nature research

Corresponding author(s): Xu G. Yu

Last updated by author(s): Jun 18, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Quantify One (version 4.4.1), ChemiDoc MP Image Lab software (BioRad, version 6.0.1), BD FACSDiva software (version 8.0.1), QuantaSoft (version 1.7.4.0917)
Data analysis	Los Alamos HIV Sequence Database Hypermut 2.0 (https://www.hiv.lanl.gov/content/sequence/HYPERMUT/hypermut.html), MEGA (https://www.megasoftware.net, version 7.0.26), MUSCLE (http://www.drive5.com/muscle, version 3.8.1551), Graphpad prism (https://www.graphpad.com/scientific-software/prism/, version 8.2.1), UltraCycler v1.0 (Brian Seed and Huajun Wang from MGH CCIB DNACore, unpublished), R (https://www.r-project.org, version 3.5.3), UCSC Genome Browser (https://genome.ucsc.edu), GENCODE (https://www.gencodegenes.org, version 29), Ensembl (https://ensembl.org, version 86), RepeatMasker (www.repeatmasker.org), RSEM (https://deweylab.github.io/RSEM/, version 1.2.22), STAR (https://github.com/alexdobin/STAR,version 2.5.1b), FastQC (https://www.bioinformatics.babraham.ac.uk, version 0.11.9), Trimmomatic (http://www.usadellab.org, version 0.39), Samtools (http://www.htslib.org/, version 1.3.1), MACS2 (https://pypi.python.org/pypi/MACS2, version 2.1.1.20160309), iMethyl (http://imethyl.iwate-megabank.org), ROADMAP (http://www.roadmapepigenomics.org/), MAFFT (https://mafft.cbrc.jp/alignment/software, version 7), Highlighter (https://www.hiv.lanl.gov/content/sequence/HIGHLIGHT/highlighter_top.html), SPICE (https://niaid.github.io/spice/), Recombinant Identification Program (https://www.hiv.lanl.gov/content/sequence/RIP/RIP.html), FlowJo software (version 10.6), Bowtie2 (http://bowtie-bio.sourceforge.net/bowtie2/index.shtml, version 2.2.9), in-house intactness pipeline (https://github.com/BWH-Lichterfeld-Lab/Intactness-Pipeline)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

RNA-Seq and ATAC-Seq data have been deposited in a public repository (NCBI GEO, accession number GSE144334). Due to study participant confidentiality concerns, full-length viral sequencing data cannot be publicly released, but will be made available to investigators upon reasonable request and after signing a coded tissue agreement. The Los Alamos HIV Sequence Database Hypermut 2.0 and the Los Alamos HIV Immunology Database 2.0 are available at www.hiv.lanl.gov. The iMethyl database is available at http://imethyl.iwate-megabank.org. ROADMAP epigenomic data are available at http://www.roadmapepigenomics.org.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	A total of n=64 EC and n=41 ART-treated individuals were analyzed in data described in Figure 1. In Figure 2-4, n=11 EC are described in detail. No computational approach was used to determine these sample sizes, testing was based on availability of more than 50 million PBMC per study participant.
Data exclusions	No data from the described individuals were excluded.
Replication	Viral and integration site sequencing was performed once for each individual proviral sequence. To test the accuracy of our sequencing approach, we repeated sequencing of near full-length HIV-1 DNA from the 8E5 cell line 50 consecutive times, which resulted in 100% identical sequences in all runs.
Randomization	No randomization was performed, because we performed a cross-sectional analysis of study participants enrolled in an observational study.
Blinding	Coded samples from study participants were used throughout the study; laboratory personnel was not blinded with regard to the respective study cohorts, since this was a non-interventional, observational study. All sequencing reactions were performed at a local core facilities; core facility employees were fully blinded.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	CD3 (clone OKT3, BioLegend, catalog 317332) CD4 (clone RPA-T4, BioLegend, catalog 300518) CCR7 (clone G043H7, BioLegend, catalog 353216) CCD45RO (clone UCHL1, BioLegend, catalog 304236) CD3/CD8 bi-specific antibody (NIH AIDS Reagent Program #12277)

Validation	CD3 (clone OKT3, BioLegend, catalog 317332): Reactivity: Human Host Species: Mouse Application: FC - Quality tested Application Notes: The OKT3 monoclonal antibody reacts with an epitope on the epsilon-subunit within the human CD3 complex. Clone OKT3 can block the binding of clones SK7 and UCHT1.4 The OKT3 antibody is able to induce T cell activation. Additional reported applications (for the relevant formats) include: immunohistochemical staining of acetone-fixed frozen sections and activation of T cells. The LEAF™ purified antibody (Endotoxin <0.1 EU/µg, Azide-Free, 0.2 µm filtered) is recommended for functional assays (Cat. No. 317304). For highly sensitive assays, we recommend Ultra-LEAF™ purified antibody (Cat. No. 317326) with a lower endotoxin limit than standard LEAF™ purified antibodies (Endotoxin <0.01 EU/µg). Application References: Schlossman S, et al. Eds. 1995. Leucocyte Typing V. Oxford University Press. New York. Knapp W. 1989. Leucocyte Typing IV. Oxford University Press New York. Barclay N, et al. 1997. The Leucocyte Antigen Facts Book. Academic Press Inc. San Diego. Li B, et al. 2005. Immunology 116:487. CD4 (clone RPA-T4, BioLegend, catalog 300518): Reactivity: Human, Chimpanzee Host Species: Mouse Application: FC - Quality tested Application Notes: The RPA-T4 antibody binds to the D1 domain of CD4 (CDR1 and CDR3 epitopes) and can block HIV gp120 binding and inhibit syncytia formation. Additional reported applications (for the relevant formats) include: immunohistochemistry of acetone-fixed frozen sections3,4,5, and blocking of T cell activation1,2. This clone was tested in-house and does not work on formalin fixed paraffin-embedded (FFPE) tissue. The LEAF™ purified antibody (Endotoxin <0.1 EU/µg, Azide-Free, 0.2 µm filtered) is recommended for functional assays (Cat. No. 300516). Application References: Knapp W, et al. 1989. Leucocyte Typing IV. Oxford University Press. New York. (Activ) Moir S, et al. 1999. J. Virol. 73:7972. (Activ) Deng MC, et al. 1995. Circulation 91:1647. (IHC) Friedman T, et al. 1999. J. Immunol. 162:5256. (IHC) Mack CL, et al. 2004. Pediatr. Res. 56:79. (IHC) Lan RY, et al. 2006. Hepatology 43:729. Zenaro E, et al. 2009. J. Leukoc. Biol. 86:1393. (FC) PubMed Yoshino N, et al. 2000. Exp. Anim. (Tokyo) 49:97. (FC) Stoeckius M, et al. 2017. Nat. Methods. 14:865. (PG) CCR7 (clone G043H7, BioLegend, catalog 353216): Reactivity: Human, African Green, Baboon, Cynomolgus, Rhesus Host Species: Mouse Application: FC - Quality tested CCD45RO (clone UCHL1, BioLegend, catalog 304236): Reactivity: Human, Chimpanzee, Cynomolgus, Common Marmoset Host Species: Mouse Application: FC - Quality tested Application Notes: The UCHL1 antibody is commonly used in combination with antibodies against CD45RA to discern memory and naïve T cells. Additional reported applications (for the relevant formats) include: immunohistochemical staining of acetone-fixed frozen tissue sections5 and formalin-fixed paraffin-embedded tissue sections4, Western blotting2, and immunoprecipitation3. Application References: Knapp W, et al. Eds. 1989. Leucocyte Typing IV. Oxford University Press. New York. (FC) Ishii T, et al. 2001. P. Natl. Acad. Sci. USA 98:12138. (WB) Ponsford M, et al. 2001. Clin. Exp. Immunol. 124:315. (IP) Yamada M, et al. 1996. Stroke 27:1155. (IHC) Sakkas LI, et al. 1998. Clin. Diagn. Lab. Immunol. 5:430. (IHC) CD3/CD8 bi-specific antibody (NIH AIDS Reagent Program #12277) The bi-specific CD3/8 (CD3.8) monoclonal antibody was generated by fusing the anti-CD3 mAb producing hybridoma (12F6) with the anti-CD8 mAb producing hymbridoma (OKT8). The resulting anti-CD3/8 antibody, when added to long term peripheral blood co-cultures results in the potent elimination of CD8+ T cells. The remaining cells are highly activated and serve as a reliable source of purified activated cells of interest.

Human research participants

Policy information about studies involving human research participants

Population characteristics	Please see Extended Data Table 1.
Recruitment	EC and ART-treated individuals were recruited based on referral by HIV clinicians and infectious disease physicians. The enrollment protocols allowed recruited of men and women >18 years old, of any race or ethnicity.
Ethics oversight	The Partners Human Research Committee approved all sample collection at MGH and BWH; the IRB of UCSF supervised sample collection at UCSF.

Note that full information on the approval of the study protocol must also be provided in the manuscript.