1 **HIV reservoir size is determined prior to ART initiation and linked to CD8 T cell**

2 **activation and memory expansion**

3 Genevieve E Martin[1], Matthew Pace[1], Freya Shearer[2], Eva Zilber[1], Jacob Hurst[1], Jodi

4 Meyerowitz[1], John P Thornhill[1,3], Julianne Lwanga[4], Helen Brown[1], Nicola Robinson[1],

5 Emily Hopkins[1], Natalia Olejniczak[1], Nneka Nwokolo[5†], Julie Fox[4,6†], Sarah Fidler[3,7†],

6 Christian B Willberg[1,8†], and John Frater[1,8†*] on behalf of the CHERUB investigators

7 [1] Peter Medawar Building for Pathogen Research, Nuffield Department of Medicine, University
8 of Oxford, Oxford, UK
9 [2] Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of
10 Oxford, Oxford, UK
11 [3] Division of Medicine, Wright Fleming Institute, Imperial College, London, UK
12 [4] Department of Genitourinary Medicine and Infectious Disease, Guys and St Thomas' NHS
13 Trust, London, UK
14 [5] Chelsea and Westminster Hospital, London, UK
15 [6] King's College NIHR Biomedical Research Centre, London, UK
16 [7] Imperial College NIHR Biomedical Research Centre, London, UK
17 [8] National Institute of Health Research Biomedical Research Centre, Oxford, UK
18
19 †These authors contributed equally

20 *Current affiliations:*
21 Genevieve E Martin: Alfred Hospital, Melbourne, Australia
22 Freya Shearer: Modelling and Simulation Unit, School of Population and Global Health,
23 University of Melbourne, Melbourne, Australia
24

25 ***Corresponding author:**

26 John Frater, Nuffield Department of Medicine, Peter Medawar Building for Pathogen

27 Research, John Radcliffe Hospital, Oxford, UK.

28 Email: john.frater@ndm.ox.ac.uk

29

30 **Conflict of interest statement:** The authors have declared that no conflict of interest

31 exists

1

32 **Abstract**

33 Initiation of antiretroviral therapy (ART) in early compared with chronic HIV infection is

34 associated with a smaller HIV reservoir. This longitudinal analysis of 63 individuals who

35 commenced ART during primary HIV infection (PHI) investigates which pre- and post-

36 therapy factors associate most closely with reservoir size (HIV DNA) following

37 treatment initiation during PHI. The best predictor of reservoir size at one-year was pre-

38 ART HIV DNA which was in turn significantly associated with CD8 memory

39 differentiation (effector memory, naïve and T-bet^{neg}Eomes^{neg} subsets), CD8 T cell

40 activation (CD38 expression) and PD-1 and Tim-3 expression on memory CD4 T cells.

41 No associations were found for any immunological variables following one-year of ART.

42 HIV reservoir size is determined around the time of ART initiation in individuals treated

43 during PHI. CD8 T cell activation and memory expansion are linked to HIV reservoir

44 size, suggesting the importance of the initial host-viral interplay in eventual reservoir

45 size.

Introduction

HIV persists despite long-term suppressive antiretroviral therapy (ART) in a reservoir of latently infected cells (1-3). Reducing the size of this reservoir is a focus of curative interventions as a smaller HIV reservoir is associated with delayed clinical progression and viral rebound following treatment interruption (4-6).

While there are only a few studies that explore the longitudinal formation of the HIV reservoir, there is some evidence that T cell immunity prior to the start of ART may be key. In the SPARTAC trial, both pre-therapy CD4 and CD8 T cell responses, as well as protective HLA class I alleles, were linked with lower levels of HIV DNA (7). Additionally, recent work from a different cohort treated during acute infection showed that HIV-specific T cell responses could reduce viral load (VL), and limit reservoir seeding, when antiretroviral therapy (ART) was initiated at peak viraemia (8).

T cell activation – a mediator of disease progression and non-AIDS co-morbidities (9-12) – has also been shown in cross-sectional studies to be related to HIV reservoir size (7, 13-15). Two recent longitudinal studies aimed to clarify this relationship. In individuals who initiated ART during chronic infection, no link between pre-ART inflammation/activation was observed with reservoir size over time (16). Contrasting with this, the finding that several soluble biomarkers in acute infection, including many linked to interferon-α (IFN-α) signalling, were identified as correlates of subsequent reservoir size (17) suggests a link between early immune activation and the formation of the reservoir.

3

67   Immune checkpoint receptor (ICR) expression has also been linked to the size of the HIV

68   reservoir (7, 13, 14, 18, 19). A limitation of many of these studies is that ICR expression

69   has often been measured on bulk T cells. The HIV reservoir is preferentially found in

70   memory subsets (19-22) which also express higher levels of ICRs (18, 23-26). As such,

71   ICR expression on bulk T cell subsets could be a surrogate for altered memory

72   distribution and a potential confounder (27). Furthermore, many of these relationships

73   were demonstrated for a single marker, even though the expression of many ICRs is

74   closely related to one another – and also to T cell activation (23-26, 28-31).

75   The 'HIV Reservoir Targeting with Early Antiretroviral Therapy' (HEATHER) cohort is

76   a prospective study of individuals commencing ART during primary HIV infection

77   (PHI), and provides the opportunity to study reservoir dynamics over time in well-

78   characterised participants. Here, we define the relationship of several key immunological

79   and clinical parameters with the size of the HIV reservoir and show that the key events in

80   setting the reservoir occur early, prior to starting ART.

81

82   **Results**

83   *Baseline clinical characteristics*

84   We studied 63 individuals enrolled in HEATHER, a longitudinal observational cohort of

85   more than 300 participants who commenced ART during PHI. Individuals were selected

86   based on availability of samples and access to clinical records at the time of the study.

87   Clinical and demographic details are listed in Table 1. All were male and had a median

88   age of 36 (IQR 28-41) years at the time of ART start. They commenced ART a median of

89   29 (IQR 14-46) days following a confirmed HIV diagnosis and a median of 49 (IQR 33-

90   90) days following estimated seroconversion. Different methods for diagnosing PHI

91   (Table 1) were used; 26 participants (41%) were P24 antigen positive without detectable

92   antibodies, consistent with Fiebig stages I or II at the time of diagnosis.

93   The participants had a high median baseline VL (5.5 $\log_{10}$ copies/ml; IQR 4.6-6.5), which

94   declined on ART (Fig. 1A). 60/63 (95%) were virologically suppressed (<50 copies/ml)

95   after one year of ART (the frequency of viral load sampling and time to suppression is

96   shown in Supplementary Fig. 1). The three individuals who had not achieved VL<50

97   copies/ml by one year had estimated VLs (average of those either side of the 1 year time-

98   point) of 40, 76 and 190 copies/ml, and achieved undetectability shortly after. The

99   dynamics of CD4 and CD8 count, as well as CD4 to CD8 ratio following ART initiation

100   are shown in Fig. 1B-D.

101   There was a relationship between the first measured (or 'baseline') pVL and the method

102   used to diagnose PHI (Fig. 1E). Baseline VL was higher when measured closer to

103   estimated seroconversion (r_s=-0.58, p=5.0x10^{-7}), suggesting that viral load is of limited

104   utility as a predictive variable in PHI compared with chronic HIV infection, as a stable

105   'set point' has not yet been reached.

106

107   *HIV reservoir size after one year of ART is predicted by pre-ART HIV DNA levels*

5

108   Quantitation of total HIV DNA (copies per 10^6 CD4 T cells) is used here as the surrogate

109   measure of reservoir size on ART. Other measures (e.g. integrated HIV DNA, cell-

110   associated RNA, quantitative viral outgrowth) were not used for this analysis due to the

111   high numbers of predictive variables being tested. Of note, the term "reservoir" is only

112   used for HIV DNA measurements on suppressive ART as pre-ART DNA measurements

113   will capture cells with active viral replication. Compared with pre-ART levels, HIV DNA

114   decreased a mean of 0.9 $\log_{10}$ copies following one year of therapy (Fig. 2A; p=2.2x10$^-$

115   ^{16}). HIV DNA levels pre-therapy and following one year of ART were highly correlated

116   (Fig. 2B; r=0.76, p=1.2x10^{-12}). For a subset of 18 individuals, levels of total HIV DNA

117   were also available three years post-ART initiation, and had declined a further 0.3 $\log_{10}$

118   copies since year one. (HIV DNA levels were also correlated between those two

119   measurements (Supplementary Fig. 2; r-0.57, p=0.013)).

120

121   *Immunological and clinical variables associated with HIV DNA are highly correlated*

122   We next determined which clinical and immunological variables were most closely

123   associated with the size of the HIV reservoir. Clinical variables included CD4 count,

124   CD8 count, VL, CD4:CD8 ratio, time to ART start, and time to VL suppression on ART.

125   Immunological measures included flow cytometric quantitation (Fig. 3A) of CD4 and

126   CD8 memory subsets, CD38 expression, ICR expression (PD-1, TIGIT, Tim-3 on

127   memory CD4 and CD8 T cells), the proportion of T-bet/Eomes expressing CD8

128   populations (32) and soluble plasma ICRs (sPD-1 and sTim-3).

6

129   Several parameters were highly correlated with HIV DNA. Corrgrams were used as a

130   screen to explore the relationship of variables measured prior to ART initiation (baseline;

131   Fig. 3B) and after one year of ART (Fig. 3C), with the HIV reservoir at one year. Each

132   row or column in the corrgram represents a different variable ordered by the strength of

133   the correlation with reservoir size at one year (in the top left corner). A circle indicates a

134   correlation between two variables (p<0.05), the size and intensity corresponding to the

135   magnitude of the correlation coefficient. Variables with a statistically significant

136   relationship with reservoir size at 1 year are marked by a green square.

137   Two important observations can be made. Firstly, the corrgrams for variables at baseline

138   and one year look very different (Figs. 3B, C). When exploring variables measured

139   immediately prior to ART initiation, we can see that those variables which were most

140   closely related to reservoir size (top left hand corner) were also highly correlated with

141   each other. Not surprisingly, the variable with the strongest correlation with HIV

142   reservoir size at one year was the level of HIV DNA measured at baseline. However, 26

143   other variables are associated with the reservoir and/or each other (green box).

144   When using variables measured after one year of ART – at the same time as the HIV

145   reservoir was measured (Fig. 3C) – there is little evidence of any correlation with

146   reservoir size or between variables. These data suggest that certain clinical and

147   immunological variables are the key determinants of the HIV DNA level just prior to

148   ART initiation, which is, in turn, the main predictor of HIV reservoir size on ART.

149

150   *CD8 T cell activation and memory expansion are the key determinants of HIV DNA levels*

7

151 To more formally assess the strength and independence of these relationships, regression

152 models were fitted to the data. Correlative analyses reinforced the role for baseline total

153 HIV DNA as the key predictor of reservoir size at one year. For this reason, we

154 concentrated on models to assess which variables were most highly related with baseline

155 total HIV DNA (Fig. 4A).

156 The dataset poses several challenges in the fitting of multivariable regression models,

157 especially as 8.4% of observations are missing (Supplementary Fig. 3) due to unavailable

158 samples, poor sample viability or low cell count. The large number of parameters

159 measured relative to observations, as well as the strong correlations between many of

160 these variables, was also problematic. To ensure the robustness of any conclusions, two

161 different models (boosted regression tree (BRT) and least absolute shrinkage and

162 selection operator (LASSO)) with different approaches to complex data were fitted and

163 their outputs compared.

164 BRT is a machine learning approach that builds a series of regression trees, with each

165 subsequent tree iteratively aiming to improve the previous fit by focusing on data poorly

166 modelled by the existing set of trees. This approach is able to handle missing data, does

167 not make prior assumptions about the effect of potential predictor variables and can

168 handle high-dimensional interactions (33). A BRT model was fitted with baseline total

169 HIV DNA as the outcome, and all other baseline variables as predictors. The relative

170 importance of each predictor variable included in this BRT model is shown in Fig. 4B.

171 The relative influence of each variable is estimated based on the number of times that

172 variable is selected for splitting and the improvement to the model as a result of that split

173 averaged across all trees; a higher number indicates a greater effect of the variable. We

8

174   defined influential predictors as those with a relative influence value greater than 100

175   divided by the total number of variables (indicated by the dashed line). The figure shows

176   that 10 of the 62 predictors had a consistent influence in predicting baseline reservoir

177   size. Notably, CD8 memory subsets (the proportion of EM and naïve cells), as well as

178   CD8 CD38 expression were the variables with the highest relative influence.

179   LASSO is a multivariable regression designed to cope with multi-collinearity and large

180   numbers of predictors by adding a penalty to the coefficient of each term (with the ability

181   to penalise coefficients to zero) thus performing variable selection. Missing values were

182   imputed using a random forest based method. Linear LASSO models were fitted to the

183   data (Fig. 4C) and selected six variables which were independently predictive of baseline

184   HIV DNA. All six were also selected by the BRT model. The variables with greatest

185   influence on baseline HIV DNA were associated with CD8 memory expansion (the

186   proportion of naïve and EM, as well as T-bet^{neg}Eomes^{neg} CD8 T cells) and CD4 memory

187   ICR expression (PD-1 on TM and Tim-3 on CM CD4 T cells), as well as CD38

188   expression on CD8 T cells.

189   A sensitivity analysis was conducted by constructing the same model using only

190   observations that were complete (no imputation). The overall results of this model were

191   the same (although due to decreased power fewer variables were selected) and are shown

192   in Supplementary Table 1.

193

194   *HIV DNA pre-ART is the dominant predictor of reservoir size at one year on ART*

9

195   Having established which variables were related to baseline HIV DNA, regression

196   models were then fitted to explore if any of clinical or immunological variables had

197   additional, independent relationships with reservoir size at one year. Total HIV DNA at

198   baseline was the most influential variable when combined with other pre-ART or one

199   year variables (Table 2; Supplementary Fig. 4 shows the corresponding BRT models and

200   Supplementary Table 1 the LASSO models with no imputation).

201   No immunological variables measured at one year impacted reservoir size independently

202   of the baseline HIV DNA levels, consistent with the modest correlations observed in the

203   Fig. 3C corrgram (Model B in Table 2, Supplementary Fig. 4B). Of the clinical variables

204   and immunological variables measured at baseline, only 'time from ART start to VL

205   suppression' influenced reservoir size independently of the baseline HIV DNA (Model A

206   in Table 2, Supplementary Fig. 4A).

207

208   *Reservoir size is related to HLA class I*

209   Although we did not have access to HIV-specific immune responses for HEATHER,

210   participants were typed for HLA class I. These alleles have the strongest consistent

211   relationship with viral control during HIV infection, and HLA type can be considered as

212   a surrogate marker of potential CD8-driven HIV-specific immunity (34, 35). Figure 5

213   shows the relationship between HLA class I alleles and total HIV DNA at one year post-

214   ART initiation. Alleles associated with viral control (red) can be seen clustering towards

215   the left hand side of the plot associated with low HIV DNA levels. The converse is seen

216   for alleles associated with progression (blue). The same relationship was observed with

217   baseline HIV DNA (data not shown) and supports previous findings in a different PHI

218   cohort (SPARTAC) (7). Together, these data are consistent with HIV specific immunity,

219   the general immune landscape and clinical parameters all contributing to the size of the

220   HIV reservoir on ART.

221

222   Figure 6 summarises our findings. The only two independent variables that predicted

223   HIV reservoir size after one year of ART were total HIV DNA at pre-therapy baseline

224   and the time taken to achieve viral load suppression after starting therapy. Baseline HIV

225   DNA was associated with HLA class I type and specific markers of T cell activation,

226   expansion and exhaustion.

227

228   **Discussion**

229   This work demonstrates the importance of immunological events before ART initiation in

230   determining the subsequent reservoir size. The level of HIV DNA prior to ART was the

231   most important predictor of reservoir size a year later, suggesting that reservoir size is

232   "set" early on. Of note, the level of HIV DNA pre-ART was more closely related to CD8

233   T cell activation, memory expansion and CD4 ICR expression than any clinical

234   parameters, including pre-ART VL. Several of the associations with reservoir size

235   presented here have been observed previously – but generally in isolation. This is the first

236   study to define the independence of relationships between T cell activation, memory

237   expansion and ICR expression with eventual reservoir size.

11

238   One of the key findings of this analysis is that the HIV DNA level prior to ART initiation

239   was the most influential predictor of subsequent reservoir size. Several other studies have

240   shown a relationship between pre-therapy HIV DNA levels and those once virologically

241   suppressed on ART (16, 36, 37). These findings also support previous research

242   suggesting that set point HIV DNA levels in untreated individuals are reached shortly

243   after peak VL (38), and extend it to show that this is true even if ART is started early.

244   Higher levels of initial T cell activation were associated with increased reservoir size.

245   This is consistent with the SPARTAC study where CD38 expression at baseline was

246   significantly associated with total HIV DNA at that time (7). It is possible that this is

247   driven by higher initial viral burden; it could also reflect poorer CD8 effector function

248   suggesting that activation and dysfunction may already be coupled during PHI. A recent

249   study of soluble biomarkers during acute infection demonstrated that several of these

250   were related to HIV reservoir size (also measured by total HIV DNA) prior to and

251   following 96 weeks of ART, independently of VL (17). The soluble biomarkers identified

252   can all be produced by myeloid cells in response to IFN-α/γ signalling; the authors

253   speculate that this could indirectly reflect innate and/or T cell responses to viral

254   replication (17). The analysis presented here supports these findings, and directly

255   demonstrates a link between T cell activation and reservoir size.

256   Cross-sectional studies of primary and chronic HIV infection have shown relationships

257   between contemporaneous HIV DNA and CD8 T cell activation during ART (13, 15, 39),

258   although this has not been consistently demonstrated (37, 40). Such a link between CD8

259   activation following ART and HIV DNA levels was not observed here. Of note, in this

260   work CD38 expression was used to define activation and most prior studies have used

12

261   HLA-DR/CD38 co-expression. Work from Cockerham *et al.* suggests that HIV DNA

262   may only relate to HLA-DR and not CD38 expression (15) and could explain this

263   discrepancy.

264   Gandhi *et al.* observed a relationship between T cell activation during chronic infection

265   and HIV DNA in individuals commencing treatment at this time (16). Unlike in early

266   infection studies ((17) and this analysis), this relationship did not persist following

267   adjustment for baseline VL - suggesting time-dependent differences in the relative

268   influence of baseline VL and T cell activation as predictors.

269   It is interesting to note that the VL was only modestly correlated with baseline HIV

270   DNA, as we had hypothesised that these would be closely linked. Indeed, VL was not a

271   significant predictor here. Several studies have demonstrated a relationship between pre-

272   therapy VL and HIV DNA (16, 37, 41-44). Notably, many of these studies included

273   individuals treated during chronic infection, where the VL will have reached set point. In

274   contrast, during PHI the viral load is substantially more labile. Within this cohort there

275   was a link between baseline VL and the time after estimated seroconversion that this was

276   measured (Fig. 1E), whereby individuals with more recent seroconversion had a higher

277   VL. This has also been observed in another PHI cohort with similar time since infection

278   (45) and suggests that VL measures taken over this time capture the decline from peak,

279   rather than a steady state. Within HEATHER there is some variation in the time interval

280   between when baseline VL and HIV DNA were measured (median 16 days; range 0-82

281   days). Both of these factors may contribute to the relatively weak relationship between

282   VL and baseline HIV DNA observed here.

13

283   It is also possible that the overall pre-ART viral burden (both duration and magnitude of

284   viraemia) influences HIV reservoir size, and is poorly captured by a single measurement

285   during PHI. Indeed, the consistent observation that earlier ART limits reservoir size, even

286   over the short time periods between the first three Fiebig stages (8, 36) implies a role for

287   the total pre-therapy viral burden. Two findings here suggest an influence of viral burden

288   on overall reservoir size. The first is the previously reported relationship between HLA

289   class I alleles and HIV reservoir size (7), which was confirmed. The well-characterised

290   relationship between HLA class I alleles and VL is mediated via CD8 T cell killing of

291   virally infected cells; stronger CD8 responses could result in smaller HIV reservoir size

292   by both limiting seeding (via VL reduction) and contributing to decay of the pool of

293   infected cells. The second relevant finding is the observation that time to viral load

294   suppression had an influence on reservoir size independently of baseline HIV DNA level.

295   This may be because individuals with longer time to VL suppression may have a window

296   following ART initiation for reservoir seeding to continue. Alternatively, slower time to

297   viral suppression may reflect higher pre-therapy viral burden not captured in the baseline

298   VL measurement. Whilst most individuals commenced ART at or shortly after their

299   baseline visit (median 0 days, 82% within 1 week), a small proportion had a larger

300   interval between these (maximum 48 days) providing additional time for reservoir

301   seeding not captured in the baseline HIV DNA measurement.

302   It is notable that the two measures of ICR expression which related to reservoir size were

303   measured on TM and CM CD4 T cells, the subsets most enriched for proviral DNA (19-

304   22). Several cross-sectional studies have shown a relationship between PD-1 expression

305   on bulk CD4 T cells during ART and overall reservoir size (13, 14, 18). CM and TM

306   CD4 T cells expressing PD-1 are also enriched for proviral DNA (18, 19). The findings

307   presented here demonstrate a relationship between PD-1 expression on CM CD4 T cells

308   supporting these prior observations and showing that this relationship exists during

309   untreated PHI. Contrasting with prior work, here this relationship did not continue during

310   ART. It is possible that previous studies, which have measured PD-1 on bulk T cells (13,

311   14, 18), are actually capturing this shift in memory proportions. This discrepancy could

312   also be due to differences between individuals who commence ART in primary,

313   compared with chronic, infection.

314   Whilst a comprehensive analysis of ICR expression, activation and T cell differentiation

315   was performed, one limitation of this work is that reservoir size was only measured using

316   total HIV DNA. This measure is clinically relevant as lower levels have been associated

317   with delayed viral rebound following treatment interruption (4). Much of the HIV

318   reservoir, however, is not replication competent (46, 47). Whilst total HIV DNA has been

319   suggested to reflect replication-competent reservoir size in cohort studies (48), this work

320   has not assessed if there is any impact of these immunological measures on the quality of

321   proviruses comprising the reservoir. A major strength of this study is its longitudinal

322   design. It is possible, however, that 1 year of follow up is not sufficient time for ART to

323   stabilise the HIV reservoir and that these results may not reflect the eventual long-term

324   reservoir. For individuals for whom samples were available, we have shown that HIV

325   DNA levels one year following ART initiation correlated with those at three years,

326   suggesting the validity of this approach. Several other studies have shown relationships

327   between HIV DNA levels at early time points and those much later after ART initiation

328   (16, 37, 49), further supporting the use of this time point here.

15

329   This work has shown that the magnitude of the early immunological insult, reflected in

330   CD8 T cell activation and memory expansion drives HIV reservoir size. These results

331   suggest that targeting of host or viral factors which lead to early viral expansion and T

332   cell activation may be a way of limiting HIV reservoir size, and confirm the importance

333   of starting ART as early as possible.

334

335   **Methods**

336   *Participant information*

337   HEATHER is a prospective observational cohort study of individuals who commence

338   ART (and remain on uninterrupted therapy) within 3 months of the date of HIV diagnosis

339   during PHI. Individuals are considered to have PHI if they meet any of the following

340   criteria: HIV-1 positive antibody test within 6 months of a HIV-1 negative antibody test,

341   HIV-1 antibody negative with positive PCR (or positive P24 Ag or viral load detectable),

342   RITA (recent incident assay test algorithm) assay result consistent with recent infection,

343   equivocal HIV-1 antibody test supported by a repeat test within 2 weeks showing a rising

344   optical density or having clinical manifestations of symptomatic HIV seroconversion

345   illness supported by antigen positivity. The time of seroconversion was estimated as the

346   midpoint between the most recent negative or equivocal test and the first positive test for

347   those who met relevant criteria, the date of RITA test minus 120 days and as the date of

348   test for all other participants. Individuals with co-existent hepatitis B or C infection are

349   not eligible for inclusion in HEATHER. Individuals for inclusion in this analysis were

350   selected at random and based on sample availability.

16

351   Cryopreserved peripheral blood mononuclear cells (PBMCs) were used from the closest

352   sample to seroconversion (baseline) and from a sample 9–15 months after

353   commencement of ART (1 year). CD4 count, CD8 count and VL were measured as part

354   of routine clinical care with baseline CD4 and CD8 counts defined as the earliest value

355   prior to the initiation of ART. Similarly, the baseline VL was taken as the earliest value

356   prior to the initiation of ART. In both cases, usually only one value was available.

357   *Flow cytometry*

358   Cryopreserved PBMCs were thawed in RPMI-1640 medium supplemented with 10%

359   FBS, L-glutamine, penicillin and streptomycin (R10) containing 2.7 Kunitz units/ml of

360   DNAse (Qiagen). Cells were stained in BD Horizon Brilliant Stain Buffer (BD)

361   containing all antibodies and Live/Dead Near IR at 1 in 300 dilution (Life Technologies)

362   at 4°C for 30 minutes.

363   Panel 1 - PBMCs were stained with the following antibodies: CD3 Brilliant Violet (BV)

364   570 (UCHT1), CCR7 Pacific Blue (G043H7), CD27 AlexaFluor700 (M-T271)[all

365   BioLegend], CD4 BV605 (RPA-T4), CD8 BV650 (RPA-T8)[all BD], PD-1 PE-

366   eFluor610 (eBioJ105), CD45RA FITC (HI100), TIGIT PerCP-eFluor710 (MBSA43)[all

367   eBioscience] and Tim-3 PE (344823)[R&D].

368   Panel 2 - PBMCs were stained with the following antibodies: CD3 BV570, CD38

369   AlexaFluor700 (HB-7) [BioLegend], CD4 BV605, CD8 BV650, PD-1 PE-eFluor610 and

370   Tim-3 PE. Following this, cells were washed twice prior to fixation and permeabilisation

371   with Foxp3 Buffer Set (BD) as per manufacturer's directions with reduced volumes to

372   facilitate staining in 96-well plates. Staining for intracellular epitopes was performed at

373   room temperature for 30 minutes in PBS containing 0.5% bovine serum albumin (BSA)

374   and 0.1% sodium azide with the following antibodies: T-bet FITC (4B10)[BioLegend]

375   and Eomes eFluor660 (WD1928)[eBioscience].

376   For both panels, cells were washed twice, fixed with 2% formaldehyde for 30 minutes,

377   then washed twice and resuspended in phosphate buffered saline (PBS) for acquisition.

378   All samples were acquired on a LSR II (BD) the day after staining. The same machine

379   was used for all experiments with daily calibration with Cytometer Setup & Tracking

380   beads (BD) to maximise comparability between days. Rainbow Calibration Particles

381   (BioLegend) were also used for cohort phenotyping to minimise batch-to-batch

382   variability. Data were analysed using FlowJo Version 10.8.0r1 (Treestar).

383   *Soluble PD-1 and Tim-3 quantification*

384   The concentration of sPD-1 and sTim-3 was measured in thawed plasma samples by

385   enzyme linked immunosorbent assay (ELISA) using Human PD-1 (PDCD1) ELISA kit

386   [EHPDCD1] (Thermo Fisher Scientific, Waltham, MA USA) and Quantikine ELISA

387   Human TIM-3 Immunoassay kit [DTIM30] (R&D Systems, Minneapolis, MN USA) as

388   per manufacturers' instructions at 1:2 and 1:5 dilutions respectively.

389   *Total HIV DNA quantification*

390   HIV DNA was quantified relative to cell number using qPCR as previously described

391   (50). In brief, cryopreserved PBMCs were thawed (as above) and CD4 T cells were

392   isolated by negative selection using the EasySep Human CD4 Enrichment Kit (Stemcell

393   Technologies) before DNA extraction with the QiaAMP Blood Mini Kit (Qiagen). Cell

394 copy number was initially quantified using an albumin qPCR. 25,000 cell equivalents

395 (and no less than 10,000 cell equivalents) of DNA was then used in a total HIV DNA

396 qPCR, performed in triplicate. The mean number of copies of DNA was normalised to

397 cell number and expressed as copies/10^6 CD4 T cells.

398 *HLA-typing*

399 HLA typing was performed to intermediate resolution using PCR with sequence specific

400 primers (PCR-SSP).

401 *Statistics*

402 Analyses were performed using R (v3.2.2 or v3.4.3) and GraphPad Prism (v7.0b). Except

403 where otherwise specified, p-values <0.05 were considered statistically significant.

404 Simple comparisons were performed using parametric or non-parametric tests as

405 appropriate and are described alongside the results.

406 Corrgrams were generated using the package corrplot (v0.84). Where some data were

407 missing, pairwise complete observations were used to calculate the correlation

408 coefficients.

409 Boosted regression tree models were fitted with a Gaussian outcome using the package

410 gbm3 (v2.2). Models were fitted with an interaction depth of 5 with a minimum of 5

411 observations in a terminal node. The shrinkage parameter was adjusted between 0.0001 -

412 0.001 to aim for optimal number of trees to fall in the range of 3000-10000. The optimal

413 number of trees was determined using 10 fold cross-validation, with the number of trees

19

414 that minimised cross-validation error chosen. Results presented are summarised outcomes

415 of 100 models.

416 LASSO models (51) were fitted using the R package glmnet (v2.0-16) (52). Gaussian

417 regression models were fitted with an additive linear model (no interactions). The tuning

418 parameter λ was determined using 10-fold cross-validation, with the λ value used being

419 that which minimised cross-validation error plus one standard error. Where data were

420 imputed this was performed using the package MissForest (53), which employs a random

421 forest based method for multiple imputation and has been shown to be superior to other

422 multiple imputation methods in biological datasets with comparable levels of missingness

423 (53, 54).

424 *Study approval*

425 Recruitment for the HEATHER cohort was approved by the West Midlands-South

426 Birmingham Research Ethics Committee (reference 14/WM/1104). All participants have

427 given informed consent for their participation in these studies.

Author contributions

GEM, JFr devised the study, analysed data and wrote the manuscript; FS and JH advised

on the analysis of data; GEM, MP, EZ, JT, HB, NR, EH and NO performed laboratory

analyses; JM managed the patient cohorts and collated demographic data; JT, JL, NN,

JFo and SF managed clinical sites and recruitment; NN, JFo, SF, CBW, JFr led on study

design and management.

Acknowledgments

Funding:

449 **References**

450 1.   Finzi D, Hermankova M, Pierson T, Carruth LM, Buck C, Chaisson RE, et al.

451   Identification of a reservoir for HIV-1 in patients on highly active antiretroviral

452   therapy. *Science.* 1997;278(5341):1295-300.

453 2.   Chun TW, Carruth L, Finzi D, Shen X, DiGiuseppe JA, Taylor H, et al.

454   Quantification of latent tissue reservoirs and total body viral load in HIV-1

455   infection. *Nature.* 1997;387(6629):183-8.

456 3.   Wong JK, Hezareh M, Gunthard HF, Havlir DV, Ignacio CC, Spina CA, et al.

457   Recovery of replication-competent HIV despite prolonged suppression of plasma

458   viremia. *Science.* 1997;278(5341):1291-5.

459 4.   Williams JP, Hurst J, Stohr W, Robinson N, Brown H, Fisher M, et al. HIV-1

460   DNA predicts disease progression and post-treatment virological control. *Elife.*

461   2014;3:e03821.

462 5.   Li JZ, Etemad B, Ahmed H, Aga E, Bosch RJ, Mellors JW, et al. The size of the

463   expressed HIV reservoir predicts timing of viral rebound after treatment

464   interruption. *AIDS.* 2016;30(3):343-53.

465 6.   Sneller MC, Justement JS, Gittens KR, Petrone ME, Clarridge KE, Proschan MA,

466   et al. A randomized controlled safety/efficacy trial of therapeutic vaccination in

467   HIV-infected individuals who initiated antiretroviral therapy early in infection.

468   *Sci Transl Med.* 2017;9(419).

469 7.   Hurst J, Hoffmann M, Pace M, Williams JP, Thornhill J, Hamlyn E, et al.

470   Immunological biomarkers predict HIV-1 viral rebound after treatment

471   interruption. *Nat Commun.* 2015;6:8495.

472   8.   Takata H, Buranapraditkun S, Kessing C, Fletcher JL, Muir R, Tardif V, et al.

473   Delayed differentiation of potent effector CD8(+) T cells reducing viremia and

474   reservoir seeding in acute HIV infection. *Sci Transl Med.* 2017;9(377).

475   9.   Sousa AE, Carneiro J, Meier-Schellersheim M, Grossman Z, and Victorino RM.

476   CD4 T cell depletion is linked directly to immune activation in the pathogenesis

477   of HIV-1 and HIV-2 but only indirectly to the viral load. *J Immunol.*

478   2002;169(6):3400-6.

479   10.   Giorgi JV, Hultin LE, McKeating JA, Johnson TD, Owens B, Jacobson LP, et al.

480   Shorter survival in advanced human immunodeficiency virus type 1 infection is

481   more closely associated with T lymphocyte activation than with plasma virus

482   burden or virus chemokine coreceptor usage. *J Infect Dis.* 1999;179(4):859-70.

483   11.   Deeks SG, Kitchen CM, Liu L, Guo H, Gascon R, Narvaez AB, et al. Immune

484   activation set point during early HIV infection predicts subsequent CD4+ T-cell

485   changes independent of viral load. *Blood.* 2004;104(4):942-7.

486   12.   Utay NS, and Hunt PW. Role of immune activation in progression to AIDS. *Curr

487   Opin HIV AIDS.* 2016;11(2):131-7.

488   13.   Khoury G, Fromentin R, Solomon A, Hartogensis W, Killian M, Hoh R, et al.

489   Human Immunodeficiency Virus Persistence and T-Cell Activation in Blood,

490   Rectal, and Lymph Node Tissue in Human Immunodeficiency Virus-Infected

491   Individuals Receiving Suppressive Antiretroviral Therapy. *J Infect Dis.*

492   2017;215(6):911-9.

493   14.   Hatano H, Jain V, Hunt PW, Lee TH, Sinclair E, Do TD, et al. Cell-based

494   measures of viral persistence are associated with immune activation and

23

495   programmed cell death protein 1 (PD-1)-expressing CD4+ T cells. *J Infect Dis.*

496   2013;208(1):50-6.

497   15.   Cockerham LR, Siliciano JD, Sinclair E, O'Doherty U, Palmer S, Yukl SA, et al.

498   CD4+ and CD8+ T cell activation are associated with HIV DNA in resting CD4+

499   T cells. *PLoS One.* 2014;9(10):e110731.

500   16.   Gandhi RT, McMahon DK, Bosch RJ, Lalama CM, Cyktor JC, Macatangay BJ, et

501   al. Levels of HIV-1 persistence on antiretroviral therapy are not associated with

502   markers of inflammation or activation. *PLoS Pathog.* 2017;13(4):e1006285.

503   17.   Teigler JE, Leyre L, Chomont N, Slike B, Jian N, Eller MA, et al. Distinct

504   biomarker signatures in HIV acute infection associate with viral dynamics and

505   reservoir size. *JCI Insight.* 2018;3(10).

506   18.   Fromentin R, Bakeman W, Lawani MB, Khoury G, Hartogensis W, DaFonseca S,

507   et al. CD4+ T Cells Expressing PD-1, TIGIT and LAG-3 Contribute to HIV

508   Persistence during ART. *PLoS Pathog.* 2016;12(7):e1005761.

509   19.   Chomont N, El-Far M, Ancuta P, Trautmann L, Procopio FA, Yassine-Diab B, et

510   al. HIV reservoir size and persistence are driven by T cell survival and

511   homeostatic proliferation. *Nat Med.* 2009;15(8):893-900.

512   20.   Yukl SA, Shergill AK, Ho T, Killian M, Girling V, Epling L, et al. The

513   distribution of HIV DNA and RNA in cell subsets differs in gut and blood of

514   HIV-positive patients on ART: implications for viral persistence. *J Infect Dis.*

515   2013;208(8):1212-20.

516   21.   Josefsson L, von Stockenstrom S, Faria NR, Sinclair E, Bacchetti P, Killian M, et

517   al. The HIV-1 reservoir in eight patients on long-term suppressive antiretroviral

518   therapy is stable with few genetic changes over time. *Proc Natl Acad Sci U S A.*

519   2013;110(51):E4987-96.

520   22.   Soriano-Sarabia N, Bateson RE, Dahl NP, Crooks AM, Kuruc JD, Margolis DM,

521   et al. Quantitation of replication-competent HIV-1 in populations of resting CD4+

522   T cells. *J Virol.* 2014;88(24):14070-7.

523   23.   Hoffmann M, Pantazis N, Martin GE, Hickling S, Hurst J, Meyerowitz J, et al.

524   Exhaustion of Activated CD8 T Cells Predicts Disease Progression in Primary

525   HIV-1 Infection. *PLoS Pathog.* 2016;12(7):e1005661.

526   24.   Cockerham LR, Jain V, Sinclair E, Glidden DV, Hartogenesis W, Hatano H, et al.

527   Programmed death-1 expression on CD4(+) and CD8(+) T cells in treated and

528   untreated HIV disease. *AIDS.* 2014;28(12):1749-58.

529   25.   Tauriainen J, Scharf L, Frederiksen J, Naji A, Ljunggren HG, Sonnerborg A, et al.

530   Perturbed CD8(+) T cell TIGIT/CD226/PVR axis despite early initiation of

531   antiretroviral treatment in HIV infected individuals. *Sci Rep.* 2017;7:40354.

532   26.   Yamamoto T, Price DA, Casazza JP, Ferrari G, Nason M, Chattopadhyay PK, et

533   al. Surface expression patterns of negative regulatory molecules identify

534   determinants of virus-specific CD8+ T-cell exhaustion in HIV infection. *Blood.*

535   2011;117(18):4805-15.

536   27.   Olesen R, Leth S, Nymann R, Ostergaard L, Sogaard OS, Denton PW, et al.

537   Immune checkpoints and the HIV-1 reservoir: proceed with caution. *J Virus Erad.*

538   2016;2(3):183-6.

539   28.   Jones RB, Ndhlovu LC, Barbour JD, Sheth PM, Jha AR, Long BR, et al. Tim-3

540   expression defines a novel population of dysfunctional T cells with highly

541   elevated frequencies in progressive HIV-1 infection. *J Exp Med.*

542   2008;205(12):2763-79.

543 29. Vollbrecht T, Brackmann H, Henrich N, Roeling J, Seybold U, Bogner JR, et al.

544   Impact of changes in antigen level on CD38/PD-1 co-expression on HIV-specific

545   CD8 T cells in chronic, untreated HIV-1 infection. *J Med Virol.* 2010;82(3):358-

546   70.

547 30. Bengsch B, Ohtani T, Khan O, Setty M, Manne S, O'Brien S, et al. Epigenomic-

548   Guided Mass Cytometry Profiling Reveals Disease-Specific Features of

549   Exhausted CD8 T Cells. *Immunity.* 2018;48(5):1029-45 e5.

550 31. Sauce D, Almeida JR, Larsen M, Haro L, Autran B, Freeman GJ, et al. PD-1

551   expression on human CD8 T cells depends on both state of differentiation and

552   activation status. *AIDS.* 2007;21(15):2005-13.

553 32. Buggert M, Tauriainen J, Yamamoto T, Frederiksen J, Ivarsson MA, Michaelsson

554   J, et al. T-bet and Eomes are differentially linked to the exhausted phenotype of

555   CD8+ T cells in HIV infection. *PLoS Pathog.* 2014;10(7):e1004251.

556 33. Elith J, Leathwick JR, and Hastie T. A working guide to boosted regression trees.

557   *J Anim Ecol.* 2008;77(4):802-13.

558 34. Pereyra F, Jia X, McLaren PJ, Telenti A, de Bakker PI, Walker BD, et al. The

559   major genetic determinants of HIV-1 control affect HLA class I peptide

560   presentation. *Science.* 2010;330(6010):1551-7.

561 35. Carrington M, Nelson GW, Martin MP, Kissner T, Vlahov D, Goedert JJ, et al.

562   HLA and HIV-1: heterozygote advantage and B*35-Cw*04 disadvantage.

563   *Science.* 1999;283(5408):1748-52.

564   36.   Ananworanich J, Schuetz A, Vandergeeten C, Sereti I, de Souza M, Rerknimitr R,

565   et al. Impact of multi-targeted antiretroviral treatment on gut T cell depletion and

566   HIV reservoir seeding during acute HIV infection. *PLoS One.* 2012;7(3):e33948.

567   37.   Besson GJ, Lalama CM, Bosch RJ, Gandhi RT, Bedison MA, Aga E, et al. HIV-1

568   DNA decay dynamics in blood during more than a decade of suppressive

569   antiretroviral therapy. *Clin Infect Dis.* 2014;59(9):1312-21.

570   38.   Ananworanich J, Chomont N, Eller LA, Kroon E, Tovanabutra S, Bose M, et al.

571   HIV DNA Set Point is Rapidly Established in Acute HIV Infection and

572   Dramatically Reduced by Early ART. *EBioMedicine.* 2016;11:68-72.

573   39.   Yukl SA, Gianella S, Sinclair E, Epling L, Li Q, Duan L, et al. Differences in

574   HIV burden and immune activation within the gut of HIV-positive patients

575   receiving suppressive antiretroviral therapy. *J Infect Dis.* 2010;202(10):1553-61.

576   40.   Poizot-Martin I, Faucher O, Obry-Roguet V, Nicolino-Brunet C, Ronot-

577   Bregigeon S, Dignat-George F, et al. Lack of correlation between the size of HIV

578   proviral DNA reservoir and the level of immune activation in HIV-infected

579   patients with a sustained undetectable HIV viral load for 10 years. *J Clin Virol.*

580   2013;57(4):351-5.

581   41.   Fourati S, Flandre P, Calin R, Carcelain G, Soulie C, Lambert-Niclot S, et al.

582   Factors associated with a low HIV reservoir in patients with prolonged

583   suppressive antiretroviral therapy. *J Antimicrob Chemother.* 2014;69(3):753-6.

584   42.   Lambert-Niclot S, Flandre P, Valantin MA, Soulie C, Fourati S, Wirden M, et al.

585   Similar evolution of cellular HIV-1 DNA level in darunavir/ritonavir

586   monotherapy versus triple therapy in MONOI-ANRS136 trial over 96 weeks.

587   *PLoS One.* 2012;7(7):e41390.

588   43.   Burgard M, Boufassa F, Viard JP, Garrigue I, Ruffault A, Izopet J, et al. Factors

589   influencing peripheral blood mononuclear cell-associated HIV-1 DNA level after

590   long-term suppressive antiretroviral therapy in 236 patients. *AIDS.*

591   2009;23(16):2165-71.

592   44.   Hoen B, Cooper DA, Lampe FC, Perrin L, Clumeck N, Phillips AN, et al.

593   Predictors of virological outcome and safety in primary HIV type 1-infected

594   patients initiating quadruple antiretroviral therapy: QUEST GW PROB3005. *Clin*

595   *Infect Dis.* 2007;45(3):381-90.

596   45.   Trautmann L, Mbitikon-Kobo FM, Goulet JP, Peretz Y, Shi Y, Van Grevenynghe

597   J, et al. Profound metabolic, functional, and cytolytic differences characterize

598   HIV-specific CD8 T cells in primary and chronic HIV infection. *Blood.*

599   2012;120(17):3466-77.

600   46.   Ho YC, Shan L, Hosmane NN, Wang J, Laskey SB, Rosenbloom DI, et al.

601   Replication-competent noninduced proviruses in the latent reservoir increase

602   barrier to HIV-1 cure. *Cell.* 2013;155(3):540-51.

603   47.   Bruner KM, Murray AJ, Pollack RA, Soliman MG, Laskey SB, Capoferri AA, et

604   al. Defective proviruses rapidly accumulate during acute HIV-1 infection. *Nat*

605   *Med.* 2016;22(9):1043-9.

606   48.   Sharaf R, Lee GQ, Sun X, Etemad B, Aboukhater LM, Hu Z, et al. HIV-1 proviral

607   landscapes distinguish posttreatment controllers from noncontrollers. *J Clin*

608   *Invest.* 2018;128(9):4074-85.

609   49.   Jain V, Hartogensis W, Bacchetti P, Hunt PW, Hatano H, Sinclair E, et al.

610   Antiretroviral therapy initiated within 6 months of HIV infection is associated

611   with lower T-cell activation and smaller HIV reservoir size. *J Infect Dis.*

612   2013;208(8):1202-11.

613   50.   Jones M, Williams J, Gartner K, Phillips R, Hurst J, and Frater J. Low copy target

614   detection by Droplet Digital PCR through application of a novel open access

615   bioinformatic pipeline, 'definetherain'. *J Virol Methods.* 2014;202:46-53.

616   51.   Tibshirani R. Regression Shrinkage and Selection via the Lasso. *Journal of the*

617   *Royal Statistical Society Series B (Methodological).* 1996;58(1):267-88.

618   52.   Friedman J, Hastie T, and Tibshirani R. Regularization Paths for Generalized

619   Linear Models via Coordinate Descent. *J Stat Softw.* 2010;33(1):1-22.

620   53.   Stekhoven DJ, and Buhlmann P. MissForest--non-parametric missing value

621   imputation for mixed-type data. *Bioinformatics.* 2012;28(1):112-8.

622   54.   Waljee AK, Mukherjee A, Singal AG, Zhang Y, Warren J, Balis U, et al.

623   Comparison of imputation methods for missing laboratory data in medicine. *BMJ*

624   *Open.* 2013;3(8).

625

626 **Figures**

627 **Figure 1**. Measures of clinical progression during treated primary HIV infection
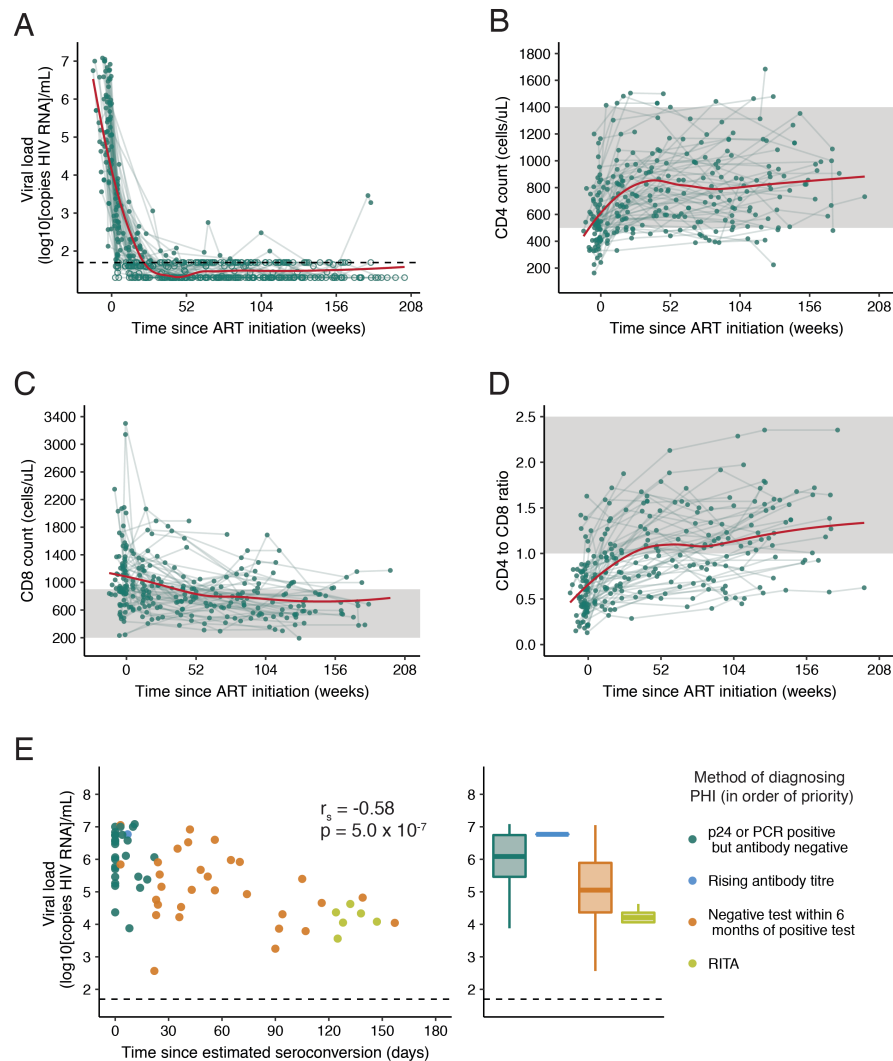
628

629 **(A)** Viral load (VL), **(B)** CD4 count **(C)** CD8 count and **(D)** CD4 to CD8 ratio in the 4 years
630 following ART initiation (n=63). A trend line (shown in red) has been fitted using local
631 polynomial regression fitting (LOESS) smoothing with an α value of 0.75. For **(A)** exact values
632 are shown as closed circles, and those below the limit of detection are shown as open circles; the
633 black dashed line indicates 50 copies/ml. In **(B-D)** the shaded region shows the normal range for
634 these parameters. **(E)** Baseline VL relative to the number of days this was measured after
635 estimated seroconversion (left panel; Spearman's correlation) with the same data as histograms
636 stratified by method of diagnosing PHI (right panel). If individuals met multiple diagnostic
637 criteria, they are plotted as the criterion with the most reliability in estimating date of
638 seroconversion; these are shown in the legend in the order of priority used.

639   **Figure 2.** Total HIV DNA during treated primary HIV infection
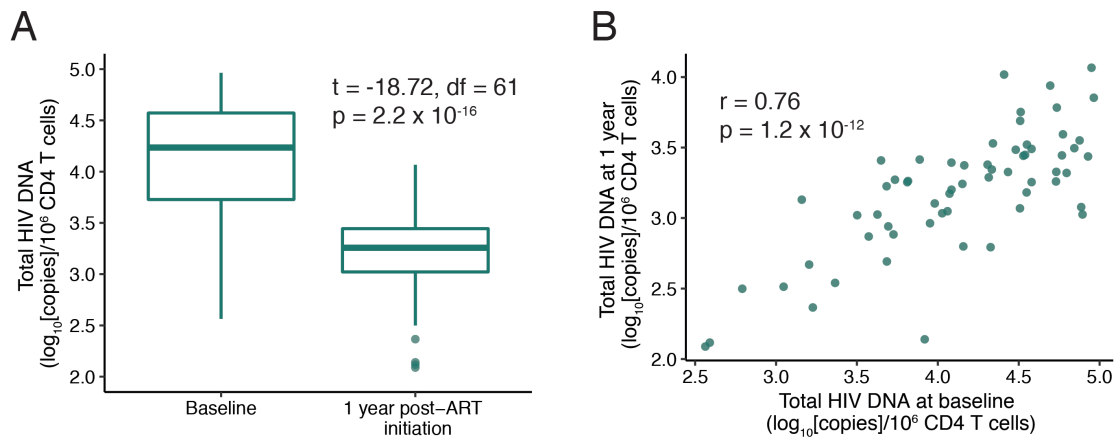

641   **(A-B)** Relationship between total HIV DNA measured at baseline and 1 year following ART
642   initiation (n=62). For **(A)** comparison was made using a paired t-test and in **(B)** a Pearson's
643   correlation was performed.

Figure 3. Immunologic and clinical variables associated with HIV reservoir size are highly correlated with one another
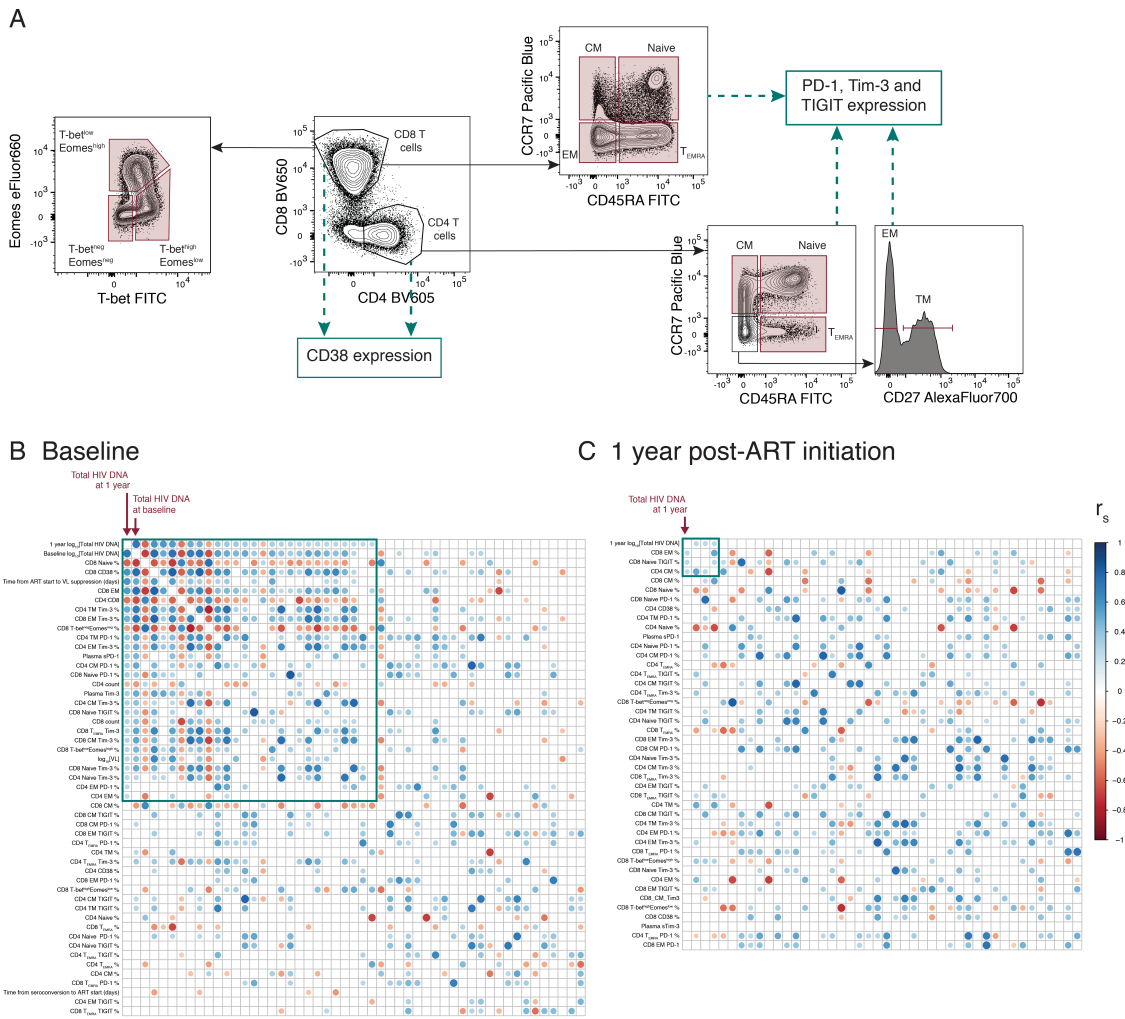
(A) Schematic showing the T cell subsets and surface markers measured by flow cytometry in this analysis. The frequency of populations gated in red was included in analysis, as well as the expression of CD38, PD-1, Tim-3 and TIGIT on populations marked. **(B-C)** shows the correlations between clinical/immunological variables and HIV reservoir size. Corrplots showing the relationship between HIV reservoir size at 1 year ($\log_{10}$ [Total HIV DNA]) and immunological/clinical variables (n=63) measured at baseline **(B)** or following 1 year of ART **(C)**. The same immunological variables were included at both time points, and clinical variables at baseline only. Reservoir size at 1 year (1 year $\log_{10}$[Total HIV DNA]) is shown in the top left corner and is marked. For both B and C, variables have been ranked based on the magnitude of absolute correlation coefficient with 1 year $\log_{10}$[Total HIV DNA] in decreasing order from the top left hand corner. The size and colour of each circle corresponds to the correlation coefficient between any two variables. Correlation coefficients were calculated using the Spearman method with pairwise complete observations; only correlations significant at a 0.05 level are shown (other boxes are left blank). Green boxes enclose variables which have a significant correlation with 1 year $\log_{10}$ [Total HIV DNA] at a 0.05 level. Abbreviations: CM, central memory; EM, effector memory.

32

663 **Figure 4.** Immunologic and clinical variables which relate to baseline total HIV DNA
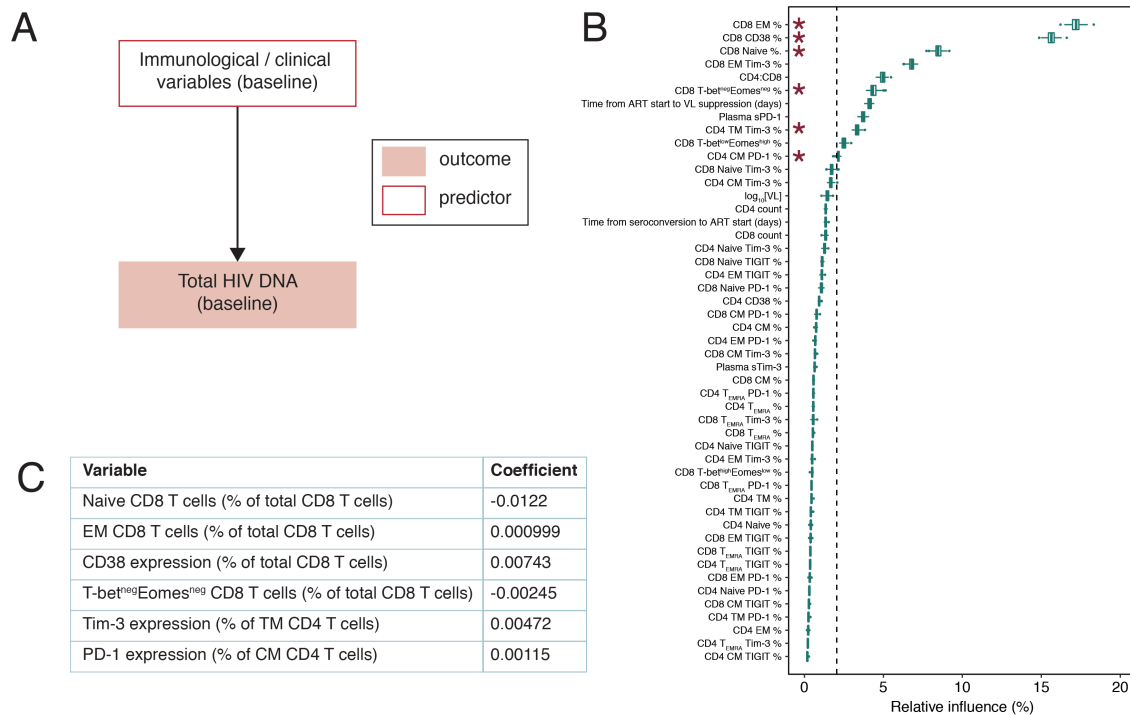
664

665 **(A)** Structure for models shown in B and C. **(B)** Boosted regression tree model to assess
666 predictors of baseline total HIV DNA (49 predictors, n=62); boxplots show the summary of 100
667 model runs. Influential predictors were defined as those whose relative contribution was greater
668 than 100 divided by the total number of covariates, this value is indicated by the dashed vertical
669 line. Variables which were also selected in C are marked with an asterisk. **(C)** Least absolute
670 shrinkage and selection operator (LASSO) output for predictors of baseline total HIV DNA (49
671 predictors, n=59, deviance explained 0.67). Variables which do not significantly contribute to the
672 model have a coefficient of zero; only those with a non-zero coefficient are shown. Missing
673 values were imputed for individuals with some but not all immunologic measures available.

33

674 **Figure 5.** Relationship between HLA class I alleles and total HIV DNA at 1 year post-

675 ART initiation
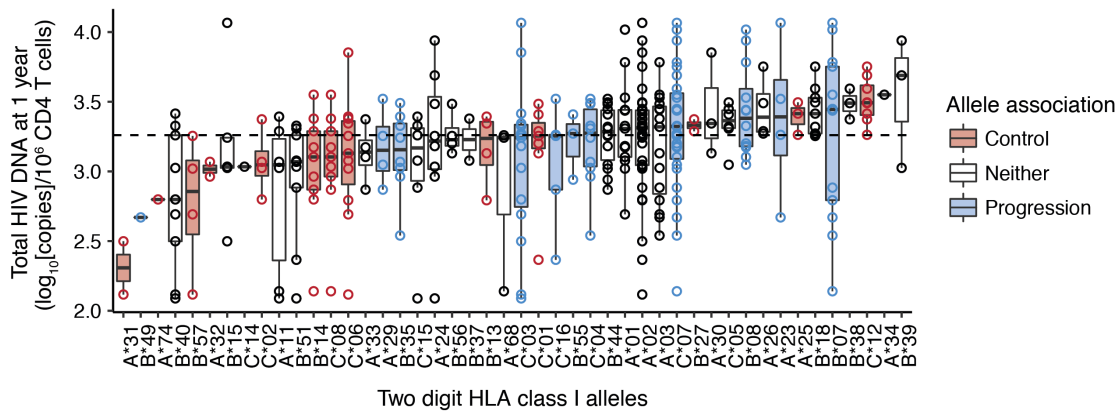
676

677 Relationship between HLA class I alleles and total HIV DNA at 1 year post-ART initiation.
678 Alleles, shown on the x-axis, are ordered based on the median value of total HIV DNA for all
679 individuals possessing that allele. Boxplots show the distribution of total HIV DNA in amongst
680 individuals possessing that allele, and each observation is shown as an open circle. Data is shown
681 for 61 individuals and a total of 333 alleles. For one individual only B alleles were available and
682 are included here; similarly for another individual only A and C alleles were available. Where
683 individuals were homozygous for a given allele, this is only shown once. The dashed line shows
684 the median value of total HIV DNA for the entire cohort. Alleles were classified as being
685 associated with disease progression (blue) or control (red), or neither (white), based on those
686 identified in the International HIV Controllers Study at a significance level of 0.05 (34).

34

687   **Figure 6.** Model for factors which influence reservoir size in treated PHI

688

35
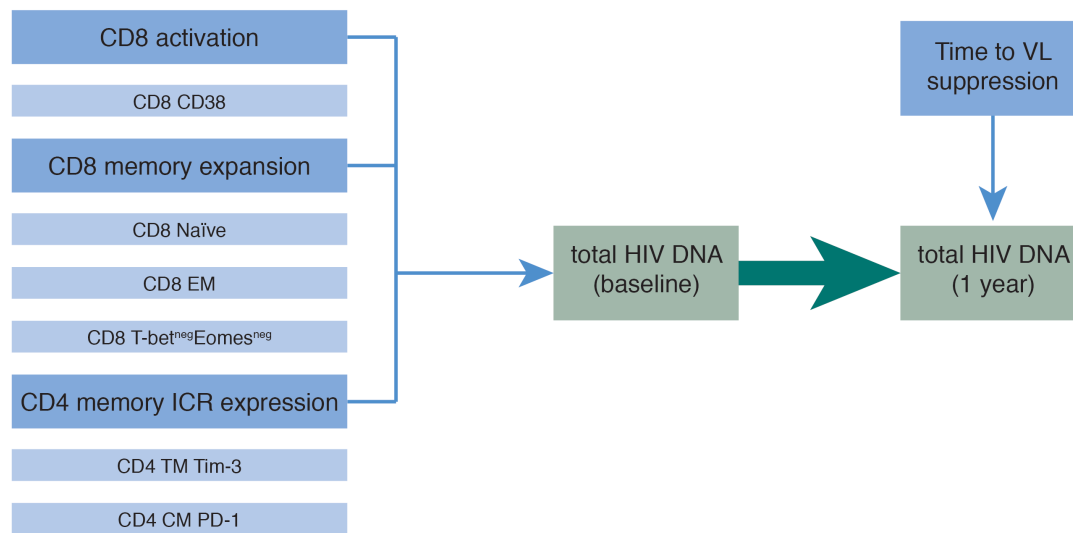
689   **Tables**

690   **Table 1.** Demographic and clinical characteristics of participants

Sex	
• Male	63 (100%)
Age (years)	36 [28 - 41]
Time between date of confirmed HIV+ test and ART initiation (days)	29 [14 - 46]
Time between estimated date of seroconversion and ART initiation (days)	49 [33 - 90]
Time between ART initiation and first VL <50 copies/ml (days)[A]	133 [93 - 241]
Baseline CD4 T cell count (cells/μL)[B]	527 [406 - 645]
Baseline CD8 T cell count (cells/μL)[B]	1033 [831 - 1326]
Baseline CD4 to CD8 ratio[B]	0.52 [0.37 – 0.76]
Baseline viral load ($\log_{10}$ copies[HIV RNA]/ml)	5.5 [4.6 – 6.5]
Method for diagnosing primary HIV infection	
• Antigen positive (p24 or PCR) but antibody negative	26 (41%)
• Rising antibody titre	1 (1.5%)
• Negative test within 6 months of positive test	30 (48%)
• Recent incidence testing algorithm	6 (9.5%)
Mode of acquisition	
• MSM	56 (89%)
• MSW	1 (1.5%)
• Unknown/unrecorded	6 (9.5%)
Initial ART regimen	
• Unknown/unrecorded	3 (4.8%)
Backbone	
• tenofovir containing	55 (87%)
• abacavir containing	5 (7.9%)
Additional agent(s)	
• protease inhibitor	36 (57%)
• NNRTI	11 (17%)
• integrase inhibitor	12 (19%)
• protease inhibitor + integrase inhibitor	1 (1.5%)

691

692   Demographic and baseline clinical characteristics of included participants from the HEATHER

693   cohort. Values given represent n (%) for categorical variables and median (interquartile range) for

694   continuous variables. Abbreviations: MSM, men who have sex with men; MSW, men who have

695   sex with women, NNRTI, non-nucleoside reverse-transcriptase inhibitor.

696   [A] 60/63 individuals were virologically suppressed to <50 copies/ml by 1 year visit

697   [B] Data available for 62/63 individuals

698   **Table 2.** Predictors of reservoir size at one year

	Model A	Model B
Predictors included	Baseline clinical and immunological variables (including baseline total HIV DNA)	1 year immunological variables and baseline total HIV DNA
	Coefficient	*Coefficient*
Baseline $\log_{10}$[total HIV DNA]	0.339	0.311
Time from ART start to VL suppression (days)	0.000342	-

699

700   Least absolute shrinkage and selection operator (LASSO) output for predictors of reservoir size

701   (total HIV DNA at 1 year). Model A includes all baseline clinical and immunological variables,

702   including baseline total HIV DNA (50 predictors, n=60, deviance explained 0.58). Model B

703   includes all immunological measures at 1 year along with baseline total HIV DNA (1 year; 44

704   predictors, n=61, deviance explained 0.47). Variables which do not significantly contribute to the

705   model have a coefficient of zero; only those with a non-zero coefficient are shown. Missing

706   values were imputed for individuals with some but not all immunologic measures available.

707   **Supplementary Materials**

708   Supplementary Figure 1 - Viral load sampling frequency

709   Supplementary Figure 2 – Reservoir size at 3 years post-ART initiation

710   Supplementary Figure 3 – Missing immunological and clinical data

711   Supplementary Figure 4 – Boosted regression tree results to assess the relative influence

712   of predictors of reservoir size at 1 year

713   Supplementary Table 1 – LASSO models using data without imputation