# Microbiome and Metabolome driven differentiation of TGF-β producing Tregs leads to Senescence and HIV latency

**Khader Ghneim[1*], Ashish Arunkumar Sharma[1*ϕ], Susan Pereira Ribeiro[1], Slim Fourati[1], Jeffrey Ahlers[2], Deanna Kulpa[3], Xuan Xu[1], Jessica Brehm[4], Aarthi Talla[1], Sahaana Arumugam[5], Samuel Darko[5], Benigno Rodriguez[6‡], Carey Shive[7], Razvan Cristescu[8], Andrey Loboda[8], Robert Balderas[9], I-Ming Wang[10], Peter Hunt[11], Daniel Lamarre[1], Daniel Douek[5], Daria Hazuda[8], Michael M. Lederman[6], Steven Deeks[11], Rafick-Pierre Sekaly[1ϕ]**

**1. Department of Pathology, University Hospitals of Cleveland and Case Western Reserve University, 2103 Cornell Road, Cleveland, Ohio 44106, USA.**
**2. Calimmune, Inc. 35 North Lake Ave., Suite 600, Pasadena, CA 91101, USA**
**3. Department of Pediatrics, Emory University, Atlanta, GA, USA**
**4. VIIV Healthcare, 5 Moore Drive, Research Triangle Park, North Carolina, NC 27709, USA**
**5. Vaccine Research Centre (NIH), 40 convent drive, Bethesda, Maryland, 20892, USA**
**6. Department of Medicine, Division of Infectious Diseases and HIV Medicine, Center for AIDS Research, Case Western Reserve University/University Hospitals Case Medical Center, Cleveland, Ohio, USA**
**7. Veterans Administration Medical Center, Cleveland, Ohio USA**
**8. MRL, Merck & Co., Inc., West Point, PA 19486, USA**
**9. BD Biosciences, San Jose, California, 95131, USA**
**10. Pfizer Inc., New York, New York, United States of America**
**11. Department of Medicine, University of California, San Francisco, California 94110, USA**

**Both authors equally contributed to the manuscript*
*ϕRafick-Pierre Sekaly and Ashish Arunkumar Sharma are the corresponding authors for this manuscript. Please direct all inquiries to: rafick.sekaly@case.edu.*
*‡Deceased.*

**Summary**

Current therapeutic interventions to eradicate latent HIV ("reservoir") and restore immune function in ART-treated HIV infection have yet to show efficacy. To explore mechanisms of HIV persistence, we apply an integrated systems biology approach and identify a distinct group of individuals with poor CD4 T-cell reconstitution (Immunologic non-responders, "INRs") and high frequencies of cells with inducible HIV. Contrary to the prevailing notion that immune activation drives HIV persistence and immune dysfunction, peripheral blood leukocytes from these subjects have enhanced expression of a network of genes regulated by cellular senescence driving transcription factors (TFs) FOXO3, SMAD2 and IRF3. In these subjects, increased frequencies of regulatory T cells and expression of the TGF-β signaling cascade are complimented by the downregulation of cell cycle, metabolic and pro-inflammatory pathways. Lactobacillaceae family and metabolites (members of the butyrate family – i.e. α-ketobutyrate) were correlated with Treg frequencies in "Senescent-INRs" *ex vivo*, triggered the differentiation of TGF-β producing Tregs and promoted HIV latency establishment *in vitro*. These cascades, downstream of PD-1/TGF-β, prevent memory T cell differentiation and are associated with an increase in frequencies of cells with inducible HIV *ex vivo*. Our findings identify cellular senescence responses that can be targeted by PD-1 or TGF-β specific interventions that have shown safety and efficacy in cancer, and may prove to be crucial for HIV eradication.

## Introduction

HIV infection remains a major global public health concern (WHO, https://www.who.int/news-room/fact-sheets/detail/hiv-aids). Although effective combined anti-retroviral therapy (cART) has altered the course of HIV disease by preventing viral replication, it is not curative and does not fully restore immune function[1]. Depending on the population studied and outcome definitions[2], up to 20% of cART-treated subjects fail to reconstitute CD4+ T-cell numbers despite years of effective treatment with sustained HIV suppression, and are predisposed to excessive risk of non-HIV comorbidities[3].

These immune non-responder subjects (INRs) show severe homeostatic alterations in CD4 T-cells, including lower frequencies of naïve CD4 T-cells, accumulation of highly differentiated T-cells, T-cell activation and apoptosis[4-6]. Impairment in IL-7/L-7R signal transduction axis[7], heightened chronic inflammation and persistent type I interferon production; are thought to be important mediators of poor CD4 T cell homeostasis and reduced thymic function[8] in the INRs. Additionally, CD4 T cells in the INRs have higher expression of replicative senescence[9] and exhaustion markers[10], and show an increase in frequencies of immunosuppressive CD4+ regulatory T-cells (Tregs)[11]. Using multivariate analyses, we previously reported that naive T-cell depletion, higher frequencies of cycling CD4+ central memory (CM) and effector memory (EM) T-cells, expression of the T-cell activation markers (CD38, HLA-DR, CCR5 and/or PD-1 ($p<0.0001$)), and levels of soluble CD14 (sCD14) distinguished INRs from immune responder subjects (IRs). These groups are distinguishable even

when adjusted for CD4+ T-cell nadir, age at cART initiation, and other clinical indices[12-14]. Collectively these studies highlight the contribution of dysregulated pro-inflammatory pathways to the lack of immune reconstitution in INRs.

Despite the ability of cART to inhibit HIV replication, the proviral DNA integrates into the human genome and persists within memory CD4 T-cells[15]. This cellular "reservoir" reignites rounds of virus replication if cART is interrupted[16], stressing the need for long-term cART administration. Indeed, reconstitution of CD4 counts during uninterrupted cART leads to a decrease in intact pro-viral DNA[17], whereas elevated levels of inflammatory markers characteristic of INRs (IL6, TNF□, IL1β, IFN□/β, MIP1□/β, RANTES) are associated with HIV persistence[18]. The activated state of the innate immune system can trigger HIV replication, alter homeostatic proliferation and sustain the HIV reservoir by occasional expansion and contraction of individual CD4 T-cell clones[19-21]. Specifically, increased expression of co-inhibitory receptors (PD-1, LAG-3, and TIGIT) that can drive cellular quiescence upon receptor engagement and inhibit HIV replication, have been shown to contribute to the magnitude of the HIV reservoir [21-23].

In this study of two independent cohorts of HIV-infected cART-treated individuals, we identified INRs with high serum levels of anti-inflammatory molecules (triggered by elevated FOXO3 expression, TGF-β associated genes), a senescent CD4 T cell profile with absence of homeostatic proliferation (and lack of CD4 recovery), and heightened frequencies of cells with inducible HIV RNA.

## Results

We used an unbiased approach that integrates data-sets from peripheral blood transcriptome, high density flow cytometry and plasma cytokine measures, to identify cellular and molecular drivers of HIV persistence and lack of CD4 T cell recovery (Supplementary Fig. 1a), in two independent cohorts of HIV-infected subjects (see Study Participants in Material and Methods; Table S1, Table S2). All subjects were under cART for at least three years and maintained consistent CD4 counts four years prior to sample collection (Supplementary Fig. 1b). The Cleveland Immune Failure cohort (CLIF) was composed of 61 subjects (17 immune responders - IRs - >500 CD4 T-cells/mm$^3$ and 44 immune non-responders - INRs - <350 CD4 T-cells/mm$^3$); whereas, the Study of the Consequences of the Protease Inhibitor Era cohort (SCOPE) included 41 subjects (20 IRs with >500 CD4 T-cells/mm$^3$ and 21 INRs with <350 CD4 T-cells/mm$^3$).

### Transcriptional profiling reveals systemic senescence as a driver of poor immune reconstitution

Exploratory analyses of whole blood transcriptomic data (CLIF cohort subjects; using unsupervised clustering described in the methods) (Fig. 1a and Supplementary Fig. 1c) identified three groups of subjects with unique transcriptional profiles: IRs and two distinct INR groups, “INR-A” and “INR-B”.

INR-As exhibited the highest transcriptomic variation from IRs with approximately 3000 differentially expressed genes (DEGs). In contrast, INR-Bs were proximal to IRs with <400 DEGs (Table S3-5). Age, years on cART, CD4 counts, CD4 nadir, markers of gut barrier dysfunction (sCD14), and inflammation (IL-6, IP10) - all known predictors of morbidity and mortality in INRs[24] - failed to distinguish the two INR groups (Supplementary Fig. 1d-p). Interestingly, DEGs specific to INR-As were mostly down-regulated (72% and 66% when contrasted against IRs and INR-Bs, respectively), suggesting a quiescent transcriptional state within these subjects (Supplementary Fig. 2a).

Comprehensive pathway analyses using MSigDb's Hallmark module[25-27] (Fig. 1b, Table S6) revealed that several pathways, reflective of significantly elevated CD4 numbers, were upregulated in IRs when compared to both INR-As and INR-Bs. Notably, when comparing INR-As and INR-Bs (where CD4 numbers are comparable), INR-As showed significantly decreased expression of inflammation, cell cycling, apoptosis, and cellular metabolism genesets, indicative of a senescent state[28,29]. INR-As had reduced expression of genes regulated by Myc, a primordial TF that regulates pathways involved in proliferation and metabolism (glycolysis, oxidative phosphorylation and fatty acid metabolism) of activated T cells[30].

Cell subset deconvolution (using gene signatures from Nakaya et al[31]) revealed a significant down-regulation in INR-As of genesets specific to all major innate and adaptive immune cell subsets (Fig. 1c, Table S7). Apoptosis, transcription/cell migration, and cellular/lipid metabolism (lipid storage, GTP

metabolic process) were enriched in gene signatures down-regulated in T-cells, myeloid dendritic cells (mDCs) and monocytes, respectively (Supplementary Fig. 2b). Similarly, down-regulation of genesets (extracted from Novershtern et al[32] - see methods for detail) specific to CD4 and CD8 T-cells naive, memory and effector cell subsets (Fig. 1d, Table S8). Altogether, these data provide supporting evidence for decreased global transcription and systemic senescence in immune cells from INR-As, when compared to INR-Bs or IRs.

**Transcriptional profiles downstream of FOXO3 and SMAD2/3 signaling are enriched in senescent INRs**

To identify mechanisms underlying poor immune reconstitution observed in INR-As, we first identified TFs that were specifically enriched in INR-As ($p < 0.05$, Supplementary Fig. 2c, Table S9). These TFs included IRF3 (driver of type I IFN production[33]), FOXO3 (transcriptional repressor[34]), SMAD2 (TGF-β signaling[35]) and CCNT2 (Negative regulator of HIV Tat protein; i.e. driver of HIV latency induction[36]) (Supplementary Fig. 2c). Next, we used the Reactome pathway database (using ClueGO, FDR <0.05)[37,38] to map the functions of upregulated genes in the INR-As (vs IRs and INR-Bs; FDR < 0.05, Table S10). We observed the enrichment of cellular processes downstream of the above-mentioned TFs including heme metabolism, TGF-β signaling, IRF3 activation, reactive oxygen species (ROS) production and inhibition of NF-κB activation (Fig. 1e) in these subjects. Specifically, features of senescence regulated by FOXO3 including SOD/CAT[39-41] (driver of ROS production) (Fig 1e); and FOXO4/TP53[42]

driven anti-apoptotic genes (BCL2L1; FDR < 0.01)[43-45] were upregulated in INR-As. Interestingly, we observed the enrichment of key senescence associated genes which are known targets of SMAD2/3 including 'Inducer of Promyelocytic Leukemia' (PML; conductor of TGF-β signaling via SMAD2/3, and regulator of HIV latency[46,47]), WEE1[48] (cell cycle regulator) and GLUL[49,50] (metabolic regulator) (Supplementary Fig. 2d). These senescent features of INR-As were further confirmed by the downregulation MYC target genes (known to control ribosomal biosynthesis and translation)[51], genes of the electron transport chain, and genes regulating major metabolic pathways (including glycolysis, oxidative phosphorylation and fatty acid metabolism) (Supplementary Fig. 2d, Fig. 1c).

The senescent/anti-inflammatory profile of INR-As contrasted the pro-inflammatory profile of INR-Bs where the enrichment of inflammatory pathways downstream of TLR/IL1, cell cycling, and apoptosis/pyroptosis (Supplementary Fig. 2d) was observed. Increased expression of several members of the pro-inflammatory NF-κB family of TFs (NFKB1, NFKB2 and RELB) (Supplementary Fig. 2d), upregulation of NF-κB target genes including chemokines (e.g. CCL2 and CCL17)[52], genes of the inflammasome complex (NLRP3, IL1B and IL18)[53] and effector genes driving apoptosis/pyroptosis (e.g. CASP4, DIABLO, CASP2 and CASP3)[54] were characteristics of this group (Supplementary Fig. 2d).

Overall, our data suggest that a signaling cascade downstream of TGF-β signaling (SMAD2/3; PML) which culminates by the upregulation in the INR-As of FOXO3/4 driven senescence downstream of TGF-β and characterized by impaired cell metabolism and cell cycle arrest. The expression of the above listed

transcriptional repressors in INR-As could further explain the inhibition of cell metabolism, cell cycle and inflammation all associated to cell senescence[55]. INR-As will henceforth be referred to as “Senescent-INRs”.

**A gene-based classifier confirms the generalizability of the Senescent-INR phenotype in HIV-infected subjects**

Our next objective focused on demonstrating that other cohorts of HIV infected cART treated subjects also include senescent INRs. We built and validated a gene-based nearest shrunken centroid classifier (see methods) to identify Senescent-INRs in an independent cohort (SCOPE) of cART-treated HIV infected subjects. The classifier, trained on the CLIF cohort, consisted of 352 features (genes, Table S11) and provided a misclassification error rate of 0.26 (Supplementary Fig. 3a). Using the unsupervised approach (initially applied in the CLIF cohort), we confirmed two distinct INR groups in the SCOPE cohort (Supplementary Fig. 3b). The 352-gene classifier segregated these two INR groups in the SCOPE cohort with an accuracy of 81% (Fig. 1f, Supplementary Fig. 3c), thus validating the generalizability and reproducibility of two INR groups (PCA representation of the 352-gene classifier discriminates senescent INRs in Fig. 1g) and highlighting the potential use of this classifier to distinguish Senescent-INRs in clinical settings. This classifier (Table S11) included genes associated with the induction of a senescent profile (FOXO3, FOXO4, and TGFBR2)[43,56], genes indicative of senescence biology (RIOK3, IRF3, BCL2L1)[57,58] and regulators of mitochondrial activity (NDUFS3, CYB5R3, ATP5G2)[59]. Several transporters of macromolecules: SLC6A8 (Sodium- and

chloride-dependent creatine transporter 1[60]), SLC48A1 (Heme transporter[61]), SLC4A1 (Band 3 anion transporter[62]), SLC25A23 (Mitochondrial Calcium carrier[63]) and SLC38A5 (glucose[64]) were also upregulated in Senescent-INRs and were features of the classifier (Table S11). These results higjlight several molecular pathways specific to cellular senescence[65-68] and recapitulate the pathways (senescence, metabolism, and cell cycle) that discriminate between the two INR groups. These data counter-intuitively suggest that failure to increase CD4 T-cell numbers in senescent INRs to levels observed in IRs could be due to senescence and not to previously described pro-inflammatory cytokines and pathways[12].

**The classifier gene set highlights the role of senescence in driving the magnitude of the inducible HIV reservoir**

Although poor immune reconstitution has been associated with HIV persistence[69], the cellular effectors and molecular mechanisms driving this association remain to be identified. To assess the role of senescence in HIV persistence, we measured frequencies of CD4 T-cells with inducible multi-spliced HIV RNA ("inducible HIV") on all subjects of the SCOPE cohort using the Tat/rev Induced Limiting Dilution Assay (TILDA)[70]. Our data show that the Senescent-INRs have significantly higher inducible HIV when compared to IRs (~4.3-fold increase in median values, $p<0.021$) and INR-Bs (~5.5-fold increase in median values, $p = 0.046$) (Fig. 1h). As might be expected, inducible HIV levels were negatively correlated with CD4 counts, selectively in Senescent-INRs (rho = -

0.52; $p < 0.05$) while this correlation was not significant in inflammatory INRs and in IRs. In addition, SCOPE cohort subjects with the highest levels of inducible HIV showed a significant enrichment in (i.e. could be predicted by) the classifier geneset (Fig. 1i). Comprehensive analyses of the whole transcriptome confirmed that genesets characteristic of Senescent-INRs (i.e. cell cycle arrest and ROS production) were associated with higher inducible HIV (Table S12).

To better characterize the impact of senescence on intrinsic mechanisms of anti-viral activity, we studied the interplay of three master-regulators (IRF3: Type I IFN response[33], IRF7: anti-viral response[71], FOXO3: Senescence[42]) that showed significant alterations in senescent INRs (Fig 2). Increased expression of IRF3 (Fig. 2a) in Senescent-INRs was strongly associated with higher levels of FOXO3 (Fig. 2b), and contrasted the significantly reduced expression of IRF7 (Fig. 2c). These findings are in line with previous reports where a negative feedback loop between FOXO3 and IRF7[72] has been ascertained, and suggest that Senescent-INRs could exhibit heightened inducible HIV when compared to IRs and INR-Bs as a consequence of the FOXO3 driven downregulation of innate antiviral immunity. Indeed, restriction factor genes known to curb HIV life cycle and known tobe transcriptional targets of IRF [71,73] were reduced (examples include SAMHD1, BST2, APOBEC3G and MX2) in subjects with the lowest inducible HIV (Fig. 2d, Table S12). In contrast, ChIP-seq validated targets of FOXO3/IRF3[74,75] (Fig. 2d, Table S12) and type I IFN regulated genes[76] (Table S12) that were correlated with higher inducible HIV included genes involved in promoting HIV replication (AFF1[77], MT2A[78], DARC[79]), induction of latency

(BRD4[80], GMPR[81]) and persistence of HIV infected cells (BCL2L1[82]). These data show that lower levels of IRF7 as a consequence of higher IRF3/FOXO3 lead to dysregulated antiviral innate immunity and heightened inducible HIV in the senescent INRs.

**Signalling via TGF-β is a hallmark of senescent INRs and directly drives HIV latency in memory CD4 T cells in vitro**

Transcriptional profiling identified the TGF-β pathway (PML, SMAD2/3; Fig. 1e, Supplementary Fig. 2c,d) as a driver of cellular senescence in our cohort. T-regulatory cells (Tregs) that express Forkhead box P3 (FOXP3) are a primary source of TGF-β and are critical in maintaining immune homeostasis[83,84]. To ascertain the role of Tregs in driving senescence and HIV persistence, we developed a high-dimensional flow cytometry panel (Table S13) to quantify master regulators of Treg function (SATB1 and FOXP3[85]), to discriminate differentiated Tregs (CD45RA, CD49B and CD39[86]), to assess proliferation (Ki67) and to measure the capacity to activate TGF-β by expression of GARP and/or LAP (two molecules involved in promoting latent (LAP) or activated (GARP) forms of TGF-β[87,88]). Density projection of CD4 T cells expressing these markers on a two-dimensional UMAP (https://arxiv.org/abs/1802.03426) revealed distinct distribution of cells within Senescent-INRs (Fig. 3a). Next, we used unbiased clustering[89] (Fig. 3a) to identify three clusters of CD4 T cells (Fig. 3b, Table S14) that were enriched in senescent INRs (vs IRs) and expressed markers characteristic of TGF-β producing Tregs (FOXP3, CD25, lowCD127,

LAP and GARP) (Fig. 3c). The most abundant of these clusters (i.e. Cluster 7; 0.81-5.34% of CD4 T cells) also showed an effector Treg phenotype (high CD39, low CD45RA) and low levels of markers that define IL-10 producing Tregs (Tr1) (i.e. CD49B, LAG3[90]) (Fig. 3c).

We assessed the systemic impact of these Treg clusters by studying their association with groups of cytokines that define the host plasma milieu. To this end, unsupervised clustering of the 43 different plasma cytokines using independent methods (k-means, and hierarchical clustering[91]) was performed and revealed four clusters of plasma cytokines across all subjects of the SCOPE cohort (Table S14, Supplementary Fig. 4a,b,c). Of these clusters, the overall centroid score of cytokine "Cluster 3" showed significant association with the 352 classifier genes in Senescent-INRs (NES = 2.9, FDR = 0, Supplementary Fig. 4d, Table S15). No significant correlation between cytokine Cluster 3 centroid score and the classifier genes was observed INR-Bs or IRs (Table S15). Cluster 3 included several anti-inflammatory cytokines: TGF-β1, TGF-β2 (triggers of the TGF-β pathway[92]), IL13 (known to inhibit inflammatory cytokine production[93]), KC/GRO (CXCL1) (another anti-inflammatory cytokine[94]), VEGF (the inhibitor of apoptosis[95]) as well as growth and homeostatic cytokines (IL3, IL7)[96,97] (Supplementary Fig. 4c). Importantly, our data showed that frequencies of GARP+ Tregs (Treg cluster 7) were univariately correlated with members of cytokine cluster 3 (Fig. 3d, including TGF-β2 and IL-7; Table S16), and were associated with an enrichment in the classifier genes that were predictive of

inducible HIV (significantly overlapped leading edge genes represented in Fig. 3e, Table S17).

The downstream impact of heightened frequencies of TGF-β secreting Tregs on HIV latency was confirmed by the observation that ChIP-Seq validated SMAD2/3 targets[98] and HDAC1/2 targets (induced after siRNA knockdown)[99] were more abundant in subjects with the highest inducible HIV and frequencies of GARP+ Tregs (Fig. 3e). These genesets were characterized by enhanced expression of genes that activate latent TGF-β (i.e. FURIN[100]), restrict T cell differentiation (LGAL3; inducer of TGF-β driven activation of β-catenin[101,102]) and reduce cell cycling (i.e. GADD45A[103]). The co-operation between HDAC1/2 targets (i.e. BCL2L1, FKBP1B) and SMAD2/3 can restrict chromatin accessibility and induce cell quiescence (Fig. 3e).

**Microbial metabolites drive Treg differentiation and TGF-β production in Senescent-INRs**

Bacterial translocation is a known feature of HIV pathogenesis in ART-treated and untreated HIV-infected subjects[104]. INRs of the CLIF cohort (Supplementary Fig. 1m) exhibited significantly higher levels of sCD14 (marker of disease progression and leaky gut[105]) compared to IRs. However, Senescent-INRs and INR-Bs showed comparable sCD14 levels (Supplementary Fig. 1m, p=0.567). We therefore monitored for qualitative changes in the microbiome (prevalence of different bacterial quasispecies), which could be associated to the different transcriptomes, cytokine environments and T-cell subset distributions

that distinguished the two groups of INRs. Of note, several bacterial metabolites produced by commensal bacteria (i.e. acetates, butyrates, and propionate) have been shown to regulate Treg differentiation[106,107] and T-cell homeostasis. We examined the host microbiome of INRs using PathSeq[108,109] to obtain an overall quantitative and qualitative measure of the viral, bacterial, fungal, parasite and helminth burden in subjects of the SCOPE cohort. We performed mass spectrometry (MS) to measure plasma metabolite levels and to identify bacterial metabolites that would be associated (*ex vivo*) or could trigger (*in vitro*) Treg differentiation, TGF-β production or senescence, and monitor their association with Treg frequencies and levels of senescent cytokines.

Using PathSeq, we first observed higher Shannon diversity (Supplementary Fig. 5a) and increased abundance of the Firmicute phylum in the Senescent-INRs (Fig. 4a, p=0.05). All other bacterial phyla that were detectable by Path-Seq were not significantly different in abundance between Senescent and inflammatory INRs (Fig. 4a). Furthermore, we observed that overall species diversity of the Firmicute phylum (assessed by the Bray-Curtis dissimilarity statistics[110]) significantly distinguished the two groups of INRs (Supplementary Fig. 5b). To granularly assess the differences between the Senescent-INRs and inflammatory INRs, we generated a NCBI Taxonomy tree (newick tree files created using the phyloT - http://phylot.biobyte.de - and visualized using the iTOL - http://itol.embl.de - web-interfaces, respectively) of all bacterial genera with median abundance > 0 TPM in either of the three groups in the SCOPE cohort. We observed that the Senescent-INRs (vs inflammatory INRs) were significantly

enriched in genera from all major phyla. These included general like *Moraxella*/*Bradyrhizobium* (Proteobacteria), and Actinobacteria like *Kytococcus*/*Mycobaacterium* (Fig. 4b). More importantly, we observed that the abundance of *Lactobacillus* (an anti-inflammatory genus and a member of the Firmicute phylum – i.e. phylum higher in Senescent-INRs) and the Lactobacillaceae family were significantly correlated with inducible HIV levels (Fig. 4b) (Supplementary Fig. 5c). Altogether, these data support microbiome changes specific to Senescent-INRs; and pinpoint the presence of Firmicute species in plasma that are associated with magnitude of the HIV reservoir in Senescent-INRs.

Bacterial translocation is known to have a major impact on host and systemic microbial metabolome which leads to alterations of innate immune responses. Hence, we assessed the drivers of variation in total plasma metabolite levels in the SCOPE cohort. The primary components of variation (Principal Component 1 and Principal Component 2; derived from PCA analyses of ~750 detectable metabolites) were associated with inducible HIV (PC2 vs TILDA p-value <0.05; Fig. 4c) and the classifier geneset (PC1 vs classifier genes that predict inducible HIV - p-value <0.05; Fig. 4c). The PCA analyses confirmed the role of the systemic metabolome in driving inducible HIV. Given the association of the higher *Lactobacilli* abundance with inducible HIV, we performed univariate analyses to identify plasma microbiome derived metabolites (i.e. short-chain fatty acids and bile acids) that correlated with inducible HIV levels, These metabolites included members of the liver-biliary axis (i.e. primary

liver/bile metabolites like bilirubin, biliverdin, cholate and glycol-beta-muricholate; secondary liver/bile metabolites like ursodeoxycholate), carnitine derivatives and members of the hydroxybutyrate family (i.e. α-ketobutyrate and hydroxybutyryl carnitine) (Figure 4d). Several, but not all, of these metabolites were also associated with SMAD2/3 and HDAC1/2 target genesets, C7 GARP+ Tregs and inducible HIV (Fig. 4d). Importantly, this analysis showed that α-ketobutyrate was significantly correlated ($p = 0.089$) with C7 GARP+ Treg frequencies, emphasizing the association between Tregs, metabolome and inducible HIV. Together these data indicate that microbial metabolites and the microbiome could constitute integral components of the mechanisms that fuel HIV persistence.

**TGF-β production resulting from alpha-ketobutyrate driven Treg differentiation causes increased HIV latency in vitro**

To establish a causal link between butyrate metabolites on GARP+LAP+ Treg differentiation, we used the approach described by Ohno and Rudensky[106,107]. Briefly, we assessed the impact of increasing concentrations of alpha-ketobutyrate on sorted naïve CD4 T-cells obtained from uninfected subjects in the presence of IL-2, anti-CD3/28 and/or TGF-β stimulation (Fig. 5a). TCR stimulation in the presence of TGF-β led to a profound increase in frequencies of FOXP3+ cells[111]. Addition of increasing concentrations of alpha-ketobutyrate in the absence of TGF-β preferentially led to the differentiation of naïve T-cells into GARP+FOXP3+ cells (Fig. 5a,b). Of note, TGF-β synergized

with alpha-ketobutyrate to drive the differentiation of GARP+ Tregs (Supplementary Fig. 5d). This was supported by increased TGF-β1 cytokine secretion in the supernatant of naïve CD4 T-cell cultures stimulated with alpha-ketobutyrate (Fig. 5c). Furthermore, increasing concentrations of alpha-ketobutyrate led to significantly heightened levels of PD1 expressing CD4 T-cells (Fig. 5d). alpha-ketobutyrate induced GARP+ Treg (TGF-β producing) and PD1+ CD27+ phenotypes were also associated with loss of effector function - shown by reduced production of T- helper cytokines (i.e. IL17A, IFN-γ, IL9) (Supplementary Fig. 5f). Altogether, these data indicate that in addition to enhancing Treg differentiation, the abundance of hydroxybutyrates could drive the upregulation of TGF-β processing/suppressive activity of Tregs. The latter, as shown in Fig 3, are critical for the maintenance of the HIV reservoir.

To provide evidence for a direct role of TGF-β in the induction of HIV latency, we developed an *in vitro* culture model[112] where TGF-β was added to HIV-infected CD4 T-cells. Increased numbers of CD4 T-cells with integrated proviral DNA were observed after 14 days of culture with increasing concentrations (0 to 50 ng/mL) of TGF-β ($r= 0.7$; $p = 0.009$) (Fig. 5e), demonstrating that TGF-β contributes to heightened levels of non-replicative forms of HIV DNA and establishment of the HIV reservoir. We reversed latency in this system by stimulating CD4 T cells induced to become latent by addition of increased concentrations of TGF-β ( 0.2-20ng/mL) induced with immobilized CD3 and soluble CD28 specific antibodies. Our data show that frequencies of HIV p24+ cells were significantly higher in conditions where the highest

concentrations of TGF-β (0.2-20ng/mL) were added (Fig. 5f). Altogether, these *in vitro/ex vivo* observations validate the critical role of TGF-β producing Tregs in driving the mechanistic establishment and/or maintenance of the HIV reservoir.

**Increased frequencies of senescent PD1+ TCMs with impaired mitochondrial function drive lower CD4 counts in the senescent INRs**

Given that signaling via TGF-β/FOXO3 axis drives surface PD1 expression[113], cell cycle arrest and impaired metabolism[114] - we used high-dimensional flow cytometry, transcriptomics and cytokine data to identify the cellular drivers of immune reconstitution or lack thereof ; we hypothesized that the upregulation of protein levels of PD-1 and reactive oxygen species (ROS; measured by intracellular CellROX staining) in CD4 T-cells could be associated with lower CD4 counts[115]. Using the analytical strategy described in the above section, we identified a cluster of $PD1^{hi}$ $ROS^{hi}$ CD4 central memory T-cells (TCM, Cluster 9; Fig. 6a,b,c, Table S18) that was uniquely up-regulated in Senescent-INRs (i.e. not higher in INR-Bs vs IRs) when compared with IRs. Conversely, a cluster of $PD1^{hi}$ $ROS^{lo}$ CD4 effector memory T-cells (TEM, Cluster 17; Fig. 6a,b,c, Table S18) was uniquely upregulated INR-Bs. In line with previous studies[21,116], these data confirm the negative impact of PD-1 expression on T-cell homeostasis and immune reconstitution.

A granular assessment of the plasma cytokine milieu revealed that levels of cytokines that drive effector differentiation (i.e. IL-15[112]) were directly correlated with increased frequencies of $PD1^{hi}$ $ROS^{lo}$ CD4 TEMs in INR-Bs (vs

IRs) ($p<0.05$, Table S19). Whereas, increased frequencies of PD1$^{hi}$ ROS$^{hi}$ CD4 TCMs in the senescent INRs (vs IRs) were directly associated with higher plasma IFNα/IL-29 (type I and III IFN; two cytokines that drive PD1 expression[117] and lower levels of fractalkine (a chemotactic factor downstream of IL-15 driven effector differentiation[118]) (Fig. 6d), indicating that unlike in the IRs, CD4 TCMs in senescent INRs express PD1, have impaired mitochondria and are stalled at the TCM stage. Evidence of impaired metabolism and cellular senescence in these PD1$^{hi}$ ROS$^{hi}$ TCMs was obtained by identifying metabolic and cell cycling genesets that associate with this cell subset (in senescent INRs vs IRs). Our data show that individuals with higher levels of PD1$^{hi}$ROS$^{hi}$ TCMs had poor mitochondrial metabolism profiles (higher ROS and lower oxidative phosphorylation) (Fig. 6e, Table S20), reduced cMyc activity (Fig. 6e, Table S20) and increased expression of genes that drive cellular senescence (Fig. 6f, Table S20). Here, expression of catalase (CAT), peroxidins (PRDXs) and cell cycle inhibitory genes (i.e. CDKN2D) was increased in senescent-INRs; while genes that regulate oxidative phosphorylation (NDUFs and COXs) and MYC target genes that drive transcription (RPSs)/translation (EIFs) were expressed at lower levels. Concurrently, frequencies of PD1$^{hi}$ROS$^{hi}$ TCMs were also associated with higher cellular senescence pathways downstream of impaired mitochondrial activity (i.e: oxidative stress induced senescence, SASP) (Fig. 6f, Table S20), and were observed to be the highest in subjects with higher inducible HIV (Fig. 6e, Table S20). Altogether, these results indicate that PD1 expressing TCMs with impaired mitochondrial metabolism are senescent cells lacking the capacity to

cycle and to differentiate into effector cells, thereby driving low CD4 cellular counts.

**Integrated multi-omic model highlights cellular and molecular effectors of senescence as drivers of HIV persistence and lack of immune reconstitution**

Using unbiased and holistic approaches, we highlighted gene expression profiles, cytokines and T-cell subsets that were independently associated with HIV persistence and lack of CD4 reconstitution in the Senescent INRs. To investigative the interplay of pathways driving these cellular and viral phenotypes and HIV persistence, we integrated multi-omic signatures above (Figures 2, 3 and 4, Table S21) across all subjects (n=41). Our data show that the levels of inducible HIV were positively correlated to senescence as characterized by impaired metabolism[115] (increased ROS, decreased OX/PHOS), reduced transcription/translation (RPSs and EIFs) downstream of poor cMYC activity and TGF-β signaling (SMAD 2/3 target genes) (Fig. 6g, Supplementary Fig. 6). The induction of these cascades was consistent with higher expression of the transcriptional repressors that induced senescence (i.e. FOXO3, FOXO4)[42], and with lower expression of master regulators of innate antiviral activity (i.e. IRF7 and restriction factors of viral replication) (Fig. 6, Supplementary Fig. 6).

Lack of CD4 reconstitution was driven by senescence and impaired metabolism and as well by higher frequencies of GARP+ Tregs, increased expression of SMAD2/3 targets (including FURIN, SMOX) and heightened levels

of IRF3 mediated type I IFN production and induced genes (i.e. AFF, DARC, GUK1) (Fig. 6, Supplementary Fig. 6). These pathways (IRF3 and SMAD2/3) were associated with increased frequencies of cells with inducible HIV RNA. Our integrated analysis indicates that most pathways that drive HIV persistence were negatively associated with the recovery of CD4 numbers upon cART initiation in senescent subjects. Altogether, the integration of these multi-omic results provided further evidence for the direct interplay between Treg frequencies, TGF-β production, heightened FOXO3 expression, interferon signaling, establishment of cellular senescence, impaired T cell homeostasis, and quiescent cellular and metabolic state as conditions that favor the maintenance of HIV reservoir in the unique senescent INR population described here.

## Discussion

Studies of two independent cohorts of HIV infected cART-treated subjects described in this manuscript identified cellular senescence as a mechanism that underlies HIV persistence and failure to reconstitute CD4 T-cell numbers. Our results provide a mechanistic framework where anti-inflammatory cytokines, TGF-β, VGEF and IL-13, trigger the upregulation of a transcriptional network that drives Treg differentiation. Our integrated analysis associates the increased frequencies of TGF-β producing Tregs and metabolically impaired PD-1+ TCMs with inducible HIV and poor CD4 immune reconstitution. This establishes a systemic increase in senescent cells, which is associated with HIV persistence. Notably, we demonstrate that senescence triggered by gut dysbiosis as

characterized by a selective increase in the abundance of firmicutes and their metabolites including butyrates is observed only in Senescent-INRs. To our knowledge, this is the first report that provides a direct mechanistic role for the microbiome and bacterial metabolites in HIV persistence. These metabolites drive the differentiation of TGF-β producing Tregs and PD-1 expression, both of which are associated with the highest frequencies of cells with inducible HIV RNA and poor CD4 immune reconstitution.

Transcriptional profiling data underscore the critical role of signaling downstream of TGF-β that activates SMAD2/3 and the transcriptional repressor FOXO3 to induce senescence. Senescent-INRs, in our study, show increased expression of several molecules which are features of senescent cells. They include: FOXO4, a TF specifically upstream of the induction of senescence[42]; PML, triggers cellular quiescence, HIV latency and is also downstream of TGF-β; as well as molecules with anti-apoptotic activity (BCL-2, BCL-xL[119]) that ensure the survival of senescent cells. Moreover, features known to be characteristic of the senescent phenotype[68] including decreased cell metabolism (oxidative phosphorylation), upregulation of transporters of ions/salts and inhibition of pathways that regulate cell cycle entry were characteristic of peripheral immune cell subsets in senescent INRs.

In addition to their senescent profile, these subjects are also characterized by the downregulation NF-κB and IRF7, two TFs that control pro-inflammatory pathways. The decrease in the transcriptional activity of IRF7 likely results from increased activity of the transcriptional repressor FOXO3[72] and results in

diminished expression of IRF7 target genes (including innate antivirals like CD74, TRIM4, BST2, and APOBEC3G)[71]. This impaired expression of intrinsic mechanisms of innate antiviral immunity may in part explain the increased frequencies of cells with inducible HIV in Senescent-INRs. Interestingly, while expression of FOXO3 led to poor IRF7 transcriptional activity, it was associated with an increase in IRF3 transcriptional activity. Target genes of IRF3 and FOXO3 include TFs that regulate senescence associated pathways (i.e. SOX4, a TF that inhibits T-helper cell development and is triggered by TGF-β[120]) and RIOK3, an inhibitor of NF-κB[58]. Hence, while IRF7 inhibits the anti-inflammatory activity of FOXO3[121], IRF3 will inhibit TNF induced NF-κB activation[122] and as well trigger the inflammasome[123]; both( decreased IRF-7 and NFκ-b) could lead to poor immune responses and heightened HIV persistence. A senescent phenotype downstream of IRF3 driven type I IFN production and signaling to induce IFN stimulated genes (ISGs) has been described in aging cells[124,125], and could contribute to CD4 T cell senescence shown in our study. A role for type I IFN signaling and ISGs in enhancing HIV persistence was confirmed by recent studies in humanized mice whereby the blockade of IFN-α/β receptor (IFNAR) led to a reduction of HIV-1 reservoir size and to the restoration of antiviral immunity[126]. These findings are supported by the significant correlation between a type I IFN signature and HIV reservoir observed in our data. Together, this previously undescribed transcriptional cooperation between FOXO3 and IRF3 suggests novel mechanisms to understand the transcriptional regulation of

senescent immune cells associated with the lack of CD4 T-cell immune reconstitution and HIV persistence.

FOXO3 is also known to transcriptionally regulate genes involved in redox balance and anti-oxidant defenses[127,128]. Our data highlights FOXO3 target gene SNCA (α-syneculin[129]), mitophagy related genes (PINK1 and PARKIN[130]) and GABARAPL2 (the autophagy regulatory gene[131]) and SLCs that are upregulated in Senescent-INRs as part of a mitochondrial damage response. The latter may help sustain processes responsible for oxidized lipid biogenesis and ROS production. Furthermore, the downregulated oxidative phosphorylation and glycolysis observed in these subjects could be driven by the upregulation of FOXO3 target genes CYB5R3 and peroxiredoxins (PRDX2, PRDX3, PRDX5) involved in maintaining redox balance and anti-oxidant defense[132]. PRDXs, specifically PRDX2, can get oxidized by ROS due to a lack of electron transport in the mitochondria[132]. Oxidized PRDX2 can then activate kinases including p38 that initiate stress responses and protect cells from ROS mediated cell death, thereby supporting the survival of senescent cells[132]. In our data, the expression of genes associated with ROS production (Catalase and superoxide dismutase) and oxidized lipid biogenesis (CD36) were positively correlated with HIV reservoir size confirming the link between senescence and HIV reservoir establishment.

The aforementioned senescence associated impaired mitochondrial metabolism was found to be enhanced in a subset of PD-1+ TCMs that were enriched in the senescent-INRs. The heightened expression of PD1+ and poor metabolic profile is suggestive of an accumulation of these cells in an early

memory differentiated state specifically within Senescent-INRs. We and others have shown that PD-1 expression leads to the accumulation of cells in the early memory differentiated stage[133] while the genetic ablation of PD-1 resulted in their increased differentiation to the effector memory stage[134]. Additionally, T-cells from Senescent-INRs show downregulation of pro-apoptotic pathways and increased expression of anti-apoptotic molecules that potentially restrain cell turnover[135]. The blockade of T-cell differentiation from TCM to TEM as a consequence of PD-1 expression could explain the low CD4 T cell numbers observed in these subjects (in contrast to TEM cells which are known to expand and proliferate thereby increasing CD4 cell numbers[112]). This limited T-cell turnover in Senescent-INRs could explain the high frequencies of cells with the multiply spiced HIV RNA, as these cells would accumulate with time.

Finally, we provide direct evidence for TGF-β producing Tregs as drivers of impaired T-cell homeostasis and HIV persistence in Senescent-INRs. In our study, the observed increase of active plasma TGF-β in Senescent-INRs is supported by the increased expression of GARP by these Tregs and activation of signal transduction cascades that lead to upregulation of FOXO3, PML and CCNT2 – all transcriptional targets of SMAD2/SMAD3 and downstream of TGF-β[35]. PML, an interferon induced anti-viral protein[136] is triggered by TGF-β[46] and establishes cellular quiescence by stabilizing FOXO3 and FOXO4[137]. This stabilized FOXO4 binds to p53 and plays a crucial role in the downregulation of pro-apoptotic machinery[42,43], a major hallmark of senescent cells. *In vitro* validation experiments, in addition to molecular and cellular mechanisms

described here implicate TGF-β in driving senescence and dysregulated type I IFN-driven pathways (via FOXO3 and IRF3) which impact T-cell homeostasis and promote HIV persistence. Further role of TGF-β in HIV persistence is provided by the presence of CCNT2 in the transcriptional network that is specific to senescent-INRs. CCNT2 and AFF1 are both components of the pTEFb complex which is critical for HIV replication[77,138]. Heightened levels of pTEFb in CD4 T cells could provide a mechanism to explain the higher levels of inducible HIV in these subjects.

In addition, to the direct impact of α-ketobutyrate on FOXO3 upregulation, we describe a Treg dependent induction of senescence in Senescent-INRs. Using both *ex vivo* and *in vitro* experiments, we demonstrate that the differentiation of naive CD4 T-cells into GARP+ Tregs by exposure to α-ketobutyrate and the prevalence of butyrate metabolism specific genes in Tregs. These novel findings confirm previous *in vivo* studies that have shown increased Treg differentiation upon exposure to butyric acid[106,107]. In our study, the observed increase of active plasma TGF-β in Senescent-INRs is supported by the increased expression of GARP by these Tregs and activation of signal transduction cascades that lead to upregulation of FOXO3 and PML. The elevated levels of butyrates can be attributed to an increase in ketone bodies (produced by the liver during periods of poor dietary intake[139]). These observations are supported by increased detection of both primary and secondary liver/bile acids in plasma from subjects with higher inducible HIV (i.e. senescent INRs). Alternately, the increase in plasma butyrates could also be

attributed to increased bacterial metabolism[107] and is supported by the positive association of butyrates/propionates with plasma LPS and IFABP in Senescent-INRs - providing strong support for microbial translocation as a source of these metabolites. Heightened sCD14 and increased Firmicute phylum abundance (known producers of butyrates[140,141]) in the Senescent-INRs confirms a direct interplay between microbiome/metabolic changes and establishment of the Senescent phenotype. Together these data highlight the role of butyrate induced Treg differentiation[106] in TGF-β production which leads to subsequent CD4 T-cell senescence, lack of CD4 reconstitution and higher inducible HIV production. They reveal for the first time a mechanism downstream of commensal bacteria (exemplified by the association of the Lactobacillaceae family and alpha-ketobutyrate with inducible HIV) that can trigger the establishment of the HIV reservoir.

Our findings provide a strong rationale for the evaluation of senolytic drugs that would target senescent cells[142,143] as promising therapeutic interventions to reduce the HIV reservoir size. Combination interventions targeting PD-1 and TGF-β, or Senolytics, may have an improved therapeutic impact in Senescent-INRs where frequencies of PD-1+ cells are a correlate of high HIV reservoir and poor CD4 T-cell reconstitution. Interestingly, senescence and TGF-β signaling have been highlighted as hallmarks of cancer where anti-PD-1 therapy has shown efficacy[144]. Our gene classifier should help identify subjects who will not respond to checkpoint therapies in HIV and cancer[145], and could benefit from interventions that target both PD-1 and TGF-β. Senescent-INRs could also

benefit from interventions that promote T-cell differentiation (e.g. IL-15) or rescue the activity of deficient metabolic pathways allowing T-cells to overcome the senescent state and help restore both CD4 T-cell numbers and cell-mediated immunity, which could decrease reservoir size[146]. The multi-omics senescent features and the classifier gene-set described here thus represent critical tools to identify Senescent-INRs. Additionally, the integrated approach used herein holds potential for the identification of novel targets and the design of optimal trials that combine different drugs/biologics targeting senescence and exhaustion to restore immune homeostasis and eradicate HIV.

## Experimental Procedures

**Study Participants:** These studies were approved by the Institutional Review Boards at University Hospitals/Case Medical Center and the Cleveland Clinic Foundation, the Vaccine and Gene Therapy Institute of Florida and University of California San Francisco; all patients provided written informed consent in accordance with the Declaration of Helsinki. ***Cohort 1 (CLIF):*** A total of 45 immunologic non-responders (INRs) and 17 immune responders (IRs) were identified from the Cleveland Immune Failure cohort. The Cleveland Immune Failure study examined immunologic indices in healthy controls and 2 groups of patients who had been receiving ART for at least 2 years with plasma HIV RNA levels below detectable levels using routine clinical assays; typically, less than 50 copies per milliliter. Transient blips in HIV RNA levels did not exclude participation if flanked by levels below limits of detection. Immune failure patients

(INRs) had CD4 T cells <350/µL and immune success patients (IRs) had CD4 T cells >500/µL. Detailed clinical indices, anti-retroviral regimens and basic demographics are listed in Table S1. ***Cohort 2 (SCOPE):*** A total of 21 INRs and 20 IRs were identified from the University of California San Francisco (UCSF) SCOPE cohort. Viral suppression was defined by at least two longitudinal tests that showed plasma HIV RNA levels below the limit of detection using standard assays (<40 Abbot RealTime HIV-1 assay, <40 Roche COBAS® Ampliprep/COBAS® Taqman® HIV-1 Test, <50 branched DNA); these tests were done 3 months prior to and on the date of specimen collection. Subjects with viral blips below 200 copies/ml in the time period since beginning ART (or since the end of last treatment gap) 2 years prior to the specimen date were not excluded from the study. Basic demographics and clinical readouts at the time of collection are listed in Table S2.

**Microarray Pre-processing and differential gene expression analyses:** Whole blood was collected and lysed in RLT for RNA extraction (Qiagen, Valencia, CA, USA). For the CLIF cohort, cDNA obtained after reverse transcription reaction was hybridized to the Illumina Human HT-12 version 4 Expression BeadChip according to the manufacturer's instruction and quantified using the Illumina iScan System. The data were extracted using the GenomeStudio software. Similarly, the SCOPE cohort samples were run using an Affymetrix microarray system. Detailed analysis of the genome array output data was conducted using the R statistical language[147] and the Bioconductor

suite[148]. Arrays displaying unusually low median expression intensity and variability across all probes relative to all arrays were discarded from the analysis. Probes that do not map to annotated RefSeq genes and control probes were removed. Quantile normalization followed by a log2 transformation using the Bioconductor package LIMMA was applied to raw microarrays intensities. To determine differences between groups or against a continuous variable, the LIMMA package[149] was used to fit a linear model to each probe. A (moderated) Student's *t*-test to assess significance in difference between groups, whereas a Pearson correlation test was used to assess significance against a continuous variable. The proportions of false positives were defined using the Benjamini and Hochberg method[150]. Unless indicated otherwise, genes that satisfied FDR <0.05 were selected for data mining and functional analyses.

**Class identification, classifier training and validation*:*** *Class identification* **-** In the CLIF cohort, 62 samples were hybridized by the Vaccine Genome Research Institute Genomics Core and 61 samples made it to the downstream analysis (44 INRs and 17 IRs) after outlier identification and removal. Initial exploratory analysis (MDS plot – Fig. 1a) followed by unsupervised analysis (heatmap of top 200 most variable probes – Supplementary Fig. 1c) revealed the discovery of two INR classes in the dataset. These presence of these subject cluster was confirmed when the Gap statistic technique[91] was used on whole transcriptomic data to determine the optimal number of INR classes. The two classes of INRs were labeled: INR-A (an extreme INR phenotype, n = 15) and INR-B (INR

phenotype proximal to IRs, n = 29). Differential expression analyses (described above) were used to identify the differences between IR and INR-A, IR and INR-B and INR-A and INR-B. *Classifier training* **-** Construction of the INR-A – INR-B Classifier (Fig. 1f-g, Supplementary Fig. 3) was done using the Prediction Analysis of Microarrays for R (PAM) package from R. PAM uses the nearest shrunken centroid methodology[151]. A standardized centroid for each class (INR-A and INR-B) was computed. Ten-fold cross-validation was used to optimize the shrunken criteria. The-shrunken criteria that resulted in the lowest cross-validated misclassification error rate was used to generate the optimized classifier on the CLIF cohort (training cohort). Supplementary Fig. 3a represents training of the 352 features used to build the classifier on the CLIF dataset (see features in Table S11). *Classifier validation* - Testing the classifier on the validation set (SCOPE cohort) is illustrated in Fig. 1f-g. There are no "true" labels that define INR-A or INR-B in the SCOPE testing cohort as they both fall under the label immune non-responders. In order to assess the performance of the classifier on the validation set, the two INR groups had to be defined. An unbiased unsupervised approach similar to the one used to discover the classes in the CLIF cohort was applied on the normalized Affymetrix microarray data of the SCOPE cohort. Supplementary Fig. 3b shows a heatmap representation of the expression of the top varying probesets (top 200 variant transcripts) on the full SCOPE dataset. Two clusters of INRs were identified using unsupervised clustering (color of branches in Supplementary Fig. 3b denote the discovered classes). This discovery allowed us to label the two subgroups of INRs in the

SCOPE cohort. The 352-gene classifier was then applied on the SCOPE cohort. ROC analysis was then used to assess the accuracy (80%) of the classifier on the SCOPE cohort (Fig. 1f). Throughout the study the INR classes in the SCOPE cohort were defined based on the 352-gene classifier (PCA representation of classifier genes in Fig. 1g). Whole transcriptome profiling of classifier defined classes in the SCOPE cohort is shown in Supplementary Fig. 3c.

**Pathway analyses.** *Gene Set Enrichment Analysis (GSEA)*[27] throughout the study was performed to assess enrichment in pathways that discriminated between classes (Fig. 1b-d), were associated with inducible HIV (Fig. 2d and Fig. 3e), were associated with Treg cluster 7 (Fig. 3e), were associated with low CD4 counts (Fig. 4e) and were associated with $PD1^{hi}ROS^{hi}$ CD4 TCMs (Fig. 4e, f). Briefly, the whole transcriptome probe-set was collapsed to genes by assessing the most variable probes, pre-ranking of this collapsed transcriptome was done (t-statistic) and enrichment of various genesets was tested after running 1000 permutations of enrichment. *Hallmark MSigDB database*[25] (v 7.0) was used to identify pathways that differentiated IR, INR-A and INR-B subjects in the CLIF cohort (Fig. 1b). This database was also used to assess pathways driving lower CD4 counts in the Senescent INRs (Fig. 4e, Table S20), frequencies of $PD^{hi}ROS^{hi}$ CD4 TCMs in the Senescent INRs (Fig. 4e, Table S20) and inducible HIV (Table S12) in the SCOPE cohort. *Cell subset deconvolution*: Major peripheral immune subset-specific gene expression signatures were obtained from Nakaya et. al[31], and their enrichment was assessed (Fig. 1c, Table S7). In

addition, data from a comprehensive study with sorted hematopoietic stem cell-subsets[32] was used to generate gene-signatures specific to 38 unique subsets (each subset-specific signature was defined as genes found to be >2-fold higher with $p<0.05$ than in the pool of all other subsets) (Table S8). These signatures were used to assess variations in differentiated CD4 and CD8 T cell subsets in whole blood in Fig. 1d. *Custom genesets*: Senescence based gene signatures were extracted from Reactome database in MsigDB v7.0's c2 module (Table S20 lists of all pathways, Fig. 4f). Genesets specific to host antiviral restriction factors (Fig. 2d, Table S12) and IFN signaling (Fig. 1b, Fig. 2d, Table S12) were extracted from Abdel-Mohsen et. al[73] and Interferome database[76], respectively. SMAD2/3 and HDAC1/2 genesets were also extracted from MsigDB's c2 module by searching this module for 'SMAD2', 'SMAD3', 'HDAC1' and 'HDAC2'. The results of these analyses are shown in Fig. 3e (Table S17). The enrichment of the 352-geneset classifier gene-set was done where specified (Fig. 1i, Fig. 3e, Table S17). *Leading Edge overlap and Sample Level Enrichment Analyses (SLEA)*: Leading edge genes from the analyses above were overlapped (significance of overlap was determined using Fischer's Exact test; $p < 0.05$) to define new gene-lists that were associated with both "inducible HIV and Treg cluster 17" (Fig. 3e, Table S17) OR "CD4 counts and $PD1^{hi}ROS^{hi}$ CD4 TCMs" (Fig. 4e, Table S20). SLEA was used to generate a z-score normalized value for genelists in Fig. 2d, 3e, 4e and 4f. These scores were correlated (spearman correlation test) with each other, minor (Treg frequencies and $PD1^{hi}ROS^{hi}$ CD4

TCM frequencies) and major (inducible HIV and CD4 counts) outcomes to generate the integrated correlation network model in Fig. 5 and Table S21.

**Over-representation, gene-mania/correlation networks and transcription Factor (TF) analyses*:*** *Over-representation* of Reactome pathways in genes that are upregulated in the INR-As (vs both IRs and INR-Bs; FDR < 0.05) was assessed using the ClueGO plug-in from Cytoscape (http://apps.cytoscape.org/apps/cluego[37]), and the results of these analyses were corrected for multiple comparisons using Bonferroni step-down correction (Fig. 1e, see Table S10 for full results). *GeneMania and correlations network:* GeneMania Networks[152] (http://genemania.org) were plotted to represent co-expression of genes (Supplementary Fig. 2b and Supplementary Fig. 2d). Overlap between the genes included in the networks and Gene Ontology (GO) biological process was assessed using a Fisher exact test. Correlation networks in Fig. 3d, Fig. 4d, Fig. 5a and Supplementary Fig. 5a are based on significant Spearman correlations (p-value < 0.05) between the features indicated in each figure. All networks were plotted using the Cytoscape Plug-in – specific value/category for all nodes and edges are available upon request. *Transcription factor list* was generated by combining genes with known ChIP-seq validated targets from CHEA[74] (https://amp.pharm.mssm.edu/Harmonizome/resource/ChIP-X+Enrichment+Analysis) and ENCODE (https://amp.pharm.mssm.edu/Harmonizome/dataset/ENCODE+Transcription+F

actor+Targets) databases. The expression of these TFs was tested in INR-As vs INR-Bs or IRs (Supplementary Fig. 2c). IRF3 targets and FOXO3 targets were extracted from the ENCODE and CHEA databases cited above, respectively. Co-expressed targets of FOXO3 and IRF3 that correlated with inducible HIV are represented in Fig. 2d (Table S12).

**Cell Preparation and flow cytometry:** PBMCs were prepared from whole blood by ficoll-hypaque density sedimentation and cryopreserved in 10% dimethyl sulfoxide and 90% FBS. Two panels to evaluate Treg function and mitochondrial dysfunction in memory CD4s were run using previously titrated antibodies summarized in Table S13, S14 and S18. The cells were surface stained for 20 minutes in the dark at room temperature, washed, fixed and permeabilized using the eBioscience™ Foxp3 / Transcription Factor Staining Buffer Set (Cat# 00-5523-00), as per manufacture instructions. Intracellular staining was performed in Perm-Wash provided by the kit for 45 minutes at $4^{o}$C. Samples were washed and re-suspended in staining buffer for acquisition. ~500,000 live-gated events were collected per sample on the LSRII flow cytometer (Becton Dickinson, San Jose, CA). Initial data cleaning and pre-gating was done using the Flow-Jo X software (TreeStar, Ashland, OR) (Supplementary Fig. 6). Briefly, lymphocyte gate based on FSC-A and SSC-A was defined. Single cells were then selected using a FSC-A x FCS-H gate. Live CD3+CD4+CD8- cells were gated and exported for unbiased clustering analyses (Supplementary Fig. 6). For both panels, projection of the density of cells expressing markers of interest (Table S14, Table S18)

were visualized/plotted on a 2-dimensional UMAP (https://arxiv.org/abs/1802.03426, https://github.com/lmcinnes/umap). Clusters of cells using the RPhenograph package after (https://github.com/jacoblevine/PhenoGraph) after concatenating all samples per panel and bi-exponentially transforming each marker (Fig. 3a and c, Fig. 4a). Differences in cluster frequencies per group and MFI for each marker per cluster, for each are shown in Table S14 and Table S18.

**Measurement of inducible reservoir:** The HIV reservoir was measured using the tat/rev induced limiting dilution assay (TILDA). This method measures the inducible reservoir by RT-qPCR using the limiting dilution assay previously described[70]. Briefly, 2 x$10^6$ CD4 T cells from 20x$10^6$ PBMCs were enriched using a CD4 T cell negative selection kit (Stem Cell). These cells were stimulated with PMA/Iono (100ng/ml and 1µg/ml, respectively) for 12 hours. A total of 744,000 stimulated cells were then distributed in a limiting dilution in a 96-well plate and tat/rev msRNA expression was directly quantified (without prior RNA extraction) by semi-nested real-time PCR[70]. The frequencies of cells positive for inducible tat/rev msRNA per $10^6$ CD4+ T cells were determined by using the maximum likelihood method (http://bioinf.wehi.edu.au/software/elda) (Fig. 1h,i).

**Measurement of plasma biomarkers:** Plasma IL-6 in the CLIF cohort was measured using a high sensitivity ELISA kit for human IL-6 (Quantikine HS) from R&D Systems (Minneapolis, MN). Plasma levels of IL-7 in the CLIF cohort were

measured by high sensitivity IL-7 ELISA (Quantikine HS, R&D Systems, Minneapolis, MN). Plasma levels of IP10 in the CLIF cohort were measured by IP-10 ELISA (Quantikine, R&D Systems, Minneapolis, MN). Plasma levels of human Intestinal Fatty Acid Binding Protein (I-FABP) in the CLIF cohort were measured using a DuoSet ELISA Development kit from R&D Systems (Minneapolis, MN) following the manufacturer's protocol. Levels of D-dimers in the CLIF cohort were measured using the Asserachrom D-DI immunoassay (Diagnostica Stago, Asnieres France) (Supplementary Fig 1. d-m). *Measurement of LPS in the CLIF cohort:* Plasma samples were diluted to 10% or 20% with endotoxin free water and then heated to 85$^{o}$C for 15 minutes to denature plasma proteins. We then quantified plasma levels of LPS with a commercially available Limulus Amebocyte Lysate (LAL) assay (QCL-1000, Lonza, Walkersville, MD) according to the manufacturer's protocol. *Multiplex ELISA (Mesoscale):* U-PLEX assay (Meso Scale MULTI-ARRAY Technology) commercially available by Meso Scale Discovery (MSD) was used for plasma cytokine detection. The assay was performed according to the manufacturer's instructions (https://www.mesoscale.com/en/technical_resources/technical_literature/techncal_notes_search). 25µL of plasma from each donor was combined with the biotinylated antibody plus the assigned linker and the SULFO-TAG™ conjugated detection antibody; in parallel a multi-analyte calibrator standard was prepared by doing 4-fold serial dilutions. Both samples and calibrators were mixed with the Read buffer and loaded in a 10-spot U-PLEX plate, which was read by the MESO

QuickPlex SQ 120. The plasma cytokines values (pg/mL) were extrapolated from the standard curve of each specific analyte. Cytokine clustering (Supplementary Fig. 4a,b,c) was performed using independent methods: (gap statistic method to identify and characterize optimal number of k-means clusters, and hierarchical clustering (ward clustering; Euclidean distance[91]).

**Microbiome and Metabolome.** *Pathogen-Sequencing:* Host microbiomes were analyzed using the PathSeq technology. PathSeq involves deep sequencing of fragmented RNA and DNA (~100-1000 bases) in plasma to obtain an overall quantitative and qualitative measure of the viral, bacterial, fungal, parasite and helminth burden. Following RNA-sequencing, the reads that map onto host genomic and transcriptomic regions were identified and removed from the dataset. De novo assembly is then performed on the remaining, non-host sequences. The assembled contigs were aligned against the comprehensive "nt" database from NCBI to determine the microorganisms for which assembled sequences were derived. We used the Bray-Curtis dissimilarity statistic [110] (Supplementary Fig. 5a,b) to assess the beta diversity and assess differential distribution of the abundance or total read counts of all phyla to species in the microbiome PathSeq data. *Metabolomics:* In collaboration with Metabolon, plasma metabolite levels for up to 1300 metabolites were measured. Metabolon uses four ultra-high-performance liquid chromatography/tandem accurate mass spectrometry (UHPLC/MS/MS) methods in a highly controlled environment to reduce noise and produce accurate results. The data generated using

UHPLC/MS/MS were referenced against a well-established library of known and novel metabolites.

**Treg differentiation in the presence of alpha-ketobutyrate:** Naïve CD4 T cells extracted from 8 healthy donors were isolated using the EasySep™ Human Naïve CD4+ T Cell Isolation Kit (StemCell Technologies Catalog# 19555) and differentiated in vitro using the protocols described by Rudensky[106]. Briefly, 100-200,000 naïve CD4 T cells were stimulated with Dynabeads Human T-Activator CD3/CD28 beads (ThermoFisher Scientific Cat# 111.31D, 1 bead/3 cells), 100 U/mL of IL-2 (R&D Systems Cat# 202-IL-500), 0 to 1ng/mL of TGF-β (R&D Systems Cat# 240-B-010) and alpha-ketobutyrate (Cat#, from 0mM to 5mM) for 3 days in 96-well round bottom plates. Viability post-stimulation, assessed post-stimulation by staining with a fixable viability dye, was found to be >80% in all stimulation conditions. Secreted cytokines in the supernatants were quantified using the Mesoscale Discovery platform/kits described above. CD4 T cells were intracellularly stained with the Treg phenotyping panel described (Table. S8).

**Latency establishment and reactivation assay:** The latency and reactivation assays developed by our group[112] was used to assess the latency establishment in memory CD4 T cells post-stimulation with increasing doses of TGF-β. Briefly, memory CD4 T cells spinoculated with HIV (HIV strain information listed[112]) were incubated in the presence of antiretrovirals (effavirenz , Saquinavir and Ralegravir) and latency establishment media (supplemented with 0-40ng/mL of

TGF-β) for 13 days (or when the frequencies of HIV-p24+, measured by flow cytometry, CD4 T cells was negligible). Integrated HIV DNA was measured at this stage using the protocol described[112]. Reactivation of HIV post-latency was done after stimulation with 1ug/mL anti-CD3 and 1ug/mL of anti-CD28 for ~48 hours; reactivated cells were quantified by monitoring the frequencies of p24+ cells using flow cytometry.

**Other statistical analyses:** All univariate group-differences were analyzed using a non-parametric Wilcoxon-ranked test. All univariate correlation analyses were done using a non-parametric Spearman's test. $P<0.05$ is reported as significant.

**Full Data Availability Statement:** The raw and normalized data matrices for gene expression analysis of the CLIF and SCOPE cohorts were deposited into the Gene Expression Omnibus (GEO) database; GSE143742.

**Financial support:** This work was supported by the National Institutes of Health (grants UO1 AI 105937 and RO1 AI 110334 and RO1 AI 11444201), the CWRU Center for AIDS Research (grant AI 36219), DARE (U19 AI 096109), and the Fasenmyer Foundation. Rafick-Pierre Sekaly is the Richard J. Fasenmyer Professor of Immunopathogenesis.

**Potential conflicts of interest:** All authors: No reported conflicts

**Figure Legends (F1-F6):**

**Figure 1 | Whole blood transcriptional profiling identifies a 352-geneset signature that discriminates ART-treated Senescent-INRs and is associated with high inducible HIV**. **a**, Multi-dimensional scaling (MDS) was used to reduce the Euclidean distance between whole blood samples of the CLIF cohort (n=61) into two dimensions that summarize the largest transcriptomic variation in the dataset. These two dimensions were used to represent the differences between samples (circles) in the scatter plot. Three groups were identified: IR (Grey; n=17), INR-A (Red; n=14), and INR-B (Blue, n=20). The first dimension of the plot, representing 23% of the transcriptomic variance between samples, depicts the observed transcriptomic difference between the three groups of subjects. **b-d**, Heatmaps illustrating the normalized enrichment score (NES; red (high) to blue (low) scale shown on the heatmap) of the top genesets (GSEA p-value ≤ 0.05) MSigDB: Hallmark[25] (Table S6) + Interferome TypeI/II[76] Pathways (**b**) Major immune cell Subsets[31] (Table S7) (**c**) and T cell subsets[32] (Table S8) (**d**) between *IR vs INR-A, IR vs INR-B and INR-A vs INR-B* subjects. Each row depicts a genesets, and columns represent the contrast between subject groups. **e**, Network highlighting the top biological functions associated with genes upregulated in INR-A subjects (vs IRs and INR-Bs; FDR < 0.05). As indicated, node color (dark red to orange) highlights the p-value resulting from over representation analyses, while the node size represents the number of genes per gene-set. Member genes are represented as white circles and connected to member geneset by grey edges (see Table S10 for details). Reactome database

was used to annotate the biological functions and the network was plotted using the ClueGo plug-in within cytoscape. **f**, Receiver operating characteristics (ROC) curve for the genes as predictors of INR groups. X- and Y- axes represent the False and True positive rates of prediction on the SCOPE validation dataset. A 352 gene-based classifier (see Table S11 for list of genes) trained on the CLIF cohort segregates INR-A and INR-B subjects in the SCOPE cohort with an accuracy of 81%. The classifier which was tested across different microarray platforms confirmed the heterogeneity of ART-treated INR subjects in these two independent cohorts (see methods for details on approach used to build the classifier). **g**, Multi-dimensional Scaling was used to summarize the variation of the 352 genes of the classifier among subjects of the SCOPE cohort (n=41). Senescent-INR (Red), and INR-B (Blue) were identified along the first dimension of the MDS plot. **h**, Jitter plot illustrates significantly (Wilcoxon rank test p-values shown on plot) higher levels of inducible HIV (measured by TILDA[70]) in Senescent-INRs compared to IRs and INR-Bs. **i**, Heatmap illustrating the leading edge from the gene-based classifier that predict inducible HIV among subjects of the SCOPE cohort (n = 41; linear regression: classifier genes ~ inducible HIV, GSEA: NES=3.22, $p<0.0001$). Rows represent the z-score normalized genes in the leading edge and columns represent samples of the SCOPE cohort: IR (Grey), Senescent-INR (Red), and INR-B (Blue). As indicated in the legend, each row is z-score normalized. The magnitude of CD4 counts (purple) and the size of inducible HIV (black) are plotted as annotations at the top of the heatmap, as

indicated. Selected genes are highlighted in boxes to the left of the heatmap (See Table S12 for full leading-edge gene list).

**Figure 2 | Altered interferon signaling drives inducible HIV in Senescent INRs**. **a-b**, Jitter plot illustrating significantly elevated levels of IRF3 expression in Senescent-INR subjects (a), scatter plot showing a significant positive correlation between FOXO3 and IRF3 expression (b) and diminished levels of IRF7 expression in Senescent-INR subjects (c) of the CLIF cohort. A Wilcoxon rank sum test was used to assess significant differences between groups (**$p < 0.01$, ****$p<0.0001$), and a spearman correlation test was used to assess significance of association (p-value and rho indicated on plot). **d**, Heatmap illustrating the dichotomy of interferon signaling driving the HIV reservoir size among subjects of the SCOPE cohort (Top block: linear regression: Restriction Factors ~ inducible HIV, GSEA: NES= - 1.6, $p<0.05$, Bottom block: positive correlates of inducible HIV that are targets of FOXO3 and IRF3). Rows represent the z-score normalized genes in the leading edge and columns represent samples of the SCOPE cohort: IR (Grey), Senescent-INR (Red), and INR-B (Blue). As indicated in the legend, each row is z-score normalized. The magnitude of CD4 counts (purple) and the size of inducible HIV (black) are plotted as annotations at the top of the heatmap, as indicated. See Table S12 for full gene-lists.

**Figure 3 | TGF-β signaling cascade exemplified by an increase in SMAD2/3 targets and an increase in TGF-β producing Tregs is associated with an**

**increase in inducible HIV and drives latent HIV in vitro. a**, Treg panel clusters determined using the PhenoGraph method (see Table S14) and visualized using the Uniform Manifold Approximation and Projection (UMAP) analysis were used to represent the distribution of Treg phenotypes in total CD4 T cells (using high-dimensional flow cytometry) in IR, Senescent-INR, and INR-B subjects of the SCOPE cohort. FOXP3 expressing clusters that were abundant in Senescent-INR subjects (vs IRs) are highlighted on the UMAP plot (See Table S14 for details). **b**, Violin plot illustrates significantly higher frequencies of Treg cluster 7 (in total CD4 T-cells) in Senescent INR subjects. Wilcoxon rank test was used to assess significance and p-values are indicated on the plot. **c**, Heatmap highlighting the differences in markers between clusters. FOXP3+ clusters that were abundant in senescent INRs (i.e. clusters 1,7, and 19) are highlighted. Each column is z-score normalized value of the mean fluorescence intensity (raw intensities for every cluster can be found in Table S14). **d**, Network highlighting a positive correlation between FOXP3 expressing Treg clusters that are abundant in Senescent INRs (Clusters 1, 7 and 19) with plasma cytokine levels (of cytokines in cytokine cluster 3; See Supplementary Fig. 4b) measured on the same subjects. Triangular blue nodes depict plasma cytokines IL-7, TGF-β2, KC.GRO and VEGF; red edges highlight a positive correlation between those cytokines and the Treg clusters. Orange squares reflect high-density flow cytometry clusters, light blue triangles reflect plasma cytokine levels, red edges indicate significant positive correlation. A Spearman correlation test was used to assess significance (p-value <0.05) across IR, Senescent-INR, and INR-B

subjects of the SCOPE cohort (Table S16 lists the details of correlations between plasma cytokine cluster 3 members and Treg subsets). **e**, Heatmap highlighting overlapping leading-edge genes from the association of inducible HIV and frequencies of Treg cluster 7 with the 352 gene-based classifier genes (top block), SMAD2/3 targets (middle block) and HDAC1/2 targets (bottom block). Rows represent features and columns represent samples of the SCOPE cohort: IR (Grey), INR-A (Red), and INR-B (Blue). The gene-expression per row was centered at zero and to a standard deviation of 1 (z-score). As indicated in the figure, a red-white-blue gradient is used to depict the relative expression of the features. The magnitude of CD4 counts (purple), frequencies of Treg Cluster7 (GARP+, dark blue) and size of HIV inducible reservoir (black) are plotted as annotations at the top of the heatmap. GSEA was used to assess the association between the features and inducible HIV/Treg cluster 7 frequencies (linear regression, $p<0.05$). Genes that showed a significant overlap (GSEA vs inducible HIV and Treg cluster 7 frequencies; significance (p-value $<0.05$ - assessed by Fischer exact test) amongst the leading-edge gene list/pathway are represented on the heatmap (See Table S17 for all the details). **f-g** The LARA in vitro model[112] was used to characterize the impact of dose-dependent increase in TGF-β on the establishment of HIV latency in memory CD4+ T cells. Increasing concentrations of TGF-β (0.2-20ng/mL) during the latency phase of the assay led to heightened frequencies of CD4 T cells with integrated HIV DNA **(f)** and resulted in significantly higher frequencies of p24+ CD4 T cells upon reactivation with anti-CD3/28 antibodies **(g)**. Spearman's test was used to assess

significance for correlation in (f) (p-value and rho as indicated on graph) and Wilcoxon rank-sum test was used to assess significance between timepoints in (g) (p-value as indicated on graph).

**Figure 4 | Microbial metabolites drive Treg differentiation and TGF-β production in Senescent INRs.**

**a**, Box plots illustrating levels of bacterial phylum Firmicutes are heightened in Senescent-INRs (p=0.06) of the SCOPE cohort. Subject groups are plotted along the X-axis: INR-A (Red), and INR-B (Blue) and $\log_{10}$ of relative Firmicute abundance based on PathSeq plotted along the Y-axis. A Wilcoxon tank sum test was used to assess significance.

**b**, Taxonomy tree of all genera with abundance >0 TPM in all subgroups (IRs, Senescent and Inflammatory INRs) of the SCOPE cohort. Each genus label and tree leaf is colored based on bacterial phyla annotation. Bar plots in the concentric circles represent log(TPM+1) value for each genus in Senescent or Inflammatory INRs form the SCOPE cohort. Stars beside the bars highlight significant differences between Senescent or Inflammatory INRs (P<0.05: large start, P<0.1: small star). Heatmap strip in the concentric circle shows the correlation value (Spearman's rho) of each genus with inducible HIV. Stars beside the heatmap strip highlight significant correlations (P<0.05: large star, P<0.1: small star).

**c**, Principal Component Analysis of all plasma metabolome highlights a cluster of subjects with high TILDA levels and reveals a significant negative correlation

between PC1 (x-axis) and the Classifier gene-set and a significant positive correlation between PC2 (y-axis) and TILDA. Spearman Correlation was used to assess significance, rho and p-values provided on the figure.

**d,** Correlation matrix heatmap between abundance of plasma bile acid/microbial metabolites (plotted along the y-axis) and variables of interest including TILDA levels, GARP+ Tregs, SMAD2/3 and HDAC1/2 Targets (plotted along the y-axis). Analysis highlights α-ketobutyrate as the metabolite correlated with levels of TILDA and GARP+ Tregs. A spearman correlation was used to assess significance. Stars on the heatmap highlight significant correlations ($P<0.05$: large star, $P<0.1$: small star).

**Figure 5 | alpha-keto butyrate stimulation drives Treg differentiation and enhances TGF-β production.**

**a**, *in-vitro* experiment to assess the impact of increasing concentrations of alpha-ketobutyrate on sorted naïve CD4 T cells from healthy subjects in the presence of IL-2, anti-CD3/28 antibodies and/or TGF-β. Dimension Reduction of high density flow cytometry analysis highlights UMAP1 and UMAP2 on X- and Y- axes respectively and illustrates that stimulation in the presence of TGF-β led to a profound increase in frequencies of GARP+FOXP3+ cells (**a**, Cluster 7) and GARP+PD1+ cells (**a**, Cluster 13).

**b, d** Jitter plots highlighting that increasing concentrations of alpha-ketobutyrate preferentially led to differentiation into GARP+FOXP3+ cells (**b**) and increased levels of PD1 expressing quiescent cells (**d**). GARP+FOXP3+ and PD1+CD27+

levels plotted along the Y-axis, while alpha-ketobutyrate increasing concentration levels are indicated on the X-axis. Conditions in the absence of TFG-b are designated with (-TGF-β). A Mann-Whitney U-test was used to assess significance across concentrations (* represents p-value <0.05 between groups).

**c**, Plasma cytokine levels of TGF-β1 significantly increase with enhanced alpha-ketobutyrate stimulation. TGF-β1 levels plotted along the Y-axis, while alpha-ketobutyrate concentration levels are indicated on the X-axis. A Mann-Whitney U-test was used to assess significance across concentrations (* represents p-value <0.05 between groups).

**e- f,** LARA assay in vitro model was used to characterize the impact of dose-dependent increased in TGF-β on the establishment of HIV latency in memory CD4+ T cells. Increasing concentrations of TGF-β (0.2-20ng/mL) during LARA latency culture results in heightened frequencies of CD4 T cells with integrated HIV DNA **(e)** and results in augmented numbers of p24+ CD4 T cells triggered by reactivation using anti-CD3/28 antibodies stimulation **(f)**.

**Figure 6 | Heightened frequencies of PD1$^{hi}$ROS$^{hi}$ CD4 TCMs and lower CD4 counts in Senescent INRs are associated with poor mitochondrial metabolism and senescence. a**, Clusters determined using the PhenoGraph method and visualized using the Uniform Manifold Approximation and Projection (UMAP) analysis were used to represent the distribution of CD4 T cells subsets, surface expression of PD-1 and markers of mitochondrial activity in CD4 T cells (using high-dimensional flow cytometry) in IR, Senescent-INR, and INR-B

subjects of the SCOPE cohort (See Table S18 for details). Density Plots highlight the enrichment of a specific cluster “Cluster 9: PD1+TCM” in Senescent-INR subjects. **b**, Box plots illustrate unique and significantly higher frequencies of Cluster 9 (PD1+TCM in total CD4 T-cells) in Senescent INR subjects and Cluster 17 (PD1+TEM in total CD4 T cells) in the Inflammatory INR. A Wilcoxon rank test was used to assess significance (*p-value of <0.05) (Table S18). **c**, MFIs of cell surface (CD45RA, CD27, CCR7, PD1) per cluster are shown. Clusters 9 (red) and 17 (blue) are overlaid to highlight the heightened levels of CD27 and CCR7 in Cluster 9 (i.e. TCM), and low levels of CD27 and CCR7 in Cluster 17 (i.e. TEM). Both clusters showed relatively low levels of CD45RA and high levels of PD1. Black line across each plot discriminated traditionally gated negative population from the its positive counterpart. A head-to-head comparison between clusters 9 and 17 revealed a distinct upregulation in CellROX within cluster 9 (i.e. PD1+ TCMs that are increased in the senescent INRs) (Raw MFIs of all clusters are listed in Table S18). **d**, Correlation network highlighting significant associations between PD1+TCM cluster frequencies and plasma cytokine (positive: IL6, IL29, IFNa2a, negative: Fractalkine) in the pool of Senescent INRs and IRs (i.e. driven by lower CD4 counts in Senescent INRs vs IRs). Triangular blue nodes depict plasma cytokines; orange squares reflect high-density flow cytometry clusters; and red edges highlight a positive correlation between cytokines and cell clusters. A Spearman correlation test was used to assess significance (all edges have a significant p-value <0.05) across IR and Senescent-INR subjects of the SCOPE cohort (details of all correlations are

listed in Table S19). **e**, Heatmap highlighting a positive association between Reactive Oxygen Species and CD4 numbers (top block) and a negative association between Oxphos/targets of Myc and CD4 numbers (middle and bottom blocks) respectively. GSEA was used to assess the association between the features and CD4 numbers/PD1+ TCM frequencies (linear regression, $p<0.05$). Genes that showed a significant overlap (GSEA vs inducible HIV and PD1+ TCM frequencies; significance (p-value $<0.05$) assessed by Fischer exact test) amongst the leading-edge gene list/pathway are represented in the heatmap (See Table S20 for overlapping gene-lists and leading-edge gene lists). Rows represent z-score normalized (red-white-blue gradient) genes and columns represent subjects from the SCOPE cohort: IR (Grey), INR-A (Red), and INR-B (Blue). The magnitude of CD4 counts (purple), frequencies of PD1+TCM (dark red) and inducible HIV (black) are plotted as annotations at the top of the heatmap. **f**, Heatmap demonstrating the association of senescent specific pathways PD1+TCM cluster. Leading-edge genes from the association of PD1+TCM and senescence pathway were summarized into a SLEA representation (z-score per pathway per sample) (See Table S20 for details). Rows represent senescent-specific pathways and columns represent samples of the SCOPE cohort: IR (Grey), INR-A (Red), and INR-B (Blue). A red-white-blue gradient is used to depict the relative SLEA score of the features, where blue represent a low relative-expression and red a high relative-expression of the feature. The magnitude of CD4 counts (purple), frequencies of PD1+TCM (dark

red) and inducible HIV (black) are plotted as annotations at the top of the heatmap.

**g**, Circle plot highlighting an integrated model associating leading/overlapping edges from genesets identified (classifier gene-set and biological pathways) in Figs 1 to 4, GARP+ Treg frequencies (grey), and PD1+TCM frequencies (grey) and with outcomes (CD4 counts, TILDA inducible HIV (orange)) across SCOPE cohort (n = 41). SLEA scores (z-score per pathway per sample) were calculated for each of these gene-lists (see Table S21 for the scores) and spearman's test was used to assess correlation between all of the features mentioned. Color gradient of the Pathway node (white to red, white to blue) reflects relative enrichment (positive or negative; respectively) of pathway in Senescent-INRs compared to IR subjects. Edges between nodes represent a significant positive or negative correlation (red or blue respectively) (details for the full network are summarized in Table S21).

**Supplementary Figures (S1-S7):**

**Supplementary Figure 1 | Gene Expression profiling provides a unique tool to identify two heterogeneous INR subject groups. a**, Flow chart illustrating the steps in the multi–omics data analyses and integration utilized across two independent cohorts of HIV-infected ART-treated subjects. **b**, INR groups remain relatively stable over time and show slow CD4 T cell reconstitution. Plot represents absolute CD4 counts over 3 years for all samples in each group. X-axis represents the draw date (in years) and y-axis represents the absolute CD4

count in cells/mm3. Red = INR-A, Blue =INR-B, Grey= IR. The fitted lines represent the rate of CD4 reconstitution over time (y=mx + b). **c**, Unsupervised analysis of gene expression data identifies three groups of HIV subjects treated with ART (IR, INR-A and INR-B) in the CLIF cohort. Gap Statistics revealed n=2 optimal INR clusters (bootstrap n = 500, p= 0.00293). Heatmap shows the expression of the top 200 genes that distinguish the 3 groups. Hierarchical clustering (Euclidian distance, complete linkage) was used to re-group samples with similar expression of the top 200 varying genes. Gene Expression is represented as a gene-wise standardized expression (Z-score). **d-o**, several parameters associated with poor immune reconstitution in ART- treated subjects including: Duration of ART treatment (d), Age (e), CD4 Nadir (f) and CD4 counts at draw (g), Levels of D-dimer (h), Levels of IL6 (i), IP10 (j), IFABP (k), LPS (l), sCD14 (m) were assessed for their ability to distinguish the two groups of ART-treated subjects. A Welch t-test was used to assess significance. INR-As and INR-Bs could not be distinguished for CD4 numbers and age (m), CD4 Nadir (n), or Years on ART (o). A Spearman correlation was used to assess significance.

**Supplementary Figure 2 | Diminished global transcriptional activity within INR-A subjects maps to down-regulation of T-cell, mDC and monocyte subset specific gene signatures and lack of effector CD8 function. a**, Pie charts illustrating the majority of the transcriptional signal (~70%) specific to Senescent-INR subjects is due to down-regulation of genes. The majority of genes differentially expressed between INR-As and IRs (left) or INR-As and INR-

Bs (right) are down-regulated (details of DEGs are in Table S3-S5). **b**, Networks highlighting the genes and functional annotations of T-cell, MDC, and Monocyte specific gene expression signatures down-regulated in INR-A subjects. GSEA ($p<0.05$) were used to assess significance of the down-regulation of subset gene-sets, Gene Ontology (GO) was used to infer the biological function (over-representation test $p<0.05$) and Gene Mania was used to plot co-expression networks. **c**, Volcano Plots illustrating the top Transcription Factors (TFs, squares) up- and down-regulated in INR-A vs IR (top panel) and INR-A vs INR-B (bottom panel) subjects. X-axis represents the $\log_{10}$FC of TF expression, and Y-axis represents the $-\log_{10}$(p-value); dotted lines highlight the p-value ≤ 0.05 cutoff. Top TFs up-regulated in INR-As are colored in red, whereas TFs specific to IRs or INR-Bs are colored in grey or blue respectively (See Table S9 for details). **d,** A functional network of the core TFs (diamonds) and their target genes (circles) up-regulated in INR-A subjects (red, left portion) or INR-Bs (blue, left portion). Network Inference and Cytoscape were used to plot the network. Reactome and GO terms were used to annotate the biological functions of the top downstream targets of the TFs up-regulated in INR-A subjects. As indicated in the figure, the node color represents log2 fold change in expression between the two INR classes. Edges within the network depict a link between a TF and its targets.

**Supplementary Figure 3 | Construction and Validation of the gene-based classifier. a**, Training of the 352-gene classifier on the CLIF cohort. Results of

the 10-fold cross validation. Misclassification Error plot represents the optimal number of features (genes) that corresponds to the lowest misclassification error rate. The pamr package in R was used to train the classifier via the nearest shrunken centroid method; 352 genes were selected to segregate the two groups of immune non-responders (See Table S11 for list of classifier genes). **b**, Heatmap representation of the expression of the top 200 varying probesets in the SCOPE cohort. The expression intensities are represented using green-black-red color scale. Rows correspond to probesets and columns correspond to the profiled samples. Hierarchical clustering based on a complete linkage was used to regroup samples with similar gene-expression profiles. Two clusters of INRs (depicted in red and blue on the dendrogram) were identified using the unsupervised clustering. **c**, Multidimensional scaling (MDS) plot highlights the overall transcriptomic variation within the SCOPE validation cohort.

**Supplementary Figure 4 | Identification of a specific Senescent Cytokine Cluster. a**, K-means clustering and the GAP statistic technique[91] were used to identify the optimal number of cytokine clusters after 42 cytokines were measured using the meso-scale platform. Four clusters were observed. **b**, Hierarchical clustering technique identified similar clusters. Two stable clusters (identified concurrently by k-means and hierarchal clustering) and were observed, one of which contained TGF-β1/2, IL3/7/13, KC.GRO and VEGF. **c**, Table summarizes the cytokine Cluster members identified in (a). **d**, Heatmap demonstrating the expression of the TGF-β1/2 driven cytokine cluster “Cluster 3”

(top portion) and the expression of classifier features correlated to Cluster 3 (bottom portion) across all subjects of the SCOPE cohort. Rows represent the features (plasma cytokine levels or TF expression levels) and columns represent samples of the SCOPE cohort: IR (Grey), INR-A (Red), and INR-B (Blue). A red-white-blue gradient is used to depict the relative expression of the features, where blue represent a low relative-expression and red a high relative-expression of the feature. The magnitude of the Senescent Cytokine Cluster (Cluster 3, dark green) is plotted as annotations at the top of the heatmap. GSEA was used to assess the association between the classifier features and Cluster 3 of senescent cytokines (linear regression: classifier genes ~ Cluster 3, GSEA: NES=2.87, $p<0.01$), and the leading-edge features are represented (See Table S15 for detailed list of genes and associations of cytokine cluster 3 with transcriptome of all groups).

**Supplementary Figure 5 | Microbiome and Metabolome differences associated with Senescent-INR subjects map to Firmicutes and Butyrates.**

**a,** Violin jitter plot highlighting differences in PathSeq species diversity between Senescent-INR subjects (red) and inflammatory-INR subjects (blue) of the SCOPE cohort. The Shannon diversity index (H) accounting for both abundance and evenness of the species present was used and plotted along the y-axis.

**b**, Bray-Curtis dissimilarity statistic to assess the beta diversity of the Firmicute species abundance in the microbiome PathSeq data of the SCOPE cohort. Clustering dendrogram reflects differential distribution of Senescent-INR (right

most branch) and INR-B (left-most branch) subjects summarizing a difference in the Firmicute microbial composition between these subjects. Circles reflect subjects: Senescent-INR (Red), and INR-B (Blue). Clustering distances reflect similarities and differences at the species level.

**c**, Scatter plot highlighting a positive association between frequencies of CD4 T cells with inducible HIV as measured by TILDA plotted along the X-axis and levels of a specific family of the Firmicute phylum: Lactobacillaceae plotted along the Y-axis in -INR (red) and INR-B (blue) subjects of the SCOPE cohort.

**d-e**, Jitter plots highlighting that increasing concentrations of alpha-ketobutyrate preferentially led to differentiation into GARP+FOXP3+ cells (**d**) and increased levels of PD1 expressing quiescent cells (**e**). GARP+FOXP3+ and PD1+CD27+ levels plotted along the Y-axis, while alpha-ketobutyrate increasing concentration levels are indicated on the X-axis. Conditions in the presence of TFG-b are designated with (+TGF-β). A Mann-Whitney U-test was used to assess significance across concentrations (* represents p-value <0.05 between groups).

**f**, CD4 T cell subsets enriched after stimulation with alpha-ketobutyrate (i.e. GARP+ Tregs and PD1+ TCM; nodes in orange squares), were significantly positively associated ($p < 0.05$; Spearman's rho>0 - red edges) with an increase in secreted TGF-β1 and a significant decrease in effector cytokines like IL17A, IFNg and IL9 ($p < 0.05$; Spearman's rho<0 - blue edges).

**Supplementary Figure 6 | A multi-omic gene expression and flow cytometry model correlates with the magnitude of CD4 counts and inducible HIV.**

Detailed network depicting an integrated model highlighting the association between gene-expression outcomes (genes and pathways, circular nodes), flow cytometry readouts (square nodes) and clinical outcomes (diamond nodes). Circular nodes reflect leading edge genes of biological pathways listed and are colored by the fold change in Senescent INR subjects. Triangular nodes reflect highlight transcription factors and are colored by the fold change in Senescent INR subjects. Edges between nodes represent a significant positive or negative correlation (red or blue respectively). A spearman correlation was used to assess significance (p-value < 0.05).

**Supplementary Figure 7 | Flow cytometry gating strategy for the markers used in this study. a**, A lymphocyte gate based on FSC-A and SSC-A was defined. Single cells were then selected by FSC-A x FCS-H gate. Live cells were gated and successive gates to define T cell populations (CD3+CD4+ and CD3+CD8+) were made. **b**, Memory phenotypes gating of CD4 and CD8 subsets (Tcm: CD45RA-CCR7+CD27+, Ttm: CD45RA-CCR7-CD27+, Tem: CD45RA-CCR7-CD27-, Naive: CD45RA+CCR7+CD27+CD95- and Tscm: CD45RA+CCR7+CD27+CD95+) were performed. **c**, Tregs were identified after initial gating on CD4 T cells and the expression of SATB1 was measured in this subset. **d**, the frequencies of PD1+, LAP+, a surrogate marker for TGF-β expression and GLUT1+ cells were measured in all subsets using the gating strategy described here.

## References


1 Dimitrov, D. T., Kiem, H. P., Jerome, K. R., Johnston, C. & Schiffer, J. T. A curative regimen would decrease HIV prevalence but not HIV incidence unless targeted to an ART-naive population. *Sci Rep* **6**, 22183, doi:10.1038/srep22183 (2016).

2 Chu, C., Umanski, G., Blank, A., Grossberg, R. & Selwyn, P. A. HIV-infected patients and treatment outcomes: an equivalence study of community-located, primary care-based HIV treatment vs. hospital-based specialty care in the Bronx, New York. *AIDS Care* **22**, 1522-1529, doi:10.1080/09540121.2010.484456 (2010).

3 Yang, X. *et al.* Incomplete immune reconstitution in HIV/AIDS patients on antiretroviral therapy: Challenges of immunological non-responders. *J Leukoc Biol* **107**, 597-612, doi:10.1002/JLB.4MR1019-189R (2020).

4 Gaardbo, J. C., Hartling, H. J., Gerstoft, J. & Nielsen, S. D. Incomplete immune recovery in HIV infection: mechanisms, relevance for clinical care, and possible solutions. *Clin Dev Immunol* **2012**, 670957, doi:10.1155/2012/670957 (2012).

5 Massanella, M., Negredo, E., Clotet, B. & Blanco, J. Immunodiscordant responses to HAART--mechanisms and consequences. *Expert Rev Clin Immunol* **9**, 1135-1149, doi:10.1586/1744666X.2013.842897 (2013).

6 Pacheco, Y. M. *et al.* Increased risk of non-AIDS-related events in HIV subjects with persistent low CD4 counts despite cART in the CoRIS cohort. *Antiviral Res* **117**, 69-74, doi:10.1016/j.antiviral.2015.03.002 (2015).

7 Shive, C. L. *et al.* Inflammation Perturbs the IL-7 Axis, Promoting Senescence and Exhaustion that Broadly Characterize Immune Failure in Treated HIV Infection. *J Acquir Immune Defic Syndr* **71**, 483-492, doi:10.1097/QAI.0000000000000913 (2016).

8 Molina-Pinelo, S. *et al.* Premature immunosenescence in HIV-infected patients on highly active antiretroviral therapy with low-level CD4 T cell repopulation. *J Antimicrob Chemother* **64**, 579-588, doi:10.1093/jac/dkp248 (2009).

9 Mendez-Lagares, G. *et al.* Specific patterns of CD4-associated immunosenescence in vertically HIV-infected subjects. *Clin Microbiol Infect* **19**, 558-565, doi:10.1111/j.1469-0691.2012.03934.x (2013).

10 Grabmeier-Pfistershammer, K., Steinberger, P., Rieger, A., Leitner, J. & Kohrgruber, N. Identification of PD-1 as a unique marker for failing immune reconstitution in HIV-1-infected patients on treatment. *J Acquir Immune Defic Syndr* **56**, 118-124, doi:10.1097/QAI.0b013e3181fbab9f (2011).

11 Mendez-Lagares, G. *et al.* Severe immune dysregulation affects CD4(+)CD25(hi)FoxP3(+) regulatory T cells in HIV-infected patients with low-level CD4 T-cell repopulation despite suppressive highly active antiretroviral therapy. *J Infect Dis* **205**, 1501-1509, doi:10.1093/infdis/jis230 (2012).
12 Lederman, M. M., Funderburg, N. T., Sekaly, R. P., Klatt, N. R. & Hunt, P. W. Residual immune dysregulation syndrome in treated HIV infection. *Adv Immunol* **119**, 51-83, doi:10.1016/B978-0-12-407707-2.00002-3 (2013).
13 Boulassel, M. R. *et al.* CD4 T cell nadir independently predicts the magnitude of the HIV reservoir after prolonged suppressive antiretroviral therapy. *J Clin Virol* **53**, 29-32, doi:10.1016/j.jcv.2011.09.018 (2012).
14 Said, E. A. *et al.* Programmed death-1-induced interleukin-10 production by monocytes impairs CD4+ T cell activation during HIV infection. *Nat Med* **16**, 452-459, doi:10.1038/nm.2106 (2010).
15 Hiener, B. *et al.* Identification of Genetically Intact HIV-1 Proviruses in Specific CD4(+) T Cells from Effectively Treated Participants. *Cell Rep* **21**, 813-822, doi:10.1016/j.celrep.2017.09.081 (2017).
16 Churchill, M. J., Deeks, S. G., Margolis, D. M., Siliciano, R. F. & Swanstrom, R. HIV reservoirs: what, where and how to target them. *Nat Rev Microbiol* **14**, 55-60, doi:10.1038/nrmicro.2015.5 (2016).
17 Peluso, M. J. *et al.* Differential decay of intact and defective proviral DNA in HIV-1-infected individuals on suppressive antiretroviral therapy. *JCI Insight* **5**, doi:10.1172/jci.insight.132997 (2020).
18 Freeman, M. L. *et al.* Cytokines and T-Cell Homeostasis in HIV Infection. *J Infect Dis* **214 Suppl 2**, S51-57, doi:10.1093/infdis/jiw287 (2016).
19 von Stockenstrom, S. *et al.* Longitudinal Genetic Characterization Reveals That Cell Proliferation Maintains a Persistent HIV Type 1 DNA Pool During Effective HIV Therapy. *J Infect Dis* **212**, 596-607, doi:10.1093/infdis/jiv092 (2015).
20 Chomont, N., DaFonseca, S., Vandergeeten, C., Ancuta, P. & Sekaly, R. P. Maintenance of CD4+ T-cell memory and HIV persistence: keeping memory, keeping HIV. *Curr Opin HIV AIDS* **6**, 30-36, doi:10.1097/COH.0b013e3283413775 (2011).
21 Chomont, N. *et al.* HIV reservoir size and persistence are driven by T cell survival and homeostatic proliferation. *Nat Med* **15**, 893-900, doi:10.1038/nm.1972 (2009).
22 Fromentin, R. *et al.* PD-1 blockade potentiates HIV latency reversal ex vivo in CD4(+) T cells from ART-suppressed individuals. *Nat Commun* **10**, 814, doi:10.1038/s41467-019-08798-7 (2019).
23 Fromentin, R. *et al.* CD4+ T Cells Expressing PD-1, TIGIT and LAG-3 Contribute to HIV Persistence during ART. *PLoS Pathog* **12**, e1005761, doi:10.1371/journal.ppat.1005761 (2016).
24 Borges, A. H. *et al.* Factors Associated With Plasma IL-6 Levels During HIV Infection. *J Infect Dis* **212**, 585-595, doi:10.1093/infdis/jiv123 (2015).

25 Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**, 417-425, doi:10.1016/j.cels.2015.12.004 (2015).
26 Liberzon, A. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739-1740, doi:10.1093/bioinformatics/btr260 (2011).
27 Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550, doi:10.1073/pnas.0506580102 (2005).
28 Prieto, L. I., Graves, S. I. & Baker, D. J. Insights from In Vivo Studies of Cellular Senescence. *Cells* **9**, doi:10.3390/cells9040954 (2020).
29 Childs, B. G., Durik, M., Baker, D. J. & van Deursen, J. M. Cellular senescence in aging and age-related disease: from mechanisms to therapy. *Nat Med* **21**, 1424-1435, doi:10.1038/nm.4000 (2015).
30 Wang, R. *et al.* The transcription factor Myc controls metabolic reprogramming upon T lymphocyte activation. *Immunity* **35**, 871-882, doi:10.1016/j.immuni.2011.09.021 (2011).
31 Nakaya, H. I. *et al.* Systems biology of vaccination for seasonal influenza in humans. *Nat Immunol* **12**, 786-795, doi:10.1038/ni.2067 (2011).
32 Novershtern, N. *et al.* Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell* **144**, 296-309, doi:10.1016/j.cell.2011.01.004 (2011).
33 Douagi, I. *et al.* Role of interferon regulatory factor 3 in type I interferon responses in rotavirus-infected dendritic cells and fibroblasts. *J Virol* **81**, 2758-2768, doi:10.1128/JVI.01555-06 (2007).
34 Webb, A. E. & Brunet, A. FOXO transcription factors: key regulators of cellular quality control. *Trends Biochem Sci* **39**, 159-169, doi:10.1016/j.tibs.2014.02.003 (2014).
35 Nakao, A. *et al.* TGF-beta receptor-mediated signalling through Smad2, Smad3 and Smad4. *EMBO J* **16**, 5353-5362, doi:10.1093/emboj/16.17.5353 (1997).
36 Ramakrishnan, R. *et al.* Identification of novel CDK9 and Cyclin T1-associated protein complexes (CCAPs) whose siRNA depletion enhances HIV-1 Tat function. *Retrovirology* **9**, 90, doi:10.1186/1742-4690-9-90 (2012).
37 Bindea, G. *et al.* ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* **25**, 1091-1093, doi:10.1093/bioinformatics/btp101 (2009).
38 Jassal, B. *et al.* The reactome pathway knowledgebase. *Nucleic Acids Res* **48**, D498-D503, doi:10.1093/nar/gkz1031 (2020).
39 Chang, Z. S. *et al.* Forkhead box O3 protects the heart against paraquat-induced aging-associated phenotypes by upregulating the expression of antioxidant enzymes. *Aging Cell* **18**, e12990, doi:10.1111/acel.12990 (2019).
40 Nho, R. S. & Hergert, P. FoxO3a and disease progression. *World J Biol Chem* **5**, 346-354, doi:10.4331/wjbc.v5.i3.346 (2014).

41 Hagenbuchner, J. & Ausserlechner, M. J. Mitochondria and FOXO3: breath or die. *Front Physiol* **4**, 147, doi:10.3389/fphys.2013.00147 (2013).
42 Bourgeois, B. & Madl, T. Regulation of cellular senescence via the FOXO4-p53 axis. *FEBS Lett* **592**, 2083-2097, doi:10.1002/1873-3468.13057 (2018).
43 Baar, M. P. *et al.* Targeted Apoptosis of Senescent Cells Restores Tissue Homeostasis in Response to Chemotoxicity and Aging. *Cell* **169**, 132-147 e116, doi:10.1016/j.cell.2017.02.031 (2017).
44 Hagenbuchner, J. *et al.* FOXO3-induced reactive oxygen species are regulated by BCL2L11 (Bim) and SESN3. *J Cell Sci* **125**, 1191-1203, doi:10.1242/jcs.092098 (2012).
45 Fu, Z. & Tindall, D. J. FOXOs, cancer and regulation of apoptosis. *Oncogene* **27**, 2312-2319, doi:10.1038/onc.2008.24 (2008).
46 Lin, H. K., Bergmann, S. & Pandolfi, P. P. Cytoplasmic PML function in TGF-beta signalling. *Nature* **431**, 205-211, doi:10.1038/nature02783 (2004).
47 Lusic, M. *et al.* Proximity to PML nuclear bodies regulates HIV-1 latency in CD4+ T cells. *Cell Host Microbe* **13**, 665-677, doi:10.1016/j.chom.2013.05.006 (2013).
48 Brown, J. A. *et al.* TGF-beta-Induced Quiescence Mediates Chemoresistance of Tumor-Propagating Cells in Squamous Cell Carcinoma. *Cell Stem Cell* **21**, 650-664 e658, doi:10.1016/j.stem.2017.10.001 (2017).
49 Yang, L. *et al.* Targeting Stromal Glutamine Synthetase in Tumors Disrupts Tumor Microenvironment-Regulated Cancer Cell Growth. *Cell Metab* **24**, 685-700, doi:10.1016/j.cmet.2016.10.011 (2016).
50 van der Vos, K. E. & Coffer, P. J. Glutamine metabolism links growth factor signaling to the regulation of autophagy. *Autophagy* **8**, 1862-1864, doi:10.4161/auto.22152 (2012).
51 Dang, C. V. c-Myc target genes involved in cell growth, apoptosis, and metabolism. *Mol Cell Biol* **19**, 1-11, doi:10.1128/mcb.19.1.1 (1999).
52 Richmond, A. Nf-kappa B, chemokine gene transcription and tumour growth. *Nat Rev Immunol* **2**, 664-674, doi:10.1038/nri887 (2002).
53 He, Y., Hara, H. & Nunez, G. Mechanism and Regulation of NLRP3 Inflammasome Activation. *Trends Biochem Sci* **41**, 1012-1021, doi:10.1016/j.tibs.2016.09.002 (2016).
54 Barkett, M. & Gilmore, T. D. Control of apoptosis by Rel/NF-kappaB transcription factors. *Oncogene* **18**, 6910-6924, doi:10.1038/sj.onc.1203238 (1999).
55 Herranz, N. & Gil, J. Mechanisms and functions of cellular senescence. *J Clin Invest* **128**, 1238-1246, doi:10.1172/JCI95148 (2018).
56 Baar, M. P., Van Willigenburg, H. & de Keizer, P. L. J. Maintenance and repair of an aging life cycle. *Oncotarget* **8**, 86985-86986, doi:10.18632/oncotarget.18046 (2017).

57 Borras, C. *et al.* BCL-xL, a Mitochondrial Protein Involved in Successful Aging: From C. elegans to Human Centenarians. *Int J Mol Sci* **21**, doi:10.3390/ijms21020418 (2020).
58 Feng, J. *et al.* RIOK3 is an adaptor protein required for IRF3-mediated antiviral type I interferon production. *J Virol* **88**, 7987-7997, doi:10.1128/JVI.00643-14 (2014).
59 Suhane, S., Kanzaki, H., Arumugaswami, V., Murali, R. & Ramanujan, V. K. Mitochondrial NDUFS3 regulates the ROS-mediated onset of metabolic switch in transformed cells. *Biol Open* **2**, 295-305, doi:10.1242/bio.20133244 (2013).
60 Ji, L. *et al.* Slc6a8-Mediated Creatine Uptake and Accumulation Reprogram Macrophage Polarization via Regulating Cytokine Responses. *Immunity* **51**, 272-284 e277, doi:10.1016/j.immuni.2019.06.007 (2019).
61 Khan, A. A. & Quigley, J. G. Heme and FLVCR-related transporter families SLC48 and SLC49. *Mol Aspects Med* **34**, 669-682, doi:10.1016/j.mam.2012.07.013 (2013).
62 Reithmeier, R. A. *et al.* Band 3, the human red cell chloride/bicarbonate anion exchanger (AE1, SLC4A1), in a structural context. *Biochim Biophys Acta* **1858**, 1507-1532, doi:10.1016/j.bbamem.2016.03.030 (2016).
63 Hoffman, N. E. *et al.* SLC25A23 augments mitochondrial Ca(2)(+) uptake, interacts with MCU, and induces oxidative stress-mediated cell death. *Mol Biol Cell* **25**, 936-947, doi:10.1091/mbc.E13-08-0502 (2014).
64 Kim, J. *et al.* Amino Acid Transporter Slc38a5 Controls Glucagon Receptor Inhibition-Induced Pancreatic alpha Cell Hyperplasia in Mice. *Cell Metab* **25**, 1348-1361 e1348, doi:10.1016/j.cmet.2017.05.006 (2017).
65 Hoffmann, E. K. & Dunham, P. B. Membrane mechanisms and intracellular signalling in cell volume regulation. *Int Rev Cytol* **161**, 173-262, doi:10.1016/s0074-7696(08)62498-5 (1995).
66 Hamill, O. P. & Martinac, B. Molecular basis of mechanotransduction in living cells. *Physiol Rev* **81**, 685-740, doi:10.1152/physrev.2001.81.2.685 (2001).
67 Neurohr, G. E. *et al.* Excessive Cell Growth Causes Cytoplasm Dilution And Contributes to Senescence. *Cell* **176**, 1083-1097 e1018, doi:10.1016/j.cell.2019.01.018 (2019).
68 Yamakami, Y. *et al.* High concentrations of NaCl induce cell swelling leading to senescence in human cells. *Mol Cell Biochem* **411**, 117-125, doi:10.1007/s11010-015-2573-1 (2016).
69 Klatt, N. R., Chomont, N., Douek, D. C. & Deeks, S. G. Immune activation and HIV persistence: implications for curative approaches to HIV infection. *Immunol Rev* **254**, 326-342, doi:10.1111/imr.12065 (2013).
70 Procopio, F. A. *et al.* A Novel Assay to Measure the Magnitude of the Inducible Viral Reservoir in HIV-infected Individuals. *EBioMedicine* **2**, 874-883, doi:10.1016/j.ebiom.2015.06.019 (2015).
71 Daffis, S. *et al.* Interferon regulatory factor IRF-7 induces the antiviral alpha interferon response and protects against lethal West Nile virus infection. *J Virol* **82**, 8465-8475, doi:10.1128/JVI.00918-08 (2008).

72 Litvak, V. *et al.* A FOXO3-IRF7 gene regulatory circuit limits inflammatory sequelae of antiviral responses. *Nature* **490**, 421-425, doi:10.1038/nature11428 (2012).
73 Abdel-Mohsen, M. *et al.* Select host restriction factors are associated with HIV persistence during antiretroviral therapy. *AIDS* **29**, 411-420, doi:10.1097/QAD.0000000000000572 (2015).
74 Lachmann, A. *et al.* ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. *Bioinformatics* **26**, 2438-2444, doi:10.1093/bioinformatics/btq466 (2010).
75 Consortium, E. P. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* **306**, 636-640, doi:10.1126/science.1105136 (2004).
76 Rusinova, I. *et al.* Interferome v2.0: an updated database of annotated interferon-regulated genes. *Nucleic Acids Res* **41**, D1040-1046, doi:10.1093/nar/gks1215 (2013).
77 Lu, H. *et al.* AFF1 is a ubiquitous P-TEFb partner to enable Tat extraction of P-TEFb from 7SK snRNP and formation of SECs for HIV transactivation. *Proc Natl Acad Sci U S A* **111**, E15-24, doi:10.1073/pnas.1318503111 (2014).
78 Konig, R. *et al.* Global analysis of host-pathogen interactions that regulate early-stage HIV-1 replication. *Cell* **135**, 49-60, doi:10.1016/j.cell.2008.07.032 (2008).
79 Nibbs, R. J. HIV: getting to the heart of DARCness. *Blood* **114**, 2570-2571, doi:10.1182/blood-2009-08-235556 (2009).
80 Zhu, J. *et al.* Reactivation of latent HIV-1 by inhibition of BRD4. *Cell Rep* **2**, 807-816, doi:10.1016/j.celrep.2012.09.008 (2012).
81 Lee, S. Y. *et al.* Selection of biomarkers for HIV-1 latency by integrated analysis. *Genomics* **111**, 327-333, doi:10.1016/j.ygeno.2018.02.007 (2019).
82 Ren, Y. *et al.* BCL-2 antagonism sensitizes cytotoxic T cell-resistant HIV reservoirs to elimination ex vivo. *J Clin Invest* **130**, 2542-2559, doi:10.1172/JCI132374 (2020).
83 Haxhinasto, S., Mathis, D. & Benoist, C. The AKT-mTOR axis regulates de novo differentiation of CD4+Foxp3+ cells. *J Exp Med* **205**, 565-574, doi:10.1084/jem.20071477 (2008).
84 Wan, Y. Y. & Flavell, R. A. TGF-beta and regulatory T cell in immunity and autoimmunity. *J Clin Immunol* **28**, 647-659, doi:10.1007/s10875-008-9251-y (2008).
85 Kitagawa, Y. *et al.* Guidance of regulatory T cell development by Satb1-dependent super-enhancer establishment. *Nat Immunol* **18**, 173-183, doi:10.1038/ni.3646 (2017).
86 Salvany-Celades, M. *et al.* Three Types of Functional Regulatory T Cells Control T Cell Responses at the Human Maternal-Fetal Interface. *Cell Rep* **27**, 2537-2547 e2535, doi:10.1016/j.celrep.2019.04.109 (2019).
87 Dedobbeleer, O., Stockis, J., van der Woning, B., Coulie, P. G. & Lucas, S. Cutting Edge: Active TGF-beta1 Released from GARP/TGF-beta1 Complexes on the Surface of Stimulated Human B Lymphocytes

Increases Class-Switch Recombination and Production of IgA. *J Immunol* **199**, 391-396, doi:10.4049/jimmunol.1601882 (2017).

88 Tran, D. Q. *et al.* GARP (LRRC32) is essential for the surface expression of latent TGF-beta on platelets and activated FOXP3+ regulatory T cells. *Proc Natl Acad Sci U S A* **106**, 13445-13450, doi:10.1073/pnas.0901944106 (2009).

89 Levine, J. H. *et al.* Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis. *Cell* **162**, 184-197, doi:10.1016/j.cell.2015.05.047 (2015).

90 Huang, W., Solouki, S., Carter, C., Zheng, S. G. & August, A. Beyond Type 1 Regulatory T Cells: Co-expression of LAG3 and CD49b in IL-10-Producing T Cell Lineages. *Front Immunol* **9**, 2625, doi:10.3389/fimmu.2018.02625 (2018).

91 Tibshirani, R., Walther, G. & Hastie, T. Estimating the number of clusters in a data set via the gap statistic. *J R Stat Soc B* **63**, 411-423, doi:Doi 10.1111/1467-9868.00293 (2001).

92 Oh, S. A. & Li, M. O. TGF-beta: guardian of T cell function. *J Immunol* **191**, 3973-3979, doi:10.4049/jimmunol.1301843 (2013).

93 Wynn, T. A. IL-13 effector functions. *Annu Rev Immunol* **21**, 425-456, doi:10.1146/annurev.immunol.21.120601.141142 (2003).

94 Boro, M. & Balaji, K. N. CXCL1 and CXCL2 Regulate NLRP3 Inflammasome Activation via G-Protein-Coupled Receptor CXCR2. *J Immunol* **199**, 1660-1671, doi:10.4049/jimmunol.1700129 (2017).

95 Gupta, K. *et al.* VEGF prevents apoptosis of human microvascular endothelial cells via opposing effects on MAPK/ERK and SAPK/JNK signaling. *Exp Cell Res* **247**, 495-504, doi:10.1006/excr.1998.4359 (1999).

96 Cohen, S. B. IL-10 and IL-3 synergize to cause proliferation of human T cells. *Immunology* **85**, 351-356 (1995).

97 Li, W. Q. *et al.* IL-7 promotes T cell proliferation through destabilization of p27Kip1. *J Exp Med* **203**, 573-582, doi:10.1084/jem.20051520 (2006).

98 Koinuma, D. *et al.* Chromatin immunoprecipitation on microarray analysis of Smad2/3 binding sites reveals roles of ETS1 and TFAP2A in transforming growth factor beta signaling. *Mol Cell Biol* **29**, 172-186, doi:10.1128/MCB.01038-08 (2009).

99 Senese, S. *et al.* Role for histone deacetylase 1 in human tumor cell proliferation. *Mol Cell Biol* **27**, 4784-4795, doi:10.1128/MCB.00494-07 (2007).

100 Dubois, C. M. *et al.* Evidence that furin is an authentic transforming growth factor-beta1-converting enzyme. *Am J Pathol* **158**, 305-316, doi:10.1016/s0002-9440(10)63970-3 (2001).

101 Vallee, A., Lecarpentier, Y., Guillevin, R. & Vallee, J. N. Interactions between TGF-beta1, canonical WNT/beta-catenin pathway and PPAR gamma in radiation-induced fibrosis. *Oncotarget* **8**, 90579-90604, doi:10.18632/oncotarget.21234 (2017).

102 Xu, L. *et al.* Activation of Wnt/beta-catenin signalling is required for TGF-beta/Smad2/3 signalling during myofibroblast proliferation. *J Cell Mol Med* **21**, 1545-1554, doi:10.1111/jcmm.13085 (2017).
103 Kleinsimon, S. *et al.* GADD45A and CDKN1A are involved in apoptosis and cell cycle modulatory effects of viscumTT with further inactivation of the STAT3 pathway. *Sci Rep* **8**, 5750, doi:10.1038/s41598-018-24075-x (2018).
104 Marchetti, G., Tincati, C. & Silvestri, G. Microbial translocation in the pathogenesis of HIV infection and AIDS. *Clin Microbiol Rev* **26**, 2-18, doi:10.1128/CMR.00050-12 (2013).
105 Tabung, F. K. *et al.* Influence of Dietary Patterns on Plasma Soluble CD14, a Surrogate Marker of Gut Barrier Dysfunction. *Curr Dev Nutr* **1**, doi:10.3945/cdn.117.001396 (2017).
106 Arpaia, N. *et al.* Metabolites produced by commensal bacteria promote peripheral regulatory T-cell generation. *Nature* **504**, 451-455, doi:10.1038/nature12726 (2013).
107 Furusawa, Y. *et al.* Commensal microbe-derived butyrate induces the differentiation of colonic regulatory T cells. *Nature* **504**, 446-450, doi:10.1038/nature12721 (2013).
108 Walker, M. A. *et al.* GATK PathSeq: a customizable computational tool for the discovery and identification of microbial sequences in libraries from eukaryotic hosts. *Bioinformatics* **34**, 4287-4289, doi:10.1093/bioinformatics/bty501 (2018).
109 Kostic, A. D. *et al.* PathSeq: software to identify or discover microbes by deep sequencing of human tissue. *Nat Biotechnol* **29**, 393-396, doi:10.1038/nbt.1868 (2011).
110 Koh, H., Li, Y., Zhan, X., Chen, J. & Zhao, N. A Distance-Based Kernel Association Test Based on the Generalized Linear Mixed Model for Correlated Microbiome Studies. *Front Genet* **10**, 458, doi:10.3389/fgene.2019.00458 (2019).
111 Li, C., Ebert, P. J. & Li, Q. J. T cell receptor (TCR) and transforming growth factor beta (TGF-beta) signaling converge on DNA (cytosine-5)-methyltransferase to control forkhead box protein 3 (foxp3) locus methylation and inducible regulatory T cell differentiation. *J Biol Chem* **288**, 19127-19139, doi:10.1074/jbc.M113.453357 (2013).
112 Kulpa, D. A. *et al.* Differentiation into an Effector Memory Phenotype Potentiates HIV-1 Latency Reversal in CD4(+) T Cells. *J Virol* **93**, doi:10.1128/JVI.00969-19 (2019).
113 Park, B. V. *et al.* TGFbeta1-Mediated SMAD3 Enhances PD-1 Expression on Antigen-Specific T Cells in Cancer. *Cancer Discov* **6**, 1366-1381, doi:10.1158/2159-8290.CD-15-1347 (2016).
114 Ferber, E. C. *et al.* FOXO3a regulates reactive oxygen metabolism by inhibiting mitochondrial gene expression. *Cell Death Differ* **19**, 968-979, doi:10.1038/cdd.2011.179 (2012).

115 Shytaj, I. L. *et al.* Alterations of redox and iron metabolism accompany the development of HIV latency. *EMBO J* **39**, e102209, doi:10.15252/embj.2019102209 (2020).
116 Hatano, H. *et al.* Cell-based measures of viral persistence are associated with immune activation and programmed cell death protein 1 (PD-1)-expressing CD4+ T cells. *J Infect Dis* **208**, 50-56, doi:10.1093/infdis/jis630 (2013).
117 Bally, A. P., Austin, J. W. & Boss, J. M. Genetic and Epigenetic Regulation of PD-1 Expression. *J Immunol* **196**, 2431-2437, doi:10.4049/jimmunol.1502643 (2016).
118 Tilly, G. *et al.* IL-15 Harnesses Pro-inflammatory Function of TEMRA CD8 in Kidney-Transplant Recipients. *Front Immunol* **8**, 778, doi:10.3389/fimmu.2017.00778 (2017).
119 Janumyan, Y. M. *et al.* Bcl-xL/Bcl-2 coordinately regulates apoptosis, cell cycle arrest and cell cycle entry. *EMBO J* **22**, 5459-5470, doi:10.1093/emboj/cdg533 (2003).
120 Kuwahara, M. *et al.* The transcription factor Sox4 is a downstream target of signaling by the cytokine TGF-beta and suppresses T(H)2 differentiation. *Nat Immunol* **13**, 778-786, doi:10.1038/ni.2362 (2012).
121 Stefanetti, R. J., Voisin, S., Russell, A. & Lamon, S. Recent advances in understanding the role of FOXO3. *F1000Res* **7**, doi:10.12688/f1000research.15258.1 (2018).
122 Liu, T., Zhang, L., Joo, D. & Sun, S. C. NF-kappaB signaling in inflammation. *Signal Transduct Target Ther* **2**, doi:10.1038/sigtrans.2017.23 (2017).
123 Yanai, H. *et al.* Revisiting the role of IRF3 in inflammation and immunity by conditional and specifically targeted gene ablation in mice. *Proc Natl Acad Sci U S A* **115**, 5253-5258, doi:10.1073/pnas.1803936115 (2018).
124 Yang, H., Wang, H., Ren, J., Chen, Q. & Chen, Z. J. cGAS is essential for cellular senescence. *Proc Natl Acad Sci U S A* **114**, E4612-E4620, doi:10.1073/pnas.1705499114 (2017).
125 Yu, Q. *et al.* DNA-damage-induced type I interferon promotes senescence and inhibits stem cell function. *Cell Rep* **11**, 785-797, doi:10.1016/j.celrep.2015.03.069 (2015).
126 Cheng, L. *et al.* Blocking type I interferon signaling enhances T cell recovery and reduces HIV-1 reservoirs. *J Clin Invest* **127**, 269-279, doi:10.1172/JCI90745 (2017).
127 Klotz, L. O. *et al.* Redox regulation of FoxO transcription factors. *Redox Biol* **6**, 51-72, doi:10.1016/j.redox.2015.06.019 (2015).
128 Marinkovic, D. *et al.* Foxo3 is required for the regulation of oxidative stress in erythropoiesis. *J Clin Invest* **117**, 2133-2144, doi:10.1172/JCI31807 (2007).
129 Dong, H. *et al.* Role of FOXO3 Activated by HIV-1 Tat in HIV-Associated Neurocognitive Disorder Neuronal Apoptosis. *Front Neurosci* **13**, 44, doi:10.3389/fnins.2019.00044 (2019).

130 Jin, S. M. & Youle, R. J. PINK1- and Parkin-mediated mitophagy at a glance. *J Cell Sci* **125**, 795-799, doi:10.1242/jcs.093849 (2012).
131 Sasai, M. *et al.* Essential role for GABARAP autophagy proteins in interferon-inducible GTPase-mediated host defense. *Nat Immunol* **18**, 899-910, doi:10.1038/ni.3767 (2017).
132 Hopkins, B. L. & Neumann, C. A. Redoxins as gatekeepers of the transcriptional oxidative stress response. *Redox Biol* **21**, 101104, doi:10.1016/j.redox.2019.101104 (2019).
133 Trautmann, L. *et al.* Upregulation of PD-1 expression on HIV-specific CD8+ T cells leads to reversible immune dysfunction. *Nat Med* **12**, 1198-1202, doi:10.1038/nm1482 (2006).
134 Odorizzi, P. M., Pauken, K. E., Paley, M. A., Sharpe, A. & Wherry, E. J. Genetic absence of PD-1 promotes accumulation of terminally differentiated exhausted CD8+ T cells. *J Exp Med* **212**, 1125-1137, doi:10.1084/jem.20142237 (2015).
135 Elmore, S. Apoptosis: a review of programmed cell death. *Toxicol Pathol* **35**, 495-516, doi:10.1080/01926230701320337 (2007).
136 Masroori, N., Merindol, N. & Berthoux, L. The interferon-induced antiviral protein PML (TRIM19) promotes the restriction and transcriptional silencing of lentiviruses in a context-specific, isoform-specific fashion. *Retrovirology* **13**, 19, doi:10.1186/s12977-016-0253-1 (2016).
137 Nakahara, F., Weiss, C. N. & Ito, K. The role of PML in hematopoietic and leukemic stem cell maintenance. *Int J Hematol* **100**, 18-26, doi:10.1007/s12185-014-1518-x (2014).
138 Ramakrishnan, R. *et al.* Making a Short Story Long: Regulation of P-TEFb and HIV-1 Transcriptional Elongation in CD4+ T Lymphocytes and Macrophages. *Biology (Basel)* **1**, 94-115, doi:10.3390/biology1010094 (2012).
139 Hird, F. J. & Symons, R. H. The mechanism of ketone-body formation from butyrate in rat liver. *Biochem J* **84**, 212-216, doi:10.1042/bj0840212 (1962).
140 Vital, M., Karch, A. & Pieper, D. H. Colonic Butyrate-Producing Communities in Humans: an Overview Using Omics Data. *mSystems* **2**, doi:10.1128/mSystems.00130-17 (2017).
141 Vital, M., Howe, A. C. & Tiedje, J. M. Revealing the bacterial butyrate synthesis pathways by analyzing (meta)genomic data. *mBio* **5**, e00889, doi:10.1128/mBio.00889-14 (2014).
142 Hickson, L. J. *et al.* Senolytics decrease senescent cells in humans: Preliminary report from a clinical trial of Dasatinib plus Quercetin in individuals with diabetic kidney disease. *EBioMedicine* **47**, 446-456, doi:10.1016/j.ebiom.2019.08.069 (2019).
143 Xu, M. *et al.* Senolytics improve physical function and increase lifespan in old age. *Nat Med* **24**, 1246-1256, doi:10.1038/s41591-018-0092-9 (2018).
144 Jiao, S. *et al.* Differences in Tumor Microenvironment Dictate T Helper Lineage Polarization and Response to Immune Checkpoint Therapy. *Cell* **179**, 1177-1190 e1113, doi:10.1016/j.cell.2019.10.029 (2019).

145 Mariathasan, S. *et al.* TGFbeta attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature* **554**, 544-548, doi:10.1038/nature25501 (2018).
146 Watson, D. C. *et al.* Treatment with native heterodimeric IL-15 increases cytotoxic lymphocytes and reduces SHIV RNA in lymph nodes. *PLoS Pathog* **14**, e1006902, doi:10.1371/journal.ppat.1006902 (2018).
147 R: A language and enviornment for statistical computing (R Foundation for Statistical Computing, Vienna, Austria, 2018).
148 Gentleman, R. C. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* **5**, R80, doi:10.1186/gb-2004-5-10-r80 (2004).
149 Gentleman, R. *Bioinformatics and computational biology solutions using R and Bioconductor*. (Springer Science+Business Media, 2005).
150 Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57**, 289-300 (1995).
151 Tibshirani, R., Hastie, T., Narasimhan, B. & Chu, G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proc Natl Acad Sci U S A* **99**, 6567-6572, doi:10.1073/pnas.082099299 (2002).
152 Warde-Farley, D. *et al.* The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res* **38**, W214-220, doi:10.1093/nar/gkq537 (2010).

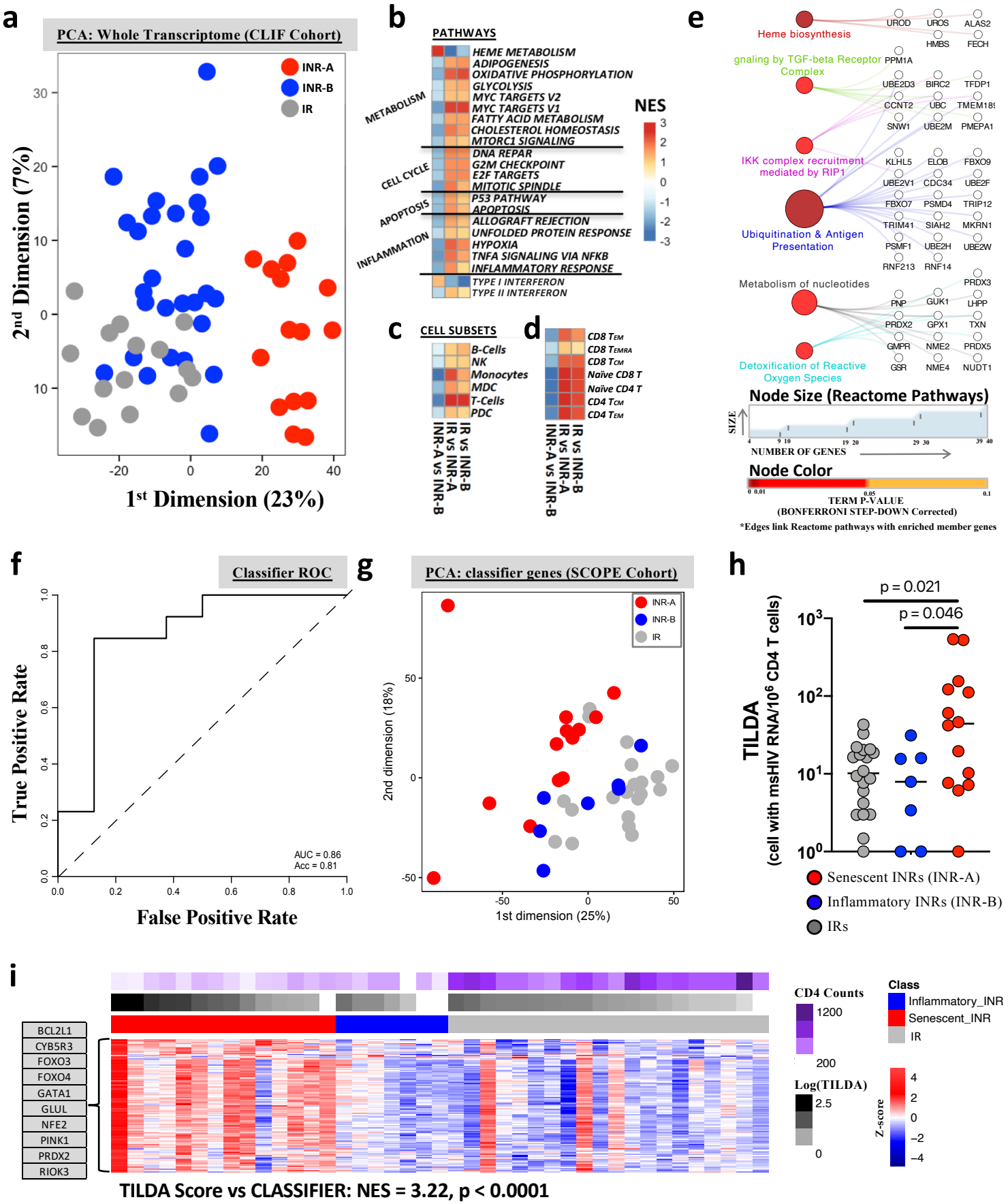

a
PCA: Whole Transcriptome (CLIF Cohort)
INR-A
INR-B
IR
2nd Dimension (7%)
1st Dimension (23%)
b
PATHWAYS
HEME METABOLISM
ADIPOGENESIS
OXIDATIVE PHOSPHORYLATION
GLYCOLYSIS
MYC TARGETS V2
MYC TARGETS V1
FATTY ACID METABOLISM
CHOLESTEROL HOMEOSTASIS
MTORC1 SIGNALING
DNA REPAR
G2M CHECKPOINT
E2F TARGETS
MITOTIC SPINDLE
P53 PATHWAY
APOPTOSIS
ALLOGRAFT REJECTION
UNFOLDED PROTEIN RESPONSE
HYPOXIA
TNFA SIGNALING VIA NFKB
INFLAMMATORY RESPONSE
TYPE I INTERFERON
TYPE II INTERFERON
METABOLISM
CELL CYCLE
APOPTOSIS
INFLAMMATION
NES
c
CELL SUBSETS
B-Cells
NK
Monocytes
MDC
T-Cells
PDC
d
CD8 TEM
CD8 TEMRA
CD8 TCM
Naïve CD8 T
Naïve CD4 T
CD4 TCM
CD4 TEM
IR vs INR-B
IR vs INR-A
INR-A vs INR-B
e
Heme biosynthesis
Ubiquitination & Antigen Presentation
Metabolism of nucleotides
Detoxification of Reactive Oxygen Species
Node Size (Reactome Pathways)
NUMBER OF GENES
Node Color
TERM P-VALUE
(BONFERRONI STEP-DOWN Corrected)
*Edges link Reactome pathways with enriched member genes
f
Classifier ROC
True Positive Rate
False Positive Rate
AUC = 0.86
Acc = 0.81
g
PCA: classifier genes (SCOPE Cohort)
INR-A
INR-B
IR
2nd dimension (18%)
1st dimension (25%)
h
p = 0.021
p = 0.046
TILDA
(cell with msHIV RNA/10^6 CD4 T cells)
Senescent INRs (INR-A)
Inflammatory INRs (INR-B)
IRs
i
BCL2L1
CYB5R3
FOXO3
FOXO4
GATA1
GLUL
NFE2
PINK1
PRDX2
RIOK3
CD4 Counts
Class
Inflammatory_INR
Senescent_INR
IR
Log(TILDA)
Z-score
TILDA Score vs CLASSIFIER: NES = 3.22, p < 0.0001

Fig. 1

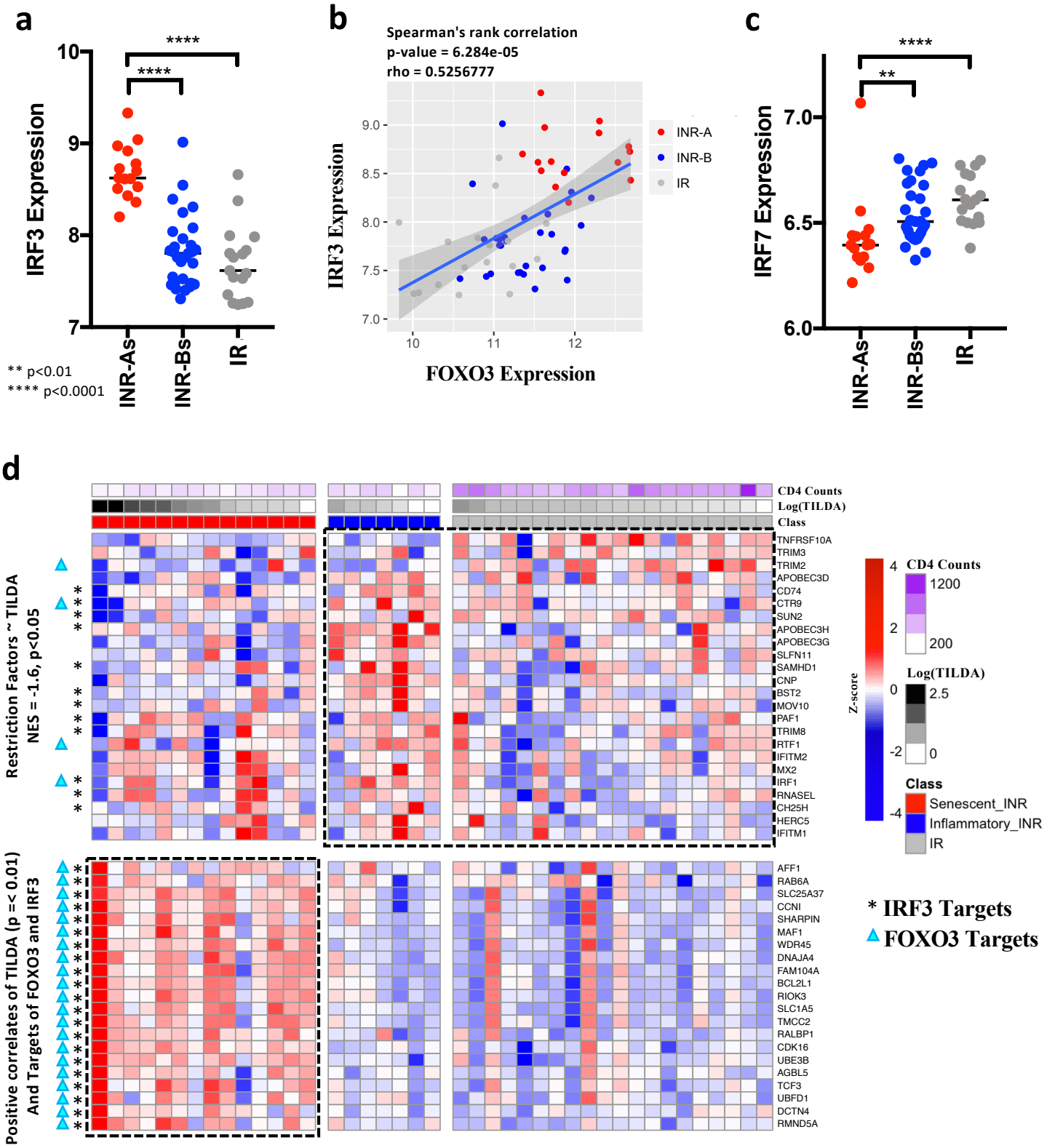

a
IRF3 Expression
INR-As
INR-Bs
IR
** p<0.01
**** p<0.0001
b
Spearman's rank correlation
p-value = 6.284e-05
rho = 0.5256777
IRF3 Expression
FOXO3 Expression
INR-A
INR-B
IR
c
IRF7 Expression
INR-As
INR-Bs
IR
d
Restriction Factors ~ TILDA
NES = -1.6, p<0.05
Positive correlates of TILDA (p =< 0.01)
And Targets of FOXO3 and IRF3
CD4 Counts
Log(TILDA)
Class
TNFRSF10A
TRIM3
TRIM2
APOBEC3D
CD74
CTR9
SUN2
APOBEC3H
APOBEC3G
SLFN11
SAMHD1
CNP
BST2
MOV10
PAF1
TRIM8
RTF1
IFITM2
MX2
IRF1
RNASEL
CH25H
HERC5
IFITM1
AFF1
RAB6A
SLC25A37
CCNI
SHARPIN
MAF1
WDR45
DNAJA4
FAM104A
BCL2L1
RIOK3
SLC1A5
TMCC2
RALBP1
CDK16
UBE3B
AGBL5
TCF3
UBFD1
DCTN4
RMND5A
Z-score
Senescent_INR
Inflammatory_INR
IR
* IRF3 Targets
FOXO3 Targets

Fig. 2

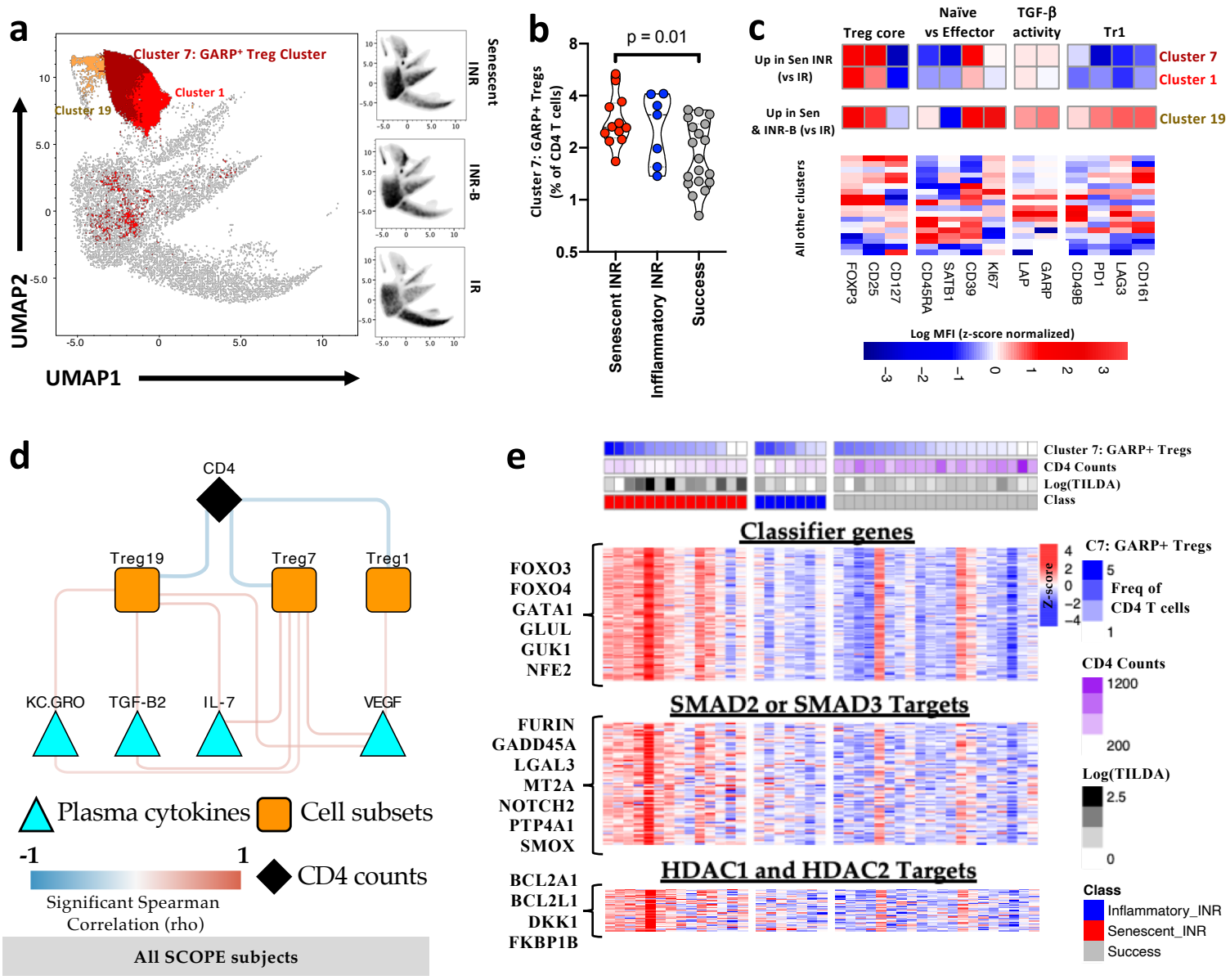

a
Cluster 7: GARP+ Treg Cluster
Cluster 1
Cluster 19
UMAP2
UMAP1
Senescent INR
INR-B
IR
b
p = 0.01
Cluster 7: GARP+ Tregs (% of CD4 T cells)
Senescent INR
Inflammatory INR
Success
c
Treg core
Naïve vs Effector
TGF-β activity
Tr1
Up in Sen INR (vs IR)
Up in Sen & INR-B (vs IR)
Cluster 7
Cluster 1
Cluster 19
All other clusters
FOXP3
CD25
CD127
CD45RA
SATB1
CD39
KI67
LAP
GARP
CD49B
PD1
LAG3
CD161
Log MFI (z-score normalized)
d
CD4
Treg19
Treg7
Treg1
KC.GRO
TGF-B2
IL-7
VEGF
Plasma cytokines
Cell subsets
CD4 counts
-1
1
Significant Spearman Correlation (rho)
All SCOPE subjects
e
Cluster 7: GARP+ Tregs
CD4 Counts
Log(TILDA)
Class
Classifier genes
FOXO3
FOXO4
GATA1
GLUL
GUK1
NFE2
SMAD2 or SMAD3 Targets
FURIN
GADD45A
LGAL3
MT2A
NOTCH2
PTP4A1
SMOX
HDAC1 and HDAC2 Targets
BCL2A1
BCL2L1
DKK1
FKBP1B
Z-score
C7: GARP+ Tregs
Freq of CD4 T cells
CD4 Counts
1200
200
Log(TILDA)
2.5
0
Class
Inflammatory_INR
Senescent_INR
Success

Fig. 3

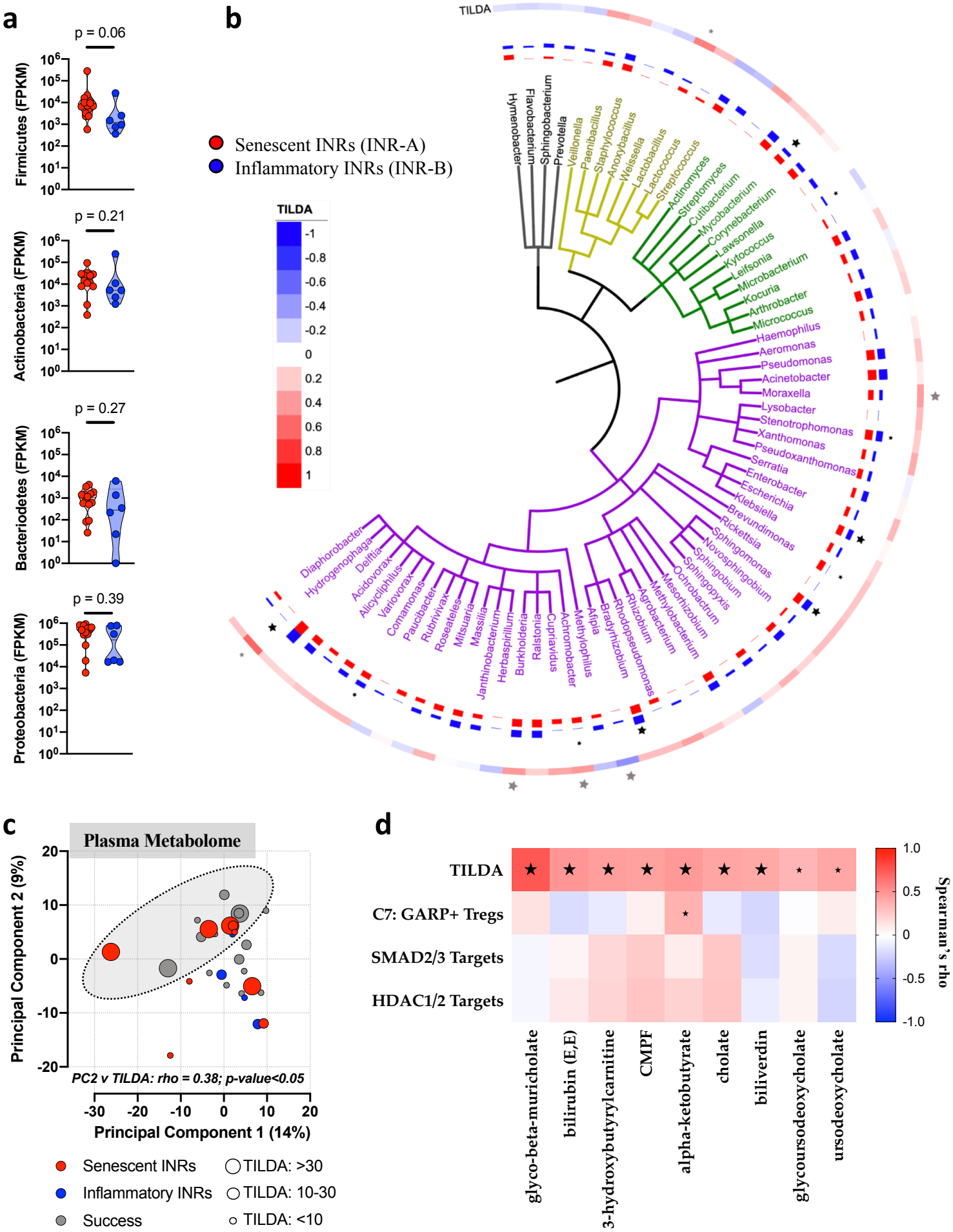
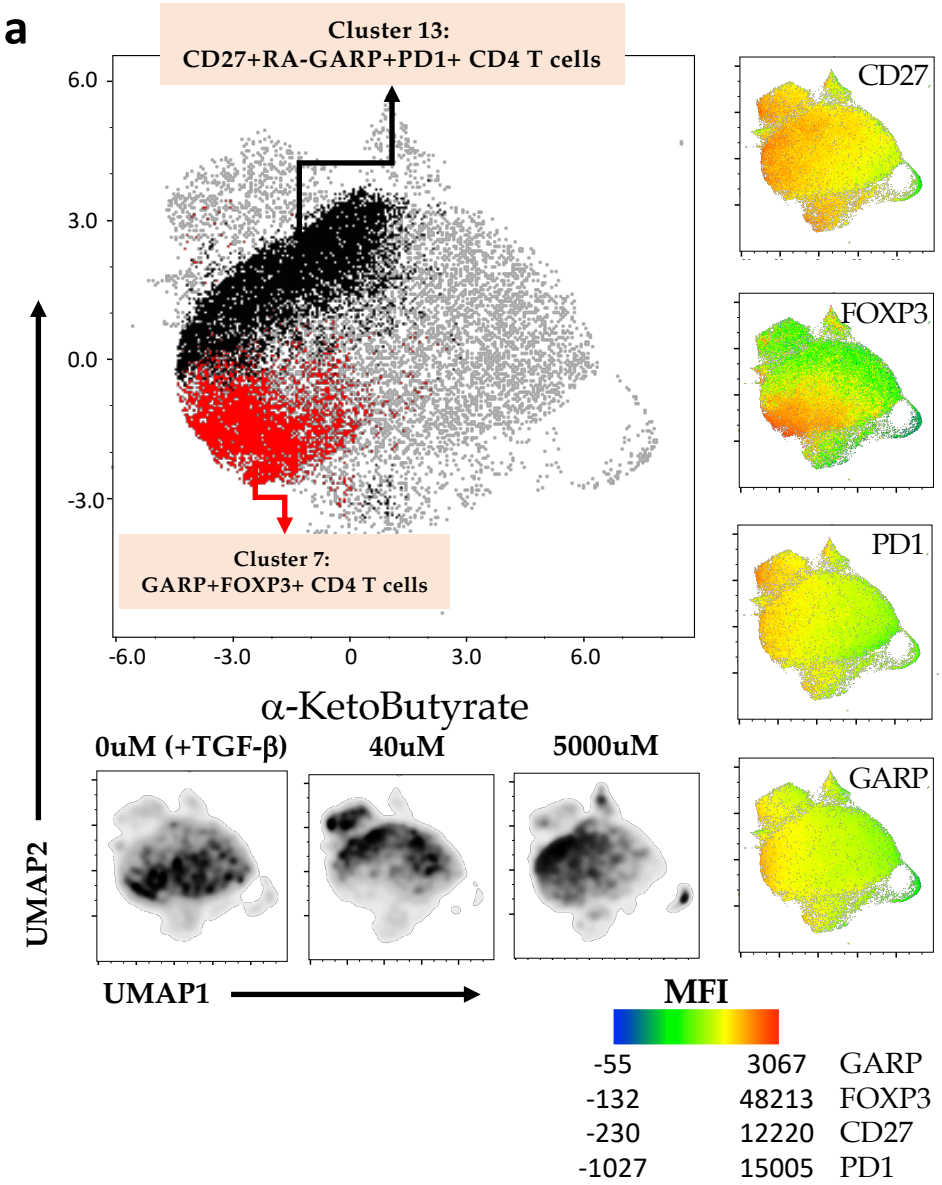
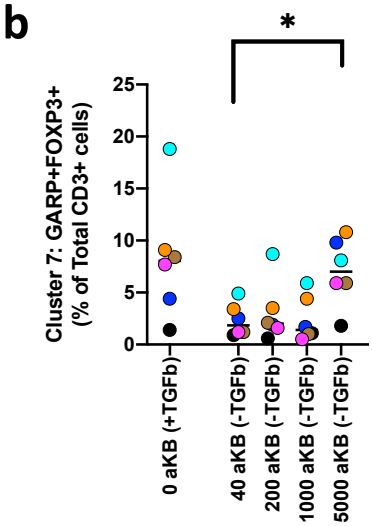
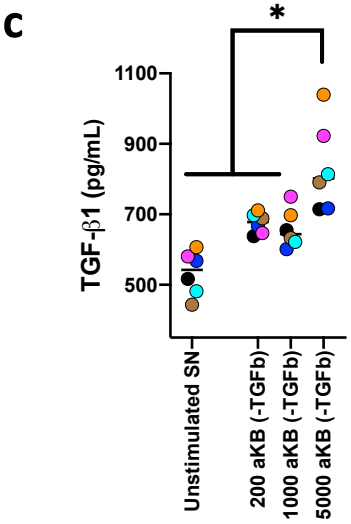
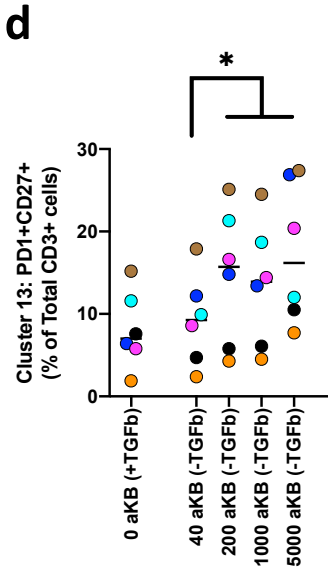

a
p = 0.06
Firmicutes (FPKM)
p = 0.21
Actinobacteria (FPKM)
p = 0.27
Bacteriodetes (FPKM)
p = 0.39
Proteobacteria (FPKM)
b
TILDA
Senescent INRs (INR-A)
Inflammatory INRs (INR-B)
TILDA
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1
Hymenobacter
Flavobacterium
Sphingobacterium
Prevotella
Veillonella
Paenibacillus
Staphylococcus
Anoxybacillus
Weissella
Lactobacillus
Lactococcus
Streptococcus
Actinomyces
Streptomyces
Cutibacterium
Mycobacterium
Corynebacterium
Lawsonella
Kytococcus
Leifsonia
Microbacterium
Kocuria
Arthrobacter
Micrococcus
Haemophilus
Aeromonas
Pseudomonas
Acinetobacter
Moraxella
Lysobacter
Stenotrophomonas
Xanthomonas
Pseudoxanthomonas
Serratia
Enterobacter
Escherichia
Klebsiella
Brevundimonas
Rickettsia
Sphingomonas
Novosphingobium
Sphingobium
Sphingopyxis
Ochrobactrum
Mesorhizobium
Methylobacterium
Agrobacterium
Rhizobium
Rhodopseudomonas
Bradyrhizobium
Afipia
Methylophilus
Achromobacter
Cupriavidus
Ralstonia
Burkholderia
Herbaspirillum
Janthinobacterium
Massilia
Mitsuaria
Roseateles
Rubrivivax
Paucibacter
Comamonas
Variovorax
Alicycliphilus
Acidovorax
Delftia
Hydrogenophaga
Diaphorobacter
c
Plasma Metabolome
Principal Component 2 (9%)
Principal Component 1 (14%)
PC2 v TILDA: rho = 0.38; p-value<0.05
Senescent INRs
Inflammatory INRs
Success
TILDA: >30
TILDA: 10-30
TILDA: <10
d
TILDA
C7: GARP+ Tregs
SMAD2/3 Targets
HDAC1/2 Targets
glyco-beta-muricholate
bilirubin (E,E)
3-hydroxybutyrylcarnitine
CMPF
alpha-ketobutyrate
cholate
biliverdin
glycoursodeoxycholate
ursodeoxycholate
Spearman's rho
1.0
0.5
0
-0.5
-1.0

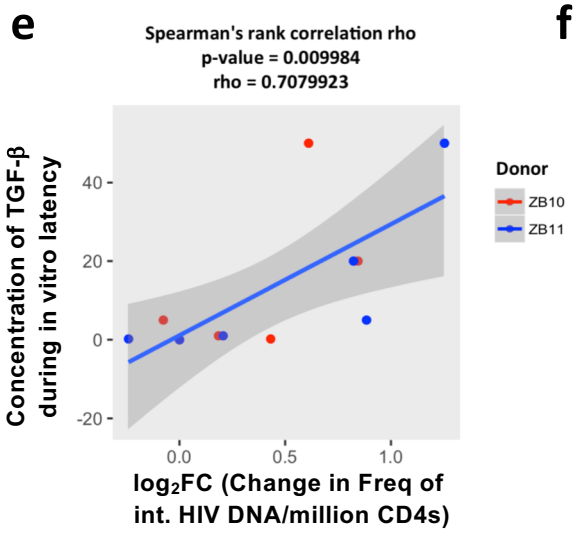
Fig. 4


a
Cluster 13:
CD27+RA-GARP+PD1+ CD4 T cells
Cluster 7:
GARP+FOXP3+ CD4 T cells
6.0
3.0
0.0
-3.0
-6.0
-3.0
0
3.0
6.0
CD27
FOXP3
PD1
GARP
α-KetoButyrate
0uM (+TGF-β)
40uM
5000uM
UMAP2
UMAP1
MFI
-55 3067 GARP
-132 48213 FOXP3
-230 12220 CD27
-1027 15005 PD1


b
*
Cluster 7: GARP+FOXP3+
(% of Total CD3+ cells)
25
20
15
10
5
0
0 aKB (+TGFb)
40 aKB (-TGFb)
200 aKB (-TGFb)
1000 aKB (-TGFb)
5000 aKB (-TGFb)


c
*
TGF-β1 (pg/mL)
1100
900
700
500
Unstimulated SN
200 aKB (-TGFb)
1000 aKB (-TGFb)
5000 aKB (-TGFb)


d
*
Cluster 13: PD1+CD27+
(% of Total CD3+ cells)
30
20
10
0
0 aKB (+TGFb)
40 aKB (-TGFb)
200 aKB (-TGFb)
1000 aKB (-TGFb)
5000 aKB (-TGFb)


e
Spearman's rank correlation rho
p-value = 0.009984
rho = 0.7079923
Concentration of TGF-β
during in vitro latency
40
20
0
-20
0.0
0.5
1.0
log2FC (Change in Freq of
int. HIV DNA/million CD4s)
Donor
ZB10
ZB11

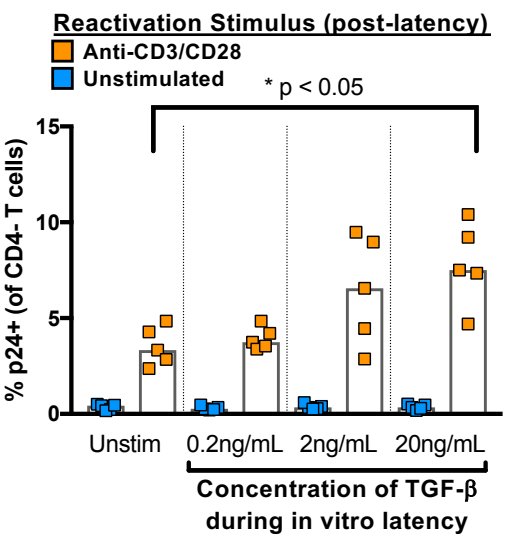

f
Reactivation Stimulus (post-latency)
Anti-CD3/CD28
Unstimulated
* p < 0.05
% p24+ (of CD4- T cells)
15
10
5
0
Unstim
0.2ng/mL
2ng/mL
20ng/mL
Concentration of TGF-β
during in vitro latency

Fig. 5

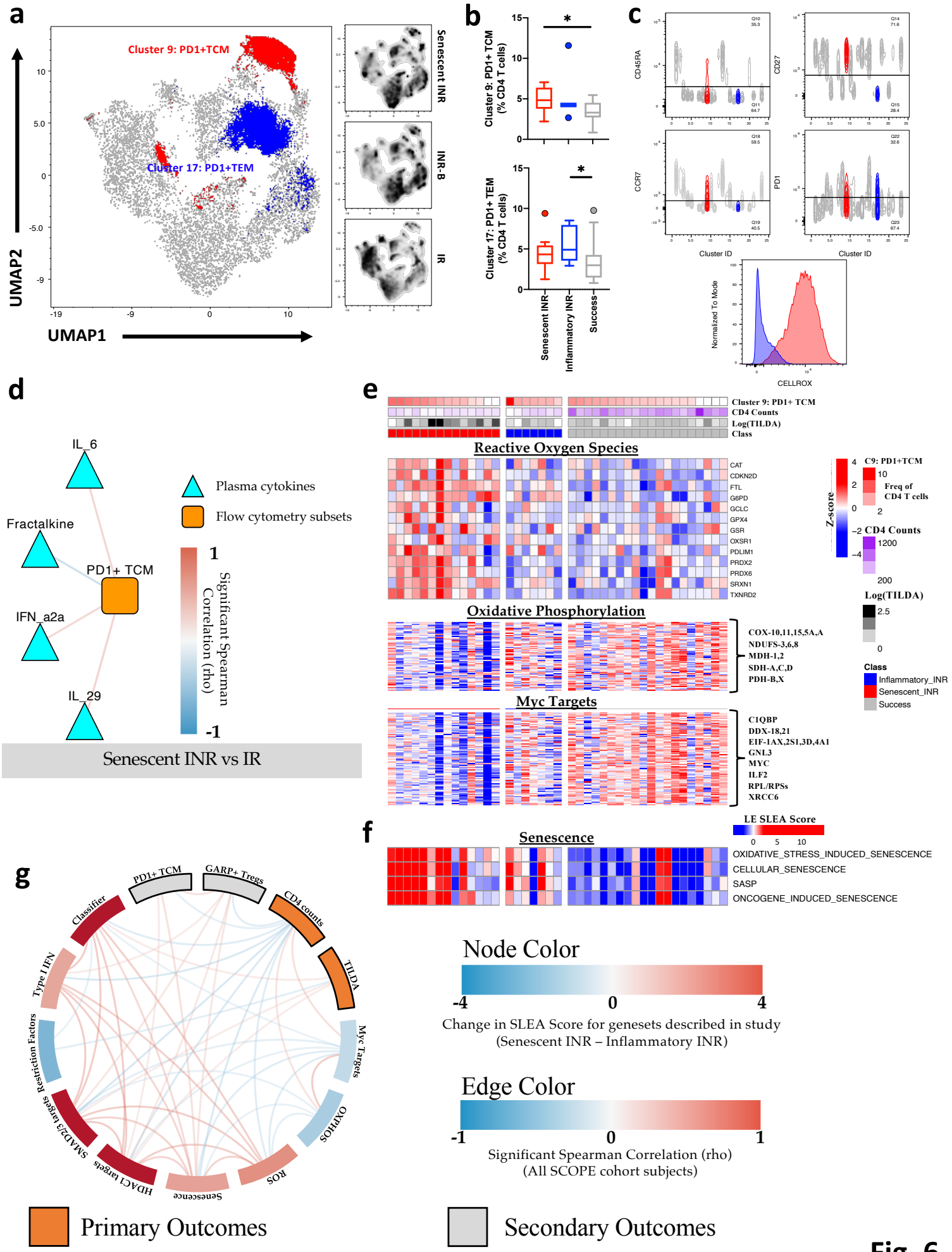
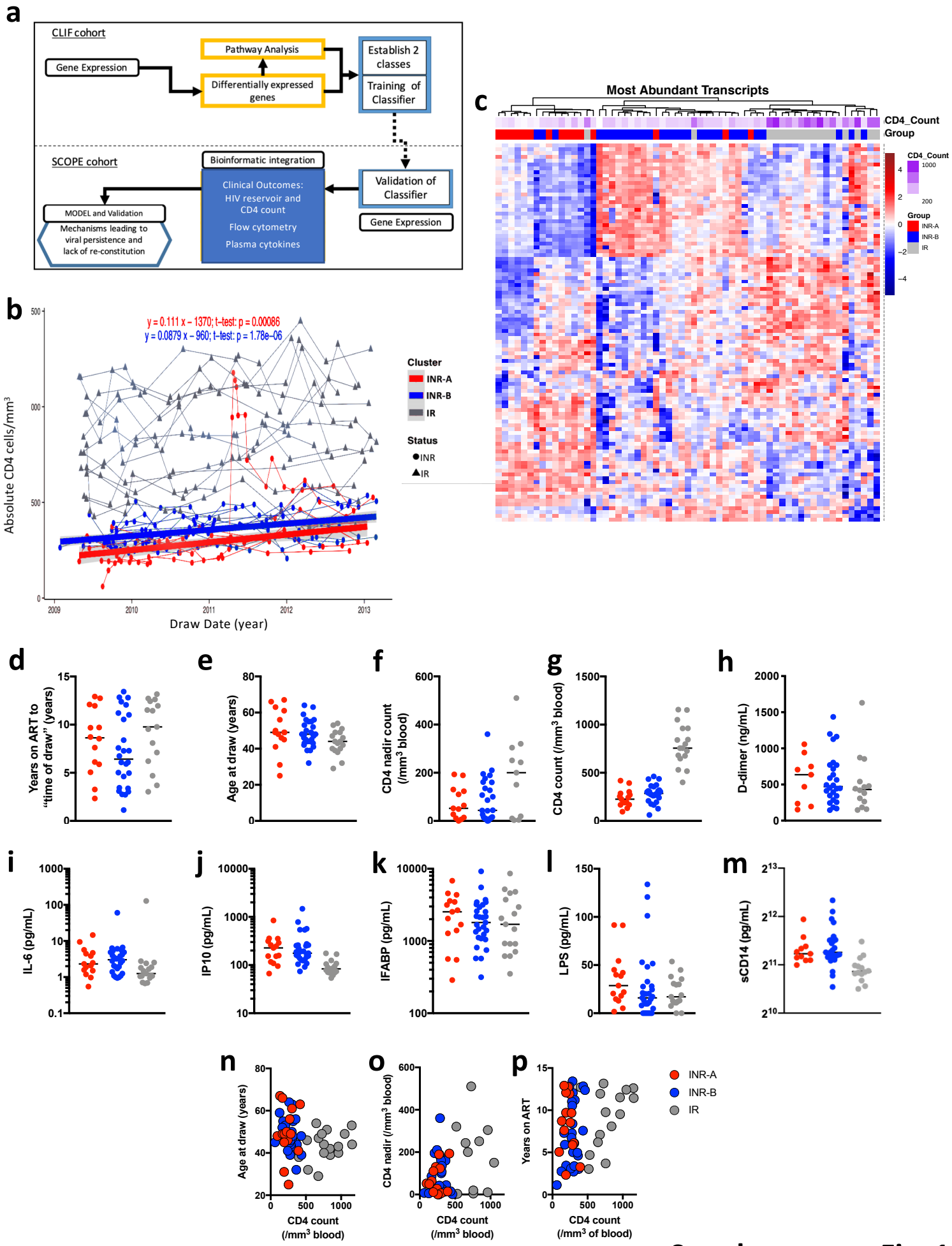
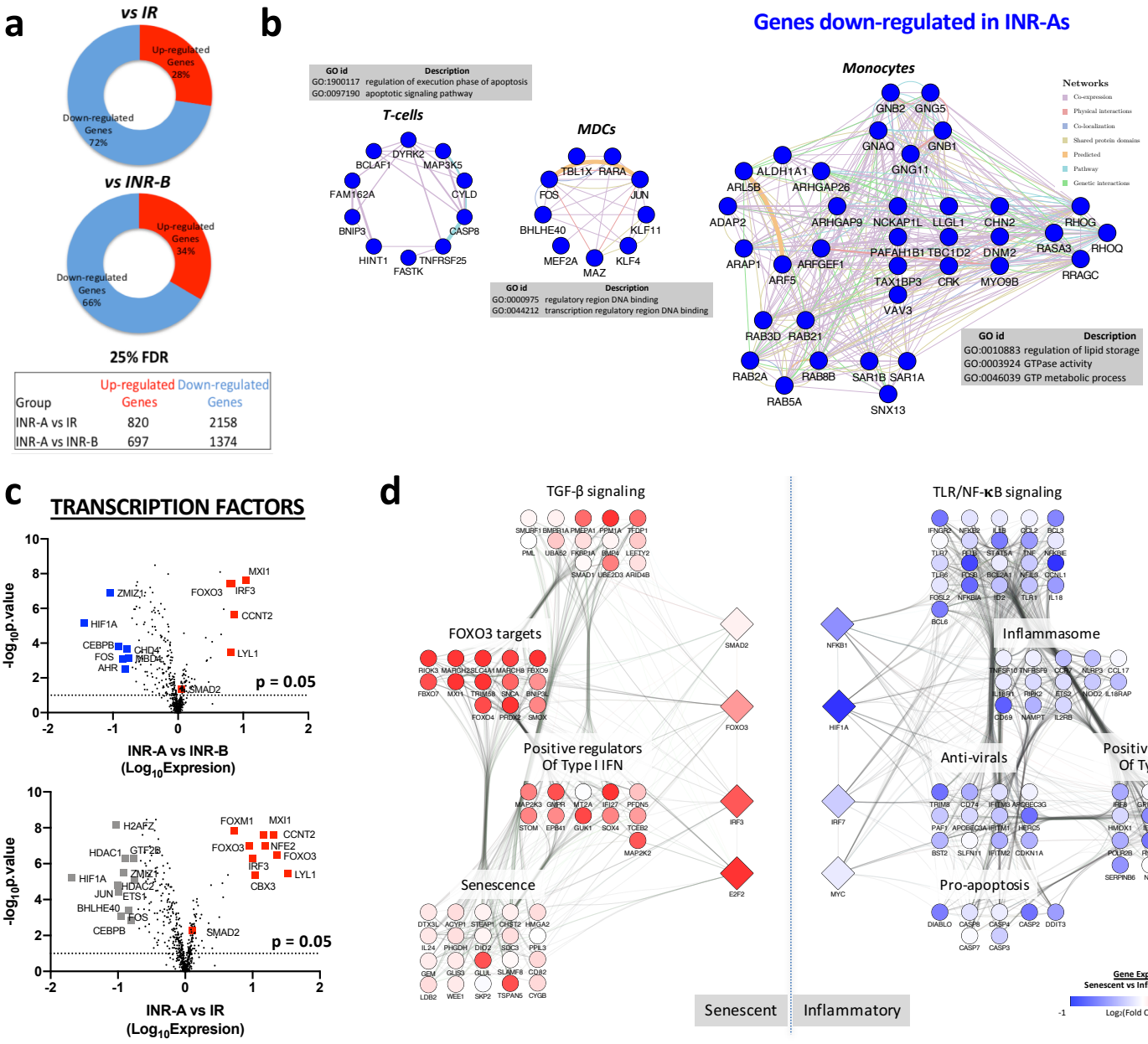

a
Cluster 9: PD1+TCM
Cluster 17: PD1+TEM
UMAP2
UMAP1
Senescent INR
INR-B
IR
b
Cluster 9: PD1+ TCM (% CD4 T cells)
Cluster 17: PD1+ TEM (% CD4 T cells)
Senescent INR
Inflammatory INR
Success
c
CD45RA
CD27
CCR7
PD1
Cluster ID
Normalized To Mode
CELLROX
d
IL_6
Fractalkine
PD1+ TCM
IFN_a2a
IL_29
Plasma cytokines
Flow cytometry subsets
Significant Spearman Correlation (rho)
Senescent INR vs IR
e
Cluster 9: PD1+ TCM
CD4 Counts
Log(TILDA)
Class
Reactive Oxygen Species
CAT
CDKN2D
FTL
G6PD
GCLC
GPX4
GSR
OXSR1
PDLIM1
PRDX2
PRDX6
SRXN1
TXNRD2
Z-score
C9: PD1+TCM
Freq of CD4 T cells
CD4 Counts
Log(TILDA)
Oxidative Phosphorylation
COX-10,11,15,5A,A
NDUFS-3,6,8
MDH-1,2
SDH-A,C,D
PDH-B,X
Myc Targets
C1QBP
DDX-18,21
EIF-1AX,2S1,3D,4A1
GNL3
MYC
ILF2
RPL/RPSs
XRCC6
Class
Inflammatory_INR
Senescent_INR
Success
f
Senescence
LE SLEA Score
OXIDATIVE_STRESS_INDUCED_SENESCENCE
CELLULAR_SENESCENCE
SASP
ONCOGENE_INDUCED_SENESCENCE
g
PD1+ TCM
GARP+ Tregs
CD4 counts
TILDA
Myc Targets
OxPHOS
ROS
Senescence
HDAC1 targets
SMAD2/3 targets
Restriction Factors
Type I IFN
Classifier
Node Color
Change in SLEA Score for genesets described in study
(Senescent INR – Inflammatory INR)
Edge Color
Significant Spearman Correlation (rho)
(All SCOPE cohort subjects)
Primary Outcomes
Secondary Outcomes

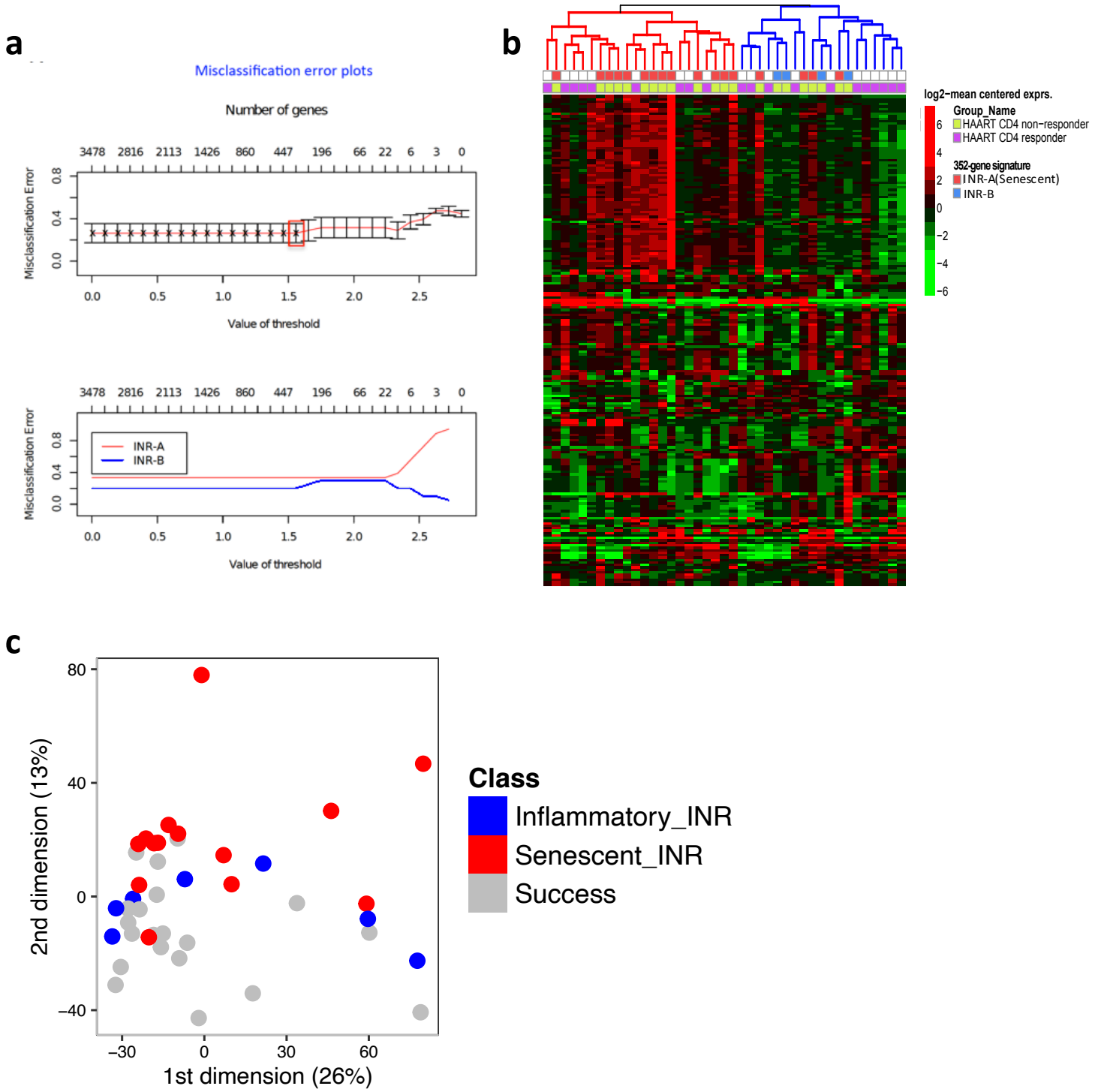
Fig. 6




**Supplementary Fig. 1**

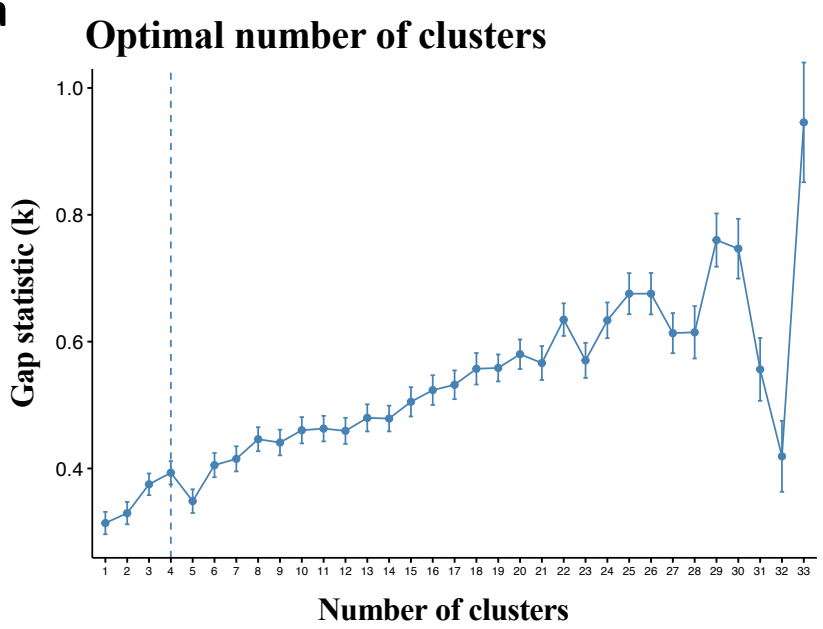



**Supplementary Fig. 2**

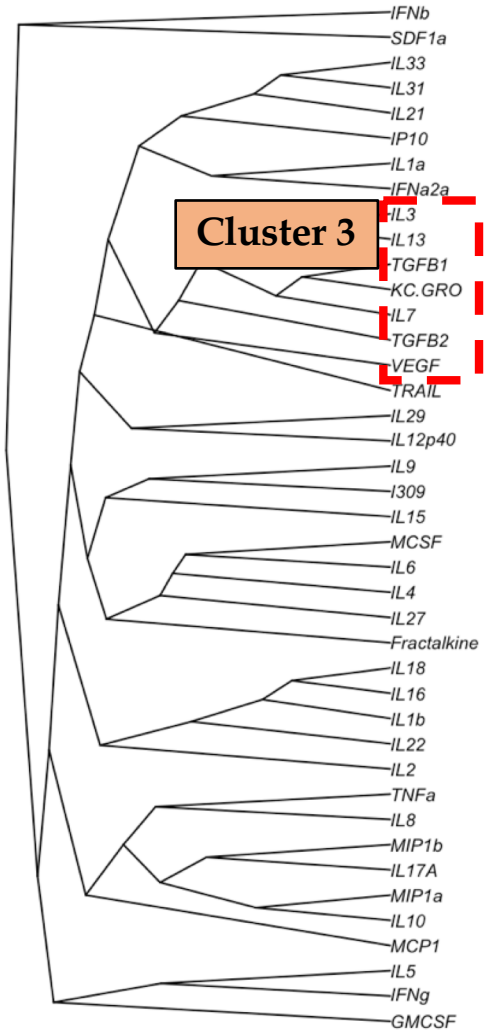



**Supplementary Fig. 3**

**a**



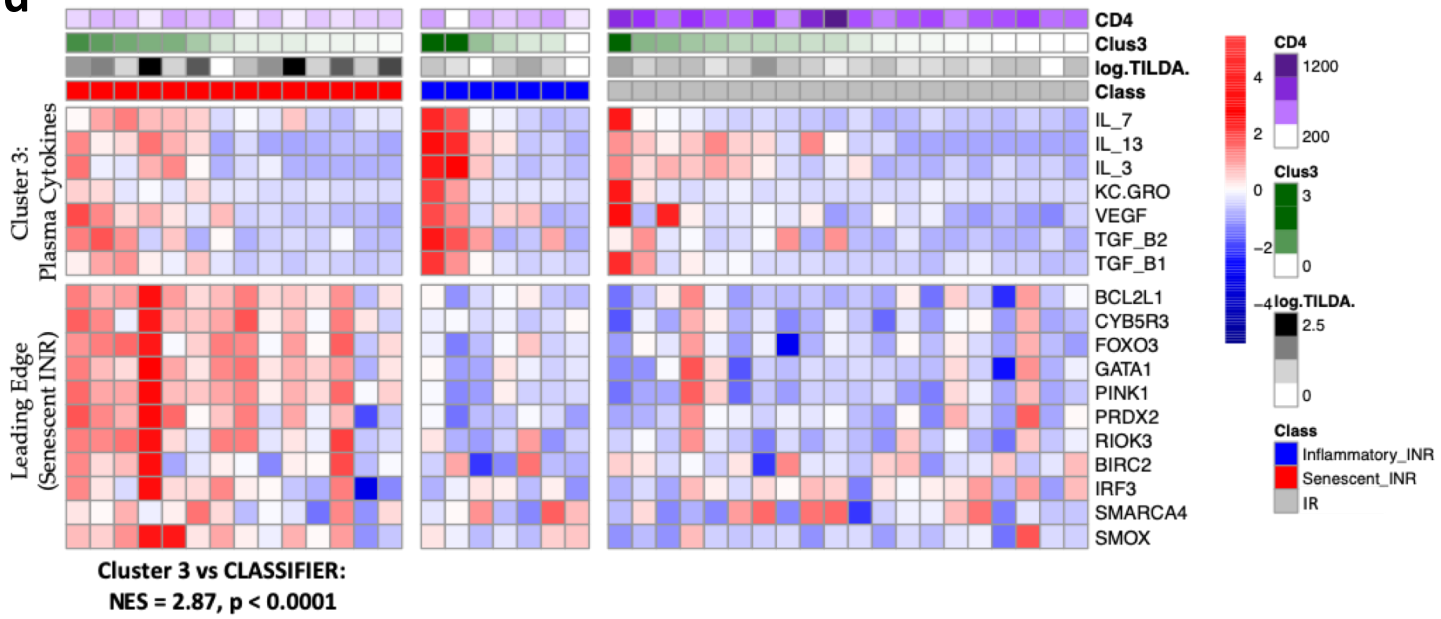

**b**



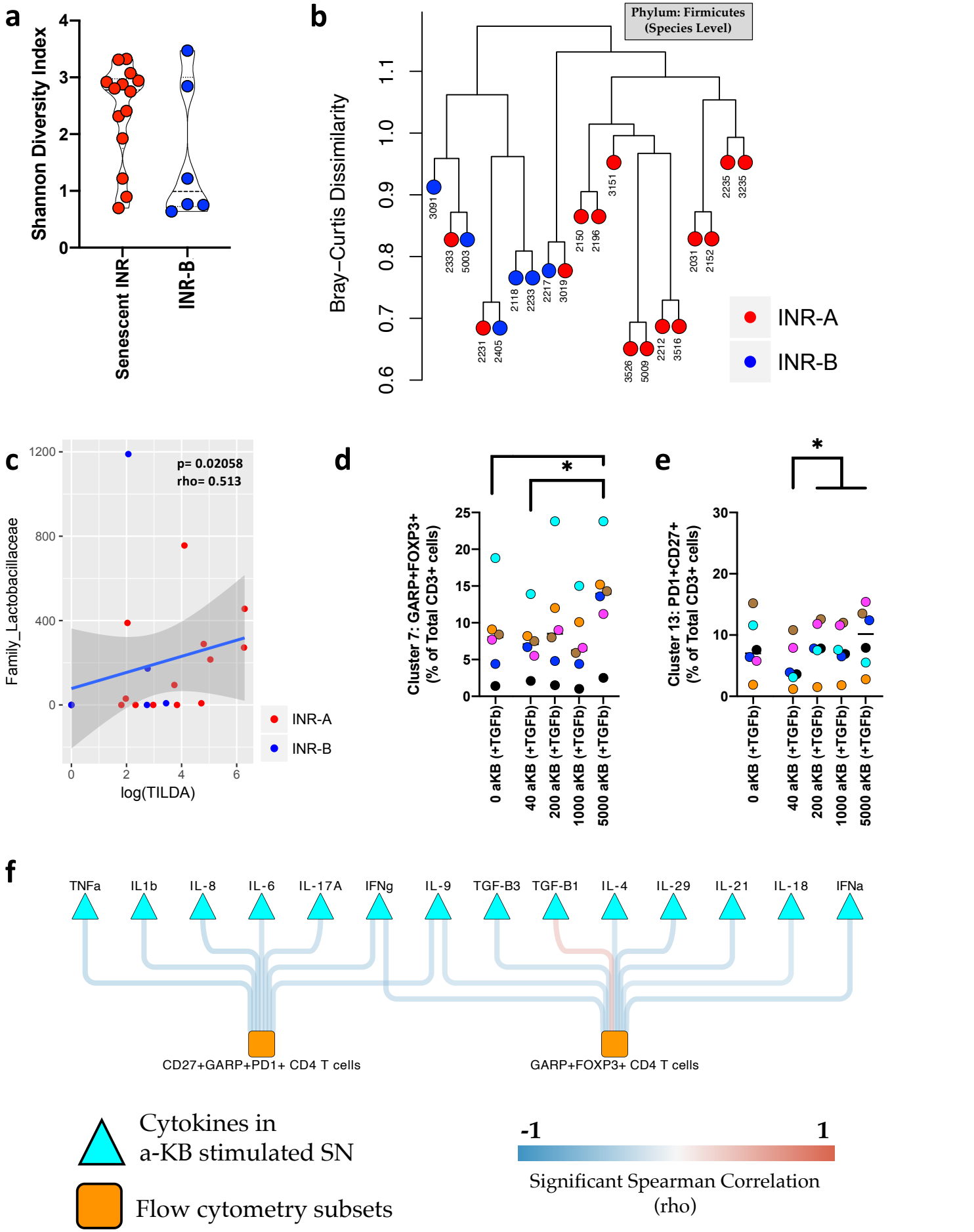

**c**

| Cytokine | CytokineSet | Cytokine | CytokineSet | Cytokine | CytokineSet | Cytokine | CytokineSet |
|---|---|---|---|---|---|---|---|
| GM_CSF | 1 | IFN_a2a | 2 | IL_7 | 3 | IL_1b | 4 |
| IFN_g | 1 | IL_12p40 | 2 | IL_13 | 3 | IL_8 | 4 |
| IL_10 | 1 | IL_9 | 2 | IL_3 | 3 | IL_16 | 4 |
| IL_4 | 1 | TRAIL | 2 | KC.GRO | 3 | IL_18 | 4 |
| IL_5 | 1 | IL_21 | 2 | VEGF | 3 | IL_1a | 4 |
| IL_6 | 1 | IL_29 | 2 | TGF_B2 | 3 | IL_22 | 4 |
| Fractalkine | 1 | IP_10 | 2 | TGF_B1 | 3 | | |
| IL_15 | 1 | | | | | | |
| SDF_1a | 1 | | | | | | |
| IFN_b | 1 | | | | | | |
| M_CSF | 1 | | | | | | |
| TNF_a | 1 | | | | | | |
| IL_27 | 1 | | | | | | |
| MCP_1 | 1 | | | | | | |

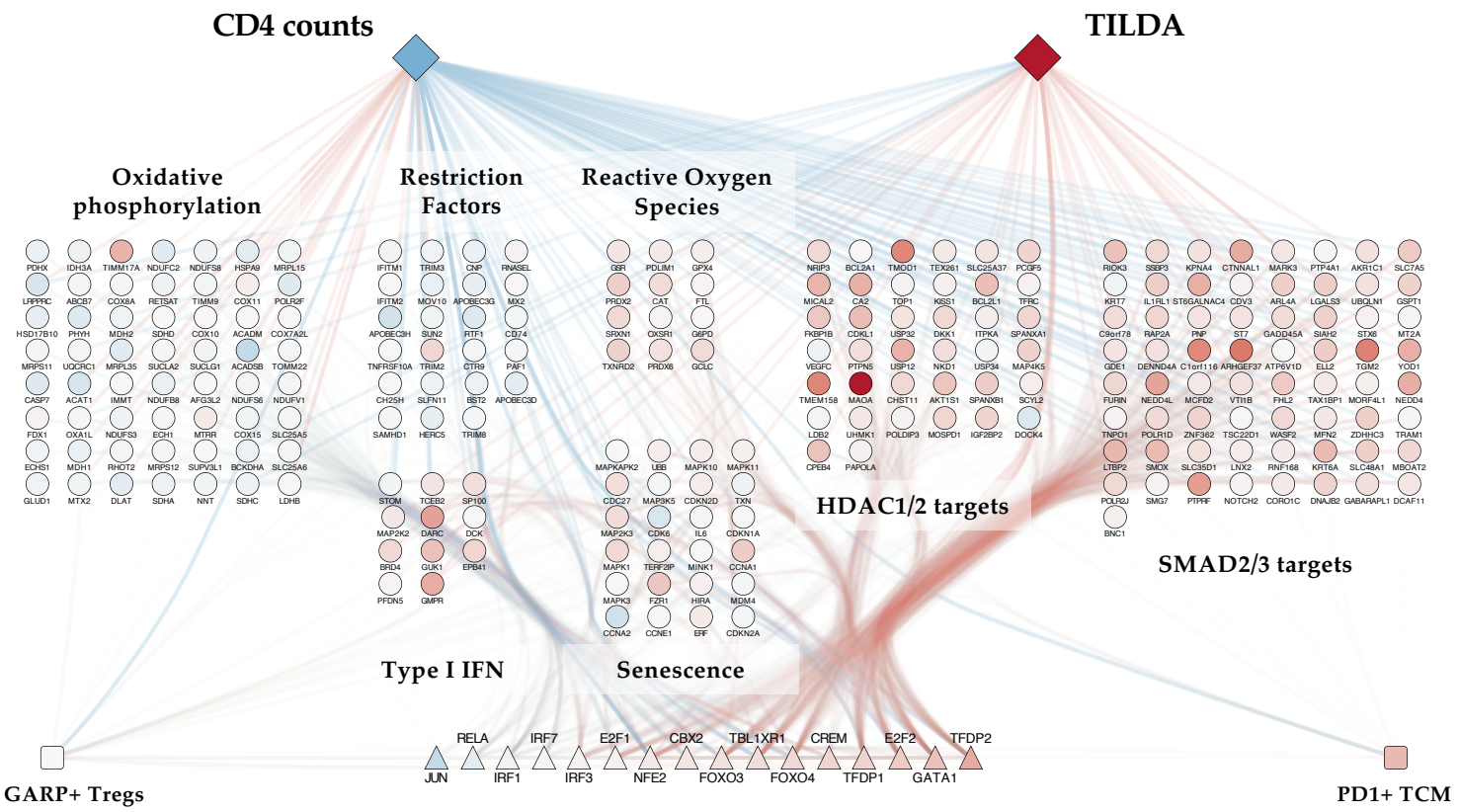
**d**




**Supplementary Fig. 4**




**Supplementary Fig. 5**



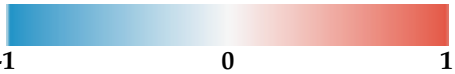
Outcomes

Transcription Factor Expression

CD4 subsets

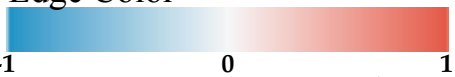
Gene expression

Node Color



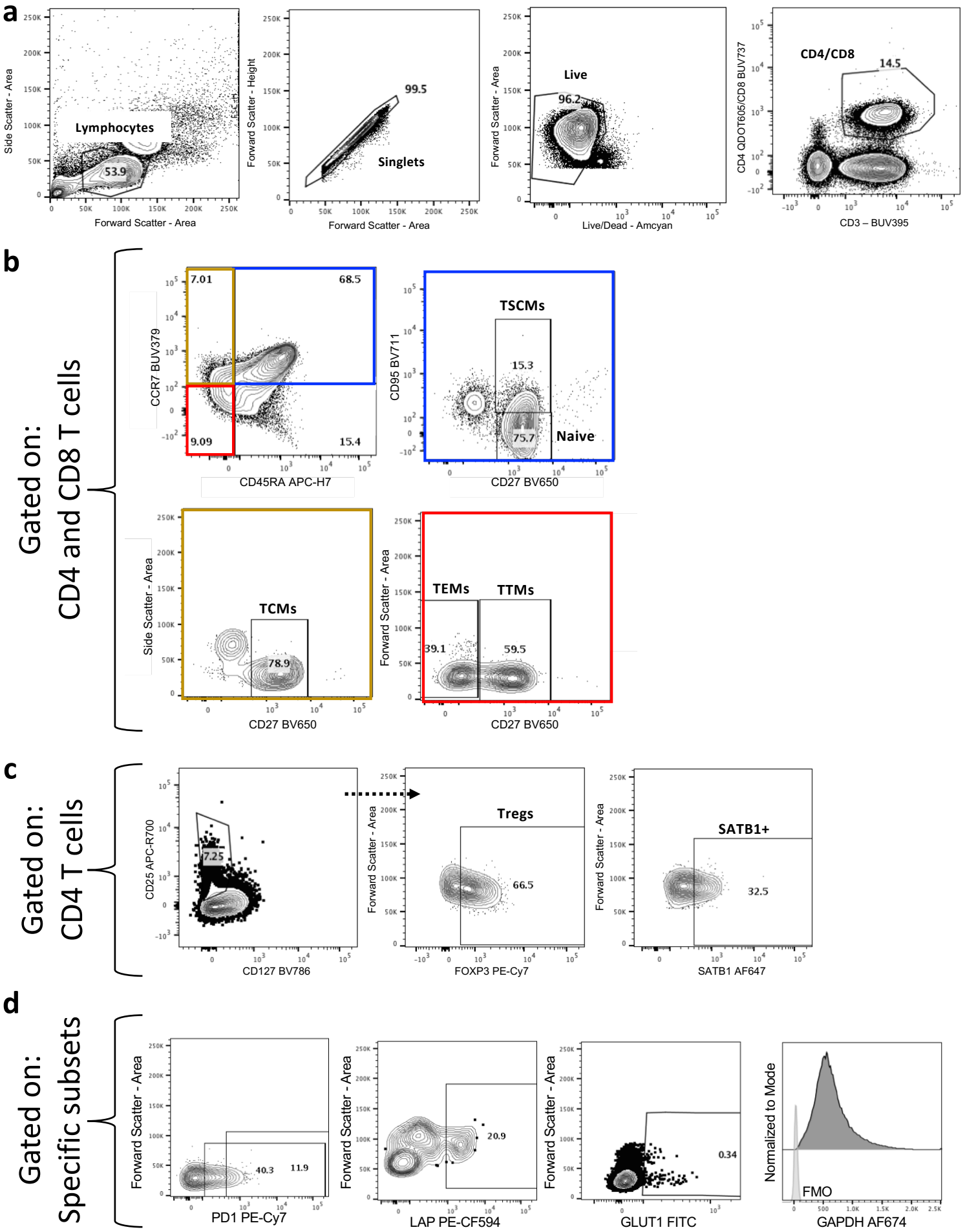

Log2FC (Senescent vs Inflammatory INRs)

Edge Color




Significant Spearman Correlation (rho)
(All SCOPE cohort subjects)

**Supplementary Fig. 6**




**Supplementary Fig. 7**