HHS Public Access

Author manuscript

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

Published in final edited form as:

J Immunol. 2022 April 01; 208(7): 1700–1710. doi:10.4049/jimmunol.2100491.

CRISPR/Cas9-mediated insertion of HIV LTR within *BACH2* promotes expansion of T regulatory-like cells

Michelle L. Christian[*], **Michael J. Dapp**[†], **Samuel C. Scharffenberger**[*], **Hank Jones**[*], **Chaozhong Song**[†], **Lisa M. Frenkel**[*,‡,§,¶], **Anthony Krumm**[†], **James I. Mullins**[†,¶,‖], **David J. Rawlings**[*,‡,#]

[*]Seattle Children's Research Institute, Seattle, WA, USA.

[†]Department of Microbiology, University of Washington, School of Medicine, Seattle, WA, USA

[‡]Department of Pediatrics, University of Washington, School of Medicine, Seattle, WA, USA

[§]Department of Laboratory Medicine, University of Washington, School of Medicine, Seattle, WA, USA

[¶]Department of Global Health, University of Washington, School of Medicine, Seattle, WA, USA

[‖]Department of Medicine, University of Washington, School of Medicine, Seattle, WA, USA

[#]Department of Immunology, University of Washington, School of Medicine, Seattle, WA, USA

Abstract

One key barrier to curative therapies for HIV is the limited understanding of HIV persistence. HIV provirus integration sites (IS) within *BACH2* are common, and almost all sites mapped to date are located upstream of the start codon in the same transcriptional orientation as the gene. These unique features suggest the possibility of insertional mutagenesis at this location. Using CRISPR/Cas9-based homology-directed repair in primary human CD4+ T cells, we directly modeled the effects of HIV integration within *BACH2*. Integration of the HIV LTR and major splice donor (MSD) increased BACH2 mRNA and protein levels, altered gene expression and promoted selective outgrowth of an activated, proliferative and T regulatory (T_{reg})-like cell population. In contrast, introduction of the HIV-LTR alone, or an HIV-LTR-MSD construct into *STAT5B,* a second common HIV IS had no functional impact. Thus, HIV LTR-driven *BACH2* expression modulates T cell programming and leads to cellular outgrowth and unique phenotypic changes; findings that support a direct role for IS-dependent HIV-1 persistence.

Materials and Correspondence.: Correspondence and requests for materials should be addressed to J.I.M. (jmullins@uw.edu) or to D.J.R. (drawling@uw.edu). Co-corresponding authors: David J. Rawlings & James I. Mullins, David J. Rawlings, Seattle Children's Research Institute, 1900 Ninth Avenue, M/S JMB-6, Seattle, WA 98101, 206-987-7319 (Direct), 206-884-3903 (Assistant), drawling@uw.edu.

AUTHOR CONTRIBUTIONS.

J.I.M. conceived the concept with input from L.M.F.; M.L.C and M.J.D. devised and led the project; J.I.M, A.K., and D.J.R. consulted on project direction and experimental design; M.L.C, M.J.D., and S.C.S. designed and performed experiments and completed the analyses; C.S., H.J., and S.C.S. performed experiments and all authors contributed to interpretation of results; M.J.D. and M.L.C wrote the manuscript with support from S.C.S., J.I.M., A.K., and D.J.R.; D.J.R. revised the manuscript in response to review.

These authors contributed equally to this work: Michelle L. Christian & Michael J. Dapp

DECLARATION OF INTERESTS

The authors declare no competing interests.

Keywords

HIV; gene-editing; CRISPR/Cas9; LTR; Treg; persistence; insertional-mutagenesis; integration-site

INTRODUCTION

The replication cycle of retroviruses, including human immunodeficiency virus (HIV), pose a unique problem for therapeutic intervention due to integration into host chromosomal DNA. While antiretroviral therapy (ART) can suppress viral replication and person-to-person spread, infected cells persist for the life of the infected individual. Thus, understanding the cellular and viral determinants that govern HIV persistence is critical for design of interventions for HIV cure.

HIV integrates in a semi-random fashion, and is enriched in transcriptional units[1,2] and near super-enhancer regions[3]. Strikingly, HIV integration site (IS) distribution studies have revealed unusually frequent integrations within the transcriptional unit of a few genes, most notably *BACH2* and *STAT5B*[4–6]. Importantly, the vast majority of *BACH2* IS have been reported in the forward direction relative to gene polarity and nearly all cluster in the intron immediately preceding the first coding exon[4–8]. In contrast to BACH2, while IS located in the *STAT5B* locus are also predominantly located upstream of the coding sequences, there is no clear directional forward orientation bias[4–8].

LTRs contain promoter, enhancer, and transcription terminating elements[9]. HIV integration into specific chromosomal loci has been associated with cell proliferation in vivo[4,5]. Based on the uniquely high frequency of detection of HIV IS in BACH2 and STAT5B in individuals with undetectable viral loads on ART, Cesana et al.[10] explored the possibility that HIV LTRs direct expression of the coding sequences of these genes. When integrated upstream of the first coding exon, expression of a cellular protein can potentially be initiated under control of the HIV promoter (LTR) by splicing from HIV splice donor sequences into the splice acceptors of the downstream genes. The major splice donor (MSD) of HIV, immediately downstream of the 5' LTR, is predicted to be the most efficient means of forming chimeric transcripts with the capacity to encode these cellular proteins. To date, proviral genome structures have been reported for only 4 integration events within BACH2. In all cases, however, the integrated proviruses retained an intact 5'LTR and MSD as would be required to generate chimeric transcripts[8]. Consistent with the concept, Cesana et al. detected LTR-driven chimeric transcripts at *BACH2* in 9% and *STAT5B* in 31% of a cohort of HIV-infected individuals, including individuals on ART[10]. Further, using a comprehensive, longitudinal RNA analysis of 44 subjects, treated with suppressive ART early after HIV infection, we identified T cells bearing *BACH2* hybrid transcripts in >40% of subjects, findings consistent with clonal expansion of T cell bearing this IS (LMF, JIM unpublished data). Together, these observations support the concept that retroviral insertion in *BACH2*, *STAT5B* and potentially other key sites, may directly alter T cells in a manner that promotes progressive clonal expansion via cell proliferation and/or survival[11]. However,

this concept and the functional impact of candidate novel transcriptional products in human primary T cells has not been directly tested.

In this study, we used gene editing to directly recapitulate the formation of hybrid transcripts reported in patient cells and, thereby, allow for interrogation of impacts at common HIV IS hotspots including *BACH2* and *STAT5B*[10]. To achieve this goal, we utilized a dual-delivery system of CRISPR/Cas9 and recombinant AAV donor templates to mediate homology directed repair (HDR)-based knock-in of alternative LTR cassettes within both *BACH2* and *STAT5B* in primary CD4+ T cells. While seeking to uncover the mechanisms behind the long-term persistence of latently infected HIV cells, our novel HIV-mediated gene expression system revealed subtle nuances of T cell biology driven by the transcription factor *BACH2*. Our combined observations demonstrate that HDR editing of *BACH2* using the LTR-MSD leads to increased BACH2 expression and generation of a unique T cell population that exhibited features of T regulatory (Treg) cells and manifested enhanced proliferation and selective outgrowth in comparision with control edited populations.

METHODS

Cloning of AAV Repair Templates.

Gibson Assembly or T4 DNA Ligase (New England Biolabs) were used to clone gene synthesized or PCR amplified fragments (IDT; Integrated DNA Technologies or Twist Bioscience). For each genomic target, plasmids were created to have 600bp homology arms homologous to the respective cut sites. All inserted nucleotide sequences and ligation products were confirmed by Sanger sequencing. Stellar competent cells (TaKaRa) were transformed with the finalized plasmids to generate stocks for AAV production.

AAV Production.

All donor templates designed for editing experiments were cloned into AAV plasmid backbones as previously described[12,13]. Recombinant AAV6 stocks were produced in HEK293T cells as previously described[13]. Viral titers were determined by qPCR using AAV-specific primers and probe as described previously[12].

Human Peripheral blood samples.

Cryopreserved PBMC isolated from volunteers were purchased from the Fred Hutchinson Cancer Research Center. Upon thawing, CD4+ cells were isolated by negative selection (EasySep Human CD4+ Enrichment Kit; STEMCELL Technologies) and cultured in RPMI 1640 media with 20% FBS, 1x GlutaMAX (Life Technologies), 10nM HEPES (Life Technologies), and 1x beta mercaptoethanol (Gibco). Cells were cultured in 50ng/ml human IL-2 (Peprotech) between 0.5–1 × 10e6/ml in flat-bottom culture plates at 37°C with 5% CO_2.

CRISPR/Cas9 target design and reagents.

Single guide RNAs (sgRNAs) were designed to target genomic sites at the *BACH2*, *STAT5B*, and *AAVS1* loci using the CCTop online tool[14], and synthesized by Synthego.

Guides were reconstituted in buffer TE (Synthego) with a final concentration of 50 pmol/ul, aliquoted and stored at −80C. Alt-R HiFi Cas9 was purchased from IDT (Newark, NJ).

Editing of Primary Human T Cells.

Following negative CD4+ selection, cells were activated for 3 days with Dynabeads Human T-Activator CD3/CD28 for T Cell Expansion and Activation (Gibco) according to the manufacturer's instructions. After activation, beads were magnetically removed and cells were cultured overnight. Cells were then washed twice in PBS and resuspended in LONZA buffer P3 prior to nucleofection. Cas9 nuclease and sgRNAs were complexed at a 1:2.5 ratio and delivered to cells by LONZA 4-D Nucleofection System according to the manufacturer's instructions, at a concentration of 2uM per 1.0×10^6 cells. Cells were then transferred into prewarmed media formulated as described above, supplemented with 2.5% FBS. AAV6 vectors carrying the corresponding repair templates were used to transduce edited cells at 30% culture volume. Twenty-four hours later, standard media supplemented with 20% FBS was added to all culture wells and cells. Cells were maintained at $0.5–1.0 \times 10^6$ cells/ml for the duration of the experiments. Media including fresh IL-2 was replaced every 2–3 days. The following sgRNA sequences were used in editing experiments: *BACH2* G1 AACTGCTTGAGCCCAAAAGG, *BACH2* G3 CCAGCAGTAAGTCTGTTGTA, *STAT5B* G3 GAGGCTACCACCTCACCTAG, and *AAVS1* P1 ATTCCCAGGGCCGGTTAATG.

Genomic and Transcriptomic Analysis.

Genomic DNA was extracted from all experimental samples (DNeasy Blood and Tissue, Qiagen) and used in ddPCR to assess HDR or in conventional PCR to quantify NHEJ). To assess NHEJ, gDNA was amplified by PCR using PrimeSTAR GXL Polymerase (Clonetech) and custom primers sitting 200–300bp flanking the predicted cut site. Amplicons were then Sanger sequenced (Genewiz) and the resulting .ab1 sequence files served as input for ICE Analysis (Inference of CRISPR Edits) developed by Synthego. Primer sequences available upon request.

Each sgRNA generated in this study was submitted to CCTop (https://crispr.cos.uni-heidelberg.de/) to examine putative off-target CRISPR/Cas9 localization to the human genome (hg38) with the additional parameters: 20bp target site length, 12bp core length, and 4 total potential mismatches (maximum of 2 core mismatches). The top 8 predicted cut sites were investigated by sequencing as described above for evidence of off target cutting (Fig. S1), and the primer sequences are available upon request.

RNA was extracted from 3 to 5×10^5 cells from all experimental samples using the RNeasy Micro Kit (Qiagen). 100 – 200 ng of RNA was used to synthesize cDNA using the iScript Advanced cDNA Synthesis Kit for RT-qPCR (Bio-Rad Laboratories) according to the manufacturer's instructions. Synthesized cDNA was used as the template for hybrid/endogenous transcript ddPCR, as well as sequencing reactions.

Flow Cytometry.

Flow cytometry was performed on an LSR II (BD Biosciences) flow cytometer, and data analyzed using FlowJo software (TreeStar). Cells were labeled with fluorescent antibodies and the viability stain Alexa Fluor 350 carboxyl acid NHS ester (Invitrogen). Surface proteins were stained by incubation with Abs for 30mins at 4°C. In cases where intracellular protein expression was assessed, cells were fixed and permeabilized following surface stain. For FOXP3, CTLA-4, and HELIOS, fixation/permeabilization was carried out with the True-Nuclear Transcription Factor Buffer Set (BioLegend) according to the manufacturer's instructions. For intracellular cytokine production, cells were cultured in media containing 50ng/ml PMA and 1ug/ml Ionomycin (Millipore Sigma), and 1ug/ml Golgistop (BD Bioscience) for 5h at 37°C. This was then followed by fixation/permeabilization with Cytofix/Cytoperm (BD Bioscience) and then incubated with cytokine-specific antibodies. To assess responsiveness to IL-2, stimulation and staining for pStat5 was performed as described[15]. 5ng/ml IL-2 (PeproTech) was used to stimulate the cells during the pStat5 stain. To assess cellular proliferation by flow cytometry, the CellTrace Far Red Cell Proliferation Kit (Invitrogen) was used according to the manufacturer's instructions. Briefly, cells were incubated in PBS with the appropriate concentration of CellTrace stain for 20 min at 37°C. 5 times the staining volume of media was then added to all stained wells to remove free dye for 5 min, and then cells were pelleted by centrifugation and resuspended in fresh media. Cells were cultured for 96h before proliferation was assessed by flow cytometry.

Antibodies used in this study include, from BioLegend, IFNg (cat. 502516), TNFa (cat. 502944), CTLA-4 (cat. 349908), FOXP3 (cat. 320208), CD4 (cat. 300530), LAG3 (cat. 369314), ICOS (cat. 313528), PD-1 (cat. 329922 and 329920), CD69 (cat. 310910), CD44 (cat. 103028); from BD Biosciences, pSTAT5 (cat. 612599), CD3 (cat. 349201), CD25 (cat. 335789), CD127 (cat. 563086), CD62L (cat. 562720), CD8 (cat. 560662); eBioscience, HELIOS (cat. 501124891/61988342), TIGIT (cat. 5011268 / 12950042); and from Life Technologies, IL-2 (cat. 25–7029-42). All antibodies were used at a 1:100 dilution except for pSTAT5 (1:10) and TIGIT (1:50).

Digital droplet (dd)PCR HDR Assays.

Droplet Digital PCR was used to assess targeting efficiency in edited samples (Bio-Rad Laboratories). All primers and probes were designed in Primer-Blast (https://www.ncbi.nlm.nih.gov/tools/primer-blast/) using the *Homo sapiens* genome as the reference database, and were purchased from IDT. Briefly, a forward primer and HEX probe were placed on the LTR region of the inserted cassette, and a reverse primer was located downstream of the 3' homology arm on the endogenous locus. The same forward primer and probe were used to assess editing at all genomic loci, while the reverse primer was specific to the gene targeted. The same primer and probe sets were used to assess editing events targeted by both the Solo-LTR and LTR-MSD constructs. A single control set of primers and a probe were designed to amplify a region of the *STAT5B* gene to assess the total number of alleles interrogated in each ddPCR reaction. All ddPCR reactions were duplexed with the insert set of primers and probes as well as the control set of primers and probes to assess the allelic frequency of HDR (homology-directed repair). Targeting efficiency was expressed as the ratio of the concentration of targeted alleles (positive for LTR insert) to that of the

total alleles (positive for *STAT5B* control) present. Primer and probe sequences are available upon request. ddPCR was performed according to the manufacturers' instructions.

Gene expression analysis.

The expression of hybrid HIV/*BACH2* and HIV/*STAT5B* transcripts from edited samples were quantified using custom ddPCR primers and probes designed by Cesana et al.[10], and the identities were confirmed by Sanger sequencing. Expression of *BACH2, STAT5B, IL10, TGFB1* and *HPRT* was quantified using TaqMan Gene Expression Assays (Applied Biosystems): *BACH2* (Hs00222364_m1, exon boundary 8–9), *STAT5B* (Hs00273500_m1, exon boundary 16–17), *IL10* (Hs00961622_m1, exon boundary 4–5), *TGFB1* (Hs00998133_m1, exon boundary 6–7), and *HPRT* (Hs99999909_m1, exon boundary 6–7). The concentration of hybrid or endogenous transcripts were normalized to concentration of *HPRT* transcripts duplexed in the same reaction.

Western blot analysis.

Chemiluminescent Western blot images and Ponceau S stain images were captured on the ChemiDoc MP imaging system. Quantification was performed using the ImageJ program. The Ponceau S stain band was used for quantification and normalization. BACH2 MW: 130 kDa. ACTB (b-actin) MW: 42 kDa.

Bulk RNA-sequencing and differential gene expression analysis.

Approximately 1 × 10e6 CD4+ T cells from a late culture timepoint (day 48 post-editing) were collected for RNA extraction (Qiagen) and poly-A selected mRNA library preparation (TruSeq RNA Library Prep Kit v2; Illumina). Cells were collected from 2 editing experiments. 33–41 million raw paired-end reads per sample were generated. Reads were mapped to a custom hg38 human reference genome which divided the *BACH2* locus into 2 regions: *BACH2_5p* reads mapping upstream of the target site (hg38 gene position: 115,649) and *BACH2_3p* reads mapping downstream of the target site. Additionally, the LTR-MSD sequence was appended to hg38. STAR 2.7 was used to filter and map reads[16]. HTSeq was used to count the overlap of reads with genes[17]. Count normalization and differential gene expression analysis was performed using DESeq2[18]. The data produced by this experiment was deposited in the NCBI BioSample database for public access with the accession number PRJNA733175 (https://www.ncbi.nlm.nih.gov/bioproject/733175).

PANTHER Gene-ontology biological processes overrepresentation test.

Gene lists determined by differential expression were passed through the gene ontology (GO)[19,20] biological process complete list and analyzed by PANTHER overrepresentation test (Released 20200407)[21] using the Fisher's exact test with FDR correction. An FDR $P<0.05$ was used as the cutoff. The following GO annotation version and release date was: GO Ontology database DOI: 10.5281/zenodo.3727280 Released 2020–03-23.

Statistical analysis.

Statistical analyses were performed in GraphPad Prism 7 (GraphPad) and R Software package. The analyses performed are indicated in figure legends. Combined experimental

values are shown as mean ± SD. (p values: ns = ≥ 0.05, * = 0.01 to 0.05, ** = 0.001 to 0.01, *** = 0.0001 to 0.001, **** = ≤ 0.00001).

RESULTS

Targeted insertion of HIV proviral elements into integration hotspots in primary human T cells.

To directly assess the impact of HIV integration into the BACH2 locus, we used HDR-based gene editing to insert different HIV component constructs into primary human CD4 T cells. Designer nucleases and rAAV6 donor templates were co-delivered using a highly efficient HDR-editing platform previously developed in our laboratory[12,13,22–24]. The CRISPR/Cas9 system was used to achieve efficient nuclease-based DNA targeting using synthetic single guide RNAs (sgRNAs) targeting a region within intron 5 of *BACH2* (*BACH2* G1 and G3) and within intron 1 of *STAT5B* (*STAT5B* G1; Fig. 1a and Fig. S1), respectively, previously found to have multiple distinct HIV integrations in patient cohorts[4].

Minor limitations in sgRNA design, due to oligo specificity and the 5'-NGG-3' Cas9 PAM site requirement, restricted targeting to 2-bp upstream (hg38 *BACH2* gene position: 115,649) of an *in vivo* recovered *BACH2* IS[6] (Fig. 1a). Similarly, targeting was directed to intron 1, the intron preceding the start of the coding sequence of *STAT5B* (hg38 *STAT5B* gene position: 29,125) 1-bp downstream of an *in vivo* recovered IS[6].

As shown in Fig. 1b, CD4+ T cells were activated, electroporated for delivery of CAS9 protein and gRNA as ribonucleoproteins (RNP) complexes and maintained in culture for up to 7 weeks for analyses. Initial Cas9 cutting efficiencies at the *BACH2* and *STAT5B* loci were assessed by non-homologous end-joining (NHEJ) repair readout, a measure of the NHEJ-mediated insertions and deletions (indels) that manifest after repair of double-strand DNA breaks. An average of 97% (+/− 2.4 SD) of indels were found at the intended *BACH2* target site and 96% (+/− 1.2) of indels at the *STAT5B* target site, indicating highly efficient on-target cutting at both loci (Fig. 2c). Off-target cleavage was also measured for *BACH2* G3 and *STAT5B* G1 based on *in silico* predictions and found to be minimal (Fig. S1). Additional *BACH2* sgRNAs were assessed for both on-target and off-target NHEJ-repair based on *in-silico* predicted guide design; however, appreciable off-target cleavage was observed for *BACH2* G1 and this guide was therefore excluded from further testing (Fig. S1).

Next, to recapitulate HIV-IS that faithfully mirrored the IS identified in relevant patient populations, four alternative AAV donor cassettes were generated. These donors contained either a single HIV-LTR (Solo-LTR) or a 5' HIV-LTR continuing a downstream major splice-donor (MSD) sequence (LTR-MSD), and were designed to target either *BACH2* or *STAT5B*. As shown in Fig. 1b, CD4+ T cells were initially activated and cultured for four days prior to electroporation for ribonucleoprotein (RNP) delivery and subsequently infected with rAAV6 donor viruses. HDR-mediated targeting of the AAV cassettes was measured by ddPCR, spanning the 600bp donor template homology arms. HDR rates at the *BACH2* locus were 33.5% (+/− 1.2%) for Solo-LTR and 23.4% (+/− 2.1%) for LTR-MSD. Slightly higher targeting rates of 49.5% (+/− 6.8%) and 37.7% (+/− 7.5%), respectively, were obtained

within the *STAT5B* (Fig. 1d). Regardless of initial editing efficiency or targeted loci, cellular viability was minimally impacted by RNP and rAAV co-delivery or the resulting genomic manipulations (Fig. S1–S2).

Insertion of HIV LTR-MSD within *BACH2* uniquely promotes outgrowth of HDR-edited T cells.

To determine the effects of HIV LTR insertions at specific cellular loci, HDR-edited and control primary CD4+ T cell populations were cultured for up to 7 weeks post-editing. Strikingly, the proportion of *BACH2* LTR-MSD-edited cells progressively increased over time, comprising ~70% of the culture at 7 weeks (Fig. 2a). When normalized to initial targeting frequency, *BACH2* LTR-MSD-edited cells expanded by nearly 3-fold (Fig. 2b). These growth characteristics were dependent on the presence of the HIV MSD, as Solo-LTR-edited cells at the *BACH2* locus did not expand (Fig. 2a–b). Absolute cell counts showed *BACH2* LTR-MSD-edited cells continued to expand over 40 days of culture, while Solo-LTR-edited cell numbers, similar to controls, plateaued after ~14 days (Fig. 2c). In contrast, *STAT5B*-edited cells did not exhibit a proliferative phenotype with either construct (Fig. 2d–e). Consistent with ddPCR-based HDR tracking, *STAT5B* edited-cells, targeted with either the LTR-MSD or the Solo-LTR AAV donor, did not differ in cell counts when compared to controls (Fig. 2f). As an additional control, we utilized HDR-based editing of *AAVS1,* a ‘safe-harbor’ chromosomal location not associated with HIV IS or hybrid transcripts[25]. Similar to control T cells and *STAT5B* edited-cells, *AAVS1* LTR-edited cells did not exhibit differences in T cell expansion during 4 weeks of culture (Fig. S2). To control for an idiosyncratic impact of the specific CRISPR/Cas9 target site within intron 5 of *BACH2*, an alternative CRISPR guide (*BACH2* G1) and AAV targeting cassette was used to introduce to the LTR-MSD cassette into an alternative location in intron 5. This led to a similar selective expansion of *BACH2* LTR-MSD-edited T cells further validating these findings (Fig. S1). Together, these results demonstrate that targeting of an HIV LTR-MSD within intron 1 of *BACH2* uniquely promotes the competitive outgrowth of HDR targeted primary CD4+ T cells.

Characterization of *BACH2* and *STAT5B* targeted hybrid mRNAs.

Next, we investigated whether LTR hybrid transcripts were produced by the targeted cell populations. Primers were positioned in the LTR trans-cassette and in either exon 6 of *BACH2* (Fig. 3a) or exon 2 of *STAT5B* (Fig. 3b). As expected, only the *BACH2* and *STAT5B*-edited cells targeted with the LTR-MSD cassette exhibited qualitative levels of LTR hybrid transcripts. The identities of these hybrid transcripts were confirmed by DNA sequencing (Data not shown).

To determine whether the selective outgrowth of *BACH2* LTR-MSD edited T cells correlated with production of hybrid transcripts, the relative amount of hybrid vs. endogenous transcript in *BACH2* or *STAT5B* HDR-edited T cell populations was quantified using RT-ddPCR (Fig. 3a–b) at early (Day 7) and late (Day 35) cell culture timepoints. In all assays, transcript levels were standardized to those of a housekeeping gene, hypoxanthine guanine phosphoribosyl transferase (*HPRT*). In *BACH2* LTR-MSD edited cells the level of hybrid transcripts, as a fraction of total BACH2 transcripts, increased 2.2-fold from early

to late culture timepoints ($p < 0.01$) (Fig. 3c). In contrast, no change in hybrid transcript was observed in the *STAT5B*-edited cells (Fig. 3c). The levels of endogenous *BACH2* and *STAT5B* transcripts were also evaluated over time, and interestingly, *BACH2* transcript levels increased modestly in *BACH2* LTR-MSD edited cells (1.3-fold +/− 0.2; $p < 0.05$) (Fig. 3d). This increase was not observed in *BACH2* Solo-LTR-edited samples, suggesting that generation of hybrid *BACH2* transcripts may facilitate this change. In contrast, there was no significant change in endogenous *STAT5B* transcripts in LTR-MSD edited cells (1.2-fold +/− 0.2; $p > 0.05$) compared to control cells (Fig. 3e). Consistent with these findings, BACH2 protein levels increased by up to 2.5-fold by day 21 in LTR-MSD edited-cells ($p < 0.01$), while there was no significant change in *STAT5B* LTR-MSD edited-cells (Fig. 3f–g). Together, these data demonstrate that expression of *BACH2* can be driven by transcription initiated at the HDR integrated LTR and is likely responsible, at least in part, for the increased proliferation and/or selective survival of HDR-edited CD4+ T cells.

HIV LTR-driven *BACH2* expression leads to a Treg-like population with increased proliferative capacity.

We hypothesized that the alterations in *BACH2* gene expression in edited cells would alter the phenotype of these populations. We performed RNA gene expression analysis in HDR edited populations at ~7 weeks post editing. Insertion of the LTR-MSD resulted in differential expression of 452 genes relative to cells edited with Solo-LTR (p-adj. < 0.05): 309 upregulated and 142 downregulated in LTR-MSD, respectively. Established gene sets were characterized including: i) Treg cell markers, ii) T-cell exhaustion cell markers, and iii) cell cycle associated genes (Fig 4a and data not shown). *IL7R* (CD127) was 7.5-fold lower (adj. p = 1.1E-07) and *IL2RB* (CD25) was 5.9-fold higher (adj. p = 7.6E-04) consistent with a Treg phenotype[26]. Markers for T cell exhaustion, *LAG-3 and TIGIT,* were also more abundantly expressed in the LTR-MSD population (11.3-fold; adj. p-value = 3.1E-04 and 3.7-fold; adj. p-value = 0.0051, respectively). Finally, at least 72 genes associated with cell cycle processes (GO:0022402) were differentially expressed in LTR-MSD-edited cells, including: *MKI67* (3.7-fold; adj. p-value = 1.8E-05), *TOP2A* (2.9-fold; adj. p-value = 0.012), *FOS* (35.6-fold; adj. p-value = 3.5E-04), *ASPM* (3.1-fold; adj. p-value = 4.5E-04) and *H2AFX* (2.4-fold; adj. p-value = 6.1E-04) (Fig. 4a and data not shown). An unbiased gene-ontology biological processes overrepresentation test for the 452 differentially expressed genes in LTR-MSD edited cells revealed a list of 246 gene-sets with a FDR < 0.02 (Data available upon request). The top 30 overrepresented biological processes gene-sets with fold-enrichment are shown, the majority of these indicating genes enriched in cell cycle processes (Fig. 4b). An additional comparison between LTR-MSD and unedited cells revealed 154 differentially expressed genes (p-adj. < 0.05): 82 upregulated and 72 downregulated in LTR-MSD, respectively (Figure 4b).

Next, differential expression at the *BACH2* locus (Fig 4c) was interrogated. Reads mapping to the inserted LTR cassette region (designated gene name: *LTR*) were 16.3-fold higher in LTR-MSD-edited cells (adj. p-value = 6.9E-04). This was associated with a 1.3-fold increase in transcript levels downstream of the insertion site (designated gene name: *BACH2_3p;* adj. p-value = 0.73), mirroring the RT-ddPCR results (Fig 4c). Interestingly, reads mapped to the region upstream of the BACH2 insertion site (designated gene name: *BACH2_5p*)

were 4.1-fold lower (adj. p-value = 3.7E-03) in LTR-MSD-edited cells, implying that LTR promoter-initiated transcription dominates over the host BACH2 promoter (Fig. 4c).

To further investigate this altered cell phenotype, multicolor flow cytometry was used to evaluate the expression of proteins associated with T-cell activation, exhaustion, and differentiation. *BACH2* LTR-MSD edited cells showed increased expression of multiple surface markers, including PD-1, CD69, LAG-3, TIGIT, and CD25 (Fig 5a and b). *BACH2* LTR-MSD edited-cells also exhibited a shift from a predominantly naïve (Tn: CD44-/CD62L+) or memory (Tmem: CD44+/CD62L-) phenotype to a predominant T effector (Teff: CD44-/CD62L-) phenotype (Fig 5a).

Notably, expression of proteins associated with a Treg phenotype were also impacted by LTR-driven *BACH2* expression: LTR-MSD *BACH2*-edited cells showed an increase in the proportion of $CD25^{high}/CD127^{low}$ cells compared to controls (1.4-fold +/− 0.21; $p < 0.001$) (Fig 5a and b). Strikingly, expression of intracellular proteins associated with Treg programming including the key transcription factors, FOXP3 and HELIOS, as well as the cytoplasmic pool of the inhibitory receptor, CTLA-4, were increased in LTR-MSD edited-cells (Fig 5a). FOXP3 and HELIOS were increased by 1.5-fold (+/− 0.3, $p < 0.05$) and 1.6-fold (+/− 0.28, $p < 0.01$) respectively, while CTLA-4 expression was increased by 2.1-fold (+/− 0.32, $p < 0.001$) compared to controls (Fig. 5b). To further evaluate this Treg-like phenotype, edited cells were evaluated for suppressive cytokine production, including IL-10 and TGFβ commonly expressed by multiple Treg subsets[27]. IL-10 transcript levels were significantly increased in the LTR-MSD *BACH2*-edited cells (1.8-fold +/− 0.29, $p < 0.001$), while TGFβ1 transcripts were decreased (1.3-fold lower +/− 0.16; $p < 0.001$) compared to controls (Fig. 5c).

The multiple surface and transcriptional phenotypic changes observed in LTR-MSD *BACH2*-edited cells prompted assessment of the capacity to express other key cytokines. Conventional Treg cells consistently exhibit a reduction in IL-2 expression and variable expression of inflammatory cytokines. IL-2 production in *BACH2*-edited cells was measured and found to be significantly reduced (1.2-fold lower +/− 0.11; $p < 0.05$) compared to controls (Fig. 5d). There were no discernable differences in the production of two other key inflammatory cytokines: INFγ or TNFα (Fig. 5d).

Interestingly, flow cytometry revealed a reduction in CD4 expression following HDR-editing, a finding specific to *BACH2* LTR-MSD edited cells. Over 3 weeks in culture, CD4+ cells decreased from 98% to 73% in the LTR-MSD edited-cells (Fig. 5e; average of 4 donors, $p < 0.01$). While CD4 loss varied amongst PBMC donors, a significant and reproducible loss of CD4 expression was observed in all but one donor, with the greatest reduction reaching 50% $CD4^{-}$ cells (Data not shown). To investigate whether this process reflected downregulation of CD4 production vs. surface internalization, a competitive flow cytometry based CD4 mAb assay was used. Intracellular CD4 expression was not detected in the CD4 surface negative population, suggesting that CD4 loss reflected reduced transcription and/or altered post-transcriptional events (Data not shown).

To begin to investigate the mechanism by which LTR-MSD *BACH2*-edited cells persisted and expanded in culture (Fig. 2c), we measured cellular proliferation using a flow cytometry-based proliferation dye dilution assay. At 5 weeks post-editing and growth in T cell expansion conditions, we observed that LTR-MSD cells proliferated 2.6-fold ($p < 0.01$) more than control cells (Fig. 5f); findings that correlate with enhanced growth rates of *BACH2* LTR-MSD-targeted cells in culture.

Next, similar to $CD25^{hi}$ conventional Tregs[28], it was posited that increased CD25 expression in HDR edited cells would facilitate an increase in signaling in response to IL-2. To initially test this idea, we performed flow analysis of phosphorylated STAT5 (pSTAT5) to measure IL-2 signaling. Cell populations were cultured in the absence of exogenous IL-2 for 24h, stimulated with IL-2 and assessed by flow cytometry[29]. No significant differences in pSTAT5 expression was observed in cells with integrated LTR-MSD over controls (Fig. S3a). However, we reasoned that IL-2 starvation might differentially impact *BACH2* LTR-MSD-targeted vs control cell populations. To this end, cells were exposed to different IL-2 doses (5ng/ml and 50ng/ml IL-2) post-editing and proliferation assessed by flow cytometry. Notably, in limiting IL-2 conditions, LTR-MSD edited cells proliferated 1.9-fold more ($p<0.001$) than the control cells (Fig. S3b–d). Additionally, as shown in Fig. S4, we performed LTR-MSD editing of *BACH2* using input CD4 T cells that were depleted of natural (n)Treg. In this setting, HDR-editing again led to the generation and expansion of a Treg-like population. Thus, while LTR-MSD edited Treg-like cells are not derived from nTreg, they exhibit phenotypic similarities to nTreg including increased competitive fitness in an IL-2 limited environment.

In summary, our combined observations support the conclusion that *BACH2* LTR-MSD editing results in increased BACH2 expression, thereby leading to generation of a unique T cell population exhibiting multiple Treg-like features: high levels of transcription factors including FOXP3 and Helios; increased expression of inhibitory receptors including CTLA-4 and LAG-3; upregulation of the high affinity IL-2 receptor ($CD25^{high}/CD127^{low}$); reduced expression of IL-2; and increased expression of IL-10. In parallel, this population exhibits evidence for increased cell activation (upregulation of CD69, PD-1, and acquisition of Teff phenotype), that correlate with enhanced proliferation and selective expansion in vitro.

DISCUSSION

Our findings provide the first direct demonstration that targeting of the HIV LTR to a defined genomic location can modulate the biological behavior of primary human T cells. These data show that two key retroviral elements, the HIV LTR and MSD, are sufficient to manipulate cellular function and phenotype through insertional mutagenesis mediated by LTR hybrid transcripts. Further, while performed in an *ex vivo* model system using truncated provirus elements, our results suggest a potential impact of these events in the context of HIV pathogenesis. Modeling a *cis*-acting LTR promoter targeting an IS frequently detected in HIV patients, we demonstrate that BACH2 RNA and protein levels increase over time, resulting in proliferation and selective outgrowth of T cell populations with an activated, Treg-like phenotype. In contrast to previous protein overexpression studies[10,30], our primary

T-cell HDR knock-in system closely recapitulates the endogenous transcriptional control at key HIV IS including both *BACH2* and *STAT5B*.

Importantly, these results help to explain the persistence and strong orientation bias of HIV proviruses at the *BACH2* locus in infected individuals. To date, the available published HIV insertion site data has documented 67 independent insertions across 7 individuals at the *BACH2* locus, all in the forward orientation with respect to gene polarity[4–7]. The integrated proviral sequences within BACH2 in HIV subjects remain to be definitively characterized. Interestingly, to date, proviral sequences from 4 BACH2 IS have been reported and each retained an intact 5'LTR and MSD[8] as required to generate hybrid transcipts. Because our results show that the MSD, immediately downstream of the 5'-LTR, is critical to this process, it suggests that proviruses within *BACH2* may be defective and lack transcription terminating elements present within the 3'-LTR. Alternatively, intact integrated proviruses in the same location could also generate hybrid transcripts secondary to equivalent splice capture events, leading to expansion of these cells and contribution to the latent HIV reservoir. In our studies, detailed mapping of RNA-seq reads at the *BACH2* locus showed that while the CDS region downstream of the intron 5 insertion site has a modest 1.3-fold increase in expression, host BACH2 transcription is reduced by over 4-fold. This suggests that *BACH2* expression is carefully regulated to physiologic levels in what appears to be a negative feedback system, i.e., LTR-driven expression of BACH2 negatively regulates BACH2-promoter driven expression. This observation is consistent with previous findings of LTR promoter dominance[31–33].

Together with the previous observations indicating that BACH2 is a ubiquitous target for HIV integration, our findings suggest that T cells with this IS proliferate *in vivo* and thereby reach detection levels due to a unique survival advantage. Consistent with this concept, *BACH2* HDR edited cells, under the control of an LTR promoter, exhibit a significantly enhanced proliferative capacity and expansion in vitro compared to unedited cells. Interestingly, this selective expansion was most evident at timepoints later than typical for primary cell culture models, which reach peak expansion rates in 7–14 days. These findings appear to model the subtleties anticipated *in vivo*, in which a modest, but physiologically relevant, increase in a transcriptional regulator manifests a physiologically relevant phenotype over time. Overall, these observations are consistent with a proposed model where HIV-driven BACH2 IS dependent proliferation leads to a gradual clonal expansion with selection observed over a scale of years in vivo[11].

BACH2 edited CD4+ T cells also progressively acquire a Treg-like phenotype in culture. This population exhibits many but not all features present in nTreg, as well as evidence for increased cell activation, enhanced proliferation and selective expansion *in vitro*. These combined phenotypic and growth properties likely reflect the complex role for *BACH2* in T cell development and functional specification. BACH2 acts as a transcriptional regulator required for thymic and peripheral Treg development and homeostasis, generation of T central memory cells, and in limiting effector T cell differentiation[26–28,31–36]. *BACH2* has also been implicated as a transcriptional repressor that limits activity of super-enhancers (SEs) proposed to drive establishment and stability of T cell lineages and subsets in human and murine T cells. While most studies of *BACH2* have relied upon loss of function

or overexpression, our studies provide insight into the impacts of a modest change in BACH2 dosage on the biology of primary human T cells. Additional work examining the transcriptional and epigenetic changes driven by these LTR hybrid transcripts is required to fully elucidate the presumed multiple impacts of these events in primary T cells.

Interestingly, LTR or LTR-MSD insertion within the *STAT5B* locus did not elicit any measurable molecular, phenotypic, or growth differences in primary CD4 T cells (Fig. S2). As previously reported in HIV infected persons, LTR initiated hybrid transcripts were detected in *STAT5B* edited cells[12]. However, in the present study, HIV-*STAT5B* hybrid transcripts did not lead to significant increases in STAT5B protein expression or measurable functional effects. Of note, HIV IS at this locus do not display a clear forward orientation bias[4–6]. While additional studies are warranted, our results are consistent with the lack of orientation bias and imply that one or more alternative (enhancer, chromatin, or yet unknown) *trans*-mediated mechanism(s) may responsible for the observed frequency and persistence of *STAT5B* IS T cell clones.

Our findings support the concept that modest increases in BACH2 expression may lead to the reprogramming of conventional CD4 T cells; thereby mimicking HIV provirus integration within a newly infected circulating CD4 T cell. Based on HDR editing of both bulk (Fig. 2) and nTreg depleted (Fig. S4) CD4 T cells, our data do not support the concept that reprogramming primarily impacts nTreg; findings that contradict earlier work suggesting over-expression of *BACH2* specifically within nTreg leads to their selective expansion[10]. That work based its conclusions upon identification of hybrid transcripts in purified $CD25^{hi}$, $CD127^{lo}$ T cells isolated from HIV patients; an approach that would similarly enrich for the novel Treg-like population identified here. Additional studies by those authors relied upon overexpression of BACH2 in isolated iTreg or nTreg - an approach that does not model the subtle impacts on BACH2 expression mediated by site-specific, HDR-based integration of a relevant LTR cassette.

The Treg-like population identified here may represent an immunosuppressive T cell subset. The frequency and retention of IS within *BACH2 in vivo* thus raises the possibility that such cells may alter immune responses and/or HIV disease outcome, whether or not these cells harbor replication competent proviruses. Collectively, previous work suggests that Tregs exert both negative and positive impacts on HIV disease outcome via limiting immune response to HIV and opportunistic infection and attenuating HIV-induced immune hyperactivation[37–39]. Moreover, HIV viral load is positively correlated with Treg percentages[40–44] and negatively correlated with absolute Treg numbers[41,45–47]. However, immune hyper-activation is a negative prognostic factor associated with HIV disease progression[48,49]. Additional work is required to understand the impact of HIV integrations at *BACH2* and the outcome on HIV disease progression and the viral reservoir.

In summary, we provide direct mechanistic support for how a specific HIV integration event may lead to expansion and persistence of clonally derived, CD4 T cell populations within infected individuals. Specifically, we show that transcriptional activation of *BACH2* via the generation of hybrid HIV-human transcripts is sufficient to generate a proliferative, activated, and Treg-like population *in vitro*. Based on these findings, we propose a model

wherein integration site drives phenotype and whereby proviruses convert a population of CD4+ T cells to a potentially immune-suppressive state, that, via parallel provision of selective advantage, may modulate disease outcome and/or HIV persistence.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS.

We acknowledge and thank I. Khan, E. Lopez, C. Stoffers, C. Zavala, and A. Ott of the Viral Production Team at SCRI for providing AAV stocks.

This work was supported by National Institutes of Health grants 1R01AI125026 (Mullins, PI), 1R61DA047010 (Mullins, MPI), R01CA206466 (Frenkel, PI), R01AI134419 (Frenkel, PI), 1R01AI122361(Mullins, PI), P30AI027757 Centers for AIDS Research Retrovirology and Molecular Data Sciences Core (C. Celu PI; Mullins, Core Director) and by the Seattle Children's Research Institute (SCRI) Program for Cell and Gene Therapy (PCGT), the Children's Guild Association Endowed Chair in Pediatric Immunology (to D.J.R.), and the Hansen Investigator in Pediatric Innovation Endowment (to D.J.R.).

References

1. Lewinski MK, Bisgrove D, Shinn P, Chen H, Hoffmann C, Hannenhalli S, Verdin E, Berry CC, Ecker JR, Bushman FD. 2005. Genome-wide analysis of chromosomal features repressing human immunodeficiency virus transcription. J. Virol. 79: 6610–6619. [PubMed: 15890899]
2. Schröder ARW, Shinn P, Chen H, Berry C, Ecker JR, Bushman F. 2002. HIV-1 Integration in the Human Genome Favors Active Genes and Local Hotspots. Cell. 110: 521–529. [PubMed: 12202041]
3. Lucic B, Chen H-C, Kuzman M, Zorita E, Wegner J, Minneker V, Wang W, Fronza R, Laufs S, Schmidt M, Stadhouders R, Roukos V, Vlahovicek K, Filion GJ, Lusic M. 2019. Spatially clustered loci with multiple enhancers are frequent targets of HIV-1 integration. Nat. Commun. 10: 4059. [PubMed: 31492853]
4. Wagner TA, McLaughlin S, Garg K, Cheung CYK, Larsen BB, Styrchak S, Huang HC, Edlefsen PT, Mullins JI, Frenkel LM. 2014. Proliferation of cells with HIV integrated into cancer genes contributes to persistent infection. Science. 345: 570–573. [PubMed: 25011556]
5. Maldarelli F, Wu X, Su L, Simonetti FR, Shao W, Hill S, Spindler J, Ferris AL, Mellors JW, Kearney MF, Coffin JM, Hughes SH. 2014. Specific HIV integration sites are linked to clonal expansion and persistence of infected cells. Science. 345: 179–183. [PubMed: 24968937]
6. Ikeda T, Shibata J, Yoshimura K, Koito A, Matsushita S. 2007. Recurrent HIV-1 integration at the BACH2 locus in resting CD4+ T cell populations during effective highly active antiretroviral therapy. J. Infect. Dis. 195: 716–725. [PubMed: 17262715]
7. Mack KD, Jin X, Yu S, Wei R, Kapp L, Green C, Herndier B, Abbey NW, Elbaggari A, Liu Y, McGrath MS. 2003. HIV insertions within and proximal to host cell genes are a common finding in tissues containing high levels of HIV DNA and macrophage-associated p24 antigen expression. J. Acquir. Immune Defic. Syndr 33: 308–320. [PubMed: 12843741]
8. Simonetti FR, Zhang H, Soroosh GP, Duan J, Rhodehouse K, Hill AL, Beg SA, McCormick K, Raymond HE, Nobles CL, Everett JK, Kwon KJ, White JA, Lai J, Margolick JB, Hoh R, Deeks SG, Bushman FD, Siliciano JD, Siliciano RF. 2021. Antigen-driven clonal selection shapes the persistence of HIV-1-infected CD4+ T cells in vivo. J. Clin. Invest. 131: e145254.
9. Jern P, Coffin JM. 2008. Effects of Retroviruses on Host Genome Function. Annu. Rev. Genet. 42: 709–732. [PubMed: 18694346]
10. Cesana D, Sio F. R. S. d., Rudilosso L, Gallina P, Calabria A, Beretta S, Merelli I, Bruzzesi E, Passerini L, Nozza S, Vicenzi E, Poli G, Gregori S, Tambussi G, Montini E. 2017. HIV-1-mediated insertional activation of STAT5B and BACH2 trigger viral reservoir in T regulatory cells. Nat. Commun. 8: 498. [PubMed: 28887441]

11. Liu R, Simonetti FR, Ho YC. 2020. The forces driving clonal expansion of the HIV-1 latent reservoir. Virol. J. 17: 4. [PubMed: 31910871]
12. Sather BD, Romano Ibarra GS, Sommer K, Curinga G, Hale M, Khan IF, Singh S, Song Y, Gwiazda K, Sahni J, Jarjour J, Astrakhan A, Wagner TA, Scharenberg AM, Rawlings DJ. 2015. Efficient modification of CCR5 in primary human hematopoietic cells using a megaTAL nuclease and AAV donor template. Sci. Transl. Med. 7: 307ra156.
13. Hung KL, Meitlis I, Hale M, Chen C-Y, Singh S, Jackson SW, Miao CH, Khan IF, Rawlings DJ, James RG. 2018. Engineering Protein-Secreting Plasma Cells by Homology-Directed Repair in Primary Human B Cells. Mol. Ther. 26: 456–467. [PubMed: 29273498]
14. Stemmer M, Thumberger T, Keyer M. d. S., Wittbrodt J, Mateo JL. 2015. CCTop: An Intuitive, Flexible and Reliable CRISPR/Cas9 Target Prediction Tool. PLoS. One. 10: e0124633. [PubMed: 25909470]
15. Anderson W, Thorpe J, Long SA, Rawlings DJ. 2019. Efficient CRISPR/Cas9 Disruption of Autoimmune-Associated Genes Reveals Key Signaling Programs in Primary Human T Cells. J. Immunol. 203: 3166–3178. [PubMed: 31722988]
16. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 29: 15–21. [PubMed: 23104886]
17. Anders S, Pyl PT, Huber W. 2015. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics. 31: 166–169. [PubMed: 25260700]
18. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15: 550. [PubMed: 25516281]
19. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. 25: 25–29. [PubMed: 10802651]
20. The Gene Ontology Consortium. 2019. The Gene Ontology Resource: 20 years and still GOing strong. Nucleic Acids Res. 47: D330–D338. [PubMed: 30395331]
21. Mi H, Muruganujan A, Ebert D, Huang X, Thomas PD. 2019. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. Nucleic Acids Res. 47: D419–D426. [PubMed: 30407594]
22. Hale M, Lee B, Honaker Y, Leung W-H, Grier AE, Jacobs HM, Sommer K, Sahni J, Jackson SW, Scharenberg AM, Astrakhan A, Rawlings DJ. 2017. Homology-Directed Recombination for Enhanced Engineering of Chimeric Antigen Receptor T Cells. Mol. Ther. Methods Clin. Dev. 4: 192–203. [PubMed: 28345004]
23. Hubbard N, Hagin D, Sommer K, Song Y, Khan I, Clough C, Ochs HD, Rawlings DJ, Scharenberg AM, Torgerson TR. 2016. Targeted gene editing restores regulated CD40L function in X-linked hyper-IgM syndrome. Blood. 127: 2513–2522. [PubMed: 26903548]
24. Honaker Y, Hubbard N, Xiang Y, Fisher L, Hagin D, Sommer K, Song Y, Yang SJ, Lopez C, Tappen T, Dam EM, Khan I, Hale M, Buckner JH, Scharenberg AM, Torgerson TR, Rawlings DJ. 2020. Gene editing to induce FOXP3 expression in human CD4+ T cells leads to a stable regulatory phenotype and function. Sci. Trans. Med 12: eaay6422.
25. Sadelain M, Papapetrou EP, Bushman FD. 2012. Safe harbours for the integration of new DNA in the human genome. Nat. Rev. Cancer. 12: 51–58.
26. Chen X, Oppenheim JJ. 2011. Resolving the identity myth: key markers of functional CD4+FoxP3+ regulatory T cells. Int. Immunopharmacol. 11: 1489–1496. [PubMed: 21635972]
27. Schmidt A, Oberle N, Krammer PH. 2012. Molecular Mechanisms of Treg-Mediated T Cell Suppression. Front. Immunol. 3: 51. [PubMed: 22566933]
28. Malek TR, Castro I. 2010. Interleukin-2 receptor signaling: at the interface between tolerance and immunity. Immunity. 33: 153–165. [PubMed: 20732639]
29. Johnston JA, Bacon CM, Finbloom DS, Rees RC, Kaplan D, Shibuya K, Ortaldo JR, Gupta S, Chen YQ, Giri JD. 1995. Tyrosine phosphorylation and activation of STAT5, STAT3, and Janus kinases by interleukins 2 and 15. Proc. Natl. Acad. Sci. U. S. A. 92: 8705–8709. [PubMed: 7568001]

30. Eipers PG, Salazar-Gonzalez JF, Morrow CD. 2011. HIV Gene Expression from Intact Proviruses Positioned in Bacterial Artificial Chromosomes at Integration Sites Previously Identified in Latently Infected T cells. Virology. 410: 151–160. [PubMed: 21115184]

31. Lenasi T, Contreras X, Peterlin BM. 2008. Transcriptional interference antagonizes proviral gene expression to promote HIV latency. Cell Host Microbe. 4: 123–133. [PubMed: 18692772]

32. Sherrill-Mix S, Ocwieja KE, Bushman FD. 2015. Gene activity in primary T cells infected with HIV89.6: intron retention and induction of genomic repeats. Retrovirology. 12: 79. [PubMed: 26377088]

33. Liu R, Yeh Y-HJ, Varabyou A, Collora JA, Sherrill-Mix S, Talbot CC, Mehta S, Albrecht K, Hao H, Zhang H, Pollack RA, Beg SA, Calvi RM, Hu J, Durand CM, Ambinder RF, Hoh R, Deeks SG, Chiarella J, Spudich S, Douek DC, Bushman FD, Pertea M, Ho Y-C. 2020. Single-cell transcriptional landscapes reveal HIV-1-driven aberrant host gene transcription as a potential therapeutic target. Sci. Transl. Med. 12: eaaz0802.

34. Grant FM, Yang J, Nasrallah R, Clarke J, Sadiyah F, Whiteside SK, Imianowski CJ, Kuo P, Vardaka P, Todorov T, Zandhuis N, Patrascan I, Tough DF, Kometani K, Eil R, Kurosaki T, Okkenhaug K, Roychoudhuri R. 2020. BACH2 drives quiescence and maintenance of resting Treg cells to promote homeostasis and cancer immunosuppression. J. Exp. Med. 217: e20190711. [PubMed: 32515782]

35. Sidwell T, Liao Y, Garnham AL, Vasanthakumar A, Gloury R, Blume J, Teh PP, Chisanga D, Thelemann C, Rivera F. d. L., Engwerda CR, Corcoran L, Kometani K, Kurosaki T, Smyth GK, Shi W, Kallies A. 2020. Attenuation of TCR-induced transcription by Bach2 controls regulatory T cell differentiation and homeostasis. Nat. Commun. 11: 1–17. [PubMed: 31911652]

36. Kim EH, Gasper DJ, Lee SH, Plisch EH, Svaren J, Suresh M. 2014. Bach2 regulates homeostasis of Foxp3+ regulatory T cells and protects against fatal lung disease in mice. J. Immunol. 192: 985–995. [PubMed: 24367030]

37. Simonetta F, Bourgeois C. 2013. CD4+FOXP3+ Regulatory T-Cell Subsets in Human Immunodeficiency Virus Infection. Front. Immunol. 4: 215. [PubMed: 23908654]

38. Chevalier MF, Weiss L. 2013. The split personality of regulatory T cells in HIV infection. Blood. 121: 29–37. [PubMed: 23043072]

39. Valverde-Villegas JM, Matte MCC, de Medeiros RM, Chies JAB. 2015. New Insights about Treg and Th17 Cells in HIV Infection and Disease Progression. J. Immunol. Res. 2015: 647916. [PubMed: 26568963]

40. Lim A, Tan D, Price P, Kamarulzaman A, Tan H-Y, James I, French MA. 2007. Proportions of circulating T cells with a regulatory cell phenotype increase with HIV-associated immune activation and remain high on antiretroviral therapy. AIDS. 21: 1525–1534. [PubMed: 17630546]

41. Schulze Zur Wiesch J, Thomssen A, Hartjen P, Tóth I, Lehmann C, Meyer-Olson D, Colberg K, Frerk S, Babikir D, Schmiedel S, Degen O, Mauss S, Rockstroh J, Staszewski S, Khaykin P, Strasak A, Lohse AW, Fätkenheuer G, Hauber J, van Lunzen J. 2011. Comprehensive analysis of frequency and phenotype of T regulatory cells in HIV infection: CD39 expression of FoxP3+ T regulatory cells correlates with progressive disease. J. Virol. 85: 1287–1297. [PubMed: 21047964]

42. Zhuang Y, Wei X, Li Y, Zhao K, Zhang J, Kang W, Sun Y. 2012. HCV coinfection does not alter the frequency of regulatory T cells or CD8+ T cell immune activation in chronically infected HIV+ Chinese subjects. AIDS Res. Hum. Retroviruses. 28: 1044–1051. [PubMed: 22214236]

43. Zhang Z, Jiang Y, Zhang M, Shi W, Liu J, Han X, Wang Y, Jin X, Shang H. 2008. Relationship of frequency of CD4+CD25+Foxp3+ regulatory T cells with disease progression in antiretroviral-naive HIV-1 infected Chinese. Jpn. J. Infect. Dis. 61: 391–392. [PubMed: 18806350]

44. Loke P. n., Favre D, Hunt PW, Leung JM, Kanwar B, Martin JN, Deeks SG, McCune JM. 2010. Correlating cellular and molecular signatures of mucosal immunity that distinguish HIV controllers from noncontrollers. Blood. 115: e20–32. [PubMed: 20160163]

45. Tenorio AR, Martinson J, Pollard D, Baum L, Landay A. 2008. The relationship of T-regulatory cell subsets to disease stage, immune activation, and pathogen-specific immunity in HIV infection. J. Acquir. Immune Defic. Syndr 48: 577–580. [PubMed: 18645514]

46. Eggena MP, Barugahare B, Jones N, Okello M, Mutalya S, Kityo C, Mugyenyi P, Cao H. 2005. Depletion of regulatory T cells in HIV infection is associated with immune activation. J. Immunol. 174: 4407–4414. [PubMed: 15778406]

47. Nilsson J, Boasso A, Velilla PA, Zhang R, Vaccari M, Franchini G, Shearer GM, Andersson J, Chougnet C. 2006. HIV-1-driven regulatory T-cell accumulation in lymphoid tissues is associated with disease progression in HIV/AIDS. Blood. 108: 3808–3817. [PubMed: 16902147]

48. Bouscarat F, Levacher-Clergeot M, Dazza MC, Strauss KW, Girard PM, Ruggeri C, Sinet M. 1996. Correlation of CD8 lymphocyte activation with cellular viremia and plasma HIV RNA levels in asymptomatic patients infected by human immunodeficiency virus type 1. AIDS Res. Hum. Retroviruses. 12: 17–24. [PubMed: 8825614]

49. Deeks SG, Kitchen CMR, Liu L, Guo H, Gascon R, Narváez AB, Hunt P, Martin JN, Kahn JO, Levy J, McGrath MS, Hecht FM. 2004. Immune activation set point during early HIV infection predicts subsequent CD4+ T-cell changes independent of viral load. Blood. 104: 942–947. [PubMed: 15117761]

KEY POINTS:

1. HIV integration frequently targets the intron preceding the first exon of *BACH2*
2. Gene editing was used to model the impact of integration in primary CD4+ T cells
3. BACH2 IS targeting led to selective expansion of a T regulatory-like population

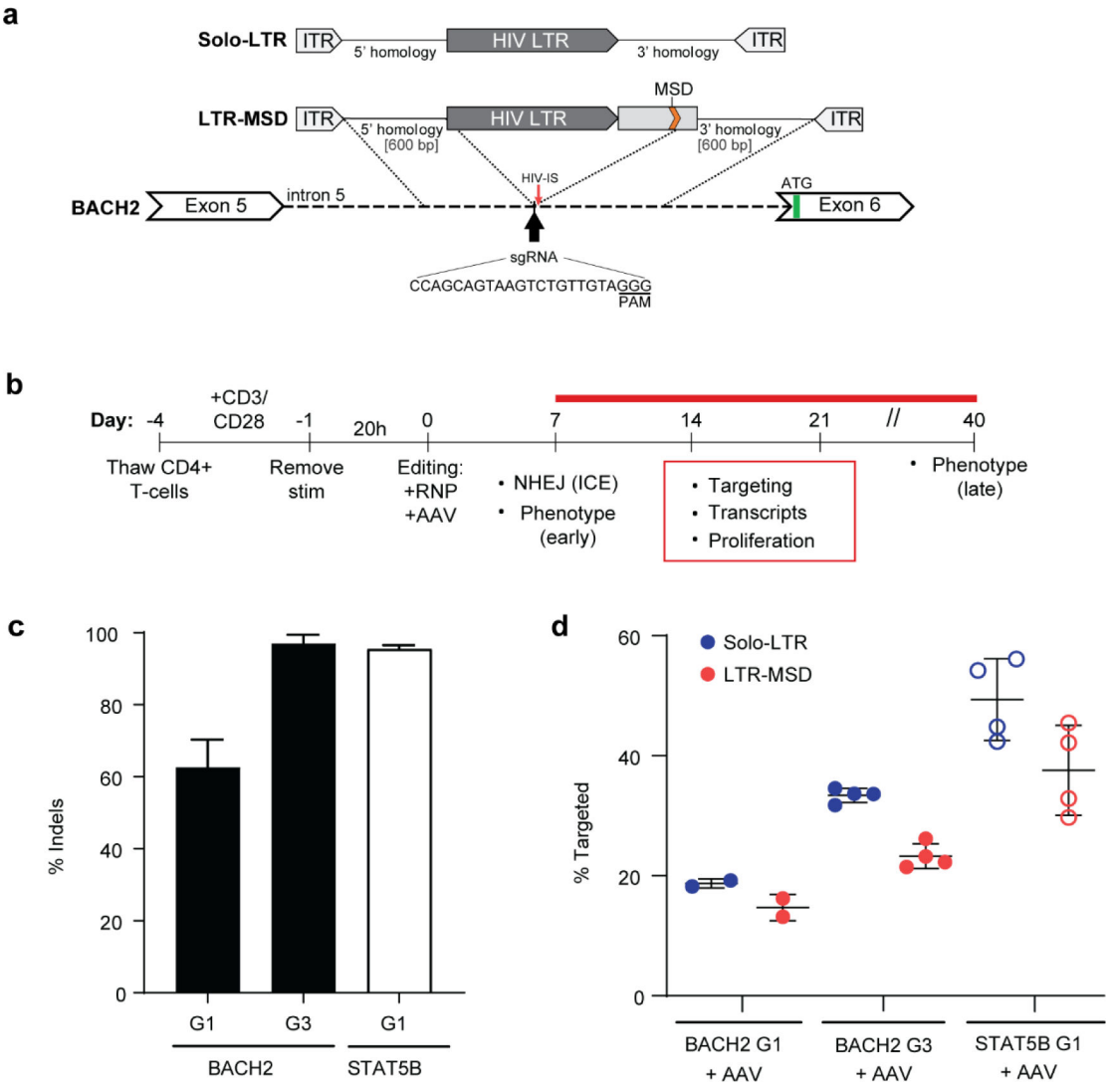

Fig. 1. Gene editing using AAV-CRISPR/Cas9 for targeted HIV-LTR insertion at the *BACH2* intron 5 locus.

a Schematic of the CRISPR/Cas9-AAV HDR-based editing approach for HIV-LTR insertion within intron 5 of the *BACH2* locus. The AAV donor constructs consist of a single HIV-LTR (long terminal repeat) element (Solo-LTR) or an LTR followed by viral sequence leading up to the major splice donor (LTR-MSD), flanked by 600 bp homology arms to BACH2 and the AAV inverted terminal repeats (AAV ITR) required for AAV packaging and viral production. The CRISPR/Cas9 target site for guide 3 (G3) is indicated by a black arrow; relative location of a known HIV-insertion site at BACH2 intron 5 is indicated by red arrow[5]. **b** Experimental protocol and timeline. **c** Frequency of indels generated by *BACH2* and *STAT5B* guides as measured by Inference of CRISPR Edits (ICE). For *BACH2* G1 n = 6, *BACH2* G3 n = 5, and *STAT5B* G1 n = 4. **d** Guide-specific HDR-editing efficiency of AAV HIV-LTR constructs as measured by ddPCR in different PBMC donors. For *BACH2* G1 n = 3; *BACH2* G3 (n = 4) is shown, n =7; and *STAT5B* G1 n = 4. All data are represented as mean and error bars indicate SD.

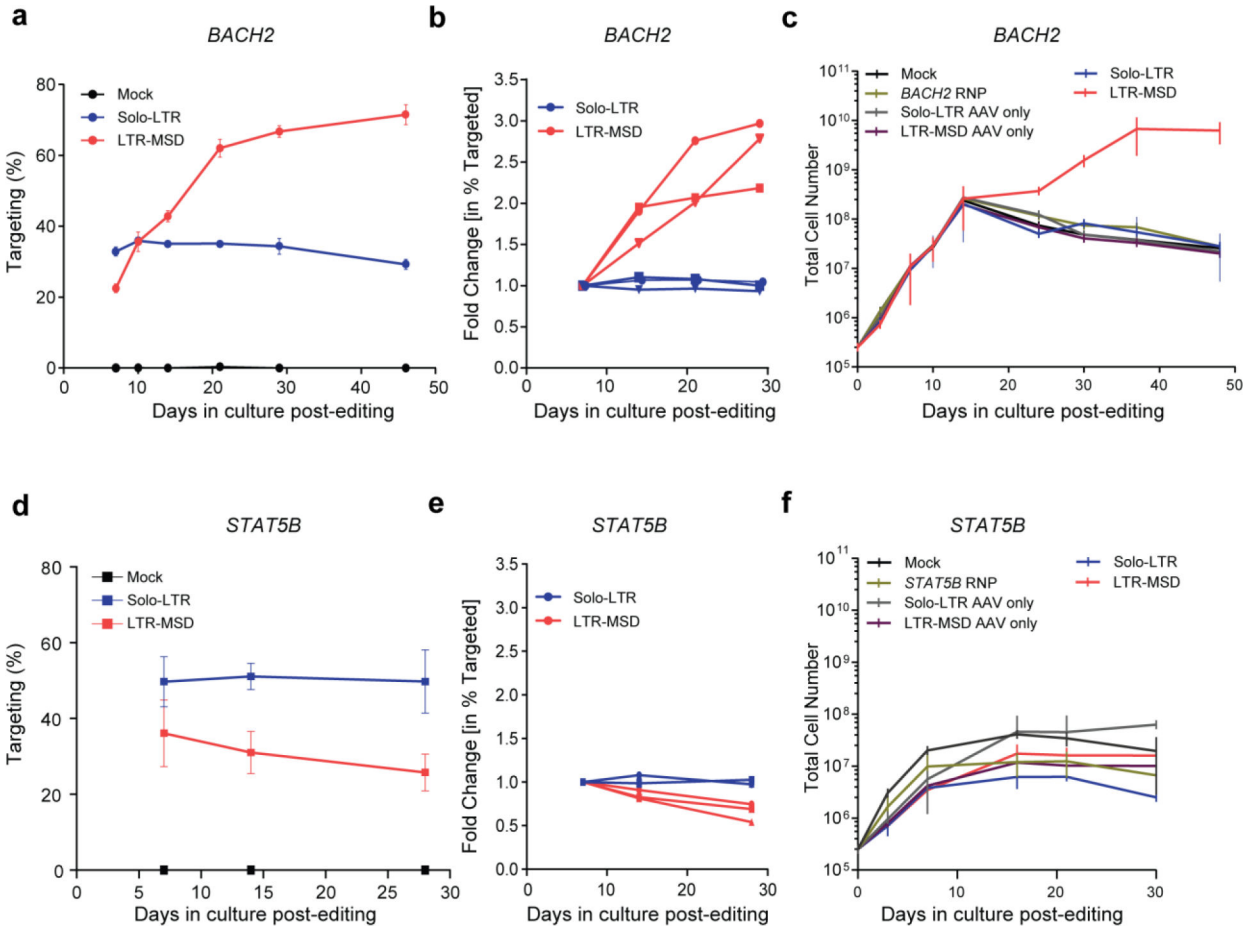

Fig. 2. Quantification of integrated HIV-LTR cassettes showing specific expansion of *BACH2* LTR-MSD targeted primary CD4+ T cells over time.

a Representative outgrowth of Solo-LTR and LTR-MSD edited populations from a single experiment targeting *BACH2* over time, measured by ddPCR. Error bars represent SD between 2 biological replicates in each experiment. **b** Fold expansion of edited populations targeting *BACH2* over time in different PBMC donor cells. Representative, (n = 3) is shown, n = 7. **c** Proliferation of CD4+ T-cell samples from *BACH2* targeting experiments assessed over time. Representative, (n = 3) is shown, n = 14. **d** Representative outgrowth of Solo-LTR and LTR-MSD edited populations from a single experiment targeting *STAT5B* over time, measured by ddPCR. Error bars represent SD between 2 biological replicates in each experiment. **e** Fold expansion of edited populations targeting *STAT5B* over time in different PBMC donor cells. Representative, (n = 3) is shown, n = 4. **f** Proliferation of CD4+ T-cell samples from *STAT5B* targeting experiments assessed over time. Representative, (n = 3) is shown, n = 5. All data are represented as mean and error bars indicate SD.

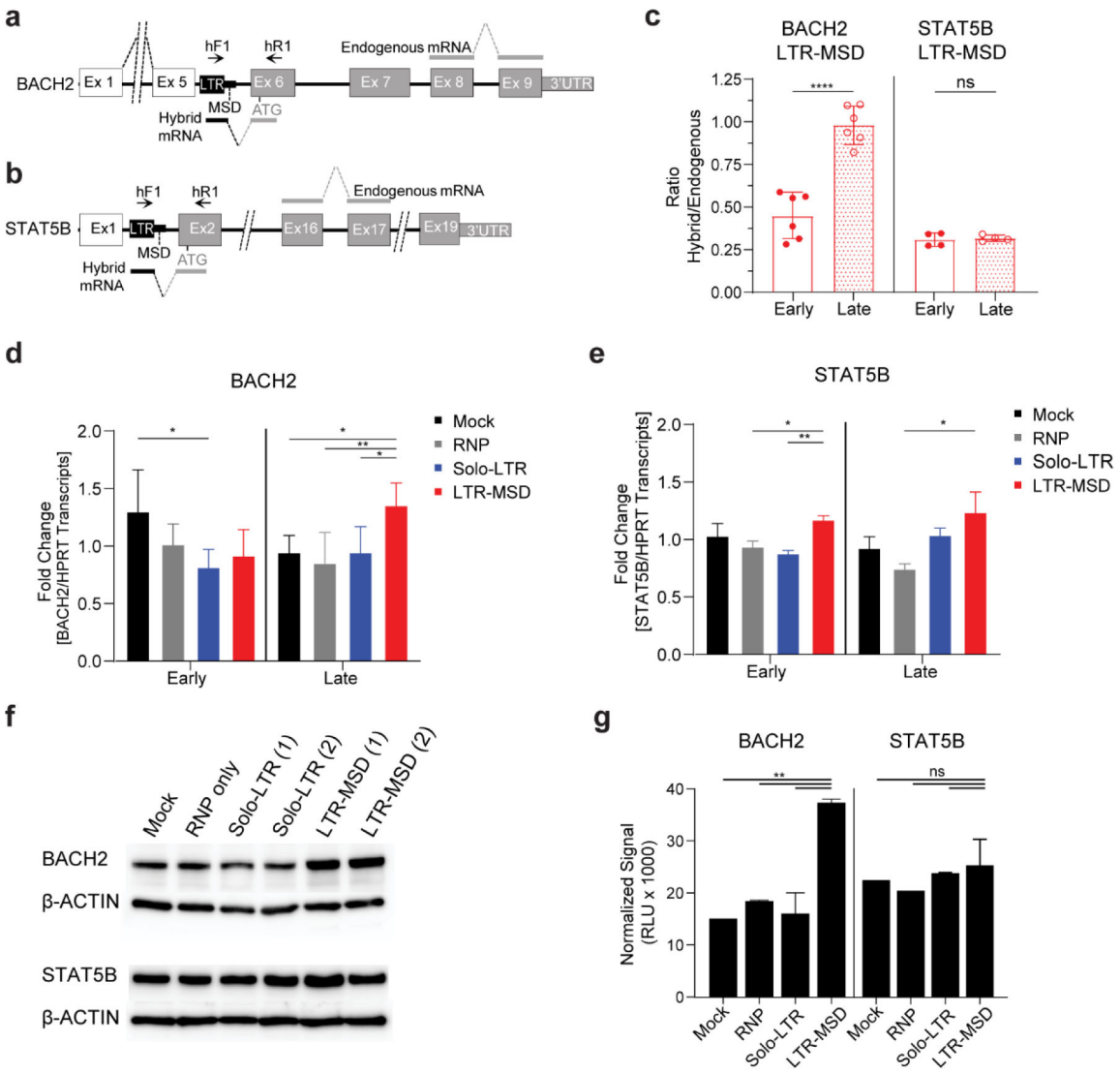

Fig. 3. BACH2 LTR-MSD editing promotes increased expression of LTR-hybrid mRNA and BACH2 protein.

a,b Schematic of BACH2/STAT5B endogenous and LTR-hybrid mRNA; white and gray boxes indicate non-coding and coding exons (Ex), respectively. The integrated LTR-MSD cassette is shown in black. Arrows represent primer sets for RT-ddPCR capturing hybrid transcripts (hF1/hR1). Bars indicate amplified cDNA products from the endogenous or hybrid primer sets with dashed lines showing spliced regions. **c** Ratio of hybrid BACH2 or STAT5B to endogenous BACH2 or STAT5B transcripts, respectively, standardized to HPRT transcripts, generated by LTR-MSD targeted samples at early (day 7) and late (day 35) timepoints. n = 6 for BACH2 and n = 4 for STAT5B. **d, e** Fold change of BACH2 and STAT5B endogenous transcripts standardized to HPRT transcripts relative to the average expression across all sample groups within an experiment. Levels at early and late timepoints measured by ddPCR, n = 6 (**c**) or 3 (**d**). **f** Western blot analysis of BACH2 (upper panel) and STAT5B (lower panel) protein expression from a representative experiment with 2 biological replicates for each HDR-targeted sample. β-actin expression was assessed

in parallel as a loading control. The representative western blots were performed twice independently. **g** Normalized and averaged signal from the Western blots in (**f**). Significance was determined by one-way ANOVAs with Tukey's multiple comparison tests (**d, e, g**), as well as Student's two-tailed t-tests (**c**). All data are represented as mean and error bars indicate SD.

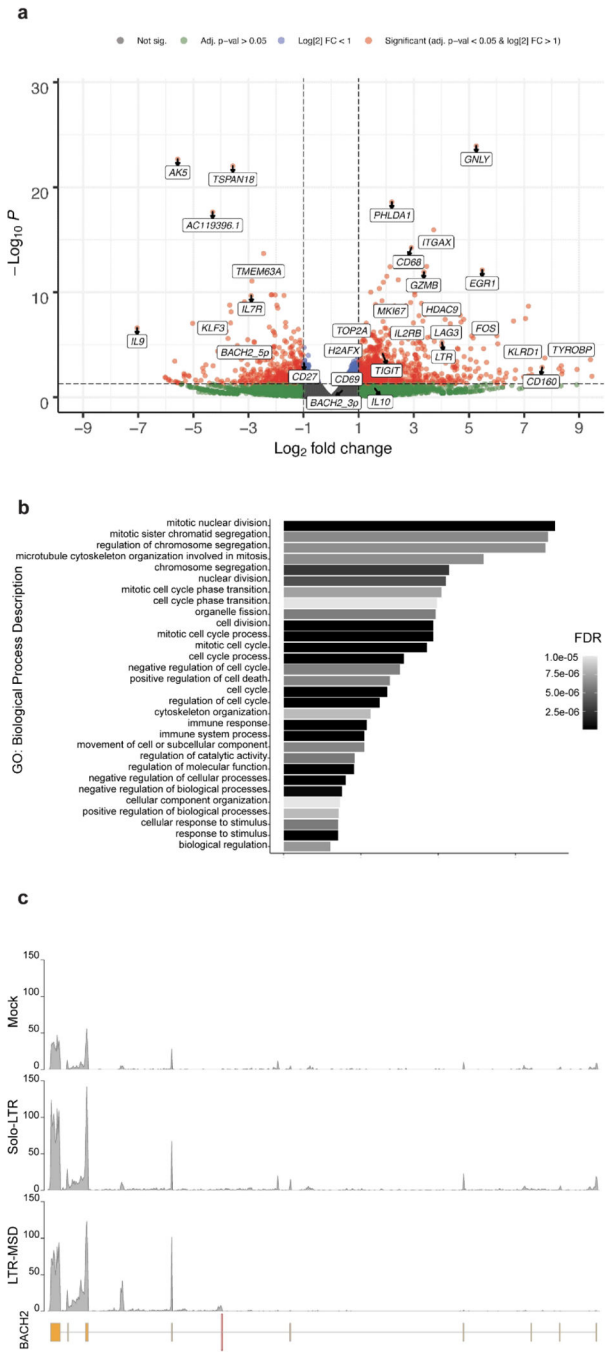

Fig 4. RNA-seq analysis of edited populations showing unique gene expression profile and hybrid transcripts in BACH2 LTR-MSD targeted T cells.

a) Volcano plot showing the 452 differentially expressed genes between BACH2 LTR-MSD and Solo-LTR edited CD4 T cell populations at 48 days of culture (309 genes upregulated in LTR-MSD and 143 down-regulated in LTR-MSD compared to Solo-LTR, respectively). Each dot represents a gene with: grey not reaching significance and less than an absolute log2 fold change of 1; green not reaching significance but an absolute log2 fold change > 1; blue having an adjusted p-value < 0.05 but less than an absolute log2 fold change of 1;

and red having an adjusted p-value < 0.05 and an absolute log2 fold change > 1. Log2 fold change > 0 means higher expression in LTR-MSD, and vice versa. A select number of genes were annotated. **b)** Shown are the top 30 alphabetically sorted genesets, fold-enrichment scores, and false discovery rate (FDR) determined from a PANTER (Protein ANalysis Through Evolutionary Relationships) overrepresentation test of *homo sapien* gene-ontology biological processes. **c)** Coverage map of RNA-seq reads mapped to the hg38 chr6 BACH2 locus. Gene polarity is displayed right to left as on chromosome 6 for mock edited control cells, LTR-solo edited, and LTR-MSD edited cells. Accumulation of reads are shown as histograms scaled to read count (y-axis). Exons 1–9 are shown as orange rectangles with the LTR insertion into intron 5 indicated as a red bar.

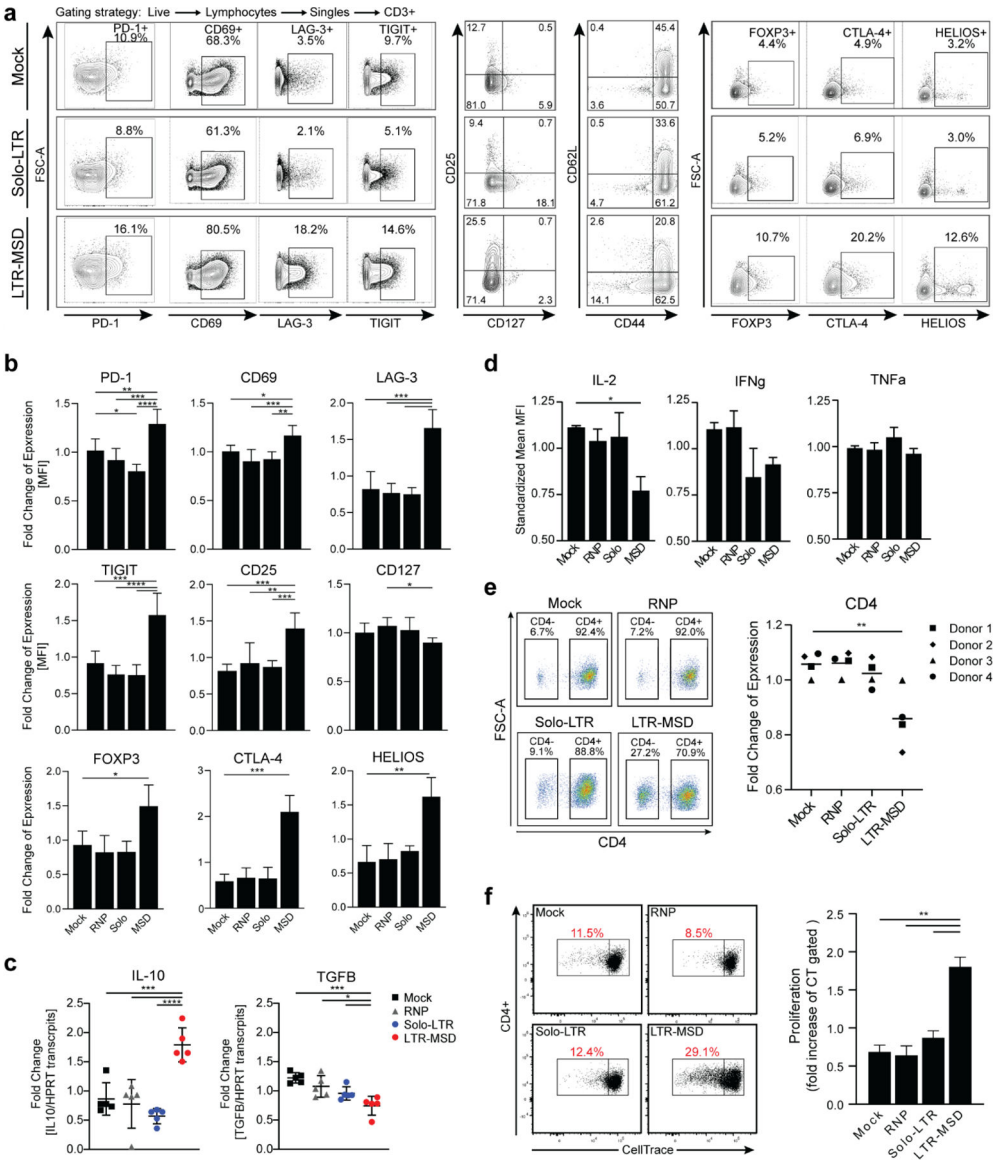

Fig 5. BACH2 LTR-MSD targeted T cells exhibit a unique proliferative, activated, Treg-like cell phenotype.

a Representative flow plots and gating strategy of edited and control samples for the indicated cell surface and intracellular markers. **b** Pooled data of markers in (**a**) with n = 6 (PD-1, CD69, CD25, CD127), n = 5 (LAG-3, TIGIT), and n = 3 (FOXP3, CTLA-4, HELIOS). Data shown as fold change compared to average mean MFI from all sample groups, within an independent experiment. **c** Combined *IL-10* and *TGFB* transcript levels from *BACH2* edited samples standardized to *HPRT* transcripts at the late timepoint, n = 5. Data shown as fold change compared to average mean expression from all sample groups, within an independent experiment. **d** Pooled cytokine production from 4 experiments, standardized as in (**b**). Briefly, cells were stimulated with PMA, Ionomycin, and Golgi-stop before fixing/perming and staining for cytokine production. **e** Left, representative flow plots tracking CD4 expression post editing. Right, combined data showing fold-change in CD4 expression from experiments using 4 independent donors **f** Left, representative flow plots of

CellTrace (CT) proliferation dye staining after 96h incubation. Briefly, cells were washed and incubated with CT dye for 20 mins, then quenched with fresh media for 5 mins. Cells were then spun down and cultured in full cytokine media for 96h at which point they were flowed for CT dilution. Right, pooled data of CT stain from 3 independent experiments. Data represented as mean and error bars indicate SD.