# CyTOF-Lec: Single-cell Glycomics Analysis Reveals Glycan Features Defining Cells Differentially Susceptible to HIV

Tongcui Ma[1, 2], Matthew McGregor[1, 2], Leila B. Giron[3], Guorui Xie[1, 2], Ashley F. George[1, 2], Mohamed Abdel-Mohsen[3, *], and Nadia R. Roan[1, 2, *]

[1] Gladstone Institutes, San Francisco, CA, USA

[2] Department of Urology, University of California, San Francisco, CA, USA

[3] The Wistar Institute, Philadelphia, PA, USA

[*]Co-Correspondence:

nadia.roan@gladstone.ucsf.edu, mmohsen@wistar.org

## ABSTRACT

High-parameter single-cell phenotyping has enabled in-depth classification and interrogation of immune cells, but to date has not allowed for glycan characterization. Here, we develop CyTOF-Lec as an approach to simultaneously characterize many protein and glycan features of human immune cells at the single-cell level. We implemented CyTOF-Lec to compare glycan features between different immune subsets from blood and multiple tissue compartments, and to characterize HIV-infected cell cultures. Using bioinformatics approaches to distinguish preferential infection of cellular subsets from viral-induced remodeling, we demonstrate that HIV upregulates the levels of cell surface fucose and sialic acid in a cell-intrinsic manner, and that memory CD4+ T cells co-expressing high levels of fucose and sialic acid are highly susceptible to HIV infection. Sialic acid levels were found to distinguish memory CD4+ T cell subsets expressing different amounts of viral entry receptors, pro-survival factors, homing receptors, and activation markers, and to play a direct role in memory CD4+ T cells' susceptibility to HIV infection. The ability of sialic acid to distinguish memory CD4+ T cells with different susceptibilities to HIV infection was experimentally validated through sorting experiments. Together, these results suggest that HIV remodels not only cellular proteins but also glycans, and that glycan expression can differentiate memory CD4+ T cells with vastly different susceptibility to HIV infection.

## INTRODUCTION

Viruses generally need to hijack multiple host cell processes to complete their replication cycle. This virus-mediated manipulation of host processes is called viral-induced remodeling, and has been studied using a variety of approaches, most entailing virally-infected cell lines analyzed in bulk. More recently, viral remodeling of primary cells was studied at the single-cell level by characterizing, via cytometry by time of flight (CyTOF), human tonsillar T cells infected *ex vivo* by varicella zoster virus (VZV) (Sen et al., 2014). Bioinformatics analyses of high-dimensional CyTOF datasets of VZV-infected cells revealed that VZV infection elicits significant host cell remodeling and alters the skin-trafficking property of subsets of infected cells. We recently implemented a follow-up approach, termed predicted precursor as determined by single-cell linkage using distance estimation (PP-SLIDE), to document HIV-induced remodeling of T cells from blood, lymph node, and genital tract, define the subsets of cells most susceptible to HIV infection, and characterize the phenotypes of HIV-infected cells in viremic and virally-suppressed people living with HIV (Cavrois et al., 2017; Ma et al., 2020; Neidleman et al., 2020b; Xie et al., 2021).

One important feature of PP-SLIDE is that it enables assessment of whether a receptor differentially expressed on HIV-infected cells reflects HIV-induced remodeling or preferential infection of cells harboring that pattern of expression of the receptor. For example, PP-SLIDE established that HIV-infected T cells express low levels of CD4 and CD28 not because HIV preferentially infects $CD4^{low}CD28^{low}$ T cells, but rather because HIV downregulates these receptors (Cavrois et al., 2017; Ma et al., 2020; Neidleman et al., 2020b; Xie et al., 2021), which were independently shown to be down-modulated by HIV accessory genes (Garcia and Miller, 1991; Swigut et al., 2001). Other HIV-remodeled surface receptors identified by PP-SLIDE include those involved in T cell migration to lymph nodes, and markers of Tfh cells (Cavrois et al., 2017; Ma et al., 2020; Neidleman et al., 2020b; Xie et al., 2021). By contrast, the low levels of surface CD127 expression on HIV-infected tonsillar T cells reflected preferential sparing of

CD127$^{high}$ T cells from productive infection (Cavrois et al., 2017; Ma et al., 2020; Neidleman et al., 2020b; Xie et al., 2021). Subsequent studies demonstrated that CD127$^{high}$ memory T cells preferentially undergo latent infection by HIV (Hsiao et al., 2020). These and other PP-SLIDE-generated findings of preferential infection of cellular subsets have been experimentally validated through a variety of sorting experiments (Cavrois et al., 2017; Ma et al., 2020; Neidleman et al., 2020b; Xie et al., 2021). Together, these studies suggest that important insights into HIV pathogenesis and persistence can be gained from characterizing HIV-induced remodeling of primary cells at a single-cell level.

However, such remodeling studies – and in fact all phenotypic characterizations of virally-infected cells to date – have only examined the cells' proteomes. Completely overlooked has been the diverse collection of glycans that are assembled on the surface of all living cells (Williams and Thorson, 2009). Cell surface glycosylation plays critical roles in regulating multiple cellular processes and immune functions (Barrera et al., 2002), as well as cell-cell (de Freitas Junior et al., 2011) and cell-pathogen (Colomb et al., 2019; Everest-Dass et al., 2012; Giron et al., 2020c) interactions. Furthermore, multiple viruses (e.g., HSV-1, CMV, and HTLV1) have been shown to alter the surface glycosylation of infected cells (Hiraiwa et al., 2003; Kambara et al., 2002; Nystrom et al., 2007; Nystrom et al., 2009). To date, studies of host glycomes have been limited to analysis of bulk cells, using techniques such as mass spectrometry, liquid chromatography, and lectin microarrays (Chen et al., 2021). A recent study analyzing bulk populations of cells with different glycan features demonstrated that they harbor different levels of HIV transcripts (Colomb et al., 2020b), suggesting that the host cell glycome can affect HIV susceptibility and/or replication. However, robust tools to deeply-characterize, at the single-cell level, the glycan features of immune cells – including HIV-infected ones – are lacking to date.

In this study, we developed a new approach taking advantage of the high-parameter analysis capabilities of CyTOF (Bendall et al., 2011), to phenotype cells simultaneously for protein and glycan features. This was achieved through conjugating a collection of lectins

(proteins that specifically bind different types of glycans) to metal lanthanides, an approach that has previously been validated at the bulk level (Leipold et al., 2009). We call our approach CyTOF-Lec – as it combines traditional CyTOF (using lanthanide-conjugated antibodies) with lanthanide-conjugated lectins to characterize surface glycosylation patterns of cells – and applied it on both blood and tissue cells. Taking advantage of the high-dimensional nature of our resulting CyTOF-Lec datasets and our PP-SLIDE analysis pipeline, we set out to address the following two fundamental questions about HIV infection: 1) Does HIV preferentially infect cells exhibiting distinct glycan features, and 2) To what extent does HIV remodel the glycan features of its host cell?

## RESULTS

### Development and validation of CyTOF-Lec

To establish a methodology that could simultaneously characterize protein and glycan features at the single-cell level, we developed a panel of lanthanide metal-conjugated antibodies and lectins compatible with CyTOF, which we refer to as CyTOF-Lec (Table S1). As tonsils provide an abundant source of both T and B cells, we used these cells for our initial validation of the panel. First we confirmed that the staining patterns of the lanthanide-conjugated antibodies were consistent with the known differential expression of their target antigens on tonsillar T vs. B cells (Figure S1), and with results previously reported using CyTOF without lectin staining (Ma et al., 2020). To ensure that the lectin staining did not displace or alter the antibodies bound to their protein targets, we developed a protocol whereby surface antibody staining was completed prior to lectin staining (see Methods). We confirmed that all five lectins (AOL, MAL-1, WGA, UEA-1, ABA, see Table S2) conjugated to lanthanides stained both T and B cells (Figure S1). Furthermore, we confirmed that antibody binding to CD3, CD4, and CD8 was the same whether or not the specimens were subsequently stained with lectins (Figure 1A). To establish the

specificity of lectin binding, we assessed the effect of sialidase, which degrades cell-surface sialic acid. As expected, binding by WGA and MAL-1, which detect different forms of sialic acid (Table S2), was decreased after sialidase treatment of the cells (Figure 1B). By contrast, binding by AOL and UEA-1, which detect different forms of fucose, and ABA, which detects T antigen (Table S2), were all increased (Figure 1C-D). This was expected as removal of sialic acid should enable better detection of these other glycan structures, and is consistent with prior reports (Giron et al., 2020b).

To determine whether the conjugated lectins could detect differences in cell-surface glycans between subsets of immune cells, we compared binding by each of the 5 lectins to B cells, memory CD8+ T cells (CD8+ Tm cells), naïve CD8+ T cells (CD8+ Tn cells), memory CD4+ T cells (CD4+ Tm cells), and naïve CD4+ T cells (CD4+ Tn cells). We assessed the expression of glycans on these subsets not only among tonsillar cells, but also among PBMCs and endometrial T cells for comparison. For all three sites, the fucose-specific lectins AOL and UEA1 bound more to CD4+ Tm and CD8+ Tm cells than to their respective naïve counterparts, although the difference only reached significance for AOL binding in tonsils (Figure 2A). These results are consistent with the known role of core surface fucosylation for T cell activation, which is more prominent within the memory compartment (Fujii et al., 2016; Liang et al., 2018). The sialic acid binding lectins WGA and MAL1 also consistently bound to memory T cells more than to their naïve counterparts, although this only reached statistical significance for MAL1 binding of endometrial CD8+ T cells (Figure 2B). Binding by ABA did not show a consistent pattern between memory vs. naïve T cells (Figure 2C). Binding by all five lectins was low on B cells, particularly as compared to memory CD8+ T cells (Figure 2A-C).

These results establish CyTOF-Lec as a panel that can quantitate glycan and protein expression at the single-cell level, and detect differential glycan expression between different subsets of immune cells.

**HIV alters expression of fucose and sialic acid in a tissue site-dependent manner**

We next applied CyTOF-Lec to determine the extent to which glycans are remodeled on the surface of HIV-infected cells. Fresh endometrial biopsies (n=6 donors) and whole tonsils from tonsillectomies (n=4 donors) from HIV seronegative individuals were processed into single-cell suspensions, and then immediately exposed to the HIV-F4.HSA, a replication-competent and Nef-sufficient virus that harbors the CCR5-tropic 109FPB4 transmitted/founder (T/F) Env (Cavrois et al., 2017). In addition, PBMCs were isolated from whole blood of HIV seronegative individuals (n=4 donors) and exposed to HIV-F4.HSA. To limit potential confounding effects of *ex vivo* induced T cell activation, infection was carried on unstimulated cells. Infection was allowed to proceed for 3 days, after which the cells, as well as cells from paired uninfected control cultures, were harvested for CyTOF-Lec analysis. CD4+ T cells were identified as intact, live singlet CD3+CD8- cells, while infected cells were identified as intact, live singlet CD3+CD8- $CD4^{Low}$ cells (Figure S2), to account for the down-regulation of cell-surface CD4 by HIV (Doms and Trono, 2000; Garcia and Miller, 1991; Lama, 2003; Piguet et al., 1999). Consistent with our prior studies (Ma et al., 2020), endometrial T cells were the most susceptible to HIV-F4.HSA infection (Figure S3A). HIV-infected cells from all three sites were remodeled, as established qualitatively by assessing their locations on a tSNE (Figure S3B) as well as quantitively using SLIDE (Sen et al., 2014) (Figure S3C). The remodeling of the infected cells is consistent with prior studies (Cavrois et al., 2017; Ma et al., 2020; Xie et al., 2021), and confirms that CyTOF-Lec is a valid panel for the analysis of remodeling.

To identify specific glycans that were remodeled, we implemented PP-SLIDE to identify the Predicted Precursor (PRE) cells (Cavrois et al., 2017; Ma et al., 2020; Neidleman et al., 2020b; Xie et al., 2021). PRE cells harbor the predicted original (pre-remodeling) features of T cells infected by HIV and are identified using k-nearest neighbor approaches by matching, in the high-parameter CyTOF space, the T cells in the uninfected culture most similar in phenotype to every HIV-infected cell (Figure S4A). As expected (Ma et al., 2020), the PRE cells from all 3

sites were preferentially memory CD4+ T cells (Figure S4B, S4C). Having identified the PRE cells, we then determined which glycans, if any, were remodeled by HIV infection. This was accomplished by assessing for lectins that differentially bound the PRE as compared to the infected cells. Glycans were quantitated by reporting the median signal intensity (MSI) of their corresponding lectins among each population of cells from each donor. Interestingly, we found that both fucose and sialic acid were upregulated during HIV infection. Infected cells from all 3 sites potently upregulated total fucose as assessed by AOL binding (Figure 3A), although this upregulation did not reach statistical significance (after correcting for multiple comparisons) for the endometrium. Binding by UEA1, however, was not significantly different between infected and PRE cells, and tended to be downregulated in the endometrium and tonsils, and upregulated in PBMCs (Figure 3A). As UEA1 binds $\alpha$1-2 branched fucose (Table S2), these results suggest that fucosylation is globally upregulated upon infection of tissue CD4+ T cells with HIV, though not the type of fucosylation that creates $\alpha$1-2 branched structure. In contrast to the fucose-binding lectins, both sialic acid-binding lectins were increased on infected compared to PRE cells (Figure 3B), and this was observed for all three sites although results did not reach statistical significance for the endometrium. These results are consistent with an upregulation of both total sialic acid (recognized by WGA; which also binds to *N*-acetylglucosamine (GlcNAc)) (Schwarz et al., 1999) and $\alpha$2-3 linked sialic acid (recognized by MAL-1) by HIV during infection. In contrast to fucose and sialic acid, we did not observe any marked upregulation of T antigen on infected cells (Figure 3C).

**HIV preferentially infects memory CD4+ T cells with higher fucose and sialic acid levels**

In addition to revealing antigens that have been remodeled by infection, the PP-SLIDE approach can also identify antigens that are differentially expressed on cells before infection. In particular, antigens more abundant on PRE than uninfected cells correspond to antigens

186 preferentially expressed on HIV-susceptible cells, while those less abundant on PRE cells
187 correspond to those preferentially expressed on HIV-resistant cells. We used as our uninfected
188 population CD4+ Tm cells, excluding CD4+ Tn, CD8+ Tm, and CD8+ Tn cells because these
189 latter three populations harbored negligible numbers of HIV-susceptible cells (Figure S4B). This
190 exclusion was important because otherwise antigens differentially expressed between
191 uninfected and PRE cells could just reflect phenotypic differences between these major subsets.
192 Significant differences in lectin binding between PRE and uninfected (CD4+ Tm) cells were only
193 observed in PBMCs, with AOL, UEA1, WGA, and MAL-1 all bound at significantly higher levels
194 on PRE cells (Figure 3). Tonsillar PRE cells also bound these lectins more than their uninfected
195 counterparts did, but these results did not reach statistical significance. Endometrial PRE cells
196 did not show significant differences in lectin binding relative to their uninfected counterparts.
197 These results together with the remodeling analysis suggest that in blood and tonsils
198 (but not endometrium), HIV preferentially infects memory CD4+ T cells with higher levels of
199 fucose and sialic acid, and then further upregulates these cell surface glycans through viral
200 remodeling.
201
202 **HIV infection alters the surface glycome of bystander immune cells in tonsils**
203 Remodeling of cells can occur in a cell-intrinsic manner as a result of direct infection, but
204 may also result from bystander effects. For example, the inflammatory environment elicited by
205 HIV replication may elicit phenotypic changes in bystander (uninfected) cells in the infected
206 culture. We therefore assessed whether HIV infection elicits any glycosylation alterations in
207 bystander cells. To identify bystander memory CD4+ T cells, we gated the infected culture on
208 CD4+ Tm cells that were HSA-negative. Increased binding by all five lectins was observed
209 among bystander tonsillar CD4+ Tm cells relative to their counterparts from uninfected cultures
210 (Figure S5), suggesting that in at least some tissue sites, remodeling of glycans on bystander
211 CD4+ T cells occurs. Interestingly, however, relative to the bystander CD4+ Tm cells, the

infected cells still exhibited higher levels of total fucose and sialic acid (as assessed by AOL and WGA binding, respectively), suggesting possible additional cell-intrinsic glycan remodeling by replicating virus (Figure S5).

To examine whether HIV alters glycan expression in other bystander cellular subsets, we compared glycan levels on multiple subsets of B and T cells from the uninfected vs. infected cultures (Figure S6). Only tonsils exhibited significant differences between uninfected vs. bystander cells, and these differences were observed among all subsets. For example, sialic acid levels (as assessed by both WGA and MAL-1 binding) were significantly higher in all the analyzed subsets of bystander B, CD8+, and CD4+ T cells, relative to their counterparts from uninfected cultures. Fucose expression was also uniformly higher among bystander cells, although the difference among CD8+ Tm cells did not reach statistical significance. Differences in levels of T antigen were also observed among B cells, CD8+ T m cells, and CD4+ Tm cells (Figure S6). These glycan changes may be elicited by HIV infection-induced inflammatory cytokines (Breen et al., 1990; Contreras et al., 2003; Sugawara et al., 2019), which can alter cell surface glycosylation patterns (Dewald et al., 2016; Giron et al., 2020a).

**HIV preferentially infects memory CD4+ T cells from tonsils and PBMCs co-expressing high levels of fucose and sialic acid**

The data presented thus far suggest that although there are differences between blood vs. the tissue sites examined, fucose and sialic acid are upregulated on HIV productively-infected cells, and CD4+ T cells expressing high levels of fucose or sialic are preferentially targeted for infection. We next conducted manual gating to assess whether the HIV-susceptible cells express high levels of both fucose or sialic acid, or whether they belong to distinct subsets of fucose+ vs. sialic acid+ cells. We focused on the AOL and WGA datasets, as they cover total fucose and different forms of sialic acid, respectively. We first examined, within the HIV-infected cultures, the infection rates among CD4+ Tm cells expressing high vs. low levels of AOL or

238 WGA. In both blood and both tissue compartments, $AOL^{High}$ and $WGA^{High}$ cells exhibited
239 significantly higher HIV infection rates than did $AOL^{Low}$ and $WGA^{Low}$ cells, respectively (Figure
240 4A). To assess the extent to which this high level of infection was due to preferential infection of
241 the $AOL^{High}$ and $WGA^{High}$ CD4+ Tm cells, we next compared, among the uninfected CD4+ Tm
242 cells and PRE cells, the percentages of cells that were $AOL^{High}$, $WGA^{High}$, or $AOL^{High}$ $WGA^{High}$.
243 Consistent with the MSI data, the percentages of cells expressing high levels of AOL or WGA
244 were higher among PRE cells in both tonsils and PBMCs (Figure 4B), suggesting preferential
245 infection of fucose- and sialic acid-expressing cells by HIV at these sites. Interestingly, $AOL^{High}$
246 $WGA^{High}$ cells were also significantly over-represented among PRE cells at these sites (Figure
247 4B), suggesting that the HIV-susceptible Tm cells co-express fucose and sialic acid. Indeed,
248 visualization of the PRE cells by tSNE revealed cells binding high levels of both AOL and WGA
249 among the tonsillar and blood compartments (Figure 4C). Although PRE cells from the
250 endometrium did not preferentially express fucose or sialic acid (Figure 4B), PRE cells co-
251 expressing AOL and WGA could be detected from this site (Figure 4C), suggesting the
252 endometrium, like the other two sites, harbors HIV-susceptible cells co-expressing fucose and
253 sialic acid. To better understand why the HIV-susceptible endometrial CD4+ Tm cells, unlike
254 their tonsillar and blood counterparts, did not *preferentially* express high levels sialic acid or
255 fucose, we compared the levels of WGA and AOL binding on uninfected CD4+ Tm cells from
256 the three sites. This analysis revealed WGA, but not AOL, to be expressed at the highest levels
257 on the endometrial cells (Figure 4D). These results suggest that the reason $WGA^{high}$ cells are
258 not preferentially targeted for infection in the endometrium may be that endometrial CD4+ Tm
259 cells all express high levels of sialic acid. Fucose levels, however, are higher in PBMCs than
260 endometrium, suggesting that the fact that $AOL^{high}$ endometrial CD4+ Tm cells aren't
261 preferentially targeted for infection cannot be explained by exceptionally high levels of fucose
262 expression.

We then implemented a more global method of subset identification, using FlowSOM (Van Gassen et al., 2015). We combined the uninfected CD4+ Tm and PRE cells from all the donors, and identified 20 clusters for each of the 3 sites (Figure 5A). Endometrial T cells, which had the highest population of PRE cells, were represented among most of the 20 endometrial cell clusters. To determine the extent of enrichment of each cluster among PRE cells, we calculated the ratio of the size of each cluster in the PRE vs. total uninfected CD4+ Tm cells. Enriched clusters identified in this manner (corresponding to those preferentially harboring HIV-susceptible cells, see Methods) were detected among all three sites, with 8 clusters from the endometrium, 9 from tonsils, and 6 from PBMCs (Figure 5B). Interestingly, the fold-enrichment was highest among the tonsils (reaching almost 20-fold in one donor), suggesting that of all 3 sites, this one exhibits the most preferential selection of subsets for infection. Compared to the other two sites, the tonsils also harbored more enriched clusters with significantly elevated levels of fucose, sialic acid, and T antigens relative to their expression levels on uninfected CD4+ Tm cells (Figure S7). To assess whether the enriched clusters co-express fucose and sialic acid, we assessed by tSNE heatmaps the levels of AOL and WGA binding on concatenated files of all the enriched clusters from each site. This analysis revealed regions of the tSNEs co-expressing high levels of fucose and sialic acid (Figure 5C), confirming the manual gating data that HIV-susceptible cells co-express these two classes of glycans.

**Total sialylated glycan is a valid marker of highly-susceptible of CD4+ Tm cells expressing HIV entry receptors and activation markers, and may play a direct role in susceptibility**

The results presented thus far suggest that CD4+ Tm cells from endometrium, tonsils, and blood are preferentially susceptible to HIV infection compared to their naïve counterparts, but only in tonsils and blood can high levels of fucose and sialic acid further distinguish HIV-susceptible CD4+ Tm cells from non-susceptible CD4+ Tm cells. To experimentally validate

these findings, we conducted sorting experiments. As endometrial and tonsillar T cells do not maintain good viability after sorting, we limited these studies to blood specimens. CD4+ Tm cells from blood expressing low ($WGA^{Low}$), medium ($WGA^{Medium}$), or high ($WGA^{High}$) levels of sialic acid were isolated through sorting (Figure 6A). These sorted populations (along with total CD4+ Tm cells as a comparison control) were then exposed to HIV-F4.HSA for 3 days and then assessed by FACS for infection rates. Infection rates directly correlated with the expression levels of sialic acid, with the $WGA^{Low}$ cells being the least susceptible and the $WGA^{High}$ the most (Figure 6B, C). These results provide experimental confirmation that in PBMCs, differentially susceptible CD4+ Tm cells can be isolated based solely on sialic acid expression levels.

To better understand the mechanism behind the differential susceptibility of cells expressing high vs. low levels of sialic acid, we returned to our CyTOF datasets and manually gated on CD4+ Tm cells expressing high vs. low levels of total sialic acid as assessed by WGA binding, to assess what was differentially expressed among these two populations. The $WGA^{High}$ cells preferentially expressed fucose (as assessed by AOL binding) (Figure 6D), consistent with earlier observations of co-expression of fucose and sialic acid on HIV-susceptible CD4+ T cells. Relative to their $WGA^{Low}$ counterparts, the $WGA^{High}$ cells also preferentially expressed higher levels of CD4 and the HIV co-receptor CCR5 (Figure 6E), potentially explaining the increased susceptibility of these cells to infection. As activated T cells are known to be preferentially susceptible to infection (Stevenson et al., 1990), we also compared expression levels of activation markers on the $WGA^{Low}$ and $WGA^{High}$ cells, and found that six markers of T cell activation (HLADR, CD69, CD38, CD25, CD28, and ICOS) were all elevated on the latter population (Figure 6F). To further validate the notion that the high susceptibility of $WGA^{High}$ CD4+ Tm cells is closely associated with the activation status of these cells, we phenotyped CD4+ Tm cells from resting vs. PHA-stimulated PBMCs. As expected, the stimulated CD4+ Tm cells expressed higher levels of multiple activation markers (Figure S8A). Importantly, the stimulated CD4+ Tm cells also bound higher levels of WGA, consistent with upregulation of

315 sialic acid upon T cell activation (Figure S8B). Furthermore, activated cells, as defined as those
316 expressing high levels HLADR, CD69, CD38, CD25, CD28, or ICOS, all expressed higher levels
317 of sialic acid as compared to cells with low levels of these activation markers (Figure S8C).
318 These data together strongly support the notion that high sialic expression identifies the most
319 activated subsets of CD4+ Tm cells.

320 We also considered the possibility that $WGA^{High}$ cells may support higher levels of
321 productive infection because these cells better survive the cytopathic effects of HIV replication.
322 Consistent with this hypothesis, we found that $WGA^{High}$ cells expressed higher levels of CD127,
323 a marker of long-lived self-renewing cells, and Ox40 and BIRC5, which are involved in
324 protecting HIV-infected cells from apoptosis (Kuo et al., 2018) (Figure 6G). In comparison,
325 markers preferentially expressed on the $WGA^{Low}$ cells were those of central memory T cells,
326 including CD27, CCR7, and CD62L (Figure 6H). Additional markers in our CyTOF panel
327 differentially expressed between the $WGA^{Low}$ and $WGA^{High}$ cells included other glycans,
328 transcription factors, homing receptors, and exhaustion markers (Figure S9).

329 Although these results suggest that high levels of sialic acid may simply be a biomarker
330 of the most HIV-susceptible CD4+ Tm cells, it did not rule out the possibility that sialic acid plays
331 a direct role in HIV susceptibility. To test this, we tried two approaches to diminish cell-surface
332 sialic acid expression: transient treatment with either sialic acid synthease inhibitor P-$3F_{AX}$-
333 Neu5Ac or sialidase. Inhibitor treatment did not decrease cell-surface sialic acid levels (Fig.
334 S8D), but sialidase did (Figure 6I). We therefore exposed mock- or sialidase-treated cells to
335 HIV-F4.HSA for 3 days and then assessed infection rates by FACS. We found that infection
336 rates were decreased in a sialidase dose-dependent manner (Figure 6J, K). These results
337 suggest that a single glycan – sialic acid – may not only be capable of distinguishing memory
338 CD4+ T cells with vastly different phenotypic features and HIV susceptibility but may also play a
339 direct role in promoting infection by the virus.
340

## DISCUSSION

We describe here the development and implementation of CyTOF-Lec, a high-parameter single-cell method to simultaneously quantitate multiple glycans and proteins on the surface of human cells. We used CyTOF-Lec to identify glycomic and phenotypic differences between immune subsets from blood and multiple tissues, and between uninfected and HIV-infected CD4+ T cells. Moreover, by performing PP-SLIDE bioinformatics on our datasets of HIV-infected blood and tissue cells, we identified unique glycan features characteristic of HIV-susceptible CD4+ T cells, and identified glycan structures that were remodeled as a result of cell-intrinsic HIV replication.

Multiple studies have demonstrated remodeling of host cells upon HIV infection. HIV infection of cell lines (Matheson et al., 2015) as well as primary cells (Cavrois et al., 2017; Ma et al., 2020; Xie et al., 2021) leads to up- and down-regulation of a variety of host proteins, reflecting the ability of the virus to hijack host processes to complete its replication cycle. We demonstrate in the current study that HIV also remodels the surface glycome of infected cells. By implementing PP-SLIDE to discern antigens differentially expressed as a result of remodeling from those differentially expressed as a reflection of HIV selection, we found that both fucose and sialic acid were upregulated on CD4+ T cells after infection. As core fucosylation on the cell surface is critical for CD4+ T cell activation (Fujii et al., 2016; Liang et al., 2018) and activated cells are more permissive to HIV infection (Stevenson et al., 1990), HIV may upregulate fucosylation to maintain an elevated state of activation facilitating the completion of its replication. We cannot rule out, however, that upregulation of fucose is a byproduct of T cell activation, and not a direct consequence of HIV infection. Of note, fucosylation is also important for lymphocyte trafficking (Colomb et al., 2020b), and the upregulation of fucose on infected cells, even if not directly promoting viral replication, may

366 facilitate the circulation of infected cells from blood into tissues where ample populations of HIV-
367 permissive CD4+ T cells reside.

368 Sialic acid, like fucose, was also upregulated as a result of HIV infection. In general,
369 sialylated glycans on the surface of cells can elicit an immunosuppressive response by effector
370 cells of the immune system. For example, binding of sialic acid on the surface of cancer cells to
371 Siglec-7 and Siglec-9 proteins on the surface of NK cells can diminish NKG2D-mediated
372 activation of the NK cells (Xiao et al., 2016). Moreover, NKG2D may play an important role in
373 NK cell-mediated killing of HIV-infected cells and NKG2D ligands are actively downregulated by
374 HIV Nef to escape immune recognition (Jost and Altfeld, 2012). We postulate that through a
375 combination of upregulating sialic acid and down-regulating NKG2D ligands, HIV may evade
376 NKG2D-mediated killing by NK cells.

377 Interestingly, sialic acid was also preferentially expressed on the memory CD4+ T cells
378 that were preferentially targeted for infection by HIV. These results suggest that the high levels
379 of sialic acid on HIV-infected cells results from preferential infection of cellular subsets with high
380 levels of cell-surface sialic acid, followed by further upregulation of this glycan. Interestingly,
381 however, the preferential selection of $WGA^{high}$ memory CD4+ T cells for HIV infection was only
382 observed in tonsils and PBMCs, but not in the endometrium (Fig. 4B). Because memory CD4+
383 T cells from the endometrium express higher levels of cell-surface sialic acid than those from
384 tonsils or PBMCs (Fig. 4D), most endometrial memory CD4+ T cells may already have the
385 $WGA^{high}$ HIV-permissive phenotype, resulting in minimal selection of these cells by HIV. Indeed,
386 this would be consistent with the many subsets of endometrial memory CD4+ T cells that are
387 susceptible to HIV infection, as compared to a markedly lower number of HIV-susceptible
388 memory CD4+ T cell subsets from the blood compartment (Ma et al., 2020). It is also worthwhile
389 to note that the endometrium is a unique tissue in that it serves as the site of blastocyst
390 implantation and development of a semi-allogenic fetus. The unusually high levels of sialic acid

391 on immune cells from this tissue may be important to prevent immune-mediated rejection of
392 spermatozoa or the developing fetus.

393 Importantly, HIV's preference for cells expressing high levels of sialic acid was validated
394 by demonstrating via sorting experiments that blood-derived memory CD4+ T cells expressing
395 high levels of WGA were significantly more susceptible to infection than those expressing low
396 levels of WGA. The $WGA^{high}$ population exhibits many features that could explain its high
397 susceptibility: higher expression levels of HIV receptor/co-receptor, higher activation status, and
398 higher expression of markers associated with survival of HIV-infected cells. However, our
399 somewhat surprising finding that sialidase-treated cells are more poorly susceptible to HIV
400 additionally suggest that sialic acid may exert a direct role in promoting HIV infection; the
401 precise step in the viral replication cycle affected by sialic acid should be interrogated in future
402 studies. Regardless, since thus far markers that distinguish memory CD4+ T cells with differing
403 susceptibilities to HIV have been difficult to identify, our findings that WGA may be useful as a
404 tool to isolate a highly susceptible population of CD4+ T cells could be of practical use for the
405 field. On a more fundamental level, our results suggest that host glycan expression can very
406 much influence HIV replication dynamics.

407 Of note, we cannot rule out the possibility that the high levels of fucose and sialic acid on
408 HIV-infected cells may also result in part from binding of HIV virions to the surface of the
409 infected CD4+ T cells. HIV particles contain cell-derived glycolipids, including the sialic acid-
410 containing GM3 ganglioside (Puryear et al., 2012). Furthermore, HIV gp120 is heavily N-
411 glycosylated, with a majority of high-mannose N-glycan structures and a lower proportion of
412 complex N-glycans carrying lactosamine residues and terminal sialic acid. Sialic acid on gp120
413 is recognized by multiple members of the Siglec family of lectins, and these lectins can facilitate
414 *trans*- or cell-to-cell infection of CD4+ T cells (Izquierdo-Useros et al., 2012; Varchetta et al.,
415 2013; Zou et al., 2011). However, we note that our studies characterizing infected cells defined
416 these cells as those that express HSA – which is not on the surface of input virions – and that

have downregulated cell-surface CD4, a hallmark of productive HIV infection. The fact that we found unique glycan features of these cells relative to bystander CD4+ T cells suggests that productively-infected cells have distinct cell-surface glycan features in a manner not attributable to just surface virion sticking.

Our study focused on characterizing cells productively-infected with HIV-1, identified as those expressing on their cell surface the LTR-driven HSA reporter protein. We envision, however, that CyTOF-Lec will also be useful for studying the HIV reservoir that persists despite antiretroviral therapy (ART). Recent observations suggest that a significant portion of the reservoir in ART-suppressed people living with HIV (PLWH) is transcriptionally active, expressing short and incomplete HIV transcripts in the absence of *ex vivo* stimulation (Yukl et al., 2018). Interestingly, we recently demonstrated using cells from ART-suppressed PLWH that CD4+ T cells expressing high levels of fucose contain significantly more transcriptionally-active HIV reservoir cells than those expressing low levels of fucose (Colomb et al., 2020a). With the development of CyTOF-Lec as a single-cell analysis tool, we will now be able to interrogate glycan expression patterns on transcriptionally-active, as well as transcriptionally-silent HIV reservoir cells. Specifically, we envision combining CyTOF-Lec with PP-SLIDE analysis of reservoir cells from PLWH. This will entail using bioinformatics approaches to trace *ex vivo* reactivated cells to their original pre-stimulation states by PP-SLIDE, an approach we previously described and validated experimentally (Neidleman et al., 2020b), but under conditions where we incorporate assessment of the glycan features of the reservoir cells with the CyTOF-Lec panel.

Beyond HIV, CyTOF-Lec has potentially wide-ranging applications. Glycosylated host receptors play important roles in enabling the attachment of virus. For example, influenza virus enters cells by binding of its hemagglutinin (HA) protein to terminal sialic acid moieties attached to surface proteins of host cells (Li et al., 2021). The novel beta-coronavirus SARS-CoV-2 may also exploit host cell glycosylation for infection, as its primary receptor ACE2 is heavily

443 glycosylated. Indeed, blockage of *N*- and *O*-glycan synthesis on the host cell can inhibit SARS-
444 CoV-2 entry by diminishing the ability of the viral spike protein to bind ACE2 (Yang et al., 2020).
445 Better understanding the surface glycans of virally-susceptible host cells through single-cell
446 glycomic analysis by CyTOF-Lec, in conjunction with identifying the protein backbones
447 harboring the glycans, will improve our understanding of viral pathogenesis and can potentially
448 lead to novel prophylactic or therapeutic agents for clinically-relevant viral pathogens.
449 Further development of CyTOF-Lec taking advantage of next-generation sequencing
450 approaches will also be valuable for developing the tools to better understand how glycan
451 expression associates with or directly regulates host cell processes. A recent report combined
452 glycomic analysis with single-cell RNAseq (Kearney et al., 2021) by incubating cells with a
453 biotinylated version of the lectin L-Pha followed by a DNA-barcoded anti-biotin antibody.
454 Subsequent droplet encapsulation and sequencing revealed the transcriptomes of cells with
455 different levels of L-Pha binding. We envision that by directly conjugating lectins to barcoded
456 DNA oligos, we can expand the numbers of lectins we can simultaneously monitor.
457 Furthermore, additional inclusion of DNA-barcoded antibodies (Peterson et al., 2017; Stoeckius
458 et al., 2017) will enable simultaneous surface proteome analysis. Such a technology, while not
459 as high-throughput, scalable, and cost-effective as CyTOF-Lec, will enable a genome-wide
460 analysis of cellular processes associated with differential surface glycosylation.
461
462 **METHODS**
463 **Tissue processing and cell isolation**
464 Endometrial tissues were obtained from the Women's Health Clinic of Naval Medical Center
465 Portsmouth (NMCP) in Virginia (CIP # NMCP.2016.0068) under standard operating procedures
466 (Fassbender et al., 2014). The biopsies were transported to San Francisco in MCDB-105
467 (Sigma-Aldrich M6395) containing 10% heat-inactivated fetal bovine (FBS) and 1%
468 Penicillin/Streptomycin (P/S). The collection of endometrial T cells was performed similar to

recently described methods (Ma et al., 2020). Briefly, endometrial tissue was washed once with SCM media, which consisted of 75% phenol red-free Dulbecco's Modified Eagle's Medium (DMEM, Life Technologies), 25% MCDB-105, 10% FBS, 1% L-glutamine with P/S (Gemini), 1 mM sodium pyruvate (Sigma-Aldrich), and 5 mg/ml insulin (Sigma-Aldrich). The tissues were then digested at 37°C for 2 hours under rotation in SCM media mixed at a 1:1 dilution with Digestion Media, which consisted of HBSS containing $Ca^{2+}$ and $Mg^{2+}$ supplemented with 3.4 mg/ml collagenase type 1 (Worthington Biochemical Corporation LS004196) and 100 U/ml hyaluronidase (Sigma-Aldrich H3631). Cells were then filtered through a Falcon 40-µm cell strainer. The filtrate was then centrifuged, washed once with R10 media (RPMI 1640 containing 10% FBS and 1% P/S), and the cells were cultured in 96-well U-bottom polystyrene plates at a concentration of $10^6$ cells/well in 200 µl R10 media.

**Processing of human lymphoid aggregate cultures (HLAC)**

Human tonsils obtained from the Cooperative Human Tissue Network (CHTN) were processed similarly to methods recently described (Ma et al., 2020). Briefly, tonsils were rinsed in tonsil media (RPMI supplemented with 15% FBS, 100 µg/ml gentamicin, 200 µg/ml ampicillin, 1 mM sodium pyruvate, 1% non-essential amino acids (Mediatech), 1% Glutamax (Thermo Fisher), and 1% Fungizone (Invitrogen)), dissected into small pieces, and then pressed through a 40-µm cell strainer using a syringe plunger. The cells were then filtered through a second 40-µm cell strainer, centrifuged, and resuspended with 200 µl tonsil media per $10^6$ cells. For sialidase treatment, $10^6$ tonsil cells were resuspended in 1 ml of tonsil media, and then treated with 20 µg of the sialidase for 1 hour at 37°C. The cells were washed once with PBS and processed for CyTOF analysis as described further below.

**PBMC isolation**

494 PBMCs were isolated from reduction chambers obtained from Vitalant Research Institute and
495 Stanford Blood Bank using Ficoll-Hypaque density gradients, and then cultured in R10. For
496 sorting experiments, CD4+ T cells were purified by negative selection using the EasySep™ CD4
497 enrichment kit (Stem cell Technologies), and further enriched for memory cells by depletion of
498 naïve T cells using CD45RA beads (Miltenyi Biotec), prior to lectin staining and sorting as
499 described further below. Where indicated, PBMCs were first stimulated for 3 days with 5 µg/ml
500 PHA in the presence of 10 IU/ml IL-2 prior to CyTOF-Lec analysis.
501
502 **Virus preparation and infection assays**
503 Viral stocks of the previously described HIV-1 reporter virus F4.HSA (Cavrois et al., 2017) were
504 prepared similarly to recently described methods (Ma et al., 2020). Briefly, 293T cells were
505 seeded in T175 flasks and transfected using polyethylenimine (PEI, Polysciences) with F4.HSA
506 proviral DNA (70 µg / flask) (Longo et al., 2013). Two days after transfection, supernatants from
507 293T cultures were harvested, filtered through a 0.22-µm filter and concentrated by
508 ultracentrifugation at 20,000 rpm (Beckman Coulter Optima XE-90) for 2 hours at 4°C. $p24^{Gag}$
509 concentrations were quantitated using the Lenti-X $p24^{Gag}$ Rapid Titer Kit (Clontech). For
510 infection, 10-20 ng/ml $p24^{Gag}$ of F4.HSA was incubated with $10^6$ cells in 200 µl R10 media in 96-
511 well U-bottom polystyrene plates. After 2 hours, cells were fed with fresh R10 media and
512 cultured for another 3 days. Where indicated, PBMCs were first treated with the sialic acid
513 inhibitor P-3$F_{AX}$-Neu5Ac (Tocris) for 24 hours, or sialidase for 1 hour, prior to infection with
514 F4.HSA.
515
516 **Flow cytometry**
517 For sorting experiments, purified memory CD4+ cells (see above) were washed once with FACS
518 buffer in 96-well V-bottom polystyrene plates, and then stained at room temperature for 15 min
519 with a 1:200 dilution of the LIVE/DEAD Zombie Aqua™ Fixable Viability Kit (Biolegend) reagent

to exclude dead cells. The cells were washed and stained for 30 min at 4°C with 5 µg/ml FITC-WGA (Vector Laboratories). After the wash, cells were resuspended at a concentration of 10 million cells/ml, and sorted on an Aria II flow cytometer (BD Biosciences) into $WGA^{low}$, $WGA^{medium}$, and $WGA^{high}$ populations. Total CD4+ Tm cells were also sorted as a control. The purity of the sorted cells was confirmed by analysis on Aria II immediately after sorting. All sorted populations were infected with F4.HSA for 3 days. For FACS analysis of the samples, 0.1-1 million cells of each sample were transferred into 96-well V-bottom polystyrene plates, washed once with FACS buffer (PBS containing 2% FBS and 2 mM EDTA), and stained for 30 min on ice with an antibody cocktail consisting of APC/Cy7-CD3 (SK7, Biolegend), PE/Cy7-CD4 (A161A1, Biolegend), APC-CD8 (SK1, Biolegend), FITC-CD24 (HSA, M1/69, BD Biosciences), and the LIVE/DEAD Zombie Aqua™ Fixable Viability Kit (Biolegend). The cells were then washed twice, fixed with 1% PFA (Electron Microscopy Sciences) in PBS, and analyzed by flow cytometry on an LSRFortessa™ (BD Biosciences).

**CyTOF data generation**

A 39-parameter CyTOF panel was designed for this study, which included antibodies against markers of T cell differentiation states, activation markers, transcription factors, and homing receptors, and an antibody against HSA to identify productively-infected cells. The panel also included numerous lectins enabling the characterization of glycan features (Table S2). X8 antibody-labeling kits (Fluidigm) were used to label antibodies that required in-house conjugation. The conjugated antibodies were quantitated for protein content by Nanodrop (Thermofisher). Prior to storage at 4ºC, specimens were diluted 1:1 using a PBS-based Antibody Stabilizer (Boca Scientific) supplemented with 0.05% sodium azide.

Preparation of samples for CyTOF staining was conducted as previously described (Ma et al., 2020; Neidleman et al., 2020a; Neidleman et al., 2020b), Briefly, 1-6 million cells were washed once with CyFACS (metal contaminant-free PBS (Rockland) supplemented with 0.1%

546 bovine serum albumin and 0.1% sodium azide). Where indicated, cells were first treated with 20
547 µg/ml of sialidase or PBS as control and incubated for 1 hour at 37°C. Sialidase was prepared
548 in-house using the *Vibrio cholerae nanH* gene cloned into the pCVD364 vector, which was
549 provided by Dr. Eric R. Vimr from the University of Illinois Urbana (Taylor et al., 1992). After
550 centrifugation, the cells were resuspended with contaminant-free PBS (Rockland) supplemented
551 with 2 mM EDTA (PBS/EDTA), and then treated with 25 µM cisplatin (Sigma-Aldrich) in 4 ml
552 PBS/EDTA for 60 seconds at room temperature. The samples were then immediately quenched
553 with CyFACS, centrifuged, resuspended in 2% PFA in CyFACS, and incubated for 10 min at
554 room temperature. The cells were then washed three times with CyFACS, resuspended in 100
555 µl of CyFACS containing 10% DMSO, and frozen at -80°C until CyTOF staining.

556 To stain multiple specimens in the same reaction, cells were barcoded using the Cell-ID
557 20-Plex Pd Barcoding Kit according to manufacturer's instructions (Fluidigm). Briefly, 1-3 million
558 cisplatin-treated cells were thawed and transferred into Nunc 96 DeepWell polystyrene plates
559 (Thermo Fisher). After two washes with Barcode Perm buffer (Fluidigm), the cells were
560 incubated with selected barcodes for 30 min. Cells were then washed with 0.8 ml Maxpar® Cell
561 Staining buffer (Fluidigm) followed by 0.8 ml CyFACS. Barcoded samples were combined and
562 pelleted, and then blocked on ice for 15 min with sera from mouse (Thermo Fisher), rat (Thermo
563 Fisher), and human (AB serum, Sigma-Aldrich). Cells were then washed twice with CyFACS,
564 and stained on ice for 45 min with a cocktail of CyTOF surface-staining antibodies (Table S1) in
565 a final volume of 100 µl / well. Cells were then washed three times with CyFACS buffer, and
566 stained on ice for 45 min with the cocktail of lanthanide-conjugated lectins (Table S1) in a final
567 volume of 100 µl/well. Cells were then washed 3 times with CyFACS buffer and fixed overnight
568 at 4°C with 2% PFA in metal contaminant-free PBS. The next day, cells were incubated at 4°C
569 for 30 min with fix/perm buffer (eBioscience), and then washed twice with Permeabilization
570 Buffer (eBioscience). After another round of Fc blocking on ice for 15 min with sera from mouse

(Thermo Fisher) and rat (Thermo Fisher), cells were washed twice with Permeabilization Buffer (eBioscience), and stained on ice for 45 min with a cocktail of CyTOF intracellular-staining antibodies (Table S1) in a final volume of 100 µl/well. Cells were then washed with CyFACS and incubated for 20 min at room temperature with 250 nM Cell-ID™ DNA Intercalator-Ir (Fluidigm) in PBS containing 2% PFA. After two more washes with CyFACS, cells were washed once with Maxpar® Cell Staining Buffer (Fluidigm), once with Maxpar® PBS (Fluidigm), and once with Maxpar® Cell Acquisition Solution (Fluidigm). Immediately prior to sample loading, cells were resuspended to a concentration of 7 x$10^5$/ml in EQ™ calibration beads (Fluidigm) diluted 1:9 in Maxpar® Cell Acquisition Solution. Cells were acquired on a Helios-upgraded CyTOF2 instrument (Fluidigm) at a rate of 250-350 events/sec, at the UCSF Flow Core Facility.

**CyTOF data analysis**

Data were normalized to EQ™ calibration beads and then exported as FCS files. The data were then de-barcoded with CyTOF software (Fluidigm) and imported into FlowJo (BD) for gating. This study's raw datasets, pre-gated on live, singlet events, are available for download via the following link in the Dryad public repository:
https://doi.org/10.7272/Q6FT8J92
Total T cells were identified by sequential gating on intact, live, singlet CD3+CD19- cells (Figure S2). Total T cells were then re-exported as FCS files and imported into Cytobank for calculations of median signal intensity (MSI), and high-dimensional analyses by t-SNE and FlowSOM. t-SNE and FlowSOM plots were generated with default settings except for a modification of total metaclusters from 10 to 20 for FlowSOM analysis. t-SNE and FlowSOM plots were generated excluding all parameters those used upstream in the gating strategy (CD19 and HSA) and all glycan characterization parameters. To map defined populations onto

t-SNE plots, subsets were defined by manual gating, and then pseudo-colored on the t-SNE plots using FlowJo software. Box plot graphs were generated using ggplot2 in R.

Identification of predicted precursor (PRE) cells by PP-SLIDE was implemented using recently described methods (Ma et al., 2020; Neidleman et al., 2020b) to match each infected cell against every CD4+ T cell in the uninfected sample and using k-nearest neighbor (kNN) calculations to identify the phenotypically most similar. The degree of enrichment of each FlowSOM cluster in PRE cells was calculated by dividing each cluster's relative size within the PRE cells by its relative size within total uninfected CD4+ Tm cells:

*Enrichment ratio (Cluster X) = Number of Cluster X cells relative to PRE cells/Number of Cluster X cells relative to uninfected Tm cells*

Clusters with ratios > 1 were designated as enriched and those with ratios > 0 and < 1 as non-enriched, while clusters with undetectable PRE cells were not shown.

**Statistical Analysis**

Expression levels were reported as median signal intensity (MSI) for each parameter (protein or glycan) within each cell population analyzed. Student's two-sided paired t-tests were used to test for differences in MSI among phenotypic subsets (B cell, CD8+ Tm, CD8+ Tn, CD4+ Tm, and CD4+ Tn cells); among uninfected and bystander cells; or among uninfected cells, PRE cells, and infected cells. P-values were adjusted for multiple testing using False Discovery Rate (FDR) via the Benjamini-Hochberg or Holm method as indicated in figure legends. FDR adjusted p-values that were less than 0.05 were considered as significant.

SLIDE analysis was conducted using the R package SLIDE (Mukherjee et al., 2018) as recently described (Ma et al., 2020). SLIDE was developed as a nearest-neighbor approach to identify and quantify viral-induced remodeling (Sen et al., 2014). The ratios between two distance measures in SLIDE (the remodeling score) provides a relative measure of remodeling,

620 and is compared to a background remodeling score generated from SLIDE analysis of non-

621 infected cells, as recently described (Mukherjee et al., 2018).

## Figure Legends

**Figure 1. Validation of CyTOF-Lec. (A)** Antibody staining for protein markers is not altered by lectins. Shown are histograms of tonsil cells expressing CD3, CD8, or CD4, as detected by CyTOF after antibody staining followed or not by staining with metal-conjugated lectins (AOL: Aspergillus Oryzae; MAL-1: Maackia Amurensis I; WGA: Wheat Germ Agglutinin; UEA-1: Ulex Europaeus I; and ABA: Agaricus Bisporus Agglutinin). Protein expression (y axis) is represented as the percentage of the maximal expression level detected for each staining. **(B-D)** Sialidase treatment elicits expected changes in lectin binding. Tonsil cells were treated with sialidase (20 μg/ml) for 1h at 37°C, and then stained with the CyTOF-Lec panel. Shown are histograms depicting the extent of interaction with sialic acid-binding **(B)**, fucose-binding **(C)** or T antigen-binding **(D)** lectins. Removal of sialic acid by sialidase decreases binding by sialic acid-binding lectins, while increasing binding by the fucose- and T antigen-binding lectins, as expected.

**Figure 2. Glycan expression in lymphocytes from human endometrium, tonsils, and blood. (A)** Box plots showing staining by fucose-binding lectins on B and T cells from the endometrium, tonsils, and PBMCs, quantified as median signal intensity (MSI). T cells were subdivided into memory CD8+ T cells (CD8+Tm), naïve CD8+ T cells (CD8+Tn), memory CD4+ T cells (CD4+Tm), and naïve CD4+ T cells (CD4+Tn) based on their expression of Tm- and Tn-specific CyTOF markers. AOL binds to total/core fucose, UEA1 binds to α1-2 branched fucose. Although there were some differences in binding between sites and between the different lectins, in all instances fucose-binding proteins bound CD4+Tm at higher levels than they did CD4+Tn. **(B)** Box plots showing binding by sialic-acid binding lectins WGA and MAL-1. Results are presented as in *panel A*. WGA binds to total sialylated glycans and MAL-1 binds to α2-3 sialylated glycans. Overall, the sialic acid-binding lectins bound CD8+ T cells at higher levels

than they did CD4+ T cells and B cells. **(C)** Box plots showing binding by the T antigen-binding lectin ABA. Results are presented as in *panel A*. Overall, ABA bound T cells at higher levels than they did B cells. *p<0.05, **p<0.01, ***p<0.001 as assessed using the Student's paired t test and adjusted for multiple testing using the Holm.

**Figure 3. HIV alters expression of fucose and sialic acid in a tissue site-dependent manner. (A)** HIV preferentially infects fucose-expressing cells and further upregulates fucose expression in a tissue site-dependent manner. Box plots showing binding by fucose-binding proteins on uninfected (UI), predicted precursor (PRE), and infected (INF) CD4+ T cells from the endometrium, tonsils, and PBMCs. All populations were pre-gated on live, singlet CD4+ Tm cells. AOL binds total/core fucose, while UEA1 binds $\alpha$1-2 branched fucose. **(B)** HIV preferentially infects sialic acid-expressing cells and further upregulates sialic acid in a tissue site-dependent manner. Box plots showing binding by sialic acid-binding lectins. Results are presented as in *panel A*. WGA binds total sialylated glycans and MAL1 binds recognizes $\alpha$2-3 sialylated glycans. **(C)** Box plots showing binding by T antigen-binding lectin ABA. Results are presented as in *panel A*. *p < 0.05, **p < 0.01 as assessed using the Student's paired t test and adjusted for multiple testing using the Benjamini-Hochberg for FDR.

**Figure 4. HIV preferentially infects memory CD4+ T cells from tonsils and PBMCs with high levels of fucose and sialic acid. (A)** The proportions of CD4+ Tm cells that were productively-infected (as assessed by HSA positivity) are higher among the $AOL^{High}$ and $WGA^{High}$ cells than among their $AOL^{Low}$ and $WGA^{Low}$ counterparts for all 3 sites. *p < 0.05, **p < 0.01, and ***p < 0.001 as assessed using the Student's paired t test. Each color corresponds to a different donor. **(B)** Proportion of uninfected CD4+ Tm and PRE cells expressing high levels of fucose, sialic acid, or both (as determined by high binding by AOL or WGA, respectively), as

672 assessed by manual gating. In tonsils and PBMCs, cells expressing fucose and sialic acid were
673 preferentially selected for infection by HIV. *p < 0.05, **p < 0.01 as assessed using the
674 Student's paired t test. **(C)** Co-expression of fucose and sialic acid on PRE cells in the indicated
675 specimens, as depicted by t-SNE heatmaps. Shown are cells concatenated from all donors
676 analyzed in the study. Regions of the tSNE co-expressing fucose and sialic acid are circled. **(D)**
677 Levels of fucose and sialic acid differ between CD4+ Tm cells from different origins, as shown
678 by MSI for binding by WGA (sialic acid-binding) and AOL (fucose-binding). *p< 0.05, **p < 0.01,
679 ***p < 0.001, ****p < 0.0001 as assessed using a one-way ANOVA and adjusted for multiple
680 testing using the Bonferroni.
681
682 **Figure 5. FlowSOM clustering confirms that HIV-susceptible subsets from tonsils and**
683 **PBMCs harbor high levels of fucose and sialic acid. (A)** tSNE plots based on FlowSOM
684 analysis of uninfected CD4+ Tm and PRE cells from endometrium, tonsil, and PBMC
685 specimens, showing 20 color-labelled clusters of cells. **(B)** Enrichment of clusters among PRE
686 cells. PRE enrichment-folds were determined by dividing the sizes of each cluster in PRE cells
687 by that in the corresponding uninfected CD4+ Tm cells. Enriched clusters (those with an
688 enrichment fold above 1) correspond to cells preferentially selected for infection. Note that the
689 highest enrichment-folds were observed in tonsils, suggesting the most preferential selection of
690 subsets for infection in this specimen type. Each color corresponds to a different donor. Labels
691 on the x-axis refer to the cluster name. **(C)** Clusters enriched among PRE cells express high
692 levels of fucose and sialic acid, as depicted by t-SNE. For each specimen set, the left-hand t-
693 SNE plot depicts clusters enriched among PRE (*red*) against total cells (*grey*), while the t-SNE
694 plots on the right depict by heatmaps the expression levels of fucose (as assessed by AOL
695 binding) and sialic acid (as assessed by WGA binding) among the enriched clusters. Note that
696 the enriched clusters from all three sites include cells expressing high levels of both fucose and
697 sialic acid (highlighted by arrows).

**Figure 6. High levels of sialylated glycans identifies highly-susceptible and activated CD4+ Tm cells, and plays a direct role in susceptibility. (A)** Histograms showing the expression of total sialylated glycans on three populations ($WGA^{Low}$ [red], $WGA^{Medium}$ [yellow], and $WGA^{High}$ [blue]) of sorted uninfected CD4+ Tm cells (CD3+CD4+CD45RA-), as assessed by WGA binding. One of three representative donors is shown. **(B-C)** The sorted uninfected cells in *panel A*, along with total CD4+ Tm cells, were exposed to F4.HSA and assessed by flow cytometry for infection rates 3 days later. Results are gated on live, singlet CD3+CD8- cells. Shown are representative FACS plots from one donor (*B*) and compiled results from 3 donors (*C*). For each donor, experimental duplicates were performed for each condition. Each datapoint shown corresponds to one donor. *$p < 0.05$ as assessed using a one-way ANOVA and adjusted for multiple testing using the Bonferroni. **(D)** $WGA^{High}$ Tm cells bind more AOL than $WGA^{Low}$ Tm cells do. Shown are the histogram plots from one representative PBMC donor (*left*) and box plots from all 4 PBMC donors (*right*). **(E)** $WGA^{High}$ Tm cells express more CD4 and CCR5 than $WGA^{Low}$ Tm cells do. Shown are the histogram plots from one representative PBMC donor (*left*) and box plots from all 4 PBMC donors (*right*). **(F)** $WGA^{High}$ Tm cells express higher levels of activation markers than $WGA^{Low}$ Tm cells do. Shown are the histogram plots from one representative PBMC donor (*top*) and box plots from all 4 PBMC donors (*bottom*). **(G)** $WGA^{High}$ Tm cells express higher levels of the pro-survival factors CD127, BIRC5, and Ox40 than $WGA^{Low}$ Tm cells do. Shown are the histogram plots from one representative PBMC donor (*left*) and box plots from all 4 PBMC donors (*right*). **(H)** The CD127, CCR7 and CD62L receptors are expressed at lower levels in $WGA^{High}$ relative to $WGA^{Low}$ Tm cells. Shown are the histogram plots from one representative PBMC donor (*left*) and box plots from all 4 PBMC donors (*right*). For panels D-G, *$p < 0.05$ as assessed using the Student's paired t test and adjusted for multiple testing using the Benjamini-Hochberg for FDR. **(I)** Transient treatment with sialidase decreases cell-surface levels of sialidase on CD4+ T cells. PBMCs were treated for 1 h with

724 sialidase prior to assessment of cell-surface WGA binding. Shown are overlaid histograms
725 demonstrating a decrease in cell-surface sialic acid levels (as reflected by WGA binding) in the
726 sialidase-treated cells. Results are gated on live, singlet CD3+CD8-CD4+ cells. Numbers
727 correspond to percent of cells within the indicated gate. **(J-K)** PBMCs treated for 1 h with the
728 indicated concentrations of sialidase were exposed to F4.HSA and assessed by flow cytometry
729 for infection rates 3 days later. Results are gated on live, singlet CD3+CD8- cells. Shown are
730 representative FACS plots from two donors (*J*) and the results of experimental triplicates from
731 each of these donors (*K*). **p < 0.01 and **** p < 0.0001 as assessed using a one-way ANOVA
732 and adjusted for multiple testing using the Bonferroni.

## ACKNOWLEDGEMENTS

This work was supported by the National Institutes of Health (R01AI127219, R01AI147777, and P01AI131374 to N.R.R., and R01DK123733, R01AG062383, R01NS117458, and R21AI143385 to M.A-M. We also acknowledge NIH for the sorter (S10-RR028962), support from CFAR (P30AI027763), and the James B. Pendleton Charitable Trust. We acknowledge the PFCC (RRID:SCR_018206) for assistance in CyTOF data acquisition, enabled by an instrument that was supported in part by the DRC Center Grant NIH P30 DK063720 and NIH S10 1S10OD018040-01. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We thank Claudia Bispo and Stanley Tamaki for CyTOF assistance at the Parnassus Flow Core, Jane Srivastava and Nandhini Raman for flow cytometry assistance at the Gladstone Flow Core. We also thank Françoise Chanut for editorial assistance, and Robin Givens for administrative assistance.

## AUTHOR CONTRIBUTIONS

T.M. designed and performed experiments, conducted data analyses, and participated in manuscript writing; M.M. wrote scripts for data analysis; L.B.G. and G.X. provided guidance and established protocols for FACS and CyTOF data analysis; A.G. processed endometrial specimens; M.A-M. conceived ideas for the study, performed supervision, conducted data analyses, and participated in manuscript writing; N.R.R. conceived ideas for the study, performed supervision, conducted data analyses, and wrote the manuscript. All authors read and approved the manuscript.

## COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

## REFERENCES

Barrera, C., Espejo, R., and Reyes, V.E. (2002). Differential glycosylation of MHC class II molecules on gastric epithelial cells: implications in local immune responses. Human immunology *63*, 384-393.

Bendall, S.C., Simonds, E.F., Qiu, P., Amir el, A.D., Krutzik, P.O., Finck, R., Bruggner, R.V., Melamed, R., Trejo, A., Ornatsky, O.I.*, et al.* (2011). Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. Science *332*, 687-696.

Breen, E.C., Rezai, A.R., Nakajima, K., Beall, G.N., Mitsuyasu, R.T., Hirano, T., Kishimoto, T., and Martinezmaza, O. (1990). Infection with Hiv Is Associated with Elevated Il-6 Levels and Production. J Immunol *144*, 480-484.

Cavrois, M., Banerjee, T., Mukherjee, G., Raman, N., Hussien, R., Rodriguez, B.A., Vasquez, J., Spitzer, M.H., Lazarus, N.H., Jones, J.J.*, et al.* (2017). Mass Cytometric Analysis of HIV Entry, Replication, and Remodeling in Tissue CD4+ T Cells. Cell Rep *20*, 984-998.

Chen, S., Qin, R., and Mahal, L.K. (2021). Sweet systems: technologies for glycomic analysis and their integration into systems biology. Crit Rev Biochem Mol Biol, 1-20.

Colomb, F., Giron, L.B., Kuri-Cervantes, L., Adeniji, O.S., Ma, T., Dweep, H., Battivelli, E., Verdin, E., Palmer, C.S., Tateno, H.*, et al.* (2020a). Sialyl-Lewis(X) Glycoantigen Is Enriched on Cells with Persistent HIV Transcription during Therapy. Cell Rep *32*, 107991.

Colomb, F., Giron, L.B., Kuri-Cervantes, L., Adeniji, O.S., Ma, T.C., Dweep, H., Battivelli, E., Verdin, E., Palmer, C.S., Tateno, H.*, et al.* (2020b). Sialyl-Lewis(X) Glycoantigen Is Enriched on Cells with Persistent HIV Transcription during Therapy. Cell Reports *32*.

Colomb, F., Giron, L.B., Premeaux, T.A., Mitchell, B.I., Niki, T., Papasavvas, E., Montaner, L.J., Ndhlovu, L.C., and Abdel-Mohsen, M. (2019). Galectin-9 Mediates HIV Transcription by Inducing TCR-Dependent ERK Signaling. Front Immunol *10*, 267.

781 Contreras, X., Bennasser, Y., Chazal, N., and Bahraoui, E. (2003). [HIV-1 Tat induces TNF-
782 alpha production by human monocytes: involvement of calcium and PKC pathways]. J Soc Biol
783 *197*, 267-275.
784 de Freitas Junior, J.C., Silva Bdu, R., de Souza, W.F., de Araujo, W.M., Abdelhay, E.S., and
785 Morgado-Diaz, J.A. (2011). Inhibition of N-linked glycosylation by tunicamycin induces E-
786 cadherin-mediated cell-cell adhesion and inhibits cell proliferation in undifferentiated human
787 colon cancer cells. Cancer chemotherapy and pharmacology *68*, 227-238.
788 Dewald, J.H., Colomb, F., Bobowski-Gerard, M., Groux-Degroote, S., and Delannoy, P. (2016).
789 Role of Cytokine-Induced Glycosylation Changes in Regulating Cell Interactions and Cell
790 Signaling in Inflammatory Diseases and Cancer. Cells *5*.
791 Doms, R.W., and Trono, D. (2000). The plasma membrane as a combat zone in the HIV
792 battlefield. Gene Dev *14*, 2677-2688.
793 Everest-Dass, A.V., Jin, D., Thaysen-Andersen, M., Nevalainen, H., Kolarich, D., and Packer,
794 N.H. (2012). Comparative structural analysis of the glycosylation of salivary and buccal cell
795 proteins: innate protection against infection by Candida albicans. Glycobiology *22*, 1465-1479.
796 Fassbender, A., Rahmioglu, N., Vitonis, A.F., Vigano, P., Giudice, L.C., D'Hooghe, T.M.,
797 Hummelshoj, L., Adamson, G.D., Becker, C.M., Missmer, S.A.*, et al.* (2014). World
798 Endometriosis Research Foundation Endometriosis Phenome and Biobanking Harmonisation
799 Project: IV. Tissue collection, processing, and storage in endometriosis research. Fertil Steril
800 *102*, 1244-1253.
801 Fujii, H., Shinzaki, S., Iijima, H., Wakamatsu, K., Iwamoto, C., Sobajima, T., Kuwahara, R.,
802 Hiyama, S., Hayashi, Y., Takamatsu, S.*, et al.* (2016). Core Fucosylation on T Cells, Required
803 for Activation of T-Cell Receptor Signaling and Induction of Colitis in Mice, Is Increased in
804 Patients With Inflammatory Bowel Disease. Gastroenterology *150*, 1620-1632.
805 Garcia, J.V., and Miller, A.D. (1991). Serine phosphorylation-independent downregulation of
806 cell-surface CD4 by nef. Nature *350*, 508-511.

807 Giron, L.B., Colomb, F., Papasavvas, E., Azzoni, L., Yin, X., Fair, M., Anzurez, A., Damra, M.,
808 Mounzer, K., Kostman, J.R.*, et al.* (2020a). Interferon-alpha alters host glycosylation machinery
809 during treated HIV infection. EBioMedicine *59*, 102945.
810 Giron, L.B., Tanes, C.E., Schleimann, M.H., Engen, P.A., Mattei, L.M., Anzurez, A., Damra, M.,
811 Zhang, H., Bittinger, K., Bushman, F.*, et al.* (2020b). Sialylation and fucosylation modulate
812 inflammasome-activating eIF2 Signaling and microbial translocation during HIV infection.
813 Mucosal Immunol *13*, 753-766.
814 Giron, L.B., Tanes, C.E., Schleimann, M.H., Engen, P.A., Mattei, L.M., Anzurez, A., Damra, M.,
815 Zhang, H., Bittinger, K., Bushman, F.*, et al.* (2020c). Sialylation and fucosylation modulate
816 inflammasome-activating eIF2 Signaling and microbial translocation during HIV infection.
817 Mucosal Immunol.
818 Hiraiwa, N., Yabuta, T., Yoritomi, K., Hiraiwa, M., Tanaka, Y., Suzuki, T., Yoshida, M., and
819 Kannagi, R. (2003). Transactivation of the fucosyltransferase VII gene by human T-cell
820 leukemia virus type 1 Tax through a variant cAMP-responsive element. Blood *101*, 3615-3621.
821 Hsiao, F., Frouard, J., Gramatica, A., Xie, G., Telwatte, S., Lee, G.Q., Roychoudhury, P.,
822 Schwarzer, R., Luo, X., Yukl, S.A.*, et al.* (2020). Tissue memory CD4+ T cells expressing IL-7
823 receptor-alpha (CD127) preferentially support latent HIV-1 infection. PLoS Pathog *16*,
824 e1008450.
825 Izquierdo-Useros, N., Lorizate, M., Puertas, M.C., Rodriguez-Plata, M.T., Zangger, N., Erikson,
826 E., Pino, M., Erkizia, I., Glass, B., Clotet, B.*, et al.* (2012). Siglec-1 Is a Novel Dendritic Cell
827 Receptor That Mediates HIV-1 Trans-Infection Through Recognition of Viral Membrane
828 Gangliosides. Plos Biol *10*.
829 Jost, S., and Altfeld, M. (2012). Evasion from NK cell-mediated immune responses by HIV-1.
830 Microbes Infect *14*, 904-915.
831 Kambara, C., Nakamura, T., Furuya, T., Nishiura, Y., Kawakami, A., Ichinose, K., Shirabe, S.,
832 and Eguchi, K. (2002). Increased sialyl Lewis(x) antigen-positive cells mediated by HTLV-1

infection in peripheral blood CD4(+) T lymphocytes in patients with HTLV-1-associated myelopathy. J Neuroimmunol *125*, 179-184.

Kearney, C.J., Vervoort, S.J., Ramsbottom, K.M., Todorovski, I., Lelliott, E.J., Zethoven, M., Pijpers, L., Martin, B.P., Semple, T., Martelotto, L.*, et al.* (2021). SUGAR-seq enables simultaneous detection of glycans, epitopes, and the transcriptome in single cells. Sci Adv *7*.

Kuo, H.H., Ahmad, R., Lee, G.Q., Gao, C., Chen, H.R., Ouyang, Z., Szucs, M.J., Kim, D., Tsibris, A., Chun, T.W.*, et al.* (2018). Anti-apoptotic Protein BIRC5 Maintains Survival of HIV-1-Infected CD4(+) T Cells. Immunity *48*, 1183-1194 e1185.

Lama, J. (2003). The physiological relevance of CD4 receptor down-modulation during HIV infection. Curr HIV Res *1*, 167-184.

Leipold, M.D., Herrera, I., Ornatsky, O., Baranov, V., and Nitz, M. (2009). ICP-MS-based multiplex profiling of glycoproteins using lectins conjugated to lanthanide-chelating polymers. J Proteome Res *8*, 443-449.

Li, Y.Q., Liu, D.Q., Wang, Y.T., Su, W.Q., Liu, G., and Dong, W.J. (2021). The Importance of Glycans of Viral and Host Proteins in Enveloped Virus Infection. Frontiers in Immunology *12*.

Liang, W., Mao, S., Sun, S., Li, M., Li, Z., Yu, R., Ma, T., Gu, J., Zhang, J., Taniguchi, N.*, et al.* (2018). Core Fucosylation of the T Cell Receptor Is Required for T Cell Activation. Front Immunol *9*, 78.

Longo, P.A., Kavran, J.M., Kim, M.S., and Leahy, D.J. (2013). Transient mammalian cell transfection with polyethylenimine (PEI). Methods Enzymol *529*, 227-240.

Ma, T., Luo, X., George, A.F., Mukherjee, G., Sen, N., Spitzer, T.L., Giudice, L.C., Greene, W.C., and Roan, N.R. (2020). HIV efficiently infects T cells from the endometrium and remodels them to promote systemic viral spread. Elife *9*.

Matheson, N.J., Sumner, J., Wals, K., Rapiteanu, R., Weekes, M.P., Vigan, R., Weinelt, J., Schindler, M., Antrobus, R., Costa, A.S.*, et al.* (2015). Cell Surface Proteomic Map of HIV

858 Infection Reveals Antagonism of Amino Acid Metabolism by Vpu and Nef. Cell Host Microbe *18*,
859 409-423.
860 Mukherjee, G., Panda, A., Arvin, A.M., Sen, A., and Sen, N. (2018). SLIDE: Single Cell Linkage
861 by Distance Estimation, R package version 1.0.0, https://CRAN.R-project.org/package=SLIDE.
862 Neidleman, J., Luo, X., Frouard, J., Xie, G., Gill, G., Stein, E.S., McGregor, M., Ma, T., George,
863 A.F., Kosters, A.*, et al.* (2020a). SARS-CoV-2-Specific T Cells Exhibit Phenotypic Features of
864 Helper Function, Lack of Terminal Differentiation, and High Proliferation Potential. Cell Rep Med
865 *1*, 100081.
866 Neidleman, J., Luo, X., Frouard, J., Xie, G., Hsiao, F., Ma, T., Morcilla, V., Lee, A., Telwatte, S.,
867 Thomas, R.*, et al.* (2020b). Phenotypic analysis of the unstimulated in vivo HIV CD4 T cell
868 reservoir. Elife *9*.
869 Nystrom, K., Grahn, A., Lindh, M., Brytting, M., Mandel, U., Larson, G., and Olofsson, S. (2007).
870 Virus-induced transcriptional activation of host FUT genes associated with neo-expression of
871 Ley in cytomegalovirus-infected and sialyl-Lex in varicella-zoster virus-infected diploid human
872 cells. Glycobiology *17*, 355-366.
873 Nystrom, K., Norden, R., Muylaert, I., Elias, P., Larson, G., and Olofsson, S. (2009). Induction of
874 sialyl-Lex expression by herpes simplex virus type 1 is dependent on viral immediate early
875 RNA-activated transcription of host fucosyltransferase genes. Glycobiology *19*, 847-859.
876 Peterson, V.M., Zhang, K.X., Kumar, N., Wong, J., Li, L., Wilson, D.C., Moore, R., McClanahan,
877 T.K., Sadekova, S., and Klappenbach, J.A. (2017). Multiplexed quantification of proteins and
878 transcripts in single cells. Nat Biotechnol *35*, 936-939.
879 Piguet, V., Schwartz, O., Le Gall, S., and Trono, D. (1999). The downregulation of CD4 and
880 MHC-I by primate lentiviruses: a paradigm for the modulation of cell surface receptors. Immunol
881 Rev *168*, 51-63.

882 Puryear, W.B., Yu, X., Ramirez, N.P., Reinhard, B.M., and Gummuluru, S. (2012). HIV-1
883 incorporation of host-cell-derived glycosphingolipid GM3 allows for capture by mature dendritic
884 cells. Proc Natl Acad Sci U S A *109*, 7475-7480.
885 Schwarz, R.E., Wojciechowicz, D.C., Picon, A.I., Schwarz, M.A., and Paty, P.B. (1999).
886 Wheatgerm agglutinin-mediated toxicity in pancreatic cancer cells. Br J Cancer *80*, 1754-1762.
887 Sen, N., Mukherjee, G., Sen, A., Bendall, S.C., Sung, P., Nolan, G.P., and Arvin, A.M. (2014).
888 Single-cell mass cytometry analysis of human tonsil T cell remodeling by varicella zoster virus.
889 Cell Rep *8*, 633-645.
890 Stevenson, M., Stanwick, T.L., Dempsey, M.P., and Lamonica, C.A. (1990). Hiv-1 Replication Is
891 Controlled at the Level of T-Cell Activation and Proviral Integration. Embo Journal *9*, 1551-1560.
892 Stoeckius, M., Hafemeister, C., Stephenson, W., Houck-Loomis, B., Chattopadhyay, P.K.,
893 Swerdlow, H., Satija, R., and Smibert, P. (2017). Simultaneous epitope and transcriptome
894 measurement in single cells. Nat Methods *14*, 865-868.
895 Sugawara, S., Thomas, D.L., and Balagopal, A. (2019). HIV-1 Infection and Type 1 Interferon:
896 Navigating Through Uncertain Waters. AIDS Res Hum Retroviruses *35*, 25-32.
897 Swigut, T., Shohdy, N., and Skowronski, J. (2001). Mechanism for down-regulation of CD28 by
898 Nef. EMBO J *20*, 1593-1604.
899 Taylor, G., Vimr, E., Garman, E., and Laver, G. (1992). Purification, crystallization and
900 preliminary crystallographic study of neuraminidase from Vibrio cholerae and Salmonella
901 typhimurium LT2. J Mol Biol *226*, 1287-1290.
902 Van Gassen, S., Callebaut, B., Van Helden, M.J., Lambrecht, B.N., Demeester, P., Dhaene, T.,
903 and Saeys, Y. (2015). FlowSOM: Using self-organizing maps for visualization and interpretation
904 of cytometry data. Cytom Part A *87a*, 636-645.
905 Varchetta, S., Lusso, P., Hudspeth, K., Mikulak, J., Mele, D., Paolucci, S., Cimbro, R., Malnati,
906 M., Riva, A., Maserati, R.*, et al.* (2013). Sialic acid-binding Ig-like lectin-7 interacts with HIV-1
907 gp120 and facilitates infection of CD4pos T cells and macrophages. Retrovirology *10*, 154.

908 Williams, G.J., and Thorson, J.S. (2009). Natural product glycosyltransferases: properties and
909 applications. Advances in enzymology and related areas of molecular biology *76*, 55-119.
910 Xiao, H., Woods, E.C., Vukojicic, P., and Bertozzi, C.R. (2016). Precision glycocalyx editing as a
911 strategy for cancer immunotherapy. P Natl Acad Sci USA *113*, 10304-10309.
912 Xie, G., Luo, X., Ma, T., Frouard, J., Neidleman, J., Hoh, R., Deeks, S.G., Greene, W.C., and
913 Roan, N.R. (2021). Characterization of HIV-induced remodeling reveals differences in infection
914 susceptibility of memory CD4(+) T cell subsets in vivo. Cell Rep *35*, 109038.
915 Yang, Q., Hughes, T.A., Kelkar, A.J., Yu, X.H., Cheng, K., Park, S., Huang, W.C., Lovell, J.F.,
916 and Neelamegham, S. (2020). Inhibition of SARS-CoV-2 viral entry upon blocking N- and O-
917 glycan elaboration. Elife *9*.
918 Yukl, S.A., Kaiser, P., Kim, P., Telwatte, S., Joshi, S.K., Vu, M., Lampiris, H., and Wong, J.K.
919 (2018). HIV latency in isolated patient CD4(+) T cells may be due to blocks in HIV
920 transcriptional elongation, completion, and splicing. Sci Transl Med *10*.
921 Zou, Z.C., Chastain, A., Moir, S., Ford, J., Trandem, K., Martinelli, E., Cicala, C., Crocker, P.,
922 Arthos, J., and Sun, P.D. (2011). Siglecs Facilitate HIV-1 Infection of Macrophages through
923 Adhesion with Viral Sialic Acids. Plos One *6*.
924

**Figure 1**



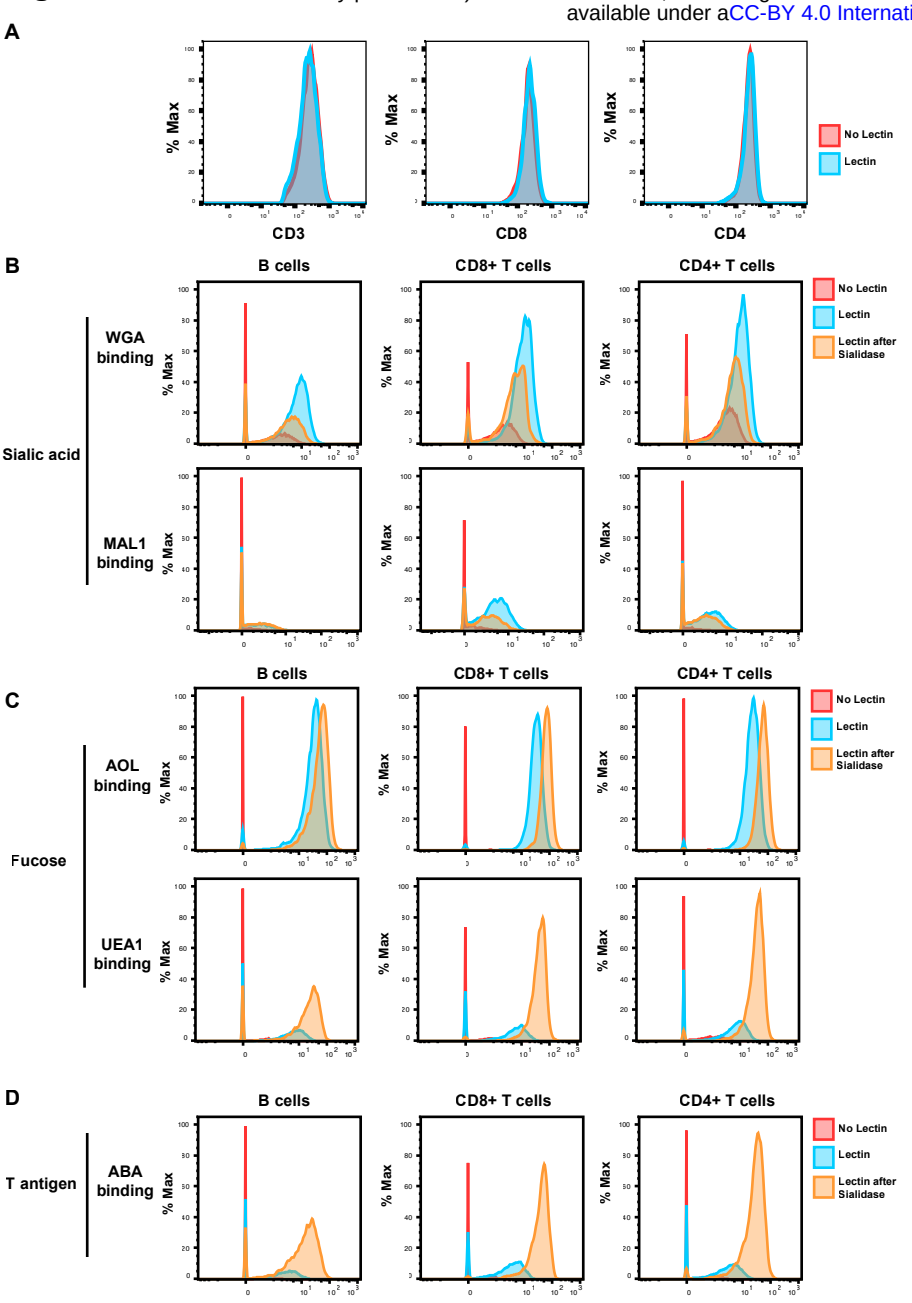


**Figure 2**



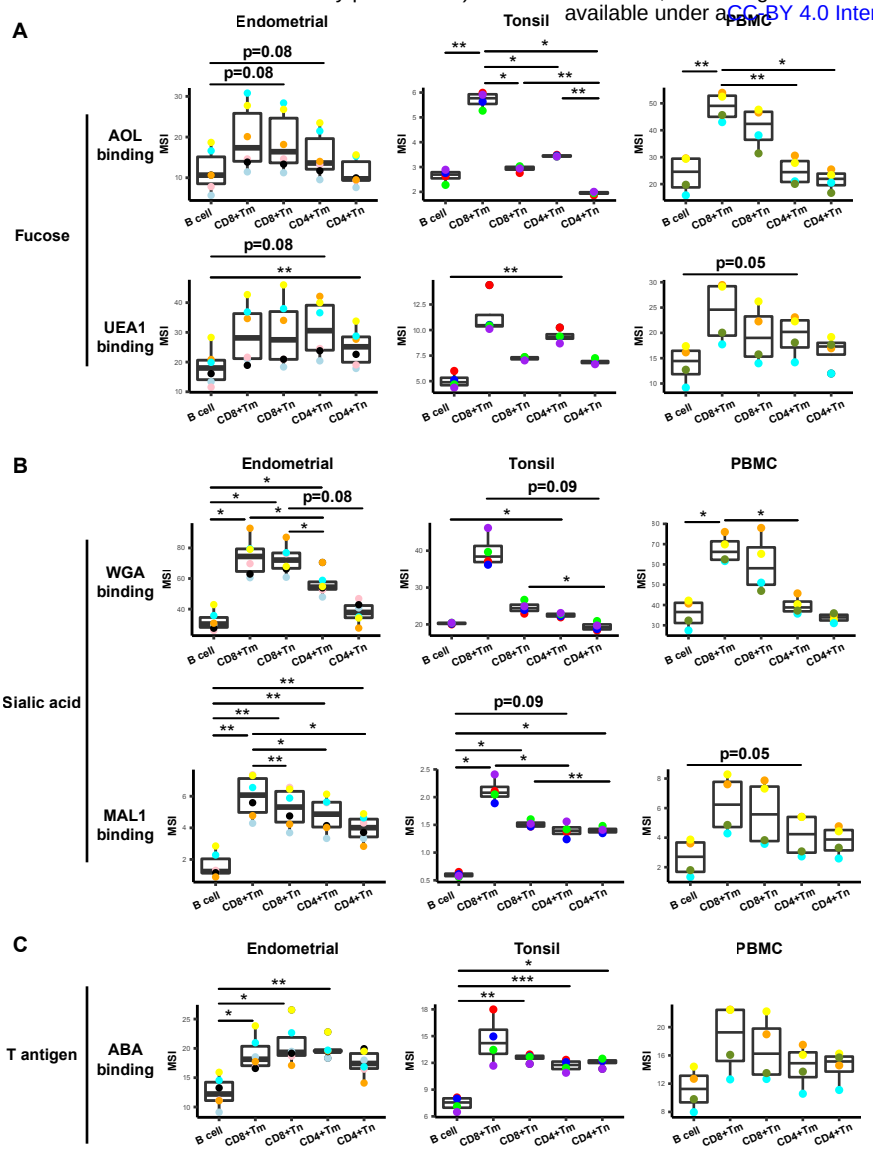


**Figure 3**

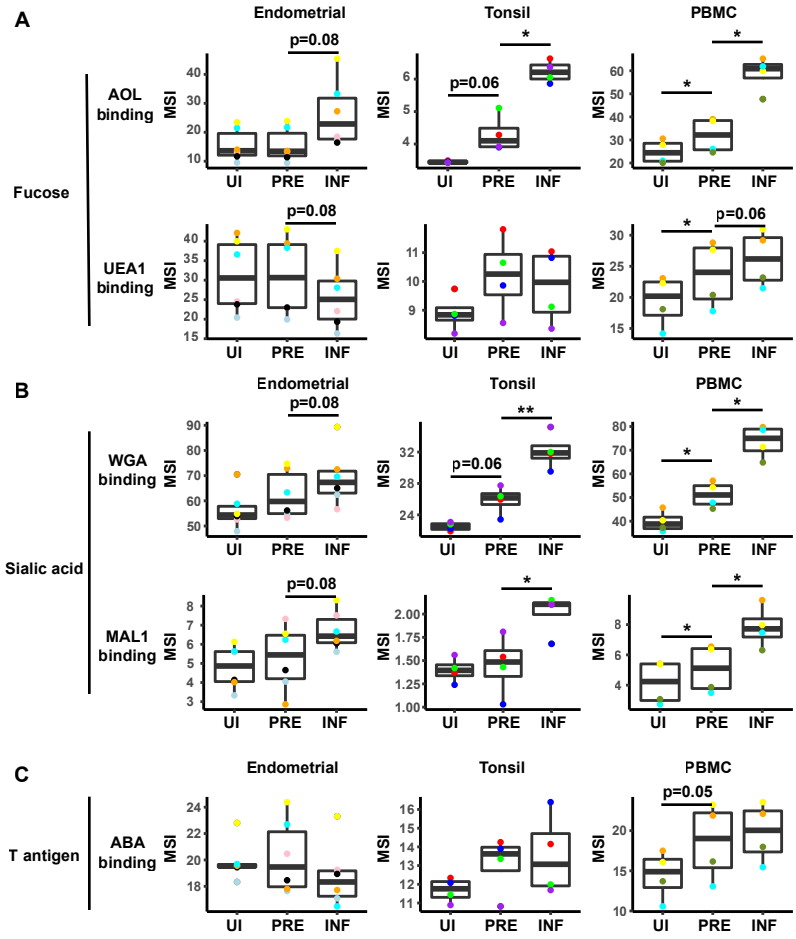


**Figure 4**



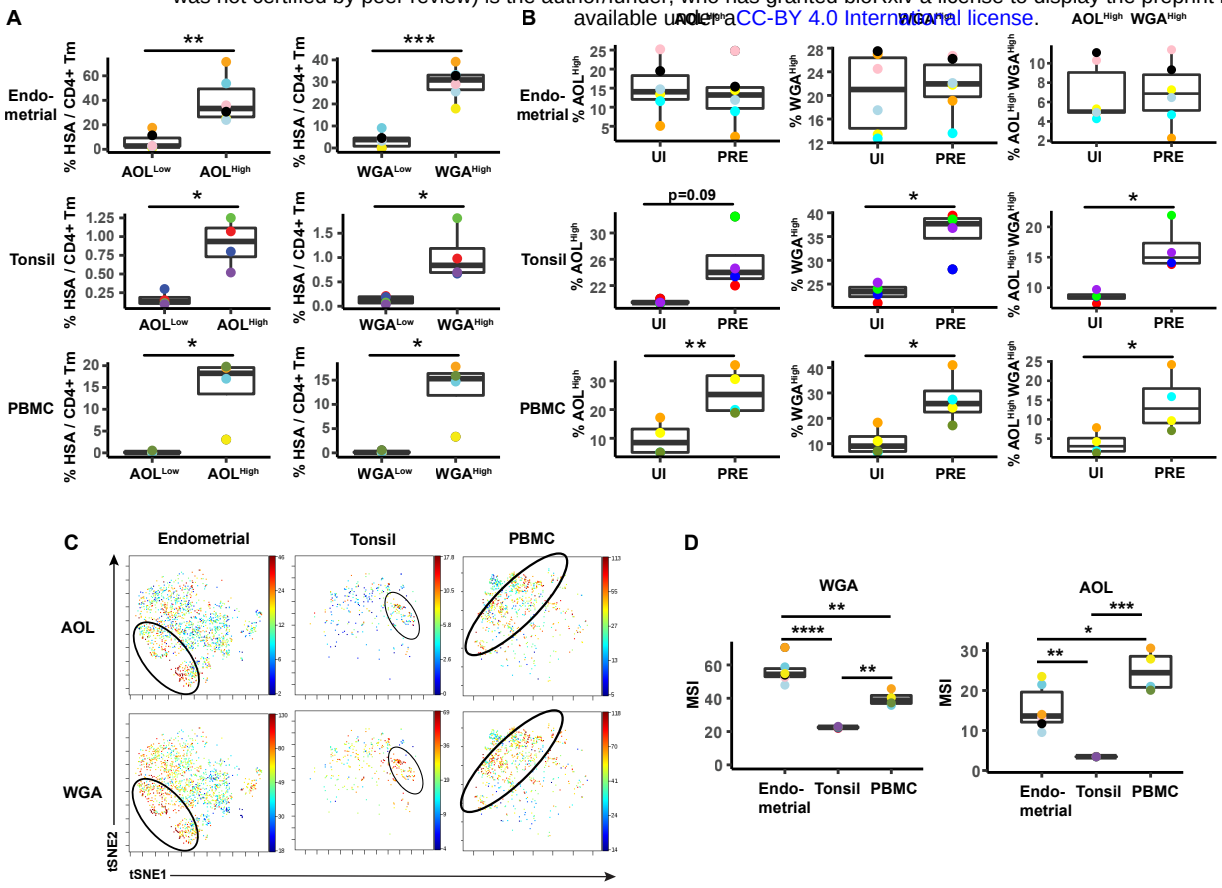


**Figure 5**



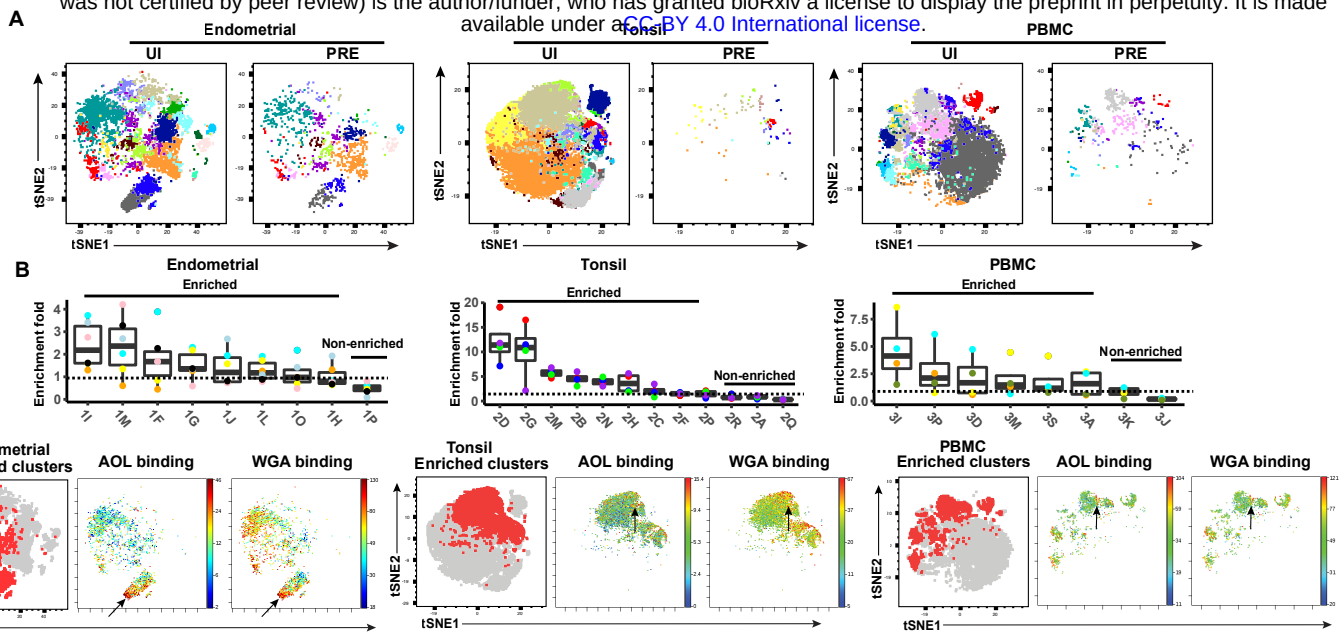
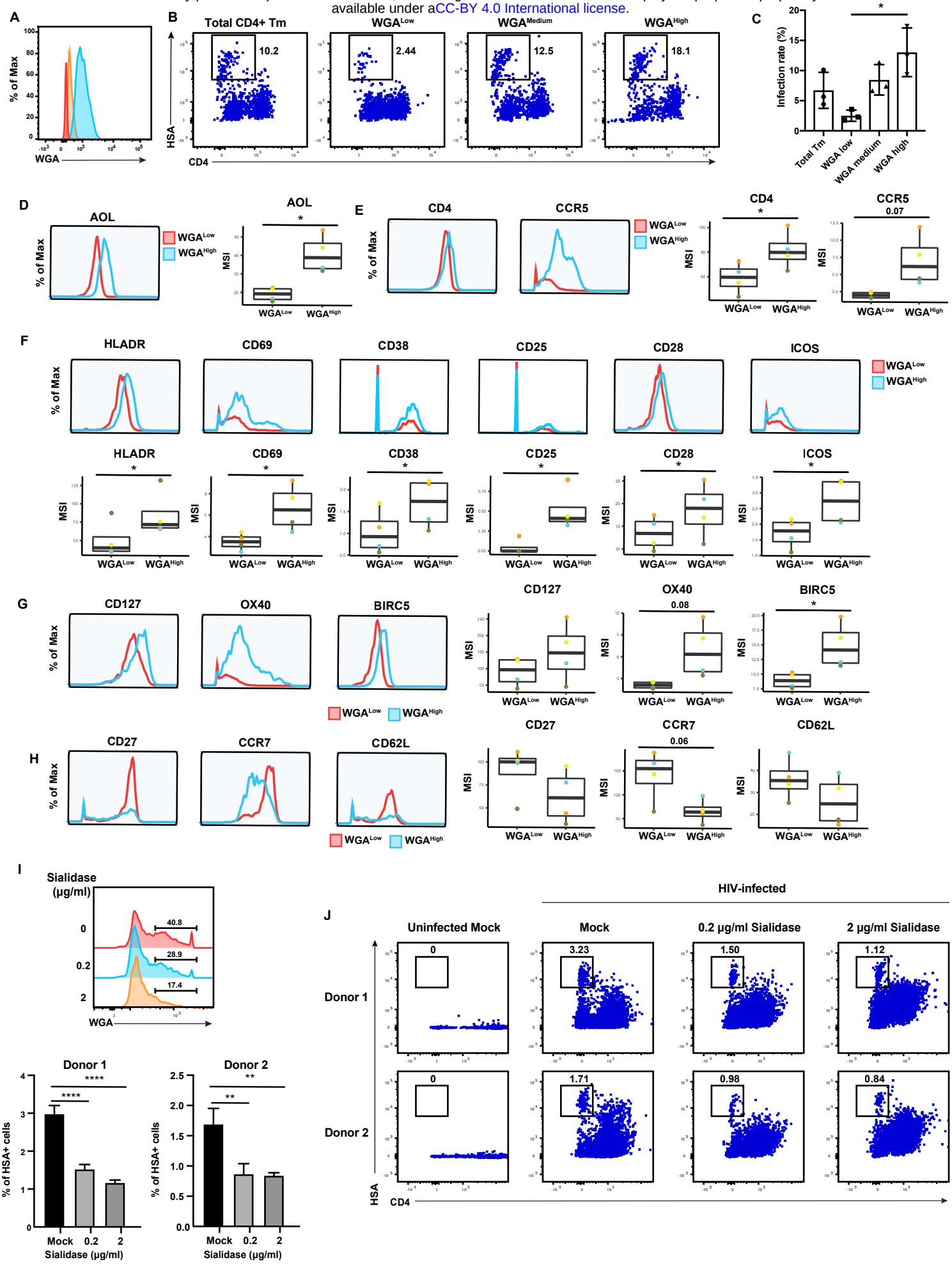


**Figure 6**





## SUPPLEMENTAL FIGURE LEGENDS

**Figure S1. Validation of antibodies and lectins used for CyTOF analysis. (A)** Expression levels of each protein antigen and glycan (the latter identified by their lectin) on tonsillar T and B cells, as assessed by CyTOF. As demonstrated schematically in the upper left, the top population of each plot corresponds to T cells (CD3+) while the bottom population corresponds to B cells (CD3-). Each antibody in the panel was validated by standard two-dimensional plots showing expression of CD3 (y-axis) versus each indicated antibody or lectin (x-axis). The observed expression patterns of the protein antigens are consistent with known expression patterns on T cells and B cells. Results are gated on live, singlet cells. **(B)** The intracellular protein antigens NFAT1, Birc5, CTAL-4, RORγt, and Tbet were further assessed for differential expression among CD4+ memory (Tm) vs. CD4+ naïve (Tn) populations by monitoring for expression levels on CD45RO+ (marker of memory) vs. CD45RA+ (marker of naïve) cells. These antigens were all preferentially expressed in the memory compartment, consistent with their known expression patterns

**Figure S2. CyTOF gating strategy to identify uninfected and HIV-infected T cells.** Productively-infected cells were defined as those that were CD3+CD8-CD4$^{low}$HSA+. The plots on the far right show specimens from a representative uninfected (*top*) and infected (*bottom*) sample.

**Figure S3. HIV remodels T cells from both tissues and blood. (A)** Cells from each of the indicated specimens from a representative donor were mock-treated or inoculated with the CCR5-tropic HIV-1 reporter virus F4.HSA, and monitored 3 days later for levels of productive infection. Results are gated on live, singlet CD3+CD8- cells. Numbers correspond to the percentage of infected cells in each sample. **(B)** t-SNE plots of uninfected and infected cells

from all donors combined. The observation that infected cells (*red*) reside in unique regions of the t-SNEs suggest viral-induced remodeling in all specimen types. Data are representative of a total of 4-6 donors per specimen type. **(C)** Quantification of viral-induced remodeling by SLIDE reveals viral-induced remodeling in all specimen types. The dashed line corresponds to the SLIDE score in the absence of remodeling (see Methods).

**Figure S4. HIV preferentially infects memory CD4+ T cells from both tissues and blood.** **(A)** Schematic showing how Predicted Precursor (PRE) cells are identified**.** Blood or tissue cells, represented as different colored ovals, were mock-treated (*top left*) or infected with the CCR5-tropic HIV reporter virus F4.HSA (*bottom left*) for 3 days. HIV-infected cells (identified as those with a black border) were identified as cells expressing HSA that had down-regulated CD4 (Figure S2). The HIV-infected cells are distinct from cells in the uninfected (UI) culture as they have been remodeled. However, by implementing PP-SLIDE, we identify, in CyTOF high-dimensional space, the phenotypically most similar CD4+ T cells in the uninfected culture (purple and blue cells in schematic) for every HIV-infected cell (red and pink cells in schematic). The “predicted precursor” (PRE) cells thus identified display the predicted phenotypes of the original cells targeted for HIV infection, prior to HIV-induced remodeling. **(B)** t-SNE plots of uninfected and predicted precursor (PRE) cells demonstrate that CD4+ Tm is the dominant population of T cells targeted for infection in all of the 3 specimen types. CD4+ Tm, CD4+ Tn, CD8+ Tm, and CD8+ Tn cells were identified by manual gating and colored as indicated. Shown is one representative donor from each site, with the remaining donors shown in Figure S5. **(C)** Quantification of preferential infection of CD4+ Tm cells among all donors reveals enrichment of CD4+ Tm cells among PRE cells. ***p<0.001, ****p<0.0001 as assessed using the Student’s paired t test.

**Figure S5. HIV infection alters expression of fucose and sialic acid in bystander cells. (A)** HIV infection upregulates fucose expression in bystander cells. Box plots showing binding by fucose-binding proteins on uninfected (UI), bystander (Bys), and infected (INF) CD4+ T cells from the endometrium, tonsils, and PBMCs. Bystander CD4+ T cells were defined as CD4+ T cells in the infected cultures that did not express the HIV reporter HSA. All populations were pre-gated on live, singlet CD4+ Tm cells. AOL binds total/core fucose, while UEA1 binds $\alpha$1-2 branched fucose. **(B)** HIV infection upregulates sialic acid in bystander cells. Box plots showing binding by sialic acid-binding lectins. Results are presented as in *panel A*. WGA binds total sialylated glycans and MAL1 $\alpha$2-3 sialylated glycans. **(C)** Box plots showing binding by T antigen-binding lectin ABA. Results are presented as in *panel A*. *p < 0.05 as assessed using the Student's paired t test and adjusted for multiple testing using the Benjamini-Hochberg for FDR. P values between 0.05-0.1 were shown in values. UI: Uninfected; Bys: Bystander.

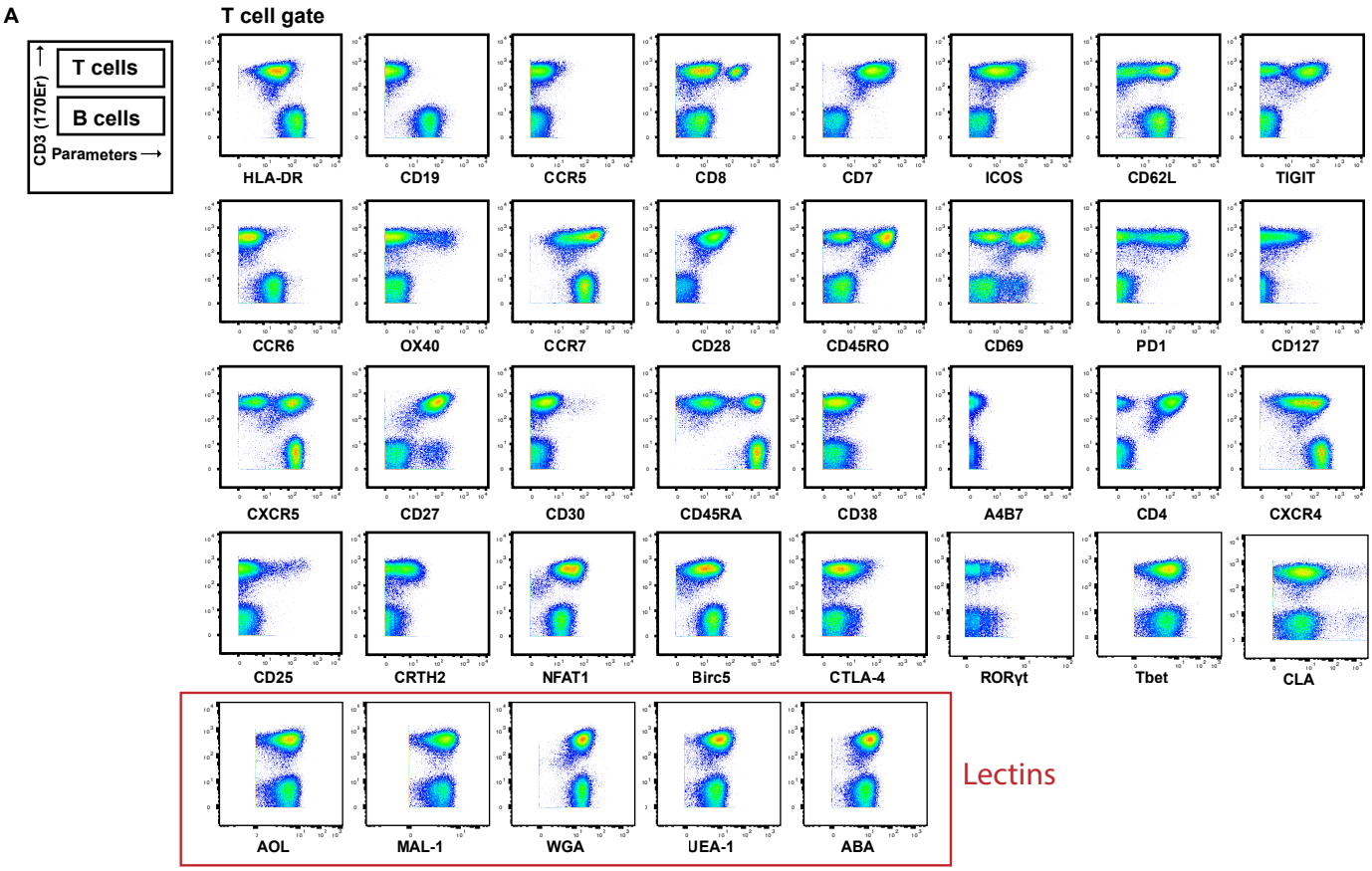
**Figure S6. HIV upregulates fucose, sialic acid, and T antigen expression in subsets of bystander cells. (A)** Box plots showing staining by fucose-binding proteins on uninfected and bystander (Bys) B and T cells from the endometrium, tonsils, and PBMCs. Bystander CD4+ T cells were defined as those that did not express the HIV reporter HSA. AOL binds to total/core fucose, and UEA1 binds to $\alpha$1-2 branched fucose. **(B)** Box plots showing staining by sialic-acid-binding lectins WGA and MAL1. Results are presented as in *panel A*. WGA binds to total sialylated glycans and MAL-1 binds to $\alpha$2-3 sialylated glycans. **(C)** Box plots showing staining by the T antigen-binding lectin ABA. Results are presented as in *panel A*. Elevation of fucose, sialic acid, and T antigen on Bys relative to uninfected (UI) cells were most apparent in tonsils. *p<0.05, **p<0.01 as assessed using the Student's paired t test and adjusted for multiple testing using the Benjamini-Hochberg for FDR. CD8+ Tm: memory CD8+ T cells; CD8+ Tn: naive CD8+ T cells; CD4+ Tm: memory CD4+ T cells; CD4+ Tn: naive CD4+ T cells

**Figure S7. Levels of glycans on FlowSOM-defined clusters.** An enrichment ratio was determined for each cluster by dividing the frequencies of the cells of that cluster among the PRE cells by their frequency among the uninfected CD4+ Tm cells. Clusters with an enrichment ratio above 1 correspond to those preferentially selected for infection. Relative binding by proteins specific for fucose *(panel A)*, sialic acid *(panel B)*, or T antigen *(panel C)* were assessed by reporting median signal intensity (MSI) of the bound lectins. Enriched clusters included those with high levels of the examined glycans. Total Tm cells are shown for comparation. *p<0.05, **p<0.01, ***p<0.001 as assessed using the Student's paired t test and adjusted for multiple testing using the Benjamini-Hochberg for FDR, and correspond to statistically significant differences between each cluster and total memory CD4+ T cells.

**Figure S8. Activated CD4+ Tm cells express high levels of sialic acid. (A)** Histograms showing the expression of activation markers on CD4+ Tm cells from resting (blue) and PHA-stimulated PBMCs (red). One of four representative donors is shown. **(B)** Stimulated CD4+ Tm cells express more sialic acid than resting CD4+ Tm cells do. Shown are the 4 PBMC donors. **(C)** $HLADR^{High}$, $CD69^{High}$, $CD38^{High}$, $CD38^{High}$, $CD25^{High}$, $CD28^{High}$, and $ICOS^{High}$ CD4+ Tm cells bind more WGA than CD4+ Tm cells expressing low levels of these activation markers. Shown are the 4 PBMC donors. For panels B-C, *p < 0.05, **p < 0.01 as assessed using the Student's paired t test and adjusted for multiple testing using the Benjamini-Hochberg for FDR. **(C)** Treatment with the sialic acid synthase inhibitor P-3$F_{ZX}$-Neu5Ac does not decrease cell-surface levels of sialidase on CD4+ T cells. PBMCs were treated for 24 h with the indicated concentrations of P-3$F_{ZX}$-Neu5Ac prior to assessment of cell-surface WGA binding. Shown are overlaid histograms demonstrating no decrease in cell-surface sialic acid levels (as reflected by WGA binding) in the inhibitor-treated cells. Results are gated on live, singlet CD3+CD8-CD4+ cells. Numbers correspond to percent of cells within the indicated gate.

**Figure S9. Expression levels of CyTOF antigens on WGA$^{Low}$ and WGA$^{High}$ Tm cells.** Shown are histogram and box plots depicting antibody and lectin staining of the WGA$^{Low}$ (*red*) and WGA$^{High}$ (*blue*) Tm cells. Shown are antigens not already depicted in Figure 6. Data from one of four representative PBMC donors is shown in histogram plots. Data from four independent PBMC donors is shown in box plots. *$p<0.05$ as assessed using the Student's paired t test and adjusted for multiple testing using the Benjamini-Hochberg for FDR.

**Figure S1**



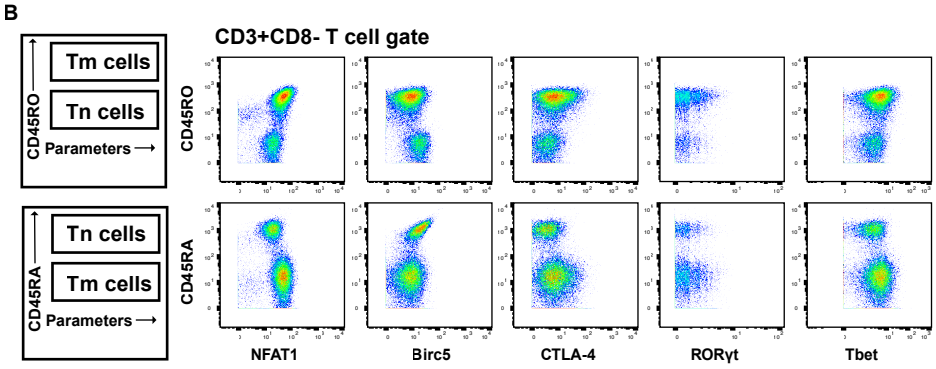



**Figure S2**

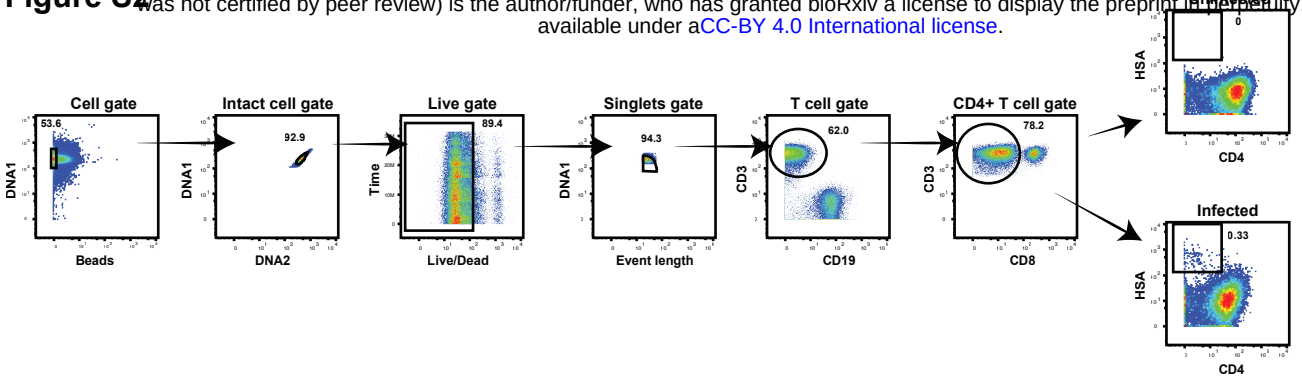


**Figure S3**



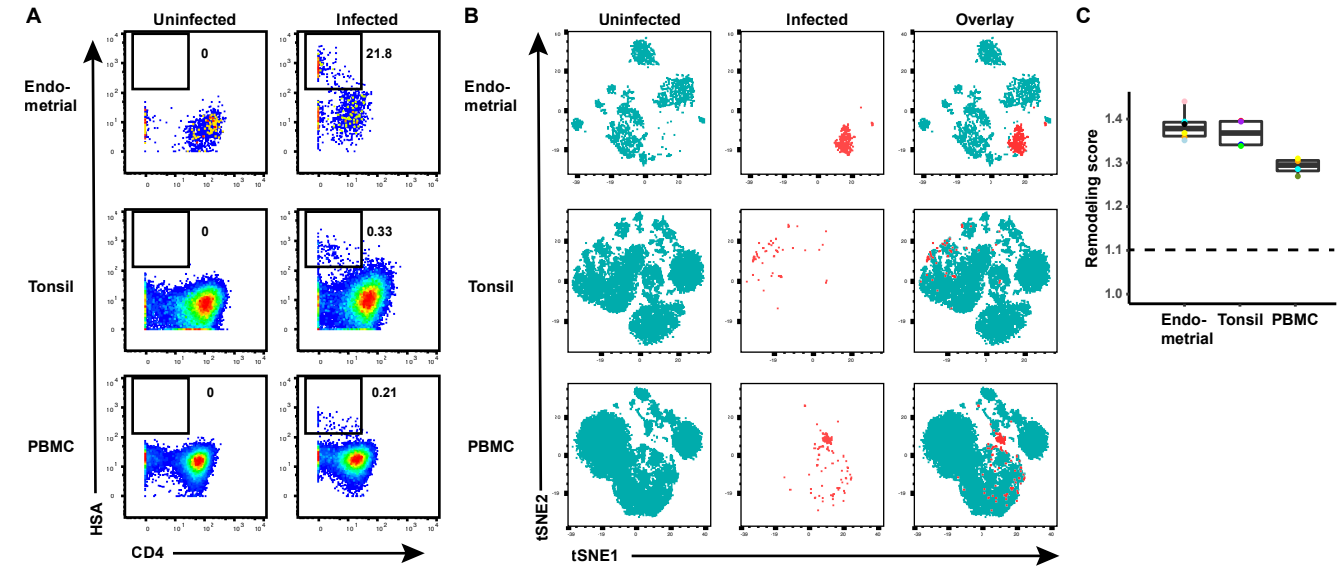
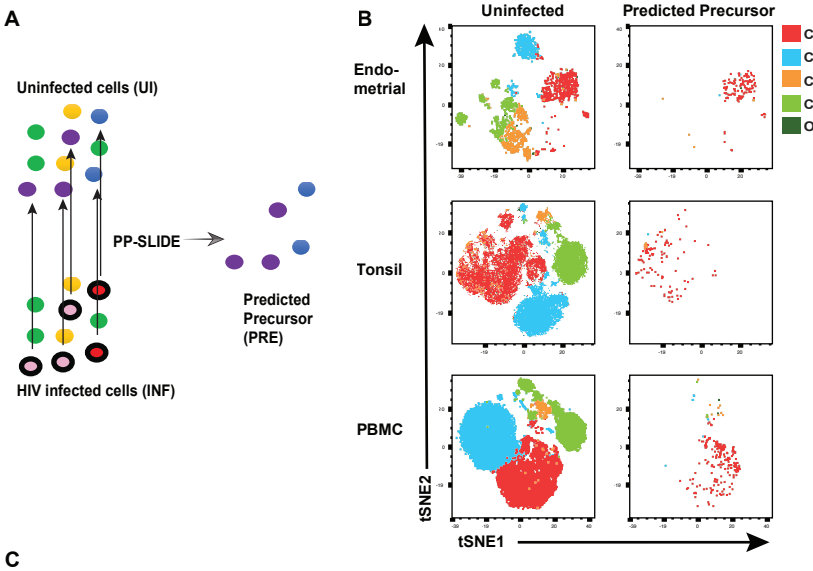


**Figure S4**



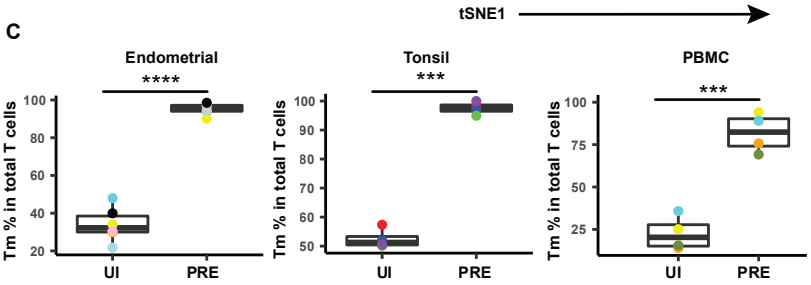
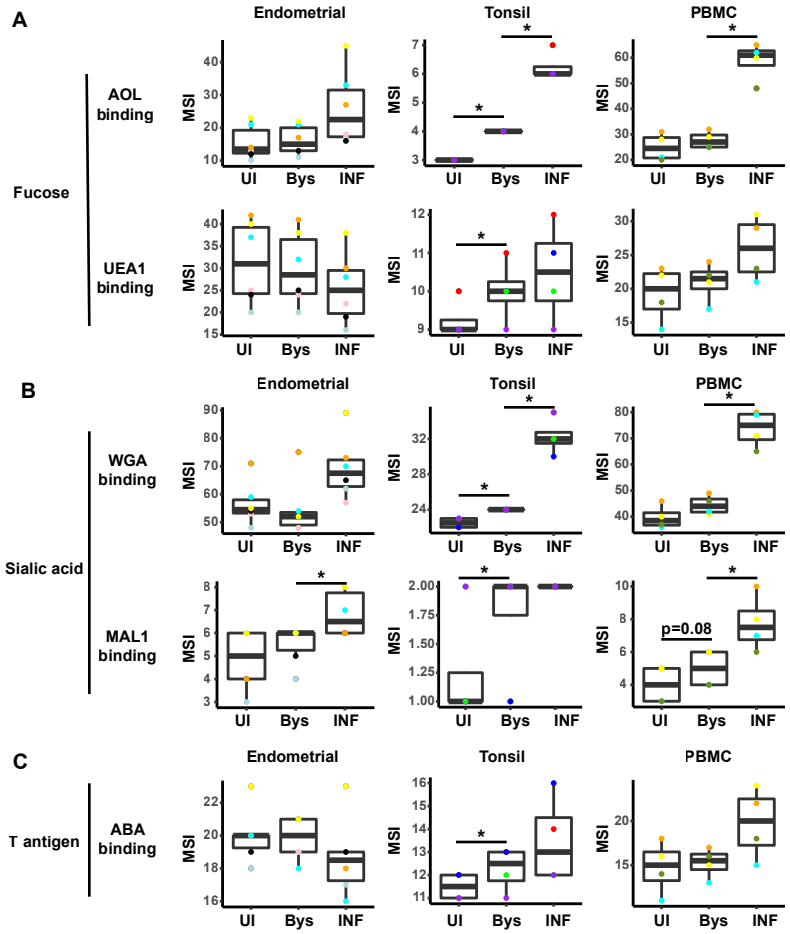



**Figure S5**



# Figure S6

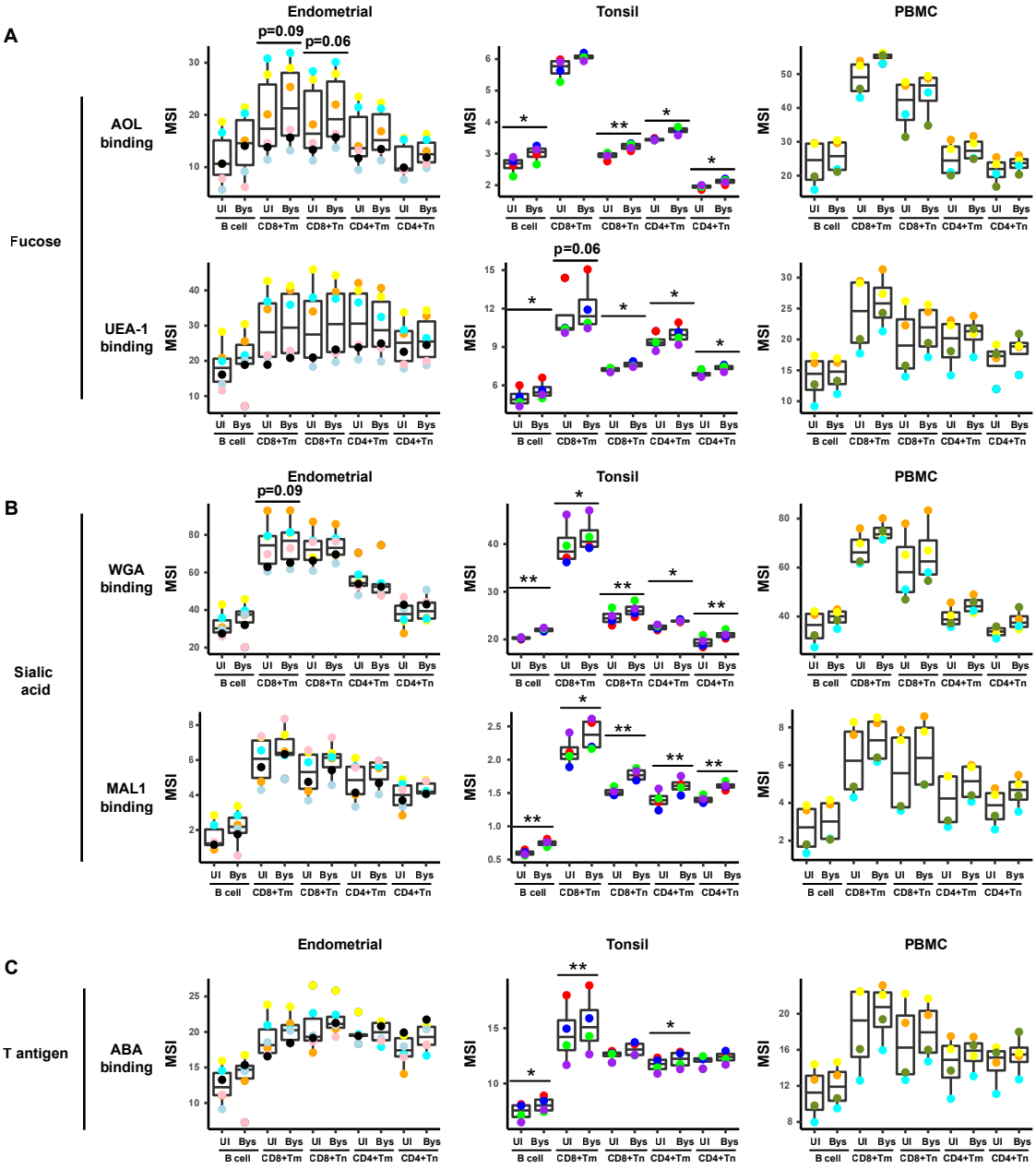
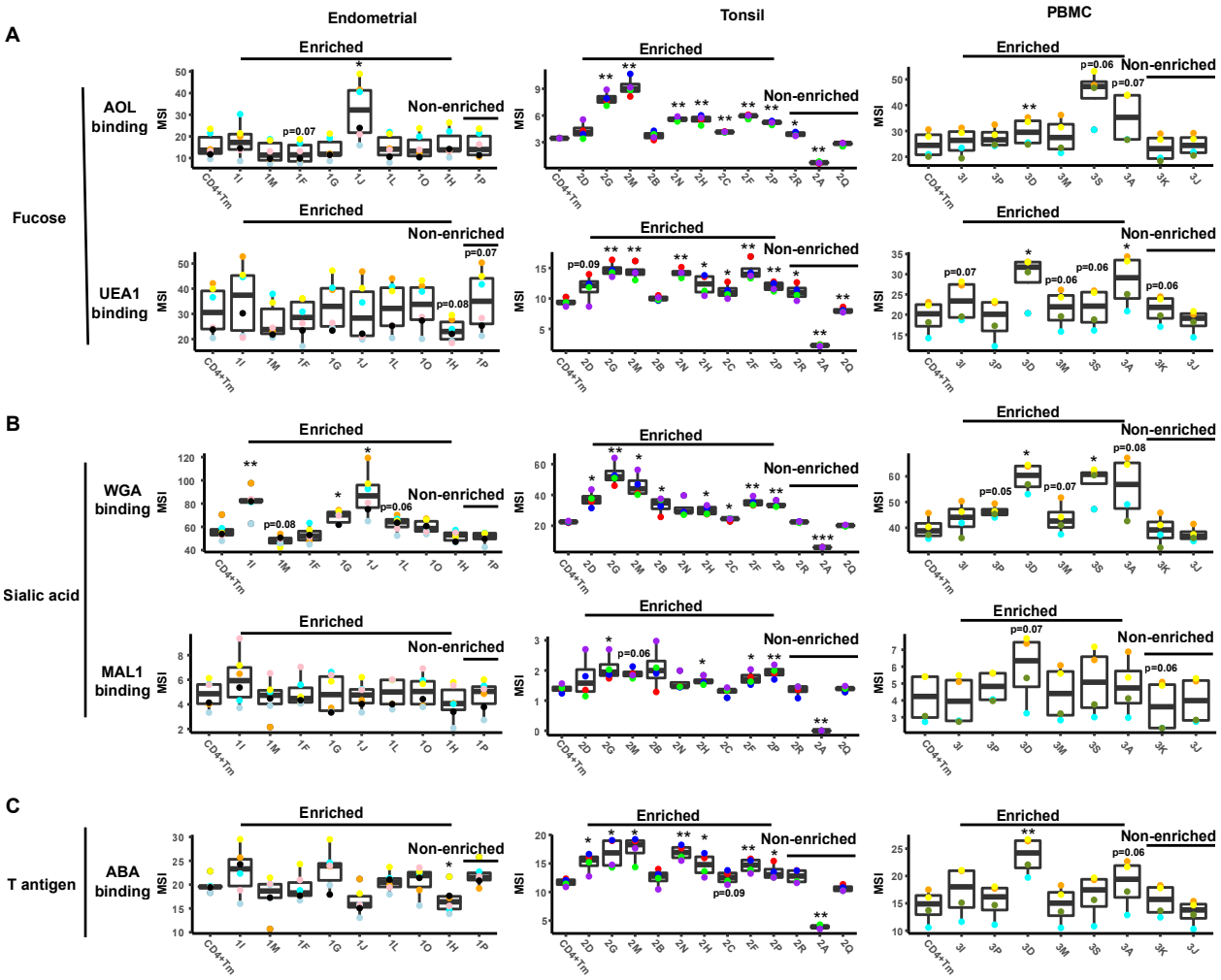


**Figure S7**





# Figure S8

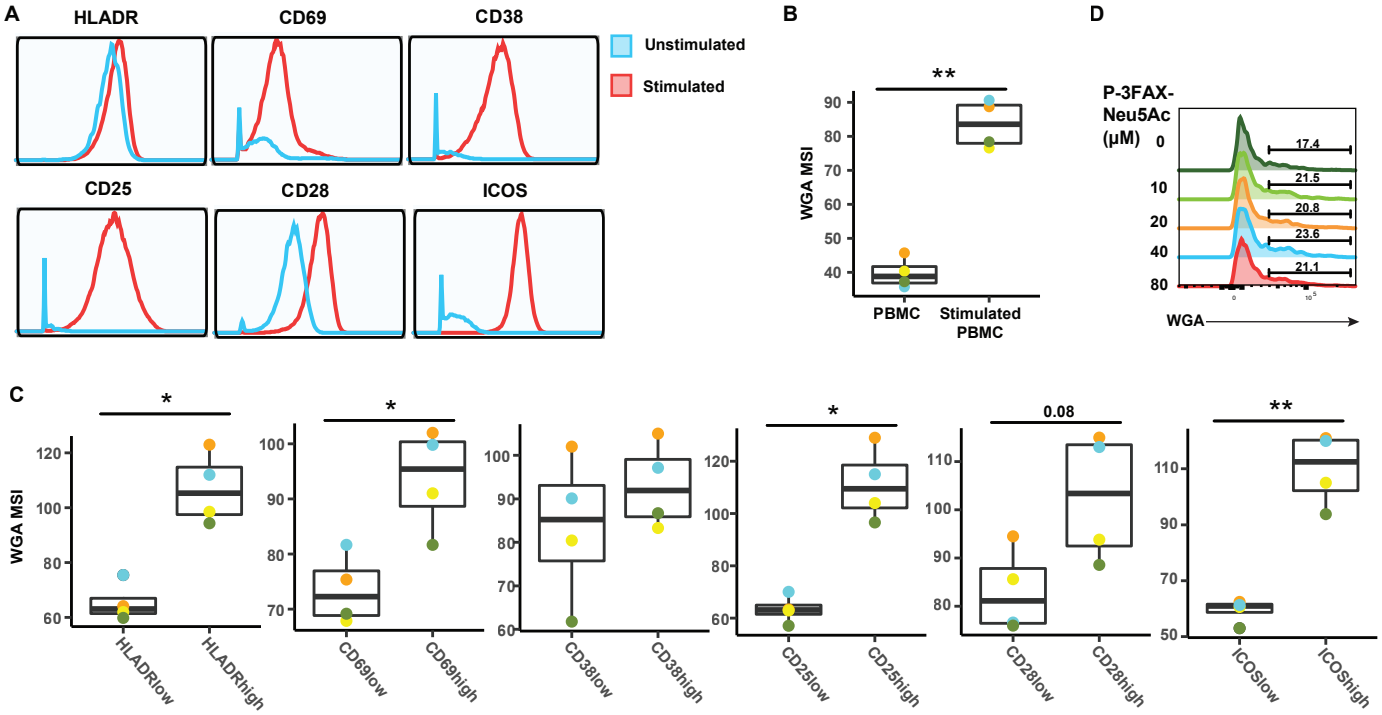
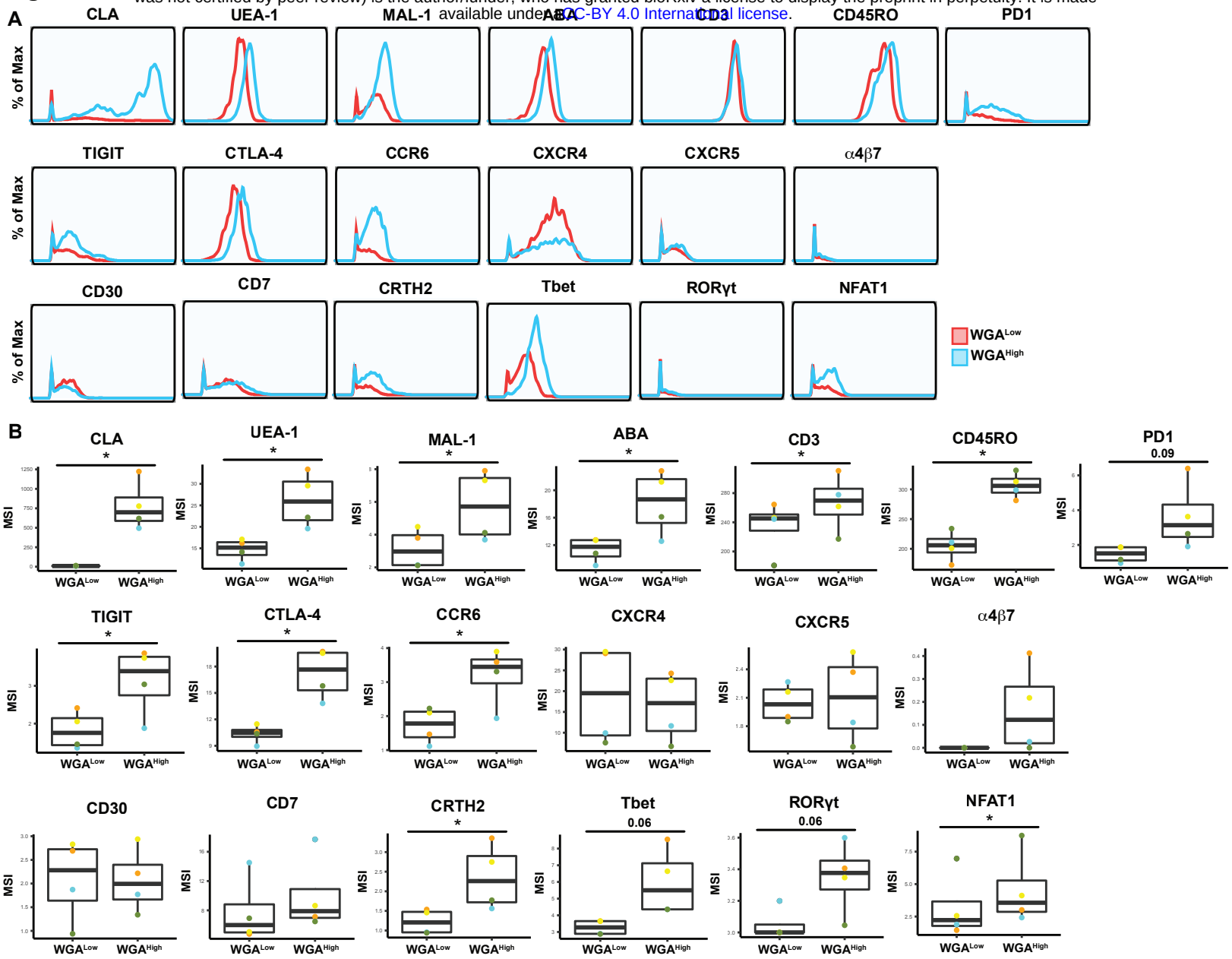


**Figure S9**



## SUPPLEMENTARY TABLES

### Table S1. List of CyTOF staining antibodies

| Antibody | Metal label | Clone | Vendor |
|---|---|---|---|
| AOL* | 141Pr | N/A | In-house |
| CD19 | 142Nd | HIB19 | Fluidigm |
| CCR5 | 144Nd | NP6G4 | Fluidigm |
| CD8 | 146Nd | RPAT8 | Fluidigm |
| CD7 | 147Sm | CD76B7 | Fluidigm |
| ICOS | 148Nd | C398.4A | Fluidigm |
| HSA | 150Nd | M1/69 | In-house |
| MAL-1* | 151Eu | N/A | In-house |
| WGA* | 152Sm | N/A | In-house |
| CD62L | 153Eu | DREG56 | Fluidigm |
| TIGIT | 154Sm | MBSA43 | Fluidigm |
| CCR6 | 155Gd | G034E3 | In-house |
| UEA-1* | 156Gd | N/A | In-house |
| OX40 | 158Gd | ACT35 | Fluidigm |
| CCR7 | 159Tb | G043H7 | Fluidigm |
| CD28 | 160Gd | CD28.2 | Fluidigm |
| CD45RO | 161Dy | UCHL1 | In-house |
| CD69 | 162Dy | FN50 | Fluidigm |
| CRTH2 | 163Dy | BM16 | Fluidigm |
| PD1 | 164Dy | EH12.1 | In-house |
| CD127 | 165Ho | A019D5 | Fluidigm |
| CXCR5 | 166Er | RF8B2 | In-house |
| CD27 | 167Er | L128 | Fluidigm |
| CD30 | 168Er | BerH8 | In-house |
| CD45RA | 169Tm | HI100 | Fluidigm |
| CD3 | 170Er | UCHT1 | Fluidigm |
| ABA* | 171Yb | N/A | In-house |
| CD38 | 172Yb | HIT2 | Fluidigm |
| $\alpha 4\beta 7$ | 173Yb | Act1 | In-house |
| CD4 | 174Yb | SK3 | Fluidigm |
| CXCR4 | 175Lu | 12G5 | Fluidigm |

| | | | |
|---|---|---|---|
| CD25 | 176Yb | M-A251 | In-house |
| CLA | 209Bi | HECA-452 | In-house |
| HLADR | 112Cd | Tu36 | Invitrogen |
| RORγt[#] | 115Di | AFKJS-9 | In-house |
| NFAT1[#] | 143Nd | D43B1 | Fluidigm |
| BIRC5[#] | 145Nd | 91630 | In-house |
| Tbet[#] | 149Sm | eBio4B10 (4B10) | In-house |
| CTLA4[#] | 157Gd | 14D3 | In-house |

**lectins, [#]Intracellular antibodies*

**Table S2. Binding properties of lectins used in this study**

| Property | MAL-1 | WGA | UEA-1 | AOL | ABA |
|---|---|---|---|---|---|
| Full Name/Source | Maackia Amuren-sis I | Wheat Germ Agglutinin | Ulex Euro-paeus I | Aspergillus Oryzae | Agaricus Bisporus Agglutinin |
| Subunits Number | 2 | 2 | 2 | 1 | 4 |
| Molecular Weight (Da) | 130 k | 36 k | 63 k | 35 k | 68 k |
| Carbohydrate specificity | Sia(α2-3)Gal (β1-4) GlcNAc | Sialic acid/ β1-4 GlcNAc | α1-2 branched fucose | total/α1-6 core fucose | T antigen, Galβ1-3GalNAcα1 |
| Sialic acids binding | α2-3 linked | Total | No | No | No |

*GlcNAc: N-acetylglucosamine, GalNAc: N-acetylgalactosamine*