Backtracking to the future: unraveling the origins of childhood leukemia

Adam J. de Smith, Joseph L. Wiemels, Adam J. Mead, Irene Robert, Anindita Roy and Logan G. Spector

CHILDHOOD LEUKEMIA EPIDEMIOLOGY

Leukemia is the most common type of childhood malignancy, with ~40 cases of acute lymphoblastic leukemia (ALL) and ~ 8 cases of acute myeloid leukemia (AML) diagnosed per million children <15 years of age in the United States [1]. For ALL, a peak in incidence in children between 2 to 5 years of age has been linked to dysregulated immune development including delayed exposure to infections during early childhood [2]. Childhood AML incidence peaks during the first two years of life, with likely distinct etiologies to ALL. Childhood ALL cure rates are > 90% in high-income countries, but survivors face life-long treatmentrelated morbidities [3]. For childhood AML, cure rates are inferior at ~70%, and patients who survive endure similar late effects of treatment and a reduced quality of life. Thus, prevention of childhood leukemia (CL) remains an essential goal to reduce this significant health burden and requires a better understanding of CL etiologies. Given the young age-at-diagnosis of most patients, a prenatal origin for CL has long been proposed and, indeed, confirmed for several subtypes (see below, and Fig. 1). Here, we give an overview of current knowledge on the early-life origins of CL and present our efforts to elucidate this through the ReCord Study.

IN UTERO DEVELOPMENT

Most subtypes, as well as most individual CL cases, are presumed to have in utero origins based both on circumstantial evidence and direct measurement of leukemia-causing somatic alterations in biological samples taken shortly after birth. Studies in monozygotic twins, for instance, have demonstrated that CL can develop in utero [4]. Identical translocation events have been detected in twin pairs concordant for KMT2A-rearranged infant ALL or AML, for ETV6::RUNX1 fusion ALL, or for BCR::ABL1 ALL, supporting that the initiating lesion arises in one twin and transfers to the other via shared placental vasculature [4]. In addition, the observed wide variation in age of leukemia onset in twins concordant for ALL [4] and analysis of the clonal evolution of leukemia cells in such twins support a two-hit model of leukemogenesis, with in utero formation of a pre-leukemic clone and postnatal development of additional mutations that drive progression to overt leukemia [2]. Twins concordant for CL are rare, and further insight into the natural history of CL has been gained through backtracking studies using newborn blood samples from children who later developed leukemia. This was first demonstrated by Gale and colleagues, who detected the presence of leukemia-forming KMT2A::AFF1 fusions in the newborn DBS of three ALL patients <2 years of age [5]. In a subsequent study by Wiemels et al., the majority of ETV6::RUNX1 fusion ALL patients were positive for the translocation sequence in newborn dried bloodspots (DBS) [6]. In addition, backtracking of KMT2A-rearranged infant ALL and AML, AML1::ETO in AML, and clonal IGH rearrangements in high hyperdiploid ALL support the prenatal development of these leukemia subtypes (Fig. 1 and reviewed by Marcotte et al. [7]; subtype frequencies derived from Pui et al. [8] and Huang et al. [9]). Results for TCF3::PBX1 fusion have been less consistent [7] and require clarification. Recently described and less common ALL subtypes, including BCR::ABL1-like and translocations of DUX4, ZNF384, and MEF2D as well as most T-cell ALL and AML-defining rearrangements, have yet to be examined at birth (Fig. 1). This is likely due, at least in part, to a lack of access to suitable stored samples. In a recent Perspective by Greaves and colleagues, the in utero origins of ALL in twins discordant for ALL was described, as the same fusion sequence was detected in newborn blood samples from the affected twins as well as their healthy co-twins [10]. Backtracking studies have largely corroborated the findings in twins, supporting the postnatal acquisition of second-hit somatic alterations [11]. Several limitations to existing backtracking data are evident, due to the reliance on investigation of bulk tissues (tumor DNA and newborn blood) rather than at the single-cell level. First, the timings of the initiating lesion and of secondary alterations have not been definitively determined across CL subtypes. Second, backtracking studies have been unable to attribute the presence of somatic alterations to particular sub-populations of developing hematopoietic cells, due in part to the almost complete reliance on newborn DBS. Indeed, to our knowledge, there is only one instance of backtracking CL to a cord blood sample [12]; this same patient showed no evidence of pre-leukemia in a newborn DBS, further supporting the limitations of DBS for definitive backtracking. Third, there is almost no data on the clonal fraction of cells with pre-leukemia that are detectable in newborn blood samples from children who later developed leukemia. Lastly, only a small number of leukemia cases have been backtracked to date, which has prevented analysis of the association between CL risk factors and prevalence of preleukemic cells at birth.

CELL-OF-ORIGIN Pediatric ALL and AML likely arise from lymphoid and myeloid precursor cells, respectively, or from hematopoietic stem cells, but the specific cells-of-origin in which pre-leukemic genetic lesions arise have not been definitively determined. Previous studies have inferred the cell-of-origin for several CL subtypes by studying the characteristics of leukemia cells and progenitors at diagnosis, including immunophenotyping and gene expression patterns [13]. For example, CD19– or multipotent stem cells have been identified as the putative cells-of-origin for some ALL subtypes, such as KMT2A-rearranged and the p210 form of BCR::ABL1, whereas other subtypes including ETV6::RUNX1 and p190 BCR::ABL1 were proposed to arise in CD19+ B-cell progenitors (reviewed by Hein et al. [13]). Identifying the precise cells-of-origin across CL subtypes will further our understanding of the prenatal origins of CL, and inform the development of early detection, therapeutic, and potentially of preventive, strategies. Examination of leukemia cells at diagnosis is limited because tumor cell phenotypes may not accurately reflect the cells-oforigin. For instance, the pre-leukemic lesion, along with secondary alterations, may reprogram the cell-of-origin's properties so that it bears markers of cells upstream or downstream in the hierarchy of hematopoiesis [13]. Mouse models examining the leukemiainitiating capacity of different cell types have also been used to investigate potential cells-of-origin. For example, mouse models of infant KMT2A-rearranged AML demonstrated the potential origins in HSCs or granulocyte–macrophage progenitors [14, 15], though studies of other pediatric AML subtypes are lacking. Furthermore, studies in mice may not necessarily identify the cell-of-origin, but instead determine cell types that have the most potential to engraft and then propagate cells harboring the leukemia-initiating lesion. A more fruitful approach may be to investigate preleukemic cells-of-origin using single-cell analyses in perinatal blood samples from children who later developed leukemia; however, such studies are currently lacking.

NEWBORN SCREENING AND PREVENTION

The possibility of screening newborns for pre-leukemic cells to identify children at risk of developing leukemia has been considered since the first backtracking studies demonstrated the prenatal origin of some leukemias. Indeed, leukemia-initiating fusions, including ETV6::RUNX1 have been detected in blood samples from as many as 5% of unselected newborns, a frequency much higher than the incidence of ETV6::RUNX1 leukemia itself [7]. Realizing the ambition of newborn screening will require knowledge of the entire natural history and particularly the prenatal origin of the full spectrum of leukemia-typical somatic alterations. This natural history should be discernible at the cellular level with much more precision in cord blood, which contains millions of hematopoietic stem and progenitor cells that can be isolated and interrogated for leukemia-associated genetic lesions. However, given the burden of collection, and cost of storing, cord blood it is unlikely to ever be screened at scale. Rather, screening would need to be implemented in the less ideal specimen, newborn DBS, because they have the advantage of already being collected population-wide in developed nations. The possibility of primary prevention of CL through early-life immunomodulation has begun to be discussed [2] concurrently with the renewed interest in the natural history of leukemia in cord blood and the development of new technologies to screen for pre-leukemia in DBS.

THE RECORD STUDY

We thus now introduce the ReCord study to the leukemia research community. ReCord is a collaboration between the University of Minnesota, University of Southern California, and University of Oxford, funded by the National Institutes of Health (R01CA262012). The study is collecting banked cord blood and paired tumor samples from CL patients, along with prenatal medical records and parental epidemiologic questionnaires. ReCord has several aims (Fig. 2). Briefly, the study will identify the presence and frequency of prenatal leukemia-initiating lesions in specific cord blood hematopoietic precursor cell populations using personalized digital PCR probes. It will investigate the cellof- origin of leukemia-initiating lesions, transcriptomic changes from pre-leukemia to overt leukemia, and whether secondary mutations arise prenatally, across CL subtypes via TARGET-seq [16] (single-cell, simultaneous DNA- and RNA-seq). It will also determine whether the presence and frequency of pre-leukemic clones correlate with risk factors for leukemia such as birthweight, sex, parental age, and genetic variants. In the United States about 20% of children were born in states which store DBS long-term and allow retrieval for research, thus the study will seek DBS when available. Comparison of the detection of pre-leukemic clones in DBS to that in cord blood, the gold standard, will provide critical data to inform the feasibility of screening for pre-leukemia. Participants are being identified through the Children's Oncology Group's Project: EveryChild but volunteer patients with banked cord blood are welcome to join ReCord: ReCord.umn.edu.

CONCLUSION

CL backtracking studies have been limited to the use of newborn DBS or the identification of identical translocations in concordant twins at diagnosis, which has precluded investigation of the frequency of pre-leukemic clones and identification of preleukemic cells-of-origin, and have examined only a fraction of the known molecular subtypes. There also is no epidemiology of pre-leukemia to date. Examination of a large number of cord blood samples from children with leukemia is poised to provide major advancements in our understanding of the cell-of-origin of diverse types of CL and to describe the clonal evolution of preleukemia to overt leukemia with great resolution.