

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see [Reporting Life Sciences Research](#). For further information on Nature Research policies, including our [data availability policy](#), see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

► Experimental design

1. Sample size

Describe how sample size was determined.

The primary data reported in this study is the DNA genome sequence of a monkey species, the Sooty Mangabey. The animal selected for sequencing was one of the original dams of a large matrilineal line of the colony. In addition, she possessed the most common MHC haplotype observed within the group. As such, her genetic constitution within the closed population was thought to be most representative of any single animal. To verify sequencing findings and estimate population penetrance of alleles, we sequenced the genome of 10 additional animals - these animals were selected as they collectively represented 76% of the offspring of the Yerkes National Primate Research colony.

For RNA-Seq, and microarray we used a power calculation performed on a model set of rhesus macaques in which RNA expression at different effect sizes (fold-changes) gave us an estimate of our sensitivity to detect changes to approximately 1.5 fold. For LPS stimulation studies, as the effect sizes (species differences) were much larger than most gene expression differences based on pilot stimulations, we also used this power calculation dataset.

2. Data exclusions

Describe any data exclusions.

To identify novel immunogenetic factors specific for SMs that may be involved in their ability to avoid progression to AIDS, we established a bioinformatic pipeline for a comparative protein analysis. During this process we excluded proteins based on certain criteria: The SM protein model alignment with the lowest e-value or highest bitscore (for equal e-values) was selected for each MacaM protein model, yielding the set of orthologous SM protein predictions most like the rhesus protein models. SM CDS sequences were screened against the de novo RNA-seq assembly transcript models by alignment with blat and an alignment score was calculated as the number of matching bases reduced by the number of CDS sequence bases missing in alignment gaps normalized by the CDS sequence length. This score penalizes bases missing from the CDS sequence without penalizing extra sequence that may have been added to the RNA-seq transcript model during the assembly process. We then excluded SM protein sequences that are not supported by the RNA-seq data (<99 % matching sequence), thereby eliminating those that had been incorrectly identified as divergent from rhesus macaque due to errors in transcript or protein models. We also excluded 6,906 SM protein models with more than 97% identity with rhesus orthologs from further analysis. Proteins of interest for differential response to lentivirus infection may be more divergent than expected on average. We thus filtered the remaining divergent SM proteins (i.e. those with less than 97% identity) that have high sequence confidence using the GO term "immune response," which after manual inspection resulted in a list of 34 candidate immune-related genes that are sequence divergent between SMs and rhesus macaques (Table 2, Extended data table 1).

Sequencing and de novo assembly of RNA-seq transcripts required a number of filtering steps to prepare threads for de novo assembly which included removing adapters, filtering for quality, removing poly A/T tails, and removing mtDNA and common mammalian rRNA.

3. Replication

Describe whether the experimental findings were reliably reproduced.

We reproduced experimental findings in the following manner:

1. To validate our findings of sequence divergence in the sooty mangabey genome, we performed full genome sequencing of an additional 10 animals. We also conducted Sanger sequencing of genomic DNA of genes of interest, and used available RNA-Seq data to verify the sequence data.
2. For the LPS stimulation data, we performed LPS stimulation of 9 RMs and 8 SMs -over three separate experiment days, using multiple dosages. Each data point shown represents the average of three technical replicates per animal. We repeated the LPS experiment in SIV-infected SMs (n = 5) and RMs (n = 4). The finding of reduced TNFa and IL6 cytokine release was measured by protein-based quantitation (CBA), by real-time PCR for RNA levels, and by RNA-Seq - with the observation of reduced production consistent across the different assays. We did not observe a failure to reliably produce our findings.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

Where possible samples were run in single batches using equal numbers of samples from species compared (i.e. rhesus macaque and sooty mangabey), thus they were balanced. For the in vivo SIV infection experiment, groups were divided 3 SM +2 RMs for processing group 1 and 2 SMs and 3 RMs for processing group 2. For the cytokine bead array experiments, ICAM2 surface expression measurements, multiple batches were measured over several days - to minimize batch effects, we would run 3-5 individuals from each species, with an equal number of RM and SM individuals run in a single batch, all samples for a given individual animal (ie. Time points or stimuli doses) were assayed together. For RNA-Seq processing, samples were run in a single batch for all processing steps, and for sequencing, bar-coded cDNA fragments from all samples were pooled and split evenly across multiple lanes to reduce lane bias.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

Blinding was incorporated for RNA-seq and microarray analysis by assigning samples de-identified inventory codes prior to submission for processing in which no information about individual, species or treatment condition was provided.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

- | | |
|--------------------------|--|
| n/a | Confirmed |
| ☐ | ☒ The <u>exact sample size</u> (n) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.) |
| ☐ | ☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ A statement indicating how many times each experiment was replicated |
| ☐ | ☒ The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section) |
| ☐ | ☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons |
| ☐ | ☒ The test results (e.g. <i>P</i> values) given as exact values whenever possible and with confidence intervals noted |
| ☐ | ☒ A clear description of statistics including <u>central tendency</u> (e.g. median, mean) and <u>variation</u> (e.g. standard deviation, interquartile range) |
| ☐ | ☒ Clearly defined error bars |

See the web collection on [statistics for biologists](#) for further resources and guidance.

► Software

Policy information about [availability of computer code](#)

7. Software

Describe the software used to analyze the data in this study.

ALLPATHS-LG, Atlas-Link (v.1.0), Atlas GapFil, Phrap, Newbler, Velvet, GraphPad Prism, PBJelly (v14.9.9), GATK, Trinity, blastp v2.2.28+, FlowJo 10, cufflinks v2.1.1,

blat (v34), GSEA, CAFE v3.0, PAML v4.9a, PASTA2, RAXML3, STAR version 2.5.2, DESeq2, Partek Genomics software version 6.6

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* [guidance for providing algorithms and software for publication](#) provides further information on this topic.

► Materials and reagents

Policy information about [availability of materials](#)

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

no restrictions

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

CD3-APC/Cy7 (SP34-2), CD14-PE/Cy7 (M5E2) and CD20-PE/Cy5 (2H7) from BD; CD4-BV650 (OKT4), CD8-BV711 (RPA-T8), ICAM-2-FITC (CBR-IC2/2), Mouse IgG2a(κ)-FITC (MOPC-173) isotype control from Biolegend; Live/Dead Fixable Aqua from Thermo Fisher Scientific. Antibodies were validated using matching isotype control.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

HEK293T from ATCC.

b. Describe the method of cell line authentication used.

Cells were purchased from ATCC who provided an authentication certificate.

c. Report whether the cell lines were tested for mycoplasma contamination.

Cells are routinely tested for mycoplasma.

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by [ICLAC](#), provide a scientific rationale for their use.

The 293T variant of HEK cells is not listed in the ICLAC database. 293T cells are widely used to study protein variants in a well-defined cellular environment by transfection.

► Animals and human research participants

Policy information about [studies involving animals](#); when reporting animal research, follow the [ARRIVE guidelines](#)

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

Blood draws were obtained from sooty mangabeys and rhesus macaques housed at the Yerkes National Primate Research Center, which is accredited by American Association of Accreditation of Laboratory Animal Care. This study was performed in strict accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health, a national set of guidelines in the U.S. and also to international recommendations detailed in the Weatherall Report (2006). This work received prior approval by the Institutional Animal Care and Use Committees (IACUC) of Emory University (IACUC protocol #2000793, entitled “Comparative AIDS Program”).

Sooty mangabey (RefSeq animal):

GenBank Bioproject, species, specimen ,sex, age (years)
PRJNA279144, sooty mangabey, blood draw, Female, 23

Sooty mangabey diversity sequencing panel:

GenBank BioSample, species, specimen, sex, age (years)
SAMN02900445, sooty mangabey, blood draw, Female, 23
SAMN02900446, sooty mangabey, blood draw, Female, 23
SAMN02900447, sooty mangabey, blood draw, Male, 31
SAMN02900448, sooty mangabey, blood draw, Male, 20
SAMN02900449, sooty mangabey, blood draw, Male, 20
SAMN02900450, sooty mangabey, blood draw, Female, 19
SAMN02900451, sooty mangabey, blood draw, Male, 17
SAMN02900452, sooty mangabey, blood draw, Female, 16
SAMN02900453, sooty mangabey, blood draw, Female, 16
SAMN02900454, sooty mangabey, blood draw, Male, 15

Rhesus macaque SIV infection study:

Animal study ID, species, specimen, sex, age (years)
RM1, rhesus macaque, blood draw, Male, 12
RM2, rhesus macaque, blood draw, Male, 12
RM3, rhesus macaque, blood draw, Male, 9
RM4, rhesus macaque, blood draw, Male, 9
RM5, rhesus macaque, blood draw, Male, 12
RM6, rhesus macaque, blood draw, Male, 5
RM7, rhesus macaque, blood draw, Male, 5
RM8, rhesus macaque, blood draw, Male, 5
RM9, rhesus macaque, blood draw, Male, 5
RM10, rhesus macaque, blood draw, Male, 5

Policy information about [studies involving human research participants](#)

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

Human biospecimens (blood) were obtained from healthy donors by blood draw into EDTA tubes at the Yerkes National Primate Research Center in accordance with internal review board protocol IRB0004582 and all relevant ethical regulations. Informed consent was obtained from all blood donors.

Flow Cytometry Reporting Summary

Form fields will expand as needed. Please do not leave fields blank.

► Data presentation

For all flow cytometry data, confirm that:

- ☒ 1. The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- ☒ 2. The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- ☐ 3. All plots are contour plots with outliers or pseudocolor plots.
- ☒ 4. A numerical value for number of cells or percentage (with statistics) is provided.

► Methodological details

- | | |
|--|---|
| 5. Describe the sample preparation. | Peripheral blood samples from SIV negative rhesus macaques and SIV negative sooty mangabeys were collected by venipuncture according to standard procedures at the Yerkes National Primate Research Center of Emory University and in accordance with U.S. National Institutes of Health guidelines. PBMCs were isolated by Ficoll density gradient centrifugation. |
| 6. Identify the instrument used for data collection. | BD LSRII |
| 7. Describe the software used to collect and analyze the flow cytometry data. | Cells were collected using BD FACSDIVA. Data was analyzed using FlowJo 10. |
| 8. Describe the abundance of the relevant cell populations within post-sort fractions. | Purity after sorting was >95% as confirmed by flow cytometry. |
| 9. Describe the gating strategy used. | Singlets were excluded using FSC-H and FSC-A. Lymphocytes were gated using FSC-A and SSC-A. Dead cells were excluded using LIVE/DEAD™ Fixable Aqua Dead Cell Stain Kit. Then CD3+CD8+ Cytotoxic T lymphocytes and CD3+CD4+ T helper cells and CD3-CD14-CD20+ B cells were defined. We present primary flow cytometry data in histograms since histograms are typically used to compare the staining of the isotype control antibody to the epitope-specific antibody in the same graph. |

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information. ☒