Unique T cell signatures associated with reduced *Chlamydia trachomatis* reinfection in a highly exposed cohort

Kacy S Yount[1], Chi-Jane Chen[2], Avinash Kollipara[1], Chuwen Liu[3], Neha V Mokashi[3], Xiaojing Zheng[1,3], C Bruce Bagwell[4], Taylor B Poston[1], Harold C Wiesenfeld[5], Sharon L Hillier[5], Catherine M O'Connell[1], Natalie Stanley[2,6], Toni Darville[1,7]

[1]Department of Pediatrics, University of North Carolina at Chapel Hill, Chapel Hill, NC; [2]Department of Computer Science, University of North Carolina at Chapel Hill, Chapel Hill, NC; [3]Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC; [4]P.O. Box 247, Topsham, ME; [5]The University of Pittsburgh School of Medicine and the Magee-Womens Research Institute, Pittsburgh, PA; [6]Computational Medicine Program, University of North Carolina at Chapel Hill, Chapel Hill, NC; [7]Department of Microbiology and Immunology, University of North Carolina at Chapel Hill, Chapel Hill, NC

Corresponding Author:

Toni Darville

lad@email.unc.edu

501-944-3130

4008B Mary Ellen Jones, 116 Manning Drive, Chapel Hill, NC, 27599

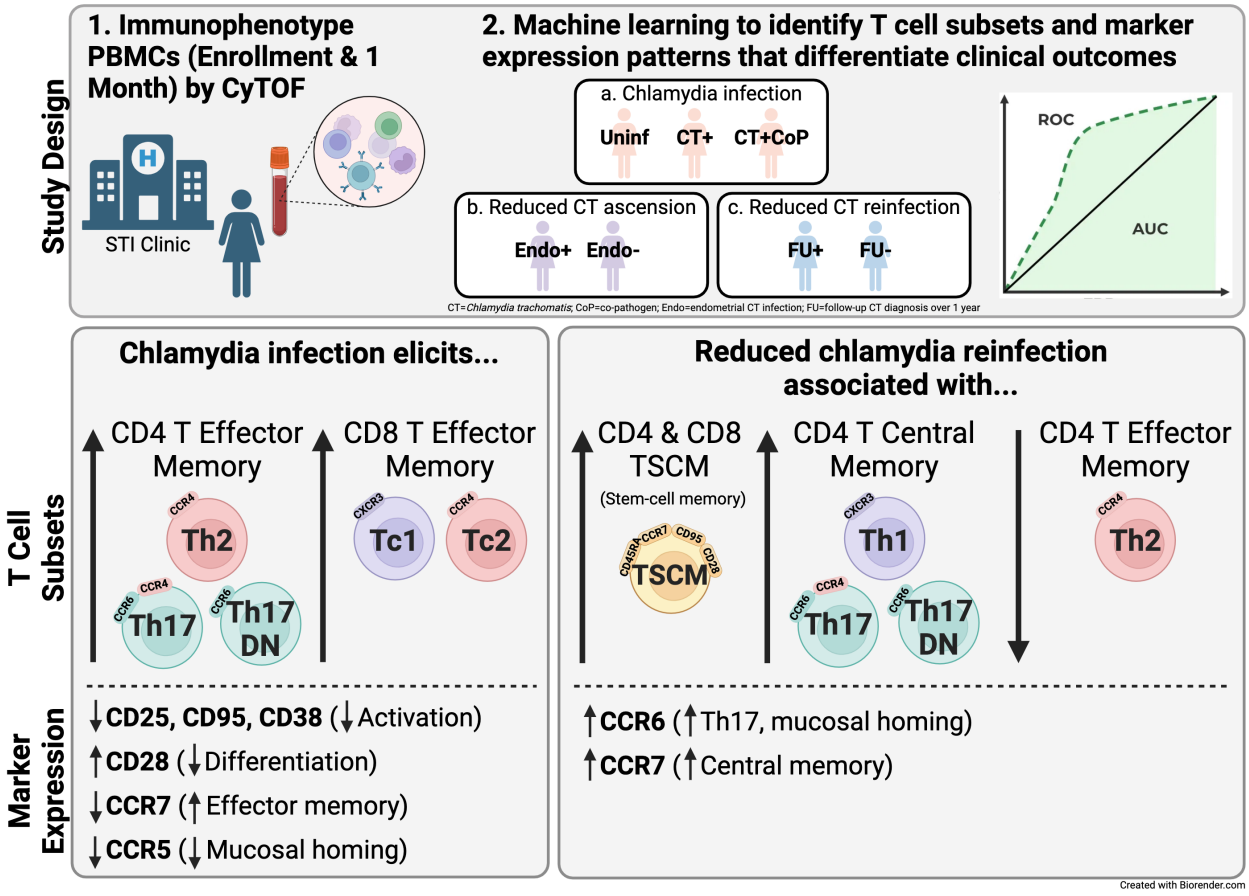
The authors have declared that no conflict of interest exists.

ABSTRACT

Chlamydia trachomatis (CT) is the most common bacterial sexually transmitted infection globally. Understanding natural immunity to CT will inform vaccine design. This study aimed to profile immune cells and associated functional features in CT-infected women, and determine immune profiles associated with reduced risk of ascended endometrial CT infection and CT reinfection. PBMCs from CT-exposed women were profiled by mass cytometry and random forest models identified key features that distinguish outcomes. CT+ participants exhibited higher frequencies of CD4+ Th2, Th17, and Th17 DN CD4 T effector memory (TEM) cells than uninfected participants with decreased expression of T cell activation and differentiation markers. No significant differences were detected between women with or without endometrial CT infection. Participants who remained follow-up negative (FU-) showed higher frequencies of CD4 T central memory (TCM) Th1, Th17, Th1/17, and Th17 DN but reduced CD4 TEM Th2 cells than FU+ participants. Expression of markers associated with central memory and Th17 lineage were increased on T cell subsets among FU- participants. These data indicate that peripheral T cells exhibit distinct features associated with resistance to CT reinfection. The highly plastic Th17 lineage appears to contribute to protection. Addressing these immune nuances could promote efficacy of CT vaccines.

43 **GRAPHICAL ABSTRACT**

44



45

46 **INTRODUCTION**

47

48 *Chlamydia trachomatis* (CT) is the most common bacterial sexually transmitted infection
49 (STI) globally (1). In women, CT infects the cervix and can ascend to the uterus and oviducts to
50 cause pelvic inflammatory disease (PID) that can result in pelvic pain, infertility, and ectopic
51 pregnancy (2). Although antibiotic treatment is effective, 80% of infections are asymptomatic and
52 may go undetected (3), contributing to the risk for chronic sequelae (4). Although repeated CT
53 infections are common, they can elicit partially protective immunity. Women who spontaneously
54 cleared infection without treatment were less likely to become reinfected (5) and persons with
55 confirmed prior infection had lower chlamydial loads (6). Determining the aspects of natural
56 immunity that drive protection against CT will advance understanding of chlamydial immunity and
57 aid design of an effective vaccine.

58 Mouse models have elucidated protective mechanisms of different arms of the immune
59 response. The importance of CD4 T cell mediated immunity is revealed from studies in severe
60 combined immune deficient (SCID), RAG-1-/-, TCRα-/-, CD4-/-, and MHCII-/- mice, which display
61 delayed or failed resolution of *Chlamydia* infection, while MHCI-deficient (B2M-/-), B cell-deficient
62 (μMT), and CD8-/- mice resolve infection with comparable kinetics to wild-type (7-11). Th1 cells
63 secrete IFNγ and TNFα that promote recruitment and activation of macrophages. IFNγ also
64 induces expression of indoleamine 2,3-dioxygenase (IDO) in epithelial cells, which catabolizes
65 tryptophan, an essential nutrient for chlamydia growth (12). IFNγ-knockout, antibody depletion,
66 and adoptive transfer studies in mice confirmed that IFNγ-producing CD4 T cells (Th1) are key
67 mediators of protection against *Chlamydia* (13-15), although this may be independent of T-bet,
68 the defining transcription factor expressed by Th1 cells (16). Th2 cells produce IL-4, IL-5, and IL-
69 13 that drive antibody responses, but a *Ct*-specific Th2 clone did not protect against murine genital
70 infection and displayed reduced trafficking to the genital mucosa, whereas a polyfunctional Th1
71 clone mediated chlamydial clearance (17, 18). Classical Th17 cells produce IL-17, recruit

72 neutrophils, promote secretory IgA, and are instrumental in protection against many mucosal
73 pathogens, but can be pathogenic. Mice lacking T-bet had Th17-skewed responses and similar
74 clearance of *C. muridarum* infection to wild-type (16) while those lacking RORγt (Th17-deficient)
75 exhibited delayed clearance (19), highlighting a potential protective role for Th17. However, we
76 found that IFNγ-deficient mice had chronic pathology related to a compensatory increase in Th17
77 cells and downstream induction of neutrophils (20), highlighting the importance for a balance in
78 Th17 responses.

79 The T Cell Response Against Chlamydia (TRAC) cohort is a unique resource containing
80 clinical samples and information from women highly exposed to CT and other STIs that was
81 established to improve understanding of human natural immunity to CT (21). In this cohort, we
82 observed that peripheral blood CD4 T cells producing CT-specific IFNγ were associated with
83 protection from reinfection (22), consistent with findings of others (23), and we additionally
84 determined that CD8 T cells producing CT-specific IFNγ were associated with protection from
85 endometrial infection (22). Further, we identified a single nucleotide polymorphism in the gene for
86 CD151, a tetrospanin molecule important for immune cell adhesion and migration and T cell
87 proliferation, as a candidate genetic marker for reduced susceptibility to endometrial infection
88 (24). In contrast, we determined that anti-CT antibodies were insufficient to protect participants
89 from endometrial or recurrent infection (25, 26). These data indicate that CD4 T cells, and to a
90 lesser extent CD8 T cells, contribute to protection against CT reinfection or ascension.

91 The objective of this study was to immunophenotype peripheral blood from TRAC
92 participants to identify specific T cell subsets and surface markers associated with T cell function
93 that associated with resistance to CT reinfection and/or endometrial infection. Therefore, a CyTOF
94 panel containing 33 markers was designed to deeply profile T-helper cell subsets such as T
95 follicular helper (Tfh; CXCR5+) cells, regulatory T (Treg; CTLA4+ CD25+ CD127lo) cells, naïve T
96 cells (TN; CD45RA+ CCR7+), central memory T (TCM; CD45RA- CCR7+) cells, effector memory
97 T (TEM; CD45RA- CCR7-) cells, terminally differentiated T effector (TEMRA; CD45RA+ CCR7-)

98 cells, and stem-cell like memory T (TSCM; CD45RA+ CCR7+ CD28+ CD95+) cells. TEM and
99 TCM T-helper cell subsets were defined by surface expression of chemokine receptors CCR6,
100 CXCR3, and CCR4 (27) because multiple studies have shown that CCR6-CXCR3+CCR4- (Th1),
101 CCR6-CXCR3-CCR4+ (Th2), and CCR6+CXCR3-CCR4+ (Th17) cells polarized or sorted from
102 human peripheral blood express the canonical cytokines and transcription factors associated with
103 their corresponding T-helper phenotype (27-31). Th17-lineage cells are also highly plastic and
104 modify chemokine receptor and effector cytokine expression depending on their environment (32,
105 33). Classical Th17 cells (CCR6+CXCR3-CCR4+) produce primarily IL-17 and are important for
106 neutrophil recruitment and defense at mucosal sites (34); Th1/17 cells (CCR6+CXCR3+CCR4-)
107 share Th1 and Th17 functionality, respond to IL-12 and IL-23, and produce mostly IFNγ (35);
108 Th17 double positive (DP) cells (CCR6+CXCR3+CCR4+) produce IFNγ and low levels of IL-17
109 (29), and double negative (DN) cells (CCR6+CXCR3-CCR4-) preferentially produce IL-17 and
110 express transcriptional markers of early Th17 development (IL-17F, STAT3), lymph-node homing
111 (CCR7, CD62L), follicular help (CXCR5, BCL6, ASCL2), and self-renewal (LEFI, MYC, TERC)
112 (29, 36).

113 To enhance the depth of our analyses, we incorporated 13 surface markers that relate to
114 T cell function into a machine learning model for predicting clinical outcomes. The panel included
115 surface markers associated with: activation/proliferation (CD71, HLA-DR, CD38, CD25);
116 degranulation/cytotoxicity (CD94, CD107a); differentiation/memory (loss of CD27, CD28,
117 CD127); Treg/checkpoint inhibitor/exhaustion (CTLA4, PD-1); Tfh/germinal center homing
118 (CXCR5); and mucosal migration (CCR5). While frequencies of T cell subsets could be analyzed
119 individually, the incorporation of expression levels of T cell functional features resulted in over
120 400 subset-marker pairs and required use of computational methods to analyze this high-
121 dimensional dataset. Using a random forest machine learning model, we determined whether T
122 cell subset frequencies and marker expression levels could distinguish clinical outcomes. When

accurate prediction was achieved, feature importance analysis by Gini score (37) determined the most discriminatory features.

Increased frequencies of CD4 TEM Th2 and Th17-lineage subsets and reduced expression of T cell activation/differentiation markers among CT+ or CT+CoP participants compared to uninfected were most informative to the model's correct classification of CT infection. The model was unable to accurately predict endometrial infection but was successful in identifying features associated with absence of reinfection. The most discriminating features were increased frequencies of CD4 TCM Th1, Th17, Th1/17, and Th17 DN cells and increased expression of markers associated with central memory and Th17-lineage. Although these data suggest a protective role for Th17 cells, their contribution to human immunopathogenesis remains undetermined.

RESULTS

CT genital tract infection elicits monocytes, plasma cells, and CD4 T effector memory cells in the periphery.

This study acquired CyTOF data from PBMCs of 82 of the 246 participants enrolled in TRAC (21). The sociodemographic characteristics, self-reported STI diagnoses, and exposure histories of this sub-cohort resemble those previously described for the full cohort (Table S1, S2) (21). Study participants were assigned to infection outcome groups according to the presence of CT or copathogens (CoP) *Neisseria gonorrhoeae* (NG) and/or *Mycoplasma genitalium* (MG) at enrollment: Uninfected (CT- NG- MG-), CT+ (CT+ NG- MG-), or CT+CoP (CT+; NG+ and/or MG+). PBMCs, collected at enrollment and 1 month post treatment (1M), were immunophenotyped by CyTOF using a 33-marker panel (Fig 1A, Table S3). The manual gating hierarchy and definitions used to determine frequencies of each cell type are outlined in Table S4. Frequencies of total myeloid cells, B cells, CD4 T cells, and CD8 T cells were similar across

149 all groups (Fig 1B-E), as were dendritic cells and NK cells (data not shown). However, classical,
150 intermediate, and non-classical monocytes were increased in CT+ and CT+CoP women at
151 enrollment and 1M, with significant increases noted in non-classical monocytes, which are
152 important for antigen presentation (Fig 1F-H). Macrophage frequencies were variable but slightly
153 increased in CT+ participants at enrollment and 1M (Fig S1A). Although frequencies of total B
154 cells were not different at enrollment and 1M, frequencies of plasma cells were increased in CT+
155 but not CT+CoP participants (Fig 1I). Previous studies have shown that CT drives a strong and
156 specific antibody response (21, 38), consistent with a high frequency of plasma cells. Although
157 frequencies of total T cells were not different, frequencies of CD4 TEM cells were significantly
158 increased in CT+ and CT+CoP participants at enrollment and 1M (Fig 1J), while CD8 TEM were
159 only slightly increased in CT+ and CT+CoP participants (Fig 1K). In contrast, CD4 and CD8 TCM
160 frequencies were similar or slightly reduced in CT+ and CT+CoP participants compared to
161 uninfected (Fig 1L, M). Frequencies of CD4 and CD8 stem cell-like memory T cells (TSCM) were
162 moderately decreased in CT+ and CT+CoP participants at enrollment and 1M, likely a reflection
163 of the proportional increase in TEM (Fig S1B, C). Overall, these data demonstrate that CT
164 infection drives increased frequencies of peripheral non-classical monocytes, plasma cells, and
165 CD4 TEM cells, and that responses remain elevated for at least one month after treatment.
166
167 **CT+ participants have increased peripheral frequencies of Th2 and Th17-lineage CD4 TEM**
168 **cells.**
169 We used differential expression of the chemokine receptors CCR6, CXCR3, and CCR4 to
170 define Th1, Th2, and Th17 subsets (Fig 2A, Table S4) (27-29). CD4 T cells, and specifically IFNγ-
171 producing Th1 cells, are broadly accepted as the primary contributors to protection against CT in
172 animals and humans (22, 39), and yet, frequencies of CD4 TEM Th1 cells were not different
173 between CT-infected and uninfected participants (Fig 2B). Instead, frequencies of CD4 TEM Th2
174 cells were significantly increased in CT+ and CT+CoP participants at enrollment and 1M (Fig 2C),

175 consistent with strong antibody responses detected in response to CT (21, 38). Frequencies of
176 CD4 TEM Th17 and Th17 DN cells were also increased in CT+ participants at enrollment and 1M,
177 whereas there was no difference in frequencies of Th1/17 or Th17 DP cells (Fig 2D-G). These
178 differences were maintained in CT+CoP participants for CD4 TEM Th2 but were dampened in the
179 case of Th17 lineage cells (Fig 2C-F), suggesting that coinfection dilutes Th17 responses elicited
180 by CT.

181 We similarly investigated cells comprising the CD8 memory compartment (Tc1, Tc2, Tc17,
182 etc.; Fig 2A, Table S4). Frequencies of CD8 TEM Tc1 and Tc2 were significantly increased in
183 CT+ participants compared to uninfected (Fig 2H-I). There was no change in CD8 TEM Tc1/17
184 (Fig 2J), but Tc17 DN cells trended toward an increase in CT+ participants (Fig 2K), both of which
185 parallel the results for CD4 TEM subsets. Again, these differences were reduced when comparing
186 uninfected to CT+CoP participants. CD8 TEM Tc17 and Tc17 DP were not detected because
187 there was very low co-expression of CCR6 and CCR4 on CD8 T cells. CD4 or CD8 TCM subset
188 frequencies were similar or slightly decreased in CT+ participants (data not shown), following the
189 trend previously observed for frequencies of total CD4 and CD8 TCM (Fig 1L-M). Overall, these
190 results suggest that CT infection drives a peripheral adaptive immune response dominated by
191 CD4 and CD8 TEM Th2/Tc2 and CD4 TEM Th17-lineage cells.

192 We then implemented an unsupervised clustering-based strategy to increase the
193 resolution of our findings obtained by supervised manual gating (Fig 3A-F). Clustering with the *k*-
194 means algorithm partitioned cells based on expression of six phenotypic markers (CD45RA,
195 CCR7, CD95, CCR6, CXCR3, CCR4). We previously defined TSCM cells as
196 CD45RA+CCR7+CD28+CD95+ by manual gating (Table S4), but since expression of CD28 was
197 similar across all clusters, this marker was excluded, and TSCM were defined as
198 CD45RA+CCR7+CD95+. Following clustering, memory and T-helper subtypes were annotated
199 based on marker expression as previously (Table S4, S5; Fig 3A-F). Annotation revealed similar
200 clusters to manual gating overall, but enhanced resolution allowed detection of heterogeneity in

201 TN subsets (Fig 3C, F). CD4 clusters 1, 20, and 22 and CD8 clusters 6, 15, and 22 expressed
202 high levels of both CD45RA and CCR7, representing the classical definition of TN. However, CD4
203 clusters 10, 14, 16 and 19 expressed lower levels of CD45RA and/or CCR7, suggesting that these
204 cells were differentiating from a naïve state to a memory state. Additionally, CD8 clusters 1, 19,
205 4, and 13 were CD45RA+CCR7+ like traditional TN cells but also expressed CXCR3, a marker
206 typically reserved for memory Th1 cells. Studies have shown that naïve human CD8 T cells
207 expressing CXCR3 have greater potential for effector differentiation (40).

208 Clusters with frequencies that were statistically significant ($P<0.05$ by Wilcoxon test)
209 between Uninfected and CT+ or CT+CoP participants at Enrollment or 1 Month are represented
210 in Fig 3G as a distribution of cluster frequencies per participant. Frequencies of CD4 clusters 10
211 (TN with low CCR7), 14 (Transitional TCM), 25 (TEM), and 3 and 9 (TEM Th2) were increased in
212 CT+ and CT+CoP participants compared to uninfected, while frequencies of clusters 1 and 8
213 (classical TN), 5 and 23 (TCM), and 17 (TCM Th2) were decreased, suggesting that CD4 T cells
214 respond to CT by evolving away from a naïve state. Similar immune signatures were observed at
215 enrollment and 1 month, suggesting that changes in immune responses in the periphery are
216 maintained after treatment. These findings were consistent with manual gating where overall CD4
217 TEM (Fig 1J), and especially TEM Th2 cells (Fig 2C), were increased in CT-infected participants
218 compared to uninfected while overall frequencies of CD4 TCM were trending towards decrease
219 (Fig 1L).

220

221 **CT+ and CT+CoP participants express reduced T cell activation and differentiation**
222 **markers on several CD4 and CD8 subsets compared to uninfected participants.**

223 While the unsupervised clustering approach lent confidence to our previous findings and
224 uncovered previously unappreciated nuance in the TN population, it demonstrated limited
225 resolution of low frequency T cell subsets defined by manual gating, such as Th17-lineage cells,
226 that were of particular interest to us. Therefore, for further analysis of marker expression on each

227 T cell subset, we used manually gated subsets. We implemented a random forest machine
228 learning model trained by T cell subset frequencies as well as expression levels of the phenotypic
229 markers used above to define T cell subsets (CD45RA, CCR7, CD95, CCR6, CXCR3, CCR4)
230 and 13 surface markers that relate to T cell function to predict clinical outcomes. The panel
231 included surface markers associated with the following T cell functions: activation/proliferation
232 (CD71, HLA-DR, CD38, CD25); degranulation/cytotoxicity (CD94, CD107a);
233 differentiation/memory (loss of CD27, CD28, CD127); Treg/checkpoint inhibitor/exhaustion
234 (CTLA4, PD-1); Tfh/germinal center homing (CXCR5); and mucosal migration (CCR5). The upper
235 quartile expression of each of these 19 markers was calculated for each CD4 or CD8 manually
236 gated subset (as previously defined, Fig 2, Table S4) and for each participant. These upper
237 quartile expression levels (functional features) were then used in addition to the manually gated
238 frequency features to train the random forest model. This machine learning approach allowed
239 interrogation of over 400 features defined by subset-marker pairs.

240 Classification of uninfected versus CT+ or CT+CoP participants subset frequency and
241 functional features was successful for CD4 (Fig 4A; AUC range 0.96-1.00) or CD8 T cells (Fig4H;
242 AUC range 0.93-0.98). Together, these results suggest that frequency and functional features of
243 CD4 and CD8 T cells contribute to distinction between CT-infected participants and uninfected
244 participants. The similarity between model success at enrollment and 1 month suggests that these
245 signatures were maintained after infection resolution.

246 To identify prominent cellular subsets that contributed to prediction, a measure of feature
247 importance, or Gini score, was assigned to each frequency feature or functional feature (37).
248 Briefly, the Gini score encodes the extent to which each frequency feature or functional feature
249 provides the model with useful information, such that subsets with high Gini scores are likely to
250 be the most useful in distinguishing between clinical outcomes. The Gini scores for predicting
251 classification of infection status are represented in Table S6 for CD4 features and Table S7 for
252 CD8 features. In general, features contributing to accurate classification of uninfected versus CT+

253 and uninfected versus CT+CoP were similar for the models informed by CD4 and CD8 features,
254 consistent with similar accuracy between the two comparisons and weak classification accuracy
255 for CT+ versus CT+CoP. A subset of upper quartile functional features with highest importance
256 by Gini score and significant difference in expression between uninfected and CT+ or CT+CoP
257 by Wilcoxon test (P<0.001) is shown in Fig 4B-G (CD4), I-M (CD8). Several subsets of both CD4
258 and CD8 T cells in CT infected participants expressed lower levels of CD25 (IL-2Rα), CD95 (Fas),
259 and CD38, markers associated with T cell activation and antigen experience, than the same
260 subsets in uninfected participants (Fig 4B-D,I-J). Loss of costimulatory molecules CD27 and
261 CD28 is associated with T cell differentiation, and CD28 expression was higher in some CD4 and
262 CD8 subsets from CT+ and CT+CoP participants than in uninfected participants (Fig 4E, K),
263 suggesting that peripheral blood from infected participants contained less differentiated T cells.
264 Activated, antigen experienced, differentiated T cells may have trafficked to the infected genital
265 tract, resulting in a reciprocal decrease in differentiation markers in the blood. Several subsets,
266 especially TSCM and TCM subsets, also expressed lower levels of CCR7 in CT+ or CT+CoP
267 participants than uninfected participants (Fig 4F, L). This aligns with increased frequencies of
268 effector memory (CCR7-) T cells in CT+ and CT+CoP participants and suggests that the central
269 memory cells of infected participants are shifted more toward an effector phenotype. Additionally,
270 expression of the mucosal homing marker CCR5, a marker known to be important for trafficking
271 of T cells to the genital tract during murine chlamydial infection (41, 42), was decreased in several
272 subsets in CT+ or CT+CoP participants compared to uninfected participants (Fig 4G, M). This
273 CCR5 expression pattern could suggest that CCR5+ T cells had migrated to the genital mucosa
274 at the time of sampling. Frequencies of T cell subsets, especially TEM Th2, TEM Th17DN, and
275 TEM Th17 among CD4 T cells (Table S6) and TEM Tc2 and generic TEM among CD8 T cells
276 (Table S7), were also important contributors to the differentiation of CT+ from uninfected
277 participants. Trends in feature importance analysis were similar for both CD4 and CD8 T cells at

278 1 month (Fig 4, Table S6,7), suggesting maintenance of the T cell signature following infection
279 resolution.

280 Discrimination of CT+ versus CT+CoP participants using CD4 or CD8 frequency
281 demonstrated only weak to moderate success, with greater success from CD4 features (Fig 4A;
282 AUC range: 0.69-0.74) than CD8 (Fig 4H; AUC range: 0.61-0.64). The top contributing feature by
283 Gini score was higher CTLA4 expression on CD4 T cell subsets among CT+CoP participants
284 compared to CT+ participants (Table S6). CTLA4 (cytotoxic T lymphocyte antigen 4) plays an
285 important role in CD4 Treg function by downregulating T cell activation. NG evades host immunity
286 by inducing TGFβ, which is important for Treg development. The increase in CTLA4 expression
287 on CD4 T cell subsets among CT+CoP participants compared to CT+ participants could be
288 reflective of NG-mediated immune evasion.

289

290 **Neither T cell subset frequencies nor functional expression features predict endometrial**
291 **CT infection.**

292 Endometrial biopsies from all participants were tested for CT, NG, and MG infection at
293 enrollment, enabling bisection of the CT+ and CT+CoP groups into CT infections that had
294 ascended to the upper genital tract (Endo+) and CT infections that were limited to the cervix
295 (Endo-). All major cell populations and T cell subsets were defined by manual gating (Table S4),
296 and comparisons of population frequencies performed between Endo+ and Endo- participants.
297 Endo+ participants exhibited significantly increased frequencies of CD4 TSCM, which were not
298 observed in Endo+ CT+CoP participants (Fig S2A). Frequencies of CD8 TEM were not different
299 between Endo+ and Endo- outcomes for CT+ or CT+CoP participants (Fig S2B). Among Endo-
300 participants, there was a trend toward increased frequencies of CD8 TEM Tc17 and TCM Tc17
301 DN cells (Fig S2C-D), suggesting expansion of these cells may contribute to protection from
302 ascending CT infection. However, the random forest model trained on manually gated subset
303 frequencies with the addition of upper quartile expression of functional markers (Fig S2E-F) failed

304 to accurately classify Endo+ from Endo- participants based on CD4 or CD8 features. Overall,
305 there were relatively minimal differences in the frequency and functional marker expression
306 patterns in peripheral blood T cell subsets between participants with or without ascended CT
307 infection.
308
309 **Increased TSCM, CD4 TCM Th1, and CD4 TCM Th17-lineage subsets but decreased CD4**
310 **TEM Th2 cells in peripheral blood are associated with absence of CT reinfection.**
311 Individuals treated for CT often experience repeat infections (6). Nevertheless, evidence
312 exists for partially protective immunity after infection (23). TRAC study participants were followed
313 for one year to determine if they became reinfected (21). We again bisected CT+ and CT+CoP
314 groups into FU+ (positive for CT by diagnostic test or self-report at any follow-up visit) and FU-
315 (completed at least 3 out of 4 follow-up visits, and CT negative by diagnostic test and self-report
316 at all visits). Five out of 20 participants in the CT+ FU+ group tested positive for CT at 1M, while
317 none of the participants in the CT+CoP FU+ group tested positive for CT at 1M. Removing these
318 5 participants in the CT+FU+ group from the analysis did not change the overall differences
319 observed between FU- and FU+ participants.
320 We compared frequencies of manually gated T cell subsets between FU- and FU+
321 individuals. Frequencies of CD4 and CD8 total T cells and CD4 and CD8 TSCM cells were
322 enriched in FU- participants without co-infection at enrollment and remained expanded 1 month
323 post treatment (Fig 5A-D). CD4 Treg cells were enriched in FU- participants at the enrollment visit
324 (Fig 5E). Interestingly, frequencies of CD4 TEM Th2 cells were increased in FU+ participants (Fig
325 5F), suggesting that although these cells are induced by CT infection (Fig 2C), they are not
326 protective. In contrast, frequencies of CD4 TCM Th1, Th17, Th1/17, and Th17 DN cells were
327 increased in FU- participants (Fig 5G-J). These data suggest CD4 and CD8 TSCM and CD4 TCM
328 Th1/Th17 lineage cells, are important for natural immunity against incident CT infection.
329 Differences in frequencies of Th17 family subsets between FU- and FU+ participants were only

330 detected among CT+ and not CT+CoP participants, which suggests that Th17 immune responses
331 specific to CT are diluted by coinfection.

332 We then repeated the comparison using frequencies of clusters generated by *k*-means
333 (Fig 3A-F) and observed trends toward higher frequencies of clusters 15 (TSCM) and 12 and 21
334 (Th17 family) and lower frequencies of clusters 3, 4, and 9 (Th2 TEM) in FU- participants
335 compared to FU+ (Fig S3; see Fig 3C for cluster annotations). These trends were consistent with
336 results obtained by traditional manual gating (Fig 5) and suggest a protective role for TSCM and
337 TCM Th17-lineage cells but not for Th2 TEM. However, there was limited resolution of Th17 family
338 subsets by clustering compared to manual gating.

339

340 **Central memory and Th17-lineage markers distinguish participants who did not become**
341 **reinfected with CT.**

342 We used a random forest machine learning model trained with manually gated T cell
343 subset frequencies and functional expression level features to classify FU+ versus FU- status (Fig
344 6). Classification of FU+ versus FU- was most accurate for CT+ participants at enrollment (CD4
345 ROC AUC = 0.72; CD8 ROC AUC = 0.70) (Fig 6A,B). Analysis of feature importances by Gini
346 score identified cell-types and subset-specific expression patterns that were most distinct (Table
347 S8). Among CT+ participants, the most prominent CD4 features contributing to successful FU-
348 versus FU+ classification were decreased frequencies of Th2 cells and increased frequencies of
349 Th17 family subsets in FU- participants (Table S8). The upper quartile functional features with
350 highest importance by Gini score and significant difference in expression between FU+ and FU-
351 by Wilcoxon test ($P<0.05$) are highlighted in Fig 6C-F. Important CD4 features included increased
352 expression of CCR7 on T cell subsets among FU- participants, consistent with increased central
353 memory, and increased CCR6 expression on Th17 subsets among FU- participants, consistent
354 with increased Th17 lineage (Fig 6C-D). Important CD8 features among FU- participants included
355 increased frequencies of CD8 TSCM, increased expression of CCR7 (similar to the CD4 model),

356 and reduced expression of CD127 (IL-7 receptor, loss of which is associated with differentiation
357 and memory) (Table S8, Fig 6E-F). Overall, the model was most successful in distinguishing FU+
358 from FU- participants among CT+ participants when informed by CD4 T cell frequencies at
359 enrollment. CD8 T cell features and T cell features at 1 month were less informative, and
360 coinfection with NG or MG diluted the success of the model. Overall, these results suggest a
361 protective role for CD4 TCM Th17-lineage cells against CT reinfection.
362

363 **DISCUSSION**

364 Using samples from a uniquely characterized and clinically relevant cohort, this study
365 immunophenotyped responses to CT infection with high dimensional analysis of surface marker
366 expression on PBMCs and found specific T cell subsets correlated with reduced incidence of CT
367 reinfection. Overall, broad immune cell populations in peripheral blood differed between
368 uninfected and CT+ participants, but were similar between CT monoinfection and coinfection with
369 NG and/or MG. CT+ participants displayed increased frequencies of non-classical (CD14-CD16+)
370 monocytes, important for antigen presentation, T cell proliferation, patrolling tissues for signs of
371 injury, and increased Fc receptor-mediated phagocytosis (43).

372 While frequencies of immune populations were largely similar in CT+ participants
373 regardless of coinfection, plasma cells were elevated in participants with CT monoinfection, but
374 their frequencies among CT+CoP participants were similar to uninfected. This is consistent with
375 robust antibody responses to CT infection (25, 44, 45), but this response appears suppressed by
376 coinfection with other STIs. NG opacity proteins bind carcinoembryonic antigen-related cellular
377 adhesion molecule 1 (CEACAM1 or CD66a) on B cells, leading to inhibitory signals which promote
378 B cell death (46). Plasma cells also express CD66a (47). Thus, opacity protein interactions with
379 CD66a could contribute to reduced numbers of plasma cells.

380 CD4 TEM frequencies were strikingly increased in CT+ and CT+CoP participants
381 compared to CD8 TEM frequencies. These data are consistent with our prior investigation of CT

382 antigen specific responses in a subset of TRAC participants where we observed higher levels of
383 IFNγ production and a greater breadth of recognition of CT antigens among CD4 T cells compared
384 to CD8 T cells (22). Lower CD8 TEM frequencies are also consistent with engagement of the PD-
385 1 receptor on CD8 T cells by antigen presenting cells inhibiting generation of CT-specific memory
386 CD8 T cells (48). Frequencies of TEM but not TCM were increased in CT+ compared to uninfected
387 participants in our study.

388 Investigation of CD4 TEM subsets revealed that CT+ participants had higher frequencies
389 of peripheral Th2, Th17, and Th17 DN cells compared to uninfected participants, with unchanged
390 frequencies of Th1 cells. CD4 Th1 cells are classically defined by production of IFNγ and
391 expression of CXCR3 and/or CCR5 (28). They are well-documented contributors to chlamydial
392 clearance and protection from reinfection in mouse models (18, 41, 42) and associate with
393 resistance to reinfection in humans (22, 23). We did not find that Th1 cells were expanded in the
394 peripheral blood of CT+ participants, potentially explaining why CT reinfection is common.
395 However, Th1 cells express the chemokine receptors CXCR3 and CCR5, which bind the
396 chemokines CXCL10 and CCL5 (RANTES), respectively (49). We previously reported detecting
397 these chemokines in cervical secretions of TRAC participants (50), and they are highly expressed
398 in genital tract tissues of *C. muridarum* infected mice (41). In this study, we found decreased
399 expression of CCR5 on T cell subsets among CT+ and CT+CoP participants. Thus, it is possible
400 that despite overall expansion of Th1 cells, they were preferentially recruited to the site of infection
401 and consequently not increased in the periphery. In contrast, peripheral CD4 Th2 cells were
402 significantly enriched in CT+ and CT+CoP participants, consistent with high antibody titers to CT
403 infection that fail to prevent reinfection (6, 21, 25).

404 Th17 cells are highly plastic. They are mainly defined by their expression of CCR6 (51)
405 but can gain or lose chemokine receptors and their associated effector functions (e.g., cytokine
406 production) (32). Th17 cells can be pathogenic, but they are also important contributors to
407 protection against mucosal pathogens (52, 53). We used expression of CCR6, CXCR3, and

408 CCR4 to define four subsets of Th17 family cells: classical Th17, Th1/17, Th17 DP, and Th17 DN
409 (29). We found increased frequencies of classical Th17 and Th17 DN cells in CT+ participants
410 compared to uninfected, which were reduced by coinfection. Frequencies of parallel populations
411 of CD8 TEM cells including Tc2 and Tc17 DN cells were also increased in CT+ and CT+CoP
412 participants, albeit to a lesser extent than CD4 counterparts. Co-expression of CCR6 and CCR4
413 was extremely rare on CD8 T cells, and therefore Tc17 and Tc17 DP populations were excluded
414 from analysis.

415 An important goal of the study was to analyze expression of additional surface markers
416 associated with activation/proliferation (CD71, HLA-DR, CD38, CD25); degranulation/cytotoxicity
417 (CD94, CD107a); differentiation/memory (loss of CD27, CD28, CD127); Treg/checkpoint
418 inhibitor/exhaustion (CTLA4, PD-1); Tfh/germinal center homing (CXCR5); and mucosal
419 migration (CCR5) on each T cell subset and to compare expression patterns between clinical
420 outcomes. While T cell subsets could be analyzed individually, investigating differences in
421 expression of phenotypic and functional markers on each subset resulted in over 400 subset-
422 marker pairs. Utilizing a random forest machine learning model allowed us to determine whether
423 T cell subset frequencies and functional expression features were distinct between observed
424 clinical outcomes of CT infection (Uninfected, CT+, CT+CoP), ascension (Endo+/-), or reinfection
425 (FU+/-). If the random forest model was successful using a combination of frequency and
426 functional features to classify clinical outcomes, subsequent feature importance analysis
427 identified which of the T cell subset frequencies and or subset-marker expression pairs were most
428 informative to the model. The model successfully classified CT+ or CT+CoP from uninfected
429 participants using CD4 or CD8 features. Increased frequencies of CD4 TEM Th2, Th17, and Th17
430 DN cells and distinct expression patterns of CD25, CD95, CD38, CCR7, CD28, and CCR5 were
431 defined as important features for the model. CD25 is the IL-2 receptor and is crucial for T cell
432 activation and proliferation. CD95, or Fas, is involved in apoptosis but is also a marker of antigen
433 experience and stem-cell memory. CD38 is a receptor involved in cell activation, adhesion, and

434 signaling, and is commonly associated with T cell activation. All three of these markers were
435 expressed at lower levels in CT+ or CT+CoP participants than uninfected participants. Lower
436 levels of CCR7 expression on T cells among CT+ and CT+ CoP participants was consistent with
437 higher frequencies of TEM. Additionally, higher levels of CD28 expression on T cell subsets
438 among CT+ or CT+CoP participants suggested less differentiation. Lower levels of CCR5, a
439 molecule important for trafficking of effector T cells to the female genital tract during murine
440 infection (42) and slightly associated with reduced chlamydia burden in humans (54), were also
441 observed on T cell subsets among CT+ or CT+CoP participants. Together, these data suggested
442 that even though CD4 TEM cells were increased in peripheral blood of participants with CT
443 infection, these T cells were not more activated or differentiated than the same subsets in
444 uninfected participants. The chronicity often associated with chlamydial infections could lead to a
445 reduced differentiation state associated with T cell suppression in the periphery, as well as
446 promote ongoing exit of activated, differentiated T cells from the peripheral blood to the infected
447 mucosa. Supporting this hypothesis of reciprocal decrease in activation/homing markers from
448 peripheral blood compared to the site of infection, other studies found that CD38 and/or CCR5
449 expression was higher among CD4 and CD8 T cells isolated from the genital tract than from
450 peripheral blood of CT+ women (55, 56).

451 Since CT ascension is a prerequisite for generation of upper tract disease, we sought to
452 identify T cell subsets and expression patterns associated with ascended infection. Peripheral
453 immune cell populations were similar overall between participants with endometrial CT infection
454 and those that had CT infection limited to the cervix. CD8 TEM Tc17 DN cells trended toward
455 increased frequencies in Endo- participants, consistent with our previous study showing that CT-
456 specific CD8 T cells weakly associated with reduced CT ascension (22). Frequencies of CD4
457 TSCM were significantly increased in Endo+ as well as FU- participants. This is consistent with
458 our prior finding that endometrial infection was associated with reduced risk of follow-up CT
459 infection (21), likely because the increased bacterial burden associated with endometrial infection

460 drives a more robust memory response. However, a random forest model trained with manually
461 gated T cell subset frequencies and functional expression features was unsuccessful in
462 classifying Endo+ from Endo- CT infections, suggesting that immune responses in the periphery
463 are too weak to distinguish whether a participant had ascended CT infection. Evaluation of
464 immune cell populations in the genital tract will likely be necessary to define the specific features
465 of immunity that prevent or limit ascending infection. Future studies will investigate immune cells
466 from cervical and endometrial TRAC samples.

467 Several T cell subsets were found to be strongly associated with the presence or absence
468 of recurrent CT infection. Sustained increased frequencies of total CD4 T cells in FU- participants
469 may indicate developing natural immunity in some participants. We also detected increased
470 frequencies of CD4 and CD8 TSCM in FU- individuals. TSCM resemble naïve T cells in their
471 expression of CD45RA and CCR7, but they also express CD95, a marker of antigen experience
472 and memory (57). They are minimally differentiated memory cells that have a high propensity for
473 stemness and self-renewal as well as the multipotency to differentiate into a broad range of
474 memory T-helper phenotypes (57). This may explain why they are associated with a FU- outcome.
475 The importance of the nature of the T cell response is demonstrated by the contrasting subsets
476 that were expanded in FU- and FU+ participants. Higher frequencies of TEM Th2 cells, which
477 were strikingly increased in CT+ participants, were associated with a FU+ outcome, while higher
478 frequencies of TCM Th1 cells, which were not different between CT+ and uninfected participants,
479 were associated with FU- status. Overall, an important role for Th1 and Th17-lineage cells in
480 protection from reinfection is suggested by the striking increase in frequencies of TCM Th1 and
481 Th17 cells among FU- participants, along with increased frequencies of TCM Th1/17 and TCM
482 Th17 DN cells. The increased stem-ness and differentiation potential of TSCM and Th17 lineage
483 cells (31, 58), especially Th17 DN (29, 36), is potentially impactful because their plasticity may
484 allow them to adapt quickly to fight infection. A recent study detected enrichment of
485 *Mycobacterium tuberculosis* (Mtb)-specific Th17-like cells in people who were resistant to Mtb

486 infection following exposure compared to those who were latently infected (59). Similar trends
487 were observed when comparing frequencies of T cell subsets as defined by *k*-means clusters.
488 Overall, clusters annotated as TSCM and Th17 lineage were increased in FU- participants and
489 clusters annotated as TEM Th2 were decreased in FU- participants.

490 Our finding that TCM subsets (TCM Th1, TCM Th17, TCM Th1/17 and TCM Th17 DN
491 cells) rather than TEM subsets were associated with FU- status, is consistent with our detection
492 of CT antigen-specific CD4 TCMs months post treatment among TRAC participants (60). In
493 addition, an analysis of *C. pneumoniae*-specific peripheral blood T cells in healthy seropositive
494 donors found that antigen-specific CD4 TCM were enriched compared to TEM (61). Thus, TCM
495 appear to make up a greater frequency of antigen-specific CD4 T cells important for protection
496 against subsequent infection, although we found that active CT infection drives expansion of TEM
497 cells that remain 1M post-treatment,.

498 A random forest machine learning model trained to classify FU- versus FU+ informed by
499 manually gated T cell subset frequencies and functional expression features confirmed that the
500 most distinct differences were increased frequencies of TCM Th17, Th17 DN, Th1/17, and Th1
501 cells and decreased frequencies of TEM Th2 cells in FU- participants. Additionally, we identified
502 distinct expression patterns of surface markers including CCR7 and CCR6. In both the CD4 and
503 CD8 models for CT+ participants, higher expression of CCR7 was observed on several subsets
504 among FU- participants. Since CCR7 is a defining marker of central memory T cells and is
505 important for trafficking to the lymph node, it suggests that central memory cells with the
506 propensity to be long-lived are associated with protection from reinfection. Higher expression of
507 CCR6 on subsets of CD4 T cells among FU- participants suggests an importance not only for
508 Th17 lineage cells, which were defined by the presence of CCR6, but also for its surface
509 expression levels. CCR6 binds CCL20 is highly expressed in mucosal tissues, driving migration
510 of Th17 cells to the site of infection. Additionally, high expression of CCR6 is associated with
511 increased secretion of Th17 effector cytokines and chemokines such as IL-17A, IL-17F, IL-22,

512 and CCL20, and increased expression of *IL17A*, *IL17F*, *IL22*, *IL23R*, *CCL20*, *RORA*, and *RORC*
513 (58).

514 The finding that CD4 Th1 and Th17 subsets were associated with reduced reinfection
515 informs vaccine development. Different vaccine platforms and adjuvants can shape T-helper
516 responses (62). Vaccination of mice with a recombinant chlamydial antigen with CAF01 (63), or
517 a STING agonist (64) elicited Th1 and Th17 responses that led to reduced chlamydial burden
518 upon challenge. In contrast, a viral vector vaccine failed to elicit antigen-specific CD4 T cell
519 responses or protection (65).

520 Our study is limited because although we defined immune cell populations increased in
521 CT+ participants compared to uninfected participants, the cells were not restimulated with CT
522 antigens to confirm their CT-specificity. However, immunophenotyping PBMCs directly without
523 restimulation allowed us to confidently define T cell subsets based on chemokine receptors
524 without the risk of downregulation upon restimulation. An important open question is whether
525 Th17 and Th17 DN cells produce CT-specific IL-17 and/or IFNγ, and this should be addressed in
526 future studies. Another limitation of this study is that cell population frequencies could not be
527 correlated with endometrial CT infection. Many of the women participating in TRAC selected
528 contraceptive methods that promote a thin uterine lining, making it difficult to acquire sufficient
529 tissue for evaluation of pathology. However, ongoing studies are investigating local T cell
530 responses in endometrial tissue sections from participants with sufficient tissue.

531 Overall, we found that frequencies of CD4 TEM, specifically Th2, Th17, and Th17 DN
532 subsets, were enriched in individuals with active CT infection. We also identified that increased
533 frequencies of CD4 TCM Th1, Th1/Th17, and Th17 were associated with individuals who had no
534 documented reinfection over the course of a year. Together, these findings increase our
535 understanding of natural immunity to CT and inform strategies for future vaccine design.

536

537 **METHODS**

538

539 **Sex as a Biological Variable**

540 The TRAC cohort was restricted to persons with a cervix and uterus because the objective of the
541 study was to determine factors associated with ascension of CT from the cervix to the
542 endometrium.

543

544 **Study Population**

545 This study involved a subset of 82 of the 246 participants enrolled in the T Cell Response Against
546 Chlamydia (TRAC) cohort (21). TRAC participants were asymptomatic and presented to the clinic
547 for STI screening due to sexual behavior that put them at risk for CT (16). None had signs or
548 symptoms of PID. At enrollment, all participants received antibiotics recommended by the Centers
549 for Disease Control and Prevention for treatment of CT and NG, and individual participants that
550 tested positive for STI at any follow-up visit were retreated (16).

551

552 **Definition of Clinical Classifications**

553 ***Uninfected, CT+, and CT+CoP Groups:***

554 Nucleic acid amplification testing (NAAT) was used to test cervical swabs collected from study
555 participants at enrollment for CT, NG, and MG. Participants whose swabs tested negatively for all
556 3 STIs at enrollment were classified as “uninfected”. Participants testing positively for CT but
557 negative for NG and MG were classified as “CT+”. Enrollees testing positively for CT and NG,
558 MG, or both were classified as “CT+CoP”. Of 82 total TRAC participants studied, 12 were
559 uninfected, 44 were CT+, and 26 were CT+CoP (Table S9). Of 26 CT+CoP participants, 11 were
560 CT+ NG+ MG-, 12 were CT+ MG+ NG-, and 3 were CT+ NG+ MG+.

561 ***Endometrial positive and endometrial negative CT:***

562 Endometrial biopsies were also tested by NAAT to determine whether infection was limited to the
563 cervix or had ascended to the upper genital tract. Within CT+ and CT+CoP groups, enrollees

564 testing positively for CT at the cervix but not the endometrial biopsy were classified as Endo-,
565 while those testing positively for CT at both sites were classified as Endo+. No participants tested
566 positive for CT solely in the endometrial sample. Of 44 CT+ participants, 19 were Endo+ and 25
567 were Endo-. Of 26 CT+CoP participants, 11 were Endo+ and 15 were Endo- (Table S10).

568 ***Follow-up CT positive and follow-up CT negative:***

569 NAAT testing was performed on cervical brush samples at enrollment and at 1, 4, 8, and 12 month
570 follow-up visits. Participants with a positive cervical CT test at enrollment and a positive cervical
571 CT test or self-reported CT diagnosis from an outside clinic in the time since last visit at any follow-
572 up visit (regardless of how many visits they completed) were classified as follow-up positive (FU+).
573 Participants that tested positive for cervical CT at enrollment, completed at least 3 out of 4 follow-
574 up visits, tested negative for cervical CT at all follow-up visits, and self-reported no CT diagnosis
575 from an outside clinic in the time since last visit at any follow-up visit were classified as follow-up
576 negative (FU-). 9 participants attended less than 3 follow-up visits (FU N/A) and were excluded
577 from follow-up analysis. Of CT+ participants, 19 were FU-, 21 were FU+, and 4 were FU N/A. Of
578 CT+CoP pariticpants, 9 were FU-, 12 were FU+, and 5 were FU N/A (Table S10). All uninfected
579 samples were FU- and the group was therefore not bisected.

580

581 **PBMC CyTOF**

582 Peripheral blood mononuclear cells (PBMCs) were isolated from whole blood via a density
583 gradient using lymphocyte separation media (Corning). Cells were frozen in 90% FBS with 10%
584 DMSO and stored in liquid nitrogen until thawed for analysis. Dead cells were stained with 1 μM
585 Cell-ID Cisplatin 198-Pt (Fluidigm). Cells were stained with an antibody cocktail comprising 33
586 antibodies (Table S3) and human Fc block (eBioscience) in cell staining buffer (CSB, Fluidigm),
587 followed by staining with secondary antibody (Anti FITC-144Nd). Stained cells were fixed with 2%
588 PFA in PBS followed by 1:2000 Cell-ID Intercalator-Ir-125 μM (Fluidigm) in fix and perm buffer
589 (Fluidigm). Cell concentrations were adjusted to 0.5 x 10^6 cells/mL in CAS (Cell Acquisition

590 Solution; Fluidigm) with Four Element Calibration beads (Fluidigm). Samples were acquired on a
591 Helios instrument (Fluidigm) by the UNC Mass Cytometry Core.
592
593 **Data Analysis**
594 ***Defining T cell subsets by manual or automated gating.***
595 Raw CyTOF files were cleaned using the Pathsetter algorithm (66).
596 *Manual gating.* Cleaned FCS files were analyzed in Cytobank (Beckman Coulter) and FlowJo
597 (BD) software. The manual gating strategy and surface marker designations selected for immune
598 cell phenotypes are summarized in Table S4. Some gated CD8 populations were undetectable
599 (<0.1% of T cells) and were excluded from further analysis (Table S4).
600 *Automated gating.* Batch effects were removed from the dataset using the combat algorithm (67).
601 A random subset of 15,000 cells was selected from each sample, resulting in 1,230,000 total cells
602 for further analysis. We then performed automated cell-population discovery followed by feature
603 engineering, as outlined in Stanley et al. (68). CD4 and CD8 T cell populations were each
604 partitioned into 25 clusters using *k*-means clustering. Briefly, each cell was represented by
605 expression of six phenotypic markers (CD45RA, CCR7, CD95, CCR6, CXCR3, CCR4), which
606 were input to the clustering algorithm. Moreover, each cell was labelled in terms of one of the 25
607 clusters, based on its expression across these six phenotypic markers (Table S5). Modeled after
608 canonical immune features that are often computed based on manual gates, we engineered per-
609 cluster frequency features, where for each sample, we computed the proportion of its cells
610 assigned across each of the 25 clusters.
611 ***Feature engineering***
612 While previous studies have created functional features based on the mean or median expression
613 of functional markers in each cluster (68, 69), some functional markers in this dataset were
614 expressed by a small proportion of cells, such that there were some cases where median
615 expression of a marker was zero. Therefore, we computed the upper quartile expression of each

616 marker across each manually gated T cell subset for each sample. For samples with 0.1% events
617 assigned to a particular subset, the upper quartile features were imputed with the mean of each
618 subset-marker feature. Additionally, subset-marker features for which upper quartile expression
619 was below 2 (range 0-500) were excluded from the training dataset. Upper quartile features were
620 arcsinh transformed with co-factor 5. In all, 19 frequency features and 233 upper quartile
621 expression features were included in the CD4 model training dataset, and 14 frequency features
622 and 174 upper quartile expression features were included in the CD8 model training dataset.

623 ***Model training***

624 We trained random forest models using engineered manually gated frequency features and upper
625 quartile expression features to classify and predict clinical outcomes for each participant. For each
626 clinical outcome comparison, we performed 30 trials of five-fold cross validation, whereby each
627 fold used 80% of samples for training and 20% for testing. Reported area under the ROC curves
628 (AUCs) represent the mean accuracy across 30 trials. For biological interpretability of each of the
629 trained random forest models, we further computed per-feature importance measures with the
630 Gini index (37). A higher per-feature Gini score implied that the feature was important in the
631 clinical prediction task. We further imposed directionality to the computed per-feature Gini scores
632 to readily identify condition-specific, predictive features by first choosing a particular clinical
633 outcome to be the *positive (negative) direction*, such that we multiplied the Gini score by 1 (-1), if
634 it was associated with the positive (negative) direction.

635

636 **Statistics**

637 Statistical calculations were computed in R and python. Statistical tests used are outlined in the
638 appropriate figure legends. $P<0.05$ was used to determine a significant result. Significance
639 symbols used in figures are defined as: *$P<0.05$, **$P<0.01$, ***$P<0.001$, ****$P<0.0001$.

640

641 **Study Approval**

642 The institutional review boards for human research of the University of Pittsburgh and the
643 University of North Carolina approved the study protocol. All participants provided written
644 informed consent at the time of enrollment and agreed to be contacted to return for follow-up visits
645 1, 4, 8, and 12 months after enrollment.
646

647 **Data and Code Availability**

648 CyTOF data are available for download via ImmPort at https://www.immport.org under study
649 accession number SDY2772. Manually gated T cell subset frequencies reported in the Figures
650 are available in Supporting Data Values. Code for (1) automated gating, (2) feature engineering,
651 (3) random forest training and Gini feature importance analysis, and (4) upper quartile expression
652 boxplots are available via GitHub at https://github.com/ksyount/TRAC_PBMC_CyTOF.
653

654 **AUTHOR CONTRIBUTIONS**

655 KSY, AK, XZ, NS, and TD designed research studies. KSY and AK conducted experiments. HCW
656 and SLH collected participant samples and performed diagnostic testing. TP advised regarding
657 mass cytometry panel design. KSY, CC, CL, NVM, and CBB analyzed data. KSY, CC, NS, CMO,
658 and TD wrote the manuscript. All authors reviewed the manuscript.
659

660 **ACKNOWLEDGEMENTS**

661 We thank the women who agreed to participate in this study; Ingrid Macio, Melinda Petrina, Carol
662 Priest, Abi Jett, and Lorna Rabe, for their efforts in the clinic and the microbiology laboratory; and
663 the staff at the Allegheny County Health Department STD Clinic. We thank Jason Whitmire for
664 critical review of the manuscript. We also thank Marie Iannone and the UNC Mass Cytometry
665 Core, which is supported by the University Cancer Research Fund (UCRF) and UNC Cancer
666 Center Core Support Grant #P30CA016086. This work was supported by the National Institute of
667 Allergy and Infectious Diseases through U19AI144181 (TD) and R01AI170959 (CMO). KSY

received funding from UNC STI/HIV postdoctoral training grant T32AI007001 and the American Association of Immunologists Intersect Fellowship Program for Computational Scientists and Immunologists.

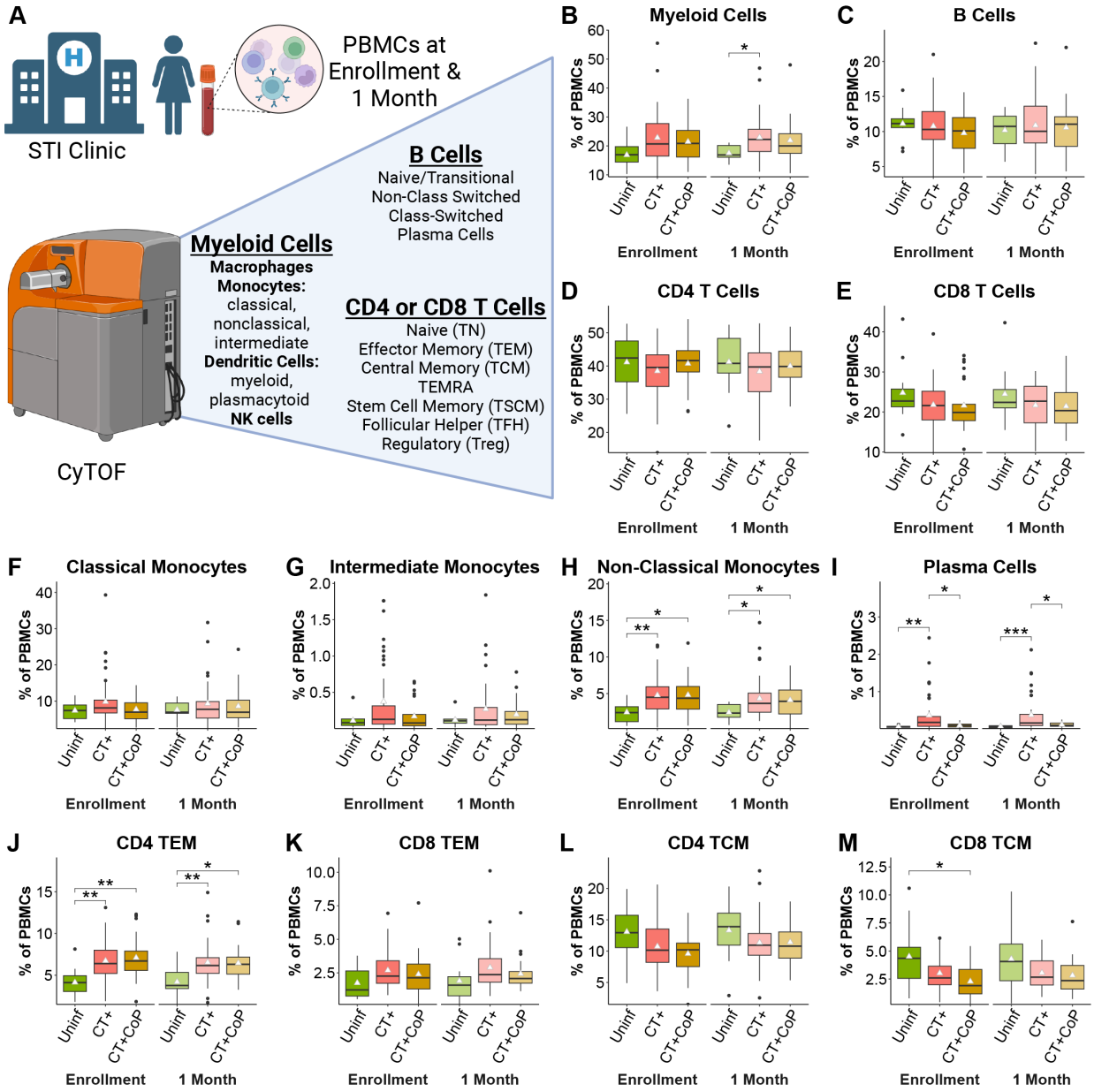


Figure 1. CT genital tract infection elicits monocytes, plasma cells, and CD4 T effector memory cells. (A) Major immune populations defined by manual gating in peripheral blood of 82 participants at enrollment and 1 month post-treatment. (Created with Biorender.com) (B-M) Population percentages of total PBMCs. White triangle represents mean. Significance determined by Dunn's test with Bonferroni correction for multiple comparisons.

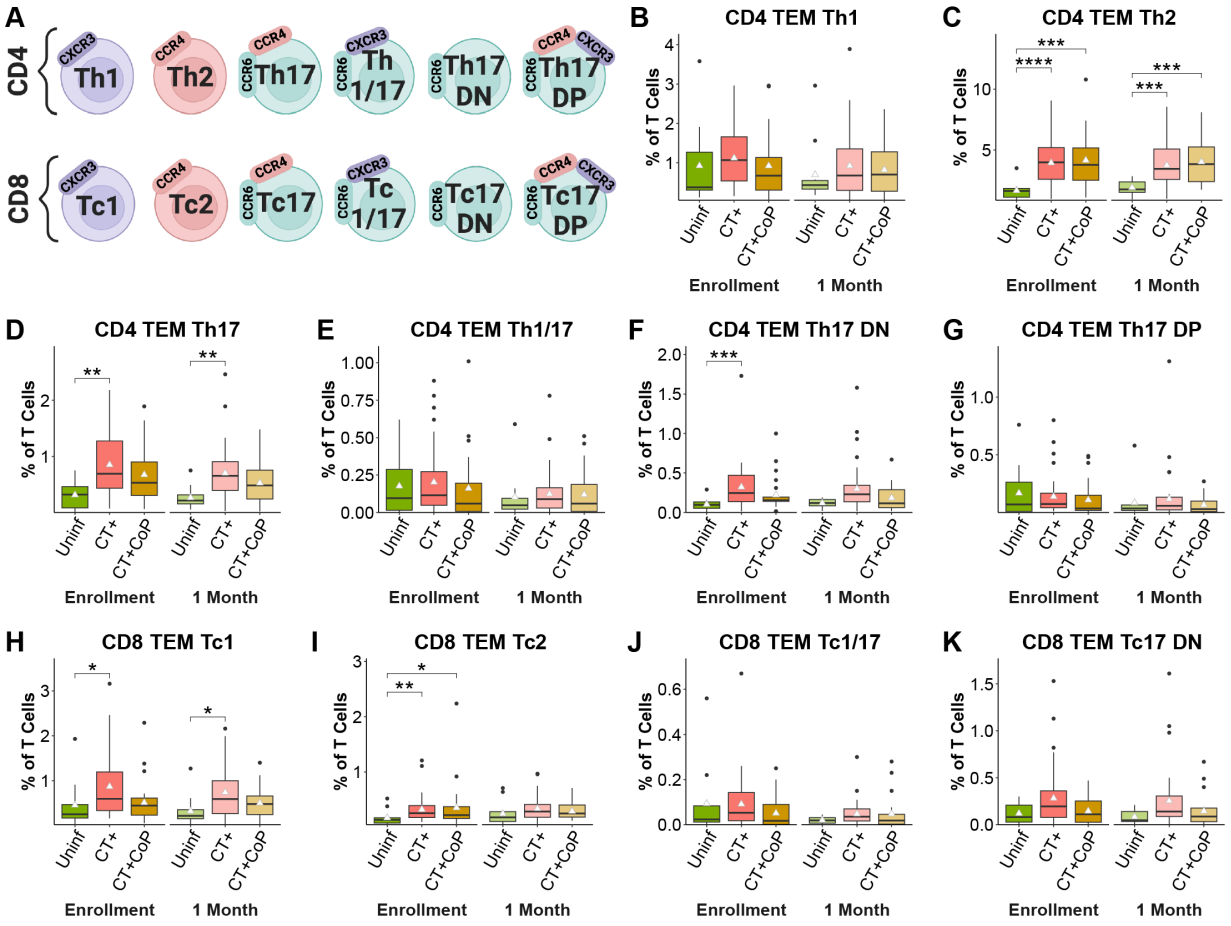


Figure 2. Increased frequencies of Th2 and Th17-lineage CD4 TEM cells among CT+ participants compared to uninfected participants. (A) Subsets of CD4 and CD8 TEM were identified by expression of chemokine receptors (Created with Biorender.com). (B-K) Population percentages of total T Cells. Triangle represents mean. Significance determined by Dunn's test with Bonferroni correction for multiple comparisons.

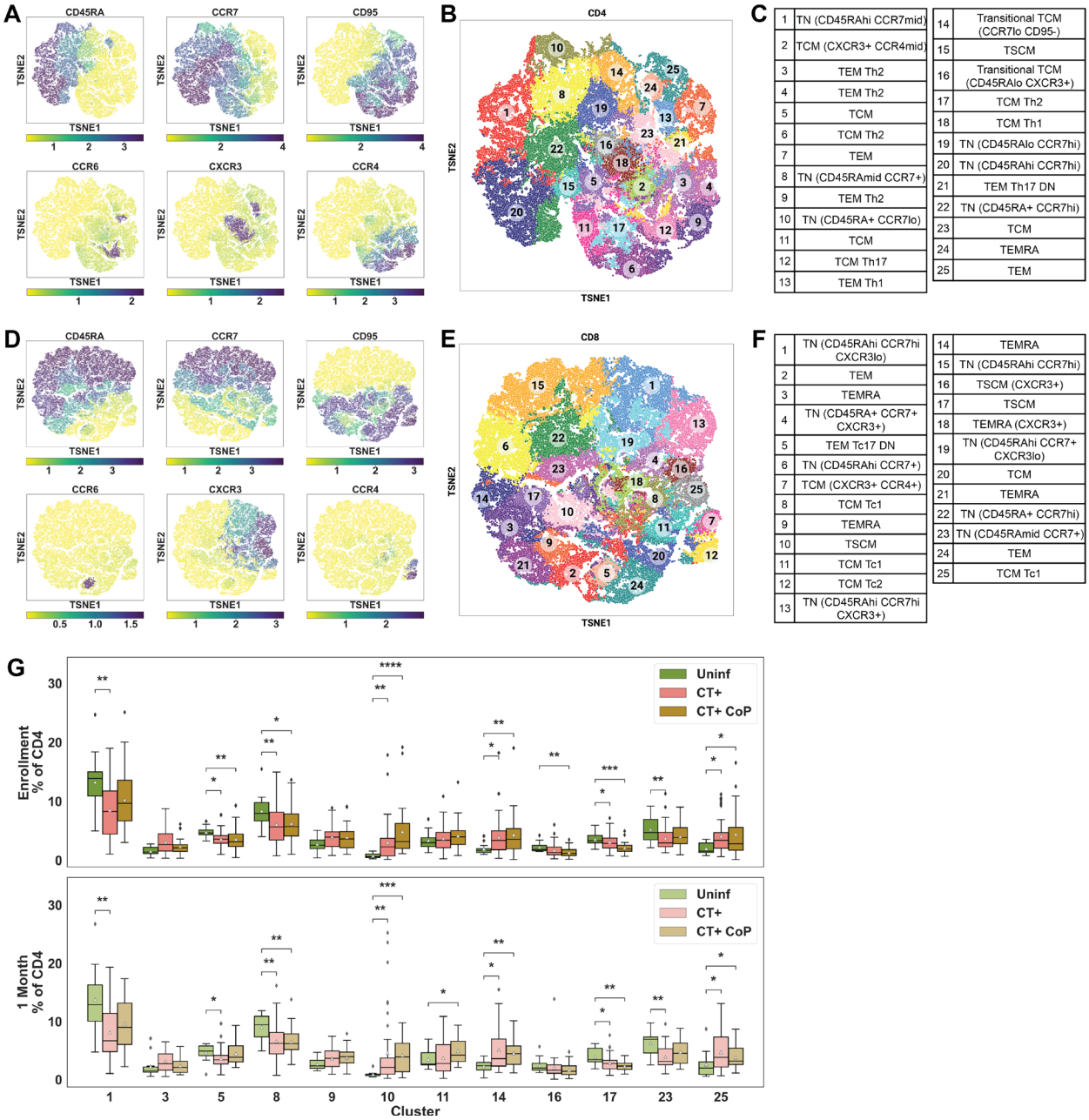


Figure 3. CD4 but not CD8 automated gating frequency features differentiate CT+ and CT+CoP participants from uninfected participants.

(A-F) Automated gating by *k*-means clustering and cluster annotation. (A-B) Expression of phenotypic markers (CD45RA, CCR7, CD95, CCR6, CXCR3, and CCR4) was used to partition CD4 T cells into 25 *k*-means clusters. (C) Resulting CD4 clusters were annotated based on marker expression (See Table S4, S11). (D-E) CD8 T cells were partitioned into 25 *k*-means clusters based on the expression of the same phenotypic markers as for CD4 T cells and (F) were annotated. (G) Boxplots representing frequencies of CD4 clusters with significant differences in frequency (Wilcoxon P<0.05). Statistical comparisons between CT+ or CT+CoP and Uninf by Wilcoxon test. White triangle represents mean.

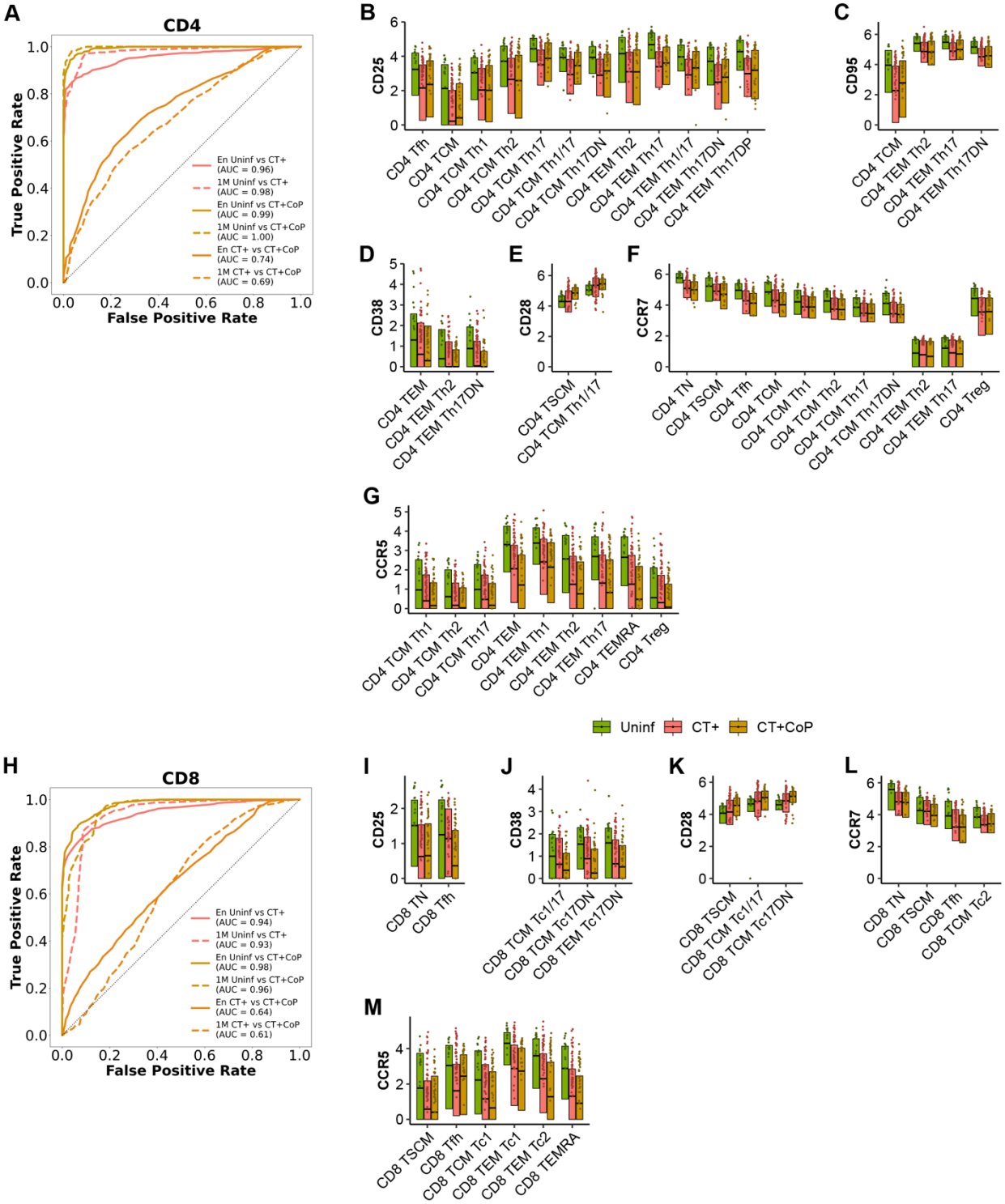


Figure 4. CT+ and CT+CoP participants express reduced T cell activation and differentiation markers on several CD4 and CD8 subsets compared to uninfected participants. A random forest model trained on (A) CD4 or (H) CD8 manually gated subset frequency features and upper quartile expression features was used to discriminate between uninfected, CT+, and CT+CoP outcomes at enrollment (En) or 1 month (1M). Receiver operating characteristic (ROC) curves describe performance of the model. Top markers of (B-G) CD4 or (I-

707 M) CD8 T cells were identified by feature importance analysis (see Suppl tables 6-7). Features
708 were included on the plot if they were statistically significant (as defined by P<0.001 by Wilcoxon
709 rank sum test between outcomes) for at least one comparison between uninfected versus CT+
710 and/or uninfected versus CT+CoP. Boxplots represent the overall distribution of marker
711 expression. Individual points represent the upper quartile expression for each participant in the
712 group. (AUC = area under the ROC curve)

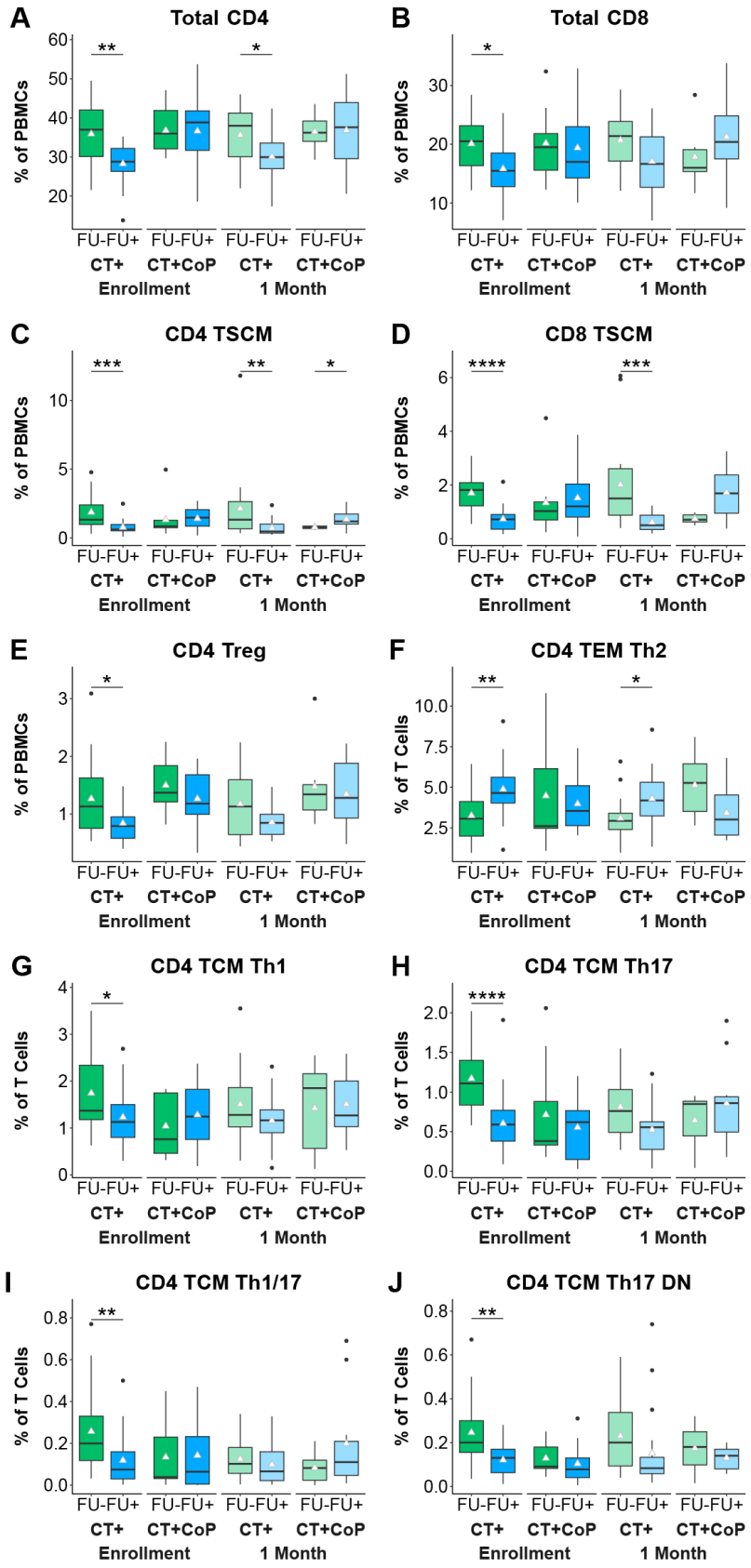


Figure 5. CD4 TCM Th1 and Th17 family subsets are enriched in peripheral blood of participants without documented CT reinfection. Frequencies of manually gated T cell subsets were stratified by follow-up negative (FU-) and follow-up positive (FU+) status. (A-J) T cell subset percentages. White triangle represents mean. Statistical significance determined by Wilcoxon rank-sum test.

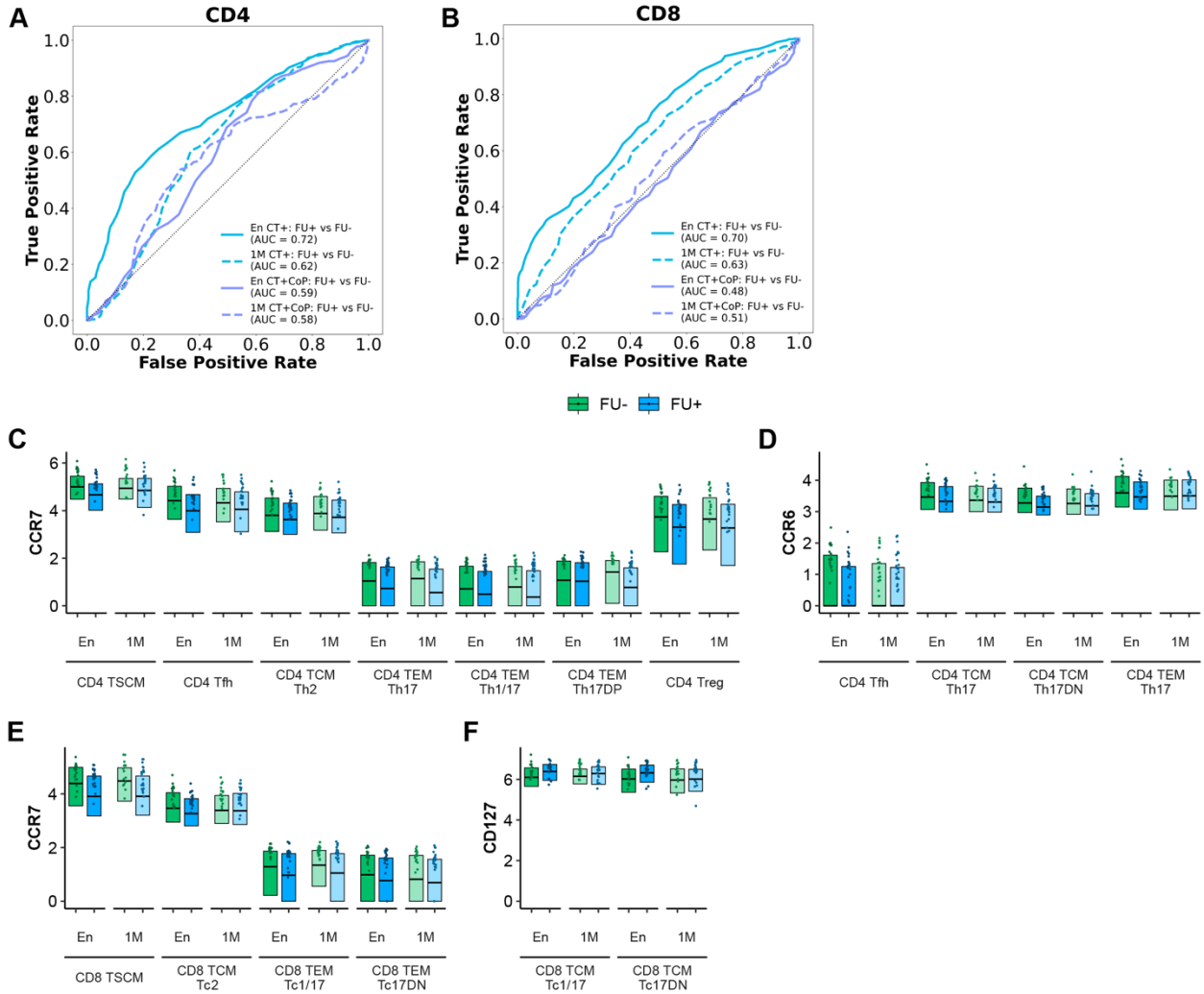


Figure 6. Increased central memory and Th17-lineage markers distinguish participants without documented CT reinfection. A random forest model trained on (A) CD4 or (B) CD8 manually gated subset frequency features and upper quartile expression features was used to discriminate between FU+ and FU- groups at enrollment (En) or 1 month (1M). Receiver operating characteristic (ROC) curves visualize the performance of the model. Top markers of (C-D) CD4 or (E-F) CD8 T cells among CT+ participants were identified by feature importance analysis (see Table S9). Features were included on the plot based on a significance threshold of P<0.05 by Wilcoxon rank sum test for FU+ verus FU- at Enrollment. Boxplots represent the overall distribution of marker expression. Individual points represent the upper quartile expression for each participant in the group. (AUC = area under the ROC curve)

743 **BIBLIOGRAPHY**

744 1. Newman L, Rowley J, Vander Hoorn S, Wijesooriya NS, Unemo M, Low N, et al.
745 Global Estimates of the Prevalence and Incidence of Four Curable Sexually
746 Transmitted Infections in 2012 Based on Systematic Review and Global Reporting.
747 *PLoS One.* 2015;10(12):e0143304.
748 2. Haggerty CL, Gottlieb SL, Taylor BD, Low N, Xu F, and Ness RB. Risk of sequelae
749 after Chlamydia trachomatis genital infection in women. *J Infect Dis.* 2010;201
750 Suppl 2:S134-55.
751 3. Stamm WE. Chlamydia trachomatis infections: progress and problems. *JInfectDis.*
752 1999;179 Suppl 2:S380-S3.
753 4. Wiesenfeld HC, Hillier SL, Meyn LA, Amortegui AJ, and Sweet RL. Subclinical
754 pelvic inflammatory disease and infertility. *Obstetrics and gynecology.*
755 2012;120(1):37-43.
756 5. Geisler WM, Lensing SY, Press CG, and Hook EW, 3rd. Spontaneous resolution
757 of genital Chlamydia trachomatis infection in women and protection from
758 reinfection. *J Infect Dis.* 2013;207(12):1850-6.
759 6. Batteiger BE, Tu W, Ofner S, Van Der Pol B, Stothard DR, Orr DP, et al. Repeated
760 Chlamydia trachomatis genital infections in adolescent women. *J Infect Dis.*
761 2010;201(1):42-51.
762 7. Yang X, Hayglass KT, and Brunham RC. Different roles are played by alpha beta
763 and gamma delta T cells in acquired immunity to Chlamydia trachomatis
764 pulmonary infection. *Immunology.* 1998;94(4):469-75.
765 8. Morrison RP, Feilzer K, and Tumas DB. Gene knockout mice establish a primary
766 protective role for major histocompatibility complex class II-restricted responses in
767 Chlamydia trachomatis genital tract infection. *Infect Immun.* 1995;63(12):4661-8.
768 9. Magee DM, Igietseme JU, Smith JG, Bleicker CA, Grubbs BG, Schachter J, et al.
769 Chlamydia trachomatis pneumonia in the severe combined immunodeficiency
770 (SCID) mouse. *Reg Immunol.* 1993;5(6):305-11.
771 10. Su H, Feilzer K, Caldwell HD, and Morrison RP. Chlamydia trachomatis genital
772 tract infection of antibody-deficient gene knockout mice. *Infect Immun.*
773 1997;65(6):1993-9.
774 11. Morrison SG, Su H, Caldwell HD, and Morrison RP. Immunity to murine Chlamydia
775 trachomatis genital tract reinfection involves B cells and CD4(+) T cells but not
776 CD8(+) T cells. *Infect Immun.* 2000;68(12):6979-87.
777 12. Beatty WL, Belanger TA, Desai AA, Morrison RP, and Byrne GI. Tryptophan
778 depletion as a mechanism of gamma interferon-mediated chlamydial persistence.
779 *Infect Immun.* 1994;62(9):3705-11.
780 13. Rank RG, and Whittum-Hudson JA. Protective immunity to chlamydial genital
781 infection: evidence from animal studies. *J Infect Dis.* 2010;201 Suppl 2:S168-77.
782 14. Perry LL, Feilzer K, and Caldwell HD. Immunity to Chlamydia trachomatis is
783 mediated by T helper 1 cells through IFN-gamma-dependent and -independent
784 pathways. *JImmunol.* 1997;158(7):3344-52.
785 15. Stary G, Olive A, Radovic-Moreno AF, Gondek D, Alvarez D, Basto PA, et al.
786 VACCINES. A mucosal vaccine against Chlamydia trachomatis generates two
787 waves of protective memory T cells. *Science.* 2015;348(6241):aaa8205.

16. Rixon JA, Depew CE, and McSorley SJ. Th1 cells are dispensable for primary clearance of Chlamydia from the female reproductive tract of mice. *PLoS Pathog.* 2022;18(2):e1010333.
17. Lampe MF, Wilson CB, Bevan MJ, and Starnbach MN. Gamma interferon production by cytotoxic T lymphocytes is required for resolution of Chlamydia trachomatis infection. *Infect Immun.* 1998;66(11):5457-61.
18. Poston TB, Qu Y, Girardi J, O'Connell CM, Frazer LC, Russell AN, et al. A Chlamydia-Specific TCR-Transgenic Mouse Demonstrates Th1 Polyfunctionality with Enhanced Effector Function. *J Immunol.* 2017;199(8):2845-54.
19. Rixon JA, Fong KD, Morris C, Nguyen AT, Depew CE, and McSorley SJ. Elimination of Chlamydia muridarum from the female reproductive tract is IL-12p40 dependent, but independent of Th1 and Th2 cells. *PLoS Pathog.* 2024;20(1):e1011914.
20. Scurlock AM, Frazer LC, Andrews CW, Jr., O'Connell CM, Foote IP, Bailey SL, et al. Interleukin-17 contributes to generation of Th1 immunity and neutrophil recruitment during Chlamydia muridarum genital tract infection but is not required for macrophage influx or normal resolution of infection. *Infect Immun.* 2011;79(3):1349-62.
21. Russell AN, Zheng X, O'Connell CM, Taylor BD, Wiesenfeld HC, Hillier SL, et al. Analysis of Factors Driving Incident and Ascending Infection and the Role of Serum Antibody in Chlamydia trachomatis Genital Tract Infection. *J Infect Dis.* 2016;213(4):523-31.
22. Russell AN, Zheng X, O'Connell CM, Wiesenfeld HC, Hillier SL, Taylor BD, et al. Identification of Chlamydia trachomatis Antigens Recognized by T Cells From Highly Exposed Women Who Limit or Resist Genital Tract Infection. *J Infect Dis.* 2016;214(12):1884-92.
23. Bakshi RK, Gupta K, Jordan SJ, Chi X, Lensing SY, Press CG, et al. An Adaptive Chlamydia trachomatis-Specific IFN-gamma-Producing CD4(+) T Cell Response Is Associated With Protection Against Chlamydia Reinfection in Women. *Front Immunol.* 2018;9:1981.
24. Zhong W, Kollipara A, Liu Y, Wang Y, O'Connell CM, Poston TB, et al. Genetic susceptibility loci for Chlamydia trachomatis endometrial infection influence expression of genes involved in T cell function, tryptophan metabolism and epithelial integrity. *Front Immunol.* 2022;13:1001255.
25. Darville T, Albritton HL, Zhong W, Dong L, O'Connell CM, Poston TB, et al. Anti-chlamydia IgG and IgA are insufficient to prevent endometrial chlamydia infection in women, and increased anti-chlamydia IgG is associated with enhanced risk for incident infection. *Am J Reprod Immunol.* 2019;81(5):e13103.
26. Liu C, Hufnagel K, O'Connell CM, Goonetilleke N, Mokashi N, Waterboer T, et al. Reduced Endometrial Ascension and Enhanced Reinfection Associated With Immunoglobulin G Antibodies to Specific Chlamydia trachomatis Proteins in Women at Risk for Chlamydia. *J Infect Dis.* 2022;225(5):846-55.
27. Mousset CM, Hobo W, Woestenenk R, Preijers F, Dolstra H, and van der Waart AB. Comprehensive Phenotyping of T Cells Using Flow Cytometry. *Cytometry Part A.* 2019;95(6):647-54.

833 28. Bonecchi R, Bianchi G, Bordignon PP, D'Ambrosio D, Lang R, Borsatti A, et al.
834 Differential expression of chemokine receptors and chemotactic responsiveness
835 of type 1 T helper cells (Th1s) and Th2s. *J Exp Med.* 1998;187(1):129-34.
836 29. Wacleche VS, Goulet JP, Gosselin A, Monteiro P, Soudeyns H, Fromentin R, et al.
837 New insights into the heterogeneity of Th17 subsets contributing to HIV-1
838 persistence during antiretroviral therapy. *Retrovirology.* 2016;13(1):59.
839 30. Singh SP, Zhang HH, Foley JF, Hedrick MN, and Farber JM. Human T cells that
840 are able to produce IL-17 express the chemokine receptor CCR6. *J Immunol.*
841 2008;180(1):214-21.
842 31. Takeshita M, Suzuki K, Kassai Y, Takiguchi M, Nakayama Y, Otomo Y, et al.
843 Polarization diversity of human CD4+ stem cell memory T cells. *Clin Immunol.*
844 2015;159(1):107-17.
845 32. Muranski P, and Restifo NP. Essentials of Th17 cell commitment and plasticity.
846 *Blood.* 2013;121(13):2402-14.
847 33. Basdeo SA, Cluxton D, Sulaimani J, Moran B, Canavan M, Orr C, et al. Ex-Th17
848 (Nonclassical Th1) Cells Are Functionally Distinct from Classical Th1 and Th17
849 Cells and Are Not Constrained by Regulatory T Cells. *J Immunol.*
850 2017;198(6):2249-59.
851 34. Khader SA, Gaffen SL, and Kolls JK. Th17 cells at the crossroads of innate and
852 adaptive immunity against infectious diseases at the mucosa. *Mucosal Immunol.*
853 2009;2(5):403-11.
854 35. Duhen T, and Campbell DJ. IL-1β promotes the differentiation of polyfunctional
855 human CCR6+CXCR3+ Th1/17 cells that are specific for pathogenic and
856 commensal microbes. *J Immunol.* 2014;193(1):120-9.
857 36. Acosta-Rodriguez EV, Rivino L, Geginat J, Jarrossay D, Gattorno M, Lanzavecchia
858 A, et al. Surface phenotype and antigenic specificity of human interleukin 17-
859 producing T helper memory cells. *Nature immunology.* 2007;8(6):639-46.
860 37. Menze BH, Kelm BM, Masuch R, Himmelreich U, Bachert P, Petrich W, et al. A
861 comparison of random forest and its Gini importance with standard chemometric
862 methods for the feature selection and classification of spectral data. *BMC*
863 *Bioinformatics.* 2009;10(1):213.
864 38. Rahman KS, Darville T, Russell AN, O'Connell CM, Wiesenfeld HC, Hillier SL, et
865 al. Comprehensive Molecular Serology of Human Chlamydia trachomatis
866 Infections by Peptide Enzyme-Linked Immunosorbent Assays. *mSphere.*
867 2018;3(4).
868 39. Johansson M, Schön K, Ward M, and Lycke N. Studies in knockout mice reveal
869 that anti-chlamydial protection requires TH1 cells producing IFN-gamma: is this
870 true for humans? *Scand J Immunol.* 1997;46(6):546-52.
871 40. De Simone G, Mazza EMC, Cassotta A, Davydov AN, Kuka M, Zanon V, et al.
872 CXCR3 Identifies Human Naive CD8(+) T Cells with Enhanced Effector
873 Differentiation Potential. *J Immunol.* 2019;203(12):3179-89.
874 41. Belay T, Eko FO, Ananaba GA, Bowers S, Moore T, Lyn D, et al. Chemokine and
875 chemokine receptor dynamics during genital chlamydial infection. *Infect Immun.*
876 2002;70(2):844-50.

42. Olive AJ, Gondek DC, and Starnbach MN. CXCR3 and CCR5 are both required for T cell-mediated protection against C. trachomatis infection in the murine genital mucosa. *Mucosal Immunol.* 2011;4(2):208-16.
43. Sampath P, Moideen K, Ranganathan UD, and Bethunaickan R. Monocyte Subsets: Phenotypes and Function in Tuberculosis Infection. *Front Immunol.* 2018;9:1726.
44. LaVerda D, Albanese LN, Ruther PE, Morrison SG, Morrison RP, Ault KA, et al. Seroreactivity to Chlamydia trachomatis Hsp10 correlates with severity of human genital tract disease. *Infect Immun.* 2000;68(1):303-9.
45. El Hakim EA, Gordon UD, and Akande VA. The relationship between serum Chlamydia antibody levels and severity of disease in infertile women with tubal damage. *Arch Gynecol Obstet.* 2010;281(4):727-33.
46. Pantelic M, Kim YJ, Bolland S, Chen I, Shively J, and Chen T. Neisseria gonorrhoeae kills carcinoembryonic antigen-related cellular adhesion molecule 1 (CD66a)-expressing human B cells and inhibits antibody production. *Infect Immun.* 2005;73(7):4171-9.
47. Lee C, Guinn BA, Brooks SE, Richardson D, and Orchard K. CD66a (CEACAM1) is the only CD66 variant expressed on the surface of plasma cells in multiple myeloma: a refined target for radiotherapy trials? *Br J Haematol.* 2010;149(5):795-6.
48. Fankhauser SC, and Starnbach MN. PD-L1 limits the mucosal CD8+ T cell response to Chlamydia trachomatis. *J Immunol.* 2014;192(3):1079-90.
49. Cockwell P, Calderwood JW, Brooks CJ, Chakravorty SJ, and Savage CO. Chemoattraction of T cells expressing CCR5, CXCR3 and CX3CR1 by proximal tubular epithelial cell chemokines. *Nephrol Dial Transplant.* 2002;17(5):734-44.
50. Poston TB, Lee DE, Darville T, Zhong W, Dong L, O'Connell CM, et al. Cervical Cytokines Associated With Chlamydia trachomatis Susceptibility and Protection. *J Infect Dis.* 2019;220(2):330-9.
51. Liu H, and Rohowsky-Kochan C. Regulation of IL-17 in human CCR6+ effector memory T cells. *J Immunol.* 2008;180(12):7948-57.
52. Ogongo P, Tezera LB, Ardain A, Nhamoyebonde S, Ramsuran D, Singh A, et al. Tissue-resident-like CD4+ T cells secreting IL-17 control Mycobacterium tuberculosis in the human lung. *J Clin Invest.* 2021;131(10).
53. Borkner L, Curham LM, Wilk MM, Moran B, and Mills KHG. IL-17 mediates protective immunity against nasal infection with Bordetella pertussis by mobilizing neutrophils, especially Siglec-F(+) neutrophils. *Mucosal Immunol.* 2021;14(5):1183-202.
54. Gérard HC, Stanich JA, Oszust CE, Whittum-Hudson JA, Carter JD, Schumacher HR, et al. Functional CCR5 receptor protects patients with arthritis from high synovial burden of infecting Chlamydia trachomatis. *Am J Med Sci.* 2010;340(6):448-51.
55. Ogendi BMO, Bakshi RK, Sabbaj S, Brown L, Lee JY, Kapil R, et al. Distinct peripheral vs mucosal T-cell phenotypes in chlamydia-infected women. *Am J Reprod Immunol.* 2017;78(6).

56. Ficarra M, Ibana JS, Poretta C, Ma L, Myers L, Taylor SN, et al. A distinct cellular profile is seen in the human endocervix during Chlamydia trachomatis infection. *Am J Reprod Immunol.* 2008;60(5):415-25.
57. Gattinoni L, Speiser DE, Lichterfeld M, and Bonini C. T memory stem cells in health and disease. *Nat Med.* 2017;23(1):18-27.
58. Singh SP, Parween F, Edara N, Zhang HH, Chen J, Otaizo-Carrasquero F, et al. Human CCR6+ Th Cells Show Both an Extended Stable Gradient of Th17 Activity and Imprinted Plasticity. *J Immunol.* 2023;210(11):1700-16.
59. Sun M, Phan JM, Kieswetter NS, Huang H, Yu KKQ, Smith MT, et al. Specific CD4(+) T cell phenotypes associate with bacterial control in people who 'resist' infection with Mycobacterium tuberculosis. *Nature immunology.* 2024;25(8):1411-21.
60. Li Y, Warren JA, Poston TB, Clutton G, Shaw FR, Conrad SZ, et al. Chlamydia trachomatis induces low-frequency, sustained CD4 T cell responses in most women, predominantly targeting chlamydial protease-like activity factor, CPAF. *J Infect Dis.* 2024.
61. Bunk S, Schaffert H, Schmid B, Goletz C, Zeller S, Borisova M, et al. Chlamydia pneumoniae-induced memory CD4+ T-cell activation in human peripheral blood correlates with distinct antibody response patterns. *Clin Vaccine Immunol.* 2010;17(5):705-12.
62. Safar HA, Mustafa AS, Amoudy HA, and El-Hashim A. The effect of adjuvants and delivery systems on Th1, Th2, Th17 and Treg cytokine responses in mice immunized with Mycobacterium tuberculosis-specific proteins. *PLoS One.* 2020;15(2):e0228381.
63. Nguyen N, Guleed S, Olsen AW, Follmann F, Christensen JP, and Dietrich J. Th1/Th17 T cell Tissue-Resident Immunity Increases Protection, But Is Not Required in a Vaccine Strategy Against Genital Infection With Chlamydia trachomatis. *Front Immunol.* 2021;12:790463.
64. Poston TB, Girardi J, Kim M, Zwarycz P, Polson AG, Yount K, et al. Intranasal immunization with CPAF combined with cyclic-di-AMP induces a memory CD4 T cell response and reduces bacterial burden following intravaginal infection with Chlamydia muridarum. *bioRxiv.* 2024.
65. Poston TB, Girardi J, Polson AG, Bhardwaj A, Yount KS, Jaras Salas I, et al. Viral-vectored boosting of OmcB- or CPAF-specific T-cell responses fail to enhance protection from Chlamydia muridarum in infection-immune mice and elicits a non-protective CD8-dominant response in naïve mice. *Mucosal Immunol.* 2024.
66. Bagwell CB, Hunsberger B, Hill B, Herbert D, Bray C, Selvanantham T, et al. Multi-site reproducibility of a human immunophenotyping assay in whole blood and peripheral blood mononuclear cells preparations using CyTOF technology coupled with Maxpar Pathsetter, an automated data analysis system. *Cytometry B Clin Cytom.* 2020;98(2):146-60.
67. Johnson WE, Li C, and Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics.* 2006;8(1):118-27.
68. Stanley N, Stelzer IA, Tsai AS, Fallahzadeh R, Ganio E, Becker M, et al. VoPo leverages cellular heterogeneity for predictive modeling of single-cell data. *Nature Communications.* 2020;11(1):3738.

967 69. Bruggner RV, Bodenmiller B, Dill DL, Tibshirani RJ, and Nolan GP. Automated
968 identification of stratifying signatures in cellular subpopulations. *Proceedings of the*
969 *National Academy of Sciences.* 2014;111(26):E2770-E7.
970