



NUDT21 limits CD19 levels through alternative mRNA polyadenylation in B cell acute lymphoblastic leukemia

Matthew T. Witkowski^{1,16,18} , Soobeom Lee^{1,2,18}, Eric Wang^{3,17,18}, Anna K. Lee¹, Alexis Talbot⁴, Chao Ma^{5,6}, Nikolaos Tsopoulidis^{7,8,9}, Justin Brumbaugh¹⁰, Yaqi Zhao¹¹, Kathryn G. Roberts¹¹, Simon J. Hogg¹², Sofia Nomikou¹, Yohana E. Ghebrechristos¹, Palaniraja Thandapani¹³, Charles G. Mullighan¹⁴, Konrad Hochedlinger¹⁵, Weiqiang Chen^{5,6}, Omar Abdel-Wahab^{13,12}, Justin Eyquem^{4,13,14,15} and Iannis Aifantis¹ 

B cell progenitor acute lymphoblastic leukemia (B-ALL) treatment has been revolutionized by T cell-based immunotherapies—including chimeric antigen receptor T cell therapy (CAR-T) and the bispecific T cell engager therapeutic, blinatumomab—targeting surface glycoprotein CD19. Unfortunately, many patients with B-ALL will fail immunotherapy due to ‘antigen escape’—the loss or absence of leukemic CD19 targeted by anti-leukemic T cells. In the present study, we utilized a genome-wide CRISPR-Cas9 screening approach to identify modulators of CD19 abundance on human B-ALL blasts. These studies identified a critical role for the transcriptional activator ZNF143 in CD19 promoter activation. Conversely, the RNA-binding protein, NUDT21, limited expression of CD19 by regulating CD19 messenger RNA polyadenylation and stability. NUDT21 deletion in B-ALL cells increased the expression of CD19 and the sensitivity to CD19-specific CAR-T and blinatumomab. In human B-ALL patients treated with CAR-T and blinatumomab, upregulation of NUDT21 mRNA coincided with CD19 loss at disease relapse. Together, these studies identify new CD19 modulators in human B-ALL.

Loss of expression of the transmembrane surface glycoprotein CD19 on B-ALL blasts drives resistance to CAR-T and bispecific T cell engager therapy^{1–3}, but the gene regulatory programs that control the abundance of CD19 on the cell surface remain unclear. To identify regulators of transmembrane glycoprotein CD19 surface abundance in transformed human B cells, we combined genome-wide clustered regularly interspaced short palindromic repeats (CRISPR)–Cas9 screening and flow-based separation based on the abundance of CD19 on the B cell surface. We generated five human Cas9-expressing CD19⁺ B cell lines, including

three B-ALL (Reh, NALM6 and 697) and two mature B cell neoplasm cell lines from chronic lymphocytic leukemia (HG3) and B cell lymphoma (TMD8) (Fig. 1a). Stable Cas9 expression was achieved using lentiviral transduction of blasticidin-selectable Cas9 constructs. Blasticidin-resistant Cas9⁺ cell lines were then transduced with the Brunello genome-wide single guide (sg)RNA library⁴ with sgRNA-expressing cells tracked by green fluorescent protein (GFP) expression (Extended Data Fig. 1a,b). Using a CD19 surface protein flow-based sorting strategy, we isolated GFP⁺CD19^{lo} and GFP⁺CD19^{hi} B cells, followed by deep sequencing of sgRNAs to generate a z-score-normalized ‘CD19 CRISPR score’ representing the average score of sgRNAs per gene within CD19^{lo} cells (CD19 activators; sgRNA targeting of CD19 activators would lead to loss of CD19 expression and a negative CRISPR score) and CD19^{hi} cells (CD19 repressors; sgRNA targeting a CD19 repressor would increase CD19 expression resulting in a positive CRISPR score) relative to non-targeting control sgRNAs (Extended Data Fig. 1a and Supplementary Table 1). To identify high-confidence and cell line-specific CD19 regulators, we selected the top 200 CD19 activators and the top 200 CD19 repressors for all five cell lines tested (Fig. 1b, Extended Data Fig. 1c,d and Supplementary Table 1). As such, we identified 22 CD19 activators and 6 CD19 repressors unique to Cas9⁺Reh, Cas9⁺697 and Cas9⁺NALM6 cells, as well as 5 CD19 activators and 3 CD19 repressors common to Cas9⁺HG3 and Cas9⁺TMD8 (Fig. 1c,d and Extended Data Fig. 1e,f). We observed the significant enrichment of CD19 sgRNA within the CD19^{lo} populations in all five human Cas9⁺ B cell lines (Fig. 1b). Consistent with CD19 protein maturation occurring via the secretory pathway-, endoplasmic reticulum- and Golgi apparatus-associated proteins such as CD81 (refs. ^{5,6}), CANX⁷ and

¹Department of Pathology and Laura & Isaac Perlmutter Cancer Center, NYU School of Medicine, New York, NY, USA. ²Department of Biology, New York University (NYU), New York, NY, USA. ³Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA.

⁴Department of Medicine, University of California San Francisco, San Francisco, CA, USA. ⁵Department of Mechanical and Aerospace Engineering, New York University, New York, NY, USA. ⁶Department of Biomedical Engineering, New York University, New York, NY, USA. ⁷Department of Molecular Biology, Massachusetts General Hospital, Boston, MA, USA. ⁸Cancer Center and Center for Regenerative Medicine, Massachusetts General Hospital, Boston, MA, USA. ⁹Department of Genetics, Harvard Medical School, Boston, MA, USA. ¹⁰Department of Molecular, Cellular, and Developmental Biology, University of Colorado Boulder, Boulder, CO, USA. ¹¹Department of Pathology, St. Jude Children’s Research Hospital, Memphis, TN, USA. ¹²Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, USA. ¹³Gladstone–UCSF Institute of Genomic Immunology, San Francisco, CA, USA. ¹⁴Department of Microbiology and Immunology, University of California San Francisco, San Francisco, CA, USA. ¹⁵Parker Institute of Cancer Immunotherapy, University of California San Francisco, San Francisco, CA, USA. ¹⁶Present address: Department of Pediatrics, University of Colorado Anschutz Medical Campus, Aurora, USA. ¹⁷Present address: The Jackson Laboratory for Genomic Medicine, Farmington, USA. ¹⁸These authors contributed equally: Matthew T. Witkowski, Soobeom Lee, Eric Wang. ✉e-mail: matthew.witkowski@cuanschutz.edu; iannis.aifantis@nyulangone.org

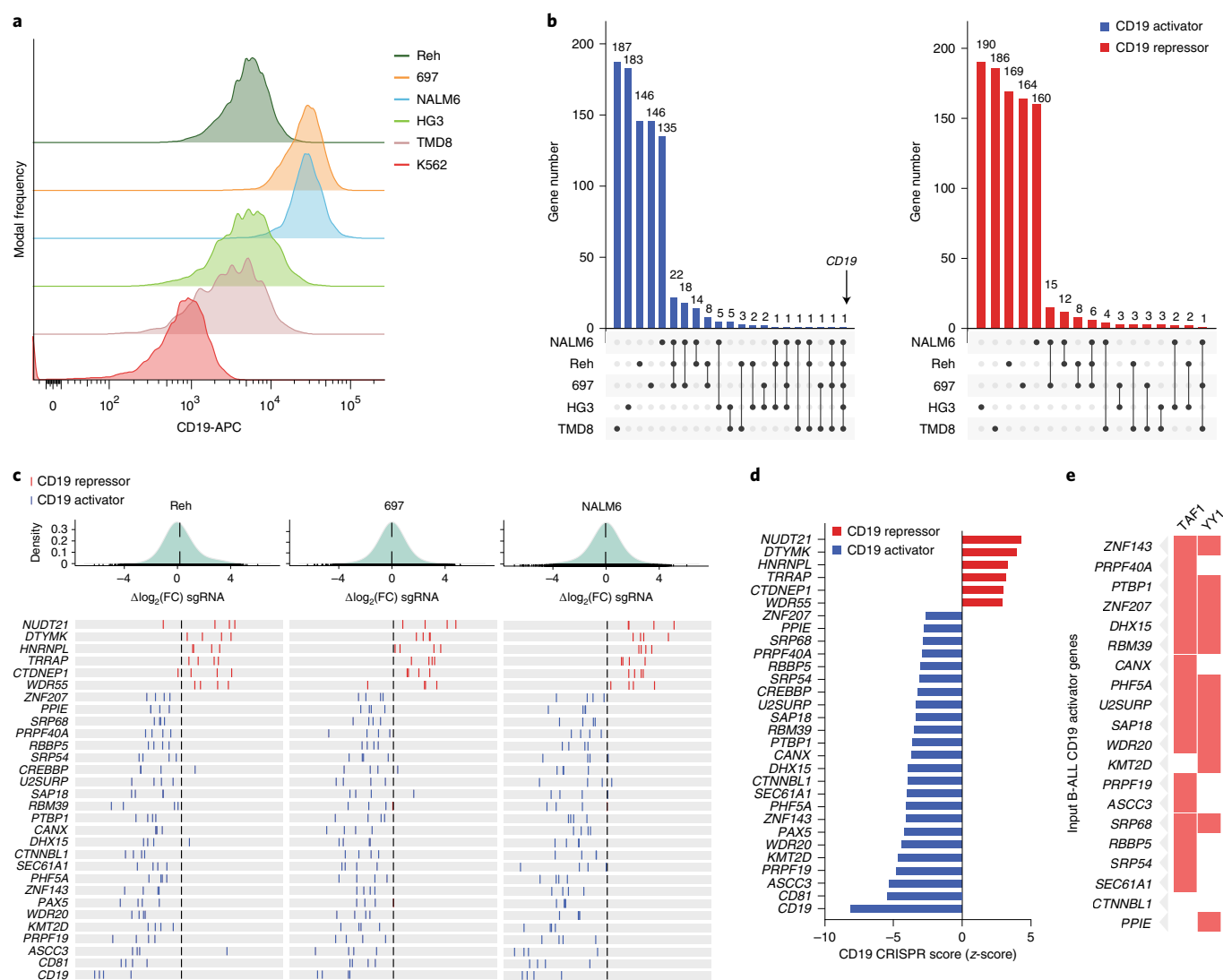


Fig. 1 | Genome-wide CRISPR screen identifies CD19 regulators in human B cell malignancies. **a**, Representative flow cytometry of CD19 expression across Cas9⁺Reh, Cas9⁺697, Cas9⁺NALM6, Cas9⁺HG3 and Cas9⁺TMD8 human B cell lines with Cas9⁺K562 erythroleukemia cell line as CD19⁻ control. **b**, Upset plot highlighting the number of overlapping CD19 activators (blue) and CD19 repressors (red) across Cas9⁺Reh, Cas9⁺697, Cas9⁺NALM6, Cas9⁺HG3 and Cas9⁺TMD8 cell lines based on the top 200 CD19 activators and repressors from each line. CD19 is indicated. **c**, Histogram showing sgRNA fold-change for individual sgRNAs targeting the top gene candidates for CD19 activators (blue) and CD19 repressors (red) in Cas9⁺Reh, Cas9⁺697 and Cas9⁺NALM6 cells. **d**, Waterfall plot showing the average CD19 CRISPR score (z-score normalized) of the top gene candidates for CD19 activators (blue) and CD19 repressors (red) in Cas9⁺NALM6, Cas9⁺Reh and Cas9⁺697 cells. **e**, Heatmap of CD19 activators associated with TAF1⁻ and YY1-interacting proteins based on the TAF1 and YY1 ENCODE ChIP-X GSEA.

SEC61A1 (ref. ⁸) were identified as CD19 activators in Cas9⁺Reh, Cas9⁺697 and Cas9⁺NALM6 lines (Fig. 1c,d and Extended Data Fig. 1c,g). Most B-ALL CD19 activators, including *PAX5* and *PTBP1*, which promote *CD19* promoter activation⁹ and *CD19* intron 2 exclusion¹⁰, respectively (Fig. 1c,d), have been implicated in the genome regulation and mRNA processing of *CD19*. Gene-set enrichment analysis (GSEA) of CD19 activators in B-ALL indicated overlap with subunits of the TAF1 and YY1 protein complexes (Fig. 1e), which are required for RNA polymerase II-mediated transcription initiation¹¹.

The DNA-binding transcription factor ZNF143, one of the CD19 activators across Cas9⁺Reh, Cas9⁺697 and Cas9⁺NALM6, is known to co-localize with YY1 and TAF1 (ref. ¹¹) and potentially regulate cell type-specific, promoter-enhancer, DNA-loop formation¹². To explore the role of ZNF143 in *CD19* gene activation, we performed RNA-sequencing (RNA-seq) analysis of Cas9⁺Reh, Cas9⁺697

and Cas9⁺NALM6 cells expressing sgRNAs targeting ZNF143 (sgZNF143-mCherry) or the *ROSA26* locus (sgROSA-mCherry) as a negative control (Fig. 2a). Immunoblot analysis confirmed that sgZNF143-mCherry reduced the expression of full-length ZNF143 relative to sgROSA in Cas9⁺NALM6 cells (Extended Data Fig. 2a). Differentially expressed gene (DEG) analysis indicated significant downregulation of *CD19* mRNA expression in sgZNF143-mCherry⁺ Cas9⁺Reh, Cas9⁺697 and Cas9⁺NALM6 cells compared with sgROSA-mCherry⁺ counterparts (Fig. 2a,b). We also observed little overlap in the ZNF143-dependent global transcriptional changes across Reh, 697 and NALM6 cell lines (Fig. 2b and Supplementary Table 2). Flow cytometry analysis indicated that the expression of CD19 was significantly reduced in sgZNF143-mCherry⁺ Cas9⁺Reh, Cas9⁺697, Cas9⁺NALM6 and the B cell lymphoblastoid cell line, Cas9⁺GM12878, compared with sgROSA-mCherry⁺ cells (Fig. 2c,d).

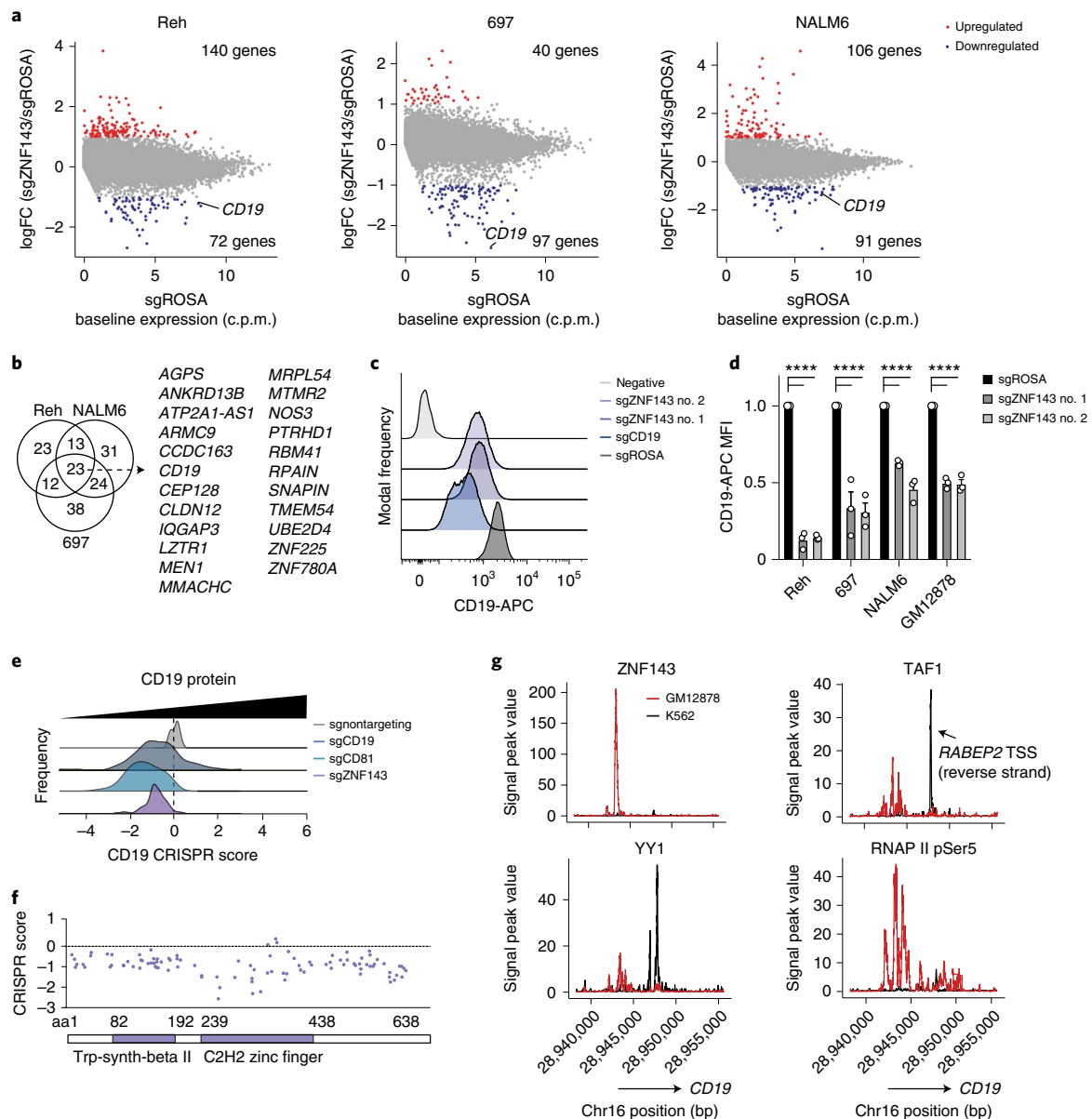


Fig. 2 | ZNF143 directly binds the CD19 locus and activates gene expression. **a**, M (log(sgZNF143:sgROSA)) and A (mean average of sgROSA) plot of Cas9⁺Reh, Cas9⁺697 and Cas9⁺NALM6 cells 2 days post-transduction with sgRNAs. ZNF143 upregulated (red) and downregulated (blue) genes are highlighted with log(FC) exceeding ± 1 cutoff. c.p.m., counts per million. CD19 is indicated. **b**, Venn diagram of overlapping downregulated genes across sgZNF143-mCherry⁺Cas9⁺Reh, -697 and -NALM6 cells. **c**, Representative flow cytometry of CD19 expression 7 days post-transduction of Cas9⁺697 cells with sgROSA, sgCD19, sgZNF143 no. 1 and sgZNF143 no. 2. **d**, Flow cytometry of CD19 expression 7 days post-transduction with sgZNF143-mCherry in Cas9⁺Reh, -697, -NALM6 and -GM12878 cells normalized to sgROSA-mCherry⁺ Cas9⁺Reh, -697, -NALM6 and -GM12878 cells (independent experiments with $n = 3$, unpaired, two-sided Student's t -test, mean and s.e.m. shown; error bars represent s.e.m.). MFI, mean fluorescence intensity. Data with statistical significance are as indicated: **** $P < 0.0001$. **e**, Histogram of individual sgRNA CD19 CRISPR scores from the Cas9⁺NALM6 cell domain screening. **f**, CD19 CRISPR score for individual sgRNAs (dots) spanning the ZNF143 locus in the Cas9⁺NALM6 cell domain screen. aa, amino acid. **g**, Histograms of signal peak value in the CD19 locus (ENCODE) for GM12878 cells (red line) and K562 cells (black line) for ZNF143, TAF1, YY1 and RNAP II Ser5.

ZNF143 promotes gene activation through its DNA-binding zinc fingers¹³. To assess the protein domains of ZNF143 required for activation of CD19, Cas9⁺NALM6 cells were transduced with a pooled sgRNA library targeting multiple exons of ZNF143, CD19 and CD81 and ten nontargeting sgRNAs, and isolated CD19^{lo} and CD19^{hi} cells, followed by sgRNA deep sequencing 12 days post-transduction (Fig. 2e and Supplementary Table 3). Using this 'domain screening' approach, we calculated that the CD19 CRISPR score for sgRNAs targeting CD19, CD81 and ZNF143 was lower relative to non-targeting sgRNAs (Fig. 2e and Supplementary Table 3),

indicating effective sgRNA targeting of each CD19 activator. Targeting exons in the CD19 extracellular domains and the CD81 transmembrane domains resulted in low CD19 CRISPR scores compared with non-targeting sgRNAs (Fig. 2e and Extended Data Fig. 2b,c), indicating that this domain screening approach effectively identifies key protein domains critical for full-length surface CD19 protein production. The sgRNAs targeting the ZNF143 C2H2 zinc finger domain resulted in a low CD19 CRISPR score compared with non-targeting sgRNAs (Fig. 2f), suggesting that DNA binding by ZNF143 may be involved in CD19 gene activation. Consistent

with this, analysis of the ENCODE chromatin immunoprecipitation sequencing (ChIP-seq) dataset¹⁴ indicated direct binding of ZNF143, YY1 and TAF1 to the actively transcribed *CD19* promoter (RNA Pol II Ser5 binding) in the CD19⁺ B cell lymphoblastoid GM12878 cell line, whereas no binding was detected in the CD19[−] K562 erythroleukemia cell line (Fig. 2g).

As ZNF143 promotes cell type-specific DNA-loop formation^{12,15}, we characterized the ZNF143-mediated looping changes within the topologically associated domain (TAD) encapsulating the *CD19* locus. Hi-C analysis of sgZNF143–mCherry⁺ Cas9⁺Reh, Cas9⁺697 and Cas9⁺NALM6 cells found minimal impact on genome-wide intra-TAD structures, including no significant intra-TAD activity changes impacting the TAD harboring the *CD19* locus, compared with the corresponding sgROSA–mCherry⁺ cells (Extended Data Fig. 2d), consistent with previous studies¹⁵. Virtual 4C analysis, using the *CD19* promoter as bait, indicated that ZNF143 targeting did not alter *CD19* promoter interactions in sgZNF143–mCherry⁺ Cas9⁺Reh, Cas9⁺697 and Cas9⁺NALM6 cells relative to sgROSA–mCherry⁺ counterparts (Extended Data Fig. 2e), suggesting that promoter-bound ZNF143 directly activated gene expression in the absence of distal genome interactions.

We then sought to identify gene regulatory networks underpinning *CD19* expression by correlating gene expression patterns of *CD19* regulators in B-ALL with *CD19* mRNA expression patterns across multiple primary human hematopoietic cell types. Using single-cell RNA-seq (scRNA-seq) data from four healthy and seven diagnosis B-ALL human bone marrow specimens¹⁶ (Fig. 3a), we confirmed the overrepresentation of human B cell progenitors (defined by Seurat reference dataset as Prog_B 1 and Prog_B 2) within the leukemic bone marrow samples (Fig. 3b). *CD19* mRNA expression was restricted to the B cell lineage with highest expression in healthy and B-ALL Prog_B 1 (characterized by high *EBF1* and *MKI67*) and Prog_B 2 (characterized by high *CD19*, *PAX5*, *EBF1* and *MKI67*) subsets compared with all other *CD19*-expressing cell populations, including naive B cells, memory B cells and plasmablasts (Fig. 3c). Notably, *CD19* mRNA expression was highest in leukemia-associated Prog_B 2 cells (Fig. 3c,d and Supplementary Table 4). Ranking the expression of individual candidate genes in the *CD19*^{hi} Prog_B 2 compartment in healthy and leukemic conditions showed that the mRNA expression of *PAX5*, a known *CD19* activator⁹, and the transcript for the RNA-binding protein *NUDT21* significantly associated with *CD19* mRNA expression (Fig. 3d, Extended Data Fig. 3a,b and Supplementary Table 4). *NUDT21* mRNA expression positively correlated with *CD19* mRNA in healthy and leukemic Prog_B 1 and Prog_B 2 B cell progenitor subsets, whereas this significant correlation was reduced in naive and memory B cells (Fig. 3e and Extended Data Fig. 3c). We also found a positive correlation between the abundance of *NUDT21* protein and the expression of *CD19* protein in *CD19*^{hi} NALM6 B-ALL cells that were sorted into quartiles based on their expression of *CD19* (Figs. 1a and 3f,g).

As *NUDT21* was one of the top-ranked repressors of *CD19* expression in our genome-wide CRISPR screen, we investigated whether *NUDT21* limits the abundance of *CD19* expression in transformed B cell progenitors. The sgRNAs targeting *NUDT21* (sgNUDT21–mCherry) reduced the expression of full-length *NUDT21* in Cas9⁺NALM6 cells relative to sgROSA–mCherry⁺ and sgCD19–mCherry⁺ Cas9⁺NALM6 cells (Extended Data Fig. 4a). At day 7 post-transduction, flow cytometry of surface *CD19* expression, normalized to the expression of the highly expressed surface protein *CD147*, which was not identified as a *CD19* regulator in our genome-wide CRISPR screen, indicated a significant increase in *CD19* abundance in sgNUDT21–mCherry⁺ Cas9⁺Reh, -697 or -NALM6 cells, but not in sgNUDT21–mCherry⁺ Cas9⁺TMD8 or Cas9⁺HG3, compared with sgROSA–mCherry⁺ counterparts (Fig. 4a,b). In growth competition assays, tracking the

percentage of sgRNA–mCherry in Cas9⁺ cells relative to non-transduced, mCherry[−] cells at days 2, 7 and 14 post-transduction indicated that expression of sgNUDT21 reduced the growth of Cas9⁺Reh, Cas9⁺697, Cas9⁺NALM6 and Cas9⁺HG3 cells when compared with sgROSA and sgCD19 (Extended Data Fig. 4b). To determine whether the phenotypes elicited by sgNUDT21 in B-ALL could be rescued, we ectopically expressed either an sgRNA-resistant form of *NUDT21* (*NUDT21*^{sgRes}–GFP) or an empty vector (EV)–GFP in Cas9⁺NALM6 cells. Co-expression of sgROSA–mCherry did not affect the expression of *CD19* or the cellular fitness in either *NUDT21*^{sgRes}–GFP⁺ or EV–GFP⁺ Cas9⁺NALM6 cells (Extended Data Fig. 4c–e). Conversely, expression of sgNUDT21–mCherry in EV–GFP⁺ Cas9⁺NALM6 cells resulted in significant upregulation of *CD19* and loss of cellular fitness compared with sgROSA–mCherry, which was rescued by overexpression of *NUDT21*^{sgRes} (Extended Data Fig. 4c–e).

Next, to assess whether *CD19* expression was dynamically regulated by the reversible inhibition of *NUDT21*, we used 697 cells transduced with the doxycycline (dox)-inducible type VI CRISPR–Cas13d¹⁷ to knock down *NUDT21* and *CD19* mRNAs in an inducible manner (Methods). Cas13d⁺697 cells lentivirally transduced with sgNUDT21–GFP, sgCD19–GFP or sgNTC (nontargeting control)–GFP were treated with dox for 6 days followed by dox withdrawal in vitro, and the expression of *CD19* was tracked every 2 days by flow cytometry (Fig. 4c,d). Dox-dependent induction of sgCD19–GFP significantly reduced the expression of *CD19* on Cas13d⁺697 cells compared with sgNTC–GFP⁺ Cas13d⁺697 cells, whereas dox withdrawal restored *CD19* expression (Fig. 4c,d). Conversely, the dox-induced knock down of *NUDT21* mRNA in sgNUDT21–GFP⁺ Cas13d⁺697 cells significantly increased the amount of surface *CD19* relative to sgNTC–GFP⁺ Cas13d⁺697 cells (Fig. 4c,d), whereas restoration of *NUDT21* by dox withdrawal resulted in a reduction of *CD19* expression (Fig. 4c,d).

To determine whether *Nudt21*-mediated repression of *CD19* is conserved in mice, we isolated whole bone marrow hematopoietic cells from *ROSA26-Cre^{ERT2}+Nudt21^{+/+}* or *ROSA26-Cre^{ERT2}+Nudt21^{fl/fl}* mice, which ubiquitously express the tamoxifen-inducible *Cre^{ERT2}* from the *ROSA26* promoter. We then retrovirally transduced bone marrow cells with murine stem cell virus (MSCV)-based constructs that allow the ectopic expression of human *BCR::ABL1^{p190}* linked to GFP and maintained cells for 4 weeks on OP9 stromal cells to generate *BCR::ABL1-GFP⁺ Cre^{ERT2}+Nudt21^{+/+}* and *BCR::ABL1-GFP⁺ Cre^{ERT2}+Nudt21^{fl/fl}*-transformed *CD19⁺IgM[−]* B cell progenitor lines, each capable of maintenance in stroma-free culture. Treatment with 4-hydroxytamoxifen (4-OHT) for 4 days, which ablated the expression of *Nudt21* protein, as determined by immunoblot analysis (Extended Data Fig. 4f), followed by 4-OHT withdrawal for a further 8 days, significantly increased expression of *CD19* protein and reduced cellular fitness in *BCR::ABL1-GFP⁺ Cre^{ERT2}+Nudt21^{fl/fl}* cells when compared with vehicle-treated cells, as measured by flow cytometry (Extended Data Fig. 4g–i). 4-OHT treatment of *BCR::ABL1-GFP⁺ Cre^{ERT2}+Nudt21^{+/+}* cells did not impact *CD19* expression compared with vehicle-treated controls (Extended Data Fig. 4g) but induced transient toxicity (Extended Data Fig. 4i). These data suggested a conserved role of *NUDT21* in limiting *CD19* expression in transformed B cell progenitor cells in humans and mice.

NUDT21 promotes pre-mRNA 3'-end cleavage and polyadenylation (pA) through the recognition of 5'-UGUA-3' sequences upstream of pA sites in the 3'-UTR of pre-mRNA molecules¹⁸. To understand the role of *NUDT21* in *CD19* mRNA processing, we analyzed the Cas9⁺NALM6 cell-based, CRISPR-based exon mutagenesis dataset. SgRNA targeting of the nucleotide hydrolase domain that encompasses the *NUDT21* RNA-binding domains resulted in high *CD19* CRISPR scores (Extended Data Fig. 5a), suggesting that this domain was involved in repression of *CD19*

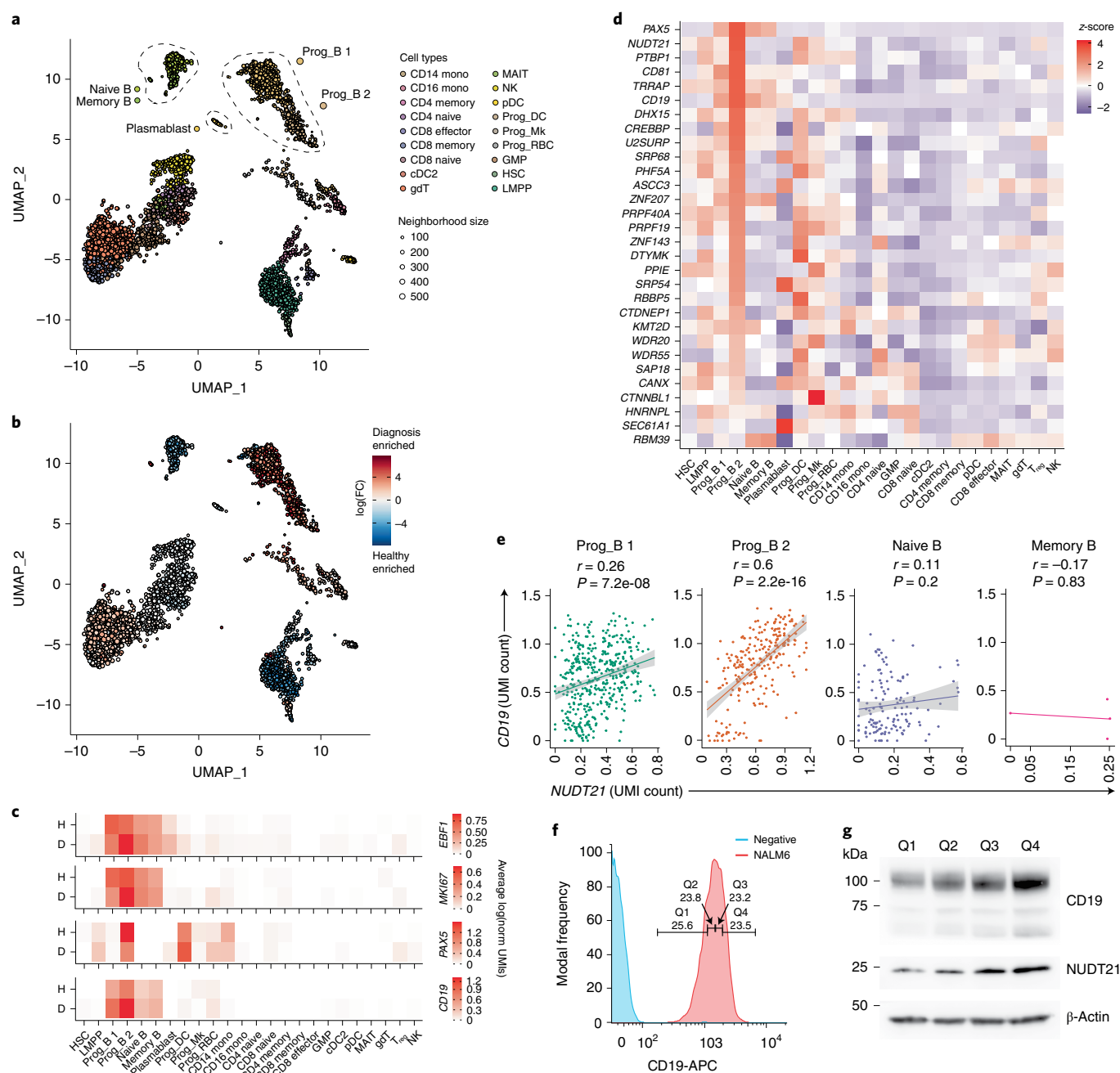


Fig. 3 | NUDT21 is highly expressed in human B cell progenitors. a, UMAP representation of primary healthy ($n=4$) and diagnosis B-ALL ($n=7$) bone marrow single-cell data (list indicated on the figure). **b**, Milo neighborhood analysis showing log(FC) representation of B-ALL diagnosis (red) and healthy (blue) neighborhood representation. **c**, Heatmap of average log(normalized UMI mRNA counts) of *CD19*, *PAX5*, *MKI67* and *EBF1* in healthy ($n=4$) and diagnosis B-ALL ($n=7$) bone marrow cell types. **d**, Heatmap representing cell type (row) z-score normalization of UMI counts for each individual CD19 candidate using scRNA-seq data from bone marrow with a confirmed B-ALL diagnosis¹⁶. Genes are ranked in descending order of individual candidate z-score across the Prog_B 2 column. **e**, Scatterplot showing correlation between *NUDT21* and *CD19* mRNA expression (UMI count) across Prog_B 1, Prog_B 2, naive B cells and memory B cells in diagnosis B-ALL bone marrow scRNA-seq data¹⁶. The r and P values are calculated on the basis of Pearson's correlation (95% confidence interval shown). P value was derived based on Student's t -test distribution; degree of freedom 416, 237, 132 and 2 for Prog_B 1, Prog_B 2, naive B and memory B, respectively. **f**, **g**, Flow cytometry (**f**) and immunoblot analysis (**g**) of NALM6 cells separated into quartiles on the basis of CD19 expression (Q1=CD19 lowest, Q4=CD19 highest). NUDT21, CD19 and β -actin are shown with a size ladder indicated.

mRNA. To assess whether *CD19* mRNA directly interacted with NUDT21, we performed enhanced crosslinking precipitation (eCLIP) of NUDT21-bound mRNA in Reh, 697, NALM6 and TMD8 cells. Significant NUDT21-binding peaks across the *CD19* mRNA were detected in Reh, 697 and NALM6, but not TMD8, cells (Fig. 4e and Extended Data Fig. 5b). Global analysis of

NUDT21 binding identified 26,609 significant peaks at 1,264 coding genes common to Reh, 697 and NALM6 cell lines (Extended Data Fig. 5c). NUDT21 binding did not enrich specifically at the 3'-UTRs, with most NUDT21 peaks mapped to intronic and exonic regions (Extended Data Fig. 5c and Supplementary Table 5), which may reflect direct interactions with RNA polymerase II

and the pre-mRNA processing machinery occurring during active transcription¹⁹.

Loss of NUDT21 may alter gene expression, partly through mRNA 3'-UTR shortening^{20,21}. We performed 3'-massive analysis of complementary DNA end sequencing (MACE-seq) to investigate the 3'-UTR changes that occurred after NUDT21 ablation in Cas9⁺Reh, Cas9⁺697, Cas9⁺NALM6 and Cas9⁺TMD8 cells. Quantification of alternative pA (QAPA) analysis indicated evidence of specific 3'-UTR shortening of *CD19* mRNA in sgNUDT21-mCherry⁺ Cas9⁺Reh, Cas9⁺697 and Cas9⁺NALM6 cells, but not Cas9⁺TMD8 cells, when compared with sgROSA-mCherry⁺ counterparts (Fig. 4f, Extended Data Fig. 5d,e and Supplementary Table 6). RNA-seq and DEG analysis of sgNUDT21-mCherry⁺ Cas9⁺Reh, Cas9⁺697, Cas9⁺NALM6, Cas9⁺K562, Cas9⁺TMD8 and Cas9⁺HG3 cells indicated 475 downregulated and 580 upregulated genes common to Cas9⁺Reh, -697 and -NALM6 ($P_{\text{adj}} < 0.1$) compared with sgROSA-mCherry⁺ cells (Supplementary Table 7). Expression of *CD19* mRNA was increased in sgNUDT21-mCherry⁺ Cas9⁺Reh, -697 and -NALM6 cells relative to sgROSA-mCherry⁺ controls; however, this effect was not observed in sgNUDT21-mCherry⁺ Cas9⁺K562, TMD8 and HG3 cells relative to sgROSA-mCherry⁺ controls (Fig. 4g). Assessment of alternative splicing using rMATS in sgNUDT21-mCherry⁺ Cas9⁺Reh, -697 and -NALM6 cells indicated that NUDT21 loss did not significantly (by threshold population stability index >0.2 and false discovery rate (FDR) <0.05) impact splicing of *CD19* mRNA compared with ROSA-mCherry⁺ cells (Extended Data Fig. 5f).

As 3'-UTR shortening may act to promote mRNA stability, we assessed *CD19* mRNA decay at multiple timepoints after actinomycin D treatment of sgNUDT21-mCherry⁺ Cas9⁺Reh, -NALM6, -HG3 and -TMD8 cells. Reverse transcription quantitative PCR (RT-qPCR) of the final coding exon (exon 14) in *CD19* mRNA found significantly delayed *CD19* mRNA decay in sgNUDT21-mCherry⁺ Reh and NALM6 cells, but not TMD8 and HG3 cells, compared with sgROSA-mCherry⁺ counterparts (Fig. 4h). Analysis of the genome-wide CRISPR screen dataset found that multiple pre-mRNA pA site regulators²², including CPSF6 and PCF11, modulated *CD19* abundance in Cas9⁺Reh, -697 and -NALM6 cells, but this effect was not clearly observed in Cas9⁺HG3 and -TMD8 cells (Fig. 4i). Targeting of factors that promote the usage of a distal pA site, such as NUDT21 and CPSF6, were associated with a high *CD19* CRISPR score (Fig. 4i), indicating increased *CD19* expression, whereas targeting of factors that facilitate proximal pA, such as PCF11 and CLP1, associated with a low *CD19* CRISPR score (Fig. 4i and Supplementary Table 1), suggesting loss of *CD19* expression relative to non-targeting control sgRNAs. To test the importance of the 3'-UTR in the expression of *CD19*, we ectopically

expressed full-length *CD19* (*CD19*^{FL}) and *CD19* lacking the 3'-UTR sequence (*CD19*^{ΔUTR}) in *CD19*^{dim} Cas9⁺Reh cells. *CD19*^{ΔUTR}-GFP⁺ Reh cells had significantly higher expression of *CD19* compared with *CD19*^{FL}-GFP⁺ cells (Fig. 4j,k), suggesting that the *CD19* 3'-UTR may limit the expression of *CD19* protein. As such, NUDT21 functions as a regulator of *CD19* 3'-UTR shortening and mRNA stability in transformed B cell progenitors.

Expression of target antigens, including *CD19* and *CD22*, directly correlates with CAR-T cell efficacy^{23–25}. To test whether loss of NUDT21 increased the abundance of *CD19* and, in turn, recognition and killing by *CD19*-directed therapeutics, we used vascularized, three-dimensional (3D) microfluidic devices that recapitulate many features of the B-ALL niche^{16,26}. Then, 5 days post-transduction of sgROSA-mCherry, sg*CD19*-mCherry or sgNUDT21-mCherry, Cas9⁺Reh cells were infused into human umbilical vein endothelial cell (HUV)EC-vascularized devices, together with either T cells harboring deletion of the endogenous *TRAC* locus (hereafter *TRAC* T cells) or T cells in which the endogenous *TRAC* locus was replaced with a *CD19*-specific CAR construct harboring a *CD28* co-stimulatory domain (hereafter *CD19* CAR-T cells)²⁷ (Extended Data Fig. 6a). After 24 hours of co-culture, we quantified the number of sgRNA-mCherry⁺ Cas9⁺Reh cells and the formation of *CD3*⁺ T cell-mCherry⁺ Reh cell synapses by confocal microscopy. We observed *CD3*⁺ T cell-mCherry⁺ Cas9⁺Reh cell synaptic junctions forming within 3D devices across all sgRNA conditions tested (Fig. 5a). Quantification of mCherry⁺ cell counts (normalized to *TRAC* T cell control conditions) showed that *CD19* CAR-T cells killed fewer sg*CD19*-mCherry⁺ Cas9⁺Reh cells than sgROSA-mCherry⁺ Cas9⁺Reh cells, although this did not reach statistical significance (Fig. 5b). Moreover, *CD19* CAR-T cells killed significantly more sgNUDT21-mCherry⁺ Cas9⁺Reh cells compared with sgROSA-mCherry⁺ and sg*CD19*-mCherry⁺ Cas9⁺Reh cells (Fig. 5d,e). In addition, confocal microscopy analysis showed CAR-T-treated sgNUDT21-mCherry⁺ Cas9⁺Reh cells engaged in a higher frequency of *CD3*⁺-mCherry⁺ synaptic events with larger synaptic size when compared with CAR-T-treated sgROSA-mCherry⁺ and sg*CD19*-mCherry⁺ Reh cells (Fig. 5c and Extended Data Fig. 6b).

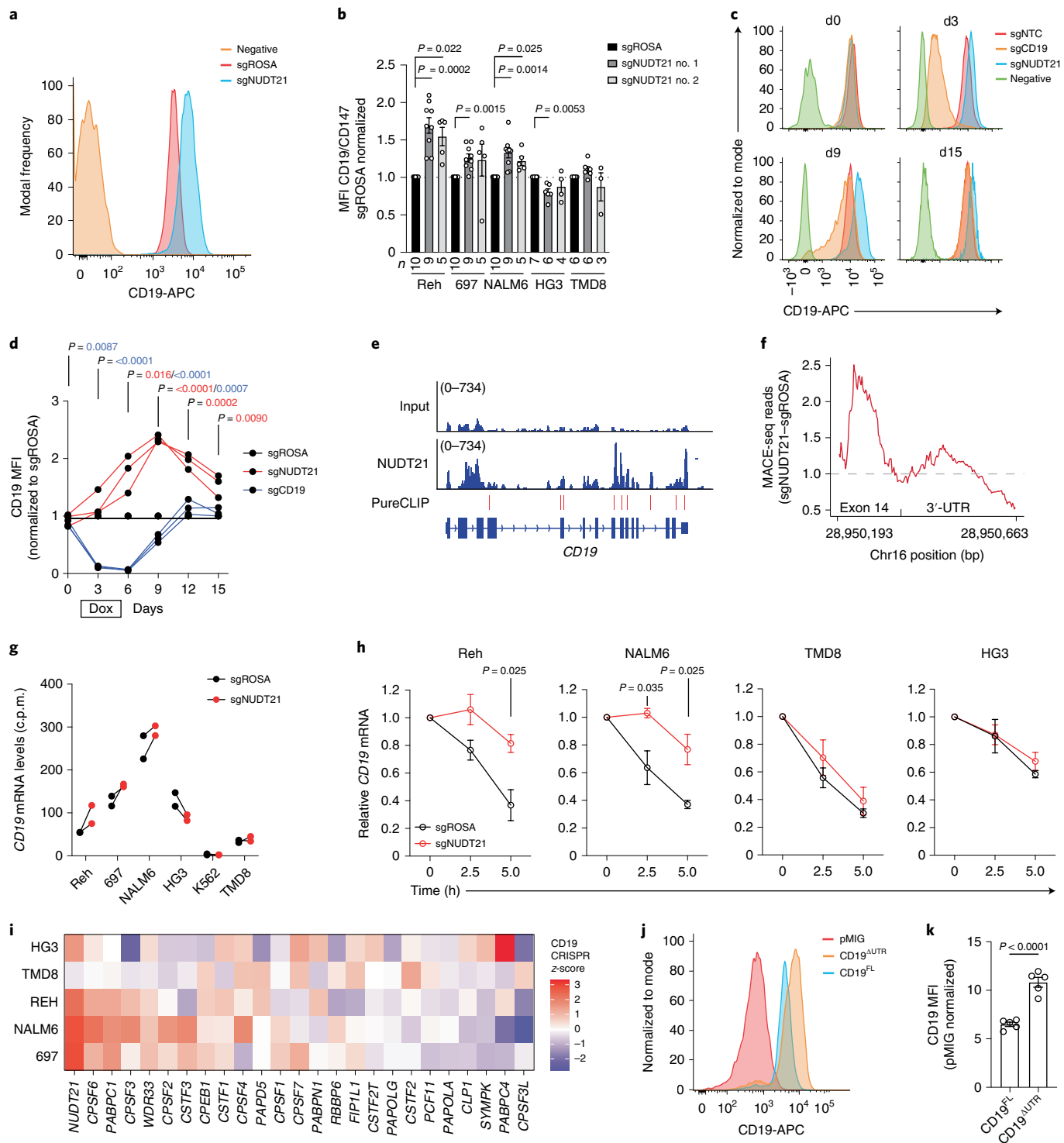
Blinatumomab is a *CD3*-*CD19* bispecific T cell engager formed by linking a distinct pair of immunoglobulin (Ig) variable domains capable of binding human *CD19* and *CD3*, respectively, with a short 25-amino acid linker sequence. On blinatumomab binding to *CD19*⁺ leukemic cells and proximal *CD3*⁺ T cells, *CD3* engagement results in T cell activation and killing of *CD19*⁺ leukemic blasts. We performed co-culture of sgROSA-mCherry⁺ Cas9⁺NALM6 cells or sgNUDT21-mCherry⁺ Cas9⁺NALM6 cells with human primary peripheral blood *CD8*⁺ T cells in the presence or absence of

Fig. 4 | NUDT21 directly represses *CD19* mRNA stability and protein expression. **a**, Representative flow cytometry of *CD19* expression 7 days post-transduction with sgNUDT21-mCherry or sgROSA-mCherry in Cas9⁺NALM6 cells. **b**, Histogram summary of surface *CD19* expression (normalized to *CD147*) 7 days post-transduction with sgNUDT21-mCherry in Cas9⁺Reh, Cas9⁺697, Cas9⁺NALM6, Cas9⁺HG3 and Cas9⁺TMD8 cells normalized to sgROSA-mCherry⁺. The *n* value indicates independent replicates shown for each condition (unpaired, two-sided Student's *t*-test, mean and s.e.m. shown; error bars represent s.e.m.). **c,d**, Representative flow cytometry (**c**) and summary of *CD19* expression (**d**) after expression of dox-induced sgNUDT21-GFP (red), sg*CD19*-GFP (blue) or sgNTC-GFP (black) in Cas13d⁺697 cells treated with dox for 6 days (days 0–6) followed by dox withdrawal (independent experiments with *n* = 3, unpaired, two-sided Student's *t*-test, mean and s.e.m. shown; error bars represent s.e.m.). **e**, The eCLIP read tracks spanning the *CD19* locus in NALM6 cells. PureCLIP significant peak signals are shown (red). **f**, MACE-seq log(FC) read counts of exon 14 and 3'-UTR junction (intron removed) in sgNUDT21 no. 1-mCherry⁺ Cas9⁺NALM6 cells and sgROSA-mCherry⁺ Cas9⁺NALM6 cells. **g**, Bulk RNA-seq of *CD19* mRNA in sgROSA-mCherry⁺ or sgNUDT21-mCherry⁺ Cas9⁺Reh, Cas9⁺697, Cas9⁺NALM6, Cas9⁺TMD8, Cas9⁺HG3 and Cas9⁺K562 cells. The experiment was performed in technical duplicate. **h**, Quantitative PCR analysis of *CD19* mRNA in sgROSA-mCherry⁺ or sgNUDT21-mCherry⁺ Cas9⁺Reh, Cas9⁺NALM6, Cas9⁺TMD8 and Cas9⁺HG3 cells at multiple timepoints (2.5 hours and 5 hours) post-actinomycin D treatment (independent experiments with *n* = 3, unpaired, two-sided Student's *t*-test, *P* value, mean and s.e.m. shown; *P* > 0.05 not shown). **i**, Heatmap of *CD19* CRISPR z-scores from Brunello genome-wide screen for Cas9⁺Reh, Cas9⁺697, Cas9⁺NALM6, Cas9⁺TMD8 and Cas9⁺HG3 cells, specifically highlighting polyadenylation regulators. **j,k**, Representative flow cytometry (**j**) and histogram summary (**k**) of *CD19*-APC expression after ectopic expression of *CD19*^{ΔUTR}-GFP and *CD19*^{FL}-GFP in Cas9⁺Reh cells. Histograms normalized to pMIG control (independent experiments with *n* = 5, unpaired two-sided Student's *t*-test, mean and s.e.m. shown; error bars represent s.e.m.).

blinatumomab and counted mCherry⁺ cells after 24 hours of co-culture as a measure of cell killing (normalized to vehicle-treated control). We found that CD8⁺ T cells combined with blinatumomab treatment led to significantly more killing of sgNUDT21-mCherry⁺ Cas9⁺NALM6 cells compared with sgROSA-mCherry⁺ Cas9⁺NALM6 cells (Fig. 5d and Extended Data Fig. 6c). In addition, limiting dilutions of CD19 CAR-T cells killed significantly more sgNUDT21-mCherry⁺ Cas9⁺697 cells compared with CAR-T-treated sgROSA-mCherry⁺ Cas9⁺697 cells (Fig. 5e and Extended Data Fig. 6d), indicating that NUDT21 limits

the sensitivity of leukemic cells to CD19-directed therapeutics *ex vivo*.

As CD19-directed blinatumomab or CAR-T cell treatment can select B-ALL blasts with reduced expression of CD19 (refs. 1,3,28), we assessed the temporal impact of CD19-directed treatment on expression of CD19 and NUDT21 in primary human B-ALL. We identified a B-ALL patient, identified as SJBALL042246 (ref. 3), who was subjected to blinatumomab treatment, having failed to respond to conventional chemotherapy. During blinatumomab therapy, this patient's leukemic blasts displayed reduced expression



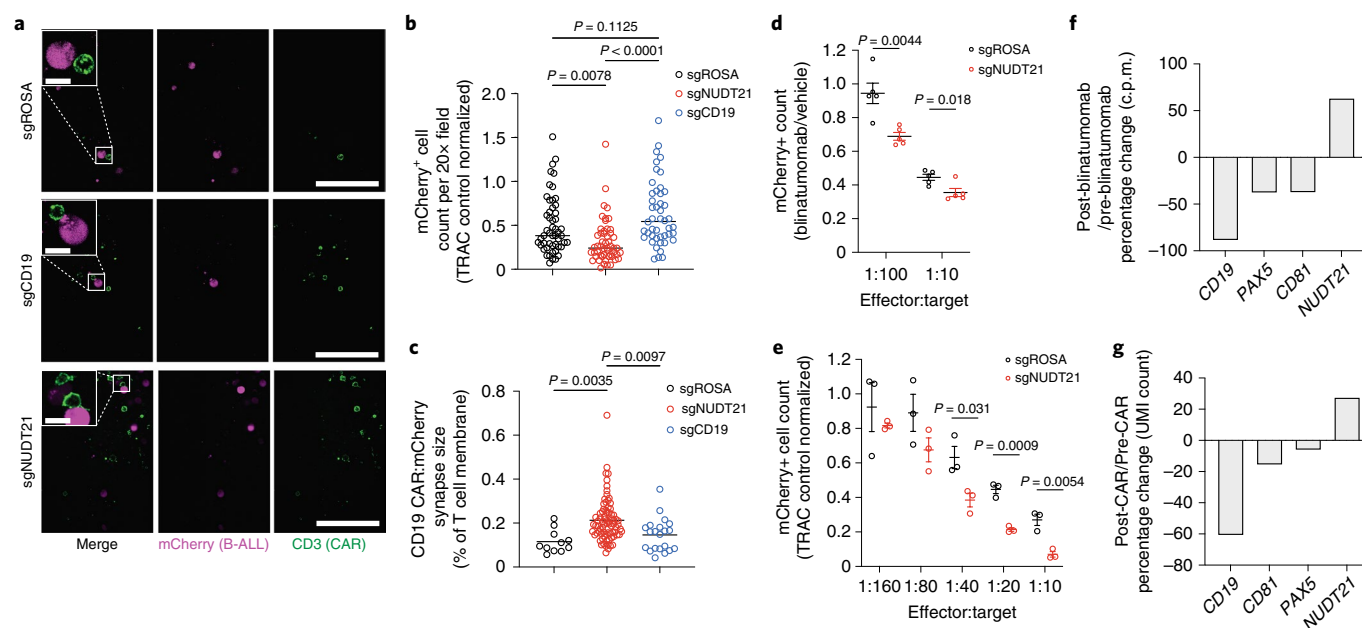


Fig. 5 | NUDT21 alters CD19-directed CAR-T cells and BiTE treatment responsiveness. **a**, Representative fluorescent images of CD3⁺ T cells (green) and sgROSA-mCherry, sgCD19-mCherry and sgNUDT21-mCherry, in Cas9⁺Reh cells (magenta) in a 3D microfluidic platform vascularized by HUVECs. Scale bars, 100 μ m and 10 μ m for magnified inset images. **b**, Quantitative comparison of live sgROSA-mCherry, sgCD19-mCherry and sgNUDT21-mCherry Cas9⁺Reh cell count after a 24-hour co-culture with TRAC CAR-T cell on-chip. Results from CAR-T groups were normalized to the respective TRAC control T cells (four independent experiments each with four devices, where each device has two to four random fields quantified ($n \geq 46$ images), unpaired Student's *t*-test, mean and s.e.m. shown). **c**, Quantitative comparison of synapse size between TRAC CAR-T cell and sgROSA-mCherry, sgCD19-mCherry and sgNUDT21-mCherry Cas9⁺Reh cells (four independent experiments each with four devices, where each device has about ten random fields quantified ($n \geq 11$ synapse events); unpaired Student's *t*-test, mean and s.e.m. shown). **d**, SgRNA-mCherry⁺ Cas9⁺NALM6 cells counts after 24 hours of co-culture with primary peripheral blood CD8⁺ T cells in the presence or absence of 0.5 ng ml⁻¹ of blinatumomab normalized to vehicle-treated control (independent experiments with $n = 5$, unpaired, two-sided Student's *t*-test, mean and s.e.m. shown). **e**, Number of sgRNA-mCherry⁺ 697 cells after 24-hour co-culture with TRAC T cells or TRAC CD19 CAR T cells. Cell counts were normalized to TRAC T cell counts (independent experiments with $n = 3$, unpaired, two-sided Student's *t*-test, mean and s.e.m. shown). **f, g**, Histogram representation of percentage change in gene expression of CD19, PAX5, CD81 and NUDT21 for bulk RNA-seq pre- and post-blinatumomab treatment in patient data (SJBALL042246, Zhao et al.²¹) (**f**) and scRNA-seq UMI counts pre- and post-CAR-T CD34⁺ treatment in B-ALL patient data (Rabilloud et al.²⁸) (**g**).

of CD19, leading to disease relapse. Using publicly available, leukemic blast-enriched, primary RNA-seq data generated from SJBALL042246 primary bone marrow mononuclear cells before blinatumomab infusion and at disease relapse, we compared the expression of CD19, PAX5, CD81 and NUDT21 mRNA throughout therapy (Fig. 5f). At disease relapse, CD19, PAX5 and CD81 mRNA in leukemia blasts was reduced compared with pre-treatment, whereas NUDT21 mRNA was elevated at disease relapse relative to pre-treatment leukemia cell gene expression (Fig. 5f). In addition, we performed analysis of public scRNA-seq data from a single B-ALL patient²⁸ obtained before CD19 CAR-T infusion and at disease relapse of a CD19^{dim} leukemia after CD19 CAR-T treatment. Consistent with the original scRNA-seq data analysis²⁸, we identified a distinct group of CD19⁺CD34⁺ cells pre-CAR-T cell treatment and CD19^{dim}CD34⁺ leukemic cells at disease relapse after CAR-T cell treatment (Extended Data Fig. 7a,b). At disease relapse, CD19^{dim}CD34⁺ leukemic blasts showed reduced expression of CD19, PAX5 and CD81 mRNA and increased expression of NUDT21 mRNA compared with pre-CAR-T treatment CD19⁺CD34⁺ leukemic blasts (Fig. 5g and Extended Data Fig. 7a,b). These data indicate that NUDT21 expression may be modulated after CD19-directed therapeutic exposure in primary human B-ALL; however, additional clinical samples with temporal gene expression profiling will be needed to validate these trends.

Surface-guided genome-wide CRISPR screening approaches are an effective tool to identify critical regulators of surface protein abundance in hematological malignancies associated with cellular differentiation²⁹ and immunotherapy responsiveness³⁰. The present study identified regulators of CD19 expression in transformed B cell progenitors at steady state, including the CD19 transcriptional activator ZNF143 and the CD19 repressor NUDT21. Public access to gene expression data from blast-purified B-ALL, CD19 CAR-T cell-treated samples remains limited. In the present study, we highlighted two clinical examples of concurrent reduction of CD19 mRNA and increased expression of NUDT21 mRNA after CD19-directed therapy with blinatumomab and in CD19 CAR-T-treated B-ALL patients. These early clinical correlations may inform future kinetic analysis of CD19 regulators throughout immunotherapy.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41590-022-01314-y>.

Received: 5 January 2022; Accepted: 11 August 2022;
Published online: 22 September 2022

References

- Sotillo, E. et al. Convergence of acquired mutations and alternative splicing of CD19 enables resistance to CART-19 immunotherapy. *Cancer Discov.* **5**, 1282–1295 (2015).
- Orlando, E. J. et al. Genetic mechanisms of target antigen loss in CAR19 therapy of acute lymphoblastic leukemia. *Nat. Med.* **24**, 1504–1506 (2018).
- Zhao, Y. et al. Tumor-intrinsic and -extrinsic determinants of response to blinatumomab in adults with B-ALL. *Blood* **137**, 471–484 (2021).
- Sanson, K. R. et al. Optimized libraries for CRISPR–Cas9 genetic screens with multiple modalities. *Nat. Commun.* **9**, 5416 (2018).
- Bradbury, L. E., Kansas, G. S., Levy, S., Evans, R. L. & Tedder, T. F. The CD19/CD21 signal transducing complex of human B lymphocytes includes the target of antiproliferative antibody-1 and Leu-13 molecules. *J. Immunol.* **149**, 2841–2850 (1992).
- Shoham, T. et al. The tetraspanin CD81 regulates the expression of CD19 during B cell development in a postendoplasmic reticulum compartment. *J. Immunol.* **171**, 4062–4072 (2003).
- David, V., Hochstenbach, F., Rajagopalan, S. & Brenner, M. B. Interaction with newly synthesized and retained proteins in the endoplasmic reticulum suggests a chaperone function for human integral membrane protein IP90 (calnexin). *J. Biol. Chem.* **268**, 9585–9592 (1993).
- Lang, S. et al. Different effects of Sec61alpha, Sec62 and Sec63 depletion on transport of polypeptides into the endoplasmic reticulum of mammalian cells. *J. Cell Sci.* **125**, 1958–1969 (2012).
- Kozmik, Z., Wang, S., Dorfler, P., Adams, B. & Busslinger, M. The promoter of the CD19 gene is a target for the B-cell-specific transcription factor BSAP. *Mol. Cell Biol.* **12**, 2662–2672 (1992).
- Cortés-López, M. et al. High-throughput mutagenesis identifies mutations and RNA-binding proteins controlling CD19 splicing and CART-19 therapy resistance. Preprint at *bioRxiv* <https://doi.org/2021.2010.2008.463671> (2021).
- Weintraub, A. S. et al. YY1 Is a structural regulator of enhancer–promoter loops. *Cell* **171**, 1573–1588.e1528 (2017).
- Bailey, S. D. et al. ZNF143 provides sequence specificity to secure chromatin interactions at gene promoters. *Nat. Commun.* **2**, 6186 (2015).
- Myslinski, E., Krol, A. & Carbon, P. ZNF76 and ZNF143 are two human homologs of the transcriptional activator Staf. *J. Biol. Chem.* **273**, 21998–22006 (1998).
- Consortium, E. P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
- Zhou, Q. et al. ZNF143 mediates CTCF-bound promoter–enhancer loops required for murine hematopoietic stem and progenitor cell function. *Nat. Commun.* **12**, 43 (2021).
- Witkowski, M. T. et al. Extensive remodeling of the immune microenvironment in B cell acute lymphoblastic leukemia. *Cancer Cell* **37**, 867–882 (2020).
- Wessels, H. H. et al. Massively parallel Cas13 screens reveal principles for guide RNA design. *Nat. Biotechnol.* **38**, 722–727 (2020).
- Yang, Q., Gilmartin, G. M. & Doublié, S. Structural basis of UGUA recognition by the Nudix protein CFI(m)25 and implications for a regulatory role in mRNA 3' processing. *Proc. Natl Acad. Sci. USA* **107**, 10062–10067 (2010).
- Bentley, D. L. Coupling mRNA processing with transcription in time and space. *Nat. Rev. Genet.* **15**, 163–175 (2014).
- Brumbaugh, J. et al. Nudt21 controls cell fate by connecting alternative polyadenylation to chromatin signaling. *Cell* **172**, 629–631 (2018).
- Sun, M. et al. NUDT21 regulates 3'-UTR length and microRNA-mediated gene silencing in hepatocellular carcinoma. *Cancer Lett.* **410**, 158–168 (2017).
- Nourse, J., Spada, S. & Danckwardt, S. Emerging roles of RNA 3'-end cleavage and polyadenylation in pathogenesis, diagnosis and therapy of human disorders. *Biomolecules* **10**, 915 (2020).
- Majzner, R. G. et al. Tuning the antigen density requirement for CAR T-cell activity. *Cancer Discov.* **10**, 702–723 (2020).
- Ramakrishna, S. et al. Modulation of target antigen density improves CAR T-cell functionality and persistence. *Clin. Cancer Res.* **25**, 5329–5341 (2019).
- Pillai, V. et al. CAR T-cell therapy is effective for CD19-dim B-lymphoblastic leukemia but is impacted by prior blinatumomab therapy. *Blood Adv.* **3**, 3539–3549 (2019).
- Ma, C. et al. Leukemia-on-a-chip: dissecting the chemoresistance mechanisms in B cell acute lymphoblastic leukemia bone marrow niche. *Sci. Adv.* **6**, eaba5536 (2020).
- Eyquem, J. et al. Targeting a CAR to the TRAC locus with CRISPR/Cas9 enhances tumour rejection. *Nature* **543**, 113–117 (2017).
- Rabilloud, T. et al. Single-cell profiling identifies pre-existing CD19-negative subclones in a B-ALL patient with CD19-negative relapse after CAR-T therapy. *Nat. Commun.* **12**, 865 (2021).
- Wang, E. et al. Surface antigen-guided CRISPR screens identify regulators of myeloid leukemia differentiation. *Cell Stem Cell* **28**, 718–731 (2021).
- Ramkumar, P. et al. CRISPR-based screens uncover determinants of immunotherapy response in multiple myeloma. *Blood Adv.* **4**, 2899–2911 (2020).

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

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2022

Methods

Materials availability. Human and murine cell lines generated and plasmids used in the present study are available on request from the lead contact.

Animal studies. Mice were maintained in individual ventilated cages and fed with autoclaved food and water at NYU School of Medicine Animal Facility. All animal experiments were done in accordance with approved protocols from the institutional animal care and use committees (IACUCs), according to national and institutional guidelines. All animal experiments were performed in accordance with protocols approved by the NYU IACUC (Aifantis protocol no. 160411, IA16-00008).

Cell lines and culture. Human cell lines Reh, 697, NALM6, HG3, TMD8, GM12878 and K562 were cultured in RPMI-1640 medium supplemented with 20% fetal bovine serum (FBS), 55 μ M 2-mercaptoethanol (Gibco, catalog no. 21985023) and penicillin–streptomycin (Gibco, catalog no. 15140122). HEK293T cells were purchased from American Type Culture Collection (ATCC, catalog no. CRL-1573), and Platinum-E cells were purchased from Cell Biolabs (catalog no. RV-101). HEK293T and Platinum-E cells were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% FBS and penicillin–streptomycin (Gibco). OP9 cells were purchased from ATCC (catalog no. CRL-2749) and cultured in Isocove's modified DMEM (IMDM) supplemented with L-glutamine (Gibco, catalog no. 31980030), 15% FBS and penicillin–streptomycin (Gibco), with supplementation of recombinant murine interleukin (IL)-7 (Peprotech, catalog no. 217-17), Flt3L (Peprotech, catalog no. 250-31L) and stem cell factor (SCF; Peprotech, catalog no. 250-03) where indicated. All cell lines were maintained at 37°C and 5% CO₂ in a humidified atmosphere. Tissue culture reagents were purchased from Gibco. Reh, 697 and NALM6 cell lines were a gift from W. Carroll's laboratory at NYU School of Medicine. HG3 and TMD8 were a gift from H.-G. Wendel's lab at Memorial Sloan Kettering Cancer Center (MSKCC).

All lentiviral transfections were performed in HEK293T cells using polyethylenimine (PEI) reagent at 4:2:3 ratios of sgRNA/Cas9/Cas13d construct:pVSVG:pPax2 in OPTI-MEM solution. All retroviral transfections were performed in Platinum-E cells using PEI reagent with 10 μ g of transfer plasmid in OPTI-MEM solution. All amphitrophic retroviral transfections were performed in PEI reagent at 5:2:5 ratios of MSCV-based construct:pVSVG:pCL-Ampho retrovirus packaging construct (Novus Biologicals, catalog no. NBP2-29541) in OPTI-MEM solution. Viral supernatant was collected 48 hours post-transfection. Spin infections were performed at room temperature and 1,500 g for 90 min with polybrene reagent (Fisher Scientific). Reh, 697, NALM6, K562, HG3 and TMD8 Cas9-expressing cell lines were generated by lentiviral transduction with retroviral Cas9-2A-blast (Addgene, plasmid no. 73310). Human GM12878 Cas9 lines were generated by lentiviral transduction with Cas9-2A-EGFP (Addgene plasmid no. 63592). The 697 Cas13d-expressing cell line was generated by transduction with lentiviral tet-inducible RfxCas13d (Addgene, plasmid no. 138149). Doxycycline (Sigma-Aldrich) treatment was performed at 0.5 μ g ml⁻¹ supplemented every 2 days in culture.

Cell lines were routinely monitored for *Mycoplasma* contamination by PCR using ATCC Universal Mycoplasma Detection Kit (20,000–1,012,000). Each leukemia line was validated by short tandem repeat analysis through ATCC. HEK293T were authenticated by the supplier (ATCC Cell Line Authentication Service Sanger Sequencing).

Murine Nudt21-conditional deletion B cell progenitor line generation. To generate condition Nudt21-knockout-transformed, murine B cell progenitor lines, whole bone marrow cells from both femurs and tibias were harvested from 8- to 12-week-old ROSA26-Cre^{ERT2};*Nudt21*^{+/+} and ROSA26-Cre^{ERT2};*Nudt21*^{fl/fl} mice. Bone marrow cell transduction with BCR-ABL1-expressing retroviral constructs (Addgene, plasmid no. 38185 (ref. 31)) was performed in IMDM buffer with GlutaMax (Gibco) and 15% FBS, murine SCF (MSCF; 100 ng ml⁻¹, Peprotech), murine Flt3 ligand (50 ng ml⁻¹, Peprotech), murine IL-7 (10 ng ml⁻¹) and polybrene (1 μ g ml⁻¹) at a cell concentration of 1 $\times$ 10⁶ cells ml⁻¹ and incubated overnight at 37°C with 5% CO₂. Transduced cells were cultured on OP9 stromal cells for 7 days in IMDM buffer with GlutaMax (Gibco) and 15% FBS, MSCF (100 ng ml⁻¹, Peprotech), murine Flt3 ligand (50 ng ml⁻¹, Peprotech), murine IL-7 (10 ng ml⁻¹), 55 μ M 2-mercaptoethanol and penicillin–streptomycin. After 7 days, SCF and Flt3L were withdrawn from culture medium and cultured on OP9 stromal cells for a further 21 days before physical separation of hematopoietic cells from OP9 stromal cells. The resulting suspension cell lines were cultured continuously in IMDM buffer with GlutaMax (Gibco) and 15% FBS and penicillin–streptomycin. 4-OHT (Sigma-Aldrich, catalog no. 68392-35-8) was supplemented at 50 nM daily for 5 days consecutively.

Flow cytometry and cell sorting. Cell-line suspensions were subjected to centrifugation at 400 g for 5 min and 4°C, and the supernatant was discarded. Pellets were re-suspended in 1 ml of phosphate-buffered saline (PBS) with 2% FBS and filtered using a 70- μ m nylon mesh (Thermo Fisher Scientific) with residual cell clumps discarded. After centrifuging at 400 g and 4°C for 5 min, the supernatant was discarded and the cell pellet was re-suspended in PBS (Corning)

with 2% FBS and placed on ice before staining for FACS. Single-cell suspensions were re-suspended in PBS with 2% FBS and Fc-blocking agents and then incubated on ice for 10 min. For human cell lines, we used Human TruStain FcX Blocking Solution (BioLegend, catalog no. 422301) and murine cell lines, Mouse BD Fc Block (BD Pharmingen, catalog no. 553142). Cell suspensions were centrifuged at 400 g and 4°C for 5 min and the supernatant discarded. Cell pellets were re-suspended in PBS with 2% FBS and antibody cocktails and incubated on ice for 30 min. For human cell lines, we used antigen-presenting cell (APC)-conjugated anti-human CD19 (BioLegend, 1:200 concentration, catalog no. 302212), PerCP-Cy5.5-conjugated anti-human CD147 (BioLegend, catalog no. 306219, 1:1,000) and FITC-conjugated anti-human CD3 (eBioscience, catalog no. 11-0038-42, 1:200). For murine studies, we used PE-Cy7-conjugated anti-mouse CD19 (eBioscience, catalog no. 25-0193-82, 1:300). After incubation, cell suspensions were subjected to two wash cycles involving re-suspension of cell pellets in 3 ml of PBS with 2% FBS, followed by centrifugation at 400 g and 4°C for 5 min, followed by supernatant removal. Finally, samples were re-suspended in 300 μ l of PBS with 2% FBS per 2 $\times$ 10⁶ cells, supplemented with DAPI (0.5 μ g ml⁻¹, Sigma-Aldrich) for live/dead staining. For CRISPR screen cell sorting, all samples were gated based on forward and side scatter, followed by exclusion of doublets, then gated on viable cells (DAPI^{low}) and sgRNA-expressing cells (GFP⁺), followed by CD19-APC surface abundance (top and bottom 15% of fluorescence) using the SY3200 highly automated parallel sorting (HAPS) cell sorter (Sony). Flow cytometry utilized the aforementioned gating strategy and was performed using the BD Fortessa. All flow cytometry data analysis was performed using Flowjo v.10 (Treestar, BD Biosciences).

CRISPR libraries and cloning. We designed two pooled sgRNA libraries for CRISPR-based exon mutagenesis targeting CD19 activators (CD19: 100 sgRNAs, CD81: 64 sgRNAs and ZNF143: 100 sgRNAs) and CD19 repressor, NUDT21 (47 sgRNAs), including 10 non-targeting sgRNAs (Supplementary Table 8). Customized sgRNAs were designed using <http://benchling.com> with a quality score of ≥ 70 to minimize off-target effects. Pooled sgRNA oligos were then synthesized by Twist Bioscience (<https://twistbioscience.com>) on a 12,000 array and amplified using customized array primers, followed by subcloning into a lentiviral sgRNA, GFP-tagged vector (LRG; Addgene, plasmid no. 65656) as previously described³². All sgRNA sequences used in the present study are provided (Supplementary Table 8). Genome-wide CRISPR screens were performed using the human Brunello knockout (KO) library (Addgene, catalog no. 73179) targeting 19,114 genes with a total of 77,441 sgRNAs (4 sgRNAs per gene). Individual sgRNA cloning was performed as described in the protocol developed by F. Zhang's laboratory³³ and subcloned into the lentiviral sgRNA vectors. Cas13d sgRNAs were cloned into RfxCas13d sgRNA vector (Addgene, catalog no. 138150), and Cas9-based individual sgRNAs were cloned into LRCherry2.1 (Addgene, catalog no. 108099) (Supplementary Table 8).

Antigen-based CRISPR screens. For CRISPR screens, Cas9-expressing cells were infected with either CRISPR exon mutagenesis (the present study) screening or Brunello genome-wide library at a low multiplicity of infection (MOI ~0.3). At day 12 post-transduction, Cas9-expressing cells were incubated with APC anti-human CD19 antibody (BioLegend, catalog no. 302212 H1B19 clone), and then cell sorting was performed using the SY3200 HAPS cell sorter (Sony) into a CD19^{high} bin (top 15%) and CD19^{low} bin (bottom 15%). Approximately 2–3 million cells from each bin were collected, and genomic DNA extraction was performed using QIAGEN DNA kit (catalog no. 51306) according to the manufacturer's protocol. PCR amplification of the Brunello library was performed based on a previous study³⁴. CRISPR mutagenesis libraries were amplified with ExTaq (Takara Bio) using the following PCR cycling conditions: an initial 1 min at 95°C, followed by 30 s at 95°C, 30 s at 53°C, 30 s at 72°C, for 26 cycles; and a final 10-min extension at 72°C. Customized P5 (staggering primers) and P7 (unique barcode primers) (Supplementary Table 9) were used to generate barcoded libraries that were subsequently sequenced using single-end 50-bp HiSeq 4000 (Illumina). Deconvolution of sgRNA reads was performed using a customized script that counts the number of reads per sgRNA from the FASTQ files by first searching for the CACCG identifier sequence found in the vector 5' to all sgRNAs. The next 20 nucleotides are then mapped to a reference file containing all the sgRNA sequences. For each sgRNA, reads were normalized to reads per million followed by log₂(transformed) by adding 1 to all values (c.p.m. + 1). The CD19 CRISPR scores were calculated by the log₂(fold-change) (log₂(FC)) (CD19^{high}/CD19^{low}) for each sgRNA.

Immunoblotting. For immunoblot analysis, cell lines were pelleted and lysed using RIPA Lysis and Extraction buffer (Thermo Fisher Scientific, catalog no. 89900). The lysates were boiled with Laemmli buffer, resolved by sodium dodecylsulfate–polyacrylamide gel electrophoresis, transferred to polyvinylidene membranes and the proteins visualized by immunoblotting. The following antibodies were used for immunoblot analysis: CD19 (1:500, Cell Signaling Technology, catalog no. 3574), NUDT21 (1:100, Santa Cruz Biotechnology, catalog no. sc-81109), YY1 (1:500, Santa Cruz Biotechnology, catalog no. H-10 sc-7341), ZNF143 (1:1,000, Novus Biologicals, catalog no. 2B4 H00007702-M01) and actin (1:1,000, Millipore, catalog no.).

Enhanced CLIP. The eCLIP studies were performed by Eclipse Bioinnovations Inc. according to the published single-end eCLIP protocol³⁵ with the following modifications: Cas9-expressing Reh, 697, NALM6 and TMD8 cells were ultraviolet crosslinked at 400 mJ cm⁻² with 254-nm radiation. Cells were lysed using 1 ml of eCLIP lysis mix and were subjected to two rounds of sonication for 4 min with 30 s ON/OFF at 75% amplitude. Then, 5 µg of pre-validated NUDT21 (Santa Cruz Biotechnology, catalog no. sc-81109) antibody was pre-coupled to 50 µl of anti-rabbit IgG dynabeads (Thermo Fisher Scientific), added to lysate equivalent to 100 µg of RNA and incubated overnight at 4°C. Before immunoprecipitation, 2% of the sample was taken as the paired input sample, with the remainder magnetically separated and washed with eCLIP high-stringency wash buffers. Immunoprecipitation and input samples were cut from the membrane at the relative band size to 75 kDa above. RNA adapter ligation, immunoprecipitation–immunoblot, reverse transcription, DNA adapter ligation and PCR amplification were performed as previously described.

MACE-seq. Cas9⁺Reh, -697 and -NALM6 cell lines were transduced with LRCherry2.1–sgROSA/NUDT21 and cultured for 7 d. The mCherry⁺ cells were isolated and RNA was extracted using the RNeasy Plus Mini Kit (QIAGEN, catalog no. 74316). MACE-seq was performed by GenXPro GmbH in Frankfurt am Main using the MACE-seq kit according to the manual of the manufacturer. Briefly, cDNA was generated from fragmented RNA with barcoded poly(A) primers during reverse transcription. After second-strand synthesis and 5'-adapter integration, a PCR with a minimum number of cycles was used to produce a library that was sequenced on an Illumina NextSeq500 machine with 1 × 75 bp³⁶.

NUDT21 rescue and ectopic CD19 variant expression. NUDT21 codon optimization was performed based on the NUDT21 full-length cDNA sequence (Ensembl NUDT21-201, catalog no. ENST00000300291.10) to exclude EcoRI, XhoI and sgNUDT21 no.1 target sequences and synthesized with terminal 5'-EcoRI and 3'-XhoI sites using gBlock synthesis (IDT; Supplementary Table 10). NUDT21 sgRNA-resistant (NUDT21^{sgRes}) was subcloned into EV pMSCV-IRES-GFP (Addgene, catalog no. 9044) via EcoRI and XhoI sites. Cas9⁺NALM6 B-ALL cells were co-transduced with amphotrophic virus harboring pMSCV-IRES-GFP; pLRCherryv2.1–sgROSA, pMSCV-NUDT21^{sgRes}–IRES-GFP; pLRCherryv2.1–sgROSA, pMSCV-IRES-GFP; pLRCherryv2.1–sgNUDT21 no.1 or pMSCV-NUDT21^{sgRes}–IRES-GFP; pLRCherryv2.1–sgNUDT21 no.1 and cultured for 6 days before flow cytometry and then cultured for a further 5 days. Fitness was assessed by calculating the percentage of GFP⁺mCherry⁺ cells at days 6 and 11 post-transduction.

Full-length (FL) CD19 cDNA (Ensembl CD19-202, catalog no. ENST00000538922.8), including (CD19^{FL}) or excluding (CD19^{ΔUTR}) the 3'-UTR sequence, was synthesized using gBlock synthesis (IDT) and then subcloned into pMSCV-IRES-GFP (Addgene, catalog no. 9044) via Gibson Assembly (New England Biolabs, catalog no. E2611) (Supplementary Table 10). Cas9⁺Reh cells were co-transduced with amphotrophic virus harboring pMSCV-IRES-GFP, pMSCV-CD19^{ΔUTR}–IRES-GFP or pMSCV-CD19^{FL}–IRES-GFP. Cells were cultured for 72 hours before flow cytometry.

RNA stability. Cas9⁺Reh, -NALM6, -TMD8 and -HG3 cell lines were transduced with LRCherry2.1–sgROSA/NUDT21 and cultured for 7 days. Cells were treated with 5 µg ml⁻¹ of actinomycin D (Cell Signaling Technology, catalog no. 15021S) at 37°C and harvested at the indicated timepoints. RNA was extracted using the RNeasy Plus Mini Kit (QIAGEN, catalog no. 74316) according to the manufacturer's instructions and converted to cDNA using the High-Capacity RNA-to-cDNA Kit (Life Technologies, catalog no. 4387406). Then, cDNA was subjected to qPCR, and gene expression was measured using LightCycler 480 SYBR Green I Master (Roche Diagnostics, catalog no. 04707516001). Data were analyzed by the $\Delta\Delta C_t$ method normalizing to 18S rRNA. The qPCR primers used were: 18S rRNA: forward 5'-ACCGCAGCTAGGAATAATGGA-3' 18S rRNA: reverse 5'-GCCTCAGTTCGGAACCA-3' CD19: forward 5'-TGAAGACCTCGAGCAGATGA-3' CD19: reverse 5'-CATGCACATCCTAAGCAAC-3'.

Generating CAR-T and TRAC T cells for ex vivo studies. For generating CAR-T and TRAC T cells for CAR-T cell studies, AAV (adeno-associated viral vector)–inverted terminal repeat (ITR) plasmids containing the 1928⁺ CAR and TRAC-targeting homology arms for homology-directed repair were used as previously described³⁷. The AAV-ITR-containing plasmid was packaged into AAV6 using PE-based co-transfection of HEK293T cells with pHelper and pAAV Rep-Cap plasmids. Viral particles were extracted from cells and purified using iodixanol-based density gradient ultracentrifugation. AAV titration was performed by qPCR after treating samples with DNase I (New England Biolabs) and proteinase K (QIAGEN), using primers targeting the left homology arm (forward: CTTTGCTGGGCTTTTCC; reverse: CTGCCACTCAAGGAAACCT). Quantitative PCR was performed using SsoFast EvaGreen Supermix (BioRad) on a StepOnePlus Real-Time PCR System (Applied Biosystems). Leukopacks with peripheral blood mononuclear cells were obtained from STEMCELL Technologies. T lymphocytes were purified using an EasySep Human T Cell

Isolation Kit (STEMCELL Technologies). T cells were activated with Dynabeads Human T-Activator CD3/CD28 (Thermo Fisher Scientific) (1:1 beads:cell) in X-VIVO 15 medium (Lonza) supplemented with 5% human serum (Gemini Bioproducts), penicillin–streptomycin (Thermo Fisher Scientific, 50 U ml⁻¹), IL-7 (Miltenyi Biotec, 5 ng ml⁻¹) and IL-15 (Miltenyi Biotec, 5 ng ml⁻¹) and cultured at 10⁶ cells ml⁻¹. Medium was exchanged every 2–3 days and cells were re-suspended at 10⁶ cells ml⁻¹.

After 48 hours of T cell activation, cells were detached from CD3/CD28 Dynabeads and the beads were magnetically removed. T cells were electroporated with ribonucleoprotein (RNP) using a 4D-Nucleofector 96-well unit (Lonza). RNP for each electroporation reaction was generated by co-incubating 60 pmol of recombinant Cas9 protein (QB3 MacroLab) with 120 pmol of TRAC sgRNA (Synthego, CAGGGUUCUGGAUAUCUGU) at 37°C for 15 min. Cells were re-suspended in P3 primary cell solution (Lonza) (2 × 10⁶ live cells per electroporation) and mixed with RNP, followed by electroporation using the EH115 Nucleofector protocol. Cells were then diluted into serum-free medium (2 × 10⁶ cells ml⁻¹) and incubated at 37°C and 5% CO₂. Recombinant AAV6 donor vector was added to the culture 30–60 min after electroporation at an MOI of 5 × 10⁴ and the cells were incubated overnight. The next day, the serum-free, AAV-containing medium was removed and the cells were re-suspended in fresh complete medium and expanded using standard culture conditions (37°C, 5% CO₂ and complete medium replenished as needed to maintain a density of 1 × 10⁶ cells ml⁻¹ every 2–3 d). Knockout and knock-in efficiency were evaluated by staining for the T cell receptor (TCR) with an anti-TCR α/β antibody (Miltenyi Biotec) and staining for CAR with a goat anti-mouse F(ab')₂ (Jackson ImmunoResearch, catalog no. 115-606-003), and flow cytometry was conducted on a BD LSRFortessa X-50 instrument.

On-chip measurement of CAR-T killing efficacy and synapse formation capability. The vascularized leukemia marrow microphysiological system was engineered following our previous protocol^{6,26}. Briefly, the microfluidic device with three distinct functional regions (that is, a central sinus region, an inner ring region and the outer ring channels) was fabricated using a standard soft lithography replica-molding technique^{16,26}. The central sinus region was vascularized by HUVECs (Lonza, catalog no. C2519A), whereas the inner ring region and the outer ring channels were, respectively, loaded with HUVECs and fibroblast cells (Lonza, catalog no. CC-2512) embedded in 3 mg ml⁻¹ of fibrin hydrogel (Sigma-Aldrich, catalog no. G2500) to maintain 3D culture. The ex vivo devices were cultured with a mixture cell culture medium of EGM-2 (Lonza, catalog no. CC-3162) and FGM-2 (Lonza, catalog no. CC-3132) at 2:1 v:v with 25 ng ml⁻¹ of vascular endothelial growth factor (RayBiotech, catalog no. 230-00012) and monitored for about 7 days to ensure the formation of a vascular network. Once the microvessel network had been successfully established, 10,000 mCherry⁺ B-ALL cells and 10,000 T cells (TRAC CAR and TRAC KO, respectively²⁷) were mixed and loaded from the central sinus into the whole microvessel network and cultured for 24 hours with RPMI-1640. The devices were then imaged with a ×20 objective using Zeiss microscopy integrated with a live-cell incubator and, using National Institutes for Health (NIH) ImageJ, the number of mCherry⁺ B-ALL cells was manually counted in each image taken. After this, the microfluidic devices were blocked with Human TruStain FcX (BioLegend, catalog no. 422302) and stained with APC-conjugated anti-human CD3 (BioLegend, catalog no. 317318) for 4 hours at 4°C, following the manufacturer's instruction. After 3× washes, the devices were imaged using Nikon Spinning Disk confocal microscopy with a ×40 objective; the number of mCherry⁺ B-ALL cells forming synapse/contact with CD3 T cells was manually counted in each image taken, and the synapse size was quantified by calculating the ratio of the length of synapse to the perimeter of CAR-T cells using NIH ImageJ. The significance of results from different groups was compared using GraphPad Prism. To quantify viability of leukemia cells after CAR T cell treatment, two to four image fields (×20) were randomly chosen for each device, and each independent experiment has four independent devices. To quantify the synapse formation, approximately ten image fields (×40) were randomly chosen for each device, and each independent experiment has four independent devices. To quantify the synapse size formed between TRAC CAR-T cell and different sgRNA-expressing Reh B-ALL, all the synapses identified were collected and quantified.

In vitro blinatumomab and CAR-T killing assays. Cas9⁺NALM6 and -697 cells were transduced with pLRCherryv2.1–sgROSA or pLRCherryv2.1–sgNUDT21 lentivirus and cultured for 6 days in RPMI-1640 supplemented with 20% FBS, 55 µM 2-mercaptoethanol and penicillin–streptomycin. Cell suspensions were centrifuged at 400 g and 4°C for 5 min and the supernatant discarded. Cell pellets were re-suspended in RPMI-1640 supplemented with 20% FBS, 55 µM 2-mercaptoethanol and penicillin–streptomycin and plated at 5 × 10⁴ cells per well in flat-bottomed, 96-well plates. For blinatumomab studies, healthy peripheral blood primary human CD8⁺ T cells were isolated by magnetic separation using CD8⁺ T Cell Isolation Kit, human (Miltenyi Biotec, catalog no. 130-096-495) according to the manufacturer's instructions.

CD8⁺ T cells were co-cultured with sgRNA–mCherry⁺ NALM6 cells at effector:target ratios indicated in the presence or absence of 0.5 ng ml⁻¹ of blinatumomab (Creative BioLabs, catalog no. BSAB-L002, lot no. Tandem

scFv-0119) for 24 hours. For mCherry⁺ cell counts, all samples were gated based on forward and side scatter, followed by exclusion of doublets, then gated on viable cells (DAPI^{low}), exclusion of T cells (CD3⁺) and gating on sgRNA-expressing NALM6 cells (mCherry⁺) using Flowjo v.10 (Treestar, BD Biosciences). In addition, CountBright Plus Absolute Counting Beads were identified based on forward and side scatter ('Bead Region' gate; Extended Data Fig. 6) and then fluorescence verification using the APC-Cy7 channel ('Beads' gate; Extended Data Fig. 6). The mCherry⁺ count was divided by the bead count for each condition to generate the normalized mCherry⁺/NALM6 cell count. TRAC or TRAC CD19 CAR T cells were co-cultured with sgRNA-mCherry⁺ Cas9⁺/697 cells at effector:target ratios indicated for 24 hours. After co-culture, cell suspensions were centrifuged at 400 g and 4°C for 5 min and the supernatant discarded. Cells were stained with APC-conjugated anti-human CD19 (BioLegend, catalog no. 302212, 1:200) and FITC-conjugated anti-human CD3 (eBioscience, catalog no. 11-0038-42, 1:200) in the presence of CountBright Plus Absolute Counting Beads (Invitrogen, catalog no. C36995). The mCherry⁺ cell counts were determined as described for blinatumomab treatment.

Bioinformatic analysis. ENCODE transcription factor binding and GSEA. Transcription factor ChIP data were downloaded from the ENCODE database. Signal *P* value, a statistical significance of the signal at a location compared with IgG control, was used as a bigwig file provided from the database. GSEA of ENCODE ChIP-X data was performed using Enrichr³⁷.

RNA-seq analysis. Cas9⁺/Reh, -697, -NALM6, -TMD8, -HG3 and -K562 cell lines were transduced with LRCheery2.1-sgRNA (sgZNF143 or sgNUDT21 in addition to negative control sgROSA) and cultured for 7 days. The mCherry⁺ cells were isolated, and RNA was extracted using the RNeasy Plus Mini Kit (QIAGEN, catalog no. 74316). Nudt21-conditional knockout B-ALL cells were subjected to 5 days of 4-OHT treatment and then cultured for an additional 2 days, and cells were isolated and RNA extracted using the RNeasy Plus Mini Kit (QIAGEN, catalog no. 74316). RNA-seq libraries were prepared using NEXTflex Rapid Illumina Directional RNA-seq Library prep kit as per the manufacturer's guidelines. The libraries were sequenced in paired-end by NovaSeq 6000 at 100 cycles. RNA-seq reads were aligned to the human reference genome (GRCh37/hg19) using STAR aligner³⁸ with default parameters. For differential expression analysis, we used the DESeq2 R package³⁹ with raw read count matrices. Default parameters of DESeq were used and batch difference was considered as a variable of the negative binomial generalized linear model when multiple batches of RNA-seq datasets were used. Counts per million (c.p.m.)-normalized expression values were used for the MA plots.

Hi-C analysis. Cas9⁺/Reh, -697 and -NALM6 cell lines were transduced with LRCheery2.1-sgROSA/NUDT21 and cultured for 7 days. Then, 100,000 mCherry⁺ cells were subjected to Hi-C sample preparation according to the manufacturer's guidelines (Arima Genomics, catalog no. A410030). Hi-C data were analyzed using the Hi-C-bench platform⁴⁰. The libraries were sequenced in paired-end by NovaSeq 6000 at 100 cycles. The reads were aligned to the human reference genome (GRCh37/hg19) by bwa-mem⁴¹ with the '-A1 -B4 -E50 -L0' parameter. After alignment, multi-mapped reads, read pairs with only one mappable read, duplicated read pairs and read pairs with a low mapping quality (MAPQ < 20) were discarded. For Virtual 4C analysis, CD19 promoter (chr16:28,943,260) was used as the viewpoint with 5,000-bp resolution. For intra-TAD activity analysis, we used the algorithm presented in our previous study⁴². Briefly, we identified common TADs between ZNF143-depleted cell lines and the control cell lines where TAD boundaries of each cell line overlapped within 120 kb. We set the minimum TAD size to 400 kb and maximum range of interaction to 2 Mb. Then we averaged the Hi-C interactions within each common TAD and performed a paired, two-sided Student's *t*-test followed by multiple testing correction using FDR estimation.

Enhanced CLIP analysis. We followed the data processing as previously described⁴³. Briefly, adapter and adapter-dimer sequences were eliminated using Cutadapt, and the trimmed reads were aligned to the human reference genome (GRCh37/hg19) using STAR aligner v.2.7.7 (ref. ³⁸) with the recommended parameter in the previous study⁴³ (--outFilterScoreMin 10, --outSAMunmapped Within, --outFilterMultimapNmax 1, outFilterMultimapScoreRange 1). After PCR duplicates were removed using Unique Molecular Identifier (UMI) tools⁴⁴, we called eCLIP peaks using PureCLIP⁴⁵ with default parameters.

MACE-seq analysis. Raw data were pre-processed using Cutadapt⁵ to eliminate poly(A) tails as well as bad-quality base reads. FastQC was used to assess the quality of sequencing after trimming. Cleaned reads were mapped to a reference genome using Bowtie2 (ref. ⁶). Quantification of mapped reads to each gene was performed using HT-seq⁴⁶. ENSEMBL-GTF data were used to provide genomic locations for quantification as well as additional data for annotation (such as gene name, gene description, gene ontology terms and so on). APA (alternative pA) usage was analyzed as described in quantification of APA (QAPA)⁴⁷ using SALMON⁴⁸ for quantifying 3'-UTRs of PolyASite. In short, QAPA quantifies APA levels of genes annotated by distinct 3'-UTR sequences, which were built from the

GENCODE database⁴⁹ and the PolyASite database⁵⁰. For every gene that has more than two distinct 3'-UTRs, we calculated the proportion of each transcript and determined the change in 3'-UTR usage when the proportion is >10%.

Alternative splicing analysis. FASTQ files were first trimmed using Trim_galore (v.0.6.4) to remove sequencing adapters and low quality (*Q* < 15) reads. Trimmed sequencing reads were aligned to the human Hg19 reference genome (GENCODE, GRCh37.p13) using STAR (v.2.7.5)³⁸. SAM files were subsequently converted to BAM files, sorted and indexed using samtools (v.1.9). For quantification of alternative RNA splicing, BAM files generated by STAR/Samtools were analyzed using rMATS (v.4.1.1)⁵¹ and the GENCODE (v.19) GTF annotation for Hg19 (GRCh37.p13). To utilize reads shorter than 51 bp resulting from adapter and/or quality control trimming by trim_galore, rMATS was programmed to accept soft-clipped reads of variable length. Enumeration of isoform counts was performed using only reads that span the splice junction directly. To identify high-confidence alternative splicing events, they were considered significant if (1) the inclusion level difference was >20% compared with sgROSA, (2) the FDR < 0.05 and (3) there was a minimum of 20 reads mapping to the splice junction. For visualization by Sashimi plot, BAM files were loaded into Integrative Genomics Viewer software (v.2.7.0) and Sashimi plots generated using the following criteria: Junction coverage min= 10.

Bulk RNA-seq: blinatumomab study. We analyzed publicly available, matched bulk, RNA-seq data of pre- and post-treatment samples of blinatumomab treatment from the previous study³. The previous study performed whole-exome sequencing and RNA-seq with a cohort including ten matched, post-blinatumomab, relapse samples. In brief, they performed whole-exome sequencing and RNA-seq using the TrueSeq DNA Exome library preparation kit and TruSeq Stranded Total RNA library preparation kit, respectively, and sequenced the libraries using HiSeq 4000 and NovaSeq 6000. In particular, of the seven samples that showed CD19 loss in relapse, one paired sample (SJBALL042246) showed a significant reduction of CD19 without the acquisition of genetic mutations. We used counts per million-normalized read counts that quantified by RSEM (RNA-seq by expectation-maximization) followed by STAR alignment against GRCh37/hg19, and analyzed the fold-change of the post-blinatumomab sample over the pre-treatment sample.

scRNA-seq data analysis. Previously published and publicly available scRNA-seq data from four healthy and seven B-ALL diagnosis bone marrow specimens were pre-processed by Cell Ranger 3.0 with default settings and aligned to the human reference genome (GRCh37/hg19). To account for biological and technical batch differences, we used the Harmony algorithm⁵² with log(transformed expression values). To determine the differential cell abundance between healthy bone marrow and B-ALL diagnosis bone marrow, we applied the Milo algorithm⁵³ to the Harmony embeddings, where we obtained a *K*-nearest neighborhood (KNN) graph with the parameters, *k* = 30, *d* = 30, and then performed differential abundance testing between two groups (healthy versus B-ALL diagnosis) based on negative, binomial, generalized linear model implemented in Milo. We used Seurat v.4 to obtain cell-type annotations, mapping our dataset to the reference pre-annotated human bone marrow dataset provided in the previous paper^{54,55}. To compare the relative expression level of CD19 regulator candidates across cell types, we calculated the average expression level of each candidate per defined cell type and standardized them using the *z*-score across the cell types. For correlation analysis between CD19 regulator candidates and CD19, we calculated the average expression level of each candidate per single-cell neighborhood defined by the KNN graph with the same parameters described above and calculated Pearson's correlation coefficient across the cell neighborhood per cell type. For the scRNA-seq of pre- and post-CAR-T exposure samples³⁸, we downloaded the publicly available pre-processed data from the Gene Expression Omnibus (GEO) database (accession no. GSE153697), used SCTransform for normalization and performed principal component analysis and Uniform Manifold Approximation and Projection (UMAP) for dimension reduction using Seurat v.4.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Data is available under the GEO SuperSeries, accession no. GSE190844. Source data are provided with this paper.

References

- Li, S., Ilari, R.L., Million, R.P., Daley, G.Q. & Van Etten, R.A. The P190, P210, and P230 forms of the *BCR/ABL* oncogene induce a similar chronic myeloid leukemia-like syndrome in mice but have different lymphoid leukemogenic activity. *J. Exp. Med.* **189**, 1399–1412 (1999).
- Shi, J. et al. Discovery of cancer drug targets by CRISPR-Cas9 screening of protein domains. *Nat. Biotechnol.* **33**, 661–667 (2015).

33. Ran, F. A. et al. Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281–2308 (2013).
34. Doench, J. G. et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* **34**, 184–191 (2016).
35. Van Nostrand, E. L. et al. Robust, cost-effective profiling of RNA binding protein targets with single-end enhanced crosslinking and immunoprecipitation (seCLIP). *Methods Mol. Biol.* **1648**, 177–200 (2017).
36. Zawada, A. M. et al. Massive analysis of cDNA ends (MACE) and miRNA expression profiling identifies proatherogenic pathways in chronic kidney disease. *Epigenetics* **9**, 161–172 (2014).
37. Kuleshov, M. V. et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–97 (2016).
38. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
39. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
40. Lazaris, C., Kelly, S., Ntziachristos, P., Aifantis, I. & Tsirigos, A. HiC-bench: comprehensive and reproducible Hi-C data analysis designed for parameter exploration and benchmarking. *BMC Genom.* **18**, 22 (2017).
41. Li, H.W. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at <https://doi.org/10.48550/arXiv.1303.3997> (2013).
42. Kloetgen, A. et al. Three-dimensional chromatin landscapes in T cell acute lymphoblastic leukemia. *Nat. Genet.* **52**, 388–400 (2020).
43. Van Nostrand, E. L. et al. A large-scale binding and functional map of human RNA-binding proteins. *Nature* **583**, 711–719 (2020).
44. Smith, T., Heger, A. & Sudbery, I. UMI-tools: modeling sequencing errors in unique molecular identifiers to improve quantification accuracy. *Genome Res.* **27**, 491–499 (2017).
45. Krakau, S., Richard, H. & Marsico, A. PureCLIP: capturing target-specific protein-RNA interaction footprints from single-nucleotide CLIP-seq data. *Genome Biol.* **18**, 240 (2017).
46. Putri, G. H., Andres, S., Pyl, P. T., Pimanda, J. E. & Zanini, F. Analysing high-throughput sequencing data in Python with HTSeq 2.0. *Bioinformatics* **38**, 2943–2945 (2022).
47. Ha, K. C. H., Blencowe, B. J. & Morris, Q. QAPA: a new method for the systematic analysis of alternative polyadenylation from RNA-seq data. *Genome Biol.* **19**, 45 (2018).
48. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419 (2017).
49. Harrow, J. et al. GENCODE: the reference human genome annotation for the ENCODE Project. *Genome Res.* **22**, 1760–1774 (2012).
50. Gruber, A. J. et al. A comprehensive analysis of 3' end sequencing data sets reveals novel polyadenylation signals and the repressive role of heterogeneous ribonucleoprotein C on cleavage and polyadenylation. *Genome Res.* **26**, 1145–1159 (2016).
51. Shen, S. et al. rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc. Natl Acad. Sci. USA* **111**, E5593–E5601 (2014).
52. Korsunsky, I. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
53. Dann, E., Henderson, N.C., Teichmann, S.A., Morgan, M.D. & Marioni, J.C. Differential abundance testing on single-cell data using k-nearest neighbor graphs. *Nat. Biotechnol.* **40**, 245–253 (2021).
54. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e3529 (2021).
55. Stuart, T. et al. Comprehensive Integration of single-cell data. *Cell* **177**, 1888–1902.e1821 (2019).

Acknowledgements

We thank all members of the Aifantis laboratory for discussions throughout this project. M.T.W. was supported by the Leukemia & Lymphoma Society Career Development Program, American Society of Hematology Restart Award, NIH/National Cancer Institute (NCI) K22 award (no. 1K22CA258520-01), Cancer League of Colorado Research Grant (AWD no. 222549) and the Jeffrey Pride Foundation for Pediatric Cancer Research and the Children's Oncology Group Foundation. P.T. was supported by the AACR Incyte Corporation Leukemia Research Fellowship and Young Investigator grant from Alex's Lemonade Stand Cancer Research Foundation. This work is supported by the National Science Foundation (grant no. CBET 2103219 to W.C.) and the US NIH (grant no. R35GM133646 to W.C.). C.M. is supported by the Cancer Research Institute Irvington Postdoctoral Fellowship (no. CRI4018). S.J.H. is supported by an investigator

grant from the National Health and Medical Research Council of Australia. J.B. is grateful for support from the NIH (grant no. 1F32HD078029-01A1). K.H. was supported by funds from Massachusetts General Hospital, the NIH (grant no. P01GM099134) and the Gerald and Darlene Jordan Chair in Regenerative Medicine. N.T. was supported by a fellowship from the German Research Foundation. A.T. and J.E. received funding from the Parker Institute for Cancer Immunotherapy and the Grand Multiple Myeloma Translational Initiative. C.G.M. was supported by the American Lebanese Syrian Associated Charities of St. Jude Children's Research Hospital, NCI (grant no. R35 CA197695) and the Henry Schueler 41&9 Foundation. This work has used computing resources at the NYU School of Medicine High Performance Computing Facility. S.N. was supported by the Onassis Foundation—scholarship ID: F ZP 036-1/2019-2020. O.A.-W. was supported by the NIH (grant nos. R01CA251138 and R01CA242020), and the NIH/NCI (grant no. 1P50 254838-01), the Leukemia & Lymphoma Society and the Edward P. Evans MDS Foundation. I.A. was supported by the NCI/NIH (grant nos. P01CA229086, R01CA252239, R01CA228135, R01CA242020 and O1CA266212), Curing Kids Cancer, the Leukemia and Lymphoma Society and the Vogelstein Foundation. We thank the NYU School of Medicine core facilities, including the Applied Bioinformatics Laboratories, Flow Cytometry and the Genome Technology Center (this shared resource is partially supported by the Cancer Center Support, grant no. P30CA016087, at the Laura and Isaac Perlmutter Cancer Center). Schematic illustrations in Extended Data Figs. 1a,g and 6a created with BioRender.com.

Author contributions

M.T.W., E.W. and S.L. conceived, planned and performed most of the experiments and co-wrote the manuscript. As such, each name can be interchanged in the first position of the manuscript. S.L. performed most of the computational analyses. E.W. and M.T.W. performed CRISPR screening and most of the experiments described. A.K. provided experimental support for RNA decay experiments and extensive conceptual support throughout the development of this project. S.J.H. provided computational support through generation of sgRNA histograms and splicing analysis. P.T., Y.G. and S.N. provided technical support for chromatin conformation experiments and data interpretation. C.M. and W.C. performed ex vivo microfluidics experiments, data analysis and support for data interpretation. A.T. and J.E. generated TRAC CD19 CAR knock-in T cells and provided guidance on experimental design. Y.Z., K.G.R. and C.G.M. provided primary patient blinatumomab gene expression data and guidance on data interpretation. N.T., J.B. and K.H. provided *Nudt21*-conditional knockout bone marrow cells and guidance on experimental design and data interpretation. O.A. provided supervision to E.W. I.A. directed and coordinated the study. All authors discussed the results and commented on the manuscript.

Competing interests

I.A. is a consultant for Foresite Labs and receives research funding from AstraZeneca. A.T. is a scientific advisor to Inteligencia AI. M.T.W. has received royalties from the Walter and Eliza Hall Institute for the development of venetoclax unrelated to the current manuscript. O.A.-W. has served as a consultant for H3B Biomedicine, Foundation Medicine Inc., Merck, Prelude Therapeutics and Janssen and is on the Scientific Advisory Board of Envisagenics Inc., AIChem, Harmonic Discovery Inc. and Pfizer Boulder. He has received previous research funding from H3B Biomedicine and LOXO Oncology unrelated to the current manuscript. C.G.M. receives research support from AbbVie and Pfizer, is a member of the advisory boards of Faze, Beam and Illumina and has accepted speaking fees from Amgen. J.E. is a compensated co-founder at Mnemo Therapeutics and a compensated scientific advisor to Cytovia Therapeutics, owns stocks in Mnemo Therapeutic and Cytovia Therapeutics and has received a consulting fee from Casdin Capital. He is also a holder of patents pertaining to but not resulting from this work. J.E.'s lab has received research support from Cytovia Therapeutic, Mnemo Therapeutics and Takeda. The remaining authors declare no competing interests.

Additional information

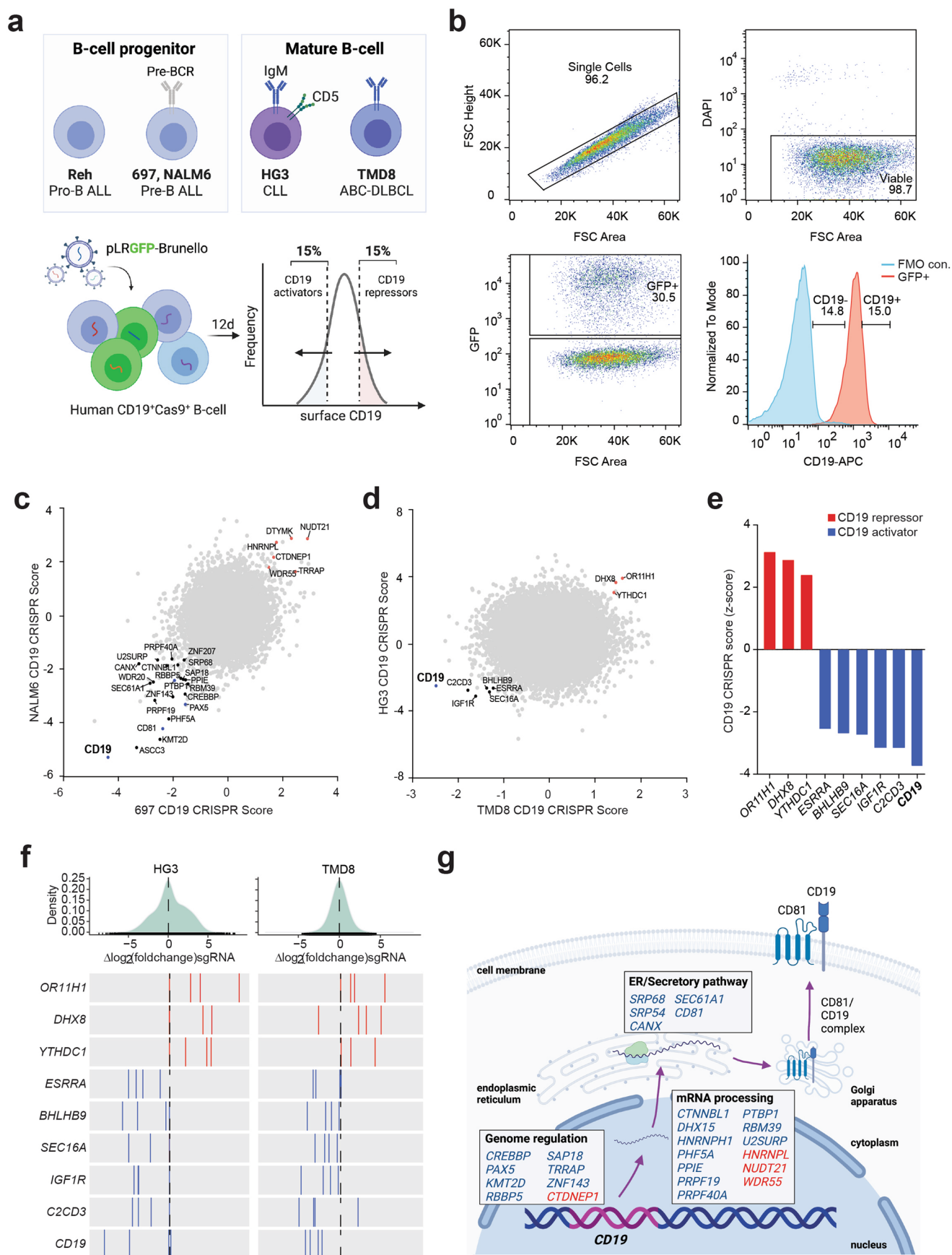
Extended data Extended data are available for this paper at <https://doi.org/10.1038/s41590-022-01314-y>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41590-022-01314-y>.

Correspondence and requests for materials should be addressed to Matthew T. Witkowski or Iannis Aifantis.

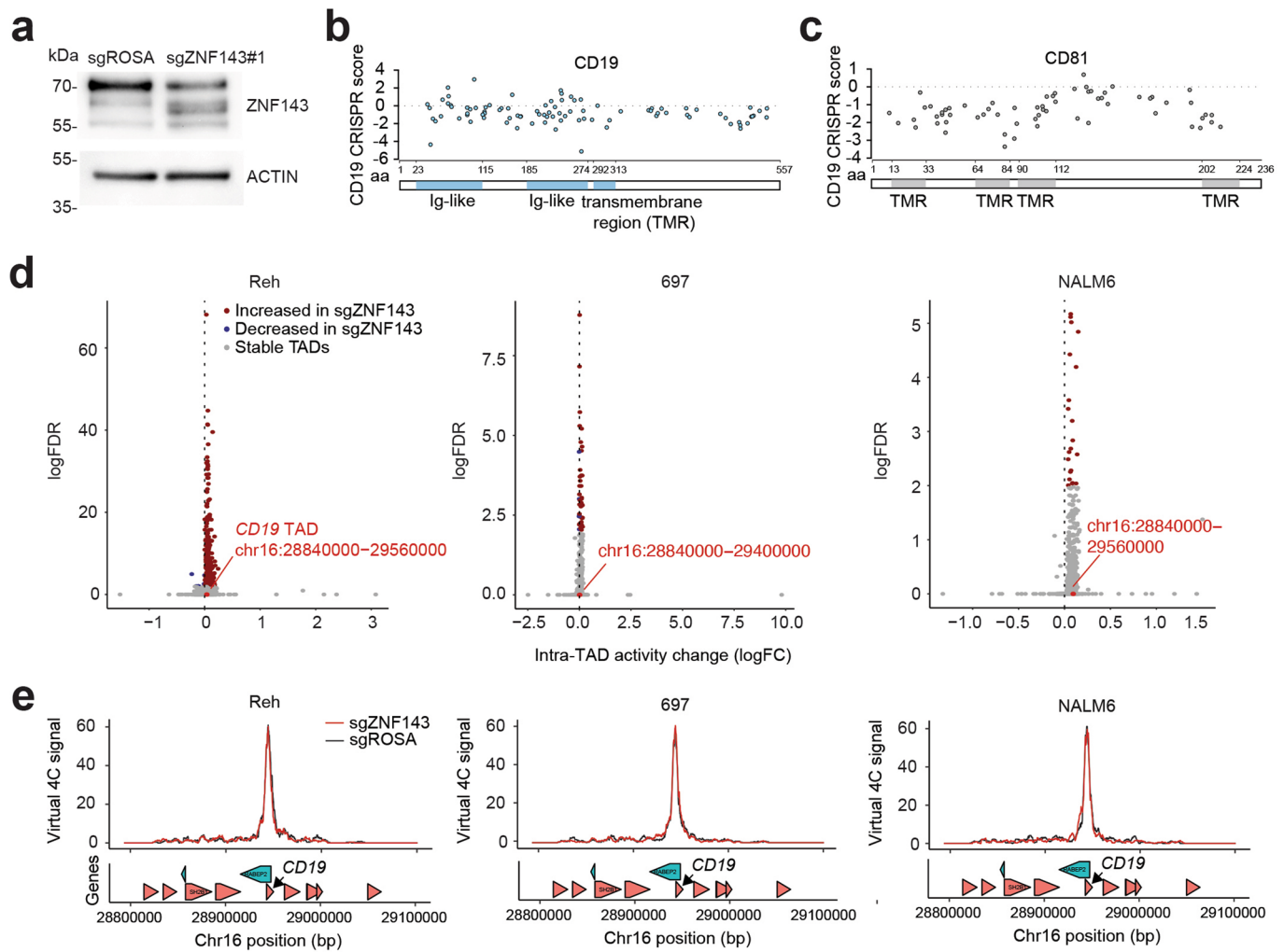
Peer review information *Nature Immunology* thanks Catriona Jamieson, Patrick Matthias and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Ioana Visan in collaboration with the *Nature Immunology* team.

Reprints and permissions information is available at www.nature.com/reprints.

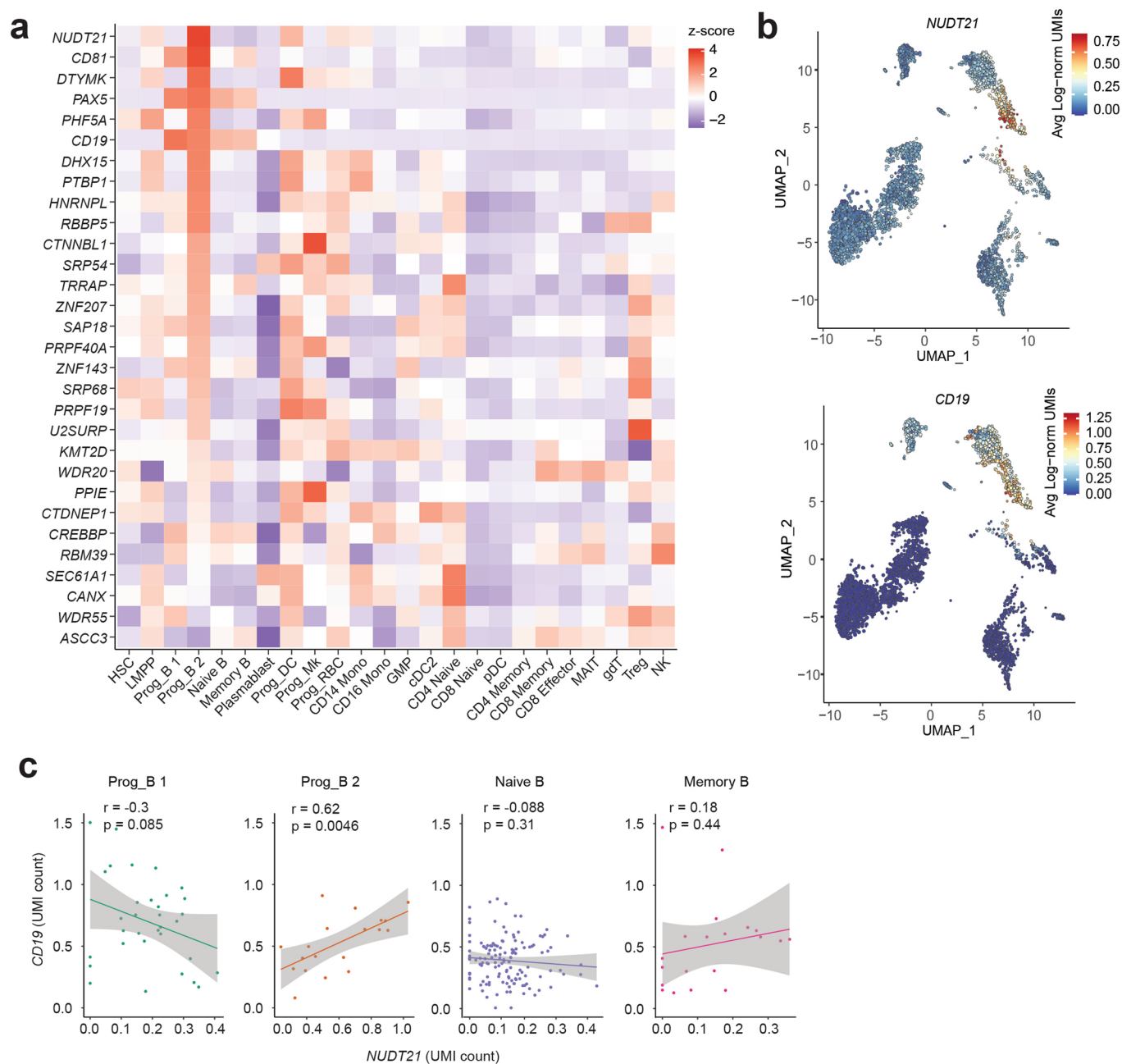


Extended Data Fig. 1 | See next page for caption.

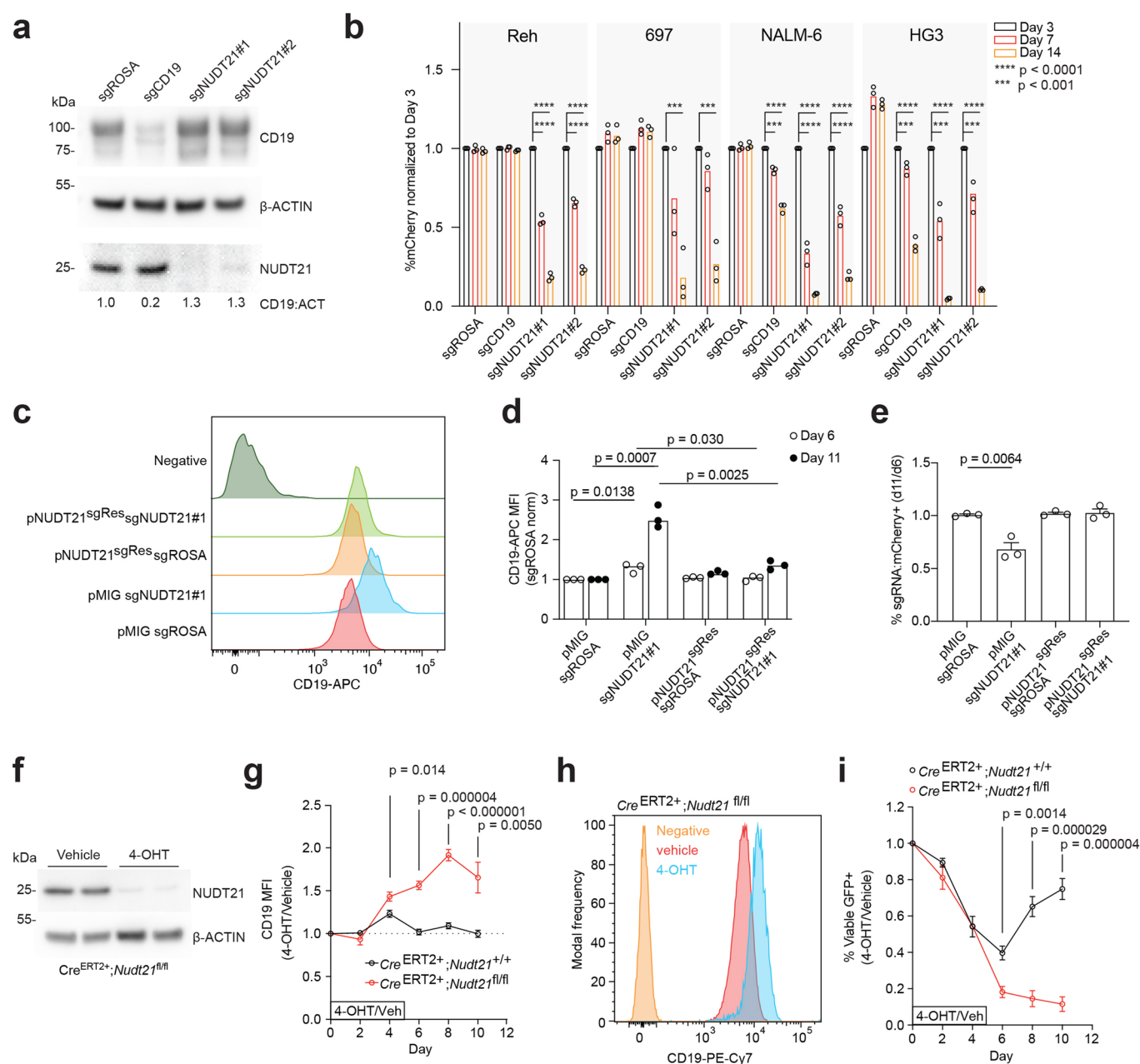
Extended Data Fig. 1 | Identifying CD19 regulatory pathways in human B cell malignancies. **(a)** Schematic of pooled genome-wide CRISPR screens across human B-cell lines. **(b)** Representative flow cytometry of CD19 separation for CRISPR screening approached in Cas9⁺ human B cell line, NALM6, transduced with Brunello sgRNA library following 12 days culture. **(c-d)** Scatterplot showing CD19 score for individual gene candidates for CD19 activators and CD19 repressors comparing **(c)** human B-cell progenitor acute lymphoblastic leukemia (BCP-ALL) cell lines (Reh, 697 and NALM6) and **(d)** mature B cell (TMD8 and HG3) lines. **(e)** Waterfall plot showing the average CD19 CRISPR z-score of top gene candidates for CD19 activators (blue) and CD19 repressors (red) in human mature B cell neoplastic lines (HG3 and TMD8). **(f)** Histogram showing sgRNA fold change for individual sgRNAs targeting top gene candidates for CD19 activators (blue) and CD19 repressors (red) in mature B cell (HG3 and TMD8) lines. **(g)** Schematic of genes involved in regulation of CD19 antigen expression in B-cell malignancies.



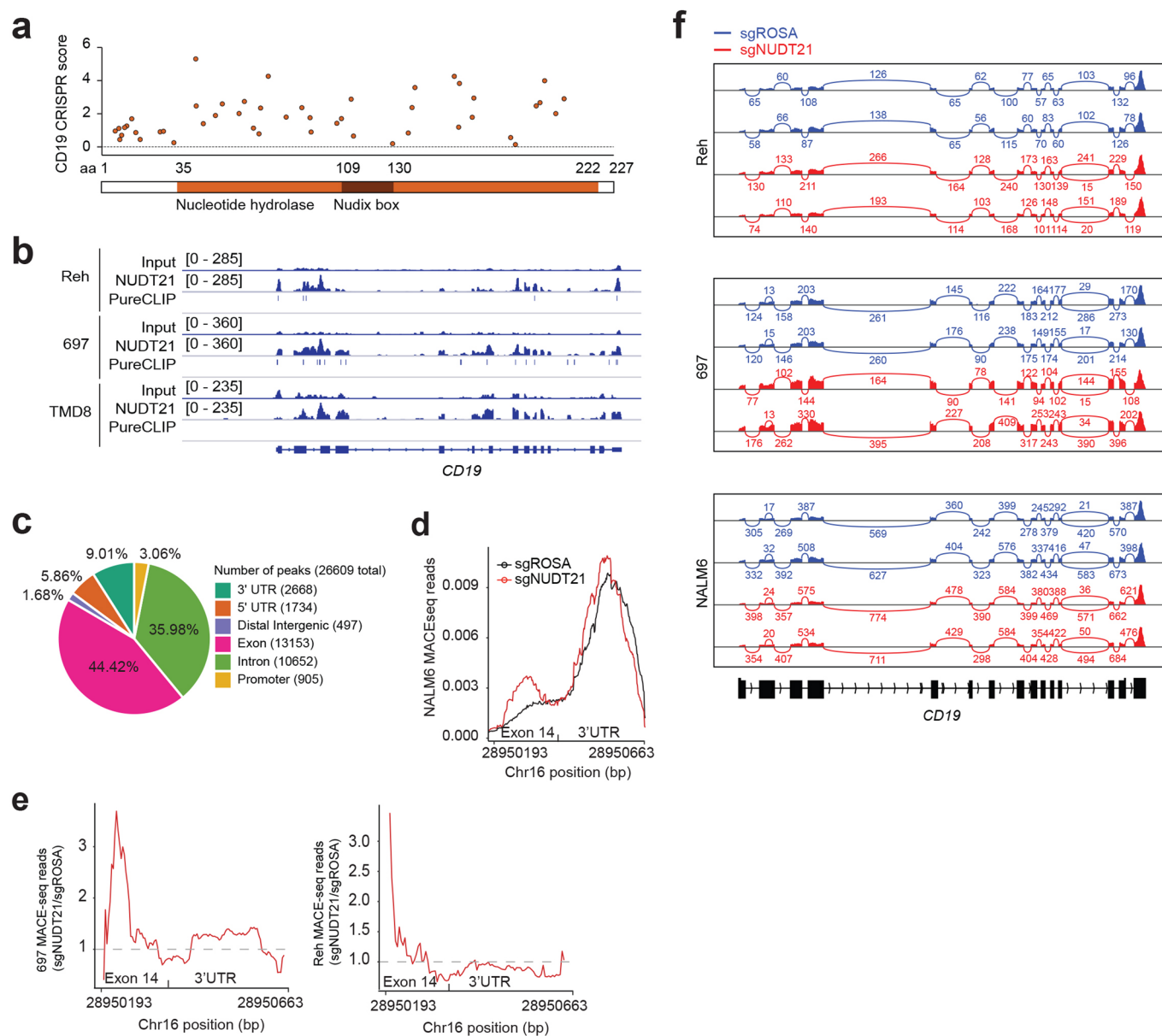
Extended Data Fig. 2 | ZNF143 ablation alters CD19 mRNA expression independent on chromatin looping. (a) Immunoblot analysis of ZNF143 and Actin in NALM6 cells expressing sgROSA or sgZNF143#1 for seven days. **(b-c)** Protein sizes indicated **(b)** CD19 scores for individual sgRNAs spanning the CD19 and **(c)** CD81 locus in NALM6 domain screen. **(d)** Volcano plots of intra-TAD activity comparing sgROSA and sgZNF143 expressing cell line, Reh, 697 and NALM6 (two-sided t-test followed by false discovery rate (FDR) correction. FDR < 0.01 cutoff). **(e)** Virtual 4C analysis of CD19 promoter viewpoint generated from Hi-C data of Reh, 697 and NALM6 expressing sgROSA and sgZNF143.



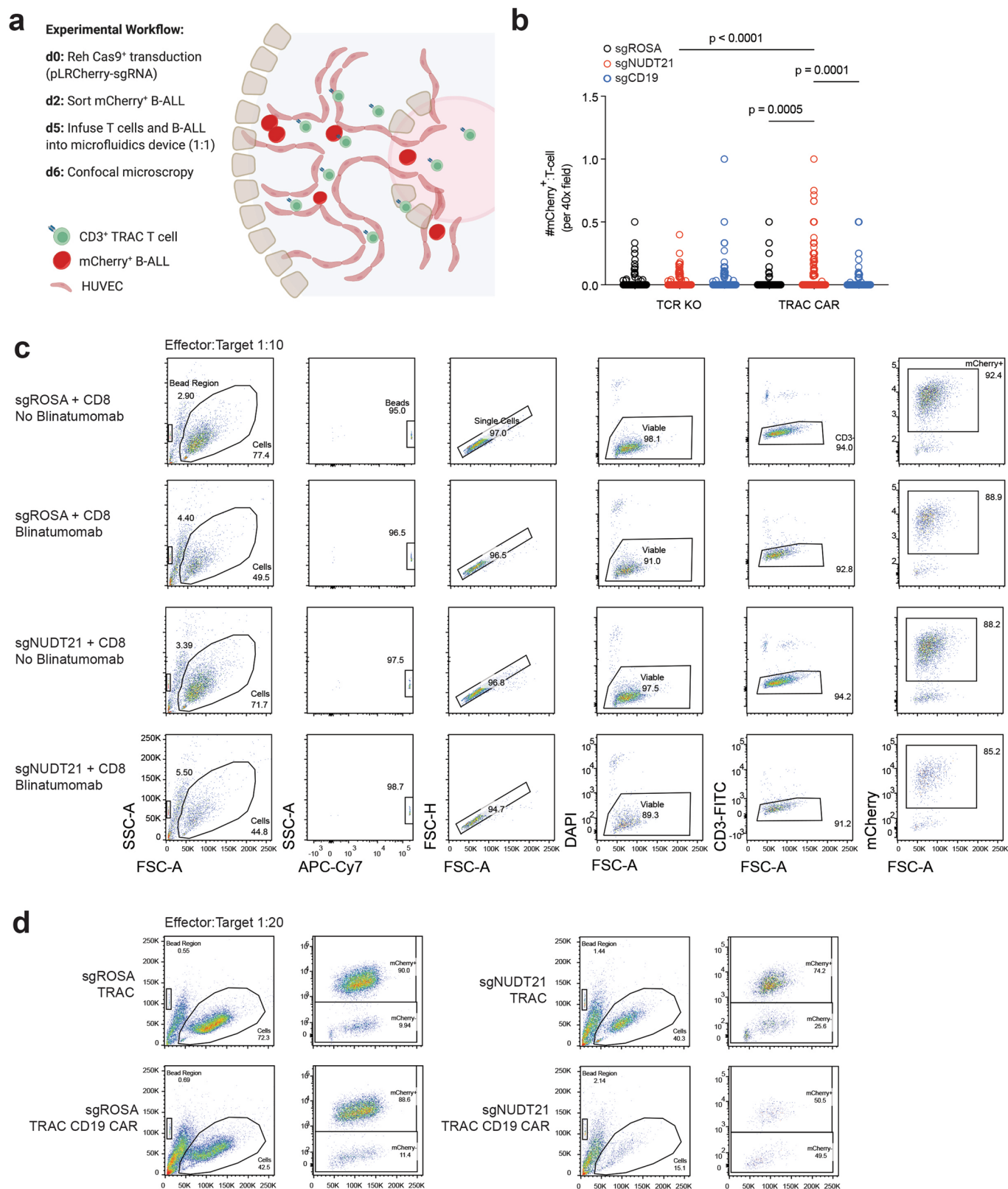
Extended Data Fig. 3 | *NUDT21* is co-expressed with *CD19* mRNA across healthy and B-ALL bone marrow cells. (a) Heatmap representation of cell type (row) z-score normalization of UMI counts for each individual *CD19* candidate using healthy bone marrow scRNA-seq data. Genes ranked in descending order or individual candidate z-score across Prog_B 2 column. **(b)** UMAP representation of *NUDT21* and *CD19* mRNA expression as measured by log normalized UMI counts. **(c)** Scatterplot showing correlation between *NUDT21* and *CD19* mRNA levels (UMI count) across Prog_B 1, Prog_B 2, Naïve B and Memory B cells in healthy bone marrow scRNA-seq data. *r*- and *p*-values calculated on the basis of Pearson's correlation.



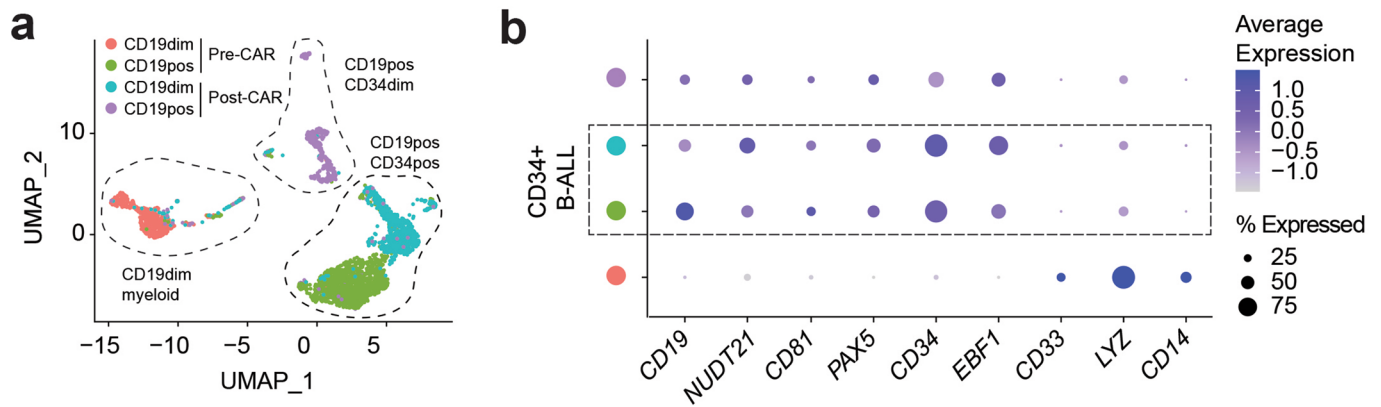
Extended Data Fig. 4 | NUDT21 represses CD19 expression and survival in human and murine B cell progenitors. (a) Immunoblot analysis of CD19, NUDT21 and β -Actin in NALM6 cells expressing sgROSA or sgNUDT21#1 for seven days. Protein sizes indicated. CD19 levels normalized to Actin by densitometry. (b) Histogram of mCherry⁺ percentages normalized to Day 3 mCherry⁺ percentage across multiple cell lines (independent experiments with $n = 3$, unpaired two-sided t-test, mean and standard error shown). Data with statistical significance are as indicated, **** $p < 0.0001$, *** $p < 0.001$. (c) Representative flow cytometry of CD19 expression in NALM6 cells expressing MSCV-IRES-GFP (pMIG), pMSCV-NUDT21^{sgRes}-IRES-GFP, sgROSA and/or sgNUDT21#1. (d) CD19-APC mean fluorescence intensity and (e) percentage GFP⁺mCherry⁺ cells comparing Day 11 to Day 6 post-transduction normalized to pMIG;sgROSA-expressing cells (independent experiments with $n = 3$, unpaired two-sided t-test, mean and standard error shown, error bars represent s.e.m.). (f) Immunoblot analysis of Nudt21 and beta-Actin whole lysate levels in ROSA26-($Cre^{ERT2+};Nudt21^{fl/fl}$) cells following five days of vehicle or 4-OHT treatment *in vitro*. (g) Kinetic summary of CD19 mean fluorescence intensity levels across $Cre^{ERT2+};Nudt21^{fl/fl}$ or $Cre^{ERT2+};Nudt21^{+/+}$ over 12 days culture period. 4-OHT values normalized to vehicle control treatment. Five days of vehicle or 4-OHT treatment (day 0 – day 5). Two independent cell lines per genotype, each performed in three independent experiments ($n = 6$ total) (unpaired two-sided t-test, mean and standard error shown, error bars represent s.e.m.). (h) Representative flow cytometry of CD19 expression at day seven culture following five days vehicle or 4-OHT treatment (day 0 – day 5) in $Cre^{ERT2+};Nudt21^{fl/fl}$ cells. (i) Kinetic summary of percentage of viable (DAPI⁻) GFP⁺ cells for $Cre^{ERT2+};Nudt21^{fl/fl}$ or $Cre^{ERT2+};Nudt21^{+/+}$ over 12 days culture period. 4-OHT values normalized to vehicle control treatment. Five days of vehicle or 4-OHT treatment (day 0 – day 5). Two independent cell lines per genotype, each performed in three independent experiments ($n = 6$ total) (unpaired two-sided t-test, mean and standard error shown, error bars represent s.e.m.).



Extended Data Fig. 5 | NUDT21 directly regulates CD19 mRNA 3' UTR length. (a) CD19 scores for individual sgRNAs spanning the NUDT21 locus in NALM6 domain screen. (b) eCLIP read tracks spanning the CD19 locus in Reh, 697 and TMD8 cells. PureCLIP significant peak signals shown. (c) Pie-chart highlighting genomic distribution of eCLIP peaks shared by BCP-ALL cell lines. (d) MACE-seq reads for NALM6 and (e) log fold-change read counts of the terminal coding exon 14 and 3-UTR junction (intron removed) in 697 and Reh cells comparing sgNUDT21#1 to sgROSA. (f) Sashimi plot of exon-exon junctions across the CD19 locus in Reh, 697 and NALM6 cells. Bulk RNA-seq experiment performed in technical duplicate.



Extended Data Fig. 6 | On-chip measurement of CAR-T killing efficacy and synapse formation and CD19-directed therapy challenge ex vivo. (a) The experimental workflow of on-chip measurement of CAR-T killing efficacy and synapse formation capability using a 3D microfluidic HUVEC vascularized model. **(b)** On-chip measurement of the frequency of synapse formation between T-cells and sgRNA-expressing BCP-ALL (independent experiments with $n = 4$, unpaired two-sided t-test, mean and standard error shown, error bars represent s.e.m.). **(c-d)** Representative flow cytometry of **(c)** blinatumomab and **(d)** TRAC CD19 CAR treatment following 24 hours of co-culture with sgRNA-expressing (mCherry⁺) BCP-ALL. Countbright beads indicated by APC-Cy7.



Extended Data Fig. 7 | Single cell identification of BCP-ALL CD34-expressing cluster throughout primary human BCP-ALL CAR-T therapy. (a) UMAP representation of primary BCP-ALL patient single cell dataset generated by Rabilloud *et al.* highlighting CD19^{dim} and CD19^{pos} cell clusters pre (T1) and post (T2) CAR-T cell therapy, with sample cluster CD19 and CD34 mRNA levels indicated. **(b)** Dot plot representation of cluster-specific mRNA expression levels from BCP-ALL CAR-T patient single cell data.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- ☐ ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐ ☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☐ ☒ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☐ ☒ A description of all covariates tested
- ☐ ☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☐ ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☐ ☒ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☐ ☒ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection ENCODEExplorer R package was used for downloading transcription factor ChIP-sequencing data from ENCODE database.

Data analysis R version 4.1.1 and Python version 3.7.2 were used throughout this study. STAR aligner version 2.7.7 was used for alignment of eCLIP data, STAR aligner version 2.7.5 was used for Bulk-RNA sequencing. bwa version 0.7.17 was used for Hi-C alignment, Bowtie2 was used for MACE-seq alignment. Samtools version 1.9 was used for throughout the study. Cutadapt version 3.4 was used for trimming eCLIP adapter sequences. UM I-tools version 1.0.0 was used for eCLIP data processing. Gene quantification for bulk RNA-seq and MACE-seq was performed by subread version 1.6.3 and HTSeq version 1.99.2, respectively. Quantification of 3'UTR in MACE-seq was performed using Salmon 1.6.0 and the QAPA python package (<https://github.com/morrislab/qapa>). Hi-C Bench (<https://github.com/NYU-BFX/hic-bench>) was used for Hi-C analysis. Seurat package version 4.0 and Milo package version 0.1.0 were used for single-cell RNA-sequencing analysis.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data reported in this paper are tabulated in the Supplementary Materials and raw data are archived on the GEO database under accession number GSE190844.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	All in vitro experiments were performed in n greater than or equal to 3, as indicated in the each figure.
Data exclusions	All replicate values have been included for visualization with no exclusion of data points.
Replication	All in vitro experiments were performed at least 2 independent times with most performed > 3 independent times and further validated in two or more cell lines. For all sequencing data-types, successful replication has been confirmed with Principal Component Analysis.
Randomization	For cell line collections and downstream processing and analysis, number labels were used to remove downstream operator bias.
Blinding	The investigators were not blinded to sample group allocation, as knowledge of the treatment groups was required.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

For western blots, the following antibodies and dilutions were used.

CD19 (1:500, Cell Signaling Technology, #3574, Lot # 5)
 NUDT21 (1:100, Santa Cruz Biotechnology, sc-81109, Lot# B1121)
 ZNF143 (1:1000, Novus Biologicals, 2B4 H00007702-M01, Lot# JC261-2B4)
 beta-Actin (1:1000, Millipore, MAB1501R)

For FACS staining; the following antibodies and dilutions were used.

APC-conjugated anti-human CD19 (1:200 concentration, #302212, BioLegend, Lot# B271032)
 PerCP.Cy5.5-conjugated anti-human CD147 (1:1000, #306219, BioLegend, Lot# B266296)
 FITC-conjugated anti-human CD3 (1:200, #11-0038-42, eBioscience, Lot# 2007254)

For microfluidics staining:

Validation

APC-conjugated anti-human CD3 (BioLegend, catalog no. 317318)
PE-Cy7 conjugated anti-mouse CD19 (1:300, 25-0193-82, eBioscience, Lot# 2082894)

All antibodies were validated by the manufacturer and used per their instructions. Additional information on validation can be found on the manufacturers' websites listed below:

CD19 (Cell Signaling Technology, #3574)
<https://www.cellsignal.com/products/primary-antibodies/cd19-antibody/3574>

NUDT21 (Santa Cruz Biotechnology, sc-81109)
<https://www.scbt.com/p/nu dt21-antibody-2203c3>

ZNF143 (Novus Biologicals, 2B4 H00007702-M01)
https://www.novusbio.com/products/znf143-antibody-2b4_h00007702-m01

Actin (1:1000, Millipore, MAB1501R)
https://www.emdmillipore.com/US/en/product/Anti-Actin-Antibodyclone-C4,MM_NF-MAB1501R

APC-conjugated anti-human CD19 (#302212, BioLegend)
<https://www.biolegend.com/en-gb/products/apc-anti-human-cd19-antibody-715>

PerCP.Cy5.5-conjugated anti-human CD147 (#306219, BioLegend)
<https://www.biolegend.com/en-us/products/percp-cyanine5-5-anti-human-cd147-antibody-16340>

FITC-conjugated anti-human CD3 (#11-0038-42, eBioscience)
<https://www.thermofisher.com/antibody/product/CD3-Antibody-clone-UCHT1-Monoclonal/11-0038-42>

PE-Cy7 conjugated anti-mouse CD19 (25-0193-82, eBioscience)
<https://www.thermofisher.com/antibody/product/CD19-Antibody-clone-eBio1D3-1D3-Monoclonal/25-0193-82>

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)

HEK293T were purchased from ATCC CRL-1573, OP9 were purchased from ATCC CRL-2749 and Platinum-E cells were purchased from Cell Biolabs RV-101. HUVEC and Human Fibroblasts purchased from Lonza. Reh, 697, NALM6 cell lines were a gift from Dr. William Carroll's lab at NYU School of Medicine. HG3 and TMD8 were a gift from Hans-Guido Wendel's lab at MSKCC.

Authentication

HEK293T and OP9 were authenticated by the supplier (ATCC Cell Line Authentication Service Sanger Sequencing). Human cell lines (Reh, 697, NALM6, K562) have been authenticated by STR testing (ATCC).

Mycoplasma contamination

Cell lines were routinely monitored for Mycoplasma contamination. All cells used in this study tested negative for Mycoplasma contamination.

Commonly misidentified lines
(See [ICLAC](#) register)

The cell lines used in this study are not listed on the ICLAC list of commonly misidentified cell lines.

Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals

Male and female ROSA26-CreERT2;Nudt21+/+ and ROSA26-CreERT2;Nudt21fl/fl 8-12 weeks age were used for ex vivo B cell transformation experiments with BCR::ABL1 P190 retrovirus. All mice were housed at 22 +/- 1 degrees C, 30-70% relative humidity, in a temperature controlled ABSL-2 facility with 12 hour day-night light cycles. Mice were cared for by the husbandry staff at NYULH DCM (division of comparative medicine), and diet and water were provided ad libitum.

Wild animals

This study did not include wild animals.

Field-collected samples

This study did not include field-collected samples.

Ethics oversight

All animal experiments were performed in accordance with protocols approved by the NYULMC IACUC (Institutional Animal Care and Use Committee) protocol IA16-00008.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Flow Cytometry

Plots

Confirm that:

- ☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- ☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- ☒ All plots are contour plots with outliers or pseudocolor plots.
- ☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

Cell line suspensions were subjected to centrifugation at 400 x g for 5 min at 4°C and supernatant was discarded. Pellets were re-suspended in 1 mL PBS with 2% FCS, and filtered using a 70 µm nylon mesh (Thermo Fisher Scientific) with residual cell clumps discarded. After centrifuging at 400 x g at 4°C for 5 min, the supernatant was discarded, and the cell pellet was re-suspended in PBS (Corning) with 2% FCS and placed on ice prior to staining for FACS. Single-cell suspensions were re-suspended in PBS with 2% FCS and Fc blocking agents then incubated on ice for 10 mins. For human cell lines, we used Human TruStain FcX Blocking Solution (BioLegend, #422301) and murine cell lines, Mouse BD Fc Block (BD Pharmingen, #553142). Cell suspensions were centrifuged at 400 x g at 4°C for 5 min, and supernatant discarded. Cell pellets were re-suspended in PBS with 2% FCS with antibody cocktails and incubated on ice for 30 mins. For human cell lines, we used APC-conjugated anti-human CD19 (1:200 concentration, #302212, BioLegend), PerCP.Cy5.5-conjugated anti-human CD147 (1:1000, #306219, BioLegend) and FITC-conjugated anti-human CD3 (1:200, #11-0038-42, eBioscience). For murine studies, we used PE-Cy7 conjugated anti-mouse CD19 (1:300, 25-0193-82, eBioscience). Following incubation, cell suspensions were subjected to two wash cycles involving re-suspension of cell pellets in 3 mL PBS with 2% FCS, followed by centrifuging at 400 x g at 4°C for 5 min, followed by supernatant removal. Finally, samples were re-suspended in 300 µL of PBS with 2% FCS per 2 x 10⁶ cells, supplemented with DAPI (0.5 µg/mL, Sigma) for live/dead staining. For CRISPR screen cell sorting, all samples were gated based on forward and side scatter, followed by exclusion of doublets, then gated on viable cells (DAPIlow) and sgRNA-expressing cells (GFP+), followed by CD19-APC surface abundance (top- and bottom-15% of fluorescence) using the SY3200™ highly automated parallel sorting (HAPS) cell sorter (Sony). Flow cytometry utilized the aforementioned gating strategy and was performed using the BD Fortessa. All flow cytometry data analysis was performed using FlowJo version 10 (Treestar, BD Biosciences).

Instrument

BD Fortessa analyzer or SY3200™ highly automated parallel sorting (HAPS) cell sorter (Sony).

Software

Data were analyzed by FlowJo v10 (Tree Star) software.

Cell population abundance

All samples were gated based on forward and side scatter, followed by exclusion of doublets, then gated on viable cells (DAPIlow). sgRNA or ectopic expression gating was based on fluorescence (GFP or mCherry, as indicated in text). Mean fluorescence intensity was calculated using FlowJo software. This quantification was, in some cases, normalized between two surface proteins (e.g. CD19-APC MFI divided by CD147-PerCP.Cy5.5 MFI). Non-targeting control (sgROSA, sgNTC) was used for normalization of values to allow comparison between replicates. For mCherry+ cell counts, all samples were gated based on forward and side scatter, followed by exclusion of doublets, then gated on viable cells (DAPIlow), exclusion of T cells (CD3-) and gating on sgRNA-expressing BCP-ALL cells (mCherry+) using FlowJo version 10 (Treestar, BD Biosciences). In addition, Countbright Plus Absolute Counting Beads were identified based on forward and side scatter (Bead Region, Extended Data Figure 6) then fluorescence verification using the APC-Cy7 channel. The mCherry+ count was divided by the bead count for each condition to generate the normalized BCP-ALL count value.

Gating strategy

All samples were gated based on forward and side scatter, followed by exclusion of doublets, then gated on viable cells (DAPIlow). sgRNA or ectopic expression gating was based on fluorescence (GFP or mCherry, as indicated in text).

For mCherry+ cell counts, all samples were gated based on forward and side scatter, followed by exclusion of doublets, then gated on viable cells (DAPIlow), exclusion of T cells (CD3-) and gating on sgRNA-expressing BCP-ALL cells (mCherry+) using FlowJo version 10 (Treestar, BD Biosciences). In addition, Countbright Plus Absolute Counting Beads were identified based on forward and side scatter (Bead Region, Extended Data Figure 6) then fluorescence verification using the APC-Cy7 channel. The mCherry+ count was divided by the bead count for each condition to generate the normalized BCP-ALL count value.

- ☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.