SUPPLEMENTAL METHODS

Bulk and sc-RNAseq

Bulk RNAseq: A previously published dataset was used [1]. Briefly, 100 cells from each hematopoietic population from second trimester FBM (n=3) were flow-sorted using a Becton Dickinson FACS Aria III as previously described [2], and RNA-seq performed as described [1].

High-throughput scRNAseq (10x Chromium): lineage negative (Lin^-) $CD34^+$ cells from matched second trimester FL and FBM from the same fetuses (n=2) were flow-sorted using a Becton Dickinson FACS Aria III as previously described [2]. 15,000-24,000 flow-sorted Lin^-CD34^+ cells were processed and analyzed using SingCellaR (https://github.com/supatt-lab/SingCellaR) as previously described [2].

FACS immunophenotyping

Primary BM samples from B-ALL patients, CB and adult BM were stained using the fluorochrome-conjugated monoclonal antibodies (MoAb) anti-CD34-PerCPCy5.5 (clone 8G12, BD Biosciences), anti-CD22-APC-AF700 (clon S510.1H11, Beckman Coulter) and anti-CD19-PECy7 (clone J3-119 Immunostep), and analyzed, after erythrocyte lysis, using a FACS CantoNon leukemic samples were analyzed with the FACS DIVA (BD Biosciences) software and B-ALL samples with the Infinicyt software (Cytognos S.L), as previously described [3-5].

FL and FBM samples were processed and analysed by flow cytometry as described in [1,2] using the following antibodies: Lineage2 cocktail [composed by anti-CD2-PerCPcy5.5 (clone RPA-2.10), anti-CD3-PerCPcy5.5 (clone okt3), anti-CD14-PerCPcy5.5 (clone M5E2), anti-CD16-PerCPcy5.5 (clone 3G8), anti-CD56-PerCPcy5.5 (clone HCD56), and anti-CD235a-PerCPcy5.5 (clone HIR2), all from Biolegend], anti-CD34-BV421 (clone 561, Biolegend), anti-CD19-APC (clone HIB19, Biolegend), anti-CD123-BV650 (clone 6H6, Biolegend), anti-CD22-PECy7 (clone HIB22, BD Biosicences), anti-CD45RA-ACPef780 (clone HI100, Life Technologies), anti-CD127-PE (clone eBioRDR5, Life Technologies), anti-CD38-af700 (clone HIT2, Life Technologies), anti-CD10-FITC (clone ebioCB-CALLA, Life Technologies), and Fixable Viability Dye ef506 (Life Technologies). The frequency of $CD22^+CD19^-$ cells within the $CD34^+$ population was calculated for each B-ALL sample and also for each FL, FBM, CB and adult BM specimen.

FISH on flow-sorted CD34+ cell subpopulations

B-ALL cells from three independent patients were stained for CD34, CD19 and CD22 as indicated above and both the $CD34^+CD19^+CD22^+$ and the $CD34^+CD19^-CD22^+$ were flow-purified using a FACSAria-III flow cytometer (BD Biosciences). Sorted blast subpopulations were immediately fixed in 3:1 (v:v)

methanol/acetic buffer and stored at 4ºC for FISH. FISH was performed on interphase nuclei according to well-standardized protocols [6,7] using the following chromosomal probes: i) LSI MLL Dual Color Break Apart probe (Vysis) for MLL rearrangements (11q23) and copy number alterations and ii) LSI IGH/MYC/CEP 8 Tri-Color Dual Fusion probe (Vysis) for t(8;14).

***In vivo* engraftment assays in NSG mice**

Eight-to-12-week-old nonobese diabetic/LtSz-scid IL-2Ry$^{-/-}$ mice (NSG) mice housed under pathogen-free conditions were used in this study. All experimental procedures were approved by the Animal Care Committee of The Barcelona Biomedical Research Park (DAAM7393). Due to the rarity of the CD34^{+}CD19^{-}CD22^{+} population, 1.3x10^{5} CD34^{+}CD19^{-}CD22^{+} or CD34^{+}CD19^{+}CD22^{+} cells were flow-sorted from a pool of ten CD19 CAR-treated B-ALL patients and transplanted into sublethally irradiated NSG mice (n=3/cell subset) by intra-osseous transplantation [8]. Leukemic engraftment was monitored every other week in peripheral blood (PB) by flow cytometry. Leukemic mice showing disease symptoms or leukemic engraftment in PB>10% were sacrificed for leukemia immunophenotype and molecular characterization [8].

Cells retrieved from BM and were stained with HLA-ABC-FITC and CD45-APC-H7 MoAbs and analysed by flow cytometry to detect the human graft (HLA-ABC^{+}CD45^{+} population). Within the human graft, the leukemia was immunophenotyped using CD33-PE, CD19-BV421 and CD22-APC-AF700 MoAbs as mentioned above on a FACSCanto-II cytometer running FACSDiva software (BD Biosciences). The identity of the B-ALL graft from CD34^{+}CD19^{-}CD22^{+} cells was confirmed using targeted-next generation sequencing (NGSeq) and optical genome mapping (OGM), as previously described [9,10].

Molecular characterization of the B-ALL graft by targeted NGSeq and OGM

Library preparation was performed from the flow-sorted leukemic graft using the AmpliSeq™ Childhood Cancer Panel kit following manufacturer's instructions and sequencing was performed on a MiSeq instrument (Illumina, San Diego, CA). Data obtained from the sequencer was analyzed by using the DNA amplicon app, RNA amplicon app and OncoCNV caller from BaseSpace™ Sequence Hub (Illumina, Inc). Somatic variants and copy number variations were analyzed and classified according to the AMP/ASCO/CAP Standards and Guidelines for Somatic Variant Interpretation [10].

OGM was performed on the Saphyr System (Bionano Genomics, San Diego, CA, USA). High molecular weight DNA was extracted according to the manufacturer's manual (Bionano Prep SP Frozen Cell Pellet DNA Isolation Protocol). gDNA was labelled following manufacturer's instructions (Bionano Prep Direct

Label and Stain Protocol). Labelled gDNA was loaded on a Saphyr G2.3 chip and scanned on the Saphyr instrument. The chip was run, and 1526 Gbp were collected reaching a 298.77x genome coverage. Rare variant analysis was performed with the Bionano Access. In order to remove contamination by mouse cells, a cut off of 100Kbp was used for structural variants (SVs) and 500Kbp for copy number variants (CNVs) [9]

Clinical impact of the CD34+CD19-CD22+ population on relapse

The percentage of $CD34^{+}CD19^{-}CD22^{+}$ cells in BM was analyzed before therapy in 53 patients treated with CD19-directed immunotherapy (n=27 CD19-CAR T-cells, n=16 blinatumomab; **Table 2**) to assess its impact on disease relapse risk within a median follow-up of 24-months. Patients were treated (and BM samples analyzed) at the following local hospitals: Hospital Clinic, Hospital Sant Joan de Deu (pediatrics) and Hospital Germans Trias i Pujol-ICO Badalona. The cumulative incidence of relapse (as a competitive event with death without relapse) was analyzed by splitting our cohort in two groups according to the frequency of this cell population above or below the mean (%=2.45). Differences in the cumulative incidence of relapse rate were evaluated with the Gray test. All statistical analyses were performed with R statistics version 4.0.2.

REFERENCES

1. O'Byrne S, Elliott N, Rice S, et al. Discovery of a CD10-negative B-progenitor in human fetal life identifies unique ontogeny-related developmental programs. *Blood*. 2019;134(13):1059-1071.
2. Roy A, Wang G, Iskander D, et al. Developmental stage- and site-specific transitions in lineage specification and gene regulatory networks in human hematopoietic stem and progenitor cells. *bioRxiv*. 2021:2021.2004.2020.440420.
3. Menendez P, Prosper F, Bueno C, et al. Sequential analysis of CD34+ and CD34- cell subsets in peripheral blood and leukapheresis products from breast cancer patients mobilized with SCF plus G-CSF and cyclophosphamide. *Leukemia*. 2001;15(3):430-439.
4. Theunissen P, Mejstrikova E, Sedek L, et al. Standardized flow cytometry for highly sensitive MRD measurements in B-cell acute lymphoblastic leukemia. *Blood*. 2017;129(3):347-357.
5. van Dongen JJ, Lhermitte L, Bottcher S, et al. EuroFlow antibody panels for standardized n-dimensional flow cytometric immunophenotyping of normal, reactive and malignant leukocytes. *Leukemia*. 2012;26(9):1908-1975.
6. Bueno C, Catalina P, Melen GJ, et al. Etoposide induces MLL rearrangements and other chromosomal abnormalities in human embryonic stem cells. *Carcinogenesis*. 2009;30(9):1628-1637.
7. Catalina P, Montes R, Ligero G, et al. Human ESCs predisposition to karyotypic instability: Is a matter of culture adaptation or differential vulnerability among hESC lines due to inherent properties? *Mol Cancer*. 2008;7:76.
8. Prieto C, Lopez-Millan B, Roca-Ho H, et al. NG2 antigen is involved in leukemia invasiveness and central nervous system infiltration in MLL-rearranged infant B-ALL. *Leukemia*. 2018;32(3):633-644.
9. Cao H, Hastie AR, Cao D, et al. Rapid detection of structural variation in a human genome using nanochannel-based genome mapping technology. *Gigascience*. 2014;3(1):34.
10. Li MM, Datto M, Duncavage EJ, et al. Standards and Guidelines for the Interpretation and Reporting of Sequence Variants in Cancer: A Joint Consensus Recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists. *J Mol Diagn*. 2017;19(1):4-23.

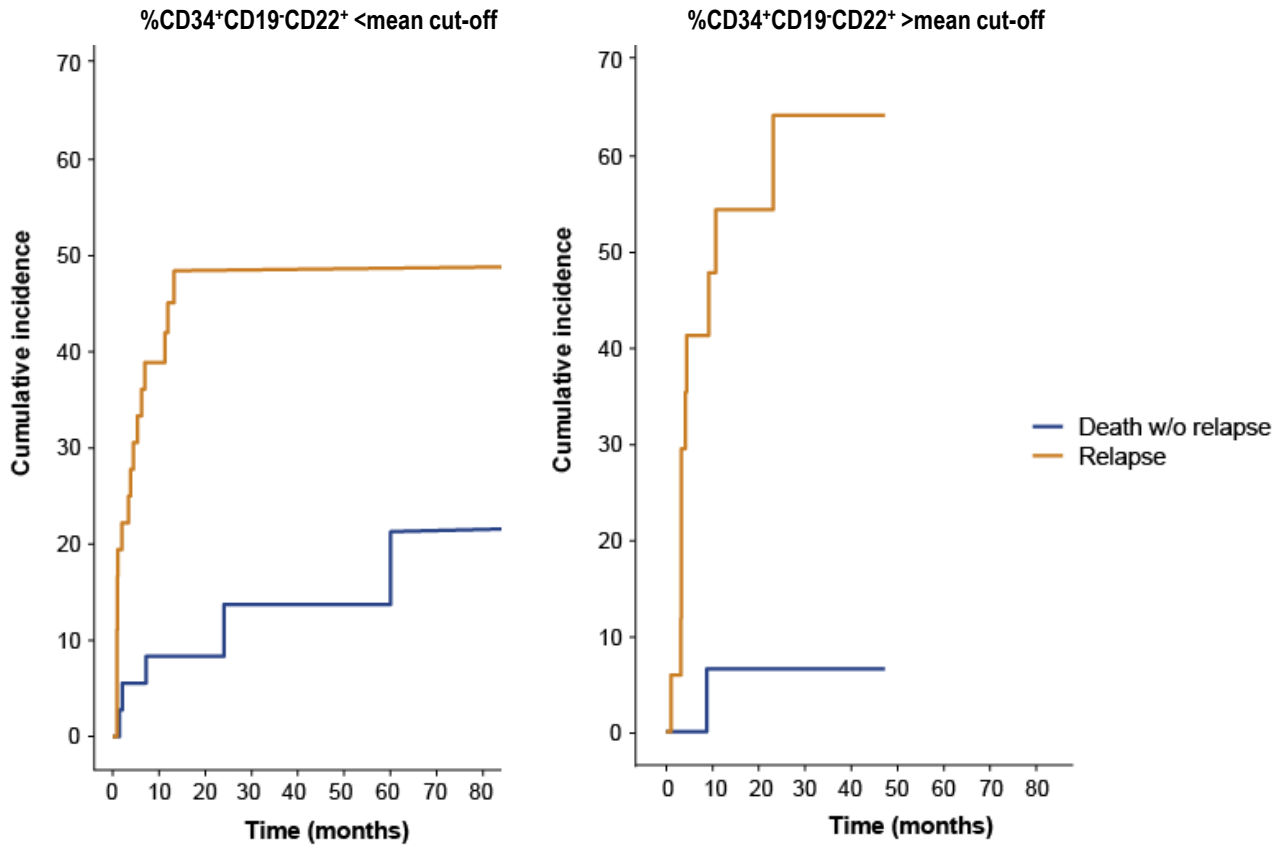

Figure S1. Cumulative relapse incidence of R/R B-ALL patients treated with CD19-directed immunotherapy according to the frequency of CD34+CD19-CD22+ cells (above or below the mean) analyzed in BM before treatment.