Title

Combined protein and transcript single cell RNA sequencing in human peripheral blood mononuclear cells

Authors

Jenifer Vallejo,[1#] Ryosuke Saigusa,[1#] Rishab Gulati,[1] Yanal Ghosheh,[1] Christopher P. Durant,[1] Payel Roy,[1] Erik Ehinger,[1] Vasantika Suryawanshi,[1] Tanyaporn Pattarabanjird,[2] Lindsey E. Padgett,[1] Claire E. Olingy,[1] David B. Hanna,[3] Alan L. Landay,[4] Russell P. Tracy,[5] Jason M. Lazar,[6] Wendy J. Mack,[7,8] Kathleen M. Weber,[9] Adaora A. Adimora,[10] Howard N. Hodis,[7,8] Phyllis C. Tien,[11] Igho Ofotokun,[12] Sonya L. Heath,[13] Rafael Blanco-Dominguez[14], Huy Q. Dinh,[15] Avishai Shemesh,[16] Coleen A. McNamara,[2] Lewis L. Lanier,[16] Catherine C. Hedrick,[1] Robert C. Kaplan,[3,17] Klaus Ley.[1,18*]

Affiliations

[1]La Jolla Institute for Immunology, La Jolla, CA, USA.

[2]Carter Immunology Center, Cardiovascular Division, Department of Medicine, University of Virginia, Charlottesville, VA, USA.

[3]Albert Einstein College of Medicine, Department of Epidemiology and Population Health, Bronx, NY, USA.

[4]Rush University Medical Center, Department of Internal Medicine, Chicago, IL, USA.

[5]University of Vermont Larner College of Medicine, Departments of Pathology & Laboratory Medicine and Biochemistry, Colchester, VT, USA.

[6]SUNY Downstate Health Sciences University, Department of Medicine, Brooklyn, NY, USA.

[7]Keck School of Medicine, University of Southern California, Department of Medicine and Preventive Medicine, Los Angeles, CA, USA.

[8]Atherosclerosis Research Unit, University of Southern California, Los Angeles, CA, USA.

[9]Cook County Health/Hektoen Institute of Medicine, Chicago, IL, USA.

[10]Department of Medicine, University of North Carolina School of Medicine, The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA.

[11]Department of Medicine, University of California, San Francisco, San Francisco, CA and Department of Veterans Affairs Medical Center, San Francisco, CA, USA.

[12]Emory University School of Medicine, Department of Medicine, Infectious Disease Division and Grady Health Care System, Atlanta, GA, USA.

[13]University of Alabama at Birmingham, Department of Medicine, Birmingham, AL, USA.

[14]Centro Nacional de Investigaciones Cardiovasculares (CNIC), Madrid, Spain.

[15]McArdle Laboratory for Cancer Research, University of Wisconsin-Madison School of Medicine and Public Health, Madison, Wisconsin, USA.

[16]Parker Institute for Cancer Immunotherapy, University of California, San Francisco, CA, USA; Department of Microbiology and Immunology, University of California, San Francisco, CA, USA.

[17] Fred Hutchinson Cancer Research Center, Public Health Sciences Division, Seattle, WA, USA.

[18] Department of Bioengineering, University of California San Diego, San Diego, CA, USA.

[#] Contributed equally to this work

Corresponding Author

Klaus Ley, MD

La Jolla Institute for Immunology

9420 Athena Circle

La Jolla, CA 92037, USA

(858) 752-6661 (tel)

(858) 752-6985 (fax)

klaus@lji.org

Abstract

Cryopreserved peripheral blood mononuclear cells (PBMCs) are frequently collected and provide disease- and treatment-relevant data in clinical studies. Here, we developed combined protein (40 antibodies) and transcript single cell (sc)RNA sequencing in PBMCs. Among 31 participants in the WIHS Study, we sequenced 41,611 cells. Using Boolean gating followed by Seurat UMAPs and Louvain clustering, we identified 58 subsets among CD4 T, CD8 T, B, NK cells and monocytes. This resolution was superior to flow cytometry, mass cytometry or scRNA-sequencing without antibodies. Since the transcriptome was not needed for cell identification, combined protein and transcript scRNA-Seq allowed for the assessment of disease-related changes in transcriptomes and cell type proportion. As a proof-of-concept, we showed such differences between healthy and matched individuals living with HIV with and without cardiovascular disease. In conclusion, combined protein and transcript scRNA sequencing is a suitable and powerful method for clinical investigations using PBMCs.

Key Words

CVD, HIV, scRNA-seq, transcriptomes, antibodies, human.

81 **MAIN TEXT**

82 **Introduction**

83 PBMCs are a rich source of disease- and treatment-relevant information.[3, 9, 22, 30, 38,
84 53, 60–62] PBMCs can be analyzed without mechanical or enzymatic dissociation, which
85 are known to alter cell surface markers and transcriptomes.[56] PBMCs can be
86 cryopreserved without loss of viability. At the most basic level, lymphocytes and
87 monocyte can be distinguished by morphology using automated cell counters (CBC).[4]
88 Current practice is to use flow cytometry of between 8 and 16 markers.[40, 43, 52] More
89 recently, mass cytometry became available,[1, 11, 47, 57] allowing for analysis of up to
90 40 markers. Single cell RNA-sequencing (scRNA-Seq) allows the interrogation of
91 expressed genes[8, 28, 35, 48, 63] and surface markers.[48, 64]

92 In immune cells, the correlation between mRNA and surface expression of any given
93 surface marker is weak.[25] This is because cell surface expression is not only
94 determined by gene expression, but also by posttranslational protein modifications,[26]
95 trafficking to the cell surface, protein stability, and proteolytic modifications. Cell types
96 in PBMCs have been defined by flow cytometry, and the surface markers of the major
97 cell types are very well known. Yet, it is difficult to call even major cell types by scRNA-
98 Seq. For example, CD4 T cells are not resolved from CD8 T cells and natural killer (NK)
99 cells.[59] To capitalize on the vast flow and mass cytometry literature, it is necessary to
100 assess cell surface phenotype along with transcriptomes.

101 Only two publications reported single cell transcriptomes from patients with
102 atherosclerosis (carotid endarterectomy specimens and matched PBMCs).[9, 58] 1,652

103 PBMCs from one patient were analyzed by 10x Genomics 3' and cellular indexing of
104 transcriptional epitope sequencing (CITE-Seq),[35, 48] using a panel of 21 antibodies.
105 No healthy control PBMCs were studied. A recent study reported the effect of HIV
106 infection on PBMC transcriptomes,[20] focusing on acute HIV infection (before
107 antiretroviral therapy started) and reporting PBMC transcriptomes in four patients at 8
108 defined time points (average of 1,976 PBMC transcriptomes per participant and
109 condition). No scRNA-Seq or CITE-Seq studies of PBMCs of people living with chronic
110 HIV infection have been reported. No single cell studies of the interaction between HIV
111 and CVD are available.
112 Here, we report transcriptomes and cell surface phenotypes of almost 42,000 PBMCs
113 using the targeted scRNA-Seq BD Rhapsody platform[8, 28] that simultaneously
114 provides single cell surface phenotype (40 monoclonal antibodies, mAbs) and
115 transcriptomes (485 immune and inflammatory transcripts) in the same cells. As a proof-
116 of-concept, we show significant differences in cell proportions and cell transcriptomes
117 between healthy subjects and matched subjects living with HIV or cardiovascular disease
118 from the WIHS cardiovascular sub-study. WIHS is an ongoing multi-center, prospective,
119 observational cohort study of women with or at risk of HIV infection. PBMCs were
120 cryopreserved on liquid N_2, following strict standard operating procedures that ensured
121 preservation of cell surface phenotype, viability, and transcriptomes.
122
123
124
125
126
127

Material and Methods

Study characteristics and sample selection. The Women's Interagency HIV Study (WIHS) was initiated in 1994 at six (now expanded to ten) U.S. locations.[13, 16] It is an ongoing prospective study of over 4,000 women with or at risk of HIV infection. Recruitment in the WIHS occurred in four phases (1994-1995, 2001-2002, 2010-2012, and 2013-2015) from HIV primary care clinics, hospital-based programs, and community outreach and support groups. Briefly, the WIHS involves semi-annual follow-up visits, during which participants undergo similar detailed examinations, specimen collection, and structured interviews assessing health behaviors, medical history, and medication use. All participants provided informed consent, and each site's Institutional Review Board approved the studies.

All participants in the current analysis were part of a vascular sub-study nested within the WIHS.[13, 16, 18] The baseline visit for the vascular sub-study occurred between 2004 and 2006, and a follow-up visit occurred on average seven years later. Participants underwent high-resolution B-mode carotid artery ultrasound to image six locations in the right carotid artery: the near and far walls of the common carotid artery, carotid bifurcation, and internal carotid artery. A standardized protocol was used at all sites,[18] and measurements of carotid artery focal plaque, a marker of subclinical atherosclerosis, were obtained at a centralized reading center (U. of Southern California). Subclinical CVD (sCVD) was defined based on the presence of one or more carotid artery lesions.[18]

152 From the initial 1,865 participants in the WIHS vascular sub-study, 32 participants were
153 selected for scRNA-seq analysis. CVD was defined as presence of carotid artery focal
154 plaque at either vascular sub-study visit to define four groups of eight participants each:
155 HIV-, HIV+CVD-, HIV+CVD+, HIV+CVD+ on CRT. Because we were interested in the
156 joint relationships of HIV infection and sCVD with surface marker and RNA expression
157 by different cell subtypes, we selected matched samples based on HIV, CVD and
158 cholesterol-reducing treatment (CRT, mostly statins). The latter was done because we
159 found that CRT had a major impact on monocyte transcriptomes.[7]. HIV infection status
160 was ascertained by enzyme-linked immunosorbent assay (ELISA) and confirmed by
161 Western blot. Non-CVD participants with self-reported coronary heart disease or current
162 lipid-lowering therapy use were excluded. Participants were formed in quartets matched
163 by race/ethnicity (except one quartet), age (± 5 years) at the baseline vascular sub-study
164 (except one quartet where the age difference was more but all the women were post-
165 menopausal), visit number, smoking history, and date of specimen collection (within 1
166 year).
167
168 Demographic, clinical, and laboratory variables were assessed from the same study visit
169 using standardized protocols. The median age at the baseline study visit was 55 years,
170 and 96% of participants were either of Black race or Hispanic ethnicity. Most (86%)
171 reported a history of smoking. Substance use was highly prevalent, with 43% of HIV+
172 and 50% of HIV- participants reporting either a history of injection drug use; current use
173 of crack, cocaine, or heroin; or alcohol use (≥14 drinks per week). Among HIV+
174 participants, over 80% reported use of HAART at the time PBMCs were obtained, and

175 59% reported an undetectable HIV-1 RNA level. The median CD4+ T-cell count was
176 585 cells/μL (IQR 382-816) in HIV+ women without sCVD and 535 cells/μL (IQR 265-
177 792) in HIV+ women with sCVD.
178
179 **Preparation of PBMC samples for combined protein and RNA-seq.** To avoid batch
180 effects, sixteen samples each were processed on the same day. PBMC tubes were thawed
181 in a 37°C water bath and tubes filled with 8 mL of complete RPMI-1640 solution (**Table**
182 **S1**; cRPMI-1640 contains human serum albumin, HEPES, sodium pyruvate, MEM-
183 NEAA, penicillin-streptomicyn, GlutaMax, and mercaptoethanol). The tubes were
184 centrifuged at 400 xg for 5 minutes and pellets resuspended in cold staining buffer (2 %
185 fetal bovine serum (FBS) in phosphate-buffered saline (PBS)). All reagents,
186 manufacturers, and catalogue numbers are listed in **Table S1**. Manual cell counting was
187 performed by diluting cell concentration to achieve 100-400 cells per hemocytometer
188 count. Cells were aliquoted to a count of 1 million cells each and incubated on ice with
189 Fc Block (BD, **Table S1**) at a 1:20 dilution, centrifuged at 400 xg for 5 minutes,
190 resuspended in 180 μL of SB and transferred to their respective sample multiplexing kit
191 tubes (BD). The cells were incubated for 20 minutes at room temperature, transferred to 5
192 mL polystyrene tubes, washed 3 times and centrifuged at 400 xg for 5 minutes. The cells
193 were resuspended in 400 μL of staining buffer and 2 μL of DRAQ7 and calcein AM were
194 added to each tube. The viability and cell count of each tube was determined using the
195 BD Rhapsody scanner (**Table S2**). Tube contents were pooled in equal proportions with
196 total cell counts not to exceed 1 million cells. The tubes were then centrifuged at 400 xg
197 for 5 minutes and resuspended in a cocktail of 40 AbSeq (**Table S3**) antibodies (2 μL

198 each and 20 μL of staining buffer) on ice for 30-60 minutes per manufacturer's
199 recommendations. The tubes were then washed with 2 mL of SB followed by
200 centrifugation at 400 xg for 5 minutes. This was repeated two more times for a total of 3
201 washes. The cells were then counted again using the scanner.
202
203 **Library preparation.** Cells were loaded at 800-1000 cells/μL into the primed plate per
204 the BD user guide. The beads were isolated with a magnet and the supernatant removed.
205 Reverse transcription was performed at 37 ℃ on a thermomixer at 1200 rpm for 20
206 minutes. Exonuclease I was incubated at 37 ℃ on a thermomixer at 1200 rpm for 30
207 minutes and then immediately placed on a heat block at 80 ℃ for 20 minutes. The tube
208 was placed on ice followed by supernatant removal while beads were on a magnet. The
209 beads were resuspended in BD bead resuspension solution. Then, the tubes were stored at
210 4 ℃ until further processing. Per BD's protocol, the reagents for PCR1 including the BD
211 Human Immune Response Panel and a custom panel of ~100 genes (**Table S4**) were
212 added to the beads. Samples were aliquoted into four 0.2 mL strip PCR tubes and
213 incubated for 10 cycles according to BD's protocol for PCR1. A double size selection
214 was performed to remove high genomic DNA fragments by adding 0.7x volume AMPure
215 XP SPRI beads to the PCR products. After incubation, the supernatant is recovered and
216 transferred to a new tube followed by purifying the supernatant with an additional 100 μL
217 of AMPure XP beads (sample tags and antibodies). The RNA tube was washed twice
218 with 500 μL of 80 % ethanol. 550 μl of supernatant were removed from the antibody tube
219 followed by two washes with 500 μL of 80 % ethanol. The cDNA was eluted off the
220 beads using 30 μL of BD elution buffer and then transferred to a 1.5 mL tube.

221

222 **Pre-sequencing quality control (QC).** QC/ and quantification was performed on the
223 tube containing AbSeq and Sample Tags using Agilent TapeStation high sensitivity
224 D1000 screentape. 5 μL from each tube (mRNA and Ab/ST) was then added to their
225 respective tubes containing the reagents for PCR2. Each tube had 12 cycles of PCR
226 performed according to BD's user guide. Each tube was cleaned with AMPure XP beads
227 at 0.8X for mRNA and 1.2X for sample tags. Two 200 μL washes were performed during
228 the clean-up using 80 % ethanol per sample. The cDNA was eluted off using BD elution
229 buffer. QC/ and quantification was performed using Agilent TapeStation high sensitivity
230 D1000 screen tape and Qubit double stranded high sensitivity DNA test kit. The mRNA
231 was then diluted, if necessary, to a concentration of 1.2-2.7 ng/μL and the antibody and
232 sample tag libraries from PCR2 were diluted, if needed, to a concentration of 0.5-1.1
233 ng/μL. From each sample 3 μL were added to a volume of 47 μL of reagents for PCR3 as
234 described by BD's user guide following the protocol and number of cycles listed, except
235 for AbSeq, which had 9 cycles of PCR performed as determined by previous
236 optimization. The three libraries were then cleaned with AMPure XP beads at 0.7X for
237 AbSeq and 0.8X for sample tags. Samples were washed twice with 200 μL of 80 %
238 ethanol. The cDNA was eluted off the beads using BD's elution buffer. Final QC and
239 quantification was performed using TapeStation and Qubit kits and reagents.

240

241 **Sequencing.** The samples were pooled and sequenced to the following nominal depth
242 recommended by BD: AbSeq: n x 1000 reads per cell, where n is the plexity of AbSeq
243 used; mRNA: 20,000 reads per cell; Sample Tags: 600 reads per cell. Thus, a total of

244 60,600 reads per cell were desired for sequencing on the NovaSeq. The samples and
245 specifications for pooling and sequencing depth, along with number of cells loaded onto
246 each plate was optimized for S1 and S2 100 cycle kits (Illumina) with the configuration
247 of 67x8x50 bp. Once sequencing was complete, a FASTA file was generated by BD as a
248 reference for our AbSeq and genes we targeted with these assays. The FASTA file and
249 FASTQ files generated by the NovaSeq were uploaded to Seven Bridges Genomics
250 pipeline, where the data was filtered and matrices and csv files were generated. This
251 analysis generated draft transcriptomes and surface phenotypes of 54,078 cells (496
252 genes, 40 antibodies). 11 genes were not expressed, leaving 485 genes for analysis.
253
254 **Doublet Removal.** Based on the 4 sample tags used per plate, 8,359 doublets were
255 removed. The remaining 45,719 cells were analyzed using the Doublet Finder package on
256 R (https://github.com/chris-mcginnis-ucsf/DoubletFinder) with the default doublet
257 formation rate (7.5%). This removed another 3,322 doublets, leaving 42,397 Cells.
258 Finally, we removed all cells that had less than 128 (2^7) antibody molecules sequenced.
259 This removed 786 noisy cells, resulting in 41,611 cell transcriptomes. All antibody data
260 were CLR (centered log-ratio) normalized and converted to $\log_2$ scale. All transcripts
261 were normalized by total UMIs in each cell and scaled up to 1000.
262
263 **Thresholding.** Preliminary experiments showed that each antibody had both specific and
264 non-specific binding, as expected. To remove the non-specific signal, each antibody
265 threshold (**Table S5**) was obtained by determining its expression in a known negative
266 cell. To identify the thresholds, biaxial plots of mutually exclusive markers were used to

267 best separate the positive populations from the noise. In combined protein and transcript
268 panel single cell sequencing, non-specific background staining is caused by incomplete
269 Fc block and oligonucleotide-tagged antibody being trapped in the nanowell.[48]
270 Ridgeline plots of the unthresholded and thresholded antibody expressions for each main
271 cell type are shown in **Supplemental Figure S1**, which indicates how the thresholding
272 worked on each antibody expression.
273
274 **Clustering.** Clustering was performed using UMAP (Uniform Manifold Approximation
275 and Projection) and Louvain clustering.[50] UMAP is a manifold learning technique for
276 dimensionality reduction. It is based on neighborhood graphs, which captures the local
277 relationship in the data. UMAP is able to maintain local structure and also preserve
278 global distances in the reduced dimension, i.e. the cells that are similar in the high
279 dimension remain close-by in the 2 dimensions and the cells that different are apart in the
280 2 dimensions. The clustering parameters used were: n_neighbors = 100, n_pcs = 50,
281 min_dist = 1, spread = 1, random state = 42. Louvain resolution was set at 0.8.
282 Subclustering of each major cell type was based on all non-negative antibodies (**Table**
283 **S6**). Gates were overlaid and used in all subsequent UMAP figures (cell numbers in
284 **Table S7**)
285
286 **Cluster Assignment.** In CD4 T cells, 4 of the initial clusters were further divided based
287 on the expression of CD11c, CD56, CD25, CD127, CXCR3, and CCR2. CD8 T cells had
288 two clusters that were divided based on CD11c, CD16, and CXCR3 surface marker
289 expression. One cluster from classical monocytes and one cluster from intermediate

290 monocytes were further divided based on CCR7 and CD152 expression, respectively. In
291 non-classical monocytes, one cluster showed differential expression of CD36 and CD152
292 expression and was divided in two. In B cells, one cluster was split because it showed
293 differential expression of CD25 and CXCR3 within the cluster. Finally, two clusters from
294 NK cells were split due to CD16, CD56, and CD11c expression.
295
296 **Comparing Gene Expression among Participant Types.** To determine differential
297 expression (DE) among the four types of participants, we use the Seurat package [49] in
298 R with no thresholds over avg_logFC, minimum fraction of cells required in the two
299 populations being compared, minimum number of cells and minimum number of cells
300 expressing a feature in either group. We filtered for adjusted $p<0.05$ and compared HIV-,
301 HIV+CVD-, HIV+CVD+, and HIV+CVD+CRT+. From this data, volcano plots were
302 generated using ggplot2 and ggrepel packages in R. Axes were restricted to the range of
303 (-2,2) on the x-axis and (0,20) on the y-axis. Genes outside these ranges were bounded to
304 the corresponding limit of the axes. Exact p-values of differential gene expression in all
305 major cell types and the top 10 highly regulated genes for the main cell types are shown
306 in **Table S8** and **S9**, respectively.
307
308 **Comparing Cell Proportions.** To find changes in proportions, we identified the cell
309 numbers for each participant in each cluster (**Table S10)**. Statistical differences in cell
310 proportions were calculated by log-odds ratio defined as p/(1-p) where p is the proportion
311 of cells, followed by ANOVA and Tukey's multiple comparison test between the four
312 groups. For clarity, the data are presented as percentage of cells.

Correlation Analysis. We correlated each antibody to its corresponding gene(s) using Spearman rank correlation and significance (R package). For each combination of gene-antibody, we discarded cells that had values below the corresponding threshold for that antibody as well as cells with zero counts for that gene. After this filter, any gene-antibody combination that had 10 cells or less was deemed insignificant. Finally, all non-significant (p-value > 0.05) were designated a nominal value of zero as the Spearman rank correlation coefficient and we selected only those genes or antibodies that had at least one correlation whose coefficient >= 0.25 or whose coefficient <= -0.25. All significant non-negative correlations are reported in **Table S11**.

Results

Identification of main cell types based on antibody expression. To identify the major known cell types, we used biaxial gating on CD3, CD19, CD4, CD8, CD14, CD16, and CD56. This approach defines (**Figure 1A-E**):

- B cells: CD19+ CD3-
- T cells: CD19- CD3+
- CD4 T cells: CD4+ CD8- T cells
- CD8 T cells: CD8+ CD4- T cells
- Monocytes (M): CD19-CD3-CD56-
- Classical (CM): CD14+CD16-
- Intermediate (INT): CD14+CD16+
- Nonclassical (NCM): CD14-CD16+CD56-
- NK cells (NK): CD4- CD56+ CD14- CD20- CD123- CD206-

CD3 and CD19 expression are mutually exclusive and specific for T and B cells, respectively. As is standard in the NK cell field,[31] the CD16- immature NK cells were gated based on higher levels of CD56 as shown in **Figure 1E**. The mature NK cells were CD19-CD3-CD16+CD56+. One CD16+CD56- cluster was also identified as NK cells. This resulted in 2,919 B cells, 11,045 CD4 T cells, 12,843 CD8 T cells, 5,145 CM, 1009 INT, 475 NCM and 1,843 NK cells. Each of these major cell types was then re-clustered separately, using Seurat [49] to construct UMAPs with Louvain clustering (**Figure 1F**). Like in flow or mass cytometry, we clustered on antibody staining only. This “preserves” the transcriptomes for investigations into disease- and treatment-related changes. Using this approach, we identified 16 CD4 T cell subsets, 14 CD8 T cell subsets, 8 CM subsets,

348 3 NCM subsets, 5 INT subsets, 6 B cell subsets and 6 NK cell subsets (**Figure 1F**). The
349 corresponding feature maps are shown in **Figure S2**. Trying to find these cell types based
350 on transcriptomes was unsatisfactory (**Figure S3**).
351
352 **Cell subsets calling using 40 surface markers.** Next, we constructed heat maps for all
353 antibodies that were significantly differentially expressed in at least one subset (**Figure**
354 **2**). This information allowed us to call all CD4 and CD8 T cell subsets in accordance
355 with published immunology work. Among CD4 T cells, CD2 was expressed in almost all
356 cells, as expected. The high affinity IL2 receptor IL2RA (CD25) was expressed in about
357 a third of the CD4 T cells and was strikingly high in cluster 13, which was also low for
358 IL7 receptor (CD127), defining cluster 13 as regulatory T cells (Tregs). CD45RA and RO
359 were mutually exclusive, separating naive and antigen-experienced CD4 T cells. CXCR3
360 (CD183) identifies T-helper-1 (Th1) cells and was highly expressed in clusters 5, 14, and
361 16. Cluster 14 co-expressed CXCR5 (CD185) with CXCR3. Cluster 7 expressed CXCR5
362 as the only chemokine receptor, suggesting it may contain follicular helper (TFH) T cells.
363 Based on surface marker information, all CD4 T cell clusters were called (**Figure 2A**).
364 All CD8 T cells expressed CD2. Cluster 3 exclusively expressed CD9 and CD36,
365 identifying these cells as NK-like CD8 cells. Clusters 7 and 13 were identified as NK-like
366 T cells with a CD45RA+ terminally differentiated memory (EMRA) phenotype (**Figure**
367 **2B**).
368
369 Among monocytes, we were able to call 5 of the 8 classical monocyte subsets based on
370 published data.[11] All CM were CD11b+ (**Figure 2C**). There were gradients of CD9,

371 CD69, CD137, CD142 (tissue factor), and CD163 (hemoglobin-haptoglobin receptor)
372 expression. The scavenger receptor CD36, the antigen presentation co-receptor CD86 and
373 the chemokine receptor CCR2 were expressed in all classical monocytes. Based on these
374 markers, 5 of the 8 CM subsets were called (**Figure 2C**) and related to subsets described
375 by mass cytometry. INT CD14+CD16+ monocytes have been considered pro-
376 inflammatory and are known to be increased in people with HIV[12] and with CVD.[41,
377 51] All INT highly expressed the inflammation-induced costimulatory molecule CD86
378 (**Figure 2C**). Cluster INT3 highly expressed CD142 (tissue factor), which has previously
379 been implicated in people living with HIV.[45] Since INT subsets have not been
380 described before, we propose a provisional naming suggestion (**Figure 2C**). NCM
381 formed 3 clusters (**Figure 2C**). Strikingly, expression of CD9 and CD36 was limited to
382 cluster 3, suggesting that this cluster corresponds to the previously described
383 CD9+CD36+ NCM.[11] CD11c, CD74, CD86, and CD141 were expressed in all NCMs
384 (**Figure 2C**).
385
386 We were able to call all 6 B cell subsets. As expected, CD20 and CD74 (HLA-DR) were
387 expressed in all B cells (**Figure 2D**). CD27, IgM and IgD are used to identify naïve B
388 cells (CD27-IgM+IgD+). Clusters 1, 3, and 4 were negative for CD27 with high
389 transcript expression for IgM and IgD, consistent with naïve B cells. Clusters 3 and 4
390 expressed CCR6, a subset found in HIV+ subjects.[32] B cell cluster 2 expressed CD25,
391 which is a known marker for B cell proliferation and exhaustion, and CD27, identifying
392 cluster 2 as a likely activated memory B cell. Cluster 5 had high CD11c levels, known to
393 increase in HIV-infected subjects,[19] and expressed some CXCR3 and CCR6, but was

394 CD27low. These features together with moderate expression of CD22 transcript suggest
395 that cluster 5 may contain CD11c+ pathologic B cells. (**Figure 2D**). Most NK cells were
396 mature ($CD56^{dim}$/CD16+), as expected (**Figure 2E**). Cluster 3 also contained immature
397 ($CD56^{bright}$CD16-) NK cells. The $CD56^{low}$CD16- cells (clusters 4 and 5) expressed CD2
398 and CD45RA. Cluster 5 was CD56-$CD16^{high}$, an NK cell subset known to be elevated in
399 chronic HIV infection.[17] Taken together, this demonstrates the power of combined
400 antibody and transcriptome sequencing.
401
402 **Changes in PBMC subset abundance on disease or treatment.** Based on this data, it is
403 possible to address shifts in cell proportion based on disease or treatment. We found
404 significant differences in cell proportions in 3 intermediate monocyte subsets, one CD8 T
405 cell, one B cell and one NK cell subset (**Figure 3**). Strikingly, three subsets of
406 intermediate monocytes (**Figure 3A**) showed significantly different abundances. INT2
407 (IL7R-) and INT3 (TF^{hi}) were significantly elevated in WIHS participants living with
408 HIV and drastically reduced in those that also had subclinical CVD. INT4 had an
409 opposite pattern: These CD163- cells were rare in WIHS participants living with HIV,
410 but more abundant in those that also had subclinical CVD. Among B cells, activated
411 memory B2 cells (**Figure 3C**) were severely lower in all WIHS participants with HIV
412 with or without subclinical CVD.
413
414 **Differential gene expression in each of the clusters.** Since the transcriptomic
415 information was not used for UMAPs and clustering shown in **Figures 1** and **2**, we were
416 able to compare the gene expression patterns in each cell subset within the same cell type.

417 We filtered for genes that were significantly differentially expressed in at least one of the
418 subsets (**Figure 4, Data S1**). This analysis revealed gene signatures for most subsets.
419 Such gene signatures can then be used to determine the presence of each subset in bulk
420 transcriptomes, and to determine their proportions using Cibersort.[34] As an example,
421 we applied the classical monocyte transcriptomes (8 subsets) to bulk transcriptomes of
422 sorted classical monocytes from 92 subjects.[7] We found 1 of the CM subsets in all
423 subjects and others at varying proportions (**Figure 5**).
424
425 **Transcriptomes shift with HIV, CVD and cholesterol control.** Transcriptomes may
426 also change with disease state. To test this possibility, we constructed volcano plots (log
427 fold change on the x axis versus -log p on the y axis (**Figure 6, Data S2**). To test for
428 changes with cardiovascular disease (CVD), we plotted genes significantly different
429 between subjects with and without CVD. All these subjects were HIV+. Many genes in
430 CD4 and CD8 T cell subsets showed significant differences. Some genes in monocyte
431 and B cell subsets showed significant differences. To test whether our method could
432 detect effects of treatment, we interrogated transcriptomes of subjects with CVD that
433 received CRT. Again, many genes in T cell and monocyte subsets and some in B cell
434 subsets showed significant differences subjects (**Figure 6**). In CD4T1, 2, and 8, IL-32
435 was highly significantly increased by CVD, but not in CVD+ women on CRT (**Figure 6**).
436 IL-32 is an inflammatory cytokine that is known to be important in CVD.[6, 21] In
437 CD4T2, L-selectin (SELL), PSGL-1 (SELPLG), and CCR7 were also highly significantly
438 increased in WIHS participants with HIV and CVD. In addition to SELL and SELPLG,
439 CD4T8 showed strong upregulation of TNFSF10 (TRAIL). In CD8T1 and 2, IL32 was

440 high in women with CVD, but less so in women receiving CRT. Other genes highly
441 induced by CVD in CD8T1 included CD52, TRAC and HOPX. Several killer cell lectin
442 receptors (KLRC4, KLRD1, KLRG1 and KLRK1) were also significantly upregulated in
443 CVD. In CD8T3, CD52, CCL5, IL32 and CD160 were all significantly higher in CVD+
444 participants. CCL5 encodes the chemokine RANTES, known to be important in
445 atherosclerosis.[54] In CD8T4, CVD was associated with significantly increased IL32,
446 TRAC, HOPX, CCL5 and the killer lectin receptors KLRK1, KLRC4, KLRD1.
447
448 In CM1, CVD was associated with significantly increased CCL4, SLC2A3, SOD2, and
449 SELPLG. CRT was associated with lower expression of these genes. In CM2, TNF,
450 DUSP1, and 2 were highly associated with CVD (**Figure 6**), as were TNFSF10 (TRAIL),
451 TNFSF13 (APRIL), and TNFSF13B (BAFF), important B cell regulators. In addition to
452 CCL3, CCL4, and DUSP2, IL1B, known to be highly relevant in atherosclerosis, was
453 highly upregulated in CM3 of HIV+CVD+ participants. The Toll-like receptor TLR2,
454 which is known to be involved in atherosclerosis, was upregulated by CVD in CM3. In
455 INT3, CCL3, CCL4, TNF, IL1B, and DUSP2 were associated CVD in the participants
456 that did not receive CRT.
457
458
459
460
461
462
463
464
465
466
467
468

469 **Discussion**

470

471 In immunology, surface markers are widely used to define and distinguish cell types.[5,
472 47, 57] Flow cytometry is the discipline-defining method of immunology.[40] Similar to
473 flow cytometry, in CyTOF, single-cell suspensions are stained with antibody panels to
474 detect cellular antigens. Unlike CyTOF, scRNA-seq allows the detection of single-cell
475 transcriptomes. Since the correlation between cell surface protein and mRNA expression
476 is weak in immune cells[25], the transcriptome provides a valuable additional dimension.
477 scRNA-Seq without surface phenotype information has led to much frustration in the
478 field, because the expression of many genes encoding well-known surface markers
479 remains undetected in scRNA-Seq.[27, 57] It is still difficult to call cell types based on
480 gene expression data alone, which emphasizes the need for cell surface phenotypes in
481 addition to transcriptomes. Here, we correlated gene expression with cell surface
482 expression for 41 pairs of genes and proteins. CD74 surface expression was well
483 correlated with the expression of both the *CD74* and the *HLA-DRA* genes. CD4 and
484 CD16 surface and gene expression were reasonably well correlated across all cell types.
485 A few other genes including *CD14*, *CD16*, IL-3 receptor (*CD123*), and *CD27* were
486 somewhat correlated with the surface expression of their proteins in some cell types. For
487 most markers, we confirm weak correlations,[25] which illustrates the value of
488 monitoring cell surface phenotype in scRNA-Seq.

489

490 PBMCs can be analyzed without mechanical or enzymatic dissociation, which are known
491 to alter cell surface markers and transcriptomes.[56] PBMC are attractive for single cell
492 RNA sequencing (scRNA-Seq) studies, because they are available in many clinical

493 studies of specific populations with defined diseases and outcomes. The participants
494 sampled for the present study were part of a sub-study nested within the WIHS,[13, 16,
495 18] which provided detailed information on subclinical atherosclerosis. Participants
496 underwent high-resolution B-mode carotid artery ultrasound to image six locations in the
497 right carotid artery.[16] Although our study is not definitive, it is suggestive of significant
498 changes in cell proportions and transcriptomes in subjects with cardiovascular disease.
499
500 scRNA-Seq has been applied to human PBMCs in diseases including cancers,[3, 60–62]
501 inflammatory bowel disease,[30, 53] and autoimmune disease,[22, 38] as well as
502 atherosclerosis.[9, 58] The foundational paper for the 10x Genomics drop-Seq
503 method[63] demonstrated the feasibility of using scRNA-Seq on PBMCs. Other studies
504 reported scPBMC transcriptomes in colorectal cancer,[61] γδ T cells,[36] liver
505 cancer,[62] in vitro salmonella infection,[2] and memory T cells.[28] Only two
506 publications reported single cell transcriptomes from patients with atherosclerosis
507 (carotid endarterectomy specimens and matched PBMCs).[9, 58] 1,652 PBMCs from one
508 patient were analyzed by 10x Genomics 3' and cellular indexing of transcriptional
509 epitope sequencing (CITE-Seq),[35, 48] using a panel of 21 antibodies. No healthy
510 control PBMCs were studied. ScRNA-Seq revealed that the process of smooth muscle
511 cell phenotypic modulation *in vivo* can be altered by the expression of *Tcf21*, a gene
512 causally associated with reduced risk of coronary artery disease. The loss of *Tcf21* results
513 in fewer fibromyocytes in the lesions and the protective fibrous cap.[58] A recent study
514 reported the effect of HIV infection on PBMC transcriptomes,[20] focusing on acute HIV
515 infection (before antiretroviral therapy started) and reporting PBMC transcriptomes in

516 four patients at 8 defined time points (average of 1,976 PBMC transcriptomes per
517 participant and condition). No scRNA-Seq or CITE-Seq studies of PBMCs of people
518 living with chronic HIV infection have been reported. No single cell studies of the
519 interaction between HIV and CVD are available.
520
521 Six clusters showed significantly different abundance of cells in the four groups of
522 participants, three of them intermediate monocyte subsets, which underscores the
523 extraordinary importance of this cell type in chronic HIV infection[14, 29] and CVD.[10,
524 24] Intermediate monocyte numbers have previously been found increased in non-HIV
525 subjects with peripheral artery occlusive disease[55] and significantly predicted
526 cardiovascular events.[15, 41, 42] Cells in INT1, the largest cluster, shared CD11b,
527 CD11c, CD9, CD36, CD38, CD56, CD69, CD83, IL-3RA, IL6R, CD137, CD141,
528 CD142 (tissue factor), CXCR4 and CD74 (HLA-DR) with other intermediate monocytes.
529 We found no single positive marker that was specific for INT1 and thus refrained from
530 naming this cluster. We found the INT2 and INT3 increased in women living with HIV.
531 Both express tissue factor (CD142). Tissue factor expression on monocytes has
532 previously been shown to be increased in HIV+ subjects.[45] Intermediate monocytes are
533 considered pro-atherogenic,[44] and tissue factor expression provides a plausible reason
534 for this. We found that in INT2 and INT3, the inflammatory chemokines CCL3 and
535 CCL4 and the known pro-atherogenic cytokine IL-1β were significantly upregulated in
536 participants with CVD, but not in those receiving CRT. INT4 uniquely lack expression of
537 CD163, the receptor for hemoglobin-haptoglobin complexes. Thus, we call INT4 CD163-

538 intermediate monocytes. INT5, called (CTLA4+CXCR3hi) uniquely expressed CTLA4
539 (CD152) and highly expressed CXCR3.
540
541 In CD4T cells clusters 1, 2 and 8, IL-32 was highly significantly increased by CVD,
542 which was reversed by cholesterol lowering in CD4T1 and 2 (**Figure 6**). IL-32 is a 27
543 kDa cytokine expressed in T cells and monocytes that is secreted after apoptosis.[33] It is
544 an inflammatory cytokine that drives IL-1β, clinically important in CVD,[39] TNF, IL-6
545 and IL-8 expression.[6, 21, 33] IL-32 activates the leukocyte surface protease PR3, which
546 in turn triggers the G-protein coupled receptor PAR2[33] and is known to be important in
547 viral infections.[23, 33, 37, 46] We found IL-32 highly expressed in most T and NK cell
548 clusters. Since IL-32 appears to be CVD-specific, we advocate for future prospective
549 studies in larger cohorts to determine whether IL32 mRNA is a useful biomarker.
550
551 Our discovery study will encourage prospective epidemiological studies to address which
552 PBMC subset and transcriptomes are best suited as clinical biomarkers to stratify risk and
553 guide treatment in subjects with coronary or peripheral artery disease. The current
554 findings also present some limitations. They need to be extended to men (the current data
555 is based on women) and other races and ethnicities (the current data is based on mostly
556 African American and Hispanic women). Studies of CVD in non-smokers are also
557 needed (the current data is based on smokers), and the age range needs to be broadened.
558
559 In conclusion, we demonstrate the utility of scRNA-Seq with cell surface phenotype
560 assessment in the same cells. The identification of 58 distinct clusters of CD4 and CD8 T

561 cells, B cells, NK cells and monocytes helps to gain a deeper understanding of PBMCs, a
562 rich and readily accessible source of biological and clinical information. The discovery of
563 subsets of intermediate monocytes calls for identifying such subsets in model organisms
564 to test their function *in vivo*.
565
566
567
568
569
570
571
572
573
574
575

576 **References**
577
578
579 1. Bekele Y, Lakshmikanth T, Chen Y, Mikes J, Nasi A, Petkov S, Hejdeman B, Brodin P,
580 Chiodi F (2019) Mass cytometry identifies distinct CD4+ T cell clusters distinguishing
581 HIV-1-infected patients according to antiretroviral therapy initiation. JCI insight 4:
582 e125442. doi: 10.1172/jci.insight.125442.
583 2. Bossel Ben-Moshe N, Hen-Avivi S, Levitin N, Yehezkel D, Oosting M, Joosten LAB,
584 Netea MG, Avraham R (2019) Predicting bacterial infection outcomes using single cell
585 RNA-sequencing analysis of human immune cells. Nat Commun 10:3266. doi:
586 10.1038/s41467-019-11257-y.
587 3. Brown CC, Gudjonson H, Pritykin Y, Deep D, Lavallée V-P, Mendoza A, Fromme R,
588 Mazutis L, Ariyan C, Leslie C, Pe'er D, Rudensky AY (2019) Transcriptional Basis of
589 Mouse and Human Dendritic Cell Heterogeneity. Cell 179:846-863. doi:
590 10.1016/j.cell.2019.09.035.
591 4. Buttarello M, Plebani M (2008) Automated blood cell counts: state of the art. Am J Clin
592 Pathol 130:104–116. doi: 10.1309/EK3C7CTDKNVPXVTN.
593 5. Cole JE, Park I, Ahern DJ, Kassiteridi C, Danso Abeam D, Goddard ME, Green P, Maffia
594 P, Monaco C (2018) Immune cell census in murine atherosclerosis: cytometry by time of
595 flight illuminates vascular myeloid cell diversity. Cardiovasc Res 114:1360–1371. doi:
596 10.1093/cvr/cvy109.
597 6. Damen MSMA, Popa CD, Netea MG, Dinarello CA, Joosten LAB (2017) Interleukin-32
598 in chronic inflammatory conditions is associated with a higher risk of cardiovascular
599 diseases. Atherosclerosis 264:83–91. doi: 10.1016/j.atherosclerosis.2017.07.005.
600 7. Ehinger E, Ghosheh Y, Bala PA, Durant CP, Lin J, Hanna DB, Mueller K, Baas L, Qi Q,
601 Wang T, Buscher K, Liu Y, Anastos K, Lazar JM, Mack WJ, Tien PC, Cohen MH,
602 Ofotokun I, Gange S, Heath SL, Hodis HN, Tracy RP, Landay AL, Kaplan RC, Ley K
603 (2020) Classical Monocyte Transcriptomes Reveal Significant Anti- Inflammatory Statin
604 Effect in Women with Chronic HIV. Cardiovasc Res 117:1166–1177. doi:
605 10.1093/cvr/cvaa188.
606 8. Fan HC, Fu GK, Fodor SPA (2015) Expression profiling. Combinatorial labeling of single
607 cells for gene expression cytometry. Science 347:1258367. doi: 10.1126/science.1258367.
608 9. Fernandez DM, Rahman AH, Fernandez NF, Chudnovskiy A, Amir ED, Amadori L, Khan
609 NS, Wong CK, Shamailova R, Hill CA, Wang Z, Remark R, Li JR, Pina C, Faries C,
610 Awad AJ, Moss N, Bjorkegren JLM, Kim-schulze S, Gnjatic S, Ma A, Mocco J, Faries P,
611 Merad M, Giannarelli C (2019) Single-cell immune landscape of human atherosclerotic
612 plaques. Nat Med 25:1576–1588. doi: 10.1038/s41591-019-0590-4.
613 10. Funderburg NT, Zidar DA, Shive C, Lioi A, Mudd J, Musselwhite LW, Simon DI, Costa
614 MA, Rodriguez B, Sieg SF, Lederman MM (2012) Shared monocyte subset phenotypes in
615 HIV-1 infection and in uninfected subjects with acute coronary syndrome. Blood
616 120:4599–4608. doi: 10.1182/blood-2012-05-433946.
617 11. Hamers AAJ, Dinh HQ, Thomas GD, Marcovecchio P, Blatchley A, Nakao CS, Kim C,
618 McSkimming C, Taylor AM, Nguyen AT, McNamara CA, Hedrick CC (2019) Human
619 Monocyte Heterogeneity as Revealed by High-Dimensional Mass Cytometry. Arterioscler
620 Thromb Vasc Biol 39:25–36. doi: 10.1161/ATVBAHA.118.311022.
621

622 12. Han J, Wang B, Han N, Zhao Y, Song C, Feng X, Mao Y, Zhang F, Zhao H, Zeng H
623 (2009) CD14(high)CD16(+) rather than CD14(low)CD16(+) monocytes correlate with
624 disease progression in chronic HIV-infected patients. J Acquir Immune Defic Syndr
625 52:553–559. doi: 10.1097/qai.0b013e3181c1d4fe.
626 13. Hanna DB, Post WS, Deal JA, Hodis HN, Jacobson LP, Mack WJ, Anastos K, Gange SJ,
627 Landay AL, Lazar JM, Palella FJ, Tien PC, Witt MD, Xue X, Young MA, Kaplan RC,
628 Kingsley LA (2015) HIV Infection Is Associated With Progression of Subclinical Carotid
629 Atherosclerosis. Clin Infect Dis 61:640–650. doi: 10.1093/cid/civ325.
630 14. Hearps AC, Maisa A, Cheng W-J, Angelovich TA, Lichtfuss GF, Palmer CS, Landay AL,
631 Jaworowski A, Crowe SM (2012) HIV infection induces age-related changes to
632 monocytes and innate immune activation in young men that persist despite combination
633 antiretroviral therapy. AIDS 26:843–853. doi: 10.1097/QAD.0b013e328351f756.
634 15. Heine GH, Ulrich C, Seibert E, Seiler S, Marell J, Reichart B, Krause M, Schlitt A, Köhler
635 H, Girndt M (2008) CD14(++)CD16+ monocytes but not total monocyte numbers predict
636 cardiovascular events in dialysis patients. Kidney Int 73:622–629. doi:
637 10.1038/sj.ki.5002744.
638 16. Hodis HN, Mack WJ, Lobo RA, Shoupe D, Sevanian A, Mahrer PR, Selzer RH, Liu Cr
639 CR, Liu Ch CH, Azen SP (2001) Estrogen in the prevention of atherosclerosis. A
640 randomized, double-blind, placebo-controlled trial. Ann Intern Med 135:939–953. doi:
641 10.7326/0003-4819-135-11-200112040-00005.
642 17. Hong HS, Eberhard JM, Keudel P, Bollmann BA, Ahmad F, Ballmaier M, Bhatnagar N,
643 Zielinska-Skowronek M, Schmidt RE, Meyer-Olson D (2010) Phenotypically and
644 functionally distinct subsets contribute to the expansion of CD56-/CD16+ natural killer
645 cells in HIV infection. AIDS 24:1823–1834. doi: 10.1097/QAD.0b013e32833b556f.
646 18. Kaplan RC, Kingsley LA, Gange SJ, Benning L, Jacobson LP, Lazar J, Anastos K, Tien
647 PC, Sharrett AR, Hodis HN (2008) Low CD4+ T-cell count as a major atherosclerosis risk
648 factor in HIV-infected women and men. AIDS 22:1615–1624. doi:
649 10.1097/QAD.0b013e328300581d.
650 19. Karnell JL, Kumar V, Wang J, Wang S, Voynova E, Ettinger R (2017) Role of CD11c(+)
651 T-bet(+) B cells in human health and disease. Cell Immunol 321:40–45. doi:
652 10.1016/j.cellimm.2017.05.008.
653 20. Kazer SW, Aicher TP, Muema DM, Carroll SL, Ordovas-Montanes J, Miao VN, Tu AA,
654 Ziegler CGK, Nyquist SK, Wong EB, Ismail N, Dong M, Moodley A, Berger B, Love JC,
655 Dong KL, Leslie A, Ndhlovu ZM, Ndung'u T, Walker BD, Shalek AK (2020) Integrated
656 single-cell analysis of multicellular immune dynamics during hyperacute HIV-1 infection.
657 Nat Med 26:511–518. doi: 10.1038/s41591-020-0799-2.
658 21. Kim S-H, Han S-Y, Azam T, Yoon D-Y, Dinarello CA (2005) Interleukin-32: a cytokine
659 and inducer of TNFalpha. Immunity 22:131–142. doi: 10.1016/j.immuni.2004.12.003.
660 22. Kotliarov Y, Sparks R, Martins AJ, Mule MP, Lu Y, Goswami M, Kardava L, Banchereau
661 R, Pascual V, Biancotto A, Chen J, Schwartzberg PL, Bansal N, Liu CC, Cheung F, Moir
662 S, Tsang JS (2020) Broad immune activation underlies shared set point signatures for
663 vaccine responsiveness in healthy individuals and disease activity in patients with lupus.
664 Nat Med 26:618–629. doi: 10.1038/s41591-020-0769-8.
665 23. Li W, Sun W, Liu L, Yang F, Li Y, Chen Y, Fang J, Zhang W, Wu J, Zhu Y (2010) IL-32:
666 a host proinflammatory factor against influenza viral replication is upregulated by aberrant
667 epigenetic modifications during influenza A virus infection. J Immunol 185:5056–5065.

668 doi: 10.4049/jimmunol.0902667.
669 24. Liang H, Xie Z, Shen T (2017) Monocyte activation and cardiovascular disease in HIV
670 infection. Cell Mol Immunol 14:960–962. doi: 10.1038/cmi.2017.109.
671 25. Liu Y, Beyer A, Aebersold R (2016) On the Dependency of Cellular Protein Levels on
672 mRNA Abundance. Cell 165:535–550. doi: 10.1016/j.cell.2016.03.014.
673 26. Londino JD, Gulick DL, Lear TB, Suber TL, Weathington NM, Masa LS, Chen BB,
674 Mallampalli RK (2017) Post-translational modification of the interferon-gamma receptor
675 alters its stability and signaling. Biochem J 474:3543–3557. doi: 10.1042/BCJ20170548.
676 27. Lundberg E, Fagerberg L, Klevebring D, Matic I, Geiger T, Cox J, Algenas C, Lundeberg
677 J, Mann M, Uhlen M (2010) Defining the transcriptome and proteome in three
678 functionally different human cell lines. Mol Syst Biol 6:450. doi: 10.1038/msb.2010.106.
679 28. Mair F, Erickson JR, Voillet V, Simoni Y, Bi T, Tyznik AJ, Martin J, Gottardo R, Newell
680 EW, Prlic M (2020) A Targeted Multi-omic Analysis Approach Measures Protein
681 Expression and Low-Abundance Transcripts on the Single-Cell Level. Cell Rep
682 31:107499. doi: 10.1016/j.celrep.2020.03.063.
683 29. Martin GE, Gouillou M, Hearps AC, Angelovich TA, Cheng AC, Lynch F, Cheng W-J,
684 Paukovics G, Palmer CS, Novak RM, Jaworowski A, Landay AL, Crowe SM (2013) Age-
685 associated changes in monocyte and innate immune activation markers occur more rapidly
686 in HIV infected women. PLoS One 8:e55279. doi: 10.1371/journal.pone.0055279.
687 30. Martin JC, Chang C, Boschetti G, Ungaro R, Giri M, Grout JA, Gettler K, Chuang L-S,
688 Nayar S, Greenstein AJ, Dubinsky M, Walker L, Leader A, Fine JS, Whitehurst CE,
689 Mbow ML, Kugathasan S, Denson LA, Hyams JS, Friedman JR, Desai PT, Ko HM,
690 Laface I, Akturk G, Schadt EE, Salmon H, Gnjatic S, Rahman AH, Merad M, Cho JH,
691 Kenigsberg E (2019) Single-Cell Analysis of Crohn's Disease Lesions Identifies a
692 Pathogenic Cellular Module Associated with Resistance to Anti-TNF Therapy. Cell
693 178:1493-1508. doi: 10.1016/j.cell.2019.08.008.
694 31. Milush JM, Long BR, Snyder-Cappione JE, Cappione AJ 3rd, York VA, Ndhlovu LC,
695 Lanier LL, Michaelsson J, Nixon DF (2009) Functionally distinct subsets of human NK
696 cells and monocyte/DC-like cells identified by coexpression of CD56, CD7, and CD4.
697 Blood 114:4823–4831. doi: 10.1182/blood-2009-04-216374.
698 32. Moir S, Fauci AS (2017) B-cell responses to HIV infection. Immunol Rev 275:33-48. doi:
699 10.1111/imr.12502.
700 33. Nakayama M, Niki Y, Kawasaki T, Takeda Y, Ikegami H, Toyama Y, Miyamoto T (2013)
701 IL-32-PAR2 axis is an innate immunity sensor providing alternative signaling for LPS-
702 TRIF axis. Sci Rep 3:2960. doi: 10.1038/srep02960.
703 34. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M,
704 Alizadeh AA (2015) Robust enumeration of cell subsets from tissue expression profiles.
705 Nat Methods 12:453–457. doi: 10.1038/nmeth.3337.
706 35. Peterson VM, Zhang KX, Kumar N, Wong J, Li L, Wilson DC, Moore R, McClanahan
707 TK, Sadekova S, Klappenbach JA (2017) Multiplexed quantification of proteins and
708 transcripts in single cells. Nat Biotechnol 35:936–939. doi: 10.1038/nbt.3973.
709 36. Pizzolato G, Kaminski H, Tosolini M, Franchini D-M, Pont F, Martins F, Valle C,
710 Labourdette D, Cadot S, Quillet-Mary A, Poupot M, Laurent C, Ysebaert L, Meraviglia S,
711 Dieli F, Merville P, Milpied P, Dechanet-Merville J, Fournie J-J (2019) Single-cell RNA
712 sequencing unveils the shared and the distinct cytotoxic hallmarks of human TCRVdelta1
713 and TCRVdelta2 gammadelta T lymphocytes. Proc Natl Acad Sci U S A 116:11906–

714 11915. doi: 10.1073/pnas.1818488116.
715 37. Rasool ST, Tang H, Wu J, Li W, Mukhtar MM, Zhang J, Mu Y, Xing HX, Wu J, Zhu Y
716 (2008) Increased level of IL-32 during human immunodeficiency virus infection
717 suppresses HIV replication. Immunol Lett 117:161–167. doi: 10.1016/j.imlet.2008.01.007.
718 38. Reyes M, Vickers D, Billman K, Eisenhaure T, Hoover P, Browne EP, Rao DA, Hacohen
719 N, Blainey PC (2019) Multiplexed enrichment and genomic profiling of peripheral blood
720 cells reveal subset-specific immune signatures. Sci Adv 5:eaau9223. doi:
721 10.1126/sciadv.aau9223.
722 39. Ridker PM, Everett BM, Thuren T, MacFadyen JG, Chang WH, Ballantyne C, Fonseca F,
723 Nicolau J, Koenig W, Anker SD, Kastelein JJP, Cornel JH, Pais P, Pella D, Genest J,
724 Cifkova R, Lorenzatti A, Forster T, Kobalava Z, Vida-Simiti L, Flather M, Shimokawa H,
725 Ogawa H, Dellborg M, Rossi PRF, Troquay RPT, Libby P, Glynn RJ (2017)
726 Antiinflammatory Therapy with Canakinumab for Atherosclerotic Disease. N Engl J Med
727 377:1119–1131. doi: 10.1056/NEJMoa1707914.
728 40. Robinson JP, Roederer M (2015) HISTORY OF SCIENCE. Flow cytometry strikes gold.
729 Science 350:739–740. doi: 10.1126/science.aad6770.
730 41. Rogacev KS, Cremers B, Zawada AM, Seiler S, Binder N, Ege P, Große-Dunker G,
731 Heisel I, Hornof F, Jeken J, Rebling NM, Ulrich C, Scheller B, Böhm M, Fliser D, Heine
732 GH (2012) CD14++CD16+ monocytes independently predict cardiovascular events: a
733 cohort study of 951 patients referred for elective coronary angiography. J Am Coll Cardiol
734 60:1512–1520. doi: 10.1016/j.jacc.2012.07.019.
735 42. Rogacev KS, Seiler S, Zawada AM, Reichart B, Herath E, Roth D, Ulrich C, Fliser D,
736 Heine GH (2011) CD14++CD16+ monocytes and cardiovascular outcome in patients with
737 chronic kidney disease. Eur Heart J 32:84–92. doi: 10.1093/eurheartj/ehq371.
738 43. De Rosa SC, Brenchley JM, Roederer M (2003) Beyond six colors: A new era in flow
739 cytometry. Nat Med 9:112–117. doi: 10.1038/nm0103-112.
740 44. SahBandar IN, Ndhlovu LC, Saiki K, Kohorn LB, Peterson MM, D'Antoni ML,
741 Shiramizu B, Shikuma CM, Chow DC (2020) Relationship between Circulating
742 Inflammatory Monocytes and Cardiovascular Disease Measures of Carotid Intimal
743 Thickness. J Atheroscler Thromb 27:441–448. doi: 10.5551/jat.49791.
744 45. Schechter ME, Andrade BB, He T, Richter GH, Tosh KW, Policicchio BB, Singh A,
745 Raehtz KD, Sheikh V, Ma D, Brocca-Cofano E, Apetrei C, Tracy R, Ribeiro RM, Sher A,
746 Francischetti IMB, Pandrea I, Sereti I (2017) Inflammatory monocytes expressing tissue
747 factor drive SIV and HIV coagulopathy. Sci Transl Med 9:eaam5441. doi:
748 10.1126/scitranslmed.aam5441.
749 46. Smith AJ, Toledo CM, Wietgrefe SW, Duan L, Schacker TW, Reilly CS, Haase AT
750 (2011) The immunosuppressive role of IL-32 in lymphatic tissue during HIV-1 infection.
751 J Immunol 186:6576–6584. doi: 10.4049/jimmunol.1100277.
752 47. Spitzer MH, Nolan GP (2016) Mass Cytometry: Single Cells, Many Features. Cell
753 165:780–791. doi: 10.1016/j.cell.2016.04.019.
754 48. Stoeckius M, Hafemeister C, Stephenson W, Houck-Loomis B, Chattopadhyay PK,
755 Swerdlow H, Satija R, Smibert P (2017) Simultaneous epitope and transcriptome
756 measurement in single cells. Nat Methods 14:865–868. doi: 10.1038/nmeth.4380.
757 49. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM 3rd, Hao Y,
758 Stoeckius M, Smibert P, Satija R (2019) Comprehensive Integration of Single-Cell Data.
759 Cell 177:1888-1902.e21. doi: 10.1016/j.cell.2019.05.031.

760 50. Subelj L, Bajec M (2011) Unfolding communities in large complex networks: combining
761 defensive and offensive label propagation for core extraction. Phys Rev E Stat Nonlin Soft
762 Matter Phys 83:36103. doi: 10.1103/PhysRevE.83.036103.
763 51. Tapp LD, Shantsila E, Wrigley BJ, Pamukcu B, Lip GYH (2012) The CD14++CD16+
764 monocyte subset and monocyte-platelet interactions in patients with ST-elevation
765 myocardial infarction. J Thromb Haemost 10:1231–1241. doi: 10.1111/j.1538-
766 7836.2011.04603.x.
767 52. Tracy RP, Doyle MF, Olson NC, Huber SA, Jenny NS, Sallam R, Psaty BM, Kronmal RA
768 (2013) T-helper type 1 bias in healthy people is associated with cytomegalovirus serology
769 and atherosclerosis: the Multi-Ethnic Study of Atherosclerosis. J Am Heart Assoc
770 2:e000117. doi: 10.1161/JAHA.113.000117.
771 53. Uniken Venema WT, Voskuil MD, Vila AV, van der Vries G, Jansen BH, Jabri B, Faber
772 KN, Dijkstra G, Xavier RJ, Wijmenga C, Graham DB, Weersma RK, Festen EA (2019)
773 Single-Cell RNA Sequencing of Blood and Ileal T Cells From Patients With Crohn's
774 Disease Reveals Tissue-Specific Characteristics and Drug Targets. Gastroenterology
775 156:812-815. doi: 10.1053/j.gastro.2018.10.046.
776 54. Virani SS, Nambi V, Hoogeveen R, Wasserman BA, Coresh J, Gonzalez F 2nd,
777 Chambless LE, Mosley TH, Boerwinkle E, Ballantyne CM (2011) Relationship between
778 circulating levels of RANTES (regulated on activation, normal T-cell expressed, and
779 secreted) and carotid plaque characteristics: the Atherosclerosis Risk in Communities
780 (ARIC) Carotid MRI Study. Eur Heart J 32:459–468. doi: 10.1093/eurheartj/ehq367.
781 55. Wildgruber M, Aschenbrenner T, Wendorff H, Czubba M, Glinzer A, Haller B,
782 Schiemann M, Zimmermann A, Berger H, Eckstein H-H, Meier R, Wohlgemuth WA,
783 Libby P, Zernecke A (2016) The "Intermediate" CD14(++)CD16(+) monocyte subset
784 increases in severe peripheral artery disease in humans. Sci Rep 6:39483. doi:
785 10.1038/srep39483.
786 56. Williams JW, Winkels H, Durant CP, Zaitsev K, Ghosheh Y, Ley K (2020) Single Cell
787 RNA Sequencing in Atherosclerosis Research. Circ Res 126:1112–1126. doi:
788 10.1161/CIRCRESAHA.119.315940.
789 57. Winkels H, Ehinger E, Vassallo M, Buscher K, Dinh HQ, Kobiyama K, Hamers AAJ,
790 Cochain C, Vafadarnejad E, Saliba AE, Zernecke A, Pramod AB, Ghosh AK, Michel NA,
791 Hoppe N, Hilgendorf I, Zirlik A, Hedrick CC, Ley K, Wolf D (2018) Atlas of the immune
792 cell repertoire in mouse atherosclerosis defined by single-cell RNA-sequencing and mass
793 cytometry. Circ Res 122:1675–1688. doi: 10.1161/CIRCRESAHA.117.312513.
794 58. Wirka RC, Wagh D, Paik DT, Pjanic M, Nguyen T, Miller CL, Kundu R, Nagao M,
795 Coller J, Koyano TK, Fong R, Woo YJ, Liu B, Montgomery SB, Wu JC, Zhu K, Chang R,
796 Alamprese M, Tallquist MD, Kim JB, Quertermous T (2019) Atheroprotective roles of
797 smooth muscle cell phenotypic modulation and the TCF21 disease gene as revealed by
798 single-cell analysis. Nat Med 25:1280–1289. doi: 10.1038/s41591-019-0512-5.
799 59. Zernecke A, Winkels H, Cochain C, Williams JW, Wolf D, Soehnlein O, Robbins CS,
800 Monaco C, Park I, McNamara CA, Binder CJ, Cybulsky MI, Scipione CA, Hedrick CC,
801 Galkina E V, Kyaw T, Ghosheh Y, Dinh HQ, Ley K (2020) Meta-Analysis of Leukocyte
802 Diversity in Atherosclerotic Mouse Aortas. Circ Res 127:402–426. doi:
803 10.1161/CIRCRESAHA.120.316903.
804 60. Zhang L, Li Z, Skrzypczynska KM, Fang Q, Zhang W, O'Brien SA, He Y, Wang L,
805 Zhang Q, Kim A, Gao R, Orf J, Wang T, Sawant D, Kang J, Bhatt D, Lu D, Li C-M,

806 Rapaport AS, Perez K, Ye Y, Wang S, Hu X, Ren X, Ouyang W, Shen Z, Egen JG, Zhang
807 Z, Yu X (2020) Single-Cell Analyses Inform Mechanisms of Myeloid-Targeted Therapies
808 in Colon Cancer. Cell 181:442-459.e29. doi: 10.1016/j.cell.2020.03.048.
809 61. Zhang Y, Zheng L, Zhang L, Hu X, Ren X, Zhang Z (2019) Deep single-cell RNA
810 sequencing data of individual T cells from treatment-naive colorectal cancer patients. Sci
811 data 6:131. doi: 10.1038/s41597-019-0131-5.
812 62. Zheng C, Zheng L, Yoo J-K, Guo H, Zhang Y, Guo X, Kang B, Hu R, Huang JY, Zhang
813 Q, Liu Z, Dong M, Hu X, Ouyang W, Peng J, Zhang Z (2017) Landscape of Infiltrating T
814 Cells in Liver Cancer Revealed by Single-Cell Sequencing. Cell 169:1342-1356. doi:
815 10.1016/j.cell.2017.05.035.
816 63. Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, Ziraldo SB, Wheeler
817 TD, McDermott GP, Zhu J, Gregory MT, Shuga J, Montesclaros L, Underwood JG,
818 Masquelier DA, Nishimura SY, Schnall-Levin M, Wyatt PW, Hindson CM, Bharadwaj R,
819 Wong A, Ness KD, Beppu LW, Deeg HJ, McFarland C, Loeb KR, Valente WJ, Ericson
820 NG, Stevens EA, Radich JP, Mikkelsen TS, Hindson BJ, Bielas JH (2017) Massively
821 parallel digital transcriptional profiling of single cells. Nat Commun 8:14049. doi:
822 10.1038/ncomms14049.
823 64. Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, Ziraldo SB, Wheeler
824 TD, McDermott GP, Zhu J, Gregory MT, Shuga J, Montesclaros L, Underwood JG,
825 Masquelier DA, Nishimura SY, Schnall-Levin M, Wyatt PW, Hindson CM, Bharadwaj R,
826 Wong A, Ness KD, Beppu LW, Deeg HJ, McFarland C, Loeb KR, Valente WJ, Ericson
827 NG, Stevens EA, Radich JP, Mikkelsen TS, Hindson BJ, Bielas JH (2017) Massively
828 parallel digital transcriptional profiling of single cells. Nat Commun 8:1–12. doi:
829 10.1038/ncomms14049.
830
831

832 **Declarations**

833 **Funding:**

834 National Institutes of Health Grant R35-HL-145241, R01-HL-121697, R01-HL-148094

835 (K.L.)

836 National Institutes of Health Grant P01-KL-136275 (C.C.H.)

837 National Institutes of Health Grant R01-HL-134236 (C.C.H.)

838 National Institutes of Health Grant R01-HL-126543, 5R01-HL-126543-05, 5R01-HL-

839 140976-02, R01-HL-148094-01, R01-HL-148094 (R.C.K.)

840 National Institutes of Health Grant K01-HL-137557 (D.B.H.)

841 National Institutes of Health Grant U01-AI-103408 (I.O.)

842 National Institutes of Health Grant U01-AI-103390 (A.A.A.)

843 National Institutes of Health Grant U01-AI-034989 (P.C.T.)

844 Cancer Research Institute (CRI) (A.S.)

845 American Heart Association (AHA) 19POST34450020 (L.E.P)

846 National Institutes of Health Grant NIAID, NICHD, NCI, NIDA, NIMH, NIDCR,

847 NIAAA, NIDCD, UL1-TR-000004, P30-AI-050409, P30-AI-050410, and P30-AI-

848 027767.

849 Formación de Profesorado Universitario (FPU) 16/02780 (R.B.D.)

850 Swedish Society for Medical Research (SSMF) (J.V.)

851 Data in this manuscript were collected by the Women's Interagency HIV study, now the

852 MACS/WIHS Combined Cohort Study (MWCCS).

853

854

855 **Author contributions:**

856 Design of the study: J.V., R.S., Y.G., C.P.D., E.E.

857 Collection of samples and data. A.L.L., R.P.T., J.M.L., W.J.M., K.M.W, A.A.A.,

858 H.N.H., P.C.T., I.O., S.L.H., and R.C.K.

859 Analysis of clinical data: D.B.H.

860 Design and collection of data for the B mode ultrasound sub-study: H.N.H.

861 scRNA-Seq experiments: C.P.D., and E.E.

862 Analysis of the data: J.V., R.S., R.G., Y.G., P.R., T.P., L.E.P., C.E.O., R.B.D.,

863 H.Q.D., A.S., C.A.M., L.L.L., T.W., C.C.H., and K.L.

864 Bioinformatics analysis: R.G., Y.G. and H.Q.D.

865 Writing-original draft: J.V., R.S., and K.L.

866 **Competing interests:** There are no conflicts of interest.

867 **Data and material availability:** All data are available in the main text or the

868 supplementary materials.

869 **Code availability:** All packages used for the analysis on this data are available in R.

870 **Ethics approval:** All participants provided informed consent, and each site's

871 Institutional Review Board approved the studies.

872 **Consent to participate**: All participants provided informed consent, and each site's

873 Institutional Review Board approved the studies.

874 **Consent for publication:** All the listed authors have reviewed the manuscript and agreed

875 with its submission.

876

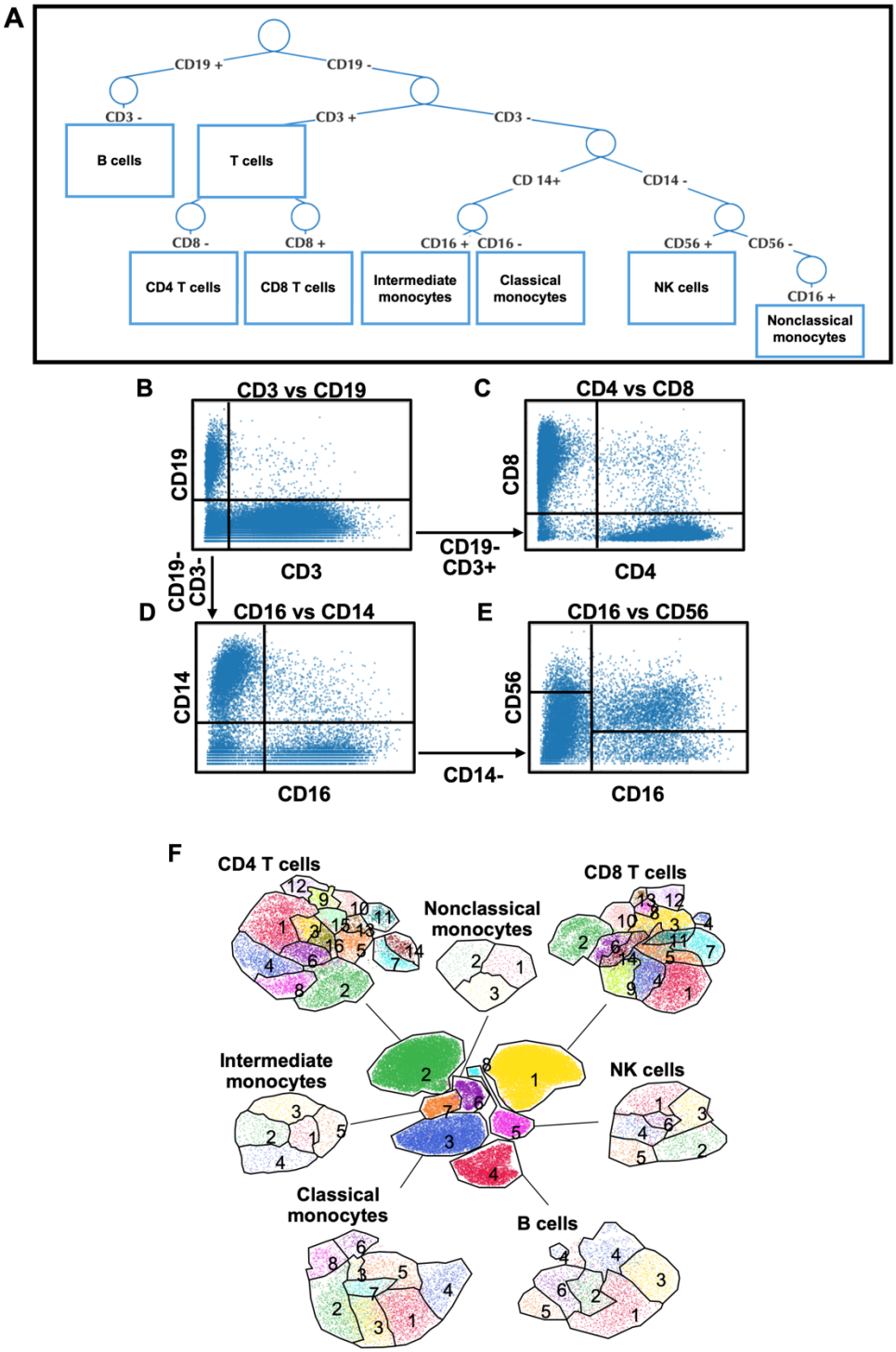
877

878 **Figures**

879

880 **Figure 1. Gating scheme (A), biaxial dot plots (B-E) to identify major known cell**

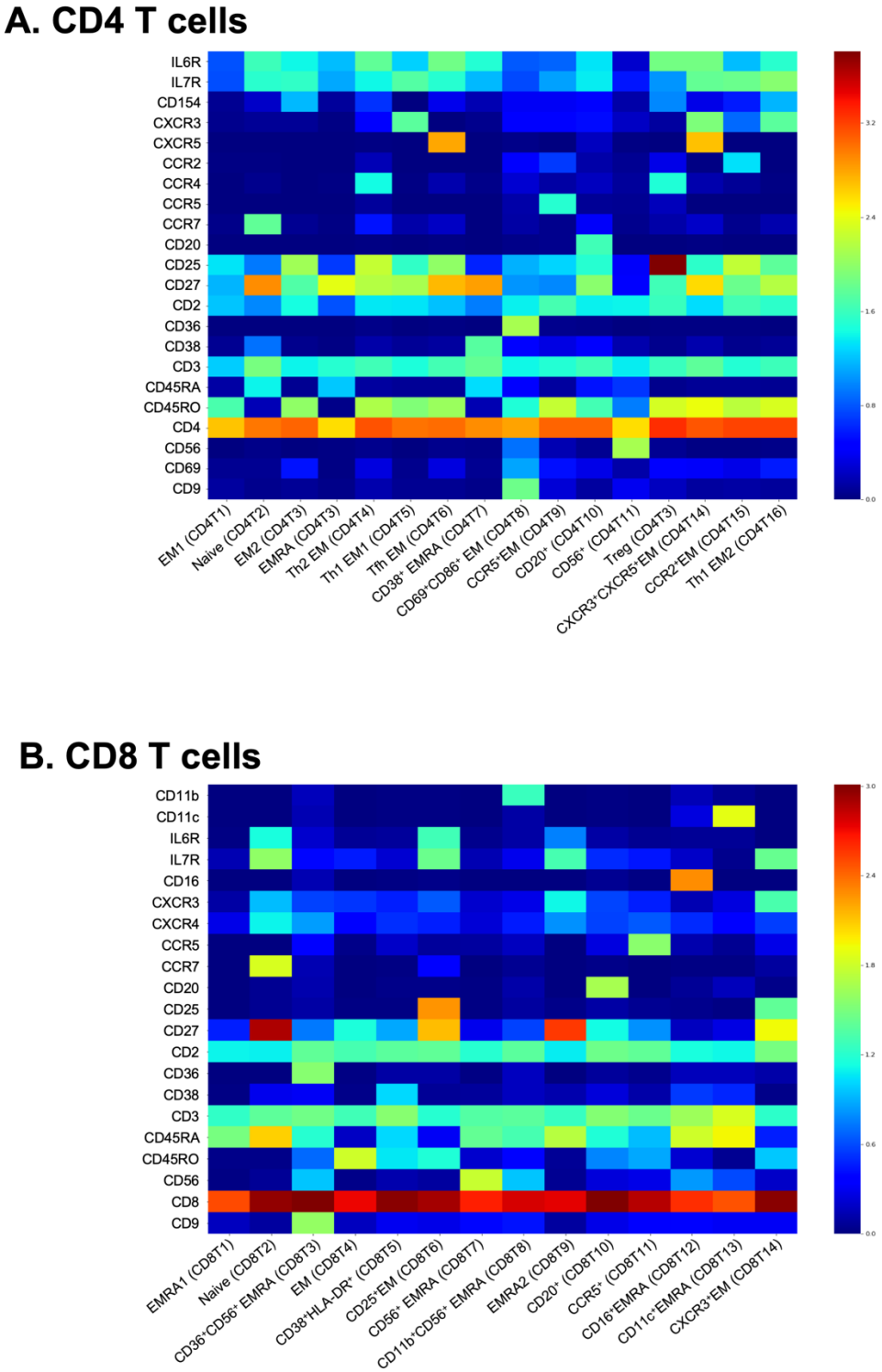
881 **types, and (F) antibody- based UMAP clustering of major cell types.**



882

PBMCs from 32 WIHS participants were hash-tagged and stained with 40 oligonucleotide-tagged mAbs (**Table S3**). (**B**) B cells were defined as CD19+CD3- and T cells as CD19-CD3+. (**C**) T cells were identified as CD4 (CD4+CD8-) or CD8 (CD4-CD8+). (**D**) All CD19-CD3- cells were gated for CD14 and CD16, with CD14+CD16- cells being classical (CM) and CD14+CD16+ being intermediate (INT) monocytes. (**E**) The CD14-CD16+ cells from panel D contain NK cells, which were identified by CD56 and defined as CD56+CD14-CD20-CD123-CD206-. Most of the remaining CD56-CD16+ cells were nonclassical monocytes (NCM). (**F**) The major known cell types were UMAP-Louvain-clustered by CD3, CD19, CD14, CD16, and CD56 surface expression (central panel). Then, each major known cell type was UMAP-Louvain-clustered by all non-negative surface markers. CD4 T cells formed 16 clusters, cluster numbers indicated; CD8 T cells formed 14 clusters; Classical monocytes (CM) formed 8, Intermediate monocytes (INT) 5, and Nonclassical monocytes (NCM) 3 clusters. B cells and NK cells formed 6 clusters each.

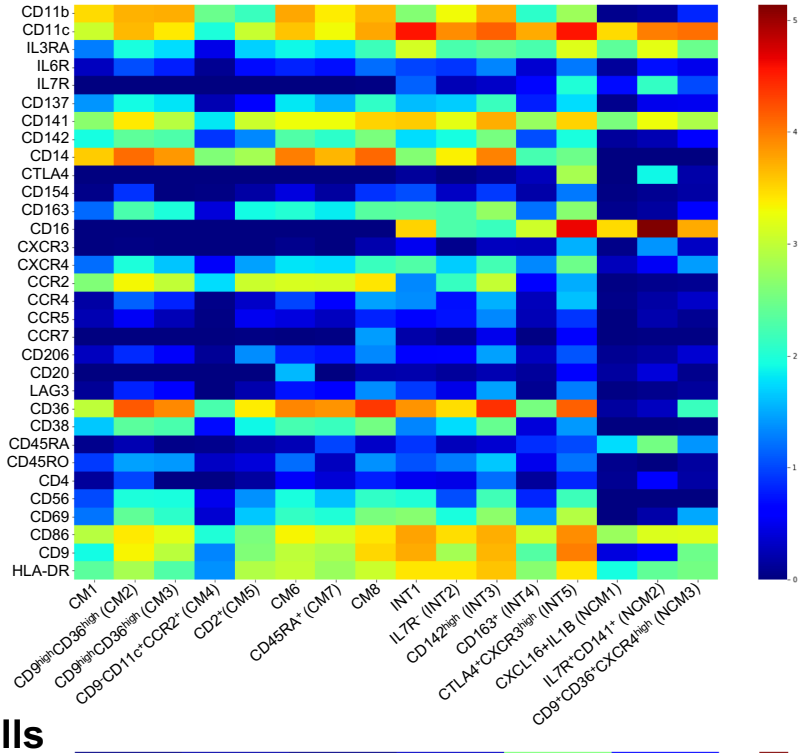
906 **Figure 2. Heatmaps of antibody expression (log_2 scale) in each main cell type.**



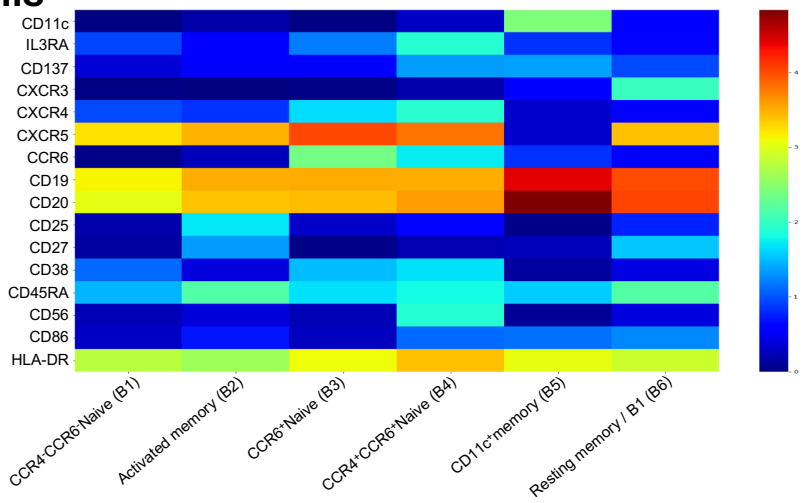
907

908

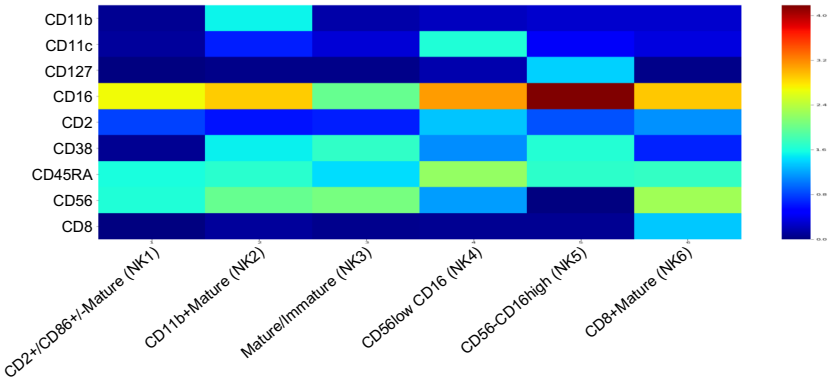
C. Monocytes



D. B cells

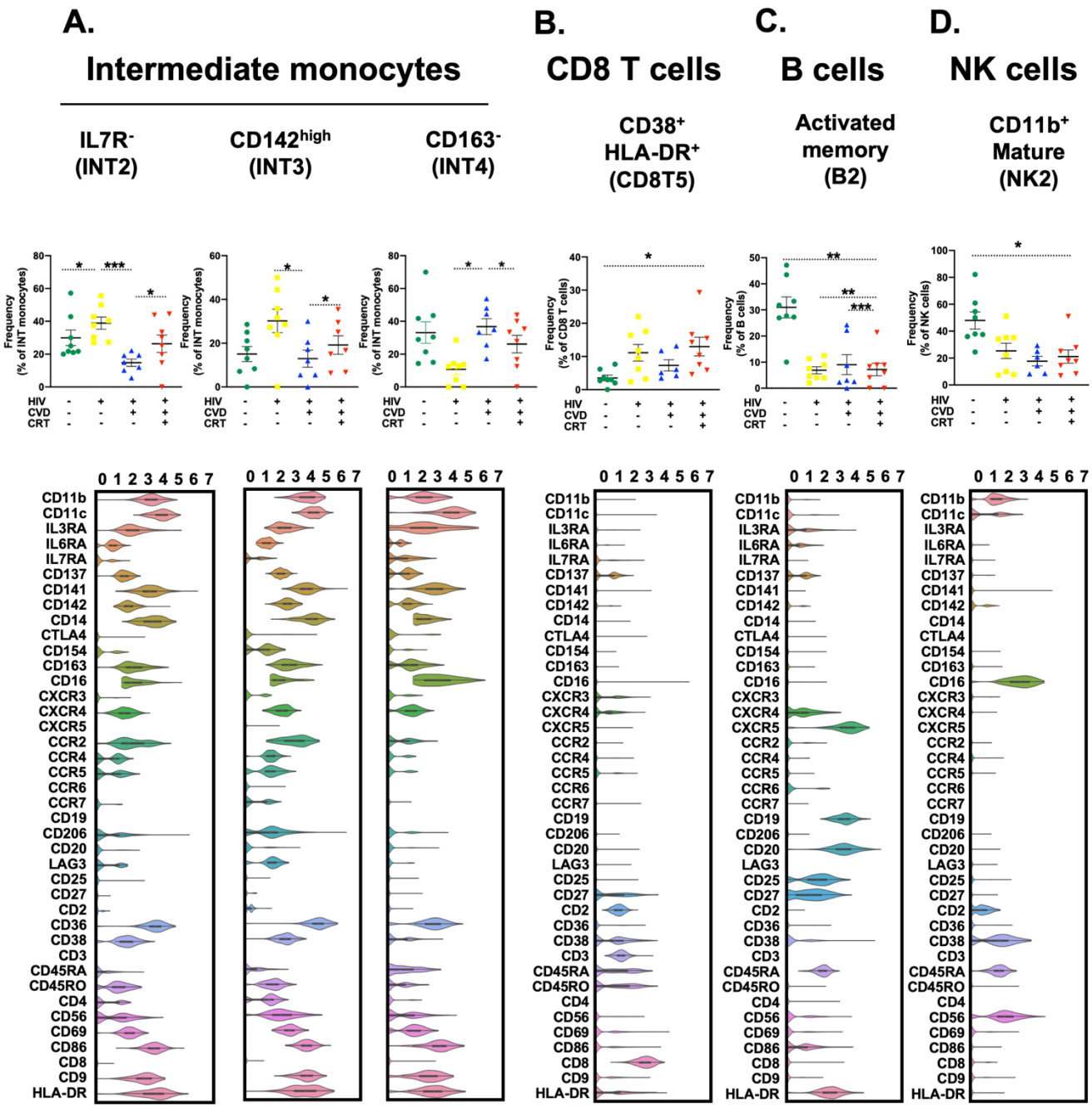


E. NK cells



910 (**A**) CD4 T cell, (**B**) CD8 T cell, (**C**) Monocytes, (**D**) B cells, and (**E**) NK cells.
911 Immunophenotypes at the bottom. EM, effector memory; EMRA, terminally
912 differentiated effector memory; CM, Classical Monocyte; INT, Intermediate Monocyte;
913 NCM, Nonclassical Monocyte.
914

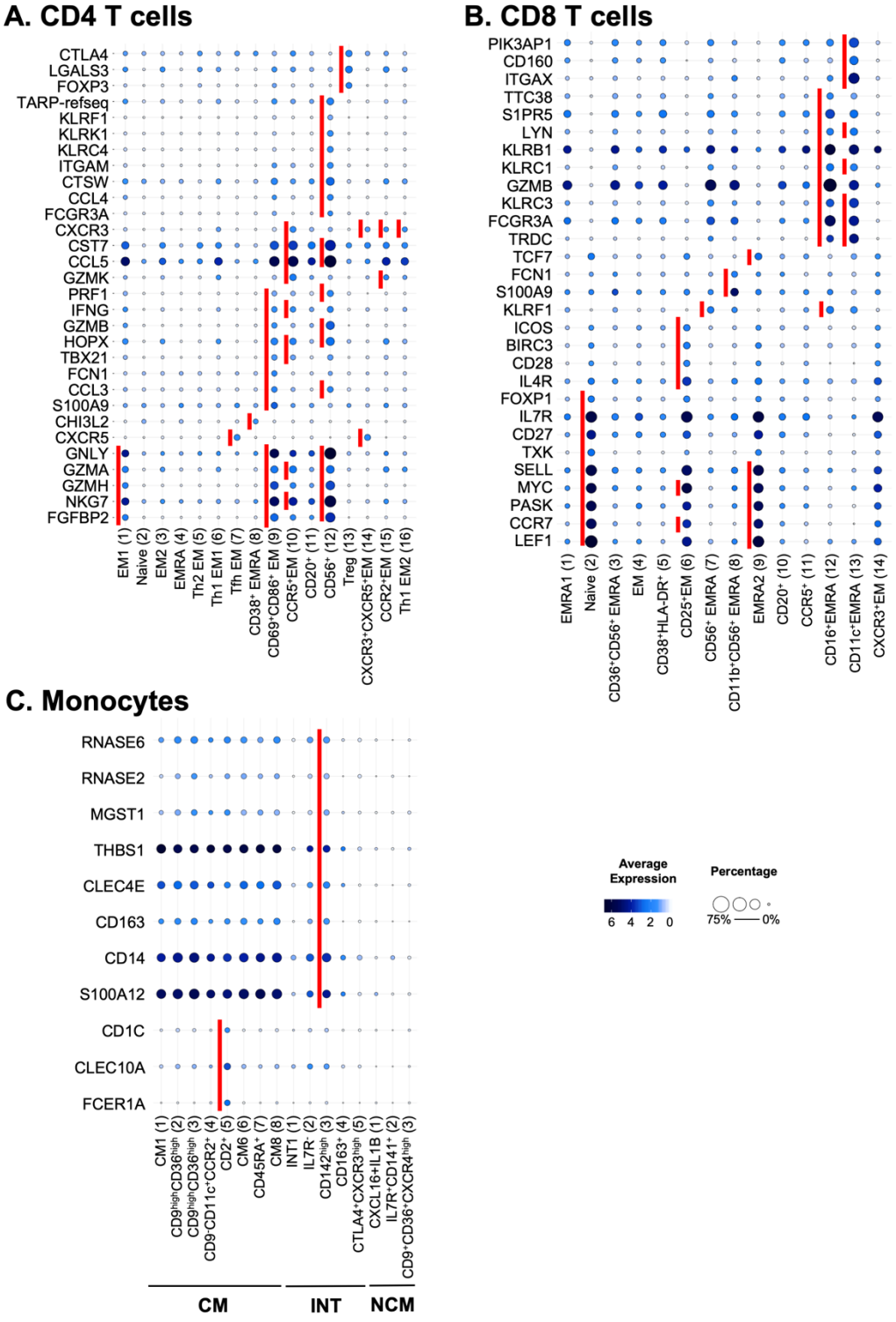
Figure 3. Cell proportions in women with HIV, CVD, both or neither.



HIV-CVD- (green), HIV+CVD- (yellow), HIV+CVD+ (blue), and HIV+CVD+ on CRT (red), from left to right. 8 samples per group except 7 for HIV+CVD+. Proportions of cells in each cluster calculated as percentage of the parent cell type as indicated in the

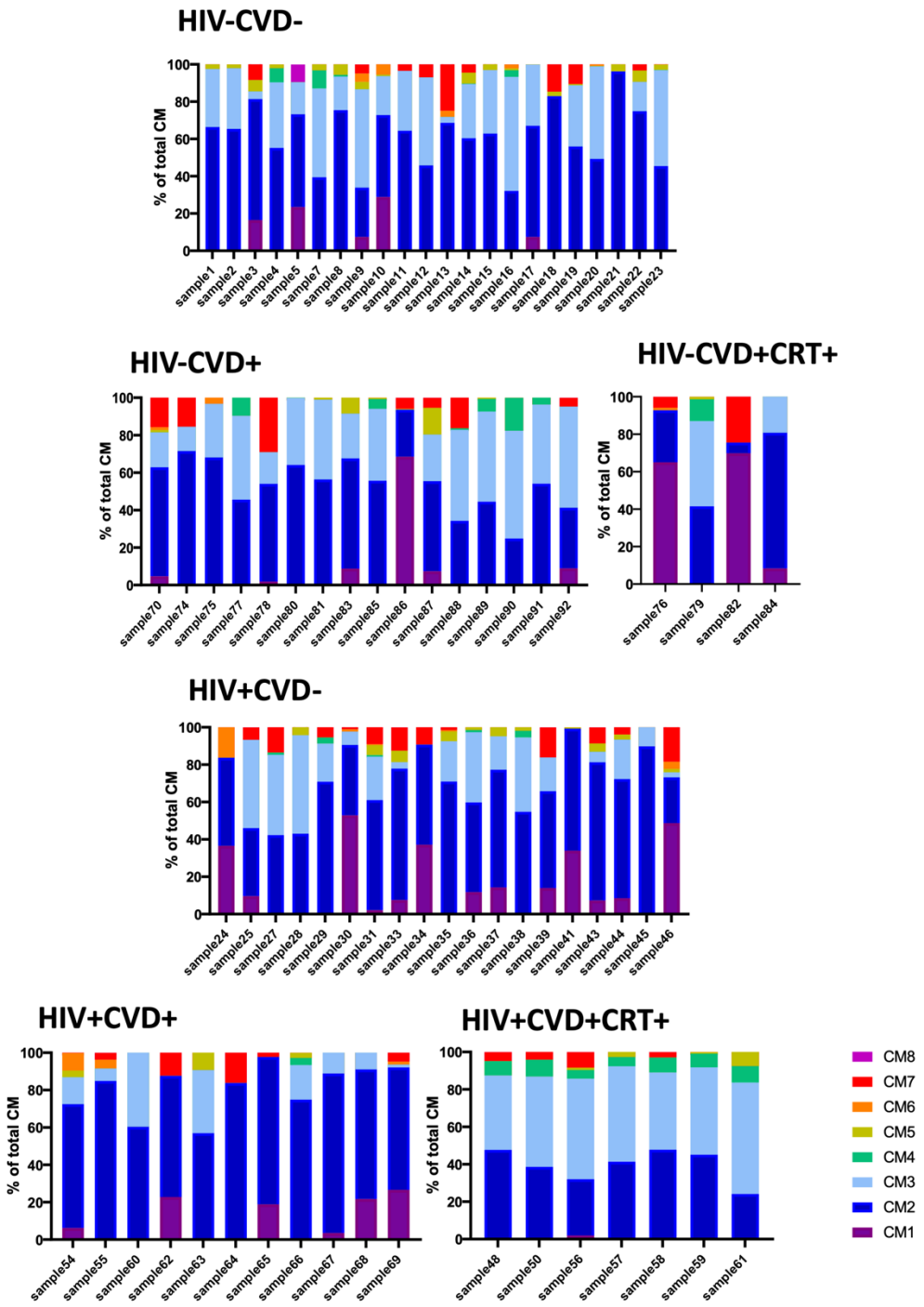
920 title of each panel. Clusters with significant differences (*, $p<0.05$, **, $p<0.01$, ***
921 $p<0.001$) in cell proportions (by log odds ratio) are shown with individual data points,
922 means and standard error of the mean (SEM). Violin plots below show expression of all
923 40 cell surface markers ($\log_2$ scale). INT, intermediate monocytes; CRT, cholesterol-
924 reducing treatment.
925

Figure 4. Significantly differentially expressed genes of cells in each cluster.



928 Expression of 485 transcripts was determined by targeted amplification (BD Rhapsody
929 system). Significant genes defined as adjusted $p<0.05$ and $\log_2$ fold change >0. Dot plot:
930 fraction of cells in cluster expressing each gene shown by size of circle and level of
931 expression shown from white (=0) to dark blue (=max, $\log_2$ scale). Red bars indicate
932 genes that were significantly higher in one cluster compared to all other clusters of the
933 parent cell type. There were no DEGs in NK cell clusters. (**A**) CD4 T cells, (**B**) CD8 T
934 cells and (**C**) monocytes. CM, Classical monocytes; INT, Intermediate monocytes; NCM,
935 Nonclassical monocytes; EM, effector memory; EMRA, terminally differentiated effector
936 memory.
937

938 **Figure 5. Cell proportions of classical monocyte subsets.**



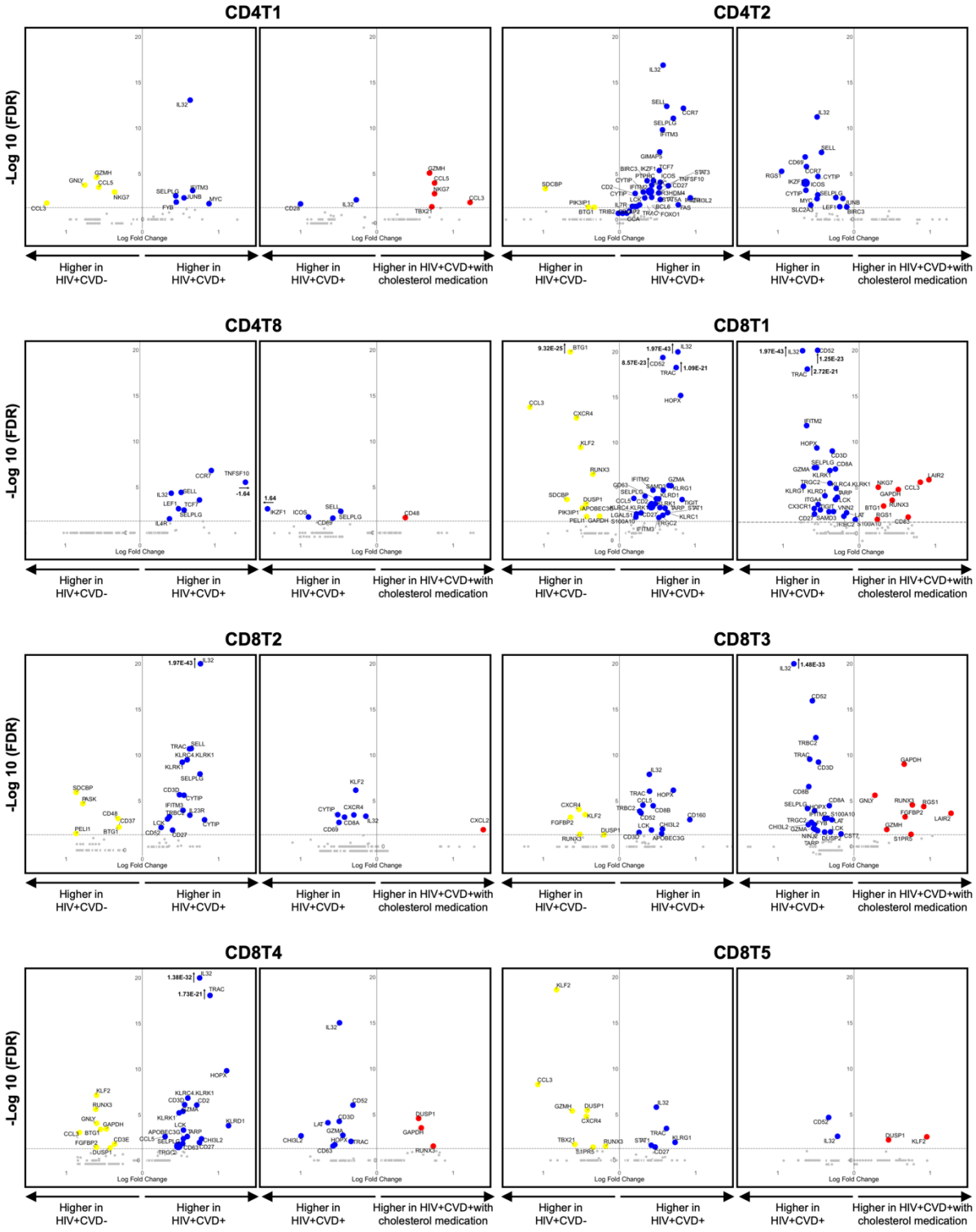
939

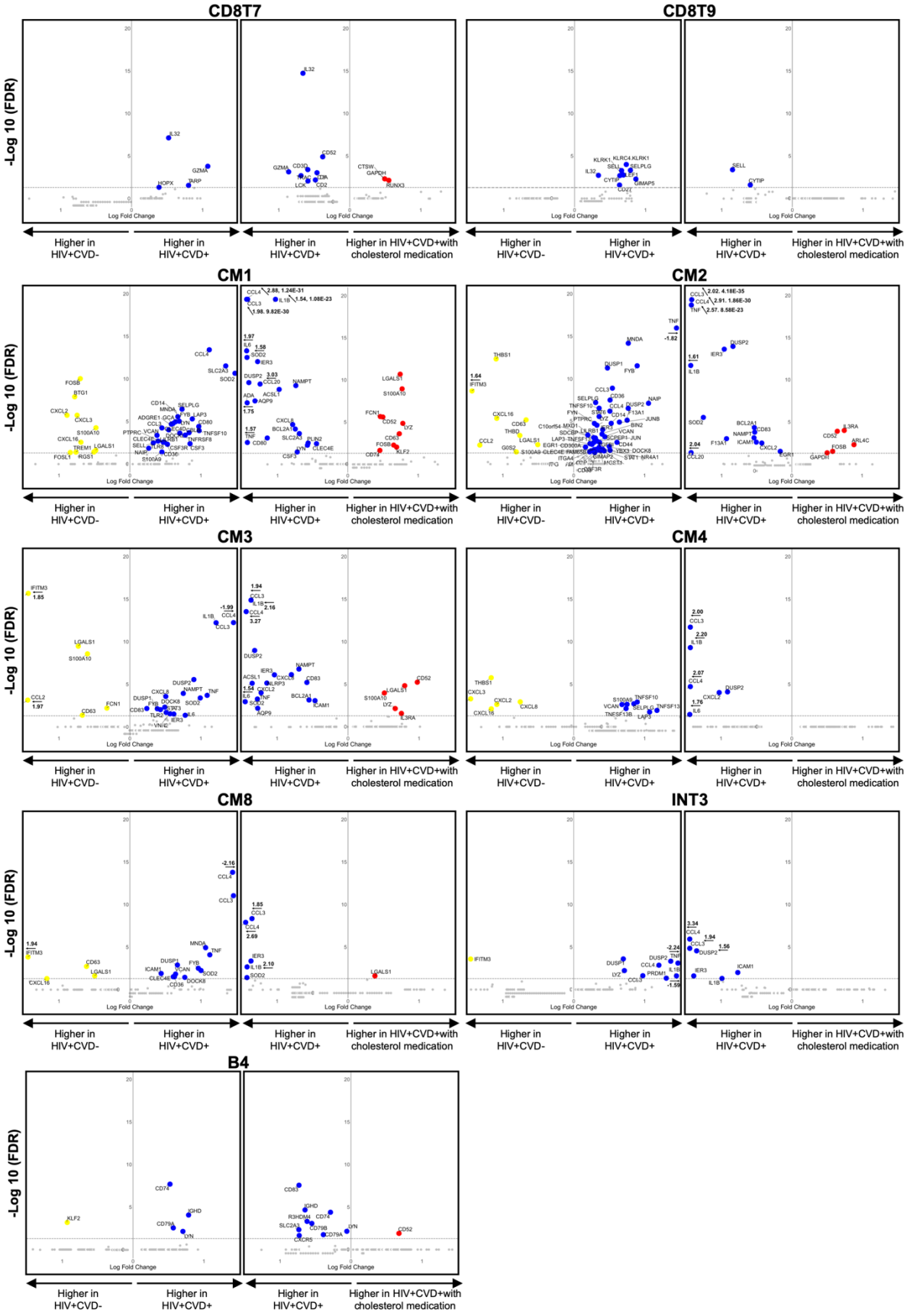
940 Cell proportions of classical monocytes (CMs, CM1-8) from the present dataset in 92

941 samples of classical monocyte transcriptomes in women with HIV, CVD, both or

942 neither.[7] CM, classical monocyte; CRT, cholesterol-reducing treatment.

Figure 6. Volcano plots comparing gene expression in single cells from WIHS participant types in each cluster.




CD8T7
CD8T9
CM1
CM2
CM3
CM4
CM8
INT3
B4
-Log 10 (FDR)
Log Fold Change
Higher in HIV+CVD-
Higher in HIV+CVD+
Higher in HIV+CVD+
Higher in HIV+CVD+with cholesterol medication

947 Gene expression focused on HIV+CVD- vs HIV+CVD+, and HIV+CVD+ vs
948 HIV+CVD+ with CRT. All clusters in which at least 10 genes were statistically
949 significant are shown. Colored dots (HIV+CVD- yellow, HIV+CVD+ blue, and
950 HIV+CVD+ with CRT red) indicate significantly differentiated expressed genes
951 (FDR<0.05 and $|\log_2 FC|>2$). 3 CD4 T and 7 CD8 T cell clusters, 5 CM and 1 each INT
952 and B cell clusters met these criteria. Dashed line indicates adjusted p-value of 0.05. CM,
953 Classical monocytes; INT, Intermediate monocytes.
954