1	Executable models of pathways built using single-cell RNA
2	seq data reveal immune signaling dysregulations in people
3	living with HIV and atherosclerosis
4	Mukta G. Palshikar <sup>1</sup> , Rohith Palli <sup>2</sup> , Alicia Tyrell <sup>3</sup> , Sanjay Maggirwar <sup>4</sup> , Giovanni Schifitto <sup>5, 6</sup> , Meera
5	V. Singh <sup>5, 8,</sup> and Juilee Thakar <sup>1, 7, 8, 9**</sup>
6	**Correspondence: Juilee_Thakar@URMC.Rochester.edu
7	<sup>1</sup> Biophysics, Structural, and Computational Biology Program, University of Rochester School of
8	Medicine and Dentistry, Rochester, USA. <sup>6</sup> Department of Biostatistics and Computational
9	Biology, University of Rochester School of Medicine and Dentistry, Rochester, USA. <sup>7</sup>
10	Department of Microbiology and Immunology, University of Rochester School of Medicine and
11	Dentistry, Rochester, USA. 8 Department of Biomedical Genetics, University of Rochester
12	School of Medicine and Dentistry, Rochester, USA
13	Full list of author information is available at the end of the article
14	Abstract:
15	Background: Atherosclerosis (AS)-associated cardiovascular disease is an important cause of mortality
16	in an aging population of people living with HIV (PLWH). This elevated risk of atherosclerosis has been

17 attributed to viral infection, prolonged usage of anti-retroviral therapy, and subsequent chronic

18 inflammation.

- 19 Methods: To investigate dysregulated immune signaling in PLWH with and without AS, we sequenced
- 20 9368 peripheral blood mononuclear cells (PBMCs) from 8 PLWH, 4 of whom also had atherosclerosis
- 21 (AS+). To develop executable models of signaling pathways that drive cellular states in HIV-associated
- 22 atherosclerosis, we developed the single-cell Boolean Omics Network Invariant Time Analysis

23	(scBONITA) algorithm. ScBONITA (a) uses single-cell RNA sequencing data to infer Boolean rules for
24	topologically characterized networks, (b) prioritizes genes based on their impact on signaling, (c)
25	performs pathway analysis, and (d) maps sequenced cells to characteristic signaling states. We used
26	scBONITA to identify dysregulated pathways in different cell-types from AS+ PLWH and AS- PLWH.
27	To compare our findings with pathways associated with HIV infection, we used scBONITA to analyze a
28	publicly available dataset of PBMCs from subjects before and after HIV infection. Additionally, the
29	executable Boolean models characterized by scBONITA were used to analyze observed cellular states
30	corresponding to the steady states of signaling pathways
31	<b>Results:</b> We identified an increased subpopulation of CD8+ T cells and a decreased subpopulation of
32	monocytes in AS+ PLWH. Dynamic modeling of signaling pathways and pathway analysis with
33	scBONITA provided a new perspective on the mechanisms of HIV-associated atherosclerosis. Lipid
34	metabolism and cell migration pathways are induced by AS rather than by HIV infection. These pathways
35	included AGE-RAGE and PI3K-AKT signaling in CD8+ T cells, and glucagon and cAMP signaling
36	pathways in monocytes. Further analysis of other cell subpopulations suggests that the highly
37	interconnected PI3K-AKT signaling pathway drives cell migratory state in response to dyslipidemia.
38	scBONITA attractor analysis mapped cells to pathway-specific signaling states that correspond to distinct
39	cellular states.
40	Conclusions: Dynamic network modeling and pathway analysis with scBONITA indicates that
41	dysregulated lipid signaling regulates cell migration into the vascular endothelium in AS+ PLWH.
42	Attractor analysis with scBONITA facilitated pathway-based characterization of cellular states that are
43	not apparent in gene expression analyses.
44	Keywords: single-cell RNA sequencing; Boolean networks; HIV; atherosclerosis; pathway analysis;

45 network modeling

# 46 Background

47 Human immunodeficiency virus (HIV) infection greatly increases the risk of atherosclerosis (AS) 48 associated cardiovascular disease (CVD), which is a leading cause of morbidity and mortality in persons 49 living with HIV (PLWH) (1-4). Several factors contribute to this elevated risk of AS. PLWH have a 50 higher prevalence of traditional risk factors for AS such as dyslipidemia, diabetes, hypertension, and 51 smoking (5-9). In addition, the off-target effects of certain classes of antiretroviral drugs lead to an 52 increase in traditional metabolic risk factors such as dyslipidemia, weight gain, and metabolic syndrome 53 (9, 10). However, PLWH have an elevated risk of developing CVD even when controlling for these risk 54 factors (4, 11, 12). Finally, HIV infection itself causes metabolic changes leading to a pro-atherogenic 55 inflammatory environment in the vasculature (13-16). 56 HIV infection and long-term antiretroviral therapy modulate signaling dynamics, leading to changes in 57 the composition of peripheral blood mononuclear cells (PBMCs) and in the expression of functionally 58 important molecules in these cells (17). HIV infection mediates an array of molecular signaling pathways, including inflammasome activation, cell migration and apoptosis. These signaling pathways contribute to 59 60 immune cell activation and inflammation in the vasculature (18). Biomarker studies also highlight the 61 importance of these processes in atherogenesis, especially in the context of activated 62 monocyte/macrophages and T cells (14-16). CD8+ T cells contribute to the atherogenic environment by 63 cytokine secretion, secretion of cytotoxic granules and formation of the necrotic core of atherosclerotic 64 plaques (reviewed in (19)). Monocyte/macrophages migrate into the intima and eventually form 65 apoptotic, atherosclerotic plaques (reviewed in (20)). The interplay between immune signaling pathways 66 and immune cell activation and inflammation in the context of HIV infection still requires further investigation. Single-cell measurements of expression profiles allow the investigation of these 67 68 perturbations simultaneously. To investigate the mechanistic link between HIV infection and 69 atherosclerosis, we used the 10X Genomics platform to sequence ~10,000 PBMCs from 8 PLWH, 4 of 70 whom have atherosclerosis.

71 Typically, analysis of single cell RNA sequencing (scRNA-seq) data uses clustering methods to define 72 cell subpopulations, followed by differential expression and gene set overrepresentation analysis (ORA) 73 to estimate modulation of molecular pathways in the condition under study. This standard analysis 74 approach discounts pathway topology and falls short of connecting molecular state to cellular state. 75 Furthermore, ORA ignores synergistic interactions among genes by, in effect, treating genes as 76 independent and equal, resulting in a failure to correctly estimate the significance of pathways (21). 77 Discrete-state network modeling facilitates prioritization of experiments by using simple logic rules such 78 as 'AND' or 'OR' to explicitly define signal integration, enabling investigation of crosstalk and 79 downstream events as shown in our previous studies. Previously, we have extensively used discrete-state 80 network modeling to investigate virus infections and have experimentally validated the predictions (22, 81 23). We have also developed an algorithm to perform Boolean rule inference and pathway analysis using 82 bulk transcriptomic data. This algorithm has been rigorously tested and compared across other widely 83 used gene-set enrichment methods (24). Here, we expand our discrete-state modeling method and present 84 single-cell Boolean Omics Network Invariant-Time Analysis (scBONITA) to (a) infer Boolean rules (e.g., 85 "AND", "OR") describing signal integration for gene interactions described by the pathway topologies 86 from scRNA-seq data and (b) use these inferred regulatory rules to identify condition-specific 87 dysregulated pathways and to prioritize genes/proteins for further investigation. Instead of simply 88 returning a list of dysregulated pathways and associated p-values, scBONITA returns precise modes of 89 dysregulation, captured by node-level impact scores that quantify the contribution of each node (a gene or 90 protein) to the overall dysregulation of a pathway, measured by *in silico* perturbation and simulation that 91 matches experimental conditions. In silico simulation and perturbation of molecular pathways allows the 92 use of scBONITA as a powerful hypothesis-generating tool.

We demonstrate that scBONITA can infer dysregulated pathways from single-cell RNA sequencing
(scRNA-seq) data upon HIV infection using publicly available data and in the context of HIV-associated
atherosclerosis using data presented in this work. scBONITA identifies dysregulated cell migration and

96 lipid metabolism related pathways such as PI3K signaling, leukocyte transendothelial migration, and 97 AGE-RAGE signaling in the subpopulations of CD8+ T cells and monocytes, known to be implicated in 98 HIV-associated atherosclerosis. ScBONITA identifies genes, such as the PI3K and PLC genes, which 99 have high impacts on signal flow through the signaling pathways named above. CD4+ T cells and B cells 100 are also known to play significant roles in the development of atherosclerosis and cardiovascular disease 101 (25-35). In both subpopulations, scBONITA identified pathways that were linked to cell migration (such 102 as proteoglycans in cancer and the regulation of actin cytoskeleton pathways), lipid metabolism (P13K-103 Akt signaling, phosphatidylinositol signaling) and pathways linked to intercellular communication (such 104 as the chemokine signaling, apelin signaling, and cytokine-cytokine receptor interaction pathways). 105 Furthermore, we used a publicly available dataset of PBMCs from persons before and after HIV infection 106 to show that cell migration pathways are also dysregulated in the early stages of HIV infection, indicating 107 a role for these pathways in both the antiviral response and in subsequent AS (36). We also present a 108 novel method for mapping cells to pathway-specific signaling states using rules identified by BONITA. In 109 conclusion, scBONITA is a powerful tool that can be used for network modeling using single-cell RNA-110 seq data. In this study, we demonstrate that scBONITA provides an insight into the mechanisms of HIV-111 associated atherosclerosis at the single-cell level.

# 112 Methods

# 113 Participant cohort summary, sample collection, and storage

Eight men living with HIV and >= 50 years of age on stable combined antiretroviral therapy (cART) for at least 1 year and with viral load <= 50 copies/mL were recruited. All methods were carried out in accordance with University of Rochester guidelines and regulations, and all experimental and study protocols were approved by the University of Rochester Institutional Review Board (#RSRB00063845). Informed consent was obtained from all subjects. Individuals were classified as having atherosclerosis

119 (AS+) if they had plaques on the carotid arteries on ultrasound imaging. Four of the 8 subjects were 120 assigned as AS+ and had plaques in both right and left carotid arteries. AS- subjects were aged between 121 47 and 57 and AS+ subjects were aged between 51 and 66. AS+ subjects had mean serum cholesterol of 122 161.5 mg/dl ( $\sigma$  = 40.9) and mean serum high-density lipid HDL of 54.7 mg/dl ( $\sigma$  = 16.3). AS- subjects 123 had mean serum cholesterol of 167.7 mg/dl ( $\sigma$  =57.2) and mean serum high-density lipid HDL of 51 124 mg/dl ( $\sigma = 7.7$ ). AS- subjects and AS+ subjects had a mean CD4+ T cell count of 518.5 cells/µl ( $\sigma = 347.8$ 125 cells/ $\mu$ l) and 838.7 cells/ $\mu$ l ( $\sigma$  = 514.5 cells/ $\mu$ l) respectively. 30 mls of blood per study participant was 126 collected in ACD vacutainers and was processed within 2 - 3 hours of collection. Peripheral Blood 127 Mononuclear Cells (PBMCs) were isolated using Ficoll density gradient centrifugation. 5 million PBMCs 128 were preserved using RNAlater (Thermo Fisher) and were used for scRNA-seq. De-identified subject 129 information is available in Supplementary File 1, Supplementary Figure 1.

### 130 Single-cell sequencing and data processing

131 Frozen vials containing cells in RNAlater were thawed quickly in a 37-degree water bath. Cell suspension 132 was transferred to a 15ml conical tube. 10 ml PBS/2% FBS was slowly added. Samples were centrifuged 133 at 1600rpm for 6 min. Washes were repeated for an additional 2 times for a total of 3 washes. Using the 134 MACS Miltenyi Biotec Dead Cell removal kit (PN130-090-101), dead cells were removed using 135 manufacturer's recommendations. Cells were counted and cellular suspensions were loaded on a 136 Chromium Single-Cell Instrument (10x Genomics, Pleasanton, CA, USA) to generate single-cell Gel 137 Bead-in-Emulsions (GEMs). ScRNA-seq libraries were prepared using Chromium Single-Cell 3' Library 138 & Gel Bead Kit (10x Genomics). The beads were dissolved, and cells were lysed per manufacturer's 139 recommendations. GEM reverse transcription (GEM-RT) was performed to produce a barcoded, full-140 length cDNA from poly-adenylated mRNA. After incubation, GEMs were broken, and the pooled post-141 GEM-RT reaction mixtures were recovered, and cDNA was purified with silane magnetic beads 142 (DynaBeads MyOne Silane Beads, PN37002D, ThermoFisher Scientific). The entire purified post GEM-143 RT product was amplified by PCR. This amplification reaction generated sufficient material to construct a

144 3' cDNA library. Enzymatic fragmentation and size selection was used to optimize the cDNA amplicon 145 size and indexed sequencing libraries were constructed by End Repair, A-tailing, Adaptor Ligation, and 146 PCR. Final libraries contain the P5 and P7 priming sites used in Illumina bridge amplification. Sequence 147 data was generated using Illumina's NovaSeq 6000. Approximately 2000 cells were sequenced from each 148 subject. Cell Ranger (version 2.1.1; 10x Genomics) was used for demultiplexing and alignment with 149 default parameters. Reads were aligned to the human reference genome GRCh38 (Ensembl 93). The 150 Seurat R package (37) was used to further process the gene counts obtained from the CellRanger pipeline. 151 Cells that express < 200 genes, > 2500 genes, and > 5% mitochondrial genes were filtered out. Genes 152 expressed in < 3 cells were filtered out. Gene counts were per-cell normalized and log<sub>2</sub>-transformed. 153 These preliminary filtering and selection procedures yielded a set of 9368 sequenced cells, approximately 154 equally distributed between subjects (and hence conditions), and 14017 genes. Note that sample 155 collection, processing and sequencing were performed in one batch, leading to extremely high-quality 156 data where no subject specific patterns were observed.

#### 157 Classification into subpopulations using modularity-optimized

#### 158 Louvain community detection, and cluster labeling

159 Cells were classified into subpopulations using modularity optimized community detection, implemented 160 in the Seurat R package (37). 664 highly variable genes were used to identify 10 principal components 161 that explained the majority of variance in the data. These principal components were used to cluster the 162 data. Clustering yielded 16 subpopulations. Cluster markers were identified using MAST (38). As 163 suggested in (39), CIBERSORT (40) was used to "deconvolute" the average gene expression of each 164 cluster into the constituent canonical cell types. A reference expression set of 22 immune cell types and 165 547 genes was used (40). Over-representation analysis was performed using the implementation of the 166 hypergeometric test in the R package clusterprofiler with Kyoto Encyclopedia of Genes and Genomes

167 (KEGG) gene sets downloaded from MSigDb (41-43). Gene sets were identified as significantly over-

168 represented if the Bonferroni-adjusted p-value was < 0.05.

#### 169 scBONITA algorithm for development of discrete-state models of

#### 170 *pathways*

171 Network topologies: ScBONITA infers Boolean regulatory rules/ logic gates for directed

172 networks wherein nodes represent genes and edges represent the regulatory relationships between those

173 genes. These networks contain edge annotations denoting activation/inhibition relationships between

174 nodes, which are exploited by scBONITA to restrict the search space for rule inference to sign-compatible

175 canalyzing functions. Such network models of biological pathways are commonly obtained from pathway

176 databases such as KEGG and WikiPathways (43-45). ScBONITA offers an interface to KEGG and

177 WikiPathways databases that allows automated download and processing of user-specified networks.

178 Users can also provide custom networks in graphml format.

#### 179 **Boolean rule determination from scRNA seq data:** The underlying principle of

180 scBONITA is that cross-sectional measurements of cells by scRNA-seq data represent states of an

181 underlying dynamic biological process. scBONITA's rule determination (scBONITA-RD) algorithm,

182 which has been extended from our previous BONITA algorithm exploits this property to infer Boolean

rules for an input biological network, using a combination of a genetic algorithm (GA) and a node-wise

184 local search (46).

185 The global search using GA infer a single candidate rule set that adequately describes the input data with 186 respect to the network topology with minimum error (24, 47). The function to be minimized is:

187 
$$\sum_{c=1}^{cells} \min\left(\sum_{n=1}^{nodes} |E_{c,n} - A_{c,n,a}| \forall a \text{ in } T_c\right)$$

Where, *c* from *1* to *cells* iterates over the number of cells in the training dataset, n iterates from 1 to number of *nodes* in the network,  $E_{c,n}$  is the binarized expression of node *n* in cell *c*,  $A_{c,n,a}$  is the value of node *n* in the attractor *a* reachable from cell *c*, and  $T_c$  is the attractor reachable from *c*. Note that  $T_c$  may have multiple repeating states in a limit cycle or only one steady state, i.e., it may be a singleton attractor.  $T_c$  is obtained after simulating the network with the candidate rule set for 100 time-steps, which causes the simulation to reach an attractor state for all tested networks.

194 The minimum error rule set identified using the above-described genetic algorithm strategy is further

refined by a node-level local search that sequentially optimizes the rule for each node keeping the rules

196 for all other nodes in the network constant. An optimal set of rules for a node n is obtained by minimizing

197 the function

198 
$$\sum_{c=1}^{cells} \min\left(|E_{c,n} - A_{c,n,a}| \forall a \text{ in } T_c\right)$$

199 where variables and constants are same as described above.

Several rules may satisfy the termination criteria with equal errors. The local search thus returns a set of equivalent rules that all satisfactorily explain the observed state in the experimental data. This set of rules is referred to as the equivalent rule set (ERS) in the text.

203 Pathway analysis (PA) with scBONITA: scBONITA performs pathway analysis in a two-

step process. In the first step, importance scores for each node in the biological network under

205 consideration are calculated. In the second step, a pathway modulation metric incorporating both

206 experiment-specific fold changes and the node importance scores calculated in step 1 is calculated.

scBONITA quantifies the influence  $I_n$  of a node *n* over the state of the network by quantifying the overall

208 effect of its perturbation on that network. This is achieved by simulating knock-in and knock-out of that

209 node.

210 
$$I_n = \sum_{c=1}^{cells} |KI_{c,n} - KO_{c,n}| * Uncertainty Factor$$

211 where  $KI_{c,n}$  and  $KO_{c,n}$  are the discrete expression vectors of network nodes in the attractors reached after a

simulation starting from cell c where the node under consideration n is knocked in and knocked out

213 respectively. The uncertainty factor is defined as follows:

214 Uncertainty factor = 
$$\frac{|Maximum ERS_i| - |Observed ERS_i| + 1}{|Maximum ERS_i|}$$

215 Where  $ERS_i$  is the ERS for a node *i*,  $|Maximum ERS_i|$  is the maximum possible size of the ERS for a

node *i* and  $|Observed ERS_i|$  is the size of the ERS for a node *i* upon optimization by scBONITA.

217 The uncertainty factor weighs  $I_n$  relative to the maximum state space for that node, to capture the

218 uncertainty in the rule determination for that node. The importance scores of the nodes in a network are

scaled to [0, 1] by dividing by the maximum calculated importance score for the network under

220 consideration.

A pathway modulation metric  $(M_p)$  is calculated by weighting the node importance score by the difference between the average gene expression in each group (relative abundance, *RA*) and the standard deviation of expression of that gene ( $\sigma$ ). A p-value is calculated by bootstrapping, where a contrastspecific distribution of weighted importance scores is generated using randomly resampled *RA* values. Pathways are described in the text as being overall upregulated in a given contrast if the sum of fold changes of all genes in the pathway is positive. Conversely, pathways are described as being downregulated if the sum of fold changes of all genes in the pathway is negative.

228 
$$M_p = \sum_{n=1}^{nodes} RA_n * \sigma_n * I_n$$

Pathways are described in the text as being overall upregulated in a given contrast if the sum of fold
changes of all genes in the pathway is positive. Conversely, pathways are described as being

downregulated if the sum of fold changes of all genes in the pathway is negative.

231

Steady-state analysis with scBONITA: scBONITA assumes that the observed cellular states 232 233 are defined by states of multiple dynamic cellular processes or signaling pathways. While observed cells 234 are samples along a dynamic trajectory of signaling cascades, analyzing attractors upon randomly 235 sampling the rules from ERS allows us to investigate most common signaling states of a network under 236 consideration. Hence, we sample ten network specific from the ERS inferred by scBONITA-RD to 237 identify a set of reachable attractors. This is achieved by simulating the network synchronously as 238 performed in other studies (48-51) starting from an observed state (i.e., a cell expression vector) until a 239 steady state (or an attractor cycle) is reached. By starting simulations from expression levels of all cells 240 (all observed states), we can ensure that these simulations cover a large fraction of available state space 241 for a given network. In this way, all reachable attractor states, corresponding to observable signaling 242 states, can be identified. The similarity between cells and attractors is quantified using the Hamming 243 distance. Cells are assigned to the attractor that is most similar to their expression data.

Implementation and availability: scBONITA is implemented in Python 3 and C. Source code,
 documentation, and tutorials are available on <a href="https://github.com/Thakar-Lab/scBONITA">https://github.com/Thakar-Lab/scBONITA</a>.

# 246 Application of scBONITA on a publicly available data set

A scRNA-seq dataset obtained from four persons living with HIV (PLWH) before and during infection
was selected to demonstrate the utility of the scBONITA pipeline on other datasets and to compare
signaling dysregulations upon atherosclerosis in PLWH with signaling dysregulations upon HIV infection
(36). Log2-transformed TPM data and metadata processed and curated by the study authors was collected
from the Single-Cell Portal database (https://singlecell.broadinstitute.org/single\_cell/study/SCP256.). The
complete scBONITA pipeline was used to compare samples collected before infection to samples

collected 1 year after infection. We retained the cluster labels assigned by the authors of the original
 study. A set of 210 KEGG networks was used with the scBONITA pipeline. Dysregulated pathways and
 steady states identified by scBONITA were compared to the original analysis, as described in the results.

#### 256 In silico evaluation of scBONITA

257 To show that scBONITA-RD is robust to training set size, we selected a cluster of B cells from the

258 HIV/AS dataset. This subset of the dataset was manipulated to either downsample or augment the size of

the training dataset (number of cells) presented to scBONITA-RD. The training dataset was downsampled

to 1% and 50% of the original number of cells for cluster 0 ("B cells nave - 1"). To augment the dataset

and thereby introduce heterogeneity, the dataset was increased to 200% of its original size by adding in

262 cells from a neighboring cluster of B cells. A set of 210 KEGG networks was used to evaluate the sizes of

the ERS obtained by scBONITA-RD using these manipulated training datasets. The size of the ERS is

used as a proxy for scBONITA's ability to successfully cut down the state space of the possible rules for

265 each node using cross-sectional scRNA-seq data.

# 266 **Results**

# 267 Single-cell sequencing identifies 16 transcriptionally distinct cell 268 subpopulations in PBMCs derived from AS+ and AS- PLWH

269 To investigate dysregulated immune signaling in People Living with HIV (PLWH), who are at an

270 increased risk of atherosclerosis (AS), we recruited a cohort of eight PLWH, four with carotid plaques on

both sides of the arteries (referred here as AS+) and four without carotid plaques (referred as AS-). Lipid

272 profiles and age, both known risk factors for cardiovascular disease, were matched and were not

significantly different between subject groups (see Methods, Supplementary File 1, Supplementary Figure

- 1). We used the 10X Genomics platform to profile transcriptional changes in ~1200 peripheral blood
- 275 mononuclear cells (PBMCs) per subject. This scRNA-seq data was processed using the Cell Ranger and

276	Seurat pipelines (37) as described in the Methods to identify sixteen transcriptionally distinct populations
277	of immune cells (Figure 1A). These clusters were then annotated using CIBERSORT (40) with a dataset
278	of sorted immune cells (Supplementary File 1, Supplementary Figure 2) and by using cell-lineage specific
279	markers (Supplementary Table 1). This scRNA-seq dataset is referred to in the text as the HIV/AS
280	dataset.
281	A population of CD8 T cells/NK resting cells was significantly lower in AS- PLWH and a population of
282	CD14+CD16+ monocytes was significantly higher in AS- PLWH (t-test, p < 0.05) (Figure 1B). The
283	cluster-specific differentially expressed genes, or cluster markers, for these populations of CD14+CD16+
284	monocytes and CD8+ T cells/NK resting cells were involved in pathways linked to cell migration, such as
285	the leukocyte transendothelial migration, regulation of actin cytoskeleton, adherens junction, and
286	chemokine signaling pathways (Supplementary Table 1), suggesting that CD8+ T cells/NK resting cells
287	and monocytes have a migratory phenotype. Indeed these cells are known to migrate into intima during
288	the formation of atherosclerotic lesions in the vascular wall (52-58).

#### 289 Characterization of peripheral blood mononuclear cells in people

### 290 living with HIV with and without atherosclerosis

291 To characterize the expression differences between these cell types in AS+ and AS- PLWH, differentially 292 expressed (DE) genes were identified using the Wilcoxon test as described in the Methods (selected cell 293 types are shown in Figure 1C-F and complete list of DE genes presented in Supplementary Table 2). 294 Genes were reported as being differentially expressed if the adjusted p-value was < 0.05 and the average 295 absolute log fold-change was > 0.3 (representative cell subpopulations shown in Figure 1C-F). Gene set 296 enrichment analysis was performed as described in the Methods. We used gene sets obtained from KEGG 297 and curated in MSigDb. Complete results of the enrichment analysis are presented in Supplementary 298 Table 2.

299 Considered together, differential expression analysis and gene set enrichment analysis provided insights into the functional differences in cells obtained from AS+ and AS- PLWH. Several genes (and gene sets) 300 301 related to cell migration and mobility were upregulated in cells derived from AS+ PLWH. MHC Class 1 302 genes (HLA-A, HLA-B and HLA-C) and MHC Class II genes (HLA-DP, HLA-DQ, HLA-DB, and HLA-303 DR) were upregulated in multiple cell subpopulations. Figure 1E shows the expression of HLA-A and 304 HLA-DRA in B cells naïve -2. Similarly, ITGB2 was upregulated in population a cluster of CD8+ T cells 305 ("T cells CD8 – 3") derived from AS+ PLWH. These cell-surface proteins are all involved in cell-cell 306 interactions and cellular adhesion.-ITGB2 interacts with ICAM2 expressed on the surface of endothelial 307 cells and is involved in leukocyte transendothelial migration. The actin gene ACTB was upregulated in 308 both populations of naïve B cells and a population of CD8+ T cells derived from AS+ PLWH. Similarly, 309 CXCR4 was upregulated in cells derived from AS+ PLWH in two populations of naïve B cells, two 310 populations of CD8+ T cells, and a population of resting NK cells. Both CXCR4 and ACTB play an 311 important role in the process of leukocyte transendothelial migration. CXCR4 activation is known to 312 drive both migration and proliferation of vascular cells. Leukocyte transendothelial migration involves 313 dynamic actin (ACTB) cytoskeleton remodeling, that is in part mediated by the key genes RHOA and 314 ROCK, which are not themselves differentially expressed between cells derived from AS+ and AS-315 PLWH. However, gene sets related to this process were not significantly over-represented in the 316 differentially expressed genes for any subpopulations, with the exception of a small population of NK 317 resting cells.

The S100A8/S100A9 genes are upregulated in T cells CD8/NK resting cells (Figure 1C), monocytes (Figure 1D), T cells CD8/CD4/CD4 naive (Figure 1E) and B cells naive – 2 cluster (Figure 1F) derived from AS- PLWH. S100B is upregulated in CD8+ T cells populations derived from AS- subjects (Figure 1C, 1F). In monocytes, S100B has been shown to engage RAGE leading to overproduction of reactive oxygen species, leading to the modulation of genes involved in inflammatory processes and adhesion to vasculature (59). In addition, several ribosomal genes are upregulated in cells derived from AS- PLWH in

the B cells naïve –2 cluster (RPL8, RPS10, RPS26, RPS4Y1, RPS9), and upregulated in cells derived
from AS+ PLWH in the T cells CD8/CD4/CD4 naïve cluster (RP11-347P5.1, RPL13, RPL39, RPS14,
RPS27, RPS29, RPS6). The ribosome gene set was over-represented in the DE genes for 12 out of the 16
identified subpopulations (Supplementary Table 2). Ribosomal genes have been associated with cellular
states (60) and with the activity of the mTOR pathway (61-64).
While DE and enrichment analysis indicated some mechanisms of HIV-associated atherosclerosis, a
cohesive understanding of how these multiple genes (DE and non-DE) jointly regulate signaling cascades

did not emerge from the analysis above. Hence, we developed the network-based pathway analysis

algorithm scBONITA to further investigate the importance of specific genes and their role in signaling.

## 333 scBONITA algorithm for development of discrete-state models of

#### 334 pathways

335 To investigate dysregulated immune signaling, we used a discrete-state modeling approach to develop 336 executable models of signaling pathways. Specifically, we developed the scBONITA (Single-Cell 337 Boolean Omics Network Invariant-Time Analysis) algorithm using our previously established and 338 validated BONITA algorithm for scRNA-seq data (24) (Figure 2). Briefly, scBONITA uses a Boolean 339 framework where "AND" or "OR" logic gates are optimized to model signal integration and flow through 340 biological networks. scBONITA requires two inputs: (a) a binarized scRNA-seq dataset, and (b) a prior 341 knowledge network (PKN) (Figure 2A). PKNs describe pathway topology defined by nodes representing 342 genes and directed edges representing the activating or inhibitory interactions between nodes. The 343 optimized logic gates for a PKN are referred to as the rule set for that PKN. scBONITA leverages the 344 principle that observed states of single cells correspond to the steady states or attractors of dynamic 345 biological networks to identify regulatory rules for the input PKNs (Figure 2B). A genetic algorithm is 346 used to perform a global search and identify a minimum-error rule set which is further optimized by a 347 node-wise local-search. This procedure returns a set of discrete-state models for pathways referred to as

the equivalent rule set (ERS). A pathway is described in the text as 'optimized' if scBONITA-RD
successfully cuts down the state space of the possible rules for at least one node in the pathway.

350 The discrete-state models learned by scBONITA can be simulated to generate time-course trajectories 351 that are analogous to signal flow. They can also be perturbed *in silico* by simulation of gene knock-ins 352 and knockouts. We quantify the difference between network states after knock-out and knock-in, and 353 weight the score by the size of the ERS to quantify the uncertainty in rule determination for each node to 354 calculate a node-importance score that describes the influence of each node over the network. scBONITA 355 uses these node importance scores and comparison-specific fold changes from the scRNA-seq data used 356 for training to identify dysregulated pathways in a pre-specified contrast (Figure 2C). The simulation 357 trajectories of these discrete-state models fall into steady states known as attractors. These attractor states 358 of biological networks are the cell states with respect to specific pathways and have been hypothesized to 359 correspond to signaling behavior characteristic of specific cell types. Cells are assigned to the attractor 360 that is most similar to their expression data (Figure 2D). In this way, cells are matched to characteristic 361 signaling states for the network under consideration. Thus, scBONITA allows in depth investigation and 362 simulation of known biological signaling pathways by incorporating network topology.

### 363 scBONITA identifies dysregulated pathways in T cell populations in

### 364 people living with HIV stratified by atherosclerosis

The scBONITA pathway analysis algorithm identifies dysregulated pathways in all subpopulations
derived from AS+ and AS- PLWH, providing an insight into mechanisms of atherosclerosis development
in PLWH (Supplementary Table 3). ScBONITA optimized multiple pathways in CD8+ T cells that are
involved in this proinflammatory and anti-viral process (Figure 3A, Supplementary Table 3). These
pathways include the JAK-STAT signaling pathway, the TGFβ signaling pathway, the NFκB signaling
pathway, the AGE-RAGE signaling pathway, and the HIF-1 signaling pathway. In conjunction, cellmigration pathways such as the regulation of actin cytoskeleton and axon guidance, the PI3K-AKT

372 signaling pathway and mTOR signaling pathway that modulate apoptosis and cell migration, are 373 identified as being dysregulated in the HIV+AS+ vs HIV+AS- contrast (Figure 3A). Only the Th17 cell 374 differentiation pathway was overall upregulated in cells derived from AS+ PLWH. All these pathways 375 were newly identified by scBONITA, i.e., the corresponding gene sets were not identified as being 376 enriched in the AS+/AS- contrast by enrichr (Supplementary Table 2). 377 Pathway analysis with scBONITA identified multiple optimized pathways in subpopulations containing 378 CD4+ T cells (T cells CD4 memory resting/T cells CD8, T cells naïve, and T cells CD4 -1 clusters, 379 Figure 1A) as being dysregulated (Bonferroni-adjusted p-value < 0.01) in the HIV+AS+ vs HIV+AS-380 contrast (Figure 3B, Supplementary Table 3). CD4+ T cells may exert either an atherogenic or 381 atheroprotective phenotype, depending on subset and interactions with antigen presenting cells (APCs), 382 such as B cells in the adventitia or macrophages in plaques (65). Of the dysregulated pathways, herpes 383 simplex virus 1 infection and MAPK signaling pathways were overall upregulated in T cells CD4 384 memory resting/ T cells CD8 and T cells CD8/CD4/CD4 naïve cells derived from AS+ PLWH. The 385 proteoglycans in cancer and mTOR signaling pathways were overall upregulated in T cells CD4 memory 386 resting/ T cells CD8 from AS+ PLWH. PI3K-AKT signaling and chemokine signaling pathways were 387 upregulated in the T cells CD8/CD4/CD4 naïve cluster derived from AS+ PLWH. The 'proteoglycans in 388 cancer' signaling pathway is involved in cell adhesion and migration is upstream of processes known to 389 be dysregulated in atherosclerosis, such as regulation of the actin cytoskeleton, mTOR signaling, and 390 apoptosis. Similarly, the herpes simplex virus infection process involves activation of the PI3K-AKT 391 signaling and apoptosis pathways, which are known to be relevant in atherosclerosis. 392 The AGE-RAGE signaling pathway was further investigated in CD8+ T cells due to their relevance in 393 atherosclerosis and to demonstrate the additional information obtained from scBONITA in comparison to 394 other enrichment methods. The AGE-RAGE signaling pathway in a cluster of CD8+ T cells had the 395 highest pathway modulation score (0.8) amongst all tested pathways for this cluster (Supplementary Table 396 3). Most of the genes in this pathway were higher in AS+ PLWH (Figure 3C). Furthermore, the

397 uncertainty score was used to identify pathways where logic gates for genes could be optimized. 398 Particularly, scBONITA learned rules for the highly central DIAPH1 node, which has a high influence 399 over signal flow in this network due to its high connectivity (uncertainty score= 0.5) (Figure 3C). The 400 combination of scBONITA's node importance score and a difference in expression between the 401 HIV+AS+ and HIV+AS- groups was used to identify key genes which may or may not have a significant 402 difference in expression but whose activity influences the flow of signal through the network and the 403 signaling behavior in atherosclerosis. These key genes in AGE-RAGE pathway had higher expression in 404 AS+ PLWH. scBONITA assigns the class 1 PI3K genes (PIK3CA, PIK3CB and PIK3CD genes), the 405 P13K regulator PI3KR1, and PLC genes (PLCB1, PLCB2) maximal importance scores in this network. 406 All these genes are highly expressed in AS+ PLWH (Figure 3C). PI3K is known to activate intracellular 407 pathways involved in the pathophysiology of atherosclerosis, such as lipid accumulation and transport, 408 macrophage autophagy, phenotypic transition, and the expression of adhesion molecules involved in the 409 inflammatory response (reviewed in (66)). As noted in the preceding section, P13K is also downstream of 410 CXCR4, which is significantly upregulated in cells derived from AS+ PLWH. While CXCR4 is not itself 411 in this network topology, scBONITA nevertheless identifies its downstream effectors PI3K and PLC, 412 which are not themselves significantly differentially expressed, as playing an important role in the 413 signaling pathways leading to the expression of atherosclerosis-related genes and hence and 414 atherosclerotic phenotype. In this manner, pathway analysis with scBONITA revealed that several 415 pathways and genes associated with atherosclerosis are dysregulated in T cells derived from PLWH. 416 Many of these pathways have been previously shown to regulate T cell migration during atherosclerosis.

#### 417 scBONITA identifies dysregulated pathways in monocytes in people

### 418 *living with HIV, stratified by atherosclerosis*

Pathway analysis with scBONITA identified multiple optimized dysregulated pathways in monocytes
(Figure 4A, Supplementary Table 3). Key pathways include the apoptosis, cAMP signaling, leukocyte

transendothelial migration, PI3K-AKT signaling and cellular senescence pathways, which appear to be
involved in the proinflammatory behavior of proatherogenic monocytes. (53, 67-76). Only the cAMP
signaling pathway and the endocrine resistance pathway are overall upregulated in cells derived from
AS+ PLWH.

425 The leukocyte transendothelial migration pathway was further investigated as monocyte migration outside 426 the vascular compartment plays a crucial role in the inflammatory cascade that leads to an atherosclerotic 427 phenotype (67, 69, 77). In addition, scBONITA was able to learn a limited set of regulatory rules for the 428 influential, highly connected RHOA gene (uncertainty factor=0.13). This pathway had a pathway 429 modulation score of 0.45, which is the third-highest pathway modulation score amongst tested pathways 430 for this cluster (Supplementary Table 3) (Figure 4B). scBONITA also assigned high importance scores to 431 the NCF genes (NCF1, NCF2, and NCF4), CYBA and CYBB. NCF genes are involved in superoxide 432 production and are a positive regulator of P13K signaling (78, 79). Once again, PLCG1 and PLCG2 are 433 assigned high importance scores by scBONITA in this network, underscoring the importance of PLC in 434 proinflammatory and proatherogenic processes. The upstream regulator of PLCG1, MSN, is involved in 435 cytoskeletal remodeling during leukocyte migration and is similarly assigned a high importance score 436 (80). ROCK2 shows only a small change across AS groups, which may be driven by feedback regulation 437 of mRNAs of ROCK genes but has a high importance score indicating a stronger role in regulating the 438 signal flow. The G protein GNAI3, which is downstream of CXCR4, is also assigned a high importance 439 score, possibly indicating a role for CXCR4-mediated activation of this pathway despite the low observed 440 fold change of CXCR4 between the AS+ and AS- groups. The downstream effectors of these high-441 importance genes have higher fold changes than the high-importance genes themselves. These genes 442 include ACTG1 and EZR, which are involved in cytoskeletal remodeling (80-82), and ITGA4, ITGB1, 443 and ITGB2, which are involved in cell adhesion. As in the case of CD8+ T cells, pathway analysis with 444 scBONITA identifies several genes known to be associated with atherosclerosis which are also 445 dysregulated in monocytes derived from PLWH.

#### 446 Pathways dysregulated by HIV infection are implicated in

#### 447 *atherosclerosis*

456

448 To identify the biological mechanisms modulated by HIV infection we analyzed PBMCs from four

449 individuals before and during acute HIV infection. Specifically, Kazer et al (36) sequenced PBMCs from

450 4 individuals before and during acute HIV infection and identified gene expression programs activated in

451 proinflammatory T cells, monocytes, and NK cells during HIV infection. Pre-clustered scRNA-seq data

and the set of KEGG networks used in the above analysis were used with scBONITA to infer Boolean

rules and perform pathway analysis. We compared pathways that were dysregulated pathways after 1 year

454 of HIV infection (36) to those identified as being dysregulated ( $p_{adj} < 0.1$ ) in the AS+HIV+ vs AS-

455 HIV+ contrast from our dataset described above (Figure 6A). Subpopulations of immune cells in the two

datasets were matched for the purpose of this comparison as shown in Supplementary Table 6, except for

457 dendritic cells, which did not have a matching subpopulation in the HIV/AS dataset.

458 scBONITA identified 10 optimized pathways that were dysregulated after 1 year of HIV infection in 459 cytotoxic T cells (36) (Figure 3A-B, Figure 6A, Supplementary File 1, Supplementary Figure 5, 460 Supplementary Table 5, Supplementary Table 6). Of these pathways, the 'Axon guidance', 'MAPK 461 signaling', 'Proteoglycans in cancer', 'Herpes simplex virus 1 infection', and 'Cytokine-cytokine receptor 462 interaction' signaling pathways were also dysregulated in PLWH with AS. Similarly, 5 out of 19 463 optimized pathways were dysregulated in monocytes upon HIV infection and in AS+ individuals (Figure 464 4A, Figure 6A, Supplementary File 1, Supplementary Figure 5, Supplementary Table 6). The herpes 465 simplex virus 1 infection, PI3K-AKT signaling, and cellular senescence pathways were dysregulated 466 upon HIV infection and in PLWH with AS. While overall dysregulation of the pathways is a crude metric 467 for activation, we note that 81 and 41 genes from these overlapping pathways were upregulated in both 468 contrasts in the cytotoxic T cell populations and monocyte population respectively (Figure 6B and 6C, 469 statistical significance was not tested). Enrichment analysis using KEGG biological pathways suggests 470 that the genes upregulated after HIV infection and in AS+ PLWH in the CD8+ T cell subpopulation were

471 enriched for viral response pathways such as herpes simplex virus 1 infection and human cytomegalovirus infection (Supplementary File 1, Supplementary Figure 6). Similarly, the genes upregulated after HIV 472 473 infection and in AS+ PLWH in the monocyte subpopulation were enriched for cell migration related 474 pathways such as the PI3K-AKT signaling pathway and the cAMP signaling pathway (Supplementary 475 File 1, Supplementary Figure 6). In this vein, mTOR signaling was dysregulated in T cells from both 476 datasets and proteoglycans in cancer, PI3K-AKT signaling, and cellular senescence were dysregulated in 477 B cells from both datasets. These dysregulated pathways suggest that the modulation of cell migration and 478 inflammation processes upon HIV infection progresses over time, leading to AS in PLWH.

#### 479 Attractor analysis reveals PI3K genes driving distinct cellular

## 480 signaling states potentially associated with atherosclerosis

481 We performed attractor analysis for the pathways significantly dysregulated between AS+ and AS-

482 PLWH in the cluster of CD8+ T cells discussed above (Cluster CD8 T cells -1 in Figure 1A). The 483 attractor analysis facilitates evaluation of cellular states across subjects and disease groups. The simplest 484 rules, which have the smallest number of "AND" terms, were chosen to simulate the network and identify 485 attractors as described in the Methods. The insulin resistance pathway, which is downstream of the AGE-RAGE signaling and the PI3K-AKT pathways, was particularly interesting because more than 50% of the 486 487 same cluster of CD8+ T cells mapped to three dominant attractors. scBONITA identified 72 attractors 488 representing cellular states with respect to the insulin resistance pathway in CD8+ T cells. Of these 72 489 signaling states, 3 dominant signaling states mapped to 16.5%, 7.5% and 27.4% of cells respectively 490 (Figure 5A-B). The insulin resistance pathway was found to have a significant association between 491 identified attractors and subjects (chi-square test, p-value < 0.05) but no significant association between 492 identified attractors and atherosclerosis status (chi-square test, p-value > 0.05). Notably, attractor analysis 493 showed that there was significant association between assigned attractors and subject from which cells 494 were derived for the T cells CD4 - 1 and T cells CD8/CD4/CD4 naïve clusters for the chemokine signaling

495	pathway (chi-square test, $p < 0.01$ ). The attractors of the PI3K-AKT signaling pathway were significantly
496	associated with subject for the T cells CD8/CD4/CD4 naive cluster (chi-square test, $p < 0.01$ ).
497	These cellular states were characterized by differences in several key genes (Figure 5C), including PI3K
498	genes (PI3KA, B, and D) and the PI3K regulators (PIK3R1, 2, and 3) that were identified as being highly
499	influential in the AGE-RAGE signaling pathway (Figure 3B). In addition, the two less abundant attractors
500	differed in the activity of the key TNFR and TNF genes. Hence, these attractors are referred to as the
501	PIK3R+ PIK3+, TNFR+TNF- and TNFR-TNF+ attractors. The activity of these PI3K genes and the
502	activity of AKT genes (AKT1, 2 and 3) was higher in the most common signaling state (PI3KR+ PI3K+
503	attractor). However, the activity of the downstream targets of AKT, such as CREB1, CREB3, CREB5,
504	NFKB1, FOXO1, CREB3L4 and CREB3L42 were lower in the PI3KR+ PI3K+ attractor. TNF, which is
505	known to be produced at a low level by some subsets of T cells, also mediates a range of pro-
506	inflammatory processes in vascular endothelial cells, particularly leukocyte adhesion and transendothelial
507	migration (83, 84). TNF-TNFR1 signaling mediates an apoptotic process that is mediated by TRADD and
508	FADD (85). In addition, TNFR1 is known to activate the PI3K signaling pathway in regulatory T cells in
509	the context of autoimmune disease (86). These differences may indicate differences in TNF production
510	and response in the cells that are assigned to these cell states. Significant differences in signaling modes
511	of cells exist across subjects, suggesting the existence of distinct modes of operation (Figure 5D).
512	Attractor analysis of the leukocyte transendothelial migration pathway in monocytes revealed 9 attractors
513	that mapped to cells in the dataset and two dominant signaling modes mapping to 51.04% and 30.73 % of
514	cells respectively (Figure 6A-B). The two dominant signaling modes of this pathway in monocytes
515	differed in the activity of the PECAM1 and F11R genes (Figure 6C) and are hence referred to as the
516	PECAM+ and F11R+ attractors. F11R is required for platelet adhesion to vascular endothelial cells (87),
517	which occurs prior to infiltration of monocytes into the endothelium and eventual plaque formation (88,
518	89). PECAM1 has been recognized as an influential signaling molecule with widespread effects on
519	vascular biology and atherosclerosis in particular (90-92). Attractors for this pathway were significantly

associated with the subjects (chi-square test, p-value < 0.05, Figure 6 B), but were not significantly</li>
associated with atherosclerosis status (chi-square test, p-value > 0.05). Similarly, the attractor activity of
individual genes was not significantly associated with atherosclerosis status (t-test, p-value > 0.05).
However, we observed significant variation in attractor activity for PECAM1 and F11R across subjects,
reflective of the different states of the cells across subjects. Thus, scBONITA allows us to investigate
cellular states potentially associated with inter-subject variability, driven by molecular signaling.

## 526 Evaluation of scBONITA performance in silico

527 The BONITA algorithm has already been rigorously validated in our previous study (24). Specifically, 528 comparison with other network-based pathway analysis tools has been performed. Here we evaluate 529 scRNA-seq specific components of the algorithm. To show that scBONITA rule determination is robust 530 to training set size, we varied the size of the training data provided to scBONITA. Specifically, the 531 number of cells in the largest cluster of cells (Naïve B cells -1) were varied by random selection from 1% 532 of cells from that cluster to 200% by adding cells from neighboring clusters. The reduced size of the ERS 533 for nodes with in-degree 3 (i.e., the most complex case considered by scBONITA) informs improved 534 certainty in rule inference by scBONITA. (Figure 7A) While there was a significant decline in 535 performance when the data was downsampled to 1% of the original cluster, there was no significant 536 increase in effect once 50% of the cells were used, or when the training dataset was augmented. This 537 indicates that scBONITA is robust to heterogeneity in the training data set. To show that scBONITA's 538 node importance score is not correlated to commonly used measures of node centrality, we compared the 539 node importance scores calculated by scBONITA for KEGG networks when it is trained on the HIV/AS 540 dataset as described above to six centrality measures – Katz centrality, degree centrality, current flow 541 centrality, eccentricity centrality, betweenness centrality, and local reaching centrality. We observed that 542 scBONITA's node importance score is not correlated to any of these metrics (Spearman Correlation 543 Coefficient < 0.15, Figure 6H). To show that network topology also significantly influences the node 544 importance score, we compared scBONITA's node importance scores for the same set of KEGG

545	pathways, assigned using different training datasets. These node importance scores were compared using
546	similar cell subpopulations only. We found that the node importance scores between the two datasets
547	were correlated as shown by a representative comparison between the node importance scores for the
548	subpopulation of cytotoxic T cells from the Kazer et al dataset and the subpopulations of CD8+ T cells
549	(Figure 6B, $0.71 < Pearson Correlation Coefficient < 0.91, p < 0.01)$ show .Similarly, the node
550	importance scores for the populations of monocytes were highly correlated (Pearson correlation
551	coefficient = 0.78, Supplementary Table 4, Supplementary File 1, Supplementary Figure 3). However, the
552	correlations were relatively lower for other pairs of subpopulations (Supplementary Table 4,
553	Supplementary File 1, Supplementary Figure 5), indicating that scBONITA learns some characteristic
554	features of a network topology, but node importance scores are still assigned in a context-dependent
555	manner.

# 556 **Discussion**

557 Among people living with HIV, widespread use of cART has significantly reduced overall mortality. 558 However, the earlier and increased incidence of cardiovascular diseases, including atherosclerosis, 559 remains the major cause of mortality in an aging HIV+ population. The causes for this are manifold and 560 include side effects of cART and lower level of HIV proliferation (4, 9-12). We and others have 561 attempted to identify the immune signaling mechanisms that lead to this increased incidence of 562 atherosclerosis (93). However, to the best of our knowledge no study has attempted to holistically study 563 the combined effects of variance in the numbers of specific immune cells and the cell-type specific 564 signaling dysregulations (93). We found that in accordance with previous studies (19, 94-100), a 565 population of CD8+ T cells was significantly increased in PBMCs from AS+ PLWH. However, contrary to our expectations (20, 68, 69, 74, 76, 101-104), a population of monocytes was significantly decreased 566 567 in PBMCs from AS+ PLWH. Interestingly, evaluation by our pathway analysis algorithm scBONITA and 568 differential expression analysis identified differences in the migratory phenotype of these cells, 569 suggesting to us that these monocytes are migrating into the vascular intima in AS+ PLWH (described

570 later in the discussion). Conventional differential expression and gene set enrichment analysis methods 571 identified genes involved in signaling pathways that are known to be linked to cell migration and cell 572 aging. Key amongst these identified genes were CXCR4 and ACTB (Figure 1 C-F) which were 573 upregulated in populations of CD8+ T cells derived from AS+ PLWH. CXCR4 activation by its ligands 574 leads to the activation of phosphatidylinositol-3-OH kinases (PI3K), which in turn leads to the activation 575 of the serine-threonine kinase AKT via PIP3 (105). PI3K/AKT signaling leads to multiple processes 576 involved in plaque formation, such as cell migration, intracellular lipid accumulation, and smooth muscle 577 cell proliferation (66). S100A8/S100A9 were found to be upregulated in monocytes derived from AS-578 PLWH (Figure 1 C-F). In monocytes, S100A8/9 have been shown to increase adhesion, migration, and 579 production of inflammatory cytokines such as TNF-alpha and IL1 $\beta$  (106, 107). However, none of these 580 processes were identified as being significantly enriched in the differentially expressed genes from these 581 subpopulations. In addition, these methods failed to provide insights into how disparate genes involved in 582 different pathways regulate cellular states. To characterize signaling dysregulations in HIV-associated 583 atherosclerosis more effectively, we developed the scBONITA algorithms for regulatory rule inference, 584 network simulation, pathway analysis, and attractor/steady-state analysis. 585 scBONITA learns condition-specific logic models using scRNA-seq data in conjunction with published 586 prior knowledge networks. This study builds on our previously published BONITA method (24) that 587 inferred logic rules from bulk RNAseq data. scBONITA exploits the bimodal nature of scRNA-seq data 588 (38, 108) and the cell-level resolution of expression to successfully learn regulatory rules and identify 589 attractors for prior knowledge networks. We show that scBONITA can successfully learn regulatory rules 590 for biologically significant signaling pathways and that these rules can be used to perturb and simulate 591 these pathways in silico. Unlike other tools utilizing Boolean networks, scBONITA is not dependent on 592 time-series data and in fact hypothesizes that scRNA-seq data represents samples in the state space of a 593 dynamic Boolean network. In addition, scBONITA uses published network topologies, thereby reducing 594 the uncertainty in the inferred rules. Other groups have published algorithms to infer logic rules and

595 reconstruct gene-regulatory networks on a small subset of genes from scRNA-seq data (51, 109, 110). 596 However, scBONITA does not depend on pre-selection of genes. Additionally, scBONITA also identifies 597 dysregulated signaling pathways in a given context, combining expression information with scBONITA 598 derived impact score to create a unique metric of pathway dysregulation that considers (a) the dynamic 599 nature of signaling pathways and (b) the impact of node perturbations. 600 scBONITA identifies interesting, dysregulated pathways across all cell subpopulations in the HIV/AS 601 dataset (Supplementary Table 3, Figures 3-4). The AGE-RAGE signaling pathway was significantly 602 dysregulated in a population of CD8+ T cells (Figure 3A). AGE-RAGE signaling elicits activation of

multiple intracellular signaling pathways such as cell proliferation and apoptosis pathways (111-119).

604 scBONITA assigned the highest importance scores for this network to the PI3K family of genes

605 (PIK3CA, PIK3CB and PIK3CD), which promote intracellular lipid deposition leading to the formation

of foam cells and atherosclerotic plaques and can also reduce the expression of lipid transporters and

reduce the efflux of intracellular cholesterol depending on upstream signals (66). PLC (PLCB1 and

608 PLCB2 genes), which is also assigned a high importance score in this network, facilitates

609 proinflammatory and proatherogenic processes. PLC activity induced by oxidized low-density lipids

610 (oxLDL) serves to stimulate proinflammatory IL-8 secretion and promote leukocyte adhesion, promoting

611 plaque development. PLC also enhances endothelial dysfunction and plaque progression by inducing

612 VEC apoptosis (80, 120). Of note, both PI3K and PLC genes were upregulated in AS+ PLWH (Figure

613 3C).

614 ScBONITA identified several pathways linked to lipid metabolism as being dysregulated in the

615 population of monocytes. Amongst these pathways were the cAMP signaling pathway, which was overall

616 upregulated in AS+ PLWH, and leukocyte transendothelial migration pathway, which was overall

617 upregulated in AS- PLWH, both of which are involved in the infiltration of monocytes into the intima

618 during the formation of atherosclerotic lesions and hence progression of atherosclerosis (52-57).

619 Similarly, scBONITA identified genes critical to the atherosclerotic process in the leukocyte

620 transendothelial migration pathway trained on the monocyte subpopulation. Specifically, ROCK1 and 621 ROCK2 are stimulated by atherogenic stimuli, such as oxLDL. ROCK activation leads to various 622 pathophysiological changes include endothelial dysfunction, migration and angiogenesis, migration, 623 proliferation and differentiation in SMCs and vascular remodeling (121, 122). ROCK inhibitors such as 624 statins have been recently shown to attenuate atherosclerosis by inhibiting ROCK1 and ROCK2 and 625 consequently inhibition of altered chemotaxis of macrophages and its transformation into foam cells 626 (123). The dysregulation of the glucagon signaling pathway, cAMP signaling, PI3K-AKT signaling and 627 proteoglycans in cancer pathways indicate that in AS+ PLWH, dysregulations in glucose metabolism 628 induce expression of adhesion molecules by the vascular endothelium resulting in increased monocyte 629 transendothelial migration (66, 124). In this manner, scBONITA revealed novel insights into pathway 630 regulation upon atherosclerosis.

631 Pathway analysis with scBONITA shows that lipid metabolism and cell migration are dysregulated in 632 AS+ PLWH across all cell subpopulations. This effect is most clearly seen in the case of the PI3K-AKT 633 signaling pathway, which is dysregulated in all cell subpopulations (Figures 4-5, Supplementary Table 634 3). The pathways that are upstream and downstream of PI3K-AKT signaling are, however, dysregulated 635 in different subpopulations, suggesting that the activation and effector mechanisms of this signaling 636 cascade vary by cell type. While there appears to be no strong evidence for apelin expression in B cells, 637 the apelin signaling pathway was upregulated in B cells naïve -1 derived from AS+ PLWH and 638 adipocytokine signaling pathway was downregulated in B cells naïve -2 derived from AS+ PLWH, along 639 with PI3K-AKT signaling and MAPK signaling. The cardioprotective effect of apelin is modulated by 640 (amongst other routes) the PI3K-AKT signaling and MAPK signaling pathways (80-82). It is also shown 641 to be upregulated in human atherosclerotic coronary arteries and colocalized with markers for 642 macrophages (125, 126). PI3K-AKT signaling is also dysregulated in CD4+ T cells (Figure 4B), as is its 643 upstream signaling pathway cytokine-cytokine receptor interaction, suggesting a different mechanism of activation of PI3K-AKT signaling in this cell type. 644

645 We used the scBONITA pipeline to infer pathways involved in the progression of HIV infection from an 646 independent, publicly available scRNA-seq dataset derived from 4 subjects before and after HIV infection 647 and compared these pathways to those identified in the progression of atherosclerosis in PLWH. 648 scBONITA-RD was able to successfully learn regulatory rules for KEGG pathways when trained on this 649 dataset (Supplementary File 1, Supplementary Figure 4). Several pathways dysregulated upon HIV 650 infection were also dysregulated in PLWH with AS (Figure 3A, Figure 4A, Supplementary File 1, 651 Supplementary Figure 5, Supplementary Table 6). Among the dysregulated pathways, the proteoglycans 652 in cancer pathway, which is linked to cell migration and adhesion, and the axon guidance pathway, which 653 is linked to cytoskeletal reorganization and Rho GTPase signaling and has been suggested to be a special 654 case of cell migration that is active in many cell types, are both also suggestive of changes occurring in 655 cell migration of cytotoxic T cells due to HIV infection (127-130). Interestingly, both these pathways are 656 also linked to the cytokine - cytokine receptor interaction pathway. cAMP signaling negatively regulates 657 the production of PI3K-AKT signaling and hence the transcription of pro-inflammatory cytokines via 658 PKA, and positively regulates the transcription of anti-inflammatory cytokines via PKA and CREB. 659 Decreased PI3K-AKT signaling results in decreased cell migration (131). The PI3K-AKT signaling, 660 cAMP signaling, and cellular senescence pathways were dysregulated in the population of monocytes 661 derived after one year of HIV infection and in monocytes derived from PLWH with AS, further indicating 662 an HIV-induced dysregulation in cell migration processes. These conclusions were further borne out by 663 the enrichment analysis of genes from these pathways that were upregulated after HIV infection and in 664 atherosclerosis (Figure 7).

To map cells to distinct signaling modes of the pathways described above, we developed scBONITA's attractor analysis capabilities. Attractors are regions in the state space of a dynamic system towards which simulation trajectories are "pulled" or attracted. These attractors and the propensity of simulation trajectories to end in these attractors (i.e., the size of the attractor basins) are characteristics of a specific network with a specific set of regulatory rules. Previous works have suggested that these steady states

correspond to observable cell states, or hallmarks of specific phenotypes such as cell type differentiation, 670 671 disease state, or drug treatment (132-137). These studies show that even simple dynamic models capture 672 rich and nuanced cell behaviors. ScRNA-seq allows the study of these dynamic landscapes and their 673 steady states or attractors at an unprecedented resolution (137-139). scBONITA uses learned rules to map 674 cells to characteristic signaling states based on the property of Boolean networks to move towards steady 675 states or attractors. This attractor analysis allows users to characterize cells based on the dynamic 676 properties of signaling networks, which dictate their phenotype. scBONITA not only identifies these 677 attractors, but it also allows identification, of the master regulators or switches that control the changes 678 between these cell states by evaluating scBONITA's node importance score. Thus, this attractor analysis 679 provides complex insights into cellular processes in a variety of conditions. 680 The importance of cell migration and dysregulated lipid signaling in the development of HIV-associated 681 atherosclerosis was further underscored by attractor analysis with scBONITA. We identified attractors for 682 all signaling pathways that were identified as being dysregulated upon AS in PLWH in CD8+ T cells and 683 monocytes, and mapped cells to these identified signaling states. Furthermore, we identified signaling 684 pathways in which (a) more than one attractor mapped to a significant proportion of cells from the cluster 685 under consideration and (b) there was a significant association between the attractor to which cells were 686 assigned and the subject from which these cells were derived. We note that in most cases, multiple 687 theoretically possibly signaling modes (attractors) were identified but only one dominant signaling state 688 existed in the dataset. Using the criteria defined above, we selected the insulin resistance pathway in 689 CD8+ T cells and the leukocyte transendothelial migration pathway in monocytes for further analysis 690 (140-144). We also note that the insulin signaling pathway exerts immunomodulatory effects on T cells. 691 Decreased insulin receptor expression (used as a proxy for insulin resistance) during viral infection has 692 been shown to downregulate components of the infection response, such as proliferation, cytokine 693 response and glycolysis (140). The three dominant signaling modes of the insulin resistance pathway 694 differed in the activity of PI3K and AKT genes, which were identified as key regulators of the pathways

695 described in the preceding sections, and their downstream effectors, such as CREB and FOXO1. This 696 suggests the existence of two distinct modes of operation for this signaling pathway corresponding to a 697 proliferative cell state (activation of PI3K and AKT) and a senescent cell state (transcription of CREB-698 and FOXO1-controlled genes) (141, 142, 145). Similarly, we identified two dominant signaling modes for 699 the leukocyte transendothelial migration pathway in monocytes. These modes differed in the activity of 700 F11R and PECAM1 genes, which were active in the most common and second most common attractor 701 respectively. While there is no significant separation between cells derived from AS+ and AS- PLWH 702 based on the attractors that they map to, the existence of these two signaling modes suggests variation in 703 the cell states with respect to this pathway in PLWH. Dysregulation in insulin signaling promotes 704 PECAM1-mediated migration of monocytes through endothelial cells (146, 147). Emerging evidence 705 suggests that PECAM1's loss contributes to atherosclerosis (92). The presence of a signaling state in 706 which PECAM1 is deficient may therefore indicate that those cells are contributing to a negative 707 feedback loop of the inflammatory process. The identification of pathway-specific signaling modes that 708 can be differentiated based on specific gene activities shows that scBONITA can shed light on signaling 709 mechanisms that are not apparent by analysis of gene expression differences alone. 710 Although BONITA algorithm has been rigorously validated in our prior publication (46) here we wanted 711 to evaluate the scRNA-seq data specific parts of the algorithm. We next assessed whether the provided

network topology also significantly influences the node importance score, and the dependence of the node

713 importance score on the training data. We demonstrated the influence of network topology on the node

714 importance score (Figure 7B), which is also an indirect measure of the similarity (though not identity) of

the rules inferred by scBONITA for a given network, using different datasets. Thus, scBONITA can

716 identify characteristic structural properties of networks and use this in conjunction with expression

717 information to identify dysregulated pathways in a specified condition. We also show that scBONITA's

node importance score is not correlated with measures of node centrality (Figure 6H). We also assessed

the dependence of scBONITA on the training data to show that scBONITA-RD can narrow down the vast

720 possible state space for a Boolean network (Figure 6A). This ability is dependent on the similarity of the 721 signaling process operating in the cells of the training dataset. scBONITA's ability to resolve rules 722 decreases slightly when the heterogeneity of the training dataset increases (Figure 7E, 200% cells). This 723 restricts the ability of scBONITA to identify regulatory rules that can be experimentally verified. We 724 expect that this capability will improve significantly when pure cell populations are sequenced, for 725 example after flow cytometry or pure cell culture. While scBONITA is not strictly dependent on the 726 clustering method used to classify scRNA-seq data into subpopulations, we used pre-classified 727 subpopulations to reduce variability and, additionally, to improve the specificity of scBONITA-PA by 728 restricting the identification of dysregulated pathways to cell types. This is consistent with the typical 729 gene set enrichment analysis in the context of scRNA-seq data. Additionally, scBONITA-RD requires a 730 longer runtime (<12 hours in our tests) and more powerful computational capabilities than a typical 731 analysis pipeline run on scRNA-seq dataset of typical size; however, these resources are usually available 732 to academic users on computing clusters.

733 The biological variability and nonspecific distortions of expression due to the technicalities of scRNA-seq 734 has been previously identified (148), leading to identification of fewer differences across conditions, as in 735 studies such as (149). In addition, even differential expression methods that are sensitive to the known 736 characteristic distributions of scRNA-seq data are prone to false discoveries (recently reviewed in 737 34584091). Our approach implemented in scBONITA allows analysis of genes in the context of their 738 function in the signaling pathways and network topology of the interactions. This additional information 739 minimizes the impact of the caveats in the scRNA-seq technology mentioned above. Thus, our analysis 740 reveals influential genes in several signaling networks in a context-specific manner, thereby predicting 741 novel targets for further experimental validation or for therapies.

# 742 Conclusions

743	To study the cellular and immunological processes involved in HIV-associated atherosclerosis, we used
744	scRNA-seq to profile PBMCs from 8 PLWH, 4 of whom had atherosclerosis. We developed the
745	scBONITA algorithm to use this scRNA-seq data to (a) infer regulatory rules for networks with a known
746	topology, (b) perturb and simulate these networks in silico to identify master regulators of these networks,
747	(c) combined this topology-specific information with expression information to identify dysregulated
748	pathways in this condition, and (d) grouped cells into characteristic signaling states based on the dynamic
749	properties of these networks. We validated scBONITA on a publicly available dataset of PBMCs from
750	persons before and after HIV infection. scBONITA identified key dysregulated pathways that drive
751	inflammation in people living with HIV. The scBONITA source code, along with documentation and
752	tutorials, is freely available on https://github.com/Thakar-Lab/scBONITA.

### **List of abbreviations**

AS: atherosclerosis; HIV: human immunodeficiency virus; PLWH: people living with HIV; PBMC:

peripheral blood mononuclear cell; scBONITA: single-cell Boolean Omics Network Invariant Time

Analysis; single-cell RNA sequencing: scRNA-seq; CVD: cardiovascular disease; ORA:

- 757 overrepresentation analysis; cART: combined antiretroviral therapy; KEGG: Kyoto Encyclopedia of
- 758 Genes and Genomes; scBONITA-RD: single-cell Boolean Omics Network Invariant Time Analysis rule
- determination; scBONITA-PA: single-cell Boolean Omics Network Invariant Time Analysis pathway
- analysis; GA: genetic algorithm; LS: local search; RA: relative abundance; ERS: equivalent rule set;
- 761 PKN: prior knowledge network

# 762 Acknowledgements

763 We would like to thank Adam Cornwell, Jiayue Meng and George Kassis for testing scBONITA at

various stages of development and for useful discussions on scBONITA functionality. We would also like

to thank Alan Grossfield, Andrew McDavid, Gourab Ghoshal, David Mathews, Lauren Benoodt, Raven
M. Osborn, and all past and present members of the Thakar Lab for helpful discussions. The Center for
Integrated Research Computing at the University of Rochester provided high-performance computing
resources and computing expertise. We are especially grateful to the participants in the HIV/AS study,
their families, and the clinical team.

# 770 Funding

- 771 The study was supported by U.S. National Institutes of Health. MGP is supported by R01 AI134058. RP
- is supported by T32 GM07356. JT was supported by UM1 AI069511, P30 AI078498, R01 AI134058 and
- 773 R21 AI136668. SBM, GS and AT were supported by R01 HL123346, SBM and MVS are supported by
- R01 HL128155, R01 NS066801. The University of Rochester Center for AIDS Research (UR-CFAR;
- P30 AI078498) provided support and core facilities.

# 776 **Declarations**

### 777 Authors' information

#### 778 Affiliations

- <sup>1</sup>Biophysics, Structural, and Computational Biology Program, University of Rochester School of
- 780 Medicine and Dentistry, Rochester, USA.<sup>2</sup> Medical Scientist Training Program, University of Rochester
- 781 School of Medicine and Dentistry, Rochester, USA.<sup>3</sup> University of Rochester Clinical & Translational
- 782 Science Institute. <sup>4</sup>Department of Microbiology, Immunology and Tropical Medicine, George
- 783 Washington University School of Medicine and Health Sciences, Washington, D.C., USA. <sup>5</sup>Department
- of Neurology, University of Rochester School of Medicine and Dentistry, Rochester, USA. <sup>6</sup>Department
- of Imaging Sciences, University of Rochester School of Medicine and Dentistry, Rochester, USA.
- <sup>7</sup>Department of Biostatistics and Computational Biology, University of Rochester School of Medicine and

- 787 Dentistry, Rochester, USA. <sup>8</sup>Department of Microbiology and Immunology, University of Rochester
- 788 School of Medicine and Dentistry, Rochester, USA. <sup>9</sup>Department of Biomedical Genetics, University of
- 789 Rochester School of Medicine and Dentistry, Rochester, USA

#### 790 Authors' contributions

- 791 Conceptualization of the study: JT, conceptualization of BONITA: MGP, RP, JT; Data curation: MGP,
- AT, JT; Formal Analysis: MGP, JT; Funding acquisition: SM, GS, JT; Investigation: MGP, JT;
- 793 Methodology: MGP, JT; Project administration: AT, JT; Resources: SM, GS, MVS, JT; Software: MGP,
- JT; Supervision: JT; Validation: MGP, JT; Visualization: MGP, JT; Writing original draft: MGP, JT;
- 795 Writing review & editing: All authors

#### 796 **Ethics Declarations**

#### 797 Ethics approval and consent to participate

All methods were carried out in accordance with University of Rochester guidelines and regulations, and
all experimental and study protocols were approved by the University of Rochester Institutional Review
Board (#RSRB00063845). All study participants gave their written informed consent.

#### 801 **Consent for publication**

802 Not applicable

#### 803 Availability of data and materials

The HIV/AS scRNA-seq dataset presented in this manuscript has been deposited in Single Cell Portal and will be made public upon publication of the manuscript. Due to the sensitive nature of HIV data, we have not made the raw data public; however, all results presented in this manuscript may be recapitulated from the data in Single Cell Portal. We also analyzed a previously published dataset that is freely accessible at

- 808 <u>https://singlecell.broadinstitute.org/single\_cell/study/SCP256</u>. All source code and documentation for the
- scBONITA package is available at <u>https://github.com/Thakar-Lab/scBONITA</u>.

### 810 Competing interests

811 The authors declare that they have no competing interests.

# 812 **References**

813 McLaughlin MM, Ma Y, Scherzer R, Rahalkar S, Martin JN, Mills C, et al. Association of Viral 1. 814 Persistence and Atherosclerosis in Adults With Treated HIV Infection. JAMA Netw Open. 815 2020;3(10):e2018099. 816 2. Shah ASV, Stelzle D, Lee KK, Beck EJ, Alam S, Clifford S, et al. Global Burden of 817 Atherosclerotic Cardiovascular Disease in People Living With HIV. Circulation. 2018;138(11):1100-12. 818 Althoff KN, McGinnis KA, Wyatt CM, Freiberg MS, Gilbert C, Oursler KK, et al. Comparison 3. 819 of risk and age at diagnosis of myocardial infarction, end-stage renal disease, and non-AIDS-defining 820 cancer in HIV-infected versus uninfected adults. Clin Infect Dis. 2015;60(4):627-38. 821 4. Drozd DR, Kitahata MM, Althoff KN, Zhang J, Gange SJ, Napravnik S, et al. Increased Risk of 822 Myocardial Infarction in HIV-Infected Individuals in North America Compared With the General 823 Population. J Acquir Immune Defic Syndr. 2017;75(5):568-76. 824 5. Mdodo R, Frazier EL, Dube SR, Mattson CL, Sutton MY, Brooks JT, et al. Cigarette smoking 825 prevalence among adults with HIV compared with the general adult population in the United States: cross-sectional surveys. Ann Intern Med. 2015;162(5):335-44. 826 827 Tien PC, Schneider MF, Cox C, Karim R, Cohen M, Sharma A, et al. Association of HIV 6. 828 infection with incident diabetes mellitus; impact of using hemoglobin A1C as a criterion for diabetes. J 829 Acquir Immune Defic Syndr. 2012;61(3):334-40. Schouten J, Wit FW, Stolte IG, Kootstra NA, van der Valk M, Geerlings SE, et al. Cross-830 7. 831 sectional comparison of the prevalence of age-associated comorbidities and their risk factors between 832 HIV-infected and uninfected individuals: the AGEhIV cohort study. Clin Infect Dis. 2014;59(12):1787-833 97. 834 8. Cotter AG, Satchell CS, O'Halloran J A, Feeney ER, Sabin CA, Mallon PW. High-density 835 lipoprotein levels and 10-year cardiovascular risk in HIV-1-infected patients. Aids. 2011;25(6):867-9. 836 Fahme SA, Bloomfield GS, Peck R. Hypertension in HIV-Infected Adults: Novel 9. 837 Pathophysiologic Mechanisms. Hypertension. 2018;72(1):44-55. Hatleberg CI, Ryom L, d'Arminio Monforte A, Fontas E, Reiss P, Kirk O, et al. Association 838 10. 839 between exposure to antiretroviral drugs and the incidence of hypertension in HIV-positive persons: the 840 Data Collection on Adverse Events of Anti-HIV Drugs (D:A:D) study. HIV Med. 2018;19(9):605-18. 841 Triant VA, Lee H, Hadigan C, Grinspoon SK. Increased acute myocardial infarction rates and 11. 842 cardiovascular risk factors among patients with human immunodeficiency virus disease. J Clin 843 Endocrinol Metab. 2007;92(7):2506-12. 844 Freiberg MS, Chang CC, Kuller LH, Skanderson M, Lowy E, Kraemer KL, et al. HIV infection 12. 845 and the risk of acute myocardial infarction. JAMA Intern Med. 2013;173(8):614-22. 846 Vos AG, Idris NS, Barth RE, Klipstein-Grobusch K, Grobbee DE. Pro-Inflammatory Markers in 13. Relation to Cardiovascular Disease in HIV Infection. A Systematic Review. PLoS One. 847

848 2016;11(1):e0147484.

- 849 14. Subramanian S, Tawakol A, Burdo TH, Abbara S, Wei J, Vijayakumar J, et al. Arterial 850 inflammation in patients with HIV. Jama. 2012;308(4):379-86. Burdo TH, Lentz MR, Autissier P, Krishnan A, Halpern E, Letendre S, et al. Soluble CD163 851 15. 852 made by monocyte/macrophages is a novel marker of HIV activity in early and chronic infection prior to 853 and after anti-retroviral therapy. The Journal of infectious diseases. 2011:204(1):154-63. 854 Dirajlal-Fargo S, Sattar A, Kulkarni M, Funderburg N, McComsey GA. Soluble TWEAK may 16. 855 predict carotid atherosclerosis in treated HIV infection. HIV Clin Trials. 2017;18(4):156-63. 856 Corbeau P, Reynes J. Immune reconstitution under antiretroviral therapy: the new challenge in 17. 857 HIV-1 infection. Blood. 2011;117(21):5582-90. 858 Longenecker CT, Sullivan C, Baker JV. Immune activation and cardiovascular disease in chronic 18. 859 HIV infection. Curr Opin HIV AIDS. 2016;11(2):216-25. 860 19. Schäfer S, Zernecke A. CD8(+) T Cells in Atherosclerosis. Cells. 2020;10(1):37. 861 20. Jaworowski A, Hearps AC, Angelovich TA, Hoy JF. How Monocytes Contribute to Increased 862 Risk of Atherosclerosis in Virologically-Suppressed HIV-Positive Individuals Receiving Combination Antiretroviral Therapy. Front Immunol. 2019;10:1378. 863 Khatri P, Sirota M, Butte AJ. Ten years of pathway analysis: current approaches and outstanding 864 21. 865 challenges. PLoS computational biology. 2012;8(2):e1002375-e. 866 Thakar J, Pilione M, Kirimanjeswara G, Harvill ET, Albert R. Modeling systems-level regulation 22. 867 of host immune responses. PLoS Comput Biol. 2007;3(6):e109. 868 23. Van Twisk D, Murphy SP, Thakar J. Optimized logic rules reveal interferon-gamma-induced 869 modes regulated by histone deacetylases and protein tyrosine phosphatases. Immunology. 870 2017;151(1):71-80. 871 24. Palli R, Palshikar MG, Thakar J. Executable pathway analysis using ensemble discrete-state 872 modeling for large-scale data. PLoS Comput Biol. 2019;15(9):e1007317. 873 Saigusa R, Winkels H, Ley K. T cell subsets and functions in atherosclerosis. Nature reviews 25. 874 Cardiology. 2020;17(7):387-401. 875 Tay C, Kanellakis P, Hosseini H, Cao A, Toh B-H, Bobik A, et al. B Cell and CD4 T Cell 26. 876 Interactions Promote Development of Atherosclerosis. Frontiers in Immunology. 2020;10. 877 27. Tse K, Tse H, Sidney J, Sette A, Ley K. T cells in atherosclerosis. Int Immunol. 2013;25(11):615-878 22. 879 28. Abdolmaleki F, Gheibi Havat SM, Bianconi V, Johnston TP, Sahebkar A. Atherosclerosis and 880 immunity: A perspective. Trends Cardiovasc Med. 2019;29(6):363-71. 881 29. Perry H, Bender T, McNamara C. B cell subsets in atherosclerosis. Frontiers in Immunology. 882 2012;3. 883 30. Srikakulapu P, McNamara CA. B cells and atherosclerosis. American Journal of Physiology-884 Heart and Circulatory Physiology. 2017;312(5):H1060-H7. 885 31. Tay C, Liu Y-H, Hosseini H, Kanellakis P, Cao A, Peter K, et al. B-cell-specific depletion of tumour necrosis factor alpha inhibits atherosclerosis development and plaque vulnerability to rupture by 886 887 reducing cell death and inflammation. Cardiovascular Research. 2016;111(4):385-97. 888 Tay C, Liu Y-H, Kanellakis P, Kallies A, Li Y, Cao A, et al. Follicular B Cells Promote 32. 889 Atherosclerosis via T Cell-Mediated Differentiation Into Plasma Cells and Secreting Pathogenic 890 Immunoglobulin G. Arteriosclerosis, Thrombosis, and Vascular Biology. 2018;38(5):e71-e84. 891 Tsiantoulas D, Diehl CJ, Witztum JL, Binder CJ. B cells and humoral immunity in 33. 892 atherosclerosis. Circ Res. 2014;114(11):1743-56. 893 Weber C, Zernecke A, Libby P. The multifaceted contributions of leukocyte subsets to 34. 894 atherosclerosis: lessons from mouse models. Nat Rev Immunol. 2008;8(10):802-15. 895 Zernecke A, Winkels H, Cochain C, Williams JW, Wolf D, Soehnlein O, et al. Meta-Analysis of 35. 896 Leukocyte Diversity in Atherosclerotic Mouse Aortas. Circ Res. 2020;127(3):402-26. 897 36. Kazer SW, Aicher TP, Muema DM, Carroll SL, Ordovas-Montanes J, Miao VN, et al. Integrated 898 single-cell analysis of multicellular immune dynamics during hyperacute HIV-1 infection. Nature
- 899 Medicine. 2020;26(4):511-8.

- Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data
   across different conditions, technologies, and species. Nat Biotechnol. 2018;36(5):411-20.
- 38. Finak G, McDavid A, Yajima M, Deng J, Gersuk V, Shalek AK, et al. MAST: a flexible
  statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell
  RNA sequencing data. Genome Biol. 2015;16:278.
- 39. Diaz-Mejia JJ, Meng EC, Pico AR, MacParland SA, Ketela T, Pugh TJ, et al. Evaluation of
  methods to assign cell type labels to cell clusters from single-cell RNA-sequencing data. F1000Res.
  2019;8.
- 40. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods. 2015;12(5):453-7.
- 41. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: A universal enrichment tool
  for interpreting omics data. Innovation (N Y). 2021;2(3):100141.
- 42. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular
  Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. 2015;1(6):417-25.
- 43. Kanehisa M, Furumichi M, Sato Y, Ishiguro-Watanabe M, Tanabe M. KEGG: integrating viruses
- and cellular organisms. Nucleic acids research. 2021;49(D1):D545-d51.
- 44. Martens M, Ammar A, Riutta A, Waagmeester A, Slenter DN, Hanspers K, et al. WikiPathways:
  connecting communities. Nucleic acids research. 2021;49(D1):D613-d21.
- Palshikar MG, Hilchey SP, Zand MS, Thakar J. WikiNetworks: translating manually created
  biological pathways for topological analysis. Bioinformatics. 2021.
- 46. Palli R, Palshikar MG, Thakar J. Executable pathway analysis using ensemble discrete-state
  modeling for large-scale data. Plos Computational Biology. 2019;15(9).
- 47. Fortin F-A, De Rainville F-M, Gardner M-AG, Parizeau M, Gagné C. DEAP: Evolutionary
  algorithms made easy. The Journal of Machine Learning Research. 2012;13(1):2171-5.
- 48. Méndez A, Mendoza L. A Network Model to Describe the Terminal Differentiation of B Cells.
  PLoS Comput Biol. 2016;12(1):e1004696.
- 49. Martínez-Méndez D, Villarreal C, Mendoza L, Huerta L. An Integrative Network Modeling
  Approach to T CD4 Cell Activation. Front Physiol. 2020;11:380.
- 928 50. Palma A, Jarrah AS, Tieri P, Cesareni G, Castiglione F. Gene Regulatory Network Modeling of
  929 Macrophage Differentiation Corroborates the Continuum Hypothesis of Polarization States. Front
  930 Physiol. 2018;9:1659.
- 51. Schwab JD, Ikonomi N, Werle SD, Weidner FM, Geiger H, Kestler HA. Reconstructing Boolean
  network ensembles from single-cell data for unraveling dynamics in the aging of human hematopoietic
  stem cells. Comput Struct Biotechnol J. 2021;19:5321-32.
- 934 52. Bäck M, Yurdagul A, Jr., Tabas I, Öörni K, Kovanen PT. Inflammation and its resolution in
- atherosclerosis: mediators and therapeutic opportunities. Nature reviews Cardiology. 2019;16(7):389-406.
- 936 53. Negreiros-Lima GL, Lima KM, Moreira IZ, Jardim BLO, Vago JP, Galvão I, et al. Cyclic AMP
   937 Regulates Key Features of Macrophages via PKA: Recruitment, Reprogramming and Efferocytosis. Cells.
- 2020;9(1).
  54. Wahlang B, McClain C, Barve S, Gobejishvili L. Role of cAMP and phosphodiesterase signaling
  in liver health and disease. Cell Signal. 2018;49:105-15.
- Sugimoto MA, Vago JP, Perretti M, Teixeira MM. Mediators of the Resolution of the
  Inflammatory Response. Trends Immunol. 2019;40(3):212-27.
- 56. Dalli J, Serhan CN. Pro-Resolving Mediators in Regulating and Conferring Macrophage
   Function. Front Immunol. 2017;8:1400.
- 57. Lorenowicz MJ, Fernandez-Borja M, Hordijk PL. cAMP signaling in leukocyte transendothelial
   migration. Arterioscler Thromb Vasc Biol. 2007;27(5):1014-22.
- 947 58. Gewaltig J, Kummer M, Koella C, Cathomas G, Biedermann BC. Requirements for CD8 T-cell
- 948 migration into the human arterial wall. Human Pathology. 2008;39(12):1756-62.

949 59. Havakawa E, Yoshimoto T, Sekizawa N, Sugiyama T, Hirata Y. Overexpression of receptor for 950 advanced glycation end products induces monocyte chemoattractant protein-1 expression in rat vascular 951 smooth muscle cell line. J Atheroscler Thromb. 2012;19(1):13-22. 952 Athanasiadis EI, Botthof JG, Andres H, Ferreira L, Lio P, Cvejic A. Single-cell RNA-sequencing 60. 953 uncovers transcriptional states and fate decisions in haematopoiesis. Nature Communications. 954 2017;8(1):2045. 955 61. Iadevaia V, Huo Y, Zhang Z, Foster Leonard J, Proud Christopher G. Roles of the mammalian 956 target of rapamycin, mTOR, in controlling ribosome biogenesis and protein synthesis. Biochemical 957 Society Transactions. 2012;40(1):168-72. 958 Shahbazian D, Roux PP, Mieulet V, Cohen MS, Raught B, Taunton J, et al. The mTOR/PI3K and 62. 959 MAPK pathways converge on eIF4B to control its phosphorylation and activity. EMBO J. 960 2006;25(12):2781-91. 961 Rosario FJ, Powell TL, Gupta MB, Cox L, Jansson T. mTORC1 Transcriptional Regulation of 63. 962 Ribosome Subunits, Protein Synthesis, and Molecular Transport in Primary Human Trophoblast Cells. 963 Frontiers in Cell and Developmental Biology. 2020;8. 964 Chau KF, Shannon ML, Fame RM, Fonseca E, Mullan H, Johnson MB, et al. Downregulation of 64. 965 ribosome biogenesis during early forebrain development. eLife. 2018;7:e36998. 966 Grönberg C, Nilsson J, Wigren M. Recent advances on CD4+ T cells in atherosclerosis and its 65. 967 implications for therapy. European Journal of Pharmacology. 2017;816:58-66. 968 66. Zhao Y, Qian Y, Sun Z, Shen X, Cai Y, Li L, et al. Role of PI3K in the Progression and 969 Regression of Atherosclerosis. Frontiers in Pharmacology. 2021;12. 970 Flynn MC, Pernes G, Lee MKS, Nagareddy PR, Murphy AJ. Monocytes, Macrophages, and 67. 971 Metabolic Disease in Atherosclerosis. Front Pharmacol. 2019;10:666. 972 Gupta Rajat M, Lee-Kim Vivian S, Libby P. The March of Monocytes in Atherosclerosis. Circ 68. 973 Res. 2020;126(10):1324-6. 974 69. Ley K, Miller YI, Hedrick CC. Monocyte and macrophage dynamics during atherogenesis. 975 Arterioscler Thromb Vasc Biol. 2011;31(7):1506-16. 976 70. Llodrá J, Angeli V, Liu J, Trogan E, Fisher EA, Randolph GJ. Emigration of monocyte-derived 977 cells from atherosclerotic lesions characterizes regressive, but not progressive, plaques. Proceedings of 978 the National Academy of Sciences of the United States of America. 2004;101(32):11779-84. 979 Merino A, Buendia P, Martin-Malo A, Aljama P, Ramirez R, Carracedo J. Senescent 71. 980 CD14+CD16+ monocytes exhibit proinflammatory and proatherosclerotic activity. J Immunol. 981 2011;186(3):1809-15. 982 Ong S-M, Hadadi E, Dang T-M, Yeap W-H, Tan CT-Y, Ng T-P, et al. The pro-inflammatory 72. 983 phenotype of the human non-classical monocyte subset is attributed to senescence. Cell Death & Disease. 984 2018;9(3):266. 985 73. Rafieian-Kopaei M, Setorki M, Doudi M, Baradaran A, Nasri H. Atherosclerosis: process, 986 indicators, risk factors and new hopes. Int J Prev Med. 2014;5(8):927-46. 987 Randolph GJ. The fate of monocytes in atherosclerosis. J Thromb Haemost. 2009;7 Suppl 74. 988 1(Suppl 1):28-30. 989 Wilson HM. Macrophages heterogeneity in atherosclerosis - implications for therapy. J Cell Mol 75. 990 Med. 2010:14(8):2055-65. 991 Woollard KJ, Geissmann F. Monocytes in atherosclerosis: subsets and functions. Nature reviews 76. 992 Cardiology. 2010;7(2):77-86. 993 Gerhardt T, Ley K. Monocyte trafficking across the vessel wall. Cardiovasc Res. 77. 994 2015;107(3):321-30. 995 Taylor JP, Tse HM. The role of NADPH oxidases in infectious and inflammatory diseases. Redox 78. 996 Biology. 2021;48:102159. 997 Zhang J, Li H, Wu Q, Chen Y, Deng Y, Yang Z, et al. Tumoral NOX4 recruits M2 tumor-79. 998 associated macrophages via ROS/PI3K signaling-dependent various cytokine production to promote 999 NSCLC growth. Redox Biology. 2019;22:101116.

Hao JJ, Liu Y, Kruhlak M, Debell KE, Rellahan BL, Shaw S. Phospholipase C-mediated 1000 80. 1001 hydrolysis of PIP2 releases ERM proteins from lymphocyte membrane. J Cell Biol. 2009;184(3):451-62. 1002 81. Gupta N, Wollscheid B, Watts JD, Scheer B, Aebersold R, DeFranco AL. Quantitative proteomic 1003 analysis of B cell lipid rafts reveals that ezrin regulates antigen receptor-mediated lipid raft dynamics. Nat 1004 Immunol. 2006;7(6):625-33. 1005 Parameswaran N, Matsui K, Gupta N. Conformational switching in ezrin regulates morphological 82. 1006 and cytoskeletal changes required for B cell chemotaxis. J Immunol. 2011;186(7):4088-97. 1007 Bradley JR. TNF-mediated inflammatory disease. The Journal of Pathology. 2008;214(2):149-60. 83. 1008 84. Mehta AK, Gracias DT, Croft M. TNF activity and T cells. Cytokine. 2018;101:14-8. 1009 85. Jones SJ, Ledgerwood EC, Prins JB, Galbraith J, Johnson DR, Pober JS, et al. TNF recruits 1010 TRADD to the plasma membrane but not the trans-Golgi network, the principal subcellular location of 1011 TNF-R1. J Immunol. 1999;162(2):1042-8. 1012 Zhang Q, Cui F, Fang L, Hong J, Zheng B, Zhang JZ. TNF-α impairs differentiation and function 86. 1013 of TGF-β-induced Treg cells in autoimmune diseases through Akt and Smad3 signaling pathway. J Mol 1014 Cell Biol. 2013;5(2):85-98. 1015 Azari BM, Marmur JD, Salifu MO, Ehrlich YH, Kornecki E, Babinska A. Transcription and 87. 1016 translation of human F11R gene are required for an initial step of atherogenesis induced by inflammatory 1017 cytokines. Journal of Translational Medicine. 2011;9(1):98. 1018 Theilmeier G, Michiels C, Spaepen E, Vreys I, Collen D, Vermylen J, et al. Endothelial von 88. 1019 Willebrand factor recruits platelets to atherosclerosis-prone sites in response to hypercholesterolemia. 1020 Blood. 2002;99(12):4486-93. 1021 Massberg S, Brand K, Grüner S, Page S, Müller E, Müller I, et al. A critical role of platelet 89. 1022 adhesion in the initiation of atherosclerotic lesion formation. J Exp Med. 2002;196(7):887-96. 1023 Woodfin A, Voisin M-B, Nourshargh S. PECAM-1: A Multi-Functional Molecule in 90. 1024 Inflammation and Vascular Biology. Arteriosclerosis, Thrombosis, and Vascular Biology. 1025 2007;27(12):2514-23. 1026 Stevens HY, Melchior B, Bell KS, Yun S, Yeh JC, Frangos JA. PECAM-1 is a critical mediator 91. 1027 of atherosclerosis. Dis Model Mech. 2008;1(2-3):175-81; discussion 9. 1028 Caligiuri G. CD31 as a Therapeutic Target in Atherosclerosis. Circ Res. 2020;126(9):1178-89. 92. 1029 93. Cornwell A, Palli R, Singh MV, Benoodt L, Tyrell A, Abe JI, et al. Molecular characterization of 1030 atherosclerosis in HIV positive persons. Sci Rep. 2021;11(1):3232. 1031 Saigusa R, Winkels H, Ley K. T cell subsets and functions in atherosclerosis. Nature Reviews 94. 1032 Cardiology. 2020;17(7):387-401. 1033 Reed JC, Preston-Hurlburt P, Philbrick W, Betancur G, Korah M, Lucas C, et al. The receptor for 95. 1034 advanced glycation endproducts (RAGE) modulates T cell signaling. PLOS ONE. 2020;15(9):e0236921. 1035 96. Schäfer S, Zernecke A. CD8(+) T Cells in Atherosclerosis. Cells. 2020;10(1). 1036 97. van Duijn J, Kritikou E, Benne N, van der Heijden T, van Puijvelde GH, Kroner MJ, et al. CD8+ 1037 T-cells contribute to lesion stabilization in advanced atherosclerosis by limiting macrophage content and 1038 CD4+ T-cell responses. Cardiovasc Res. 2019;115(4):729-38. 1039 van Duijn J, Kuiper J, Slutter B. The many faces of CD8+ T cells in atherosclerosis. Current 98. 1040 opinion in lipidology. 2018;29(5):411-6. 1041 Cochain C, Zernecke A. Protective and pathogenic roles of CD8(+) T cells in atherosclerosis. 99. 1042 Basic research in cardiology. 2016;111(6):71. 1043 Castilho JL, Shepherd BE, Koethe J, Turner M, Bebawy S, Logan J, et al. CD4+/CD8+ ratio, age, 100. 1044 and risk of serious noncommunicable diseases in HIV-infected adults on antiretroviral therapy. AIDS 1045 (London, England). 2016;30(6):899-908. 1046 101. Flynn MC, Pernes G, Lee MKS, Nagareddy PR, Murphy AJ. Monocytes, Macrophages, and 1047 Metabolic Disease in Atherosclerosis. Frontiers in Pharmacology. 2019;10(666). 1048 102. Cochain C, Zernecke A. Macrophages and immune cells in atherosclerosis: recent advances and 1049 novel concepts. Basic research in cardiology. 2015;110(4):34.

- 103. Shrestha S, Irvin MR, Grunfeld C, Arnett DK. HIV, inflammation, and calcium in atherosclerosis.
   1051 Arterioscler Thromb Vasc Biol. 2014;34(2):244-50.
- 1052 104. Crowe SM, Westhorpe CL, Mukhamedova N, Jaworowski A, Sviridov D, Bukrinsky M. The
- macrophage: the intersection between HIV infection and atherosclerosis. J Leukoc Biol. 2010;87(4):58998.
- 1055 105. Bianchi ME, Mezzapelle R. The Chemokine Receptor CXCR4 in Cell Proliferation and Tissue
   1056 Regeneration. Frontiers in Immunology. 2020;11.
- 1057 106. Vogl T, Tenbrock K, Ludwig S, Leukert N, Ehrhardt C, van Zoelen MA, et al. Mrp8 and Mrp14
- are endogenous activators of Toll-like receptor 4, promoting lethal, endotoxin-induced shock. Nat Med.
   2007;13(9):1042-9.
- 1060 107. Crowe LAN, McLean M, Kitson SM, Melchor EG, Patommel K, Cao HM, et al. S100A8 &
  1061 S100A9: Alarmin mediated inflammation in tendinopathy. Sci Rep. 2019;9(1):1463.
- 1062 108. Shalek AK, Satija R, Adiconis X, Gertner RS, Gaublomme JT, Raychowdhury R, et al. Single-
- 1063 cell transcriptomics reveals bimodality in expression and splicing in immune cells. Nature.1064 2013;498(7453):236-40.
- 1065 109. Trinh H-C, Kwon Y-K. A novel constrained genetic algorithm-based Boolean network inference
- 1066 method from steady-state gene expression data. Bioinformatics. 2021;37(Supplement\_1):i383-i91.
- 1067 110. Lim CY, Wang H, Woodhouse S, Piterman N, Wernisch L, Fisher J, et al. BTR: training
- asynchronous Boolean models using single-cell expression data. BMC Bioinformatics. 2016;17(1):355.
- 1069 111. Egaña-Gorroño L, López-Díez R, Yepuri G, Ramirez LS, Reverdatto S, Gugger PF, et al.
- 1070 Receptor for Advanced Glycation End Products (RAGE) and Mechanisms and Therapeutic Opportunities
- in Diabetes and Cardiovascular Disease: Insights From Human Subjects and Animal Models. Front
   Cardiovasc Med. 2020;7:37.
- 1073 112. Kosmopoulos M, Drekolias D, Zavras PD, Piperi C, Papavassiliou AG. Impact of advanced 1074 glycation end products (AGEs) signaling in coronary artery disease. Biochimica et Biophysica Acta
- 1075 (BBA) Molecular Basis of Disease. 2019;1865(3):611-9.
- 1076 113. Narumi K, Miyakawa R, Ueda R, Hashimoto H, Yamamoto Y, Yoshida T, et al. Proinflammatory
- 1077 Proteins S100A8/S100A9 Activate NK Cells via Interaction with RAGE. The Journal of Immunology.
  1078 2015;194(11):5539-48.
- 1079 114. Hofmann Bowman MA, Schmidt AM. S100/calgranulins EN-RAGEing the blood vessels:
- 1080 implications for inflammatory responses and atherosclerosis. Am J Cardiovasc Dis. 2011;1(1):92-100.
- 1081 115. Barlovic DP, Soro-Paavonen A, Jandeleit-Dahm KA. RAGE biology, atherosclerosis and diabetes. Clin Sci (Lond). 2011;121(2):43-55.
- 1083 116. Yan SF, Ramasamy R, Schmidt AM. The receptor for advanced glycation endproducts (RAGE) 1084 and cardiovascular disease. Expert Rev Mol Med. 2009;11:e9.
- 1085 117. Sparvero LJ, Asafu-Adjei D, Kang R, Tang D, Amin N, Im J, et al. RAGE (Receptor for
- 1086 Advanced Glycation Endproducts), RAGE Ligands, and their role in Cancer and Inflammation. Journal of
- 1087 Translational Medicine. 2009;7(1):17.
- 1088 118. Nienhuis HL, Westra J, Smit AJ, Limburg PC, Kallenberg CG, Bijl M. AGE and their receptor
- 1089 RAGE in systemic autoimmune diseases: an inflammation propagating factor contributing to accelerated
   1090 atherosclerosis. Autoimmunity. 2009;42(4):302-4.
- 1091 119. Jandeleit-Dahm K, Watson A, Soro-Paavonen A. The AGE/RAGE axis in diabetes-accelerated
   1092 atherosclerosis. Clin Exp Pharmacol Physiol. 2008;35(3):329-34.
- 1093 120. Li H, Zhang L, Yin D, Zhang Y, Miao J. Targeting Phosphatidylcholine-Specific Phospholipase
   1094 C for Atherogenesis Therapy. Trends in Cardiovascular Medicine. 2010;20(5):172-6.
- 1095 121. Zhou Q, Liao JK. Rho kinase: an important mediator of atherosclerosis and vascular disease. Curr 1096 Pharm Des. 2009;15(27):3108-15.
- 1097 122. Hartmann S, Ridley AJ, Lutz S. The Function of Rho-Associated Kinases ROCK1 and ROCK2 in 1098 the Pathogenesis of Cardiovascular Disease. Front Pharmacol. 2015;6:276.
- 1099 123. Cai A, Zhou Y, Li L. Rho-GTPase and Atherosclerosis: Pleiotropic Effects of Statins. J Am Heart 1100 Assoc. 2015;4(7).

- 1101 124. Jin SY, Kim EK, Ha JM, Lee DH, Kim JS, Kim IY, et al. Insulin regulates monocyte trans-
- endothelial migration through surface expression of macrophage-1 antigen. Biochim Biophys Acta.
  2014;1842(9):1539-48.
- 1104 125. Zhou Y, Wang Y, Qiao S, Yin L. Effects of Apelin on Cardiovascular Aging. Frontiers in1105 Physiology. 2017;8.
- 1106 126. Pitkin SL, Maguire JJ, Kuc RE, Davenport AP. Modulation of the apelin/APJ system in heart 1107 failure and atherosclerosis in man. British Journal of Pharmacology. 2010;160(7):1785-95.
- 1108 127. Ahrens TD, Bang-Christensen SR, Jørgensen AM, Løppke C, Spliid CB, Sand NT, et al. The
- 1109 Role of Proteoglycans in Cancer Metastasis and Circulating Tumor Cell Analysis. Front Cell Dev Biol.
- 1110 2020;8:749.
- 1111 128. Edwards IJ. Proteoglycans in prostate cancer. Nat Rev Urol. 2012;9(4):196-206.
- 1112 129. Aberle H. Axon Guidance and Collective Cell Migration by Substrate-Derived Attractants. Front1113 Mol Neurosci. 2019;12:148.
- 1114 130. Russell SA, Bashaw GJ. Axon guidance pathways and the control of gene expression. Dev Dyn.
  1115 2018;247(4):571-80.
- 1116 131. Tavares LP, Negreiros-Lima GL, Lima KM, E Silva PMR, Pinho V, Teixeira MM, et al. Blame
- 1117 the signaling: Role of cAMP for the resolution of inflammation. Pharmacological Research.
- 1118 2020;159:105030.
- 1119 132. Huang S. Reprogramming cell fates: reconciling rarity with robustness. BioEssays : news and 1120 reviews in molecular, cellular and developmental biology. 2009;31(5):546-60.
- 1121 133. Huang S, Eichler G, Bar-Yam Y, Ingber DE. Cell fates as high-dimensional attractor states of a 1122 complex gene regulatory network. Physical review letters. 2005;94(12):128701.
- 1123 134. Huang S, Ernberg I, Kauffman S. Cancer attractors: a systems view of tumors from a gene 1124 network dynamics and developmental perspective. Seminars in cell & developmental biology.
- 1124 network dynamics and developmental perspective. Seminars in cen & developmental biology. 1125 2009;20(7):869-76.
- 1126 135. Taherian Fard A, Ragan MA. Modeling the Attractor Landscape of Disease Progression: a
  1127 Network-Based Approach. Front Genet. 2017;8:48.
- 1128 136. Uthamacumaran A. A review of dynamical systems approaches for the detection of chaotic 1129 attractors in cancer networks. Patterns (N Y). 2021;2(4):100226.
- 1130 137. Trapnell C. Defining cell types and states with single-cell genomics. Genome Res.
- 1131 2015;25(10):1491-8.
- 1132 138. Munsky B, Neuert G, van Oudenaarden A. Using gene expression noise to understand gene 1133 regulation. Science (New York, NY). 2012;336(6078):183-7.
- 1134 139. Liu S, Trapnell C. Single-cell transcriptome sequencing: recent advances and remaining 1135 challenges. F1000Research. 2016;5:F1000 Faculty Rev-182.
- 1136140.Tsai S, Clemente-Casares X, Zhou AC, Lei H, Ahn JJ, Chan YT, et al. Insulin Receptor-Mediated1137Stimulation Boosts T Cell Immunity during Inflammation and Infection. Cell Metab. 2018;28(6):922-
- 1138 34.e4.
- 1139 141. Arasanz H, Gato-Cañas M, Zuazo M, Ibañez-Vea M, Breckpot K, Kochan G, et al. PD1 signal
  1140 transduction pathways in T cells. Oncotarget. 2017;8(31):51936-45.
- 1141 142. Zhao Y, Shao Q, Peng G. Exhaustion and senescence: two crucial dysfunctional states of T cells 1142 in the tumor microenvironment. Cellular & Molecular Immunology. 2020;17(1):27-35.
- 1142 In the funior microenvironment. Centular & Molecular minutology. 2020;17(1):27-55. 1143 143. Yi H-S, Kim SY, Kim JT, Lee Y-S, Moon JS, Kim M, et al. T-cell senescence contributes to
- 1144 abnormal glucose homeostasis in humans and mice. Cell Death & Disease. 2019;10(3):249.
- 1145 144. McLaughlin T, Liu L-F, Lamendola C, Shen L, Morton J, Rivas H, et al. T-cell profile in adipose
- 1146 tissue is associated with insulin resistance and systemic inflammation in humans. Arteriosclerosis,
- 1147 thrombosis, and vascular biology. 2014;34(12):2637-43.
- 1148 145. Singh MV, Kotla S, Le N-T, Ae Ko K, Heo K-S, Wang Y, et al. Senescent Phenotype Induced by
- 1149 p90RSK-NRF2 Signaling Sensitizes Monocytes and Macrophages to Oxidative Stress in HIV-Positive
- 1150 Individuals. Circulation. 2019;139(9):1199-216.

- 1151 146. Rattan V, Shen Y, Sultana C, Kumar D, Kalra VK. Glucose-induced transmigration of monocytes
- is linked to phosphorylation of PECAM-1 in cultured endothelial cells. American Journal of Physiology Endocrinology and Metabolism. 1996;271(4):E711-E7.
- 1154 147. Okouchi M, Okayama N, Imai S, Omi H, Shimizu M, Fukutomi T, et al. High insulin enhances
- 1155 neutrophil transendothelial migration through increasing surface expression of platelet endothelial cell
- adhesion molecule-1 via activation of mitogen activated protein kinase. Diabetologia. 2002;45(10):1449-
- 1157 56.
- 1158 148. Kharchenko PV, Silberstein L, Scadden DT. Bayesian approach to single-cell differential
- expression analysis. Nature methods. 2014;11(7):740-2.
- 1160 149. Brorson IS, Eriksson A, Leikfoss IS, Celius EG, Berg-Hansen P, Barcellos LF, et al. No
- 1161 differential gene expression for CD4(+) T cells of MS patients and healthy controls. Mult Scler J Exp
- 1162 Transl Clin. 2019;5(2):2055217319856903-.

1163

1164

# 1165 Supplementary Material

1166	1.	Supplementary	File	1:
------	----	---------------	------	----

1167		• Filename: supplementary_file_1.pdf
1168		• <b>Description:</b> Contains Supplementary Figures 1 – 6 and corresponding captions.
1169	2.	Supplementary Table 1:
1170		• Filename: <i>supplementary_table_1.txt</i>
1171		• Description of Supplementary Table 1: Cluster markers for each subpopulation
1172		identified in PBMCs derived from AS+ and AS- PLWH are listed in the file
1173		"supplementary_table_1.csv". Cluster markers were identified using methods
1174		implemented in the Seurat R package, as described in the methods.
1175	3.	Supplementary Table 2:
1176		• Filename: supplementary_table_2.xlsx
1177		• Description of supplementary table 2: supplementary_table_2 contains 2 worksheets,
1178		"DE genes in AS+ vs AS-" and "enrichr_kegg". "DE genes in AS+ vs AS-" contains a
1179		table of genes differentially expressed between cells derived from AS+ and AS- PLWH
1180		for each subpopulation in the HIV/AS dataset. The sheet "enrichr_kegg" contains a table
1181		of KEGG gene sets enriched (identified using the enrichr R package) in the DE genes
1182		from "DE genes in AS+ vs AS-".
1183	4.	Supplementary Table 3:
1184		• Filename: supplementary_table_3.txt
1185		• <b>Description of Supplementary Table 3:</b> scBONITA infers biologically meaningful
1186		dysregulated pathways for subpopulations of PBMCs derived from AS+ and AS- PLWH
1187		in the HIVAS/HIVAS- contrast. The table of dysregulated pathways identified by
1188		scBONITA in the HIV+/AS+ - HIV+/AS- contrast in all cell clusters is presented in the
1189		CSV file titled "supplementary_table_3.txt".

1190	5.	Supplementary Table 4
1191		• Filename: supplementary_table_4.txt
1192		• <b>Description:</b> Pearson correlation coefficients between importance scores for networks
1193		trained on subpopulations of PBMCs from the Kazer et al dataset and trained on the
1194		corresponding subpopulations from the HIV/AS dataset. All p values are $< 0.01$ .
1195	6.	Supplementary Table 5
1196		• <b>Filename:</b> <i>supplementary_table_5.txt</i>
1197		• <b>Description:</b> scBONITA infers biologically meaningful dysregulated pathways for
1198		subpopulations of PBMCs derived from HIV- subjects and subjects after 1 year of HIV
1199		infection (Kazer et al). The CSV file supplementary_table_5.txt lists the dysregulated
1200		pathways and p-values from scBONITA for every subpopulation.
1201	7.	Supplementary Table 6
1202		• <b>Filename:</b> <i>supplementary_table_6.txt</i>
1203		• <b>Description:</b> Comparison of dysregulated pathways, as identified by scBONITA,
1204		between subpopulations of PBMCs derived from HIV- subjects and subjects after 1 year
1205		of HIV infection (Kazer et al) and subpopulations of PBMCs derived from HIV+ subjects
1206		with and without atherosclerosis. The file <i>supplementary_table_6.txt</i> lists the
1207		dysregulated pathways for subpopulations from the Kazer et al dataset and the
1208		corresponding subpopulations from the HIV/AS dataset, along with the intersections
1209		between the pathways dysregulated in the two contrasts. We speculate that intersecting
1210		pathways and the pathways dysregulated in the HIV/AS contrast are driven by HIV-
1211		associated inflammatory processes. Similarly, pathways that are dysregulated only in the
1212		HIV-/HIV+ contrast are assumed to be driven by the immediate antiviral response to HIV
1213		infection.

# 1214 Figures



1215





1225

1226 Figure 2: scBONITA pipeline to infer Boolean rules and perform pathway analysis using single cell 1227 expression measurements (A) Input: scBONITA requires a binarized single-cell RNA-seq dataset as a 1228 text file, and a prior knowledge network (PKN) describing the activating or inhibitory relationships 1229 between genes (B) Rule determination: scBONITA infers logic rules that describe the regulatory 1230 relationships between nodes in the PKN by a global search followed by node-level rule refinement (C) 1231 Pathway analysis: scBONITA calculates a gene importance score calculated by simulating network 1232 perturbations with inferred rules and combines these scores with fold-changes from scRNA-seq to 1233 identify dysregulated pathways in a specified contrast (D) Steady-state analysis: scBONITA simulates 1234 networks using learned rules to identify steady states which correspond to observed cellular states.



Figure 3: scBONITA identifies dysregulated pathways in T cells derived from AS+ and AS- PLWH. (A) 1236 1237 Pathways (y-axis) dysregulated in the AS+ vs AS- contrast in PLWH in clusters of CD8+ T cells. Clusters 1238 are differentiated by point shape, as shown in the legend. Pathways that have Bonferroni-corrected p-1239 value < 0.01 (x-axis) and a reduced ERS (see Methods for details) are shown. Pathways labeled with "\*\*\*" were also significantly dysregulated between cytotoxic T cells derived from HIV- subjects and 1240 subjects after 1 year of HIV infection (36) (B) (Pathways (y-axis) dysregulated in the AS+ vs AS-1241 1242 contrast in PLWH in clusters of CD4+ T cells and naïve T cells. Clusters are differentiated by point 1243 shape, as shown in the legend. Pathways that have Bonferroni-corrected p-value < 0.01 (x-axis) and a reduced ERS (see Methods for details) are shown. Pathways labeled with "\*\*\*" were also significantly 1244 1245 dysregulated between T cells derived from HIV- subjects and subjects after 1 year of HIV infection (36) (C) Network representation of the AGE-RAGE signaling pathway (Bonferroni-corrected p-value < 0.01) 1246 1247 in a cluster of CD8+ T cells referred to as CD8 T cells -1 in Figure 1A. Small black intermediate nodes 1248 indicate that the downstream nodes are controlled by an AND function of the upstream nodes. The size of 1249 nodes corresponding to genes is proportional to their importance score calculated by scBONITA. Nodes 1250 are colored according to the magnitude of their fold change between the HIV+AS+ and HIV+AS- groups. 1251 Violet edges indicate inhibition edges and black edges indicate activation edges.

1252

1235



1254 Figure 4: scBONITA identifies dysregulated pathways in monocytes derived from AS+ and AS- PLWH: 1255 (A) Pathways (y-axis) dysregulated in the AS+ vs AS- contrast in monocytes derived from PLWH. Only 1256 pathways that have Bonferroni-corrected p-value < 0.01 (x-axis) and which have a reduced ERS (see 1257 Methods for details) are shown. Pathways labeled with "\*\*\*" were also significantly dysregulated in monocytes after one year of HIV infection (36) (B) Network representation of the leukocyte 1258 1259 transendothelial migration pathway. Small black intermediate nodes indicate that the downstream nodes 1260 are controlled by an AND function of the upstream nodes. The size of nodes corresponding to genes is 1261 proportional to their importance score as calculated by scBONITA. Nodes are colored according to the 1262 magnitude of their fold change between the HIV+AS+ and HIV+AS- groups. Violet edges indicate 1263 inhibition edges and black edges indicate activation edges.

1264

1253



1268 with scBONITA. (A) UMAP representation of a cluster of CD8+ T cells (CD8+ T cells – 1 in Figure 1A)

1269 colored by the attractor to which they are assigned, based on their similarity. The three dominant states

1270 (PI3KR+ PI3K+, TNFR1+TNF- and TNFR1-TNF+ attractors) are represented by green, blue and orange.

- 1271 All other attractors are collectively labeled in grey. (B) Percentages of CD8+ T cells derived from each
- 1272 subject, mapping to the three dominant and all other attractors. (C) Gene activity (ON- red, OFF- light
- 1273 blue) in the three dominant attractors. Only genes that are different between these states are shown. (D)
- 1274 Attractor gene values ranging from 0 (blue) to 1 (red) averaged for each individual subject. The top bar
- 1275 indicates AS+ (grey) and AS- (black) subjects.
- 1276



1278 Figure 6: Monocyte states with respect to the leukocyte transendothelial migration pathway identified by 1279 attractor analysis with scBONITA. (A) UMAP representation of the cluster of monocytes colored by the 1280 attractor to which they are assigned, based on their similarity. The two dominant modes (F11R+ and 1281 PECAM+ attractors) are represented by blue and orange. All other attractors are collectively labeled in 1282 grey. (B) Percentages of monocytes derived from each subject, mapping to the two dominant attractors 1283 and all other attractors for the leukocyte transendothelial migration pathway. (C) Attractor gene values 1284 for the for the leukocyte transendothelial migration pathway trained on monocytes, ranging from 0 (blue) 1285 to 1 (red), averaged for each individual subjects. The top bar indicates AS+(grey) and AS-(black)1286 subjects. The genes that differ between the two dominant attractors F11R+ and PECAM+ are highlighted 1287 by a violet box.

1288



1290 Figure 7: Performance of scBONITA rule determination. (A) The number of pathways identified as 1291 significantly dysregulated (Bonferroni-adjusted p value < 0.05) one year upon HIV infection(36), 1292 between AS+ and AS- PLWH, and the intersections between these sets. Subpopulations from the two 1293 datasets were matched as shown in Supplementary Table 5. (B) Effects of number of cells on the ERS 1294 size evaluated by downsampling and augmentation using the largest cluster ("B cells nave - 1") from the 1295 HIV/AS dataset (C) Relation between importance scores in 130 KEGG networks evaluated using CD8+ T 1296 cells from AS+ and AS- PLWH and from persons before and one year after HIV infection (36). (D) 1297 Spearman correlations (p < 0.01 for all comparisons) between scBONITA's node importance score 1298 (labeled as 'scBONITA score') and 6 measures of node centrality (along x and y axis). Correlation 1299 coefficients are depicted by colors ranging from blue (-1) to red (+1).