Title

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

Journal

Nature Immunology, 23(10)

ISSN

1529-2908

Authors

Witkowski, Matthew T

Lee, Soobeom

Wang, Eric

et al.

Publication Date

2022-10-01

DOI

10.1038/s41590-022-01314-y

Peer reviewed

HHS Public Access

Author manuscript

Nat Immunol. Author manuscript; available in PMC 2023 April 01.

Published in final edited form as:

Nat Immunol. 2022 October ; 23(10): 1424–1432. doi:10.1038/s41590-022-01314-y.

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

Matthew T. Witkowski[1,15,16], **Soobeom Lee**[1,2,15], **Eric Wang**[3,15], **Anna K. Lee**[1], **Alexis Talbot**[4,5,6], **Chao Ma**[8,9], **Nikolaos Tsopoulidis**[10,11,12], **Justin Brumbaugh**[13], **Yaqi Zhao**[14], **Kathryn G. Roberts**[14], **Simon J. Hogg**[3], **Sofia Nomikou**[1], **Yohana Ghebrechristos**[1], **Palaniraja Thandapani**[1], **Charles G. Mullighan**[14], **Konrad Hochedlinger**[10,11,12], **Weiqiang Chen**[8,9], **Omar Abdel-Wahab**[3], **Justin Eyquem**[4,5,6,7], **Iannis Aifantis**[1,16]

[1]Department of Pathology and Laura & Isaac Perlmutter Cancer Center, NYU School of Medicine, New York, NY 10016, USA

[2]Department of Biology, New York University, New York, NY 10003, USA

[3]Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, NY 10021, USA; Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY 10021, USA

[4]Gladstone-UCSF Institute of Genomic Immunology, San Francisco, CA 94158, USA

[5]Department of Medicine, University of California San Francisco, San Francisco, CA 94143, USA

[6]Department of Microbiology and Immunology, University of California San Francisco, San Francisco, CA 94143, USA

[7]Parker Institute of Cancer Immunotherapy, University of California San Francisco, San Francisco, CA, USA

[16]Correspondence to: Matthew T. Witkowski, matthew.witkowski@cuanschutz.edu (M.T.W.) and Iannis Aifantis, Ioannis.Aifantis@nyulangone.org (I.A.).

Author Contributions

M.T.W, E.W and S.L conceived, planned and performed most experiments, and co-wrote the manuscript. As such, each name can be interchanged in first position of the manuscript. S.L. performed the majority of computational analyses. E.W and M.T.W performed CRISPR screening and thee majority of 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 though 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, 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 Financial Interests

I.A. is a consultant for Foresite Labs and receives research funding from AstraZeneca. A.T. is a Scientific Advisor to Intelligencia 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., AIChemy, Harmonic Discovery Inc., and Pfizer Boulder; O.A.-W. has received prior 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. JE is a compensated scientific advisor to Cytovia Therapeutics. J.E own stocks in Mnemo Therapeutic and Cytovia Therapeutics. J.E. has received a consulting fee from Casdin Capital. The Eyquem lab has received research support from Cytovia Therapeutic, Mnemo Therapeutics and Takeda. J.E. is a holder of patents pertaining to but not resulting from this work. The remaining authors declare no competing interests.

[8]Department of Mechanical and Aerospace Engineering, New York University, New York, NY 11201, USA

[9]Department of Biomedical Engineering, New York University, New York, NY 11201, USA

[10]Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA

[11]Cancer Center and Center for Regenerative Medicine, Massachusetts General Hospital, Boston, Massachusetts 02114, USA

[12]Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA

[13]Department of Molecular, Cellular, and Developmental Biology, University of Colorado Boulder, Boulder, CO 80309, USA

[14]Department of Pathology, St. Jude Children's Research Hospital, Memphis, TN 38105, USA

[15]Authors contributes equally to this manuscript

Abstract

B cell progenitor acute lymphoblastic leukemia (BCP-ALL) treatment has been revolutionized by T cell-based immunotherapies - including chimeric antigen receptor (CAR) T cell therapy and bi-specific T cell engager therapeutic, blinatumomab - targeting surface glycoprotein CD19. Unfortunately, many B-ALL patients will fail immunotherapy due to 'antigen escape' – the loss or absence of leukemic CD19 targeted by anti-leukemic T cells. Here, we utilized genome-wide CRISPR/Cas9 screening approach to identify modulators of CD19 abundance on human BCP-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 mRNA polyadenylation and stability. NUDT21 deletion in BCP-ALL cells increased the expression of CD19 and the sensitivity to CD19-specific CAR-T and blinatumomab. In human BCP-ALL patients treated with CAR-T and blinatumomab, upregulation of *NUDT21* mRNA coincided with CD19 loss at disease relapse. Together, these studies identify novel CD19 modulators in human BCP-ALL.

Editor summary:

Aifantis and colleagues use leukemic cell lines to identify modulators of CD19 expression that have the potential to improve therapeutic strategies.

Loss of expression of the transmembrane surface glycoprotein CD19 on B cell progenitor acute lymphoblastic leukemia (BCP-ALL) blasts drives resistance to chimeric antigen receptor (CAR)-T cell and bi-specific T cell engager therapy[1, 2, 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 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 BCP-ALL (Reh, NALM6 and 697) and two mature B cell neoplasm cell lines from chronic lymphocytic leukemia (CLL,

HG3) and B cell lymphoma (TMD8) (Fig. 1a). Stable Cas9 expression was achieved using lentiviral transduction of blasticidin-selectable Cas9 constructs. Blasticidin-resistant $Cas9^+$ cells lines were then transduced with the Brunello genome-wide sgRNA library[4] with sgRNA-expressing cells tracked by 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 number of sgRNAs/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 CRIPSR score) relative to non-targeting control sgRNAs (Extended Data Fig. 1a, Supplementary Table 1). To identify high-confidence and cell line-specific CD19 regulators, we selected the top-200 CD19 activators and top-200 CD19 repressors for all the five cell lines tested (Fig. 1b, Extended Data Fig. 1c,d, 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, Extended Data Fig. 1e,f). We observed the significant enrichment of *CD19* sgRNA within the $CD19^{lo}$ populations in all 5 human $Cas9^+$ B cell lines (Fig. 1b). Consistent with CD19 protein maturation occurring via the secretory pathway, endoplasmic reticulum (ER) and Golgi apparatus associated proteins such as *CD81*[5, 6], *CANX*[7] and *SEC61A1*[8] were identified as CD19 activators in $Cas9^+$ Reh, $Cas9^+$ 697 and $Cas9^+$ NALM6 lines (Fig. 1c,d, Extended Data Fig. 1c, g). The majority of BCP-ALL CD19 activators, including *PAX5* and *PTBP1,* which promote *CD19* promoter activation[9] and *CD19* intron 2 exclusion[10], 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 BCP-ALL indicated significant overlap with subunits of the TAF1 and YY1 protein complexes (Fig. 1e), which are required for RNA polymerase II-mediated transcription initiation[11].

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[11] and potentially regulate cell type-specific promoter-enhancer DNA loop formation[12]. To explore the role of ZNF143 in *CD19* gene activation, we performed RNA-sequencing 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 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 to 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, Supplementary Table 2). Flow cytometry indicated 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 to sgROSA-$mCherry^+$ cells (Fig. 2c, d).

ZNF143 promotes gene activation through its DNA-binding zinc fingers[13]. To assess the protein domains of ZNF143 required for activation of *CD19*, transduced Cas9$^+$ NALM6 cells with a pooled sgRNA library targeting multiple exons of *ZNF143*, *CD19*, *CD81* and 10 non-targeting sgRNAs and isolated CD19^{lo} and CD19^{hi} cells followed by sgRNA deep sequencing 12 days post-transduction (Fig. 2e, Supplementary Table 3). Using this 'domain screening' approach, we calculated the CD19 CRISPR score for sgRNAs targeting *CD19, CD81* and *ZNF143* was lower relative to non-targeting sgRNAs (Fig. 2e, 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 to non-targeting sgRNAs (Fig. 2e, Extended Data Fig. 2b, c), indicating this domain screening approach effectively identifies key protein domains critical for full-length surface CD19 protein production. sgRNAs targeting the ZNF143 C2H2 zinc finger domain resulted in a low CD19 CRISPR score compared to 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 ChIP-Seq dataset[14] 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 cells line, while no binding was detected in CD19$^-$ K562 erythroleukemia cell line (Fig. 2g).

Because 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$^+$697 and Cas9$^+$NALM6 cells found minimal impact on genome-wide intra-TAD structures, including no significant intra-TAD activity changes impacting the TAD harboring the *CD19* locus, compared to the corresponding sgROSA-mCherry$^+$ cells (Extended Data Fig. 2d), consistent with previous studies[15]. Virtual 4C analysis using the *CD19* promoter as the 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 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 BCP-ALL with CD19 mRNA expression patterns across multiple primary human hematopoietic cell types. Using single cell RNA-sequencing (scRNA-seq) data from four healthy and seven BCP-ALL human bone marrow specimens[16] (Fig. 3a), we confirmed the overrepresentation of human B cell progenitors (defined by Seurat reference dataset as Prog_B1 and Prog_B2) within the leukemic bone marrow samples (Fig. 3b). *CD19* mRNA expression was restricted to the B cell lineage with highest expression in healthy and BCP-ALL Prog_B1 (characterized by high *EBF1* and *MKI67*) and Prog_B2 (characterized by high *CD19*, *PAX5*, *EBF1* and *MKI67*) subsets compared to all other CD19-expressing cell populations, including naïve B cells, memory B cells and plasmablasts (Fig. 3c). Notably, *CD19* mRNA expression was highest in leukemia-associated Prog_B2 cells (Fig. 3c,d, Supplementary Table 4). Ranking the expression of individual candidate genes in the *CD19*^{hi} Prog_B2 compartment in healthy and leukemic conditions showed the mRNA expression of *PAX5,* a known CD19 activator[9], and the transcript for the RNA-binding protein *NUDT21* significantly associated with *CD19*

mRNA expression (Fig. 3d, Extended Data Fig. 3a, b, Supplementary Table 4). *NUDT21* mRNA expression positively correlated with *CD19* mRNA in healthy and leukemic Prog_B1 and Prog_B2 B cell progenitor subsets, while this significant correlation was reduced in naïve, memory and plasmablast B cells (Fig. 3e, 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 BCP-ALL cells that were sorted into quartiles based on their expression of CD19 (Fig 1a, Fig. 3f,g).

Because 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. sgRNAs targeting of 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 to 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 day 2, day7 and day 14 post-transduction indicated that expression of sgNUDT21 reduced the growth of $Cas9^+$Reh, $Cas9^+$697, $Cas9^+$NALM6 and $Cas9^+$HG3 cells when compared to sgROSA and sgCD19 (Extended Data Fig. 4b). To determine whether the phenotypes elicited by sgNUDT21 in BCP-ALL could be rescued, we ectopically expressed either a 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 to sgROSA-mCherry, which was rescued by overexpression of $NUDT21^{sgRes}$ (Extended Data Fig 4c–e).

Next, to assess if CD19 expression was dynamically regulated by the reversible inhibition of NUDT21, we used 697 cells transduced with the doxycycline (dox)-inducible type VI CRISPR-Cas13d[17] to knockdown *NUDT21* and *CD19* mRNAs in an inducible manner (see Methods). $Cas13d^+$ 697 cells lentivirally transduced with sgNUDT21-GFP, sgCD19-GFP or sgNTC (non-targeting control)-GFP were treated with dox for six days followed by dox withdrawal *in vitro*, and the expression of CD19 was tracked every two days by flow cytometry (Fig. 4c, d). Dox-dependent induction of sgCD19-GFP significantly reduced the expression of CD19 on $Cas13d^+$697 cells compared to sgNTC-GFP+ $Cas13d^+$ 697 cells, while 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), while restoration of NUDT21 by dox withdrawal resulted in 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, retrovirally transduced them with MSCV-based constructs, which allow the ectopic expression of human *BCR::ABL1*^{p190} linked to GFP, and maintained them 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 eight days significantly increased expression of CD19 protein and reduced cellular fitness in *BCR::ABL1-GFP$^+$ Cre $^{ERT2+}$Nudt21$^{fl/fl}$* cells when compared 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 to vehicle-treated controls (Extended Data Fig. 4g) but induced transient 4-OHT toxicity (Extended Data Fig. 4i). These data suggested a conserved role NUDT21 in limiting CD19 expression in transformed B cell progenitor cells in human 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’ untranslated region (3’ UTR) of pre-mRNA molecules[18,]. To understand the role of NUDT21 in *CD19* mRNA processing, we analyzed the Cas9$^+$ NALM6 cells-based CRISPR-based exon mutagenesis dataset. sgRNAs targeting of the nucleotide hydrolase domain that encompasses the NUDT21 RNA-binding domains resulted in high CD19 CRISPR scores (Extended Data Fig. 5a), suggesting this domain was involved in repression of *CD19* 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, 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’ UTR regions, with the majority of NUDT21 peaks mapped to intronic and exonic regions (Extended Data Fig. 5c, Supplementary Table 5), which may reflect direct interactions with RNA polymerase II and the pre-mRNA processing machinery occurring during active transcription[19.]

Loss of NUDT21 may alter gene expression, partly through mRNA 3’ UTR shortening[20, 21]. We performed 3′ massive analysis of c-DNA ends sequencing (MACE-seq) to investigate the 3’ UTR changes that occurred following NUDT21 ablation in Cas9$^+$Reh, Cas9$^+$697, Cas9$^+$NALM6 and Cas9$^+$TMD8 cells. Quantification of alternative polyadenylation (QAPA) analysis indicated evidence of specific 3’ UTR shortening of *CD19* mRNA in sgNUDT21-mCherry$^+$ Cas9$^+$ Reh, Cas9$^+$697 and Cas9$^+$NALM6 cells, but not Cas9$^+$TMD8 cells, when compared to sgROSA-mCherry$^+$ counterparts (Fig. 4f, Extended Data Fig. 5d,e, 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

(adjusted *p*-value < 0.1) compared to sgROSA-mCherry$^+$ cells (Supplementary Table 7). Expression of *CD19* mRNA was increased in sgNUDT21-mCherry$^+$ Cas9$^+$ Reh, 697 and NALM6 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 PSI > 0.2 and FDR < 0.05) impact splicing of *CD19* mRNA compared to ROSA-mCherry$^+$ cells(Extended Data Fig. 5f).

Because 3' UTR shortening may act to promote mRNA stability, we assessed *CD19* mRNA decay at multiple timepoints following actinomycin D treatment of sgNUDT21-mCherry$^+$ Cas9$^+$ Reh, NALM6, HG3 and TMD8 cells. Real-time quantitative PCR 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, compared to sgROSA-mCherry$^+$ counterparts (Fig. 4h). Analysis of the genome-wide CRISPR screen dataset found multiple pre-mRNA polyadenylation site regulators[22], 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 distal polyadenylation site, such as NUDT21 and CPSF6, associated with a high CD19 CRISPR score (Fig. 4i), indicating increased CD19 expression, while targeting of factors that facilitate proximal polyadenylation, such as PCF11 and CLP1, associated with a low CD19 CRISPR score (Fig. 4i, 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$^{\Delta UTR}$) in CD19^{dim} Cas9$^+$ Reh cells. CD19$^{\Delta UTR}$-GFP+ Reh cells had significantly higher expression of CD19 compared to CD19^{FL}-GFP+ cells (Fig. 4j,k), suggesting the *CD19* 3'UTR may limit the expression of CD19 protein. As such, NUDT21 functioned 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, 24, 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 BCP-ALL niche[16, 26]. Five days post-transduction of sgROSA-mCherry, sgCD19-mCherry or sgNUDT21-mCherry, Cas9$^+$ Reh cells were infused into human umbilical vein endothelial cell (HUVEC) 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 CD19 CAR-T cells killed less sgCD19-mCherry$^+$ Cas9$^+$ Reh cells than sgROSA-mCherry$^+$ Cas9$^+$ Reh cells, although

this did to reach statistical significance (Fig. 5b). Moreover, CD19 CAR-T cells killed significantly more sgNUDT21-mCherry$^+$ Cas9$^+$ Reh cells compared to sgROSA-mCherry$^+$ and sgCD19-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 to CAR-T-treated sgROSA-mCherry$^+$ and sgCD19-mCherry$^+$ Reh cells (Fig. 5c, Extended Data Fig. 6b).

Blinatumomab is a CD3-CD19 bispecific T cell engager formed by linking a distinct pair of immunoglobulin variable domains capable of binding human CD19 and CD3, respectively, with a short 25-amino acid linker sequence. Upon 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 blinatumomab, and counted mCherry$^+$ cells following 24-hours of co-culture as a measure of cell killing (normalized to vehicle-treated control). We found that CD8$^+$ T cell combined with blinatumomab treatment led to significantly more killing of sgNUDT21-mCherry$^+$ Cas9$^+$ NALM6 cells compared to sgROSA-mCherry$^+$ Cas9$^+$ NALM6 cells (Fig. 5d, Extended Data Fig. 6c). In addition, limiting dilutions of CD19 CAR-T cells killed significantly more sgNUDT21-mCherry$^+$ Cas9$^+$ 697 cells compared to CAR-T-treated sgROSA-mCherry$^+$ Cas9$^+$ 697 cells (Fig. 5e, Extended Data Fig. 6d), indicating that NUDT21 limits the sensitivity of leukemic cells to CD19-directed therapeutics *ex vivo.*

Because CD19-directed blinatumomab or CAR-T cell treatment can select BCP-ALL blasts with reduced expression of CD19 [1, 3, 28] we assessed the temporal impact of CD19-directed treatment on expression of CD19 and NUDT21 expression in primary human BCP-ALL. We identified a BCP-ALL patient, identified as SJBALL042246[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 of CD19, leading to disease relapse. Using publicly-available, leukemic blast-enriched primary RNA-sequencing 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 to 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 single-cell RNA-seq data from a single BCP-ALL patient[28] obtained prior to CD19 CAR-T infusion and at disease relapse of a CD19^{dim} leukemia following CD19 CAR-T treatment. Consistent with the original scRNA-seq data analysis[28], we identified a distinct group of CD19$^+$CD34$^+$ cells pre-CAR-T cell treatment and CD19^{dim}CD34$^+$ leukemic cells at disease relapse following 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 to pre-CAR-T treatment CD19$^+$CD34$^+$ leukemic blasts (Fig. 5g, Extended Data Fig. 7a,b). These data indicate that NUDT21 expression may be modulated after CD19-directed therapeutic exposure in primary human BCP-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 differentiation[29] and immunotherapy responsiveness[30]. This 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 BCP-ALL, CD19 CAR-T cell treated samples remains limited. Here we highlighted two clinical examples of concurrent reduction of CD19 mRNA and increased expression of NUDT21 mRNA following CD19-directed therapy in blinatumomab and CD19 CAR-T-treated BCP-ALL patients. These early clinical correlations may inform future kinetic analysis of CD19 regulators throughout immunotherapy.

METHODS

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Iannis Aifantis (Ioannis.Aifantis@nyulangone.org).

Materials Availability

Human and Murine cell lines generated and plasmids used in this study are available upon request from the Lead Contact.

Data Availability

Data is available under the GEO SuperSeries accession GSE190844.

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, according to national and institutional guidelines. All animal experiments were performed in accordance with protocols approved by the New York University Institutional Animal Care and Use Committee (IACUC, Aifantis Protocol # 160411, IA16–00008).

Cell lines and culture

Human cell lines Reh, 697, NALM6, HG3, TMD8, GM12878 and K562 were cultured in RPMI-1640 media supplemented with 20% fetal bovine serum, 55 μM β-mercaptoethanol (Gibco # 21985023) and penicillin, streptomycin (Gibco #15140122). HEK293T were purchased from ATCC CRL-1573 and Platinum-E cells were purchased from Cell Biolabs RV-101. HEK293T and Platinum-E were cultured in DMEM media supplemented with 10% fetal bovine serum and penicillin, streptomycin (Gibco #15140122). OP9 were purchased from ATCC CRL-2749. OP9 were cultured in IMDM media supplemented L-Glutamine (Gibco # 31980030) with 15% fetal bovine serum and penicillin, streptomycin (Gibco

#15140122) with supplementation of recombinant murine IL-7 (217–17, Peprotech), Flt3L (250–31L, Peprotech) and SCF (250–03, Peprotech) 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, 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.

All lentiviral transfections were performed in HEK293T cells using Polyethylenimine (PEI) reagent at 4:2:3 ratios of sgRNA or Cas9 or Cas13d construct: pVSVG: pPax2 in OPTI-MEM solution. All retroviral transfections were performed in Platinum-E cells using Polyethylenimine (PEI) reagent with 10 μg transfer plasmid in OPTI-MEM solution. All amphitropic retroviral transfections were performed Polyethylenimine (PEI) reagent at 5:2:5 ratios of MSCV-based construct: pVSVG: pCL-Ampho retrovirus packaging construct (Novus Biologicals, NBP2–29541) in OPTI-MEM solution Viral supernatant was collected 48-hrs post-transfection. Spin infections were performed at room temperature at 1,500g for 90 mins with polybrene reagent (Fisher Scientific). Reh, 697, NALM6, K562, HG3, TMD8 Cas9-expressing cell lines were generated by lentiviral transduction with retroviral Cas9–2A-blast (Addgene, plasmid #73310). Human GM12878 Cas9 lines were generated by lentiviral transduction with Cas9–2A-EGFP (Addgene plasmid #63592). 697 Cas13d-expressing cell line was generated by transduction with lentiviral tet-inducble RfxCas13d (Addgene, plasmid # 138149). Doxycycline (Sigma) treatment was performed at 0.5 μg/mL supplemented every two days in culture.

Cell lines were routinely monitored for mycoplasma contamination by PCR using ATCC Universal Mycoplasma Detection Kit (20–1012K). 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 8–12-week-old *ROSA26-Cre^{ERT2+};Nudt21$^{+/+}$* and *ROSA26-Cre^{ERT2+};Nudt21$^{fl/fl}$* mice[31]. Bone marrow cell transduction with BCR-ABL1-expressing retroviral constructs (Addgene Plasmid #38185) was performed in in Iscove's Modified Dulbecco's Medium (IMDM) buffer with GlutaMax (Gibco) with 15% FBS, murine stem cell factor (SCF) (100 ng/mL, Peprotech), murine Flt3 ligand (50 ng/mL, Peprotech), murine IL-7 (10 ng/mL) and polybrene (1 μg/mL) at a cell concentration of 1×10^6 cells per mL, and incubated overnight at 37°C with 5% CO_2. Transduced cells were cultured on OP9 stromal cells for 7 days in (IMDM) buffer with GlutaMax (Gibco) with 15% FBS, murine stem cell factor (SCF) (100 ng/mL, Peprotech), murine Flt3 ligand (50 ng/mL, Peprotech), murine IL-7 (10 ng/mL), 55 uM β-mercaptoethanol and Penicillin/Streptomycin. Following 7 days, SCF and Flt3L were withdrawn from culture media, and cultured on OP9 stromal cells for a further 21 days prior to physical separation of hematopoietic cells from OP9 stromal cells. Resulting suspension cell lines were cultured continuously in IMDM buffer with GlutaMax (Gibco) with 15%

FBS and penicillin/streptomycin. 4-Hydroxytamoxifen (4-OHT, 68392–35-8, Sigma) was supplemented at 50 nM daily for five consecutive days.

Flow Cytometry and cell sorting

Cell line suspensions were subjected to centrifugation at 400 × 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 × 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 × 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 × 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 × 10^6 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 ($DAPI^{low}$) 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 version 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). Custom sgRNAs were designed using http://benchling.com/ with a quality score of 70 or higher to minimalize off-target effects. Pooled sgRNA oligos were then synthesized by Twist Bioscience (https://twistbioscience.com/) on a 12K array and amplified using custom array primers followed by subcloning into a lentiviral sgRNA GFP-tagged vector (LRG) (Addgene plasmid no. 65656) as previously described[32]. All sgRNA sequences used in this study are provided (Supplementary Table 8). Genome-wide CRISPR screens were performed using the human Brunello knockout library (Addgene #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 Feng Zhang's laboratory[33] and subcloned into the lentiviral sgRNA vectors. Cas13d sgRNAs were

cloned into LRGv2.1 (Addgene #108098) and Cas9-based individual sgRNAs were cloned into LRCherry2.1 (Addgene #108099) (Supplementary Table 8).

Antigen-based CRISPR screens

For CRISPR screens, Cas9-expressing cells were infected with either CRISPR exon mutagenesis (this study) screening or Brunello genome-wide library at a low MOI (~0.3). At day 12 post-transduction, Cas9-expressing cells were incubated with APC anti-human CD19 antibody (Biolegend #302212 HIB19 clone) and then cell sorting was performed using the SY3200™ highly automated parallel sorting (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 (gDNA) extraction was performed using Qiagen DNA kit (#51306) according to manufacturer's protocol. PCR amplification of the Brunello library was performed based on a previous study[34]. 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. Custom P5 (staggering primers) and P7 (unique barcode primers) (Supplementary Table 9) were used to generate barcoded libraries which were subsequently sequenced using single-end 50 bp HiSeq 4000 (Illumina). Deconvolution of sgRNA reads was performed using a custom 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 $\log_2$-transformed by adding 1 to all values (CPM+1). The CD19 CRISPR scores were calculated by the $\log_2$ fold change ($CD19^{high}/CD19^{low}$) for each sgRNA.

Immunoblotting

For immunoblot analysis, cell lines were pelleted and lysed using RIPA lysis and extraction buffer (Thermo Fisher, # 89900). The lysates were boiled with Laemmli buffer, resolved by SDS-PAGE, transferred to PVDF membranes and proteins visualized by immunoblotting. The following antibodies were used for immunoblot analysis: CD19 (1:500, Cell Signaling Technology, #3574), NUDT21 (1:100, Santa Cruz Biotechnology, sc-81109), YY1 (1:500, Santa Cruz Biotechnology, H-10 sc-7341), ZNF143 (1:1000, Novus Biologicals, 2B4 H00007702-M01) and Actin (1:1000, Millipore, MAB1501R).

eCLIP

eCLIP studies were performed by Eclipse Bioinnovations Inc (San Diego) according to the published single-end seCLIP protocol[35] with the following modifications. Cas9-expressing Reh, 697, NALM6 and TMD-8 cells were UV crosslinked at 400 mJoules/cm2 with 254 nm radiation. Cells were lysed using 1mL of eCLIP lysis mix and were subjected to two rounds of sonication for 4 minutes with 30 second ON/OFF at 75% amplitude. 5μg of pre-validated NUDT21 (Santa Cruz Biotechnology, sc-81109) antibody was then pre-coupled to 50ul of Anti-Rabbit IgG Dynabeads (ThermoFisher), added to lysate equivalent to 100μg of RNA, and then incubated overnight at 4°C. Prior to 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. IP and input samples were cut from

the membrane at the relative band size to 75kDa above. RNA adapter ligation, IP-western, 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 seven days. mCherry$^+$ cells were isolated and RNA was extracted using the RNeasy Plus Mini Kit (Qiagen, #74316). Massive Analyses of cDNA Ends (MACE-Seq) was performed by GenXPro GmbH in Frankfurt am Main using the MACE-Seq kit according to the manual of the manufacturers. 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 minimum number of cycles was used to produce a library that was sequenced on an Illumina NextSeq500 machine with 1× 75 bps[36].

NUDT21 rescue and Ectopic CD19 variant expression

NUDT21 codon-optimization was performed based on the NUDT21 full-length cDNA sequence (Ensembl NUDT21-201, ENST00000300291.10) to exclude EcoRI, XhoI and *sgNUDT21.336* 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 Empty vector (EV) pMSCV-IRES-GFP (Addgene #9044) via EcoRI and XhoI sites. Cas9$^+$ NALM6 BCP-ALL cells were co-transduced with amphotrophic virus harboring either pMSCV-IRES-GFP;pLRCherryv2.1-sgROSA, pMSCV-NUDT21^{sgRes}-IRES-GFP;pLRCherryv2.1-sgROSA, pMSCV-IRES-GFP;pLRCherryv2.1-sgNUDT21.336 or pMSCV-NUDT21^{sgRes}-IRES-GFP;pLRCherryv2.1-sgNUDT21.336 and cultured for six days prior to flow cytometry, then cultured for a further five days. Fitness was assessed by calculating the percentage GFP$^+$mCherry$^+$ cells at day 6 and day 11 post-transduction.

Full length (FL) *CD19* cDNA (Ensembl CD19–202, ENST00000538922.8) including (CD19^{FL}) or excluding (CD19$^{\Delta UTR}$) the 3' UTR sequence was synthesized using gBlock synthesis (IDT) then subcloned into pMSCV-IRES-GFP (Addgene #9044) via Gibson Assembly (New England BioLabs, E2611) (Supplementary Table 10). Cas9$^+$ Reh cells were co-transduced with amphotrophic virus harboring either pMSCV-IRES-GFP, pMSCV-CD19$^{\Delta UTR}$-IRES-GFP or pMSCV- CD19^{FL}-IRES-GFP. Cells were cultured for 72-hours prior to flow cytometry.

RNA stability

Cas9$^+$ Reh, NALM6, TMD8 and HG3 cell lines were transduced with LRCherry2.1-sgROSA/NUDT21 and cultured for seven days. Cells were treated with 5 μg/mL actinomycin D (Cell Signaling Technology, #15021S) at 37 °C and harvested at the indicated time points. RNA was extracted using the RNeasy Plus Mini Kit (Qiagen, #74316) according to the manufacturer's instructions and converted to cDNA using the High-Capacity RNA-to-cDNA Kit (Life Technologies, #4387406). cDNA was subjected to qPCR and gene expression was measured using LightCycler 480 SYBR Green I Master

(Roche Diagnostics, #04707516001). Data were analyzed by ΔΔCt method normalizing to *18S rRNA*. qPCR primers used:

18S rRNA Forward: 5'-ACCGCAGCTAGGAATAATGGA-3'

18S rRNA Reverse: 5'-GCCTCAGTTCCGAAAACCA-3'

CD19 Forward: 5'-TGAAGACCTCGAGCAGATGA-3'

CD19 Reverse: 5'-CATGCACACATCCTAAGCAAC-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-ITR plasmids containing the 1928z CAR and *TRAC*-targeting homology arms for homology directed repair were used as previously described[27]. The AAV-ITR-containing plasmid was packaged into AAV6 using Polyethylenimine-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 (NEB) and Proteinase K (Qiagen), using primers targeting the left homology arm (Forward: CTTTGCTGGGCCTTTTTCCC, Reverse: CCTGCCACTCAAGGAAACCT). qPCR was performed using SsoFast Eva Green Supermix (Bio-Rad) 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 (ThermoFisher) (1:1 beads:cell) in X-VIVO 15 medium (Lonza) supplemented with 5% human serum (Gemini Bioproducts), penicillin-streptomycin (Thermo Fisher, 50 U/ml), IL-7 (Miltenyi Biotec, 5 ng/ml), and IL-15 (Miltenyi Biotec, 5 ng/ml) and cultured at 10^6 cells per ml. Medium was exchanged every two to three days, and cells were resuspended at 10^6 cells per ml.

After 48 hours of T cell activation, cells were detached from CD3/CD28 Dynabeads, and 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 recombinant Cas9 protein (QB3 MacroLab) with 120 pmol *TRAC* sgRNA (Synthego, CAGGGUUCUGGAUAUCUGU) at 37 °C for 15 min. Cells were resuspended in P3 primary cell solution (Lonza) (2×10^6 live cells per electroporation) and mixed with RNP, followed by electroporation using the EH115 Nucleofector protocol. Cells were then diluted into serum-free medium (2×10^6 cells per ml) and incubated at 37 °C, 5% CO_2. Recombinant AAV6 donor vector was added to the culture 30–60 min after electroporation at a multiplicity of infection of 5×10^4, and cells were incubated overnight. The next day, the serum-free AAV-containing medium was removed and cells were resuspended in fresh complete medium and expanded using standard culture conditions (37 °C, 5% CO_2, and complete medium replenished as needed to maintain a density of 1×10^6 cells per ml every two to three days). Knockout and knock-in efficiency were evaluated by staining for the TCR with an anti-TCRα/β antibody (Miltenyi Biotec)

and staining for the CAR with a goat anti-mouse $F(ab')_2$ (Jackson ImmunoResearch, 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 [26, 37]. Briefly, the microfluidic device with three distinct functional regions (i.e., a central sinus region, an inner ring region and the outer ring channels) was fabricated using standard soft lithography replica molding technique[26, 37]. The central sinus region was vascularized by human umbilical vein endothelial cells (HUVECs, Lonza, catalog no. C2519A), while the inner ring region and the outer ring channels were respectively loaded with HUVECs and fibroblast cells (Lonza, catalog no. CC-2512) embedded in 3mg/ml fibrin hydrogel (Sigma, catalog no. G2500) to maintain 3D culture. The *ex vivo* devices were cultured with a mixture cell culture media of EGM-2 (Lonza, catalog no. CC-3162) and FGM-2 (Lonza, catalog no. CC-3132) at 2:1 vol/vol with 25 ng/ml VEGF (RayBiotech, catalog no. 230-00012) and monitored for about 7 days to ensure the formation of vascular network. Once the microvessel network was successfully established, 10,000 mCherry+ B-ALL cells and 10,000 T-cells (TRAC CAR and TRAC KO, respectively[27]) were mixed and loaded from the central sinus into the whole microvessel network and cultured for 24 hours with RPMI1640. The devices were then imaged with 20× objective using Zeiss microscopy integrated with a live-cell incubator and using NIH ImageJ, the number of $mCherry^+$ B-ALL cells was manually counted in each images taken. Following 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 washes for 3 times, the devices were imaged using Nikon Spinning Disk confocal microscopy with 40× objective and the number of $mCherry^+$ B-ALL cells forming synapse/contact with CD3 T-cells was manually counted in each images 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 group were compare using GraphPad, Prism. To quantify viability of leukemia cells after CAR T-cell treatment, 2–4 image fields (20x) were randomly chosen for each device, and each independent experiment has four independent devices. To quantify the synapse formation, approximately 10 image fields (40x) 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 BCP-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 six days in RPMI 1640 supplemented with 20% FBS, 55 μM β-mercaptoethanol and Penicillin/Streptomycin. Cell suspensions were centrifuged at 400 × g at 4°C for 5 min, and supernatant discarded. Cell pellets were re-suspended in RPMI 1640 supplemented with 20% FBS, 55 μM β-mercaptoethanol and Penicillin/Streptomycin and plated at 5×10^4 cells per well in flat-bottom 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, 130–096-495) according to 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 blinatumomab (BSAB-L002, Lot #Tandem scFv-0119, Creative BioLabs) 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 version 10 (Treestar, BD Biosciences). In addition, Countbright Plus Absolute Counting Beads were identified based on forward and side scatter ('Bead Region' gate, Extended Data Figure 6) then fluorescence verification using the APC-Cy7 channel (**'Beads' gate**, Extended Data Figure 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. Following co-culture, cell suspensions were centrifuged at 400 × g at 4°C for 5 min, and supernatant discarded. Cells were stained with APC-conjugated anti-human CD19 (1:200 concentration, #302212, BioLegend), and FITC-conjugated anti-human CD3 (1:200, #11–0038-42, eBioscience) in the presence of Countbright Plus Absolute Counting Beads (Invitrogen, C36995). mCherry$^+$ cell counts were determined as described for blinatumomab treatment.

Bioinformatic Analysis

ENCODE transcription factor binding and gene set enrichment analysis— Transcription factor chromatin immunoprecipitation (ChIP) data was downloaded from the ENCODE database. Signal p-value, a statistical significance of the signal at a location compared to IgG control, was used as a bigwig file provided from the database. Gene set enrichment analysis of ENCODE ChIP-X data was performed using Enrichr[38].

RNA-sequencing analysis—Cas9+ Reh, 697, NALM6, TMD8, HG3 and K562 cell lines were transduced with LRCherry2.1-sgRNA (sgZNF143 or sgNUDT21 in addition to negative control sgROSA) and cultured for seven days. mCherry$^+$ cells were isolated and RNA was extracted using the RNeasy Plus Mini Kit (Qiagen, #74316). Nudt21-conditional knockout B-ALL cells were subjected to five days 4-OHT treatment, then cultured an additional two days, then cells were isolated and RNA was extracted using the RNeasy Plus Mini Kit (Qiagen, #74316). RNA-seq libraries were prepared using NEXTflex Rapid Illumina Directional RNA-seq Library prep kit as per 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 aligner[39] with default parameters. For differential expression analysis, we used the DESeq2 R package[40] with raw read counts 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 (CPM) normalized expression values were used for the MA plots.

Hi-C analysis—Cas9$^+$ Reh, 697 and NALM6 cell lines were transduced with LRCherry2.1-sgROSA/NUDT21 and cultured for seven days. 100,000 mCherry$^+$ cells were subjected to Hi-C sample preparation according to manufacturer's guidelines (Arima Genomics, # A410030). Hi-C data was analyzed with the HiC-bench platform[41]. 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-mem[42] with "-A1 -B4 -E50 -L0" parameter. After alignment, multi-mapped reads, read-pairs with only one mappable reads, 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 study[43]. Briefly, we identified common TADs between ZNF143-depleted cell lines and the control cell lines where TAD boundaries of each cell line are overlapped within 120kb. We set the minimum TAD size to 400kb and maximum range of interaction to 2Mb. Then we averaged the Hi-C interactions within each common TAD and performed a paired two-sided t-test followed by multiple testing correction using false-discovery rate (FDR) estimation.

eCLIP analysis

We followed the data processing as previously described[44]. Briefly, adapters and adapter-dimers sequences were eliminated using Cutadapt (ref) and the trimmed reads were aligned to the human reference genome (GRCh37/hg19) using STAR aligner version 2.7.7[39] with the recommended parameter in the previous study[44] (--outFilterScoreMin 10, --outSAMunmapped Within, --outFilterMultimapNmax 1, outFilterMultimapScoreRange 1). After PCR-duplicates were removed using UMI-tools [45], we called eCLIP peaks using PureCLIP [46] with default parameters.

MACE-Seq analysis

Raw data was preprocessed using Cutadapt[5] 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[6]. Quantification of mapped reads to each gene was performed using HT-seq[7]. ENSEMBL-GTF data was used to provide genomic locations for quantification as well as additional data for annotation (such as gene name, gene description, GO-Terms etc). APA usage was analyzed as described in QAPA[47] using SALMON[48] for quantifying 3'UTRs of PolyASite. In short, QAPA quantifies APA levels of genes annotated by distinct 3'UTR sequences which were built from GENCODE database[49] and polyAsite database[50]. 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 greater than 10 percent.

Alternative Splicing Analysis

FASTQ files were first trimmed using Trim_galore (v0.6.4) 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 (v2.7.5) [39]. SAM files were subsequently converted to BAM files, sorted, and indexed using samtools (v1.9). For

quantification of alternative RNA splicing, BAM files generated by STAR/Samtools were analyzed using rMATS (v4.1.1) [51] using the GENCODE (v19) GTF annotation for Hg19 (GRCh37.p13). To utilize reads shorted than 51 b.p. resulting from adapter and/or QC 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 AS events, events were considered significant if (i) the inclusion level difference was greater than 20% compared to sgROSA, (ii) the False Discovery Rate (FDR) was smaller than 0.05, and (iii) there were a minimum of 20 reads mapping to the splice junction. For visualization by Sashimi plot, BAM files were loaded into Integrative Genomics Viewer (IGV) software (v2.7.0) and Sashimi plots generated using the following criteria: Junction coverage min= 10.

Bulk RNA-seq: Blinatumomab study

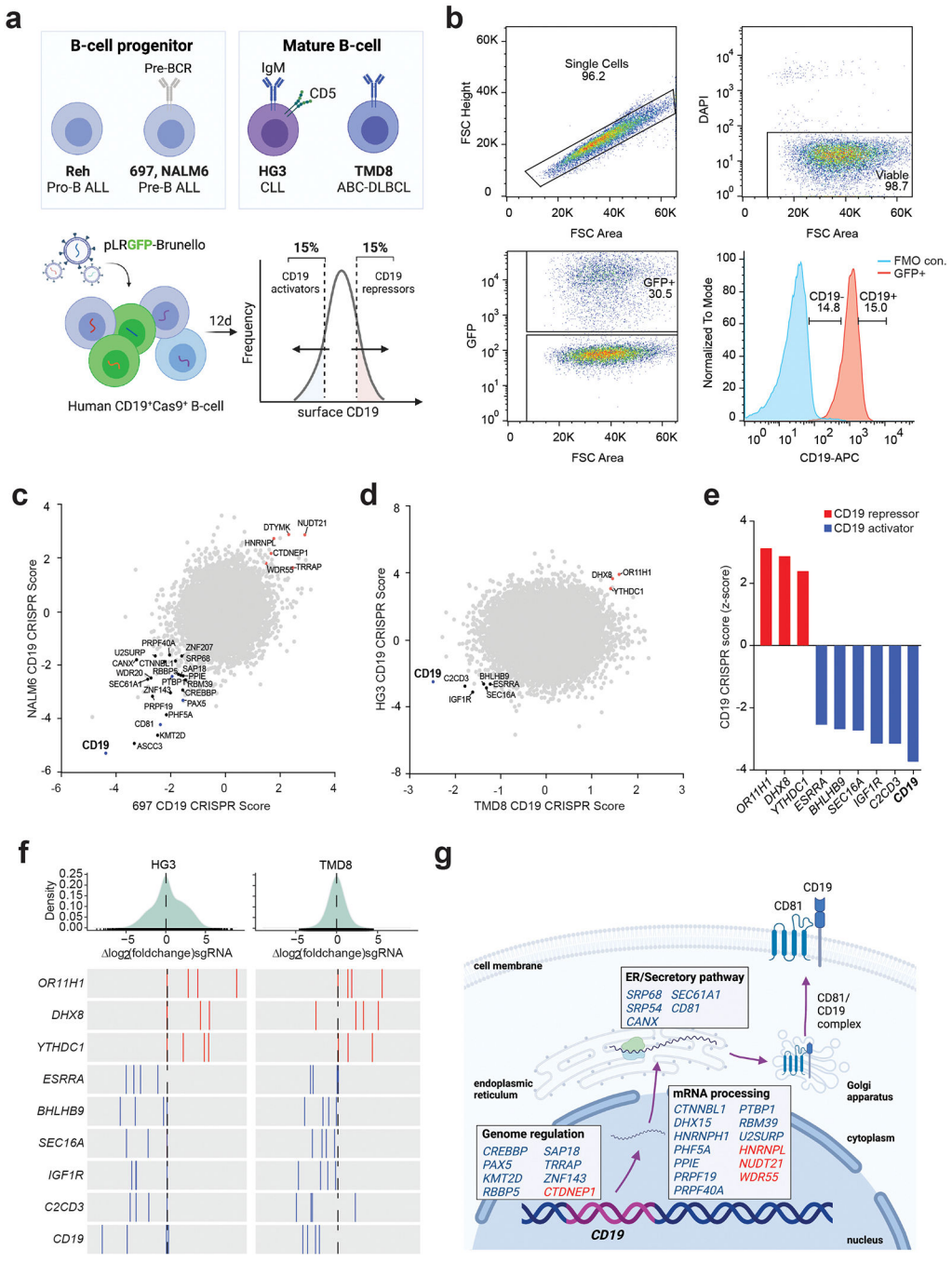
We analyzed publicly-avalaible matched bulk RNA-seq data of pre- and post-treatment samples of blinatumomab treatment from the previous study[3]. The previous study performed whole exome sequencing (WES) and RNA-sequencing with a cohort including 10 matched post-blinatumomab relapse samples. In brief, they performed WES 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 CPM normalized read counts that quantified by RSEM following by STAR alignment against GRCh37/hg19, and analyzed the fold change of post-blinatumomab sample over the pre-treatment sample.

Single-cell RNA sequencing data analysis

Previously published and publicly available single-cell RNA-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 algorithm[52] 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 algorithm[53] to the Harmony embeddings where we obtained K-nearest Neighborhood (KNN) graph with the parameters, k=30, d=30, then performed differential abundance testing between two groups (healthy vs B-ALL diagnosis) based on negative binomial generalized linear model implemented in Milo. We used Seurat v4 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 standardize them using 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 the Pearson correlation coefficient across the cell neighborhood per cell type. For the single-cell RNA-seq of pre- and post-CAR-T exposure samples[28], we downloaded the publicly available

pre-processed data from GEO database (GSE153697), used SCTransform for normalization, and performed PCA and UMAP for dimension reduction using Seurat v4.

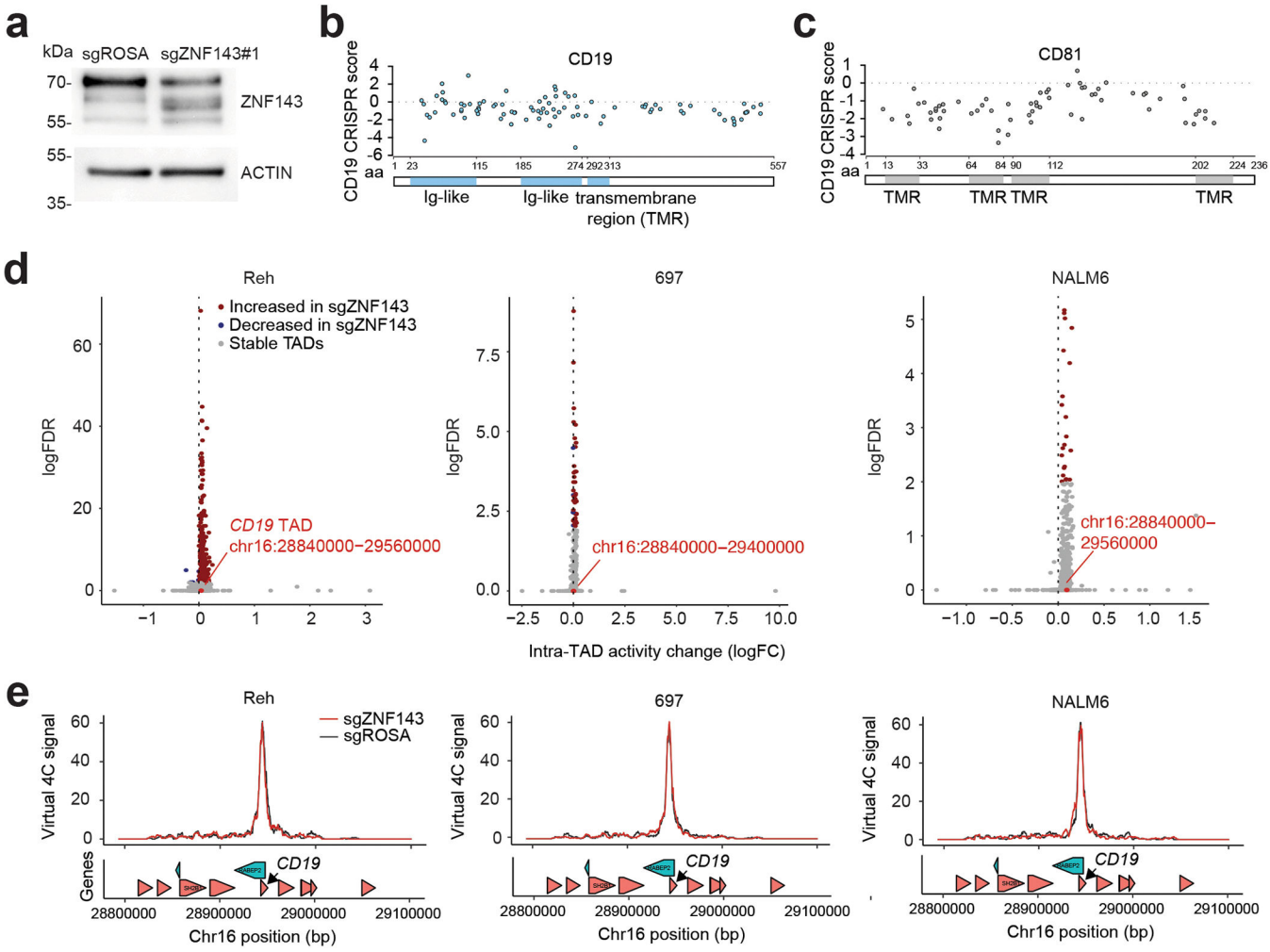
Extended Data



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

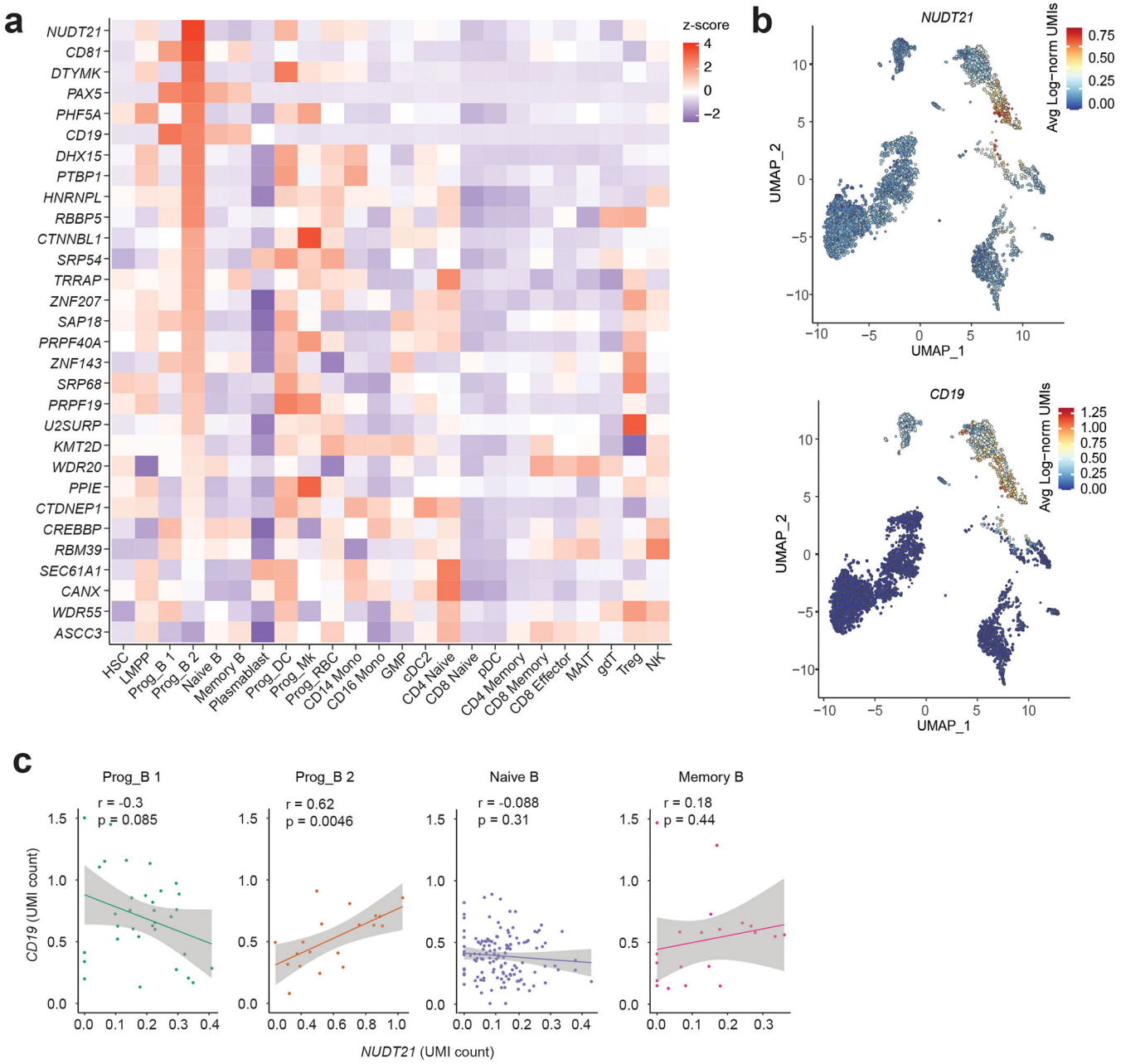
days culture. **(c-d)** Scatterplot showing CD19 score for individual gene candidates for CD19 activators and CD19 repressors comparing **(c)** human B-cell progenitor acute lymphoblastic leukemia (BCP-ALL) cell lines (Reh, 697 and NALM6) and **(d)** mature B cell (TMD8 and HG3) lines. **(e)** Waterfall 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 <

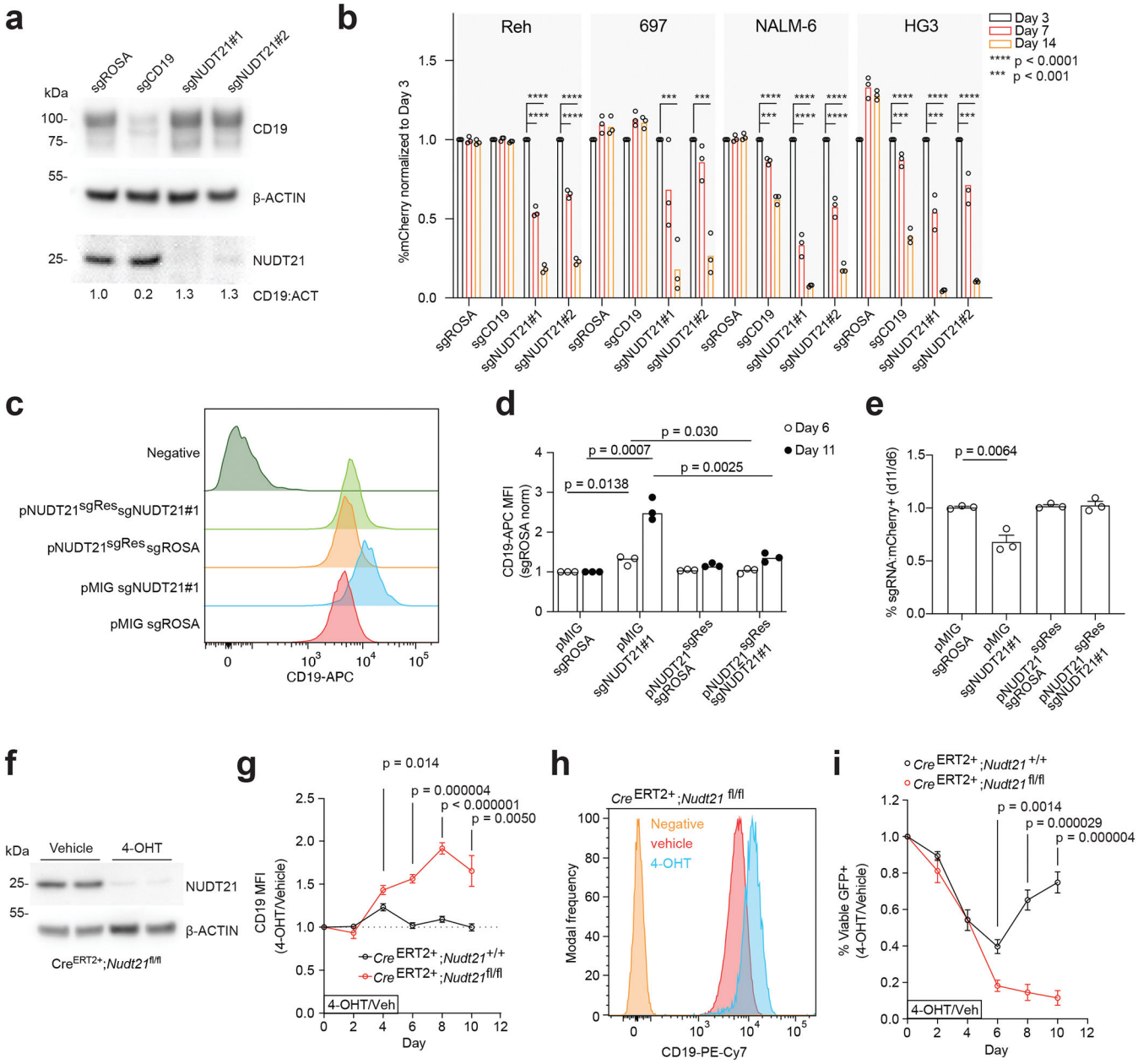
0.01 cutoff). **(e)** Virtual 4C analysis of *CD19* promoter viewpoint generated from Hi-C data of Reh, 697 and NALM6 expressing sgROSA and sgZNF143.



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

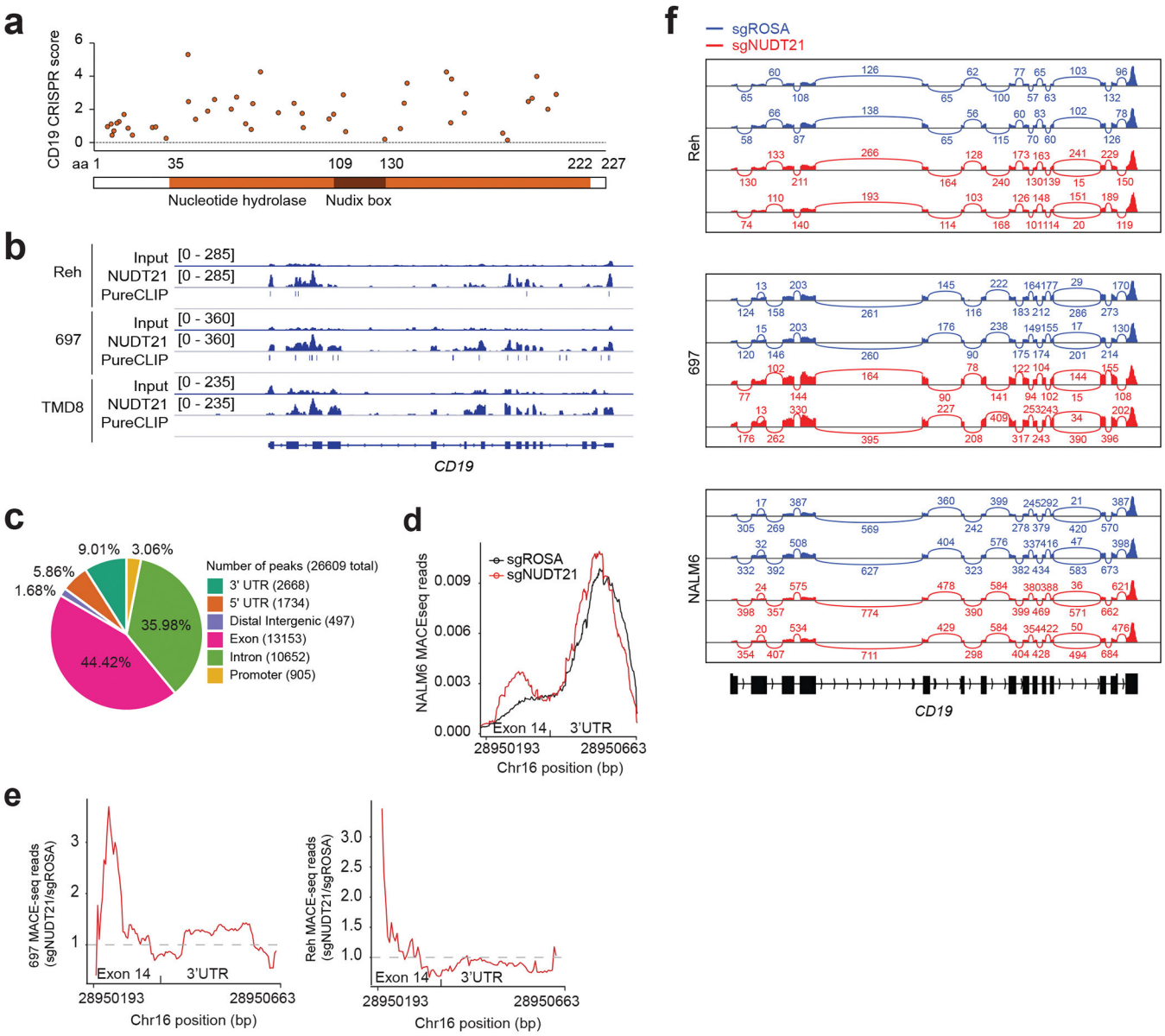
(UMI count) across Prog_B 1, Prog_B 2, Naïve B and Memory B cells in healthy bone marrow scRNA-seq data. r- and p-values calculated on the basis of Pearson's correlation.



Extended Data Fig. 4. NUDT21 represses CD19 expression and survival in human and murine B cell progenitors.

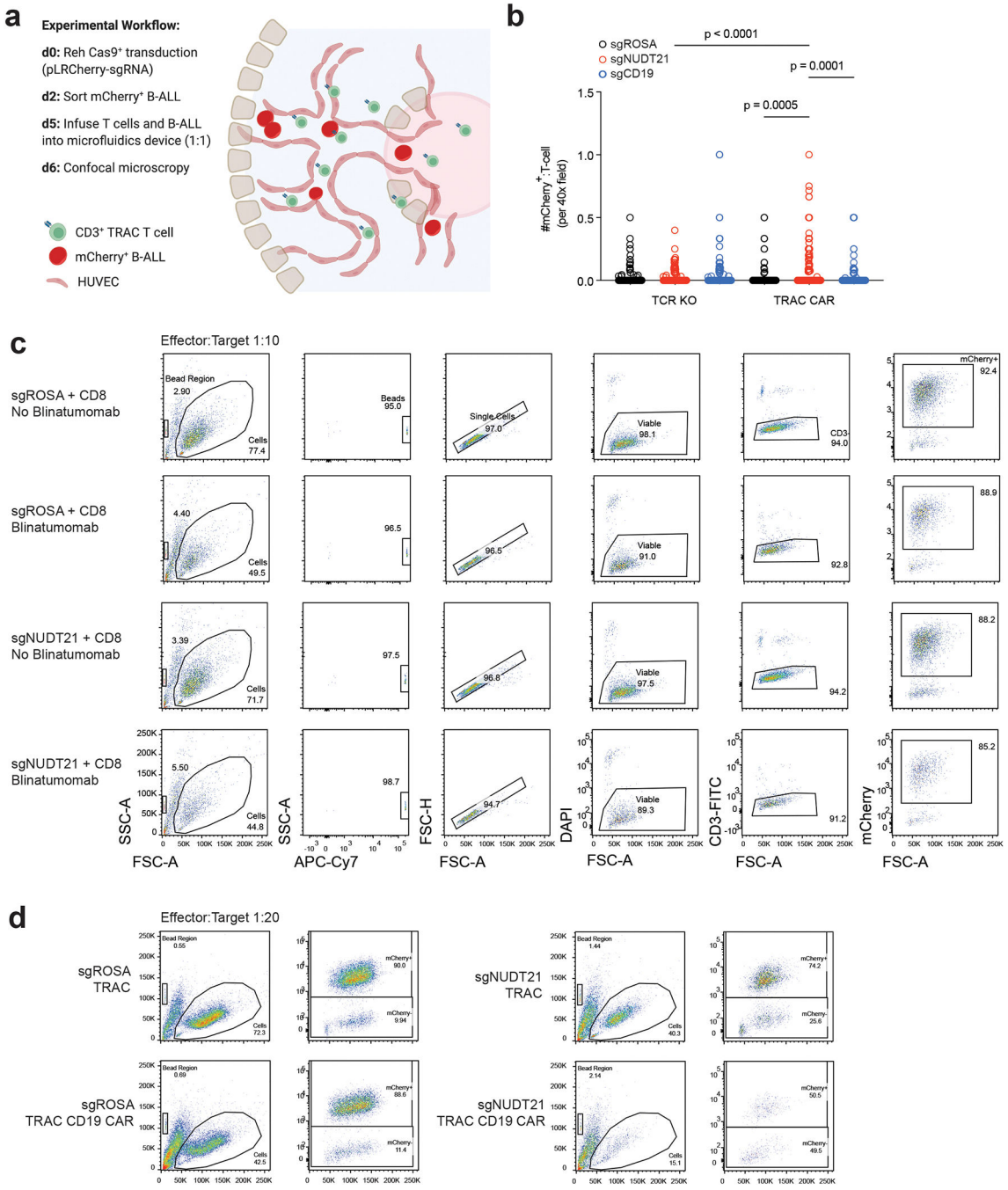
(a) Immunoblot analysis of CD19, NUDT21 and b-Actin in NALM6 cells expressing sgROSA or sgNUDT21#1 for seven days. Protein sizes indicated. CD19 levels normalized to Actin by densitometry. **(b)** Histogram of mCherry^{+} percentages normalized to Day 3 mCherry^{+} percentage across multiple cell lines (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.

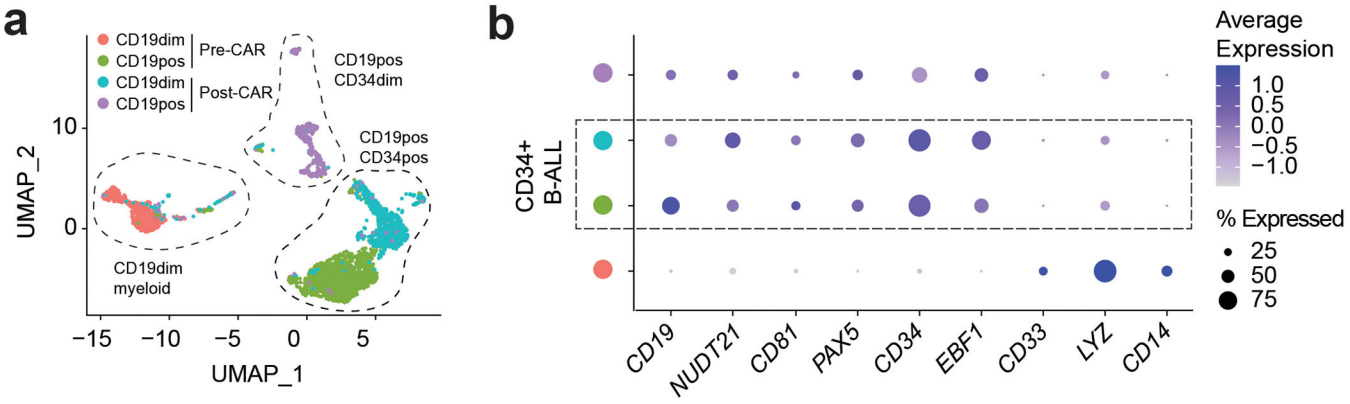
(a) CD19 scores for individual sgRNAs spanning the NUDT21 locus in NALM6 domain screen. **(b)** eCLIP read tracks spanning the CD19 locus in Reh, 697 and TMD8 cells. PureCLIP significant peak signals shown. **(c)** Pie-chart highlighting genomic distribution of eCLIP peaks shared by BCP-ALL cell lines. **(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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank all members of the Aifantis laboratory for discussions throughout this project. M.T.W. is supported by The Leukemia & Lymphoma Society Career Development Program, American Society of Hematology Restart Award, NIH/NCI K22 award (1K22CA258520–01), Cancer League of Colorado Research Grant (AWD#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 (CBET 2103219 to W.C.), the US National Institutes of Health (R35GM133646 to W.C.). C.M. is supported by the Cancer Research Institute Irvington Postdoctoral Fellowship (CRI4018). S.J.H is supported by an Investigator Grant from the National Health and Medical Research Council (NHMRC) of Australia. J.B. is grateful for support from the NIH (1F32HD078029–01A1). K.H. was supported by funds from Massachusetts General Hospital, the NIH (P01 GM099134), and the Gerald and Darlene Jordan Chair in Regenerative Medicine; N.T. was supported by a fellowship from the German Research Foundation (DFG). A.T and J.E received funding from 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 R35 CA197695, and the Henry Schueler 41&9 Foundation. This work has used computing resources at the NYU School of Medicine High Performance Computing (HPC) Facility. SN is supported by the Onassis Foundation - Scholarship ID: F ZP 036–1/2019–2020. O.A.-W. is supported by NIH R01 CA251138 (O.A.-W.), NIH R01 CA242020, NIH/NCI 1P50 254838–01 (O.A.-W.), the Leukemia & Lymphoma Society, and the Edward P. Evans MDS Foundation. I.A. is supported by the NCI/NIH (P01CA229086, RO1CA252239, R01CA228135, R01CA242020, RO1CA266212), Curing Kids Cancer, The Leukemia and Lymphoma Society and the Vogelstein Foundation. We would like to thank the NYU School of Medicine core facilities including the Applied Bioinformatics Laboratories (ABL), Flow Cytometry and the Genome Technology Center (this shared resource is partially supported by the Cancer Center Support Grant P30CA016087 at the Laura and Isaac Perlmutter Cancer Center). We acknowledge BioRender for schematic illustrations shown in Extended Data Figure 1a and Extended Data Figure 6g.

References

1. 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). [PubMed: 26516065]
2. Orlando EJ et al. Genetic mechanisms of target antigen loss in CAR19 therapy of acute lymphoblastic leukemia. Nat Med 24, 1504–1506 (2018). [PubMed: 30275569]
3. Zhao Y et al. Tumor-intrinsic and -extrinsic determinants of response to blinatumomab in adults with B-ALL. Blood 137, 471–484 (2021). [PubMed: 32881995]
4. Sanson KR et al. Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities. Nat Commun 9, 5416 (2018). [PubMed: 30575746]
5. Bradbury LE, Kansas GS, Levy S, Evans RL & Tedder TF 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). [PubMed: 1383329]
6. 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). [PubMed: 14530327]
7. David V, Hochstenbach F, Rajagopalan S & Brenner MB 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). [PubMed: 8486646]
8. 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). [PubMed: 22375059]
9. 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). [PubMed: 1375324]
10. Cortés-López M et al. High-throughput mutagenesis identifies mutations and RNA-binding proteins controlling CD19 splicing and CART-19 therapy resistance. bioRxiv, 2021.2010.2008.463671 (2021).
11. Weintraub AS et al. YY1 Is a Structural Regulator of Enhancer-Promoter Loops. Cell 171, 1573–1588 e1528 (2017). [PubMed: 29224777]
12. Bailey SD et al. ZNF143 provides sequence specificity to secure chromatin interactions at gene promoters. Nat Commun 2, 6186 (2015). [PubMed: 25645053]
13. 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). [PubMed: 9705341]
14. Consortium EP An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57–74 (2012). [PubMed: 22955616]
15. 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). [PubMed: 33397967]
16. Witkowski MT et al. Extensive Remodeling of the Immune Microenvironment in B Cell Acute Lymphoblastic Leukemia. Cancer Cell (2020).
17. Wessels HH et al. Massively parallel Cas13 screens reveal principles for guide RNA design. Nat Biotechnol 38, 722–727 (2020). [PubMed: 32518401]
18. Yang Q, Gilmartin GM & Doublie 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 U S A 107, 10062–10067 (2010). [PubMed: 20479262]
19. Bentley DL Coupling mRNA processing with transcription in time and space. Nat Rev Genet 15, 163–175 (2014). [PubMed: 24514444]
20. Brumbaugh J et al. Nudt21 Controls Cell Fate by Connecting Alternative Polyadenylation to Chromatin Signaling. Cell 172, 629–631 (2018). [PubMed: 29373832]
21. Sun M et al. NUDT21 regulates 3'-UTR length and microRNA-mediated gene silencing in hepatocellular carcinoma. Cancer Lett 410, 158–168 (2017). [PubMed: 28964783]

22. 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 (2020).
23. Majzner RG et al. Tuning the Antigen Density Requirement for CAR T-cell Activity. Cancer Discov 10, 702–723 (2020). [PubMed: 32193224]
24. Ramakrishna S et al. Modulation of Target Antigen Density Improves CAR T-cell Functionality and Persistence. Clin Cancer Res 25, 5329–5341 (2019). [PubMed: 31110075]
25. 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). [PubMed: 31738832]
26. Ma C et al. Leukemia-on-a-chip: Dissecting the chemoresistance mechanisms in B cell acute lymphoblastic leukemia bone marrow niche. Sci Adv 6 (2020).
27. Eyquem J et al. Targeting a CAR to the TRAC locus with CRISPR/Cas9 enhances tumour rejection. Nature 543, 113–117 (2017). [PubMed: 28225754]
28. 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). [PubMed: 33558546]
29. Wang E et al. Surface antigen-guided CRISPR screens identify regulators of myeloid leukemia differentiation. Cell Stem Cell (2021).
30. Ramkumar P et al. CRISPR-based screens uncover determinants of immunotherapy response in multiple myeloma. Blood Adv 4, 2899–2911 (2020). [PubMed: 32589729]

References

31. Li S, Ilaria RL, Million RP, Daley GQ & Van Etten RA 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; 1999. pp. 1399–1412. [PubMed: 10224280]
32. Shi J et al. Discovery of cancer drug targets by CRISPR-Cas9 screening of protein domains. Nat Biotechnol 33, 661–667 (2015). [PubMed: 25961408]
33. Ran FA et al. Genome engineering using the CRISPR-Cas9 system. Nat Protoc 8, 2281–2308 (2013). [PubMed: 24157548]
34. Doench JG et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nat Biotechnol 34, 184–191 (2016). [PubMed: 26780180]
35. Van Nostrand EL 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). [PubMed: 28766298]
36. Zawada AM et al. Massive analysis of cDNA Ends (MACE) and miRNA expression profiling identifies proatherogenic pathways in chronic kidney disease. Epigenetics 9, 161–172 (2014). [PubMed: 24184689]
37. Witkowski MT et al. Extensive Remodeling of the Immune Microenvironment in B Cell Acute Lymphoblastic Leukemia. Cancer Cell 37, 867–882 e812 (2020). [PubMed: 32470390]
38. Kuleshov MV et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res 44, W90–97 (2016). [PubMed: 27141961]
39. Dobin A et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21 (2013). [PubMed: 23104886]
40. Love MI, Huber W & Anders S Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15, 550 (2014). [PubMed: 25516281]
41. 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 Genomics 18, 22 (2017). [PubMed: 28056762]
42. Li HW Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv: Genomics (2013).
43. Kloetgen A et al. Three-dimensional chromatin landscapes in T cell acute lymphoblastic leukemia. Nat Genet 52, 388–400 (2020). [PubMed: 32203470]

44. Van Nostrand EL et al. A large-scale binding and functional map of human RNA-binding proteins. Nature 583, 711–719 (2020). [PubMed: 32728246]
45. 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). [PubMed: 28100584]
46. 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). [PubMed: 29284540]
47. Ha KCH, Blencowe BJ & Morris Q QAPA: a new method for the systematic analysis of alternative polyadenylation from RNA-seq data. Genome Biol 19, 45 (2018). [PubMed: 29592814]
48. Patro R, Duggal G, Love MI, Irizarry RA & Kingsford C Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods 14, 417–419 (2017). [PubMed: 28263959]
49. Harrow J et al. GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res 22, 1760–1774 (2012). [PubMed: 22955987]
50. Gruber AJ 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). [PubMed: 27382025]
51. Shen S et al. rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. Proc Natl Acad Sci U S A 111, E5593–5601 (2014). [PubMed: 25480548]
52. Korsunsky I et al. Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Methods 16, 1289–1296 (2019). [PubMed: 31740819]
53. Dann E, Henderson NC, Teichmann SA, Morgan MD & Marioni JC Differential abundance testing on single-cell data using k-nearest neighbor graphs. Nat Biotechnol (2021).
54. Hao Y et al. Integrated analysis of multimodal single-cell data. Cell 184, 3573–3587 e3529 (2021). [PubMed: 34062119]
55. Stuart T et al. Comprehensive Integration of Single-Cell Data. Cell 177, 1888–1902 e1821 (2019). [PubMed: 31178118]

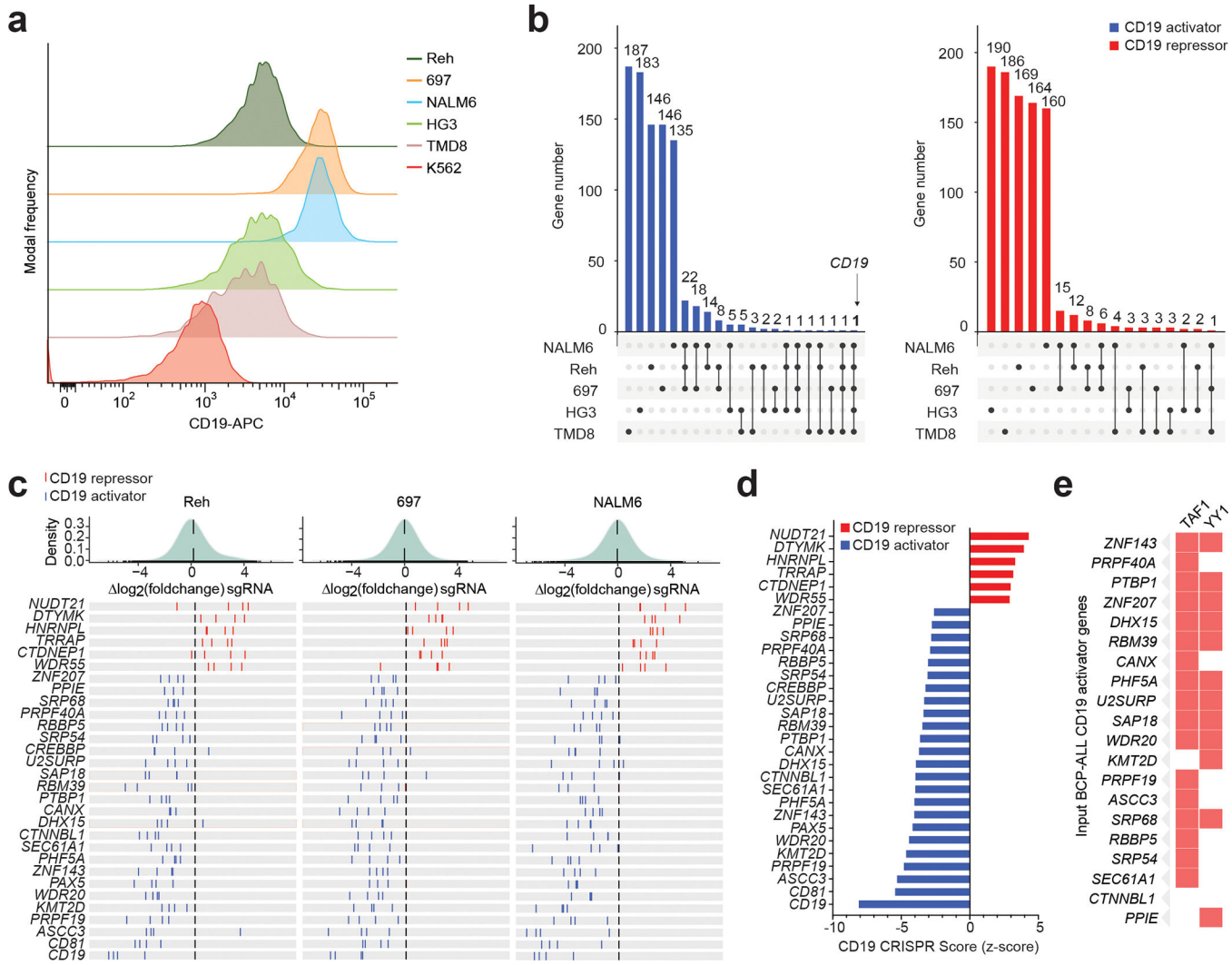


Figure 1. Genome-wide CRISPR screen identifies CD19 regulators in human B cell malignancies.

(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 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 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.

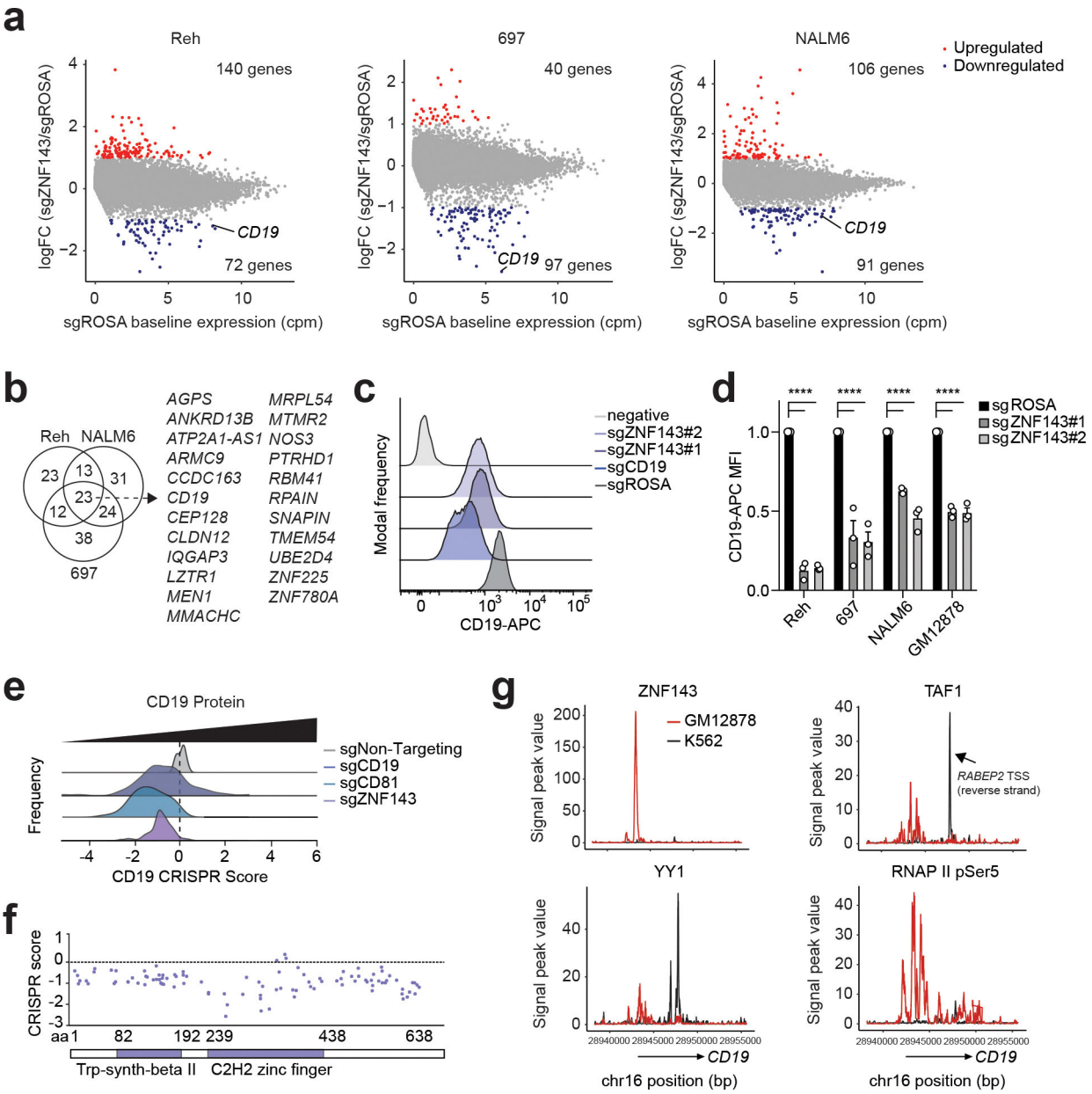


Figure 2. ZNF143 directly binds the CD19 locus and activates gene expression.

(a) M (log ratio of sgZNF143/sgROSA) and A (mean average of sgROSA,) plot of Cas9$^+$ Reh, Cas9$^+$697 and Cas9$^+$NALM6 cells seven days post-transduction with sgRNAs. CD19 upregulated (red) and downregulated (blue) genes highlighted with logFC exceeding +/−1 cut-off. Cpm, counts per million. (b) Venn diagram of overlapping downregulated genes across sgZNF143-mCherry$^+$ 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#1 and sgZNF143#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. MFI, mean fluorescence intensity. (independent experiments with n = 3, unpaired two-sided t-test, mean

and standard error shown, error bars represent s.e.m). 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. (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.

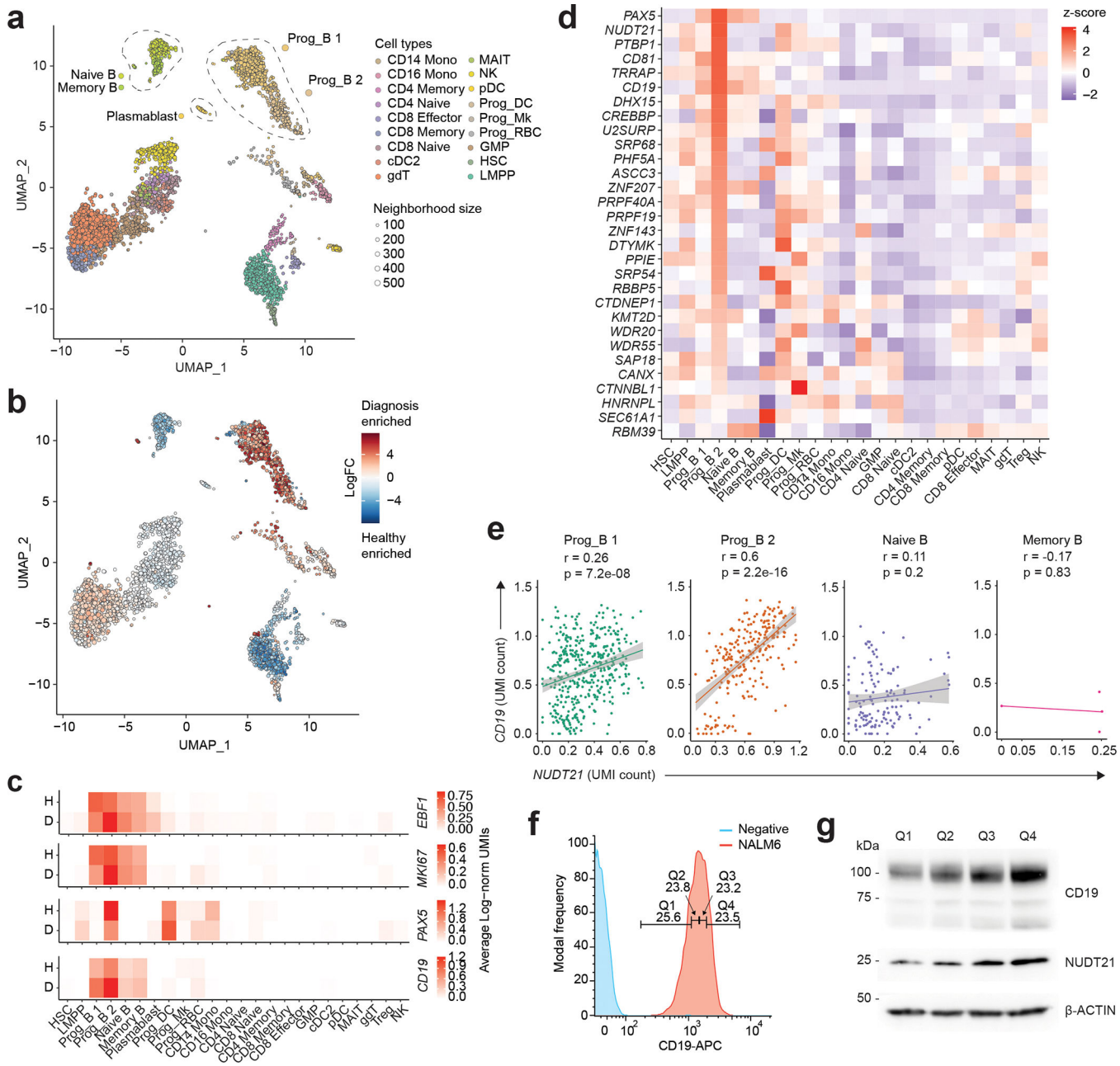


Figure 3. NUDT21 is highly-expressed in human B cell progenitors.

(a) UMAP representation of primary healthy (n = 4) and BCP-ALL diagnosed (n = 7) bone marrow single cell data. (list all indicated on the figure). **(b)** Milo neighborhood analysis showing log fold-change representation of BCP-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 BCP-ALL diagnosed (n = 7) bone marrow cell types. **(d)** Heatmap representation of cell type (row) z-score normalization of UMI counts for each individual CD19 candidate using scRNA-seq data from bone marrow with confirmed BCP-ALL diagnosis[17]. Genes ranked in descending order of individual candidate z-score across Prog_B 2 column, **(e)** Scatterplot showing correlation between

NUDT21 and *CD19* mRNA expression (UMI count) across Prog_B 1, Prog_B 2, naïve B cells and memory B cells in BCP-ALL diagnosed bone marrow scRNA-seq data[17]. r- and p-values calculated on the basis of Pearson's correlation (95% confidence interval shown. *p*-value derived based on t-distribution, degree of freedom is 416, 237, 132, 2 for Prog_B 1, Prog_B 2, Naïve B, 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 shown with size ladder indicated.

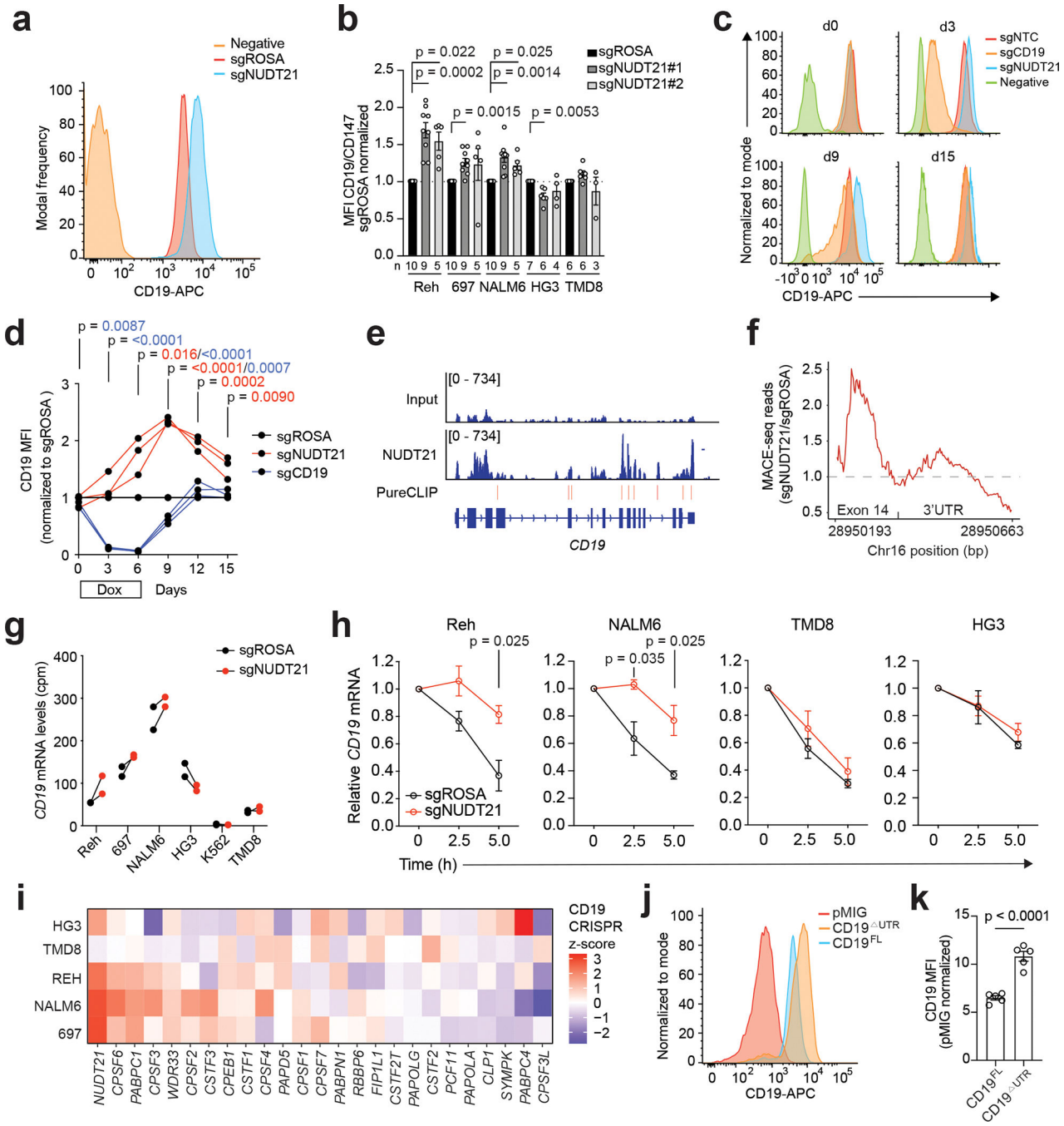


Figure 4. NUDT21 directly represses CD19 mRNA stability and protein expression.

(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 normalized to sgROSA-mCherry^{+}. n-value indicating independent replicates shown for each condition (unpaired two-sided t-test, mean and standard error shown, error bars represent s.e.m). (c,d) Representative flow cytometry (c) and summary of CD19 expression

(d) following expression of dox-induced sgNUDT21-GFP (red), sgCD19-GFP (blue) or sgNTC-GFP (black) in Cas13d^{+} 697 cells treated with dox for six days (day 0 – day 6) followed by dox withdrawal (independent experiments with n = 3, unpaired two-sided t-test, mean and standard error shown, error bars represent s.e.m). (e) eCLIP read tracks spanning the CD19 locus in NALM6 cells. PureCLIP significant peak signals shown (red). (f) MACE-seq log fold-change read counts of exon 14 and 3'-UTR junction (intron removed) in sgNUDT21#1-mCherry^{+} Cas9^{+} NALM6 cells and to sgROSA-mCherry^{+} Cas9^{+} NALM6 cells. (g) Bulk RNA-seq of CD19 mRNA in in sgROSA-mCherry^{+} or sgNUDT21-mCherry^{+} Cas9^{+} Reh, Cas9^{+} 697, Cas9^{+} NALM6, Cas9^{+} TMD8, Cas9^{+} HG3 and Cas9^{+} K562. Experiment performed in technical duplicate. Cpm, counts per million (h) Quantitative PCR analysis of *CD19* mRNA in sgROSA-mCherry^{+} or sgNUDT21-mCherry^{+} Cas9^{+} Reh, Cas9^{+} NALM6, Cas9^{+} TMD8 and Cas9^{+} HG3 at multiple timepoints (2.5 hours and 5 hours) post-actinomycin D treatment (independent experiments with n = 3, unpaired two-sided t-test, *p*-value, mean and standard error shown. *p*-value greater than 0.05 are 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 specifically highlighting polyadenylation regulators. (j,k) Representative flow cytometry (j) and histogram summary (k) of CD19-APC expression following ectopic expression of CD19 $^{\Delta UTR}$-GFP and CD19^{FL}-GFP in Cas9^{+} Reh cells. Histograms normalized to pMIG control (independent experiments with n = 5, unpaired two-sided t-test, mean and standard error shown, error bars represent s.e.m).

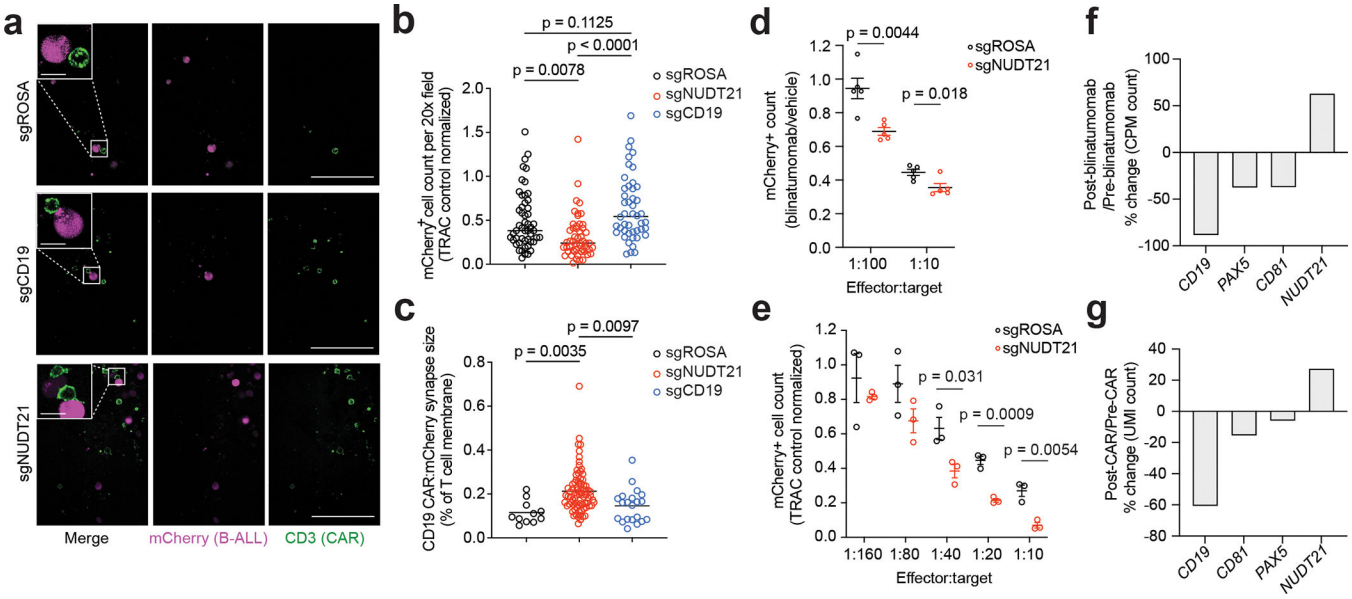


Figure 5. NUDT21 alters CD19-directed CAR-T cells and BiTE treatment responsiveness.

(a) Representative fluorescence 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 human umbilical vein endothelial cells. Scale bars represent 100 μm with 10 μm scale bar 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 4 devices, where each device has 2–4 random fields be quantified (n>=46 images), unpaired t-test, mean and standard error 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 4 devices, where each device has about 10 random fields be quantified (n>=11 synapse events), unpaired t-test, mean and standard error shown). (d) sgRNA-mCherry$^+$ Cas9+ NALM6 cells counts after 24-hour of co-culture with primary peripheral blood CD8$^+$ T cells in the presence or absence of 0.5 ng/mL blinatumomab normalized to vehicle treated control (independent experiments with n = 5, unpaired two-sided t-test, mean and standard error shown). (e) Number of sgRNA-mCherry$^+$ 697 cells after 24-hour co-culture with TRAC T cells or TRAC CD19-CAR T cells (in what conditions). Cell counts normalized to TRAC T cell counts (independent experiments with n = 3, unpaired two-sided t-test, mean and standard error 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.* Blood) (f) and scRNA-seq UMI counts pre- and post-CAR-T CD34$^+$ treatment in B-ALL patient data (Rabilloud *et al.*) (g). counts per million (cpm).