 Hauer J, Mulligan C, Morillon E, Wang G, Bruneau J, Brousse N et al. Loss of p19Arf in a Rag1$^{-/-}$ B-cell precursor population initiates acute B-lymphoblastic leukemia. (submitted).

6

39 Stubbs MC, Kim YM, Krivtsov AV, Wright RD, Feng Z, Agarwal J et al. MLL-AF9 and FLT3 cooperation in acute myelogenous leukemia: development of a model for rapid therapeutic assessment. Leukemia 2008; 22: 66–77.

40 Ford AM, Palmi C, Bueno C, Hong D, Cardus P, Knight D et al. The TEL-AML1 leukemia fusion gene dysregulates the TGF-beta pathway in early B lineage progenitor cells. J Clin Invest 2009; 119: 826–836.

41 Vicente-Duenas C, Cobaleda C, Perez-Losada J, Sanchez-Garcia IL. The evolution of cancer modeling: the shadow of stem cells. Dis Models Mechanisms 2010; 3: 149–155.

42 Kustikova OS, Geiger H, Li Z, Brugman MH, Chambers SM, Shaw CA et al. Retroviral vector insertion sites associated with dominant hematopoietic clones mark 'stemness' pathways. Blood 2007; 109: 1897–1907.

43 Modlich U, Navarro S, Zychlinski D, Maetzig T, Knoess S, Brugman MH et al. Insertional transformation of hematopoietic cells by self-inactivating lentiviral and gammaretroviral vectors. Mol Ther 2009; 17: 1919–1928.

44 Little MP, Muirhead CR, Stiller CA. Modelling lymphocytic leukaemia incidence in England and Wales using generalizations of the two-mutation model of carcinogenesis of Moolgavkar, Venzon and Knudson. Stat Med 1996; 15: 1003–1022.

45 Cobaleda C, Busslinger M. Developmental plasticity of lymphocytes. Curr Opin Immunol 2008; 20: 139–148.

46 Little MP. Cancer models, genomic instability and somatic cellular Darwinian evolution. Biol Direct 2010; 5: 19.

47 Wang PY, Young F, Chen CY, Stevens BM, Neering SJ, Rossi RM et al. The biologic properties of leukemias arising from BCR/ABL-mediated transformation vary as a function of developmental origin and activity of the p19ARF gene. Blood 2008; 112: 4184–4192.