Backtracking to the future: unraveling the origins of childhood leukemia

Adam J. de Smith[1,2], **Joseph L. Wiemels**[1,2], **Adam J. Mead**[3], **Irene Roberts**[4], **Anindita Roy**[4,5], **Logan G. Spector**[6]

[1]Center for Genetic Epidemiology, Department of Population and Public Health Sciences, University of Southern California Keck School of Medicine, Los Angeles, CA, USA.

[2]USC Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, CA, USA.

[3]Haematopoietic Stem Cell Biology Laboratory, MRC Molecular Haematology Unit, MRC Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK.

[4]Department of Paediatrics and MRC Molecular Haematology Unit, MRC Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK.

[5]Department of Haematology, Great Ormond Street Hospital for Children, London, UK.

[6]Division of Epidemiology and Clinical Research, Department of Pediatrics, University of Minnesota, Minneapolis, Minnesota, MN, USA.

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 treatment-related 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 pre-leukemic 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-of-origin. 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 leukemia-initiating 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 pre-leukemic 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 cell-of-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 pre-leukemic 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 pre-leukemia to overt leukemia with great resolution.

Author Manuscript

ACKNOWLEDGEMENTS

REFERENCES

1. American Cancer Society. Cancer Facts and Figures, 2019. American Cancer Society, Inc.; 2019. https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2019.html.

2. Greaves M. A causal mechanism for childhood acute lymphoblastic leukaemia. Nat Rev Cancer. 2018;18:471–84. 10.1038/s41568-018-0015-6. [PubMed: 29784935]

3. Phillips SM, Padgett LS, Leisenring WM, Stratton KK, Bishop K, Krull KR, et al. Survivors of childhood cancer in the United States: prevalence and burden of morbidity. Cancer Epidemiol Biomark Prev. 2015;24:653–63. 10.1158/1055-9965.EPI-14-1418.

4. Greaves MF, Maia AT, Wiemels JL, Ford AM. Leukemia in twins: lessons in natural history. Blood. 2003;102:2321–33. 10.1182/blood-2002-12-3817. [PubMed: 12791663]

5. Gale KB, Ford AM, Repp R, Borkhardt A, Keller C, Eden OB, et al. Backtracking leukemia to birth: identification of clonotypic gene fusion sequences in neonatal blood spots. Proc Natl Acad Sci USA. 1997;94:13950–4. 10.1073/pnas.94.25.13950. [PubMed: 9391133]

6. Wiemels JL, Cazzaniga G, Daniotti M, Eden OB, Addison GM, Masera G, et al. Prenatal origin of acute lymphoblastic leukaemia in children. Lancet Lond Engl. 1999;354:1499–503. 10.1016/s0140-6736(99)09403-9.

7. Marcotte EL, Spector LG, Mendes-de-Almeida DP, Nelson HH. The prenatal origin of childhood leukemia: potential applications for epidemiology and newborn screening. Front Pediatr. 2021;9:639479. 10.3389/fped.2021.639479. [PubMed: 33968846]

8. Pui CH, Nichols KE, Yang JJ. Somatic and germline genomics in paediatric acute lymphoblastic leukaemia. Nat Rev Clin Oncol. 2019;16:227–40. 10.1038/s41571-018-0136-6. [PubMed: 30546053]

9. Huang BJ, Smith JL, Farrar JE, Wang YC, Umeda M, Ries RE, et al. Integrated stem cell signature and cytomolecular risk determination in pediatric acute myeloid leukemia. Nat Commun. 2022;13:5487. 10.1038/s41467-022-33244-6. [PubMed: 36123353]

10. Ford AM, Colman S, Greaves M. Covert pre-leukaemic clones in healthy co-twins of patients with childhood acute lymphoblastic leukaemia. Leukemia. 2023;37:47–52. 10.1038/s41375-022-01756-1. [PubMed: 36536099]

11. Wiemels JL, Hofmann J, Kang M, Selzer R, Green R, Zhou M, et al. Chromosome 12p deletions in TEL-AML1 childhood acute lymphoblastic leukemia are associated with retrotransposon elements and occur postnatally. Cancer Res. 2008;68:9935–44. 10.1158/0008-5472.CAN-08-2139. [PubMed: 19047175]

12. Maia AT, Tussiwand R, Cazzaniga G, Rebulla P, Colman S, Biondi A, et al. Identification of pre-leukemic precursors of hyperdiploid acute lymphoblastic leukemia in cord blood. Genes Chromosomes Cancer. 2004;40:38–43. 10.1002/gcc.20010. [PubMed: 15034866]

13. Hein D, Borkhardt A, Fischer U. Insights into the prenatal origin of childhood acute lymphoblastic leukemia. Cancer Metastasis Rev. 2020;39:161–71. 10.1007/s10555-019-09841-1. [PubMed: 31902036]

14. Krivtsov AV, Figueroa ME, Sinha AU, Stubbs MC, Feng Z, Valk PJM, et al. Cell of origin determines clinically relevant subtypes of MLL-rearranged AML. Leukemia. 2013;27:852–60. 10.1038/leu.2012.363. [PubMed: 23235717]

15. Krivtsov AV, Twomey D, Feng Z, Stubbs MC, Wang Y, Faber J, et al. Transformation from committed progenitor to leukaemia stem cell initiated by MLL-AF9. Nature. 2006;442:818–22. 10.1038/nature04980. [PubMed: 16862118]
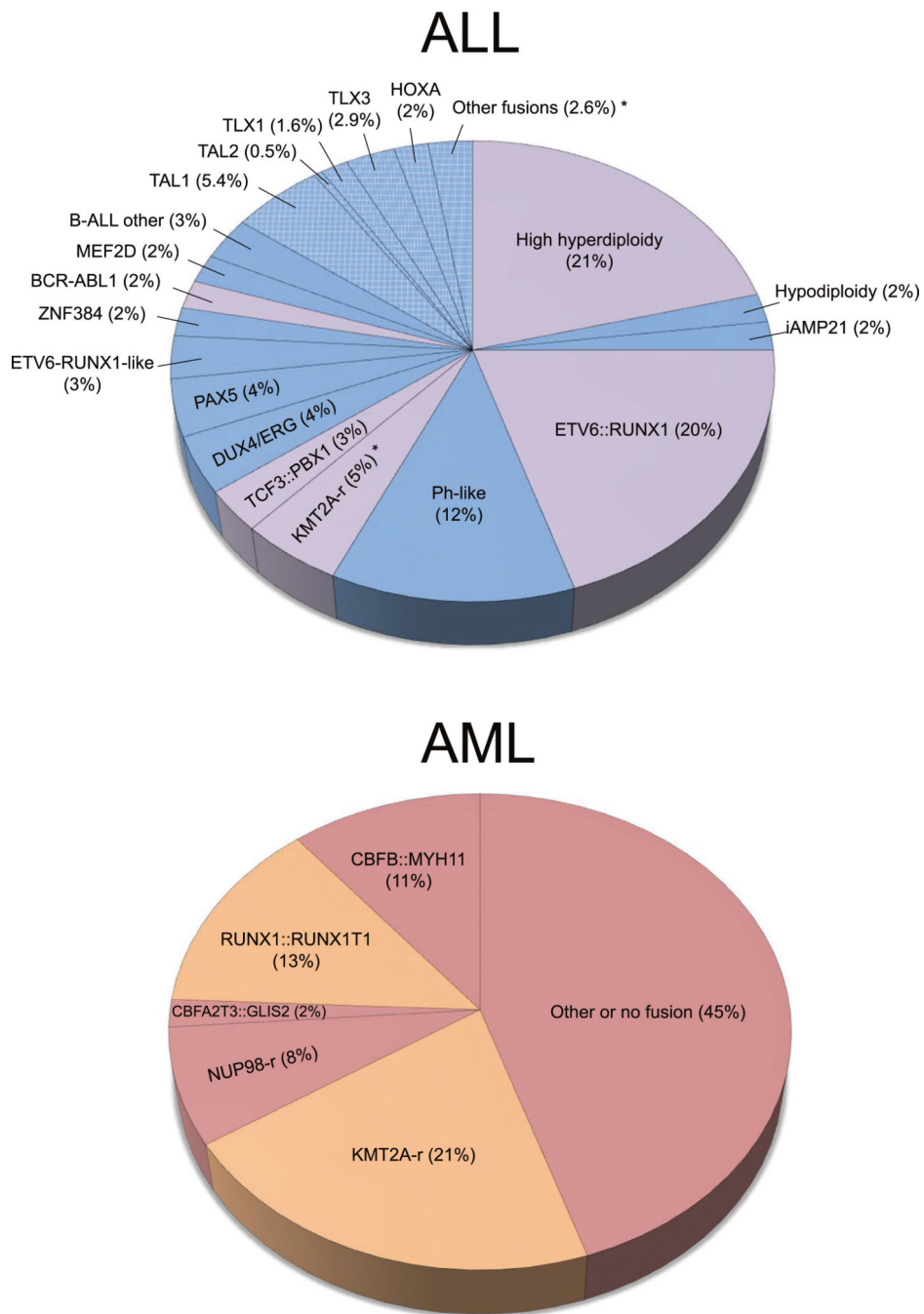
16. Rodriguez-Meira A, Buck G, Clark SA, Povinelli BJ, Alcolea V, Louka E, et al. Unravelling intratumoral heterogeneity through high-sensitivity single-cell mutational analysis and parallel RNA sequencing. Mol Cell. 2019;73:1292–1305.e8. 10.1016/j.molcel.2019.01.009. [PubMed: 30765193]

Fig. 1. Childhood leukemia subtypes and backtracking status.

Top: ALL molecular subtype frequencies were modified from Pui et al. 2019. ALL subtypes that have been previously backtracked are highlighted in purple. Remaining B-cell ALL and T-cell ALL subtypes that have not been backtracked are colored blue, with T-cell subtypes highlighted by cross-hatching. Asterisks indicate that some T-cell ALL patients with *KMT2A*-rearranged (*KMT2A*-r) and other mutation events have been backtracked. Bottom: AML subtype frequencies from Huang et al. 2022, with previously backtracked subtypes highlighted in orange and remaining non-backtracked subtypes colored red.
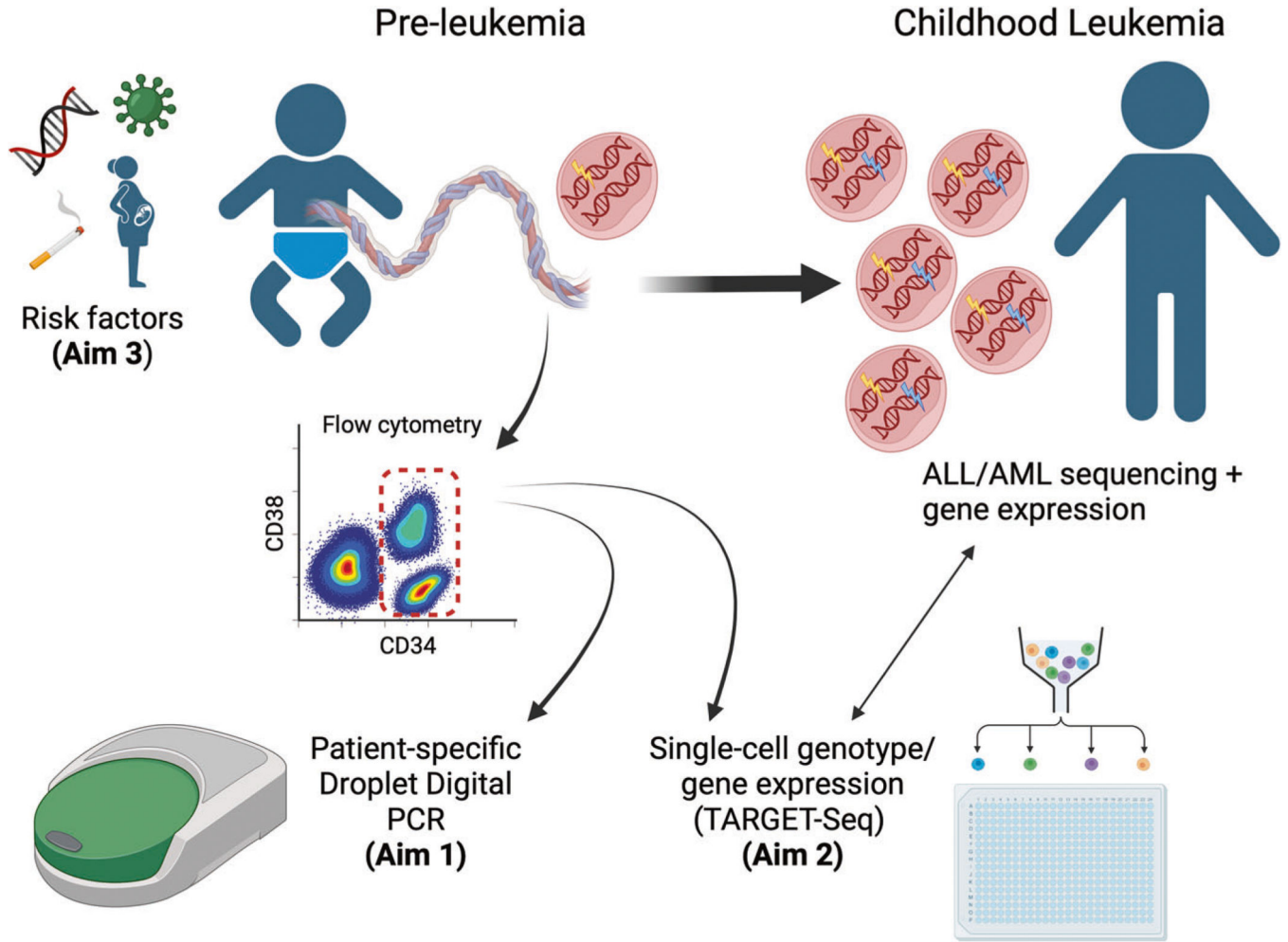
Fig. 2. Aims of the ReCord study, 2021-ongoing.
Aim 1: Enroll childhood leukemia patients with banked cord blood, then sequence
leukemia specimens to determine the driving molecular lesions in each patient, followed
by backtracking in cord blood using patient-specific droplet digital PCR (ddPCR) assays
to determine which leukemia subtypes arise prenatally and their clonal frequency across
cord blood cell compartments. Aim 2: Investigate the cell-of-origin of leukemia-initiating
lesions and transcriptomic changes from pre-leukemia to overt leukemia, across childhood
leukemia subtypes, at the single-cell level using TARGET-Seq in paired cord blood and
diagnostic leukemia samples. Aim 3: To determine whether the presence and frequency
of pre-leukemic clones correlate with leukemia risk factors, including demographic, pre/
perinatal, and genomic risk factors for ALL/AML. Created with BioRender.com.