Defective proviruses cause T cell reprogramming through promoter exaptation in HIV-1 infection

Martin V. Hamann[1]*, Lisa Brauckmann[1]*, Christoph Schwarz[1], Michael Spohn[2], Ramon Stoeck[1], Sabrina M. Leddy[1,3], Julie Frouard[4], Maisha Adiba[1], Friederike Thiele[6], Sanamjeet Virdi[1], Adam Grundhoff[1], Cedric Feschotte[3], Nadia R. Roan[4, 5], Eva Tolosa[6], Ulrike C. Lange[1,7,8]

[1]Leibniz Institute of Virology, Hamburg, Germany;
[2]Bioinformatics Core, University Medical Center Hamburg–Eppendorf, Hamburg, Germany;
[3]Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY, USA, [4]Gladstone Institutes, UCSF, San Francisco, CA, USA;
[5]University of California San Francisco, San Francisco, CA, USA;
[6]Institute of Immunology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany;
[7]Institute for Infection Research and Vaccine Development, University Medical Center Hamburg–Eppendorf, Hamburg, Germany;
[8]German Center for Infection Research (DZIF), partner site Hamburg-Luebeck-Borstel-Riems, Germany
*contributed equally

Corresponding author: Ulrike C. Lange
e-mail: ulrike.lange@leibniz-liv.de

Classification: BIOLOGICAL SCIENCES, Medical Sciences

Keywords: BACH2, HIV-1, proviruses, clonal expansion, HIV persistence

Abstract

People living with HIV (PLWH) on antiretroviral therapy (ART) accumulate primarily defective proviral sequences in genomes of often clonally expanded CD4+ HIV-1 target cells. Viral-derived DNA is preferentially found at distinct genomic loci suggesting a selective process driven by integration site-specific crosstalk between viral and host sequences. Focusing on one of the most prominent selected integration loci, the BTB Domain and CNC Homolog 2 (*BACH2*) gene, we here show evidence for exaptation of provirus-derived regulatory sequences leading to target cell reprogramming during long-term ART. Using a cellular model of *BACH2*-integrated proviruses, we find that proviral transcription drives aberrant BACH2 protein levels that escape autoregulatory feedback and impose BACH2-dependent transcriptomic changes. By mimicking these changes in primary CD4+ T lymphocytes, we observe that BACH2 drives reprogramming of cells toward a proliferative, precursor memory-like type. These reprogrammed CD4+ T cells possess traits of immune evasion and cellular survival that are signatures of HIV reservoir cells in PLWH. Inhibition of provirus transcriptional activity can mitigate exaptation, suggesting a strategy to offset HIV-driven differentiation and expansion of CD4+ T cells. Finally, our data suggest that provirus exaptation at the Signal Transducer And Activator of Transcription 5B (STAT5B) selected integration gene drives a contrary, effector-like T cell fate, suggesting a multifaceted impact of exaptation on immune homeostasis. Overall, our data demonstrate that proviruses, even if structurally defective, can modulate target cells through insertional activation of integration genes, a process which we postulate to contribute to the complex immune dysregulation experienced by ART-suppressed PLWH.

Significance statement

People living with HIV (PLWH) on suppressive therapy accumulate genomically integrated proviruses in HIV-1-targeted immune cells. We show that HIV proviruses can exapt viral regulatory elements to disrupt integration gene control, resulting in reprogramming of HIV-targeted CD4+ T cells toward different fates. The most prominent BACH2-integrated proviruses for example direct cells toward a proliferative, memory-like state with features of persistence and immune evasion. This mechanism demonstrates how even defective proviruses contribute to immune dysregulation in HIV-1 infection, and suggests that blocking proviral transcription may mitigate these effects in PLWH.

Introduction

People living with HIV-1 (PLWH) harbor chromosomally integrated viral genomes in HIV-1 reservoir cells. These proviruses are remnants of the initial infection and persist despite suppressive antiretroviral therapy (ART) (1, 2). One central challenge in current HIV research is understanding how these persistent proviruses lead to immune dysfunction that has been linked to reservoir maintenance and non-AIDS co-morbidities in PLWH and devising strategies how to halt this process (3). The provirus reservoir in PLWH on long-term ART has been extensively characterized, and displays two main features: structural defectiveness and clonality. Less than 2% of proviruses can support replication and viral re-bound upon ART cessation. The vast majority are sequence-defective, demonstrating point mutations or deletions (4–8). Although not replication-competent, defective proviruses can impact on immune responses in PLWH: Since ART does not inhibit HIV-1 transcription, defective proviruses can be the source of viral transcripts and peptides that elicit type 1 interferon responses, production of inflammatory cytokine, and deflection of CTL responses (9–13). Defective proviruses have also been suggested as source for non-suppressible residual viremia in PLWH on ART and as driver for HIV-related disease progression and lack of immunological response during therapy (7, 14, 15).

The provirus reservoir in PLWH on ART is also highly clonal (16–20). Clonal expansion has been observed for cells with intact as well as defective proviruses. Homeostatic proliferation and antigen-driven stimulation have been suggested as major driving forces for clonal expansion of HIV reservoir cells (21, 22). While these mechanisms are provirus-independent, evidence from longitudinal studies of proviral integration landscapes in PLWH on ART suggest an additional, provirus-dependent expansion process. Time on ART selects for clones with specific integration genes, termed recurrently-detected integration genes (RdIGs) (23–29). These RdIGs often appear to be associated with pathways promoting cellular proliferation and/or survival (17, 18, 28, 29). Clonal expansion of cells with proviral integrations at RdIGs has thus been suggested to result from functional crosstalk between provirus and human DNA, in which virus-derived promoter elements cause insertional activation of integration genes (22, 23). While often stated, there is so far little direct evidence in support of this process of HIV-1 promoter exaptation.

The *BTB And CNC Homolog 2* (*BACH2)* gene has been identified as one of the most prominent RdIGs (28–32). *BACH2* encodes a transcriptional master regulator in B and T lymphocytes, that has been associated with maintaining immune homeostasis by restraining terminal differentiation, promoting memory and regulatory subsets, and ensuring effective, balanced immune responses (33–37). Time on ART selects for *BACH2* provirus integrations that are clustered in congruent transcriptional orientation upstream of the gene's protein coding exons (27–31). This selection pattern is highly reminiscent of a functional crosstalk in terms of promoter exaptation. Indeed, chimeric proviral/*BACH2* transcripts, which are initiated by splicing from the HIV-1 major splice donor site (MSD) into an endogenous splice acceptor site of *BACH2,* have been detected in peripheral blood mononuclear cells from PLWH (38–40). The selective pressure for clones carrying *BACH2* proviral integrants is remarkably high: *BACH2* integrations and proviral/BACH2 transcripts have been found in up to 40% of long-term suppressed PLWH (30, 38, 39). Despite this prominence, mechanistic details of how promoter exaptation occurs at the BACH2 locus and how this impacts the host cell are largely unknown, We have previously shown that transcriptional activity of *BACH2*-integrated HIV-1 promoter sequences is not linked to *BACH2* gene expression, indicating that effects driven by promoter-exaptation are potentially governed by distinct regulatory pathways compared to endogenous BACH2 expression (30). It has also been suggested that proviral/BACH2 crosstalk could promote a Treg-reservoir population that maintains HIV persistence (38). However, no sequence-intact *BACH2*-integrated provirus has so far been reported in PLWH on ART, making contribution to the replication-competent reservoir less likely (12, 21, 41–44). A second study suggested that proviral integration into BACH2 could cause

differentiation towards a Treg-like phenotype, however there was no direct evidence that changes were driven by insertional BACH2 activation (39).

In the present study, we set out to further define how provirus integration into BACH2 affects infected CD4+ T cells and how this could lead to immune modulation in PLWH. Using targeted genome engineering, we generated a cellular model system, that replicates *BACH2* proviral integrations observed *in vivo* as unique tool to study provirus/human crosstalk at this locus. We further determine how, as a result, aberrant BACH2 expression reprograms primary CD4+ T lymphocytes and thus alters the target cell immune state. We highlight how reprogrammed cells show features of expanded HIV-1 reservoir clones in PLWH, reveal that outcomes of reprogramming are integration-gene dependent and show the potential to halt provirus-mediated insertional activation through compounds that promote HIV-1 latency. Our results emphasize a mechanism by which persistent proviruses, independent of replication competence, have the ability the modulate the immune landscape of PLWH on ART through exaptation of virus-derived regulatory elements.

Results

1. LTR exaptation drives BACH2 protein expression in a cellular model that replicates *in vivo*-selected *BACH2* proviral clones

In order to establish a comprehensive view of BACH2 proviral integrations and the effects of *in vivo* selection, we first undertook comparative profiling of proviral integration sites (ISs) detected in PBMCs of aviremic PLWH on ART and integration sites from *in vitro* infected primary CD4 T cells. Using data from the Retrovirus Integration Database v2.0 (45), we catalogued 11,579 ISs and 3,898 integration genes (IGs) reported *in vivo* (12, 28, 29, 31, 32, 42, 46–50), and 463,206 ISs and 14,192 IGs *in vitro* (24, 47, 51, 52) from a total of 15 studies. Consistent with prior reports (24, 28–31), we detected the *BACH2* locus as the most frequent and most *in vivo* selected integration gene (Fig1a, SuppFig1). We also observed that *in vivo BACH2*-integrations showed significant enrichment towards intron 5 and congruent transcriptional orientation with respect to the gene (Fig1b).

In order to study how proviral sequences interact with the BACH2 integration locus, we then generated a cellular model based on the features of *in vivo*-selected BACH2-proviruses. We used CRISPR/Cas9-mediated genome engineering of a leukemia cell line to target a minimal HIV-1 reporter into intron 5 of the *BACH2* gene, the precise region that BACH2 integrants are selected for in PLWH (Fig1c, SuppFig2a). Two clonal lines, HIV_BACH2^{D5} and HIV_BACH2^{E10} were generated, that carry a heterozygous (D5) or homozygous (E10) insertion of the HIV-1 reporter in congruent transcriptional orientation at the targeted site (Fig1c). The reporter consists of an expression cassette for blue fluorescent protein (bfp) and HIV-1 transcriptional activator tat, flanked by two HIV-1 long terminal repeat (LTR) sequences (Fig1d). Of note, the gag leader sequence containing HIV-1 major splice donor site (MSD) is included in the reporter upstream of bfp, allowing for splicing of LTR-driven transcripts. Clonal lines were validated for correct reporter insertion using Northern blotting and Sanger sequencing (SuppFig2b, c).

In both clonal lines, transcriptional activity of the LTR (as read out by bfp expression) was minimal but could be induced by exposure to phorbol myristate acetate (PMA) PMA/Ionomycin (Iono), a potent HIV-1 latency reversing treatment (Fig1e). This is consistent with our previous findings on HIV-1 reporter integrations into *BACH2* intron 5 (30). We next tested whether PMA/Iono stimulation can lead to generation of chimeric LTR-BACH2 transcripts. Using primers targeting the reporter LTR and exon 6 of *BACH2*, we could indeed amplify LTR-BACH2 transcripts in cDNA derived from PMA/Iono-treated HIV_BACH2$^{D5/E10}$ cells (Fig1f). Sanger sequencing confirmed that chimeric transcripts were generated through aberrant splicing between the reporter MSD and the *BACH2* exon 6 splice acceptor site. To assess whether chimeric transcripts were translated to full length BACH2 protein, we carried out Western blot analysis of whole cell lysates of uninduced and PMA/Iono-treated HIV_BACH2$^{D5/E10}$ cells.

As shown in Fig1g, transcriptional induction by PMA/Iono indeed resulted in induction of BACH2 protein expression. These results demonstrate that the HIV_BACH2$^{D5/E10}$ lines provide a suitable *in vitro* model to study exaptation of proviral sequences, and insertional activation at the *BACH2* RdIG.

2. BACH2 integrated proviral sequences activate the *BACH2* gene expression program

Using our HIV_BACH2 model, we next went on to study downstream effects of LTR exaptation at the *BACH2* locus. To do so, we employed a tripartite CRISPR activator (CRISPRa) system, termed synergistic activation mediator (SAM), which allows for guide (g) RNA-directed recruitment of the VP64, p65 and HSF1 transcriptional activator domains to specified targets (53). Leveraging our previously validated gRNAs for efficient activation of the HIV-1 LTR and the *BACH2* gene promoter (30, 54), we set out to assess transcriptomic changes induced by SAM-mediated activation of the *BACH2*-integrated HIV reporter and contrast these to the transcriptomic programme induced by activation of the endogenous *BACH2* promoter. In brief, HIV_BACH2^{E10} cells were transiently transfected with LTR-targeting (LTR-SAM) or *BACH2*-targeting (BACH2-SAM) SAM. Expression of SAM components was linked to a fluorescent marker to allow for selection of successfully transfected cells. Wild-type (wt) cells, which harbor the endogenous *BACH2* locus without HIV-1 reporter integrant were also treated with BACH2-SAM to control for effects related to reporter integration in BACH2_HIV^{E10} cells. 72 hours post transfection SAM-expressing cells were enriched by flow cytometry and submitted to transcriptomic profiling (Fig2a).

We first assessed effects of BACH2-SAM treatment and observed that BACH2 gene transcription was induced in HIV_BACH2^{E10} and wt control cells showing the same transcript splicing pattern (Fig2b). Only minimal transcriptional induction of the BACH2-integrated reporter was observed in BACH2-SAM HIV_BACH2^{E10} cells, which confirms our previous results that activity of integrated LTRs within the BACH2 intron 5 is independent of BACH2 locus activity (30). Next, we examined the impact of LTR-SAM treatment: In HIV_BACH2^{E10} cells, LTR-SAM induced HIV reporter transcripts, as expected, but additionally a prominent splicing activity from the reporter-inherent MSD to BACH2 exon 6 was observed, giving rise to chimeric LTR/BACH2 transcripts. The splicing pattern of these chimeric transcripts downstream of exon 6 was reminiscent of endogenous BACH2 transcripts (Fig2b).

In order to assess the extent to which *BACH2* transcription induced by the LTR versus the endogenous promoter differentially affects the overall transcriptional profile of the cells, we performed RNA-seq on the SAM-induced HIV_BACH2^{E10} cells. As shown in Fig2c, in principal component space, LTR-SAM and BACH2-SAM-treated HIV-BACH2^{E10} samples segregate close to one another together with BACH2-SAM wt cells and away from non-treated wt or HIV_BACH2^{E10} cells. This indicates that transcriptomic changes induced by LTR-driven insertional activation of BACH2 mirror changes driven by endogenous BACH2. This was confirmed by differential gene expression analysis, which showed a highly similar pattern of differentially transcribed genes driven by LTR-driven versus endogenous BACH2 activation (Fig2d,e). Gene set enrichment analysis of differentially expressed genes (DEGs) in both comparisons further demonstrated that near identical cellular pathways were affected (SuppFig3a,b). Hence, these data indicate that transcriptional activity of *in vivo* selected BACH2-integrated proviruses results in transcriptomic reprogramming driven by BACH2 expression.

3. Aberrant BACH2 expression in CD4+ lymphocytes drives cells toward a proliferative precursor memory-like fate

We next set out to explore how LTR-driven aberrant activation of the *BACH2* gene program could impact CD4+ T cells. In this context it is noteworthy that dysregulated BACH2 expression has previously been reported to cause disease: *BACH2* haploinsufficiency can lead to hypogammaglobulinemia and immune dysregulation, manifesting clinically as recurrent infections and chronic intestinal

inflammation (55). The *BACH2* gene locus has also been prominently associated with conditions linked to immune imbalance in genome-wide association studies (GWAS) (SuppFig4) (56). All characterized single nucleotide polymorphisms in these studies mapped to non-coding, intronic or upstream regions of *BACH2* with highest occurrence at the 5' end of the locus and thus coinciding with regions that likely regulate BACH2 transcription and expression levels.

In order to investigate how elevated *BACH2* expression as a result of LTR exaptation could alter functions and differentiation states of primary CD4+ T cells, we used a lentiviral overexpression system (Fig3a). In short, CD4+ lymphocytes isolated from PBMCs of HIV-seronegative donors were transduced with lentiviruses coding for cDNAs of either BACH2 (LV-BACH2) or green fluorescent protein (LV-GFP) as negative control. By fusing BACH2 and GFP to a surface tag, transduced cells could be enriched using magnetic beads (Fig3b). Transduced cells showed robust expression of BACH2 or GFP RNA and protein over a time course of 25 days as assessed by RT-qPCR and flow cytometry (SuppFig5a).

We first assessed whether constitutive *BACH2* expression affects CD4+ T cell proliferation in this system. LV-BACH2 and LV-GFP-transduced cultures were stained with a marker dye 12 days post-transduction, and cell-division-dependent dilution of the dye was measured for up to five days. We observed a significant increase in proliferation of BACH2-transduced cells compared to GFP-transduced or non-transduced control cells as indicated by raised division index (Fig3c). These data suggest that BACH2 drives clonal expansion of primary CD4+ T cells.

We next determined BACH2-induced changes in the transcriptomic profile of primary CD4+ T cells. BACH2- and GFP-transduced cells were enriched at days 4, 9, and 25 post-transduction and analyzed by RNA-seq. In principle component space, LV-BACH2- and LV-GFP cells increasingly segregate away from one another during the 25 days of analysis (Fig3d). In agreement, differentially expressed gene (DEG) analysis revealed that *BACH2* overexpression progressively drives a new transcriptional program: at day 4 post transduction, 14 genes were differentially expressed, and this number increased to 183 at day 9 and 323 at day 25 (SuppTable1, Fig3e). As expected, *BACH2* transcripts were detected as significantly upregulated at all time points in LV-BACH2 cells.

We went on to assay transcriptomic changes using gene set enrichment analysis (GSEA) and STRING functional network profiling based on DEGs between LV-BACH2 and LV-GFP cells. At day 4, no particular differentially regulated pathway was highlighted, likely due to the low numbers of DEGs at this time point. However, we observe that LV-BACH2 cells show upregulation of RAS protein activator like 2 (*RASAL2),* that has been associated with inhibition of T cell activation and cytokine production (57), as well as downregulation of adhesion g protein-coupled receptor G1 (ADGRG1; also known as *GPR56*) and heme oxygenase 1 (*HMOX1)*, two genes linked to effector fates, namely cytotoxic T cells and T regs, respectively (58, 59) (SuppTable1).

At d9, GSEA shows downregulation of cellular pathways associated with chemokine signaling, cellular migration and immune response (Fig3f). Most downregulated DEGs belong to two main networks linked to effector T cell differentiation and function (eg. pro-inflammatory cytokine IFN gamma and chemokine receptor IL17RB (60)) as well as arachidonic acid metabolism, necessary for inflammatory signaling (61–63) (Fig3g). In addition, we observe downregulation of *PRDM1* (encoding B-lymphocyte-induced maturation protein-1 (Blimp1)) a direct target of BACH2. PRDM1 is a master transcriptional regulator that promotes terminal effector differentiation of T cells and reduced *PRDM1* expression levels have been shown to support an early memory T cell phenotype (64, 65).

At d25, LV-BACH2 cells show further downregulation of now a large subset of genes associated with effector T cells and inflammatory cytokine signaling, including CXC-chemokine receptor 6 (*CXCR6*), interleukins 5 and 13 (*IL5/13*) and granzyme B and H (*GZMB/H*) (Fig4a). Furthermore, we observe upregulation of *TCF7* (encoding T cell factor-1 (TCF-1)) and lymphoid enhancer-binding factor 1 (*LEF-1*), that have been reported to synergistically support formation of T cell memory precursors and acquisition of stemness properties for long-term persistence (66, 67). We also find positive enrichment

for pathways linked to mitotic activity, DNA repair and ribonucleotide complex binding in LV-BACH2 cells at d25, driven for example by upregulation of the mitotic regulators NIMA-related kinase 2 (*NEK2*) and cytoskeleton associated protein 2 (*CKAP2*) (Fig4a, SuppFig5b). This proliferative transcriptomic signature matches our results using cell tracer assays (Fig3c).

In summary, these data suggest that aberrant BACH2 expression induces a gradual transcriptomic reprogramming, marked by initial inhibition of CD4+ T cell activation and effector type differentiation, followed by initiation of a memory-precursor like program that features a signature of persistence and proliferation.

Finally we asked, whether observed transcriptional reprogramming induced by BACH2 overexpression in primary CD4+ T cells *in vitro* correlates with characteristics of clonally expanded HIV-1 infected cells in PLWH on ART. These cells demonstrate a signature of immunoregulatory markers and immune checkpoint molecules that confer T-cell functional inhibition and resistance to immune-mediated killing (44), coupled with a pro-survival gene expression profile (68–70). As observed for expanded clones *in vivo*, LV-BACH2 cells at d25 show upregulation of genes that protect from CTL or NK cell-mediated killing, such as *HLA-E*, poliovirus receptor (*PVR*), *TNFRSF14* (encoding herpes virus entry mediator (HVEM)) and programmed death ligand 1 (*PDL-1*) (Fig4b). We also detected induction of immune checkpoint genes, such as T-cell immunoreceptor with Ig and ITM domains (*TIGIT*), programmed cell death protein 1 (*PD1*) and B- and T-lymphocyte attenuator (*BTLA*) (Fig4b). Furthermore, BACH2 overexpression activates an intrinsic anti-apoptotic state in CD4+ T cells: as reported for expanded HIV-1 infected clones *in vivo*, LV-BACH2 cells were marked for example by upregulation of apoptotic inhibitor *BCL2L1* (encoding for BCL-XL protein), inhibitor of apoptosis (IAP) family member *BIRC5* and IAP regulator *USP19* (deubiquitinase that stabilizes IAP proteins) (Fig4c). Thus, these data indicate that BACH2-overexpression drives a CD4+ T cell phenotype that mimics the clonal expansion immune signature that is selected for in PLWH.

4. Aberrant expression of the STAT5B RdIG induces an activated, pro-survival CD4+ T cell state

While *BACH2* is the most prominent RdIG in PLWH on ART, a selection of further proviral integration genes have been proposed to be subject to insertional activation that could be driving cellular expansion (18, 23, 24, 27–29, 71). We asked whether insertional activation of these genes by proviral promoter exaptation could redirect CD4+ T cell fate in a similar way as here-demonstrated for BACH2. Signal transducer and activator of transcription 5B (*STAT5B*), is like *BACH2* a highly selected gene for proviral integrations in PLWH on ART (SuppFig1) (24, 27–29). STAT5B is a central transcriptional factor in CD4+ T cells that regulates homeostasis, survival and differentiation (72–75), and has been associated CD4+ T cell differentiation into Th1, Th2, Th9 and Treg cells (76). In PLWH the *STAT5B* RdIG shows significant clustering bias towards congruent proviral integrations in intron 1 (Fig5a) (24, 27) and chimeric provirus/STAT5B transcripts (spliced into STAT5B exon two which contains the start of the open reading frame) have been detected in PBMCs of PLWH on ART, providing evidence for provirus promoter exaptation at this locus.

To address whether exaptation-driven, aberrant STAT5B activation could redirect CD4+ T cell fate in a similar way as BACH2, we used above-described overexpression assay in primary CD4+ T cells. We transduced activated CD4+ T cells from HIV-seronegative donors with lentiviral SBP-deltaLNGFR bi-cistronic expression construct for STAT5B (LV-STAT5B) (Fig5b). LV-STAT5B cells showed robust expression of STAT5B at RNA and protein level (Fig5c).

Contrary to BACH2, aberrant STAT5B expression did not alter proliferation rates of CD4+ lymphocytes in comparison to controls in a marker dye-based assay (Fig5d). Further, transcriptomic profiling and DEG analysis at 12 day post transduction revealed that STAT5B overexpression upregulated 222 and downregulated 145 genes (SuppTable1). Differing from LV-BACH2-treated cells, these transcriptomic changes positively highlighted pathways of cell activation, immune response and MHC class II antigen

presentation as shown by GSEA and STRING analysis (Fig5e, 5f). Functional networks promoting inflammatory signaling and cell survival were also found enriched (Fig5f). Taken together, aberrant STAT5B expression promotes an activated CD4+ T cell state with pro-survival signature. This contrasts BACH2-driven CD4+ T cell reprogramming and suggests that insertional activation of RdIGs through LTR exaptation likely results in clones with different T cell states.

5. A negative feedback loop autoregulates endogenous *BACH2* expression

Our results indicate that an aberrant increase in *BACH2* expression leads to changes in CD4+ T cell state. This raises the question of how transcriptional activity of *BACH2* is regulated. We set out to investigate the *BACH2* promoter region to gain further understanding. In particular we were interested in a potential autoregulatory feedback loop that had been previously hypothesized (39). Motif scanning of a 40kb region around *BACH2* exon1 identified four T-MARE consensus motifs (TGCTGA(G/C)TCA; Fig6a), which are motifs that have been found to be targeted by BACH2 (77). To understand whether these motifs could have functional relevance, we assessed their epigenetic signature in CD4+ T cells. We found that two motifs in particular had an accessible chromatin profile, enriched for transcriptional regulatory histone marks, namely trimethylation of H3K4 and acetylation of H3K27. Furthermore, these two motifs have been found by chromatin immunoprecipitation to be bound by endogenous BACH2 in B cells (Fig6a). We therefore hypothesized that endogenous *BACH2* expression could be subject to autoregulation through promoter-inherent BACH2 binding sites.

To investigate whether BACH2, that has been primarily described as transcriptional repressor (78–80), could restrict its own expression, we took advantage of the two systems for aberrant *BACH2* expression developed in this study, namely HIV_BACH2$^{D5/E10}$ cells, harboring a *BACH2*-integrated HIV-1 reporter, and primary CD4+ T cells transduced with LV-BACH2. We assessed whether transcriptional activity from the endogenous *BACH2* locus could be affected by raised BACH2 protein levels. To do so, we quantified *BACH2* transcripts that originated from the non-coding region in exons 1 to 5 and reflect endogenous *BACH2* transcriptional activity, in presence or absence of LTR- or lentiviral-driven BACH2. Indeed, in HIV_BACH2$^{D5/E10}$ cells, raised levels of BACH2 through LTR exaptation (LTR-SAM) resulted in reduced transcript levels of *BACH2* exons 1-5 compared to controls (Fig6b). The effect was more pronounced in HIV_BACH2^{E10} cells with biallelic HIV reporter integration and thus stronger effects of LTR-exaptation. In agreement, primary CD4+ T cells that were transduced with a lentiviral LV-BACH2 demonstrated significantly reduced transcription of the endogenous *BACH2* locus (Fig6c,d). These data suggest that for endogenous BACH2, restriction of expression levels occurs through a negative autoregulatory feedback loop involving binding to the *BACH2* promoter region. This feedback loop could thus prohibit a continued raise of BACH2 levels, and prevent the proliferative CD4+ T cell phenotype that we observe upon overexpression. The prominent selection of cells harboring *BACH2* proviral integrants in PLWH on ART could thus be a result of LTR-driven aberrant *BACH2* expression that escapes endogenous regulatory mechanisms.

6. Aberrant *BACH2* expression through LTR exaptation can be controlled by HIV-1 latency promoting agents

Our data suggest that LTR exaptation can affect CD4+ T cell profiles, which may contribute to immune modulation and dysregulation observed in PLWH on ART. The immune modulatory activity is dependent on promoter activity of the integrated HIV LTR. To determine whether agents that interfere with HIV transcriptional activity can halt exaptation-dependent effects on the integration site locus, we tested how latency promoting agents (LPAs) impact on LTR-driven *BACH2* expression within our HIV_BACH2 cellular model. We focused on two previously reported LPAs, Spironolactone, an aldosterone receptor antagonist that has been shown to interfere with proviral transcription through degrading components of the TFIIH transcription factor complex (81, 82); and KRIBB11, a heat shock

factor 1-pTEFb complex inhibitor that has been shown to inhibit HIV-1 latency reversal (83). HIV_BACH2$^{D5/E10}$ cells were pre-treated with Spironolactone, Kribb11, or DMSO control and then stimulated with PMA/Ionomycin to induce transcription of the *BACH2* integrated proviral reporter. Samples were processed for flow cytometry to detect reporter expression and quantitative RT-PCR for LTR/BACH2 chimeric transcript detection. Both LPAs reduced expression of the HIV_BACH2 bfp reporter (Fig7a). In addition, both LPAs also significantly reduced induction of the chimeric LTR/BACH2 transcripts (Fig7b). Thus, treatment with LPAs is able to reduce LTR exaptation activity and offers a means to control dependent effects.

Discussion

Expanded clones of HIV-1 targeted CD4+ T cells that carry mainly defective proviruses are a prominent finding in PLWH on ART, their contribution to immune modulation in persistent HIV-1 infection is however unclear. In the present study we have investigated the most abundant clones so far detected in PLWH, that is CD4+ T cells with proviral integrations in the *BACH2* locus. These cells are selected during ART in a manner that suggests dependence on proviral/host gene transcriptional crosstalk. We report that *BACH2*-integrated proviruses drive aberrant *BACH2* expression through LTR exaptation, leading to induction of a proliferative precursor memory-like state that presumably persists through intrinsic anti-apoptotic activity and shielding against immune-mediated cell clearance. Interference with provirus transcriptional activity through LPA exposure can limit exaptation and thus provides a potential strategy to restrain BACH2 provirus clones.

Two hypotheses have so far been put forward for cells carrying *BACH2* integrated proviruses in PLWH on ART: First, *BACH2* proviral clones have been proposed to contribute to the replication-competent HIV-1 reservoir (38). However, since most datasets of proviral integrations do not correlate integration site with proviral sequence, it is unclear to date whether *BACH2* proviruses can contribute to persistence. To our knowledge, six studies have so far reported near full-length sequence information on a total of 14 BACH2 proviral integrants in PBMCs of PLWH on ART (12, 21, 41–44). In all but two reported cases, integration had occurred congruently, and in all cases, proviruses integrated in the *BACH2* locus carried sequence deletions of proviral genes that imply replication incompetence. The low number of *BACH2* integrants with sequence information preclude statistical analysis as for now, but there is currently no indication that *BACH2* clones contribute to the replication competent HIV-1 reservoir.

Second, it has been suggested that cells with *BACH2* proviruses differentiate towards an activated, T-reg like phenotype. This conclusion is based on an *in vitro* model with targeted insertion of a minimal LTR/MSD cassette into *BACH2* (39). In contrast, our data suggest that BACH2-provirus clones do not present effector type-signatures but adopt a precursor memory-like fate. Discrepancies could be due to differences in technical approaches (LV overexpression vs HDR manipulation) and time points of analysis (9 to 25 days post overexpression vs 7 weeks post HDR manipulation). In support of our data, BACH2 has previously been described to restrain differentiation towards CD4+ effector fates while promoting long-lived memory phenotypes (37, 56, 84, 85). Two recent reports indeed suggest that BACH2 acts as master regulator in directing CD4+ T cells towards a long-lived memory fate, supporting a cell population essential for HIV-1 persistence (86, 87). Our study implies that proviruses themselves, through dysregulation of BACH2, drive expansion of these long-lived cells.(88)Notably, we observe BACH2-induced transcriptomic changes to be gradual, with reprogramming of T cells taking place over several weeks in our experimental set-up. While future studies are required to further investigate, this may reflect that aberrant BACH2-expression indirectly leads to cell fate alteration. This is consistent with the finding, that BACH2 does not actively instruct expression of specific target genes, but rather

acts as transcriptional repressor to limit activity of super enhancer (SE) genomic regions, which regulate genes important for T-cell identity and lineage-specific cytokine responses (88–90).

Our findings suggest that provirus-driven BACH2 promotes expansion of a population of precursor-memory-like CD4+ T cells in PLWH. While these cells likely do not contribute directly to the replication-competent HIV reservoir, they could contribute indirectly to HIV persistence or infection-induced immune changes. In fact, we observe BACH2-reprogrammed cells to show upregulated expression of negative immune regulators, such as HLA-E. HLA-E has been shown to inhibit natural killer (NK) cell function through binding lectin-like receptor NKG2A (91, 92). NK cell activity is thought to be important in controlling HIV-1 infection and HLA-E expression positively correlates with HIV viremia in PLWH (93). It has also been suggested that prolonged exposure to elevated HLA-E levels could cause NK cell dysfunction and reduced viral control *in vivo (94)*. Thus, expansion of T cell clones with *BACH2* proviral integrants, could promote an immune environment in PLWH that is permissive to HIV persistence through loss of NK cell control. In this context, future studies focusing on proviral integration gene selection in lymphoid tissues of suppressed PLWH would be of particular interest. These sites harbor the majority of intact proviruses and thus are crucial for reservoir maintenance (95, 96). To our knowledge, comprehensive profiling of provirus integration landscapes from lymphoid tissues in PLWH on ART has not been carried out, this it currently remains unclear whether BACH2-proviral clones are as prevalent in this compartment as they are in peripheral blood.

Our study also highlights a further example of how defective proviral sequences can impact host function, strengthening the rationale to further investigate and target the defective HIV reservoir. In the case of LTR exaptation, as here described, proviruses are able to impact host cells as long as LTR-inherent regulatory elements and splice donor sites are functional. This is different from so far suggested immune modulatory activity of defective proviruses linked to viral gene transcripts or proteins (9–11, 97). Notably, exaptation of virus-derived regulatory elements is a well-known mechanism of how endogenized retroviruses have been coopted in physiology or pathology (98, 99). For HIV-1 infection, our data also indicate that LTR exaptation can lead to expansion of clones with different phenotypes depending on the integration gene. A number of integration genes have been proposed to be selected during ART (23–29), of which only BACH2 [33,34] and STAT5B (this study and (38, 39)) have been investigated in further detail in the context of provirus/host crosstalk. Given a recent report, which indicates that insertional activation of proviral integration genes is a phenomenon observed in up to 5% of infection events (100), future studies are warranted to investigate the extent to which LTR exaptation might alter immune landscape and function and play a role in viral persistence in PLWH.

Finally, our findings underscore the potential benefit of supplementing current antiretroviral regimens with components that block provirus transcription. Since growing evidence suggests that the majority of proviruses in PLWH remain transcriptionally active (101, 102), there is increasing interest in development and/or repurposing of FDA-approved drugs to limit LTR-driven transcription. As we show, these drugs also limit insertional activation and thus can interfere with downstream host cell modulation. Accordingly, we suggest that integration site mapping and clonality profiling of HIV reservoir cells would be two interesting parameters to assess in studies that trial block and lock agents in context of targeting non-AIDS related co-morbidities in PLWH with virological suppression.

In summary, our data highlight that proviruses – at the BACH2 locus and potentially also further RdIGs - have the ability the modulate the immune landscape of PLWH on ART through exaptation of virus-derived regulatory elements. We speculate that this could impact HIV-1 persistence and thus should be taken into account in view of optimizing current antiretroviral therapy regimens for healthy remission of PLWH.

Materials and Methods

Generation and analysis of provirus integration site datasets

Data on provirus integration sites was extracted from the Retrovirus Integration Database (https://rid.cancer.gov) (45). Datasets of integration sites were generated combining integration site data from 11 studies on in vivo integrations mapped in PBMC samples of PLWH on ART (12, 28, 29, 31, 32, 42, 46–50), and from 4 studies on in vitro integrations mapped in in vitro HIV-1 infected CD4+ T cells (24, 47, 51, 52). The IS were annotated using the 'closest' subcommand of BEDTools v2.27.1. For gene annotation, the RefSeq GRCh37 Release 105.20220307 reference was used, selecting only the "RefSeq Select" isoform for each protein-coding transcript. For visualizing of genome-wide integration site frequency, Circos plots were generated using Circos Builder (v069-815) of Galaxy Server, based on human hg19 karyotype ideogram (103). Bar plots depicting integration site distribution with regards to intronic position and provirus orientation were generated using Ggplot2. The integration site (IS) frequency for both in vivo and in vitro datasets was calculated for each gene using the formula: IS_freq = (IS count per gene / total IS count) × 100. To compare the two datasets, a scatter plot was generated where the x-axis represents individual genes of interest, and the y-axis displays the $\log_{10}$-transformed difference in IS frequency between in vivo and in vitro.

Genome-wide association study analysis

Genome-wide association study (GWAS) data including SNP location and trait information in context with *BACH2* was downloaded from https://www.ebi.ac.uk/gwas/genes/BACH2 (accessed May 2024). SNPs in *BACH2* were visualized in JBrowser2 (104). Trait information was categorized, counted and plotted using Microsoft Excel.

Cell culture

HAP1 cells (Horizon Discovery) were used for generation of HIV_BACH2 reporter lines and cultured in IMDM medium (Thermo Fisher Scientific) with 10% FBS and 1% Pen/Strep. Buffy coats for PBMC isolation were obtained from anonymized healthy donors through collaboration with the Institute of Transfusion Medicine, University Medical Center Hamburg-Eppendorf. PBMCs were isolated using density gradient centrifugation. CD4+ T cells were then isolated using the EasySep™ Human CD4+ T Cell Isolation Kit (stem cell technologies). CD4+ T cells were cultured in X-VIVO™ 15 cell medium (Lonza) with 10% FCS, 0,5% Pen/Strep and 200U/mL IL-2 at a density of $1x10^{6}$ cells/ml. Upon isolation, cells were activated with CD3/CD28 (ImmunoCult™ Human CD3/CD28 T Cell Activator, Stemcell™).

Generation of HIV_BACH2 reporter lines

An HIV-1 reporter construct was cloned in pMK backbone (Thermo Fischer Scientific) using sequential Gibson assembly (NEBuilder HiFi DNA assembly). 5'LTR, gag leader region and 3'LTR were derived from pNL4-3 (BEI Resources), bfp 2A sequence was amplified from pNLT2DenvBLB (105). To generate the transfer plasmid for CRISPR/Cas9-mediated knock-in, 5' and 3' homology arms of 1000bp length were amplified from genomic DNA extracted from Jurkat cell line and cloned flanking the HIV reporter. An 8bp long barcode sequence was introduced upstream the 5'LTR. Guide RNA for CRISPR/Cas9 targeting of the BACH2_3710 locus (5'-ggagattgtgtgaacccagc-3') was cloned into pX330-U6-Chimeric_BB-cBh-hSpCas9 (Addgene #42230) utilizing BbsI restriction sites (yielding pX330-U6-Chimeric_3710_BACH2-cBh-hSpCas9). E-CRISP tool was used for gRNA design (http://www.e-crisp.org).
CRIPSR/Cas9-mediated insertion of the HIV reporter into BACH2 intron 5 in HAP1 cells was carried out following a workflow previously described (106, 107). In brief, HAP1 cells were co-transfected using Lipofectamine 3000 (Thermo Fisher Scientific) with transfer plasmid and pX330-U6-Chimeric_3710_BACH2-cBh-hSpCas9 plasmid. Transfected cells were plated in limiting dilution for single cell clone generation at a density of 0.5-1 cells/well in flat bottom 96-well format. Screening for targeted clones was carried out in two rounds. First, bfp-expressing clones were selected after induction with 50 ng/ml phorbol 12-myristate 13-acetate (PMA; Sigma Aldrich) and 1 µM Ionomycin

(Sigma Aldrich) for 18 hours. Positive clones based on flow cytometry were subsequently screened by PCR on whole cell lysate for presence genomic insertion of the reporter at the targeted locus. To exclude cell clones with an integrated vector backbone, clones were screened by PCR for presence of pMK backbone (negative selection). Site-specific homozygous or heterozygous BACH2 targeting events were assessed by Southern blotting using bfp-specific and genomic probes as described previously (106). In brief, high molecular weight genomic DNA from 3710_BACH2_positive clones was extracted from whole cell lysate by isopropanol precipitation, digested with *ExoRI* and blotted onto Amersham Hybond^{TM}-N^{+} blotting membrane (GE Healthcare). The corresponding probes were PCR amplified from genomic DNA or targeting plasmid DNA and labelled with ^{32}dCTP using High Prime DNA labeling (Roche). Membranes were hybridized with labelled probes in hybridization buffer at 65°C over night. After stringent washing, membranes were exposed to Amersham Hyperfilm^{TM} MP (GE Healthcare).

PMA/Iono and SPA treatment of HIV_BACH2 cells

HIV_BACH2 cells were seeded at 1 x 10^6 cells/ml and treated with 50 ng/ml phorbol 12-myristate 13-acetate (PMA; Sigma Aldrich) and 1 µM Ionomycin (Sigma Aldrich) or DMSO as a control. BFP reporter expression was analysed by flow cytometry 16 hours post treatment. The effect of silencing promoting agents (SPAs) KRIBB11 and Spironolactone was assayed in HIV_BACH2 cells as follows: 1 x 10^6 cells/6-well were seeded and simultaneously treated with KRIBB11 (final concentration 5 uM, Absource Diagnostic) or Spironolactone (final concentration 10 uM, Sigma-Aldrich). 18 h post seeding, cells were induced with PMA/Iono. One day post seeding RNAzol samples were collected and BFP expression was analyzed by flow cytometry.

Western Blot analysis of HIV_BACH2 cells

HIV_BACH2 cells were activated with PMA and Ionomycin for 16h as described before. Cells were lysed in RIPA buffer (150 mM NaCl, 1 % Triton X-100, 0.25 % NP40, 0.1% SDS, 50 mM Tris HCl pH8.0 in H2O) supplemented with oComplete protease inhibitor cocktail (Millipore) and protein concentration determined by Pierce 660 reagent according to manufacturers instructions (Thermo Fisher). Twenty ug of total protein were separated by SDS-PAGE and transferred to nitrocellulose membranes (Whatman Protran). Primary antibodies against BACH2 (clone D3T3G, Cell Signaling) and LaminA/C as control (sc-376248, Santa Cruz) were used in combination with secondary HRP-conjugated antibodies (Jackson Laboratories) for detection.

SAM-mediated targeted transcriptional activation

SAM-mediated LTR (LTR-SAM) and BACH2 (BACH2-SAM) targeted transcriptional activation of HIV_BACH2 cells was carried out as previously described (30). In brief, 5*10^5 cells were seeded in a six-well plate and co-transfected with 1.4 µg sgRNA(MS2)-gRNA for either LTR or BACH2 activation, 0,8 µg dCas9-VP64_GFP (Addgene #61422) and 1 µg MS2-P65-HSF1_GFP (Addgene #61423) per well. Transfection was performed using TransIT LT1 reagent (Mirus Bio) according to the manufacturer's protocol. Three days post transfection cells were harvested, washed with PBS, sorted for BFP/GFP-positive cells (FACS Aria, BD Biosiences), washed once more and stored in RNAzol (Sigma Aldrich) at -80°C upon further processing. Total RNA was isolated following the RNAzol manufacturer's protocol. All experiments were conducted in n=3 independent replicates.

Production of lentivirus and transduction of CD4+ T cells

Lentiviruses were produced by plasmid co-transfection of the following plasmids into LentiX cells (Takara Bio) using TransIT-LT1 (Mirus Bio) according to manufacturer's instructions: psPAX2 (Addgene #12260), pCMV-VSV-G (Addgene #8454) and lentiviral transfer vectors. Supernatants were collected 72 hours post transfection and filtered through a 0.22 µm filter (Millipore). Virus concentration was conducted by ultracentrifugation at 12,300 × g for 1.5h at 4 °C (Beckman Coulter) and resuspended in X-VIVO 15 medium (Lonza). CD4+ T cells from healthy donors were activated by CD3/CD28 stimulation 24 h prior transduction. Spin-inoculation was used to transduce the cells with lentivirus stocks at a

MOI of 20 in X-VIVO 15 medium (Lonza) containing 0.2 % protamine sulfate (Sigma) (850 x g, 1 h, 30 °C). Following centrifugation, the plates were incubated for 24 h at 37 °C before medium was replaced by X-VIVO 15 cell with 10 % FBS, 0.5 % Pen/Strep and 200 U/mL IL-2 (Biomol).

Magnetic bead-mediated enrichment for SBP-LNGFR + cells

$5x10^6$ transduced CD4+ T cells were enriched using magnetic selection of the SBP-ΔLNGFR surface marker with Dynabeads Biotin Binder (Invitrogen). In brief, cells were washed three times with incubation buffer (0.5 M EDTA, 1 mg/L BSA in PBS). Beads were added at a 10:1 ratio (beads:cells) and the mixture was incubated at 4 °C for 30 min. After incubation, the samples were placed on a magnetic stand for 3 minutes before carefully removing the supernatant. Bead-bound cells were washed with cold incubation buffer, placed into the magnetic stand again for 3 minutes before decanting the supernatant. Bead-bound cells were resuspended in pre-warmed release buffer (1 M HEPES, 20 mM biotin in medium) and incubated at room temperature for 15 min. The tube was placed on the magnetic stand for 3 minutes before collecting the released cells in the supernatant.

Analysis of cell proliferation

Division index of transduced cells was measured using CellTrace™ Far Red Cell Proliferation Kit (Thermo Fisher Scientific) according to the manufacturer's instructions. In brief, transduced cell populations and non-transduced control cells (n=3) were stained at d12 after transduction. Expression of the Far Red marker dye was measured at d12 and at three following time points (d14, d15 and d17 after transduction) by flow cytometry.

Quantitative RT PCR assays

RNA was extracted using RNAzol (Sigma Aldrich) according to manufacturer's instructions. cDNA was generated from 1ug total RNA using SuperScript III Reverse Transcriptase (Invitrogen). Quantification of endogenous *BACH2* (amplicon spanning exon 4/5, located upstream of HIV reporter integration site in HAP1 cell lines) and total BACH2 (exon 6, downstream of HIV reporter integration site in HAP1 cell lines) transcript levels was measured using droplet digital PCR with the Bio-Rad QX200 system (utilized oligonucleotides are provided in SupplTable2). Each 20 µL reaction contained 10 µL of 2× ddPCR Supermix for Probes (No dUTP) (Bio-Rad), 2 µL of each primer (10 µM stock), 1 µL probe (5 µM stock), 2 µL of cDNA template, 1µL of RPL13A primer/probe mastermix for normalization (Bio-Rad #10031255) and 2 µL nuclease-free water. The reaction mix was loaded into the QX200TM Droplet Generator (Bio-Rad) with 70 µL of Droplet Generation Oil, and droplets were generated according to the manufacturer's guidelines. PCR amplification was performed using a T100TM Thermal Cycler (Bio-Rad) under the following cycling conditions: 95 °C for 10 min, 40 cycles of 94 °C for 30 sec and 57 °C for 1 min, 98 °C for 10 min, hold at 4 °C. Following amplification, droplets were analyzed using the QX200 Droplet Reader (Bio-Rad). Resulting data was visualized using GraphPad Prism v10 software.

BACH2 binding site and chromatin landscape analysis

The datasets supporting the conclusions of this analysis are publicly available and can be accessed in NCBI repositories under the accession numbers GSE102460 [BACH2 ChIP-seq], GSE122826 [H3 ChIP-seq], and GSE144329 [PLWH-on-ART ATAC-seq]. To identify putative BACH2 binding sites across the human genome, the BACH2 motif was obtained from the JASPAR database (2024 expansion, ID: MA1101) (108). The hg38 human genome assembly was scanned for BACH2 motifs using FIMO, a tool within the MEME suite (v5.5.5) (109). FIMO was run using the MEME suite's precomputed hg38 background model based on the genome's nucleotide frequencies. Significant BACH2 motif hits were defined as those with $p < 1e-4$. Raw ChIP-seq and ATAC-seq data were first assessed for quality using FastQC (v0.12.1) (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Subsequently, reads were trimmed with Trimmomatic (v0.39) removing adapter sequences and low-quality bases (110). Next, trimmed reads were aligned to the hg38 human reference genome using Bowtie2 (v2.5.1) with the --very-sensitive-local parameter for optimal sensitivity (111). All unmapped reads and reads with

a mapping quality score below 10 were excluded from further analysis. PCR duplicates were removed using Picard MarkDuplicates (v2.26.10) (http://broadinstitute.github.io/picard). For peak calling, MACS2 (v2.2.7.1) was used with parameters tailored for each assay — broad peak calling was performed for H3K27ac ChIP to capture large acetylation domains, while narrow peak calling was used for BACH2 ChIP, H3K4me3 ChIP, and ATAC-seq open chromatin regions (112). Peak annotation to hg38 genes was carried out using Homer (v4.11) (https://homer.ucsd.edu/). Sample integration and signal normalization was performed using BEDTools (v2.29.2) and deepTools (v3.5.5) (113, 114). Tracks were imported in bigWig format and visualized in JBrowse2 (v2.18.0) (115).

RNA sequencing and RNAseq dataset analysis

RNA sequencing was carried out on total RNA, isolated from SAM-treated HAP1_BACH2_LTRbfp2Atat cells and LV-transduced CD4+ T lymphocytes. For HAP1 samples, 1 ug of RNA was subjected to library preparation utilizing the NEBNext Ultra II Directional RNA mRNA kit in combination with NEBNext Multiplex Oligos Unique Dual Index UMI Adaptors and sequenced at ~100M clusters 2x100bp per sample (Illumina NextSeq500). For primary T lymphocytes, about 5 mio transduced CD4+ T cells were enriched using magnetic selection based on SBP-ΔLNGFR surface marker expression (see above). Total RNA was isolated from selected primary CD4 T cells using RNAzol (Sigma Aldrich) and following manufacturer's instructions. Isolated RNA was subjected to library preparation using CORALL mRNA-Seq Library Prep Kit (Lexogen) and subsequent sequenced at ~20M reads, 2x50bp per sample (Illumina NextSeq500).

Sequence reads were mapped using HISAT2 (HAP1 RNAseq) or Salmon (CD4 T lymphocyte RNAseq) aligners against the human reference genome (hg38) after adapter trimming and quality filtering using TrimGalore (116–118). Gene count matrices were further analysed in R in terms of differential gene expression (DEG) and gene set enrichment analysis (GSEA) utilizing packages DESeq2 (v1.44.0), EnhancedVolcano (v1.22.0), pheatmap (v1.0.12), ClusterProfiler (v4.12.6) and ggplot2 (119–123). In CD4+ T lymphocyte over expression experiments contrasting LV-GFP vs LV-BACH2 samples, 'donor_ID' was considered as covariate. In PCA plots this was achieved utilizing DESeq2 vst transformation of gene counts and limma::removeBatchEffect() (v3.60.3) for 'donor_ID' (121, 124). Treeplots of significant gene ontology (GO) terms were generated by hierarchical clustering based on Jaccard's similarity index using ClusterProfiler v(4.12.6) (122). Analysis of enrichment of functional gene networks using MCL clustering (inflation parameter 3) was done using STRING (v12.0) (125).

Single BACH2 exon analysis in the HAP1 reporters was conducted with a filtered genome annotation file (GTF) for BACH2 reference transcript ENST00000257749 containing 9 exons. HTSeq function htseq_count (v2.0.5) with parameters -m union -i exon_number -t exon was used to get read counts for individual BACH2 exons (126). Downstream analysis and plotting in R utilizing packages ggplot2 (v3.5.1) and ggpubr (v0.6.0) (123, 127). Statistical method: unpaired t-test.

HIV-to-host transcript splicing in HAP1 reporters was analysed by inserting the HIV reporter sequence into the BACH2 locus (position chr6:90012553-90015110) of the reference genome version GRCh38. Genome annotations (GTF) were accordingly updated using reform (128). Subsequently raw FASTQ files were aligned to the updated reference genome and annotation using STAR (129). Splicing pattern plots were generated using the modified GTF annotation file and the sashimi_plot command of miso with scaling parameters for intron and exon at 50 and 3 respectively (130, 131).

Data sharing plan

Materials availability: All new unique reagents generated in this study are available from the corresponding author with a completed materials transfer agreement.

Datasets availability: The raw RNA-Seq data generated as part of this study have been uploaded in the GEO repository (accession #GSE304581) and are available upon request. The remaining data are available within the article or Supplementary Information. Any additional information required to

reanalyze the data reported in this study is available from the corresponding authors upon request. This study does not report original code.

References

1. M. Lichterfeld, C. Gao, X. G. Yu, An ordeal that does not heal: Understanding barriers to a cure for HIV-1 infection. *Trends Immunol* **43**, 608 (2022).
2. L. B. Cohn, N. Chomont, S. G. Deeks, C. Z. Biohub, S. Francisco, Cell Host & Microbe The Biology of the HIV-1 Latent Reservoir and Implications for Cure Strategies. *Cell Host Microbe* **27**, 519–530 (2020).
3. S. Zicari, *et al.*, Immune Activation, Inflammation, and Non-AIDS Co-Morbidities in HIV-Infected Patients under Long-Term ART. *Viruses* **11** (2019).
4. K. M. Bruner, *et al.*, Defective proviruses rapidly accumulate during acute HIV-1 infection. *Nat Med* **22**, 1043–9 (2016).
5. Y.-C. Ho, *et al.*, Replication-competent noninduced proviruses in the latent reservoir increase barrier to HIV-1 cure. *Cell* **155**, 540–551 (2013).
6. J. C. Botha, *et al.*, The largest HIV-1-infected T cell clones in children on long-term combination antiretroviral therapy contain solo LTRs. *mBio* e0111623 (2023). https://doi.org/10.1128/mbio.01116-23.
7. J. A. White, *et al.*, Clonally expanded HIV-1 proviruses with 5'-Leader defects can give rise to nonsuppressible residual viremia. *J Clin Invest* (2023). https://doi.org/10.1172/JCI165245.
8. J. Kuniholm, C. Coote, A. J. Henderson, Defective HIV-1 genomes and their potential impact on HIV pathogenesis. *Retrovirology 2022 19:1* **19**, 1–12 (2022).
9. R. A. Pollack, *et al.*, Defective HIV-1 Proviruses Are Expressed and Can Be Recognized by Cytotoxic T Lymphocytes, which Shape the Proviral Landscape. *Cell Host Microbe* **21**, 494-506.e4 (2017).
10. H. Imamichi, *et al.*, Defective HIV-1 proviruses produce viral proteins. *Proceedings of the National Academy of Sciences* 201917876 (2020). https://doi.org/10.1073/pnas.1917876117.
11. H. Imamichi, *et al.*, Defective HIV-1 proviruses produce novel protein-coding RNA species in HIV-infected patients on combination antiretroviral therapy. *Proceedings of the National Academy of Sciences* **113**, 8783–8788 (2016).
12. K. B. Einkauf, *et al.*, Parallel analysis of transcription, integration, and sequence of single HIV-1 proviruses. *Cell* (2022). https://doi.org/10.1016/j.cell.2021.12.011.
13. K. Singh, *et al.*, Long-term persistence of transcriptionally active "defective" HIV-1 proviruses: implications for persistent immune activation during antiretroviral therapy. *AIDS* **37**, 2119–2130 (2023).
14. N. V. E. J. Buchholtz, *et al.*, Defective proviruses significantly impact viral transcription and immune activation in men and women with HIV-1 subtype C in rural South Africa. *Front Immunol* **15**, 1484358 (2024).
15. F. Scrimieri, *et al.*, Transcriptionally Active Defective HIV-1 Proviruses and Their Association With Immunological Nonresponse to Antiretroviral Therapy. *J Infect Dis* **229**, 1786–1790 (2024).
16. L. B. Cohn, *et al.*, HIV-1 Integration Landscape during Latent and Active Infection. *Cell* **160**, 420–432 (2015).
17. C.-Y. Lau, M. A. Adan, F. Maldarelli, Why the HIV Reservoir Never Runs Dry: Clonal Expansion and the Characteristics of HIV-Infected Cells Challenge Strategies to Cure and Control HIV Infection. *Viruses* **13** (2021).
18. Y. H. J. Yeh, K. Yang, A. Razmi, Y. C. Ho, The clonal expansion dynamics of the hiv-1 reservoir: Mechanisms of integration site-dependent proliferation and hiv-1 persistence. *Viruses* **13**, 1–19 (2021).
19. M. R. Pinzone, *et al.*, Longitudinal HIV sequencing reveals reservoir expression leading to decay which is obscured by clonal expansion. *Nat Commun* **10**, 728 (2019).
20. J. I. Mullins, L. M. Frenkel, Clonal expansion of human immunodeficiency virus-infected cells and human immunodeficiency virus persistence during antiretroviral therapy. *Journal of Infectious Diseases* **215**, S119–S127 (2017).
21. F. R. Simonetti, *et al.*, Antigen-driven clonal selection shapes the persistence of HIV-1-infected CD4+T cells in vivo. *Journal of Clinical Investigation* **131** (2021).
22. R. Liu, F. R. Simonetti, Y. C. Ho, The forces driving clonal expansion of the HIV-1 latent reservoir. *Virol J* [Preprint] (2020).
23. S. H. Hughes, J. M. Coffin, What Integration Sites Tell Us about HIV Persistence. *Cell Host Microbe* **19**, 588–598 (2016).
24. J. M. Coffin, *et al.*, Integration in oncogenes plays only a minor role in determining the in vivo distribution of HIV integration sites before or during suppressive antiretroviral therapy. *PLoS Pathog* **17**, e1009141 (2021).

25. B. Lucic, *et al.*, Spatially clustered loci with multiple enhancers are frequent targets of HIV-1 integration. *Nat Commun* **10**, 4059 (2019).
26. H. AS, *et al.*, Integration features of intact latent HIV-1 in CD4+ T cell clones contribute to viral persistence. *J Exp Med* **218** (2021).
27. G. J. Bedwell, S. Jang, W. Li, P. K. Singh, A. N. Engelman, rigrag : high-resolution mapping of genic targeting preferences during HIV-1 integration in vitro and in vivo. **49**, 7330–7346 (2021).
28. T. A. Wagner, *et al.*, HIV latency. Proliferation of cells with HIV integrated into cancer genes contributes to persistent infection. *Science* **345**, 570–3 (2014).
29. F. Maldarelli, *et al.*, Specific HIV integration sites are linked to clonal expansion and persistence of infected cells. *Science* **345**, 179–83 (2014).
30. M. V Hamann, *et al.*, Transcriptional behavior of the HIV-1 promoter in context of the BACH2 prominent proviral integration gene. *Virus Res* **293**, 198260 (2021).
31. T. Ikeda, J. Shibata, K. Yoshimura, A. Koito, S. Matsushita, Recurrent HIV-1 integration at the BACH2 locus in resting CD4+ T cell populations during effective highly active antiretroviral therapy. *J Infect Dis* **195**, 716–25 (2007).
32. K. D. Mack, *et al.*, 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**, 308–20 (2003).
33. T. Sidwell, A. Kallies, Bach2 is required for B cell and T cell memory differentiation. *Nat Immunol* **17**, 744–5 (2016).
34. Y. Zhou, H. Wu, M. Zhao, C. Chang, Q. Lu, The Bach Family of Transcription Factors: A Comprehensive Review. *Clin Rev Allergy Immunol* **50**, 345–56 (2016).
35. M. J. Richer, M. L. Lang, N. S. Butler, T Cell Fates Zipped Up: How the Bach2 Basic Leucine Zipper Transcriptional Repressor Directs T Cell Differentiation and Function. *J Immunol* **197**, 1009–15 (2016).
36. F. M. Grant, *et al.*, BACH2 drives quiescence and maintenance of resting Treg cells to promote homeostasis and cancer immunosuppression. *J Exp Med* **217** (2020).
37. R. Roychoudhuri, *et al.*, Bach2 represses effector programmes to stabilize Treg-mediated immune homeostasis. *Nature* **498**, 506 (2013).
38. D. Cesana, *et al.*, HIV-1-mediated insertional activation of STAT5B and BACH2 trigger viral reservoir in T regulatory cells. *Nat Commun* **8**, 498 (2017).
39. M. L. Christian, *et al.*, CRISPR/Cas9-Mediated Insertion of HIV Long Terminal Repeat within BACH2 Promotes Expansion of T Regulatory-like Cells. *J Immunol* **208**, 1700–1710 (2022).
40. R. Liu, *et al.*, Single-cell transcriptional landscapes reveal HIV-1-driven aberrant host gene transcription as a potential therapeutic target. *Sci Transl Med* **12** (2020).
41. K. W. Joseph, *et al.*, Deep Sequencing Analysis of Individual HIV-1 Proviruses Reveals Frequent Asymmetric Long Terminal Repeats. *J Virol* **96** (2022).
42. K. B. Einkauf, X. G. Yu, M. Lichterfeld, Intact HIV-1 proviruses accumulate at distinct chromosomal positions during prolonged antiretroviral therapy. *Journal of Clinical Investigation* (2019). https://doi.org/10.1172/JCI124291.
43. S. C. Patro, *et al.*, Combined HIV-1 sequence and integration site analysis informs viral dynamics and allows reconstruction of replicating viral ancestors. *Proc Natl Acad Sci U S A* **116**, 25891–25899 (2019).
44. W. Sun, *et al.*, Phenotypic signatures of immune selection in HIV-1 reservoir cells. *Nature 2023 614:7947* **614**, 309–317 (2023).
45. W. Shao, *et al.*, Retrovirus Integration Database (RID): a public database for retroviral insertion sites into host genomes. *Retrovirology* **13**, 47 (2016).
46. Y. Han, *et al.*, Resting CD4+ T cells from human immunodeficiency virus type 1 (HIV-1)-infected individuals carry integrated HIV-1 genomes within actively transcribed host genes. *J Virol* **78**, 6122–6133 (2004).
47. H. Ode, *et al.*, Identifying integration sites of the HIV-1 genome with intact and aberrant ends through deep sequencing. *J Virol Methods* **267**, 59–65 (2019).
48. J. M. Coffin, *et al.*, Clones of infected cells arise early in HIV-infected individuals. *JCI Insight* **4**, 0–15 (2019).
49. W. R. McManus, *et al.*, HIV-1 in lymph nodes is maintained by cellular proliferation during antiretroviral therapy. *Journal of Clinical Investigation* **130** (2019).
50. H. Katsuya, *et al.*, Clonality of HIV-1- and HTLV-1-Infected Cells in Naturally Coinfected Individuals. *J Infect Dis* **225**, 317–326 (2022).
51. S. Sherrill-Mix, *et al.*, HIV latency and integration site placement in five cell-based models. *Retrovirology* **10**, 90 (2013).

52. A. L. Ferris, *et al.*, Clonal expansion of SIV-infected cells in macaques on antiretroviral therapy is similar to that of HIV-infected cells in humans. *PLoS Pathog* **15**, e1007869 (2019).
53. S. Konermann, *et al.*, Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517**, 583–588 (2015).
54. J. K. Bialek, *et al.*, Targeted HIV-1 Latency Reversal Using CRISPR/Cas9-Derived Transcriptional Activator Systems. *PLoS One* **11**, e0158294 (2016).
55. B. Afzali, *et al.*, BACH2 immunodeficiency illustrates an association between super-enhancers and haploinsufficiency. *Nat Immunol* **18**, 813–823 (2017).
56. L. Yang, S. Chen, Q. Zhao, Y. Sun, H. Nie, The Critical Role of Bach2 in Shaping the Balance between CD4+ T Cell Subsets in Immune-Mediated Diseases. *Mediators Inflamm* **2019** (2019).
57. J. Carnevale, *et al.*, RASA2 ablation in T cells boosts antigen sensitivity and long-term function. *Nature* **609**, 174 (2022).
58. A. Argyriou, *et al.*, Single cell sequencing identifies clonally expanded synovial CD4+ TPH cells expressing GPR56 in rheumatoid arthritis. *Nature Communications 2022 13:1* **13**, 1–13 (2022).
59. M. Biburger, G. Theiner, M. Schädle, G. Schuler, G. Tiegs, Pivotal Advance: Heme oxygenase 1 expression by human CD4+ T cells is not sufficient for their development of immunoregulatory capacity. *J Leukoc Biol* **87**, 193–202 (2009).
60. Q. Bie, C. Jin, B. Zhang, H. Dong, IL-17B: A new area of study in the IL-17 family. *Mol Immunol* **90**, 50–56 (2017).
61. J. A. Bibby, *et al.*, Systematic single-cell pathway analysis to characterize early T cell activation. *Cell Rep* **41**, 111697 (2022).
62. H. Liu, *et al.*, Arachidonic acid activates NLRP3 inflammasome in MDSCs via FATP2 to promote post-transplant tumour recurrence in steatotic liver grafts. *JHEP Rep* **5** (2023).
63. S. Liu, S. Liao, L. Liang, J. Deng, Y. Zhou, The relationship between CD4+ T cell glycolysis and their functions. *Trends in Endocrinology & Metabolism* **34**, 345–360 (2023).
64. H. Tanaka, *et al.*, Epigenetic Regulation of the Blimp-1 Gene (Prdm1) in B Cells Involves Bach2 and Histone Deacetylase 3. *J Biol Chem* **291**, 6316–6330 (2016).
65. T. Yoshikawa, *et al.*, Genetic ablation of PRDM1 in antitumor T cells enhances therapeutic efficacy of adoptive immunotherapy. *Blood* **139**, 2156–2172 (2022).
66. M. Mammadli, L. Suo, J. M. Sen, M. Karimi, TCF-1 Is Required for CD4 T Cell Persistence Functions during AlloImmunity. *Int J Mol Sci* **24** (2023).
67. X. Zhou, H.-H. Xue, Cutting Edge: Generation of Memory Precursors and Functional Memory CD8+ T Cells Depends on T Cell Factor-1 and Lymphoid Enhancer-Binding Factor-1. *The Journal of Immunology* **189**, 2722–2726 (2012).
68. I. C. Clark, *et al.*, HIV silencing and cell survival signatures in infected T cell reservoirs. *Nature 2023 614:7947* **614**, 318–325 (2023).
69. H. H. Kuo, *et al.*, Anti-apoptotic protein BIRC5 maintains survival of HIV-1-infected CD4+ T cells. *Immunity* **48**, 1183 (2018).
70. J. A. Collora, *et al.*, Single-cell multiomics reveals persistence of HIV-1 in expanded cytotoxic T cell clones. *Immunity* (2022). https://doi.org/10.1016/j.immuni.2022.03.004.
71. Y. L. Kok, *et al.*, HIV-1 integration sites in CD4+ T cells during primary, chronic, and late presentation of HIV-1 infection. *JCI Insight* **6**, e143940 (2021).
72. J. A. Jenks, *et al.*, Differentiating the roles of STAT5B and STAT5A in human CD4+ T cells. *Clin Immunol* **148**, 227–36 (2013).
73. D. M. Jones, K. A. Read, K. J. Oestreich, Dynamic Roles for IL-2-STAT5 Signaling in Effector and Regulatory CD4+ T Cell Populations. *J Immunol* **205**, 1721–1730 (2020).
74. A. V. Villarino, Transcriptional programing of T cell metabolism by STAT family transcription factors. *Eur J Immunol* **53** (2023).
75. D. K. Taylor, *et al.*, Constitutive Activation of STAT5 Supersedes the Requirement for Cytokine and TCR Engagement of CD4+ T Cells in Steady-State Homeostasis. *The Journal of Immunology* **177**, 2216–2223 (2006).
76. M. A. Farrar, D. L. Owen, STAT5 and CD4 + T Cell Immunity. *F1000Res* **6**, 32 (2017).
77. Q. Hu, *et al.*, Diverging regulation of Bach2 protein and RNA expression determine cell fate in early B cell response. *Cell Rep* **40** (2022).
78. T. Oyake, *et al.*, Bach Proteins Belong to a Novel Family of BTB-Basic Leucine Zipper Transcription Factors That Interact with MafK and Regulate Transcription through the NF-E2 Site. *Mol Cell Biol* **16**, 6083–6095 (1996).

79. R. Roychoudhuri, *et al.*, BACH2 regulates CD8+ T cell differentiation by controlling access of AP-1 factors to enhancers. *Nature Immunology 2016 17:7* **17**, 851–860 (2016).
80. Y. Zhou, H. Wu, M. Zhao, C. Chang, Q. Lu, The Bach Family of Transcription Factors: A Comprehensive Review. *Clin Rev Allergy Immunol* **50**, 345–56 (2016).
81. L. Mori, *et al.*, The XPB Subunit of the TFIIH Complex Plays a Critical Role in HIV-1 Transcription and XPB Inhibition by Spironolactone Prevents HIV-1 Reactivation from Latency. *J Virol* **95** (2021).
82. B. Lacombe, M. Morel, F. Margottin-Goguet, B. C. Ramirez, Specific Inhibition of HIV Infection by the Action of Spironolactone in T Cells. *J Virol* **90**, 10972–10980 (2016).
83. X.-Y. Pan, *et al.*, Heat Shock Factor 1 Mediates Latent HIV Reactivation. *Sci Rep* **6**, 26294 (2016).
84. R. Roychoudhuri, *et al.*, BACH2 regulates CD8(+) T cell differentiation by controlling access of AP-1 factors to enhancers. *Nat Immunol* **17**, 851–860 (2016).
85. C. Yao, *et al.*, BACH2 enforces the transcriptional and epigenetic programs of stem-like CD8+ T cells. *Nat Immunol* **22**, 370–380 (2021).
86. H. Gao, *et al.*, BACH2 promotes seeding and establishment of long-lived HIV-1 reservoir in memory CD4+ T cells. *Cell Rep Med* 102311 (2025). https://doi.org/10.1016/J.XCRM.2025.102311.
87. Y. Wei, *et al.*, Transcription factor BACH2 shapes tissue-resident memory T cell programs to promote HIV-1 persistence. *Immunity* (2025). https://doi.org/10.1016/J.IMMUNI.2025.07.022.
88. Y. Wei, *et al.*, BACH2-driven tissue resident memory programs promote HIV-1 persistence. *bioRxiv* (2024). https://doi.org/10.1101/2024.12.16.628794.
89. S. Witte, J. J. O'Shea, G. Vahedi, Super-enhancers: Asset management in immune cell genomes. *Trends Immunol* **36**, 519–526 (2015).
90. G. Vahedi, *et al.*, Super-enhancers delineate disease-associated regulatory nodes in T cells. *Nature 2015 520:7548* **520**, 558–562 (2015).
91. M. J. Richer, M. L. Lang, N. S. Butler, T Cell Fates Zipped Up: How the Bach2 Basic Leucine Zipper Transcriptional Repressor Directs T Cell Differentiation and Function. *J Immunol* **197**, 1009–15 (2016).
92. V. M. Braud, *et al.*, HLA-E binds to natural killer cell receptors CD94/NKG2A, B and C. *Nature* **391**, 795–799 (1998).
93. S. Sivori, *et al.*, Human NK cells: surface receptors, inhibitory checkpoints, and translational applications. *Cell Mol Immunol* **16**, 430–441 (2019).
94. L. Romero-Martín, *et al.*, Disruption of the HLA-E/NKG2X axis is associated with uncontrolled HIV infections. *Front Immunol* **13**, 1027855 (2022).
95. J. Nattermann, *et al.*, HIV-1 infection leads to increased HLA-E expression resulting in impaired function of natural killer cells. *Antivir Ther* **10**, 95–107 (2005).
96. C. Dufour, *et al.*, Near full-length HIV sequencing in multiple tissues collected postmortem reveals shared clonal expansions across distinct reservoirs during ART. *Cell Rep* **42** (2023).
97. R. Lorenzo-Redondo, *et al.*, Persistent HIV-1 replication maintains the tissue reservoir during therapy. *Nature* (2016). https://doi.org/10.1038/nature16933.
98. F. Scrimieri, *et al.*, Transcriptionally Active Defective HIV-1 Proviruses and Their Association With Immunological Nonresponse to Antiretroviral Therapy. *J Infect Dis* **229**, 1786–1790 (2024).
99. A. Babaian, D. L. Mager, Endogenous retroviral promoter exaptation in human cancer. *Mob DNA* **7**, 24 (2016).
100. G. Bourque, *et al.*, Ten things you should know about transposable elements. *Genome Biol* **19**, 199 (2018).
101. Y. Ma, J. J. Peterson, D. M. Margolis, E. P. Browne, SINGLE CELL ANALYSIS OF HIV EXPRESSION AND INTEGRATION SITES REVEALS VIRAL EXPRESSION ACROSS DIVERSE CHROMATIN ENVIRONMENTS. in *Retroviruses 2025, CSHL*, (2025).
102. S. A. Yukl, *et al.*, HIV latency in isolated patient CD4 + T cells may be due to blocks in HIV transcriptional elongation, completion, and splicing. *Sci Transl Med* **10** (2018).
103. J. Prigann, *et al.*, Silence the transcriptionally active HIV reservoir to improve treatment outcomes. *Nat Microbiol* **9**, 2470 (2024).
104. M. Krzywinski, *et al.*, Circos: An information aesthetic for comparative genomics. *Genome Res* **19**, 1639–1645 (2009).
105. C. Diesh, *et al.*, JBrowse 2: a modular genome browser with views of synteny and structural variation. *Genome Biology 2023 24:1* **24**, 1–21 (2023).
106. I. Hauber, *et al.*, Highly significant antiviral activity of HIV-1 LTR-specific tre-recombinase in humanized mice. *PLoS Pathog* **9**, e1003587 (2013).

107. U. C. Lange, J. K. Bialek, T. Walther, J. Hauber, Pinpointing recurrent proviral integration sites in new models for latent HIV-1 infection. *Virus Res* doi: 10.1016/j.virusres.2018.03.007 (2018). https://doi.org/10.1016/j.virusres.2018.03.007.
108. J. K. Bialek, T. Walther, J. Hauber, U. C. Lange, CRISPR-Cas9-based Genome Engineering to Generate Jurkat Reporter Models for HIV-1 Infection with Selected Proviral Integration Sites. *Journal of Visualized Experiments* (2018). https://doi.org/10.3791/58572.
109. I. Rauluseviciute, *et al.*, JASPAR 2024: 20th anniversary of the open-access database of transcription factor binding profiles. *Nucleic Acids Res* **52**, D174–D182 (2024).
110. C. E. Grant, T. L. Bailey, W. S. Noble, FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011).
111. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
112. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357–359 (2012).
113. J. M. Gaspar, Improved peak-calling with MACS2. *bioRxiv* 496521 (2018). https://doi.org/10.1101/496521.
114. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841 (2010).
115. F. Ramírez, *et al.*, deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**, W160–W165 (2016).
116. C. Diesh, *et al.*, JBrowse 2: a modular genome browser with views of synteny and structural variation. *Genome Biol* **24** (2023).
117. D. Kim, J. M. Paggi, C. Park, C. Bennett, S. L. Salzberg, Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology 2019 37:8* **37**, 907–915 (2019).
118. R. Patro, G. Duggal, M. I. Love, R. A. Irizarry, C. Kingsford, Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* **14**, 417–419 (2017).
119. F. Krüger, A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files. [Preprint] (2015).
120. M. Blighe, Kevin; Rana, Sharmila; Lewis, EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling. (2024).
121. R. Kolde, pheatmap: Pretty Heatmaps. (2018).
122. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology 2014 15:12* **15**, 1–21 (2014).
123. G. Yu, L. G. Wang, Y. Han, Q. Y. He, ClusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).
124. H. Wickham, ggplot2. (2016). https://doi.org/10.1007/978-3-319-24277-4.
125. M. E. Ritchie, *et al.*, limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47–e47 (2015).
126. D. Szklarczyk, *et al.*, The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res* **51**, D638–D646 (2023).
127. S. Anders, P. T. Pyl, W. Huber, HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
128. A. Kassambara, ggpubr: “ggplot2” Based Publication Ready Plots. [Preprint] (2023).
129. M. Khalfan, E. Borenstein, P. Spealman, F. Abdul-Rahman, D. Gresham, Modifying Reference Sequence and Annotation Files Quickly and Reproducibly with ref orm. (2021).
130. A. Dobin, *et al.*, STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
131. Y. Katz, E. T. Wang, E. M. Airoldi, C. B. Burge, Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat Methods* **7**, 1009–1015 (2010).
132. Y. Katz, *et al.*, Quantitative visualization of alternative exon expression from RNA-seq data. *Bioinformatics* **31**, 2400–2402 (2015).

Acknowledgements

We would like to acknowledge technical assistance in the project from Christina Herrde, Susanne Hoibian and Manuela Kolster as well as the flow cytometry and sequencing core facilities at the Leibniz Institute of Virology (LIV). This work was supported by funds from the German Federal Ministry of Research, Technology and Space award number 01KI2105 (U.C.L.), NIAID award number UM1AI164559, co-funded by NHLBI, NIDA, NIMH, NINDS, and NIDDK (U.C.L., C.F., N.R.R.) and the German Center of Infection Research (DZIF) award number TI07.005 (U.C.L.).

Author information

U.C.L. conceived and designed the project. C.F., N.R., E.T. and U.C.L. secured funding. C.S., M.V. and R.S. carried out cell line experiments. L.B. and M.K. carried out primary cell experiments. Computational data analysis was carried out by: M.S. and M.A. (integration site database), S.M.L. (promoter characterization), M.V.H., J.F., S.V. and A.G. (RNAseq). M.V.H., L.B. and U.C.L. interpreted the data and discussed the results with all authors. U.C.L. wrote the manuscript, in consultation with M.V.H., C.F., N.R.R. and E.T. with comments from all authors. M.V.H. and L.B. contributed equally to this work.

Competing interests

The authors declare no competing interests.

Materials and Correspondence

Requests for materials or correspondence should be sent to Ulrike C Lange (ulrike.lange@leibniz-liv.de).

Supplementary Information

Supplemental Table 1

Differentially expressed genes (DEGs) from RNAseq assays of HIV_BACH2 SAM-treated cell lines and LV-BACH2/LV-STAT5B-treated CD4+ T cells (in comparison controls).

Supplemental Table 2

Oligonucleotides used in this study.

FIGURE 1
Hamann et al

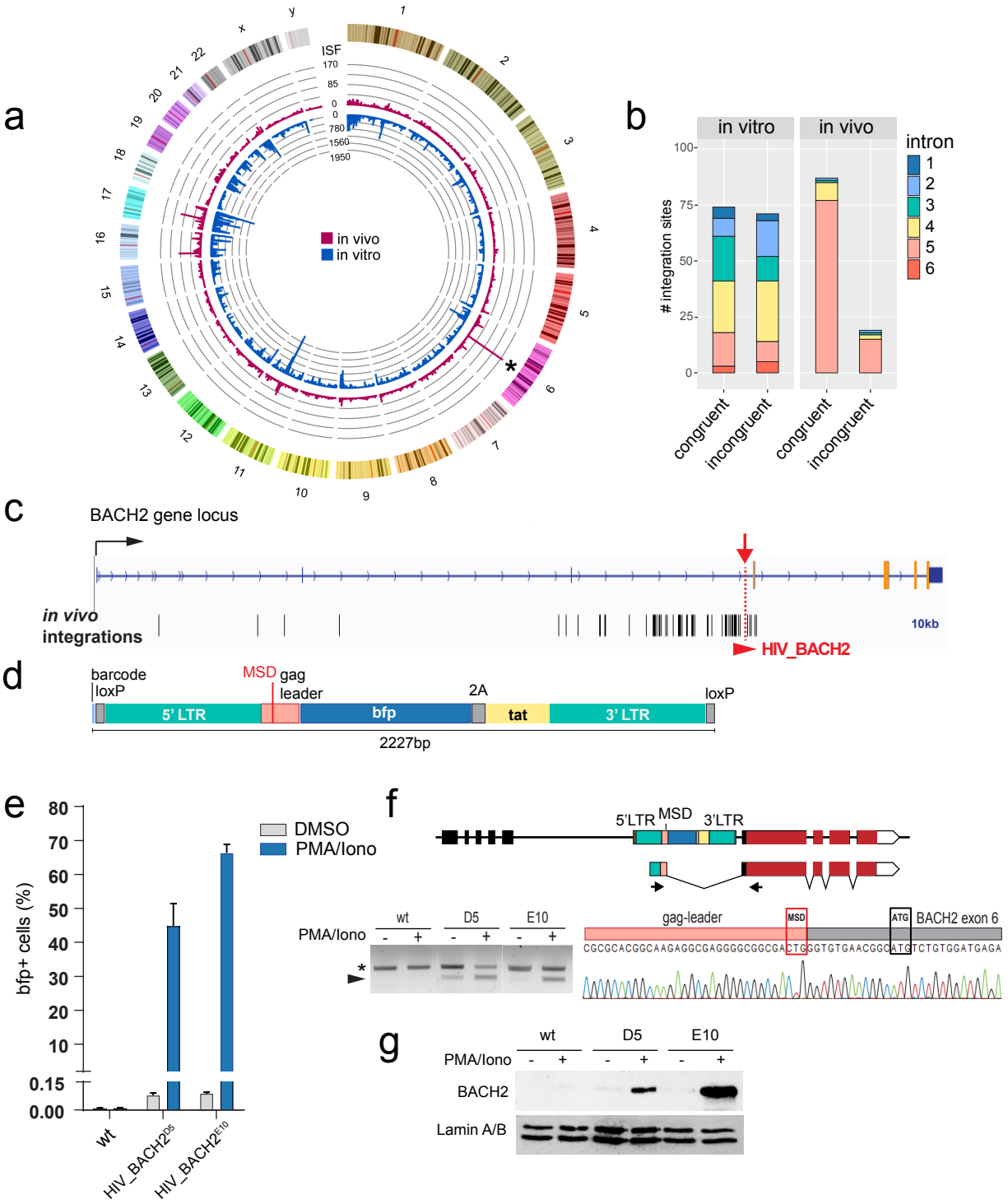


Figure 1

HIV_BACH2 cells model LTR-driven insertional activation of the BACH2 RdIG. (a) Circos plot depicting the genome-wide frequency of proviral integrations per gene locus in PLWH on ART (in vivo) vs HIV-1 infection of CD4+ T cells in vitro. Data on proviral integration sites was extracted from the Retrovirus Integration Database [45]. Asterisk marks the BACH2 gene locus. ISF: integration site frequency. (b) Frequency of in vivo and in vitro provirus integration sites in the BACH2 gene locus stratified by intron and transcriptional orientation of the provirus in context of the BACH2 gene (congruent, incongruent). (c) Schematic of the BACH2 gene locus. Black bars indicate mapped provirus integrations in PBMCs of PLWH on ART. The site of targeted integration of the HIV-1 reporter (LTRbfp2Atat) 3710bp upstream of BACH2 ATG within intron 5 is marked (red arrow). (d) Schematic of the LTRbfp2Atat reporter that was targeted into the BACH2 locus to generate HIV_BACH2 cell line models (E10 and D5 clones). (e) Percentage of bfp positive HIV_BACH2 cells (E10 and D5 clonal lines) upon exposure to PMA/Ionomycin or DMSO control. Untargeted wild-type parental cells (wt) were used as controls. (f) Detection of LTR/BACH2 chimeric transcripts in HIV_BACH2 cell models (E10 and D5 clonal lines) and untargeted wt cells by RT-PCR. Schematic indicates aberrant splicing from reporter-inherent HIV-1 major splice donor (MSD) site to acceptor site of BACH2 exon 6. Protein-coding sequence of BACH2 is indicated in red. Arrows indicate position of primers used for PCR amplification of chimeric transcripts (arrowhead). Asterisk denotes unspecific PCR product. (g) Western blot analysis of HIV_BACH2$^{D5/E10}$ cells for expression of BACH2 protein. Anti-Lamin A/B antibody was used for blotting and loading control.

FIGURE 2
Hamann et al

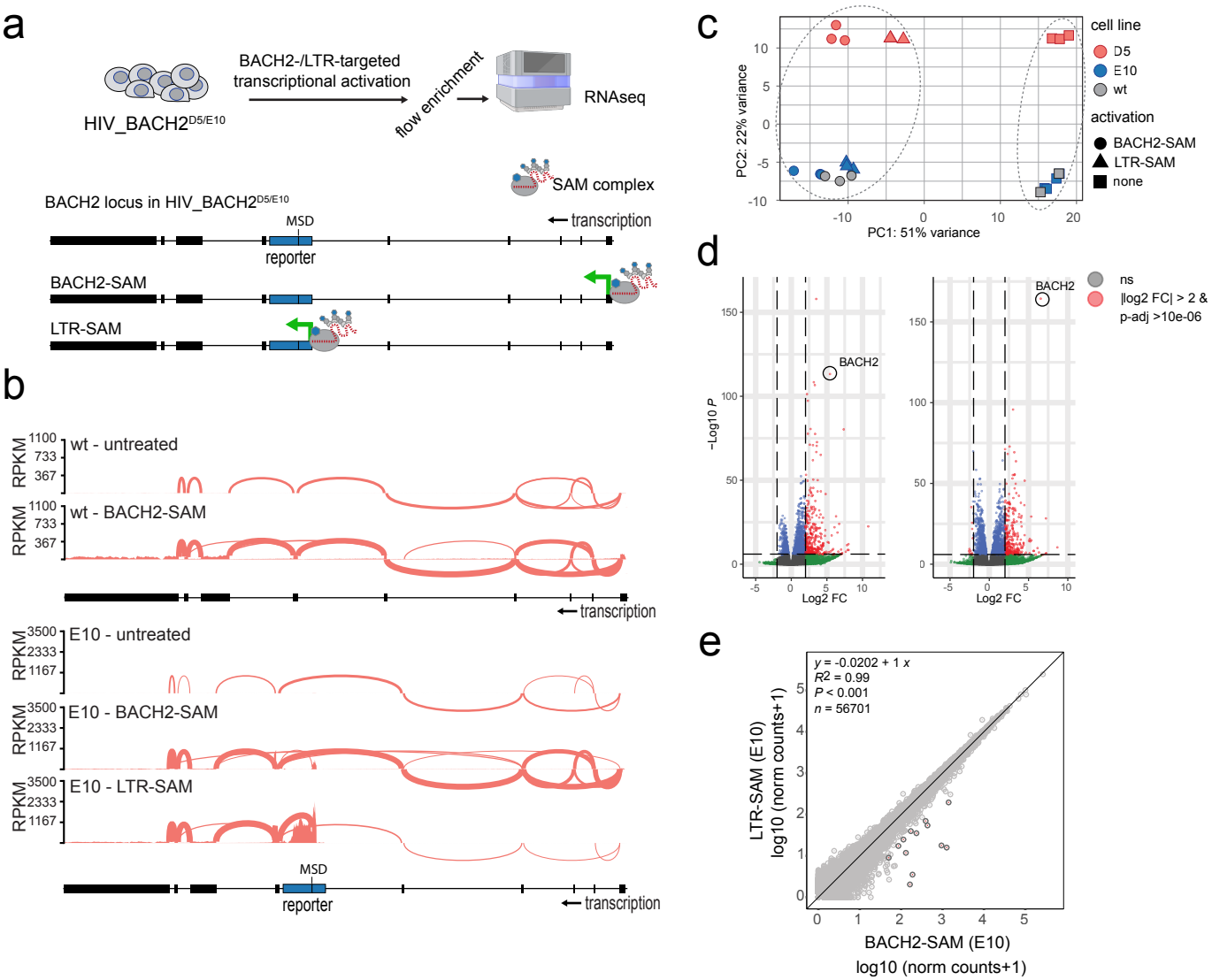


Figure 2

LTR exaptation drives BACH2 gene program in HIV_BACH2 cells. (a) Schematic depicting LTR- (LTR-SAM) and BACH2- (BACH2-SAM) targeted transcriptional activation in HIV_BACH2$^{D5/E10}$ cells and parental wild type cells (wt) using the synergistic activation mediator (SAM) system. SAM-treated cells were subjected to RNAseq analysis. (b) Sashimi plots depicting the BACH2 transcript splicing patterns in BACH2-SAM- and LTR-SAM-treated wt and HIV_BACH2^{E10} cells. (c) Principal component analysis based on RNAseq datasets of SAM-treated wt and HIV_BACH2$^{D5/E10}$ cells as well as untreated controls. (d) Volcano plots depicting differentially expressed genes (DEGs) in HIV_BACH2E10 cells upon SAM activation of endogenous BACH2 promoter (left panel) and SAM activation of BACH2-integrated LTRbfp2Atat reporter (right panel). Cutoffs set at |log2FC|>2 and padj<1x10E-6. (e) Correlation analysis of DEG genes in HIV_BACH2E10 cells upon LTR-SAM and BACH2-SAM activation. Pearson statistical method, R2 and p values indicated.

FIGURE 3
Hamann et al

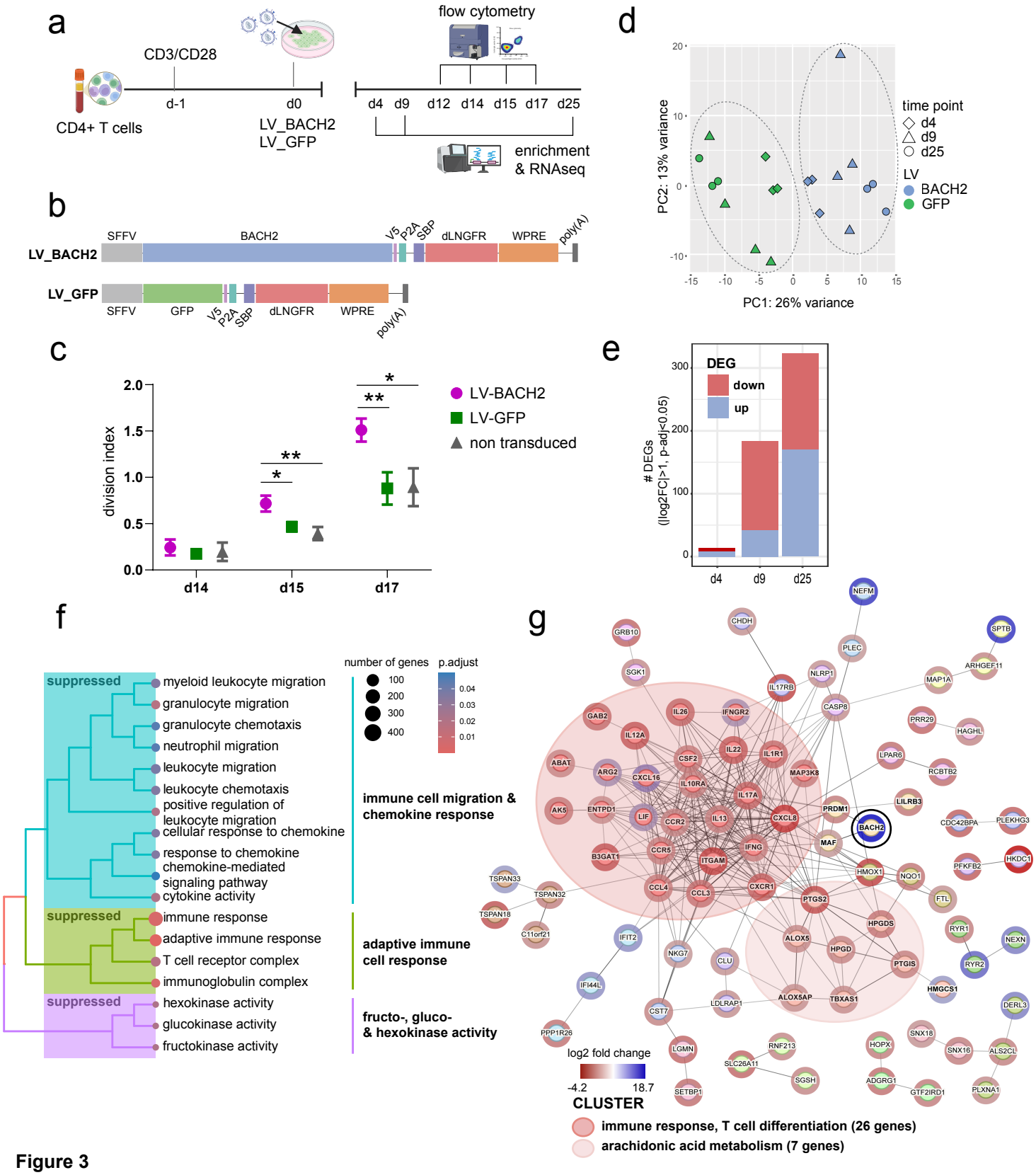


Figure 3

Aberrant BACH2 expression induces proliferation and suppresses effector fates in primary CD4+ T cells. (a) To mimic aberrant BACH2 expression, CD3/CD28-activated CD4+ T cells were transduced with LV-BACH2 and control LV-GFP lentivirus and assessed longitudinally for proliferative behavior (flow-cytometry based) and transcriptomic changes. (b) Schematic of LV-BACH2 and LV-GFP expression cassettes. (c) Cell tracer-based assay for proliferative behavior of LV-BACH2, LV-GFP and non-transduced cells. Cells were stained at d12 post transfection and followed on d14, d15 and d17. Statistical analysis was done using unpaired t-test, *p<0.05, **p<0.03. (d) Principal component analysis based on RNAseq datasets of SPB-dLNGFR-enriched LV-BACH2 and LV-GFP transduced CD4+ T cells at d4 (n=3), d9 (n=4) and d25 (n=3) post transduction. (e) Number of differentially expressed genes (DEG; cutoffs set at |log2FC|>1, padj<0.05) between LV-BACH2 and LV-GFP transduced CD4+ T cells at d4, d9 and d25 post transduction. (f) Gene set enrichment analysis (GSEA) based on DEGs between LV-BACH2 and LV-GFP transduced CD4+ T cells at d9 post transduction delineates suppressed cellular pathways in LV_BACH2 cells. (g) STRING analysis based on DEG between LV-BACH2 and LV-GFP transduced CD4+ T cells at d25 post transduction.

FIGURE 4
Hamann et al

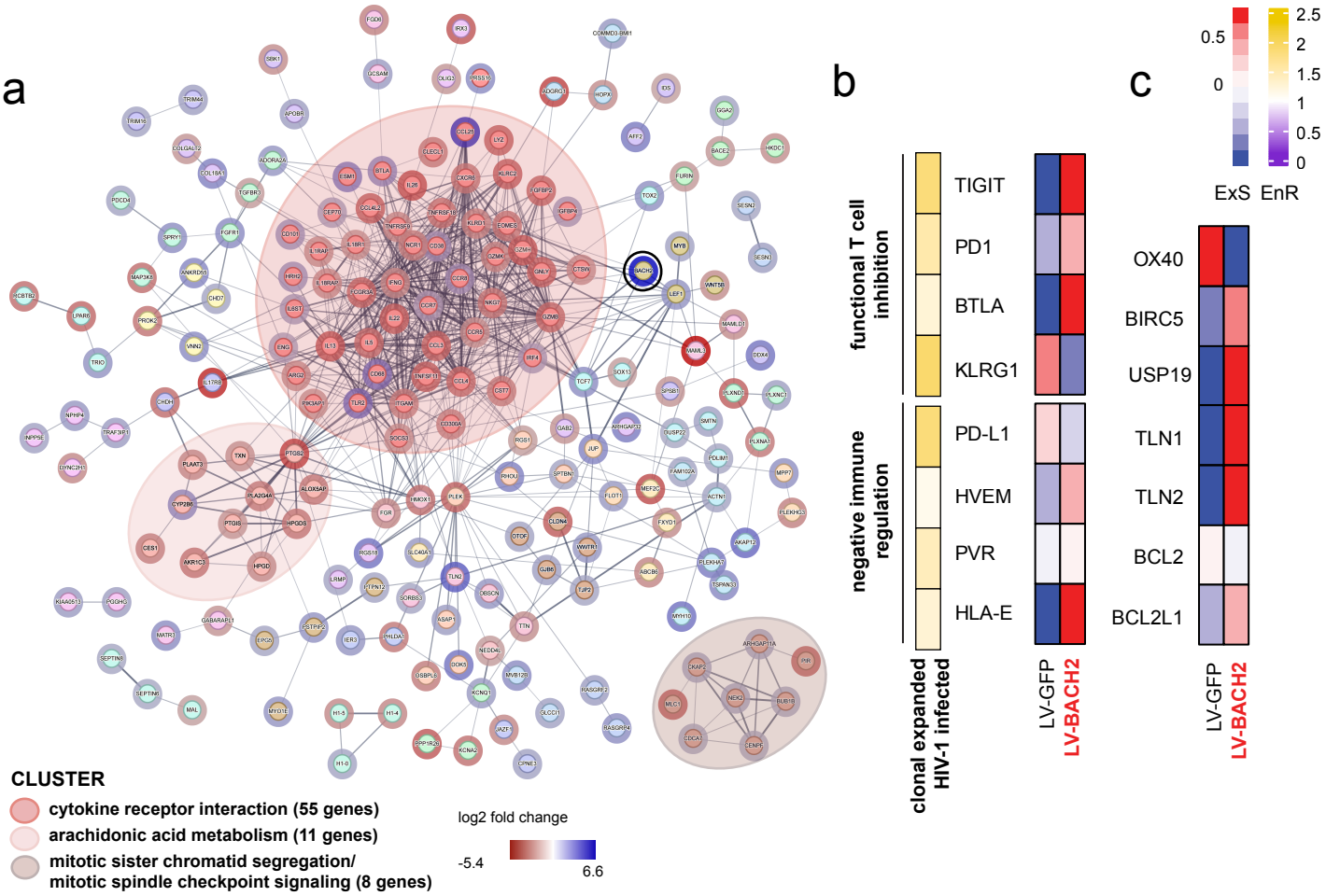


Figure 4
BACH2-reprogrammed CD4+ T cells show precursor memory signature and mimic features of expanded CD4+ T cells in PLWH on ART. (a) STRING analysis based on DEG between LV-BACH2 and LV-GFP transduced CD4+ T cells at d25 post transduction. (b) Heatmap depicting transcript expression scores (ExS) and enrichment scores (EnS) for key genes associated with T cell inhibition and negative immune regulation. ExS are derived from LV-BACH2 and LV-GFP transduced CD4+ T cells (d25 post transduction) and EnS are derived from a published dataset contrasting clonally expanded to non-clonally expanded HIV-1 infected CD4+ T cells from PLWH on ART 44. ExS: scaled normalized transcript levels; EnR: enrichment ratio calculated as the ratio of the proportion of clonal HIV-infected cells divided by the proportion of non-clonal HIV-infected cells (c) Heatmap depicting transcript expression scores (ExS) in LV-BACH2 and LV-GFP transduced CD4+ T cells (d25 post transduction) for key genes associated with anti-apoptotic activity in infected CD4+ T cells in PLWH on ART.

FIGURE 5
Hamann et al

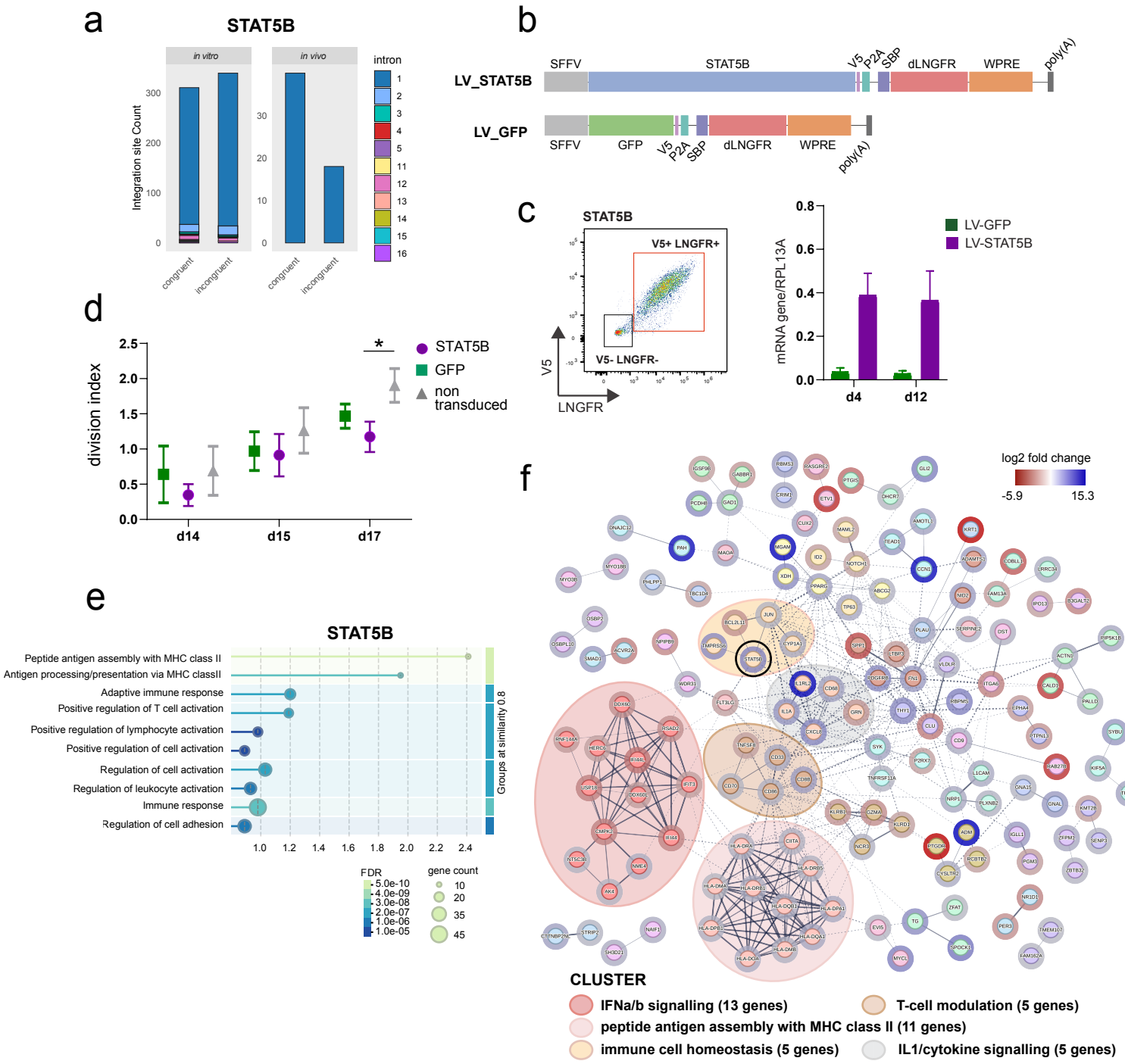


Figure 5

Aberrant expression of STAT5B alters CD4+ T cell phenoytpe towards effector type. (a) Frequency of in vivo and in vitro provirus integration sites in the STAT5B gene locus stratified by intron and transcriptional orientation of the provirus in context of the BACH2 gene (congruent, incongruent). (b) Schematic of LV-STAT5B and LV-GFP expression cassettes used to generate lentiviruses for transduction of primary CD4+ T cells to mimic aberrant expression. (c) Flow cytometric analysis of LV-STAT5B transduced CD4+ T cells stained for V5 and LNGFR expression (left panel). RT-qPCR for normalized STAT5B transcript levels in SPB-dLNGFR-enriched LV-STAT5B and LV-GFP transduced cells, at day (d) 4 and 12 post transduction (right panel). (d) Cell tracer-based assay for proliferative behavior of LV-STAT5B transduced cells as well as LV-GFP and non-transduced control cells. Cells were stained at d12 post transfection and followed on d14, d15 and d17. Statistical analysis was done using unpaired t-test, *p<0.05. (e) Enrichment analysis of functional gene networks stratified by biological processes based on DEG of LV-STAT5B transduced cells compared to LV-GFP controls at d12 post transduction. (f) STRING analysis based on DEG between LV-STAT5B and LV-GFP transduced CD4+ T cells at d12 post transduction.

FIGURE 6
Hamann et al

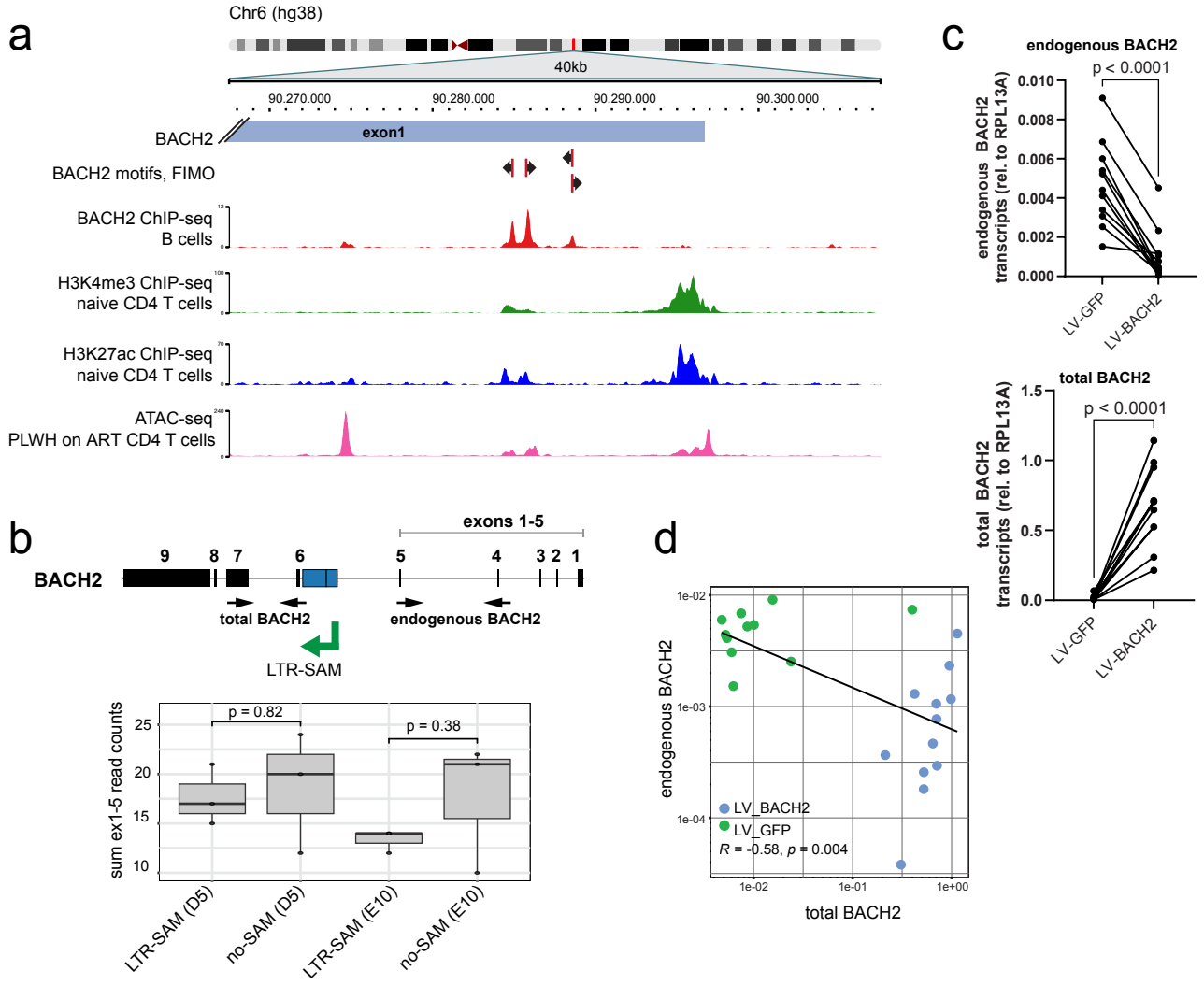


Figure 6

Autoregulation of BACH2 expression through negative feedback loop. (a) BACH2 exon 1 contains four BACH2 binding motifs, of which two are bound by BACH2 (data from B cells) and display cis-regulatory epigenetic marks and open chromatin accessible in CD4+ T cells. (b) HIV_BACH2$^{D5/E10}$ LTR-SAM and BACH2 SAM-treated cells show down-regulation of exon 1 to 5-derived transcripts upon LTR-SAM activation. Positions of primers used for total and endogenous RT-qPCR analysis in (c) are marked by arrows. Statistical analysis with paired two-tailed t-test. (c) RT-qPCR of endogenous and total BACH2 transcript levels in LV-BACH2 and LV-GFP transduced CD4+ T cells. Total BACH2 transcripts derive from endogenous BACH2 locus and lentiviral expression construct. Statistics using paired, two-sided t-test. (d) Pearson correlation analysis of endogenous and total BACH2 transcript levels in LV-BACH2 and LV-GFP transduced CD4+ T cells (R2 and p values indicated in the plot).

FIGURE 7
Hamann et al

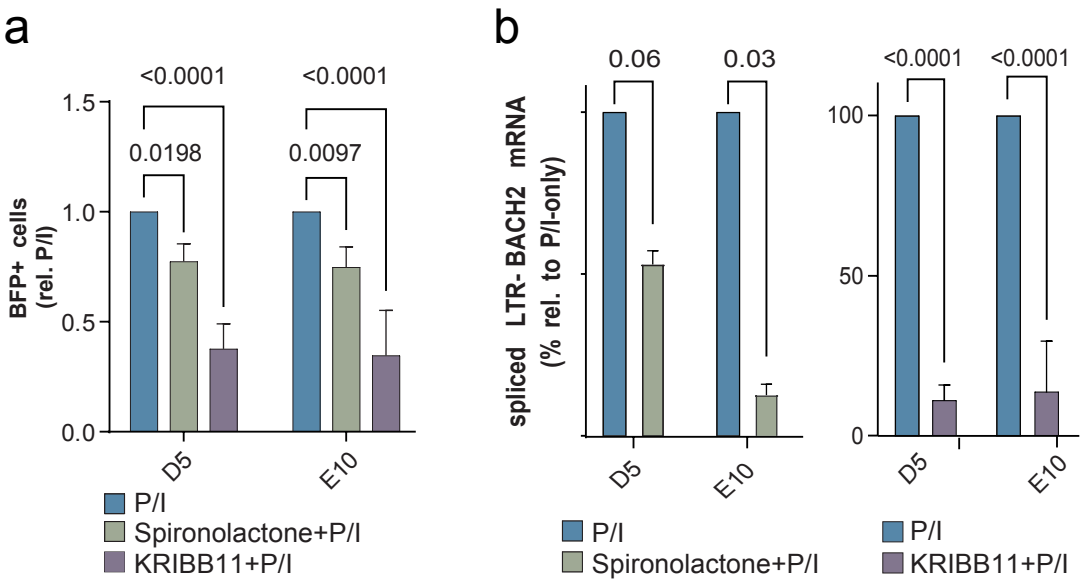
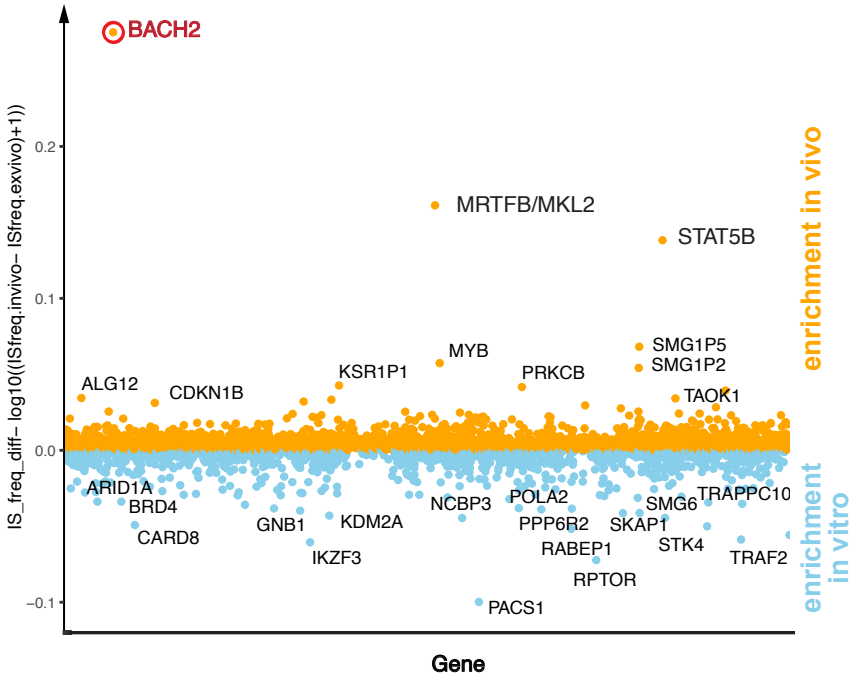


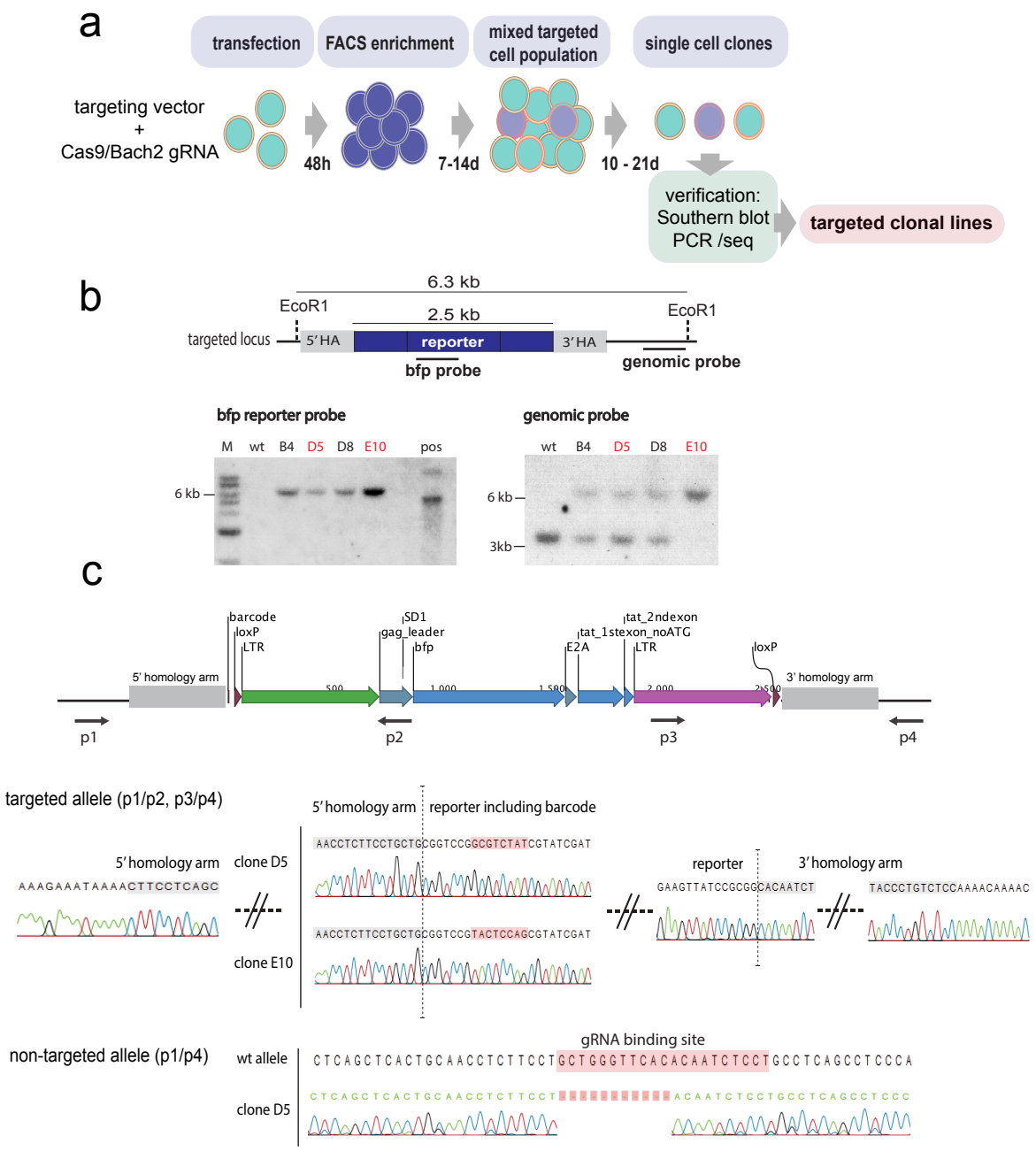
Figure 7
Latency promoting agents reduce LTR exaptation in HIV_BACH2 cells. HIV_BACH2$^{D5/E10}$ cells were treated for 24h with LPAs Spironolactone or KRIBB11 and co-induced for 6hs with PMA/Ionomycine (P/I) prior to readout. Controls were treated with 6h P/I alone. (a) Fraction of cells that express the LTR-driven bfp reporter as measured by flow cytometry (relative to P/I only control). (b) Transcript levels of LTR/BACH2 chimeric transcripts as measured by RT-qPCR relative to P/I only control. Statistics using paired, two-sided t-test.

SUPPLEMENTAL FIGURE 1
Hamann et al



Suupplemental Figure 1
Reshaping of provirus integration landscape in PLWH on ART. Manhattan plot showing enrichment of proviral integration genes in PLWH on ART (in vivo) versus HIV-1 infection of CD4+ T cells in vitro as calculated by frequency of integration sites per gene locus. Data on proviral integration sites was extracted from Retrovirus Integration Database [45].

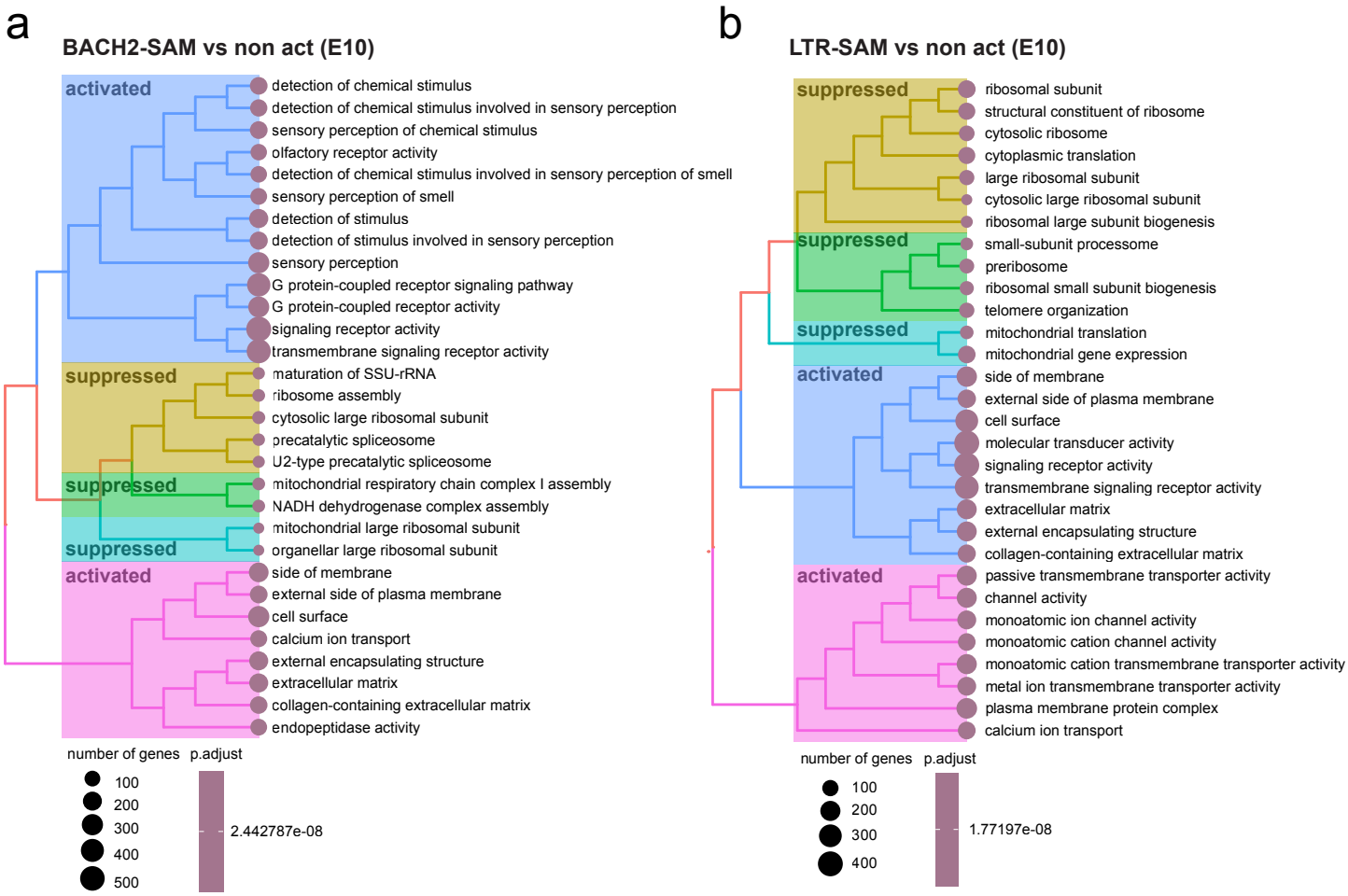
SUPPLEMENTAL FIGURE 2
Hamann et al



Supplemental Figure 2
Generation of HIV_BACH2 cell line models. (a) Workflow for generating HIV_BACH2 clonal cell lines. (b) Verification of HIV_BACH2 clonal lines for correct reporter targeting into the BACH2 locus (3710bp upstream of BACH2 open reading frame start) using Southern blotting. Genomic DNA of targeted clones was extracted, digested with EcoR1 and submitted to gel electrophoresis for size-separation of fragments. Blotted DNA was probed against bfp (bfp probe) and a genomic region outside the homology arm (HA) sequence employed for targeting (genomic probe). Clonal lines B4, D5, D8 and E10 were tested and showed heterozygous (B4, D5, D8) or homozygous (E10) pattern of targeted reporter insertion. Lines D5 and E10 were selected for further analysis. (c) Verification of HIV_BACH2 clonal lines for correct reporter targeting by Sanger sequencing. Genomic DNA of HIV_BACH2 D5 and E10 clonal lines was extracted and genomic regions containing 5' (primer p1 and p2) and 3' (p3 and p4) genome/reporter junctions were amplified by PCR and products submitted to Sanger sequencing. For D5 lines, the region spanning the non-targeted allele, containing the complementary sequence to the gRNA used for Cas9 targeting, was amplified using primers p1 and p4. Chromatograms show correct genome/reporter junctions and an 11bp deletion in the gRNA-complementary sequence on the non-targeted allele for D5 lines.

SUPPLEMENTAL FIGURE 3

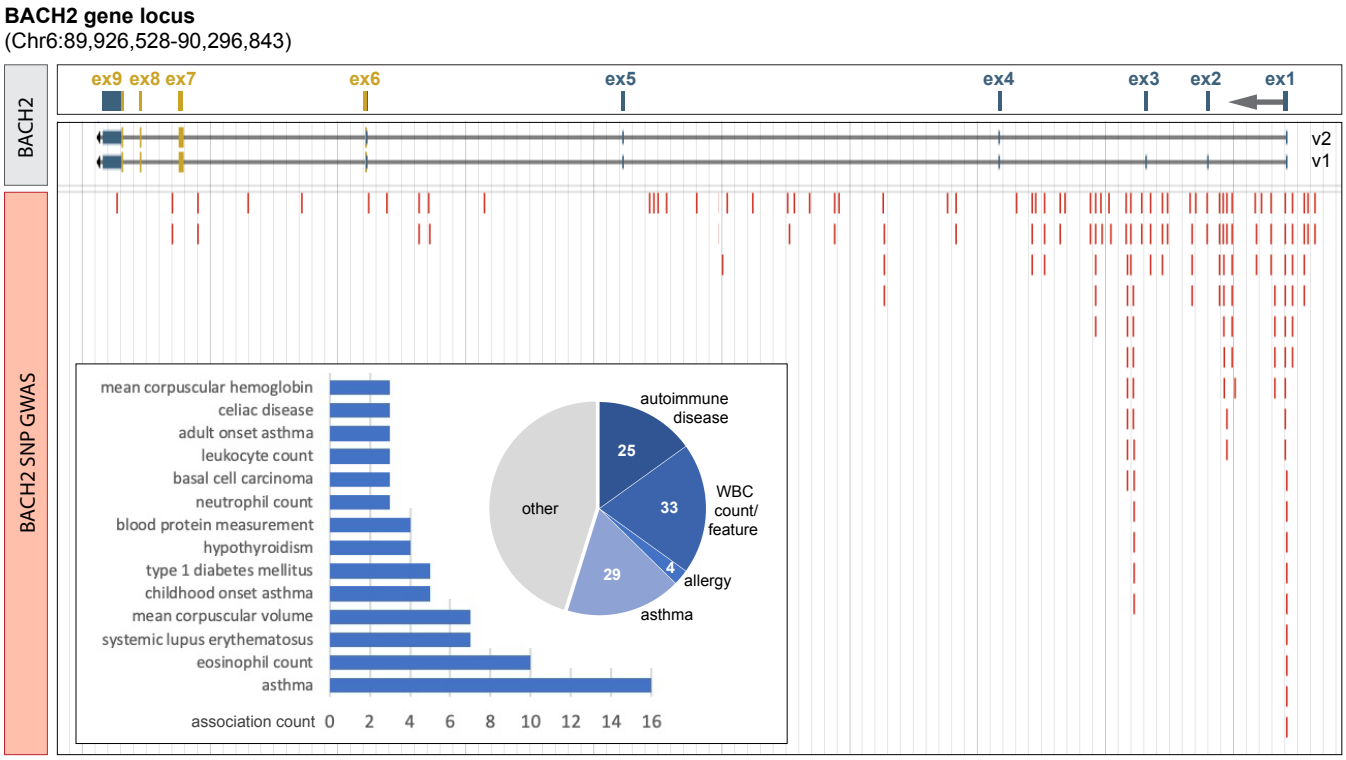
Hamann et al



Supplemental Figure 3

BACH2-induced transcriptomic changes in HIV_BACH2 cells. HIV_BACH2^{E10} clonal lines were submitted to BACH2-targeted (BACH2-SAM) or LTR-targeted (LTRT-SAM) transcriptional activation and transcriptomic changes analyzed by RNAseq. Treeplots depicting GSEA of differentially expressed genes upon (a) BACH2-targeted activation (BACH2-SAM) and (b) LTR-targeted activation (LTR-SAM) are shown.

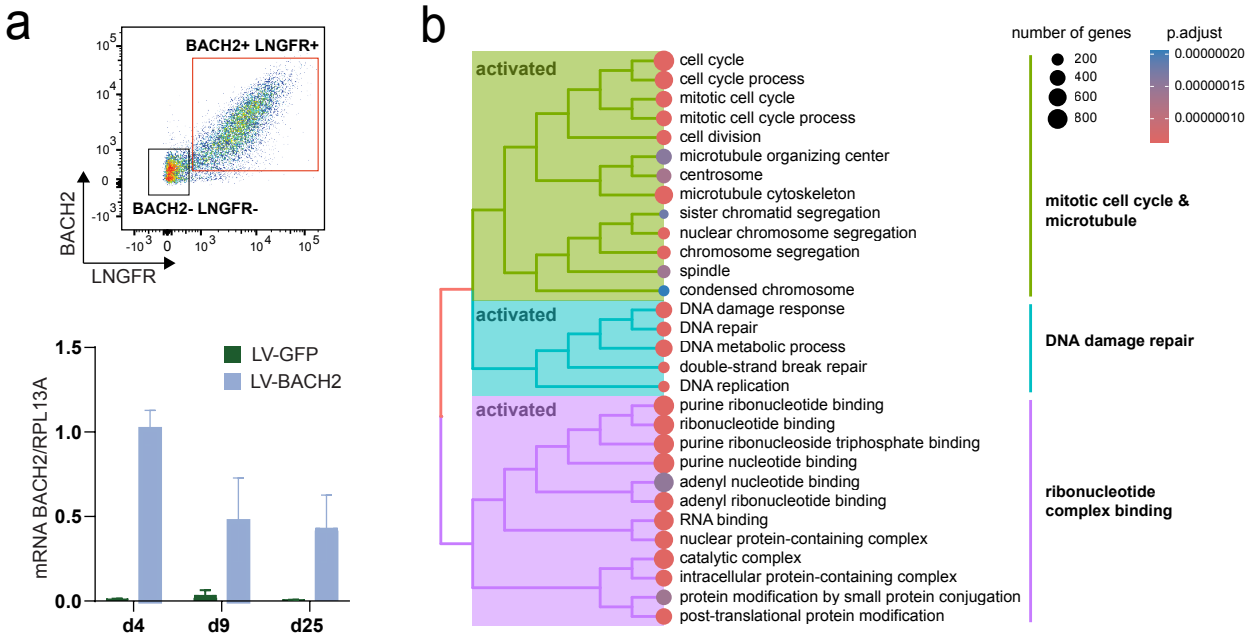
SUPPLEMENTAL FIGURE 4
Hamann et al



Supplemental Figure 4

Single nucleotide polymorphisms (SNPs) in the BACH2 locus have been associated with immune-related traits in genome-wide association studies (GWAS). The NHGRI-EBI GWAS catalog was used to retrieve information on BACH2 SNP-associated traits [131]. 84 EFO traits were associated with 166 SNPs within the BACH2 locus. Positions of SNPs relative to BACH2 gene and main BACH2 transcript variants (v1, v2) are marked (red bars). Exons coding for BACH2 open reading frame are highlighted in yellow. Bar graph insert depicts the 14 traits with association counts >3. Pie chart based on all BACH2 associations highlights associations with immune-related traits (blue).

SUPPLEMENTAL FIGURE 5
Hamann et al



Supplemental Figure 5
Transcriptomic changes induced by aberrant BACH2 expression in CD4+ T cells. (a) Flow cytometric analysis of LV-BACH2 transduced CD4+ T cells stained for BACH2 and LNGFR expression (upper panel). RT-qPCR for normalized BACH2 transcript levels in SPB-dLNGFR-enriched LV-BACH2 and LV-GFP transduced cells at day (d) 4, 12, 25 post transduction (lower panel). (b) Gene set enrichment analysis (GSEA) based on DEG between LV-BACH2 and LV-GFP transduced CD4+ T cells at d25 post transduction delineates activated cellular pathways in LV_BACH2 cells.