

## Single-cell transcriptome analysis illuminating the characteristics of species-specific innate immune responses against viral infections

--Manuscript Draft--

<b>Manuscript Number:</b>	GIGA-D-23-00007R1	
<b>Full Title:</b>	Single-cell transcriptome analysis illuminating the characteristics of species-specific innate immune responses against viral infections	
<b>Article Type:</b>	Research	
<b>Funding Information:</b>	Japan Science and Technology Corporation (CREST JPMJCR20H4)	Professor Kei Sato
<b>Abstract:</b>	<p><b>Background</b> Bats harbor various viruses without severe symptoms and act as their natural reservoirs. The tolerance of bats against viral infections is assumed to originate from the uniqueness of their immune system. However, how immune responses vary between primates and bats remains unclear. Here, we characterized differences in the immune responses by peripheral blood mononuclear cells to various pathogenic stimuli between primates (humans, chimpanzees, and macaques) and bats (Egyptian fruit bats) using single-cell RNA sequencing.</p> <p><b>Results</b> We show that the induction patterns of key cytosolic DNA/RNA sensors and antiviral genes differed between primates and bats. A novel subset of monocytes induced by pathogenic stimuli specifically in bats was identified. Furthermore, bats robustly respond to DNA virus infection even though major DNA sensors are dampened in bats.</p> <p><b>Conclusions</b> Overall, our data suggest that immune responses are substantially different between primates and bats, presumably underlying the difference in viral pathogenicity among the mammalian species tested.</p>	
<b>Corresponding Author:</b>	Kei Sato Tokyo Daigaku Minato-ku, Tokyo JAPAN	
<b>Corresponding Author Secondary Information:</b>		
<b>Corresponding Author's Institution:</b>	Tokyo Daigaku	
<b>Corresponding Author's Secondary Institution:</b>		
<b>First Author:</b>	Hirofumi Aso	
<b>First Author Secondary Information:</b>		
<b>Order of Authors:</b>	Hirofumi Aso Jumpei Ito Haruka Ozaki Yukie Kashima Yutaka Suzuki Yoshio Koyanagi Kei Sato	
<b>Order of Authors Secondary Information:</b>		
<b>Response to Reviewers:</b>	**Respond to Reviewers' is also prepared as a supplemental material (230812ResponseLetter_v12.docx). Please also find this file as well.	

Reviewer #1: Hirofumi Aso and colleagues provide a manuscript entitled 'Single-cell transcriptome analysis illuminating the characteristics of species specific innate immune responses against viral infections'. The aim was to describe differences in innate immune responses of peripheral blood mononuclear cells (PBMCs) from different primates and bats against various pathogenic stimuli (different viruses and LPS). A major conclusion from the study is that differences in the immune response between primate and bat PBMCs are more pronounced than those between DNA, RNA viruses or LPS, or between the cell types.

The topic is of interest as the immunological basis for how bats appear to be largely disease resistant to some viruses that cause severe infections in humans is not well understood. One notion by others has been that bats have a larger spectrum of interferon (IFN) type I related genes, some of which are expressed constitutively even in unstimulated tissue, and there, trigger the expression of IFN stimulated genes (ISGs). Alongside, enhanced ISG levels may need to be compensated for in bats. Accordingly, bats may exhibit reduced diversity of DNA sensing pathways, as well as absence of a range of proinflammatory cytokines triggered in humans upon encountering acute disease causing viruses.

The study here uses single-cell RNA sequencing (scRNA-seq) analysis, and transcript clustering algorithms to explore the profile of different innate immune responses upon viral infections of PBMCs from H sapiens, Chimpanzee, Rhesus macaque, and Egyptian fruit bat. Most commonly referred to cell types were detected in all four species, although naïve CD8+ T cells were not detected in bat PBMCs, which led the authors to focus on B cells, naïve T cells, killer T/NK cells, monocytes, cDCs, and pDCs. The study used three pathogenic stimuli, Herpes simplex virus 1 (HSV1), Sendai virus (SeV), and lipopolysaccharide (LPS).

#### Specific comments

The text is well written, concise, and per se interesting, but I have a few questions for clarification.

Thank you very much for your high evaluation. We have responded to the points you raised and hope you recognize the improvements we have made.

1) Can the authors provide quality and purity control data for the virus inocula to document virus homogeneity? E.g., neither the methods, nor the indicated ref 26 specify if or how HSV1 was purified. Same is true for SeV where the provided ref 34 does not indicate if virus was purified or not. If virus inocula were not purified then it remains unclear to what extent the effects on the PBMCs described in the study here were due to virus or some other component in the inoculum. Conditions using inactivated inoculum might help to clarify this issue.

Thank you for pointing this out and we apologize for the lack of clarity in our descriptions regarding this point. In this study, we did not purify either HSV-1 or SeV. Thus, it is possible that something in the supernatant of virus-producing cells (including viruses) could affect the innate immune responses in PBMCs. However, given that the use of the supernatants without purification is a conventional virological method, any unanticipated effects would be much smaller than the effects of viral infection. To make it clear, we have added a description about this information in the revised manuscript in Materials and Methods (Lines 509-510, 522-523).

Lines 509-510: "Briefly, Vero cells were infected with HSV-1 and the supernatant was collected and used without purification."

Lines 522-523: "Briefly, LLC-MK2 cells were infected with SeV and the supernatant was collected and used without purification."

2) What was the infection period? Was it the same for all viruses?

Samples were analyzed at one day post infection. This is a condition shared by all viruses and stimuli, described in Lines 539-541 (Method: Infection and stimulation) in the revised manuscript. To emphasize that they are the same conditions, minor

corrections have been made to the revised manuscript (Lines 129-133 and Lines 539-541).

Lines 129-133: "To analyze immune responses to stimuli at single-cell resolution, we performed scRNA-seq analysis of 16 types of PBMC samples: four mammalian species (Hs, Pt, Mm, and Ra) versus four conditions (mock infection/stimulation, HSV-1 infection, SeV infection, and LPS stimulation) using the 10x Genomics Chromium platform at one day post infection."

Lines 539-541: "At one day post infection, all types of infected/stimulated PBMCs were centrifuged, resuspended in PBS, and used for bulk RT-qPCR and scRNA-seq (see below)."

3) Upon stimuli application, there was a notable expansion of B cells and a compression of killer T / NK cells in the bat but not the human samples, as well as compression of monocytes, the latter observed in all four species. Can the authors comment on this observation?

Thank you for your valuable suggestion. According to the reviewer's comment, we have edited the main text in the revised manuscript as follows (Lines 168-175, 326-328):

Lines 168-175: "The frequency of monocytes decreased after stimulation in all four species, whereas the frequency of B cells and killer TNK cells changed differently among the animal species and stimuli. Upon stimulation, there was generally a notable increase in B cells and a decrease of killer TNK cells in the bat (and non-human primates) samples, but not in the human samples. After SeV infection, the frequency of B cells was decreased in all four species. On the other hand, after HSV-1 infection, the frequency of B cells was decreased in only humans."

Lines 326-328: "It is also noteworthy that changes in cell composition differed between humans and bats (Fig. 1D)."

4) Lines 78-79: I do not think that TLR9 ought to be classified as a cytosolic DNA sensor. Please clarify.

Thank you for pointing this out. It is true that this wording was incorrect. Therefore, we fixed the issues in the manuscripts (Lines 80-82).

Lines 80-82: "Cytosolic Extrachromosomal DNAs, a PAMP for DNA viruses, are recognized by cytosolic DNA sensors, such as (e.g., cGAS, AIM2, and IFI16), and endosomal DNA sensors (e.g., TLR9)5,6,9 [5, 6, 9]."

5) Line 117: please clarify that the upregulation of proinflammatory cytokines, ISGs and IFNB1 was measured at the level of transcripts not protein.

Thank you for pointing this out. According to the reviewer's suggestion, we clarified that the upregulation of those factors was measured at the level of mRNA transcripts (Lines 127-128).

Lines 127-128: "IFNB1 (Fig. S1A-C) at the level of mRNA transcripts."

6) Line 244: DNA sensors. Authors report that bats responded well to DNA viruses, although some of their DNA sensing pathways (e.g., STING downstream of cGAS, AIM2 or IFI16) were attenuated compared to primates (H sapiens, Chimpanzee, Macaque). And they elute to the dsRNA PRR TLR3. But I am not sure if TLR3 is the only PRR to compensate for attenuated DNA sensing pathways. The authors might want to explicitly discuss if other RNA sensors, such as RIG-I-like receptors (RIG-I, LGP2, MDA5) were upregulated similarly in bats as in primate cells upon inoculation with HSV1.

Thank you for the important suggestions. As you have pointed out, it is possible that PRRs other than TLR3 compensates for weak DNA sensing pathways for HSV-1. We specially mentioned TLR3 in the original manuscript because the importance of human TLR3 in the response to HSV-1 has been demonstrated (Sato, Kato et al., Nature Immunology, 2018). However, we agree with your opinion that other possibilities should also be mentioned and have added them in the revised manuscript (Lines 344-354).

Lines 344-354: "An alternative possibility is that the IFN response in response to HSV-1 infection was triggered by sensing viral molecules other than DNAs: it is known that, in humans and mice, dsRNA sensing by TLR3 plays an important role in responding to HSV-1 infection<sup>26,31</sup>. Furthermore Additionally, the Egyptian fruit bat genome encodes an intact TLR3 gene (NCBI Gene ID: 107510436), and bat immune cells express TLR3 (Fig. 3B). Furthermore, other RNA sensors, such as RIG-I, LGP2, and MDA5, were upregulated in bat cells similarly as in primate cells upon HSV-1 infection (Fig. 3B). These data suggest that in bats, bat TLR3 or other RNA sensors in bats may compensate for the immune responses induced by weakened DNA sensing pathways, leading to IFN responses to HSV-1 infection."

7) Is it known how much TLR3 protein is expressed in bat PBMCs under resting and stimulated conditions? Same question for the DNA and RNA sensor proteins, e.g., cGAS, AIM2 or IFI16, RIG-I, LGP2, MDA5, or effector proteins, such as STING. To our knowledge, there are no publications quantifying the expression of TLR3 or other sensor genes in bats at the protein level. This is mainly due to the lack of antibodies recognizing these proteins of bats.

8) Can authors clarify if cGAS is part of the attenuated DNA sensors in the bat samples under study here? And it would be nice to see the attenuated response of DNA sensing pathways in the bat samples, as suspected from the literature, including STING downstream of cGAS, or AIM2 and IFI16.

Thank you for these suggestions. We address these two points as follows:

> Can authors clarify if cGAS is part of the attenuated DNA sensors in the bat samples under study here?

In a previous study (Xie et al., Cell Host & Microbe, 2018), the cGAS signaling pathway is reduced by a mutation in STING (not in cGAS), and this mutation is conserved in bat species including *Rousettus aegyptiacus*, the bat species analyzed in this study. Therefore, the cGAS signaling pathway should be reduced in the bat PBMC samples in this study. Thus, to clarify this point, we modified the text in the revised manuscript (Lines 335-337).

Lines 335-337: "It is known that two DNA sensing pathways mediated by cGAS-STING pathway [16]16 and PYHIN proteins, including AIM2 and IFI16 [17]17, are dampened in bats, including Egyptian fruit bats."

> And it would be nice to see the attenuated response of DNA sensing pathways in the bat samples, as suspected from the literature, including STING downstream of cGAS, or AIM2 and IFI16.

We agree this is a good idea. Unfortunately, it is difficult to evaluate the activity of each DNA sensing pathway using the limited number of bat samples in this study. Further studies are required.

9) What are the expression levels of IFN-I and related genes in the bat cells among the different stimuli?

Thank you very much for your constructive comments. We agree it is important to examine the expression level of IFN-I itself when discussing innate immune responses. Unfortunately, however, we could not obtain the expression level of IFN- $\alpha$  genes in bat samples because IFN- $\alpha$  genes were not included in the gene annotations of *Rousettus aegyptiacus* (both RefSeq and Bat1K). In addition, although we tried to measure the expression level of IFN- $\alpha$  genes in bat samples using the custom gene annotation including bat IFN- $\alpha$  genes, we failed to detect the expression of these genes (Figure R1A-F). Therefore, we did not focus on the expression of IFN-I genes in this study. Instead, we discussed the expression level of ISGs as a surrogate for the activity of IFN-I pathway as shown in Fig. 3A. We added this explanation in Results in the revised manuscript (Lines 261-268).

Lines 261-268: "To test this hypothesis, we analyzed the IFN response upon HSV-1 (a DNA virus) infection. However, the expression levels of IFN- $\alpha$  genes were not examined because they were not annotated in the transcript model for the Egyptian fruit bat used in this study, and the expression of IFN- $\beta$  genes were too low. Thus, even though the expression levels of IFN-I is the primary factor to examine the activity

of the IFN response, we instead by analyzing analyzed the induced levels of “coremamm ISGs”—a set of genes that are commonly induced by type I IFNs across mammals that were defined in a previous study 25 [28].”

Note: Although HSV-1-infected bat sample showed high expression levels of IFNA3 and IFNB1 (Figure R1E), only one cell expressed most of IFNA3 and IFNB1 (Figure R1F). Therefore, this cell was determined to be an outlier.

Figure R1. Difficulties in quantitative analysis of expression levels of IFN- $\alpha$  in bat samples

(A) Schematic of the IFN-I loci near the KLHL9 locus, originated from Zhou et al. [PMID: 26903655]. (B) Information extracted from the transcript model (gtf file) of *Rousettus aegyptiacus* in the initial manuscript. (C) Results of blastn for IFN- $\alpha$  sequences near the KLHL9 locus. Query: *Rousettus aegyptiacus* IFN- $\alpha$  (GenBank: AB259762.1). Subject: mRouAeg1.p “NW\_023416306.1:61702666-62433637”. Red indicates the alignment patterns with high mapping scores. (D) Information extracted from custom gene annotation used in E and F. (E, F) Results of mapping and counting by Cell Ranger using custom gene annotation. (E) Sum of counts for each bat sample. (F) Distribution of counts per cell for IFNA3 (top) and IFNB1 (bottom) in HSV-1-infected bat sample.

10) Technical point: where can the raw scRNA-seq data be found?

As stated in Data and code availability, the raw scRNA-seq data is found in NCBI GEO (GSE218199). To clarify this, we modified the text in the revised manuscript (Lines 837-838).

Lines 837-838: “The raw and processed Ssingle-cell RNA-seq data have been deposited in the Gene Expression Omnibus (GEO) database (GSE218199) and are publicly available.”

Reviewer #2: This paper gives a good introduction on bats as reservoirs of several viral infections, which studies have shown is due to the uniqueness of their immune system. They and others suggest that bats immune system is dampened exhibiting tolerance to various viruses. This gives the study a good rationale as to why study the bats immune system, compared to other mammals. They also give a good rationale as to why they used single-cell sequencing, to allow the identification of various cell types and the differences in these cell types. From their finding the main conclusions are that differences in the host species are more impactful; than those among the different stimuli. They also suggest that bats initiate an innate immune response after infection with DNA viruses through an alternative pathway. For example, the induction dynamics of PRRs seems to be different in their dataset. They also suggest this could be due to the presence of species-specific cellular subsets.

1. Interesting model system and a good comparison of bats with other mammals.
2. Good technique in using single-cell sequencing, with a clear rationale as to why it was chosen. This advances knowledge on what was already known about bats immune system, but the species-specific cellular subsets are new.
3. Interesting technique to go through the bulk transcriptomic data in four species and four conditions. This allowed findings of the most important genes/pathways.
4. Good rationale / flow of experiments from one to another
5. I liked that they investigated stimuli from different pathogens , including DNA, RNA virus and bacteria and still show that bats had a different immune system, in the different stimuli.

Thank you for your appreciation on various points. Based on your comments, we have made modifications to improve the manuscript.

Minor comments

1. Do they speculate this occurrence in is this just in Egyptian Fruit bats or all species of bats?

Although we consider that the differences between *Rousettus aegyptiacus* and primates observed in this study mainly originate from those between bats and primates, we agreed that there should be differences among bat species. To clarify this point, we added some descriptions to the Discussion (Limitations of the study) in the revised manuscript (Lines 407-410).

Lines 407-410: "Moreover, because the results of this study rely on an analysis using a single bat species, the Egyptian fruit bat, it is unclear whether the observed bat-specific characteristics are conserved across bat species."

2. Mentioned in the introduction why they used the egyptian fruit bats - which are a model organism, but this could help people who are not in this field understand exactly why use these bats. Advantages? Location? Proximity to the various viruses based on the fact they are mostly found in endemic regions such as Africa etc.

Thank you for mentioning this important point. Indeed, the description of the reason for using the Egyptian Fruit bats was insufficient. The first reason why we selected this bat species is that they were available because they were captive bred. The second reason is that they are a species known to be asymptotically infected with human pathogenic viruses, such as Marburg virus. We have stated this point in the Introduction and Results in the revised manuscript (Lines 107-111 and Lines 119-122).

Lines 107-111: "Here, we used peripheral blood mononuclear cells (PBMCs) from four mammalian species including the Egyptian fruit bats and three pathogenic stimuli and conducted single-cell RNA sequencing (scRNA-seq) analysis to elucidate the differences in innate immune responses against pathogenic stimuli."

Lines 119-122: "In this study, the Egyptian fruit bat was used as a representative model organism for bat species because this bat species is bred and available in captivity and is known as natural host of human pathogenic viruses, such as Marburg virus [3]."

3. Can they include viral load in each species?

"Viral load" was not quantified, but the amount of RNA in the collected samples was quantified by RT-qPCR (Fig. S1B-C).

4. It is not clear which scRNAseq tools were used for data analysis in identifying the types of cells. Or did they use already established database based on markers?

Thank you for this comment. Details of the method can be found in the Cell Annotation section of Methods. In short, this study uses Seurat and Azimuth to annotate each cluster resulting from unsupervised clustering based on several pieces of information (Fig. S2A-F). In Azimuth, we also use reference data created based on markers (<https://azimuth.hubmapconsortium.org>). However, since it was written only in Methods not written in Results, we have now added a description in Results in the revised manuscript (Lines 143-145).

Lines 143-145: "We characterized the cellular composition of PBMCs from each mammalian species by annotating the cell type of individual single cells using tools available in Seurat [23, 24] and Azimuth [25] (See Methods)."

Anonymous Reviewer: I find after reading the MS that under my normal reviewer conditions I would have no choice but to reject the manuscript.

The structure and design of the MS is ok normally however the actual data has problems in that there is only n=1 sample. For each species (n=4) and each condition (n=4) there are indeed n=16 lots of single cell data. This however means there isn't even a single repeat. There is so much variation within bats, let alone between treatments or between species that simply n=1 for any scientific experiment is not acceptable. Further there are 40,717 cells detected between 16 separate experiments leaving a very low number of cells.

If you look at the raw data for fig S1D, you can see the cell number for bats is incredibly low.

Rousettus\_aeegyptiacusMockB1549  
Rousettus\_aeegyptiacusMockNaiveT541  
Rousettus\_aeegyptiacusMockKillerTNK1960  
Rousettus\_aeegyptiacusMockMono1035  
Rousettus\_aeegyptiacusMockcDC57  
Rousettus\_aeegyptiacusMockpDC10  
Rousettus\_aeegyptiacusHSV1B767  
Rousettus\_aeegyptiacusHSV1NaiveT166  
Rousettus\_aeegyptiacusHSV1KillerTNK347  
Rousettus\_aeegyptiacusHSV1Mono71  
Rousettus\_aeegyptiacusHSV1cDC3  
Rousettus\_aeegyptiacusHSV1pDC2  
Rousettus\_aeegyptiacusSeVB1160  
Rousettus\_aeegyptiacusSeVNaiveT48  
Rousettus\_aeegyptiacusSeVKillerTNK116  
Rousettus\_aeegyptiacusSeVMono137

Gigascience prides itself on high-quality datasets but this set is way too minimal. I feel the data quality is well below GigaScience's level. Its a real pity as the manuscript is otherwise written and presented well - with a logical flow and decent structure for analysing the data.

Regarding biological replicates:

We agree with the importance of biological replicates, but it is cost-prohibitive to perform scRNA-seq for 16 samples with replicates. For clarity, we have now added the fact that we did not have biological replicates in the Limitations of the study section (Line 410).

Line 410: "Furthermore, we did not perform biological replicates of scRNA-seq in this study."

Regarding your point about low cell numbers:

The average cell numbers were about 2,500 cells/run after quality control (QC). Technically, it is possible to obtain higher cell numbers (e.g. more than 10,000 cells/run) before QC, but the higher the value, the higher the percentage of cell multiplets. That's why in this study, we targeted a concentration preparation of 5,000 cells/run before QC. Furthermore, since this study used stimulation by viral infection and cells from animals other than humans, it was expected that the value would be lower than the theoretical value. Considering these factors, we believe that the average of 2500 cells/run after QC, which is 1/2 of the target value before QC, is sufficiently high. Based on these considerations and your feedback, we have now added descriptions of conditions (cell concentration and number of target cells) to Methods in the revised manuscript (Lines 557-558).

Lines 557-558: "Before loading, cell numbers and viability were confirmed. To acquire 5,000 cells recovery, 8,000 cells were loaded."

There are problems with the methods aswell when the authors made certain decisions without any explanation or discussion (or reference for similar) e.g. "Cells with 800–5,000 genes/cell or 1,200–25,000 counts/cell were extracted." "Second, we removed nontargeted cells in the present study." "Finally, regarding genes/cell and counts/cell values, cells with  $>3$  |Z score| were excluded.". This could then bias the data (and indeed seems to whereby the majority of lower-count cells for SeV are removed before and after QC - this obviously suggests Sev-infection reduces gene transcription in cells but these cells are filtered out.

Thank you for pointing this out. The cell filtering pipeline and its threshold have been set to ensure valid comparisons. As you mentioned, there was a group of cells with significantly low genes/cell and counts/cell values in Bat SeV. One possible interpretation of this could be that SeV infection may suppress the gene expression of these cells. However, it is reasonable to exclude these cells in which lower number of genes are detected from the analysis because it would be problematic to perform the analysis if these cells are included in the downstream analysis. Thus, we added the explanation why these cells were excluded in Results (Lines 133-139) and Methods (Lines 637-649) in the revised manuscript. Next, the cell types excluded from the analysis are erythrocytes, platelets, hematopoietic stem cells, and innate lymphoid

cells. For erythrocytes and platelets, we removed them because these cell types are not members of PBMCs. Hematopoietic stem cells and innate lymphoid cells were excluded because they are not major cell types when discussing innate immune responses in PBMCs. We have also now added an explanation about this in Methods in the revised manuscript (Lines 637-649).

Lines 133-139: "Next, quality control (QC) was performed to exclude both cells with lower data quality and cells not targeted in this study (Fig. S1D–G) (See Methods). Before QC, there was a group of cells with low genes-per-cell and counts-per-cell in PBMCs of SeV-infected bats (Fig. S1D–E). Although one possible interpretation of this could be that SeV infection may suppressed the gene expression in these cells, these cells were excluded to ensure the integrity of the downstream quantitative analysis."

Lines 637-649: "The thresholds were determined based on the distributions of genes/cell and counts/cell before QC (Fig. S1D–E). Second, we removed nontargeted cells in the present study. We annotated the cell type of individual cells using Azimuth (v0.4.3) [25], a reference-based cell annotation prediction program (<https://azimuth.hubmapconsortium.org>), and then, cells annotated as erythrocytes, platelets, hematopoietic stem cells, or innate lymphoid cells, and platelets were excluded as nontargeted cells in the present study. This is because erythrocytes and platelets are probably residuals after experimental PBMC extraction, and hematopoietic stem cells and innate lymphoid cells are not the major cell types in the analysis of innate immune responses using PBMCs. In this step, the gene annotation "genes shared with humans" (see Gene annotation and ortholog information) for each animal species was used. Finally, regarding genes/cell and counts/cell values, cells with  $>3$  |Z score| were excluded as outliers."

The very limited dataset then also has different methods for PBMC extraction applied to different species (particularly whereby the anaesthesia used in chimpanzee is known to affect profile of PBMCs). Yet these cells are all dumped together for comparisons of human only or bat only genes.

As you pointed out, the method of obtaining PBMCs differs from animal to animal. However, this is in accordance with the characteristics of the animals, so it is unavoidable from an ethical point of view. Also, the main analysis is about the differences between bats and mammals and does not mention chimpanzee-specific characteristics. Furthermore, the results that chimpanzee cells showed similar expression patterns to those of humans and different from those of macaques and bats are consistent with the fact that chimpanzees are evolutionally close to humans. The results also support the idea that anesthesia did not have a significant effect. In conclusion, we believe that the method of PBMC acquisition in this study was appropriate.

There are other minor problems whereby the count normalization (post-transformation) is merged for all 4 species but after that differen settings (by species) a used in FindClusters (the idea of normalization should overcome this and allow similar settings to be used).

Because the hyperparameters of FindClusters are particularly sensitive to cell number, we used the hyperparameters that were adjusted for the cell number in each animal in this study.

They also perform statistics via FIshe's exact test (despite  $n=1$ ) for GO analysis in a rather unusual and creative way for GO of X against all others (rather than mock vs treatment) due to the lack of replicates.

In this study, we used an overlap-based (over-representation-based) GO enrichment analysis, which tests whether the list of genes with a specific GO term are significantly overlapped with the list of genes of interest using Fisher's exact test. In our knowledge, this is a widely used method of GO enrichment analysis as described in Zhao and Rhee et al. (<https://doi.org/10.1016/j.tig.2023.01.003>). In Fig. 2E, GO term enrichment analysis was performed to check which GO terms are enriched in each gene list (ALL\_high, Virus\_high, TNK\_low, etc.). In Fig. 4F and 4H, GO term enrichment analysis was performed on the list of DEGs in RaC5 and RaC7. Given the nature of GO term enrichment analysis, the issue of biological replicates is irrelevant because the number of genes (not the number of samples) is considered by Fisher's exact test in this context.



Additional Information:	
Question	Response
Are you submitting this manuscript to a special series or article collection?	No
<p><b>Experimental design and statistics</b></p> <p>Full details of the experimental design and statistical methods used should be given in the Methods section, as detailed in our <a href="#">Minimum Standards Reporting Checklist</a>. Information essential to interpreting the data presented should be made available in the figure legends.</p> <p>Have you included all the information requested in your manuscript?</p>	Yes
<p><b>Resources</b></p> <p>A description of all resources used, including antibodies, cell lines, animals and software tools, with enough information to allow them to be uniquely identified, should be included in the Methods section. Authors are strongly encouraged to cite <a href="#">Research Resource Identifiers</a> (RRIDs) for antibodies, model organisms and tools, where possible.</p> <p>Have you included the information requested as detailed in our <a href="#">Minimum Standards Reporting Checklist</a>?</p>	Yes
<p><b>Availability of data and materials</b></p> <p>All datasets and code on which the conclusions of the paper rely must be either included in your submission or deposited in <a href="#">publicly available repositories</a> (where available and ethically appropriate), referencing such data using a unique identifier in the references and in the “Availability of Data and Materials” section of your manuscript.</p>	Yes

Have you have met the above requirement as detailed in our [Minimum Standards Reporting Checklist](#)?

1 **Single-cell transcriptome analysis illuminating the characteristics of species-**  
2 **specific innate immune responses against viral infections**

3

4 Hirofumi Aso<sup>1,2,3</sup>, Jumpei Ito<sup>1</sup>, Haruka Ozaki<sup>4,5</sup>, Yukie Kashima<sup>6</sup>, Yutaka Suzuki<sup>6</sup>,  
5 Yoshio Koyanagi<sup>2,3</sup>, and Kei Sato<sup>1,7,8,9,10,11,12,13,\*</sup>

6

7 1. Division of Systems Virology, Department of Microbiology and Immunology, The  
8 Institute of Medical Science, The University of Tokyo, Tokyo 1088639, Japan

9 2. Institute for Life and Medical Sciences, Kyoto University, Kyoto 6068507, Japan

10 3. Graduate School of Pharmaceutical Sciences, Kyoto University, Kyoto 6068501,  
11 Japan

12 4. Bioinformatics Laboratory, Faculty of Medicine, University of Tsukuba, Tsukuba  
13 3050821, Japan

14 5. Center for Artificial Intelligence Research, University of Tsukuba, Tsukuba  
15 3058577, Japan

16 6. Laboratory of Systems Genomics, Graduate School of Frontier Sciences, The  
17 University of Tokyo, Kashiwa 2778561, Japan

18 7. International Research Center for Infectious Diseases, The Institute of Medical  
19 Science, The University of Tokyo, Tokyo 1088639, Japan

20 8. International Vaccine Design Center, The Institute of Medical Science, The  
21 University of Tokyo, Tokyo 1088639, Japan

22 9. Graduate School of Medicine, The University of Tokyo, Tokyo 1130033, Japan

23 10. Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa  
24 2778561, Japan

25 11. Collaboration Unit for Infection, Joint Research Center for Human Retrovirus  
26 infection, Kumamoto University, Kumamoto, Japan

27 12. CREST, Japan Science and Technology Agency, Kawaguchi 3320012, Japan

28 13. Twitter: @SystemsVirology

29 \*Correspondence address. Kei Sato, Division of Systems Virology, Department of  
30 Microbiology and Immunology, The Institute of Medical Science, The University of  
31 Tokyo, Tokyo 1088639, Japan. E-mail: [keisato@g.ecc.u-tokyo.ac.jp](mailto:keisato@g.ecc.u-tokyo.ac.jp)

32

33 Hirofumi Aso [0000-0002-2826-3836];

34 Jumpei Ito [0000-0003-0440-8321];

35 Haruka Ozaki [0000-0002-1606-2762];

36 Yutaka Suzuki [0000-0003-4852-1879];

37 Yoshio Koyanagi [0000-0002-3007-6642];

38 Kei Sato [0000-0003-4431-1380]

39 **Abstract**

40

41 **Background**

42 Bats harbor various viruses without severe symptoms and act as their natural  
43 reservoirs. The tolerance of bats against viral infections is assumed to originate from  
44 the uniqueness of their immune system. However, how immune responses vary  
45 between primates and bats remains unclear. Here, we characterized differences in the  
46 immune responses by peripheral blood mononuclear cells to various pathogenic  
47 stimuli between primates (humans, chimpanzees, and macaques) and bats (Egyptian  
48 fruit bats) using single-cell RNA sequencing.

49

50 **Results**

51 We show that the induction patterns of key cytosolic DNA/RNA sensors and antiviral  
52 genes differed between primates and bats. A novel subset of monocytes induced by  
53 pathogenic stimuli specifically in bats was identified. Furthermore, bats robustly  
54 respond to DNA virus infection even though major DNA sensors are dampened in bats.

55

56 **Conclusions**

57 Overall, our data suggest that immune responses are substantially different between  
58 primates and bats, presumably underlying the difference in viral pathogenicity among  
59 the mammalian species tested.

60

61 **Keywords:** innate immunity; mammal; virus infection; single-cell RNA-sequencing;  
62 tensor

## 63 **Background**

64 Although a virus can infect various animal species, the pathogenicity of the infection  
65 can differ among host species. For example, Old World monkeys, including rhesus  
66 macaques (*Macaca mulatta*), are naturally infected with Cercopithecine herpesvirus 1  
67 (also known as B virus) without any observable disorders, while humans (*Homo*  
68 *sapiens*) exhibit severe disorders after infection [1]. Bat species are naturally infected  
69 with a variety of viruses and behave as natural reservoirs of human pathogenic viruses  
70 [2]. For example, Marburg virus infection causes severe symptoms in humans but not  
71 in Egyptian fruit bats (*Rousettus aegyptiacus*), a putative natural host of this virus [3].  
72 One possible factor that could define the differences in viral pathogenicity among host  
73 species is the difference in innate immune responses. For example, a previous study  
74 reported that Egyptian fruit bats lack the induction of proinflammatory cytokines,  
75 including *CCL8*, *FAS*, and *IL6*, which are related to disease severity in humans, upon  
76 Marburg virus infection, suggesting that the lack of cytokine induction is one of the  
77 reasons why Egyptian fruit bats exhibit asymptomatic infection with Marburg virus [4].

78 Pathogen sensing is the initial step in triggering innate immune signaling. In a  
79 broad range of animals, including vertebrates, pathogen-associated molecular  
80 patterns (PAMPs) are recognized by pattern recognition receptors (PRRs) to induce  
81 subsequent immune responses [5–8]. In humans and mice (*Mus musculus*), double-  
82 stranded RNAs (dsRNAs), a PAMP for RNA viruses, are recognized by RNA sensors,  
83 such as RIG-I, MDA5, LGP2, TLR3, and TLR7/8 [5, 6]. Extrachromosomal DNAs, a  
84 PAMP for DNA viruses, are recognized by cytosolic DNA sensors (e.g., cGAS, AIM2,  
85 and IFI16) and endosomal DNA sensors (e.g., TLR9) [5, 6, 9]. Lipopolysaccharide  
86 (LPS), a PAMP for bacteria, is recognized by TLR4 [5, 6, 10]. Once PAMPs are  
87 recognized by PRRs, type I interferons (IFNs) are produced, leading to the induction  
88 of IFN-stimulated genes (ISGs), which include many antiviral genes [5, 6].

89 In contrast to the similarities in the immune system between humans and mice,  
90 the immune system of bats is assumed to be quite different from that of humans in  
91 various aspects [11–13]. Genome analysis of Egyptian fruit bats showed expansion  
92 and diversification of immune-related genes, including type I IFN genes [14].  
93 Transcriptome analysis showed that type I IFNs in the Australian black flying fox  
94 (*Pteropus alecto*) are constitutively expressed in unstimulated tissues, leading to the  
95 constitutive expression of ISGs [15]. These observations suggest that immunity in bats  
96 may be stronger than that in other mammals. In contrast, some studies have proposed

97 that immune responses in bats are dampened, resulting in bats exhibiting stronger  
98 tolerance to various viruses [12, 14, 16]. In particular, it is known that critical molecules  
99 involved in viral DNA sensing, such as cGAS, AIM2, and IFI16, are dampened or  
100 genetically lost in some bat species, including Egyptian fruit bats [16, 17]. These  
101 differences in innate immunity between humans and bats could be one of the reasons  
102 why viral pathogenicity differs between these two mammals.

103 Previous works have highlighted the uniqueness of the bat immune system  
104 using genomic analysis [14, 15, 17], transcriptome analysis [4, 18–20], and molecular  
105 biological experiments that reconstituted a part of the bat immune system in cell  
106 culture systems [16, 21, 22]. However, it remains unclear how and to what extent the  
107 innate immune response to pathogenic stimuli varies among mammals. Particularly, it  
108 is unclear how different innate immune responses are elicited by viral infections in  
109 different cell types in each mammal. Here, we used peripheral blood mononuclear  
110 cells (PBMCs) from four mammalian species including the above-mentioned Egyptian  
111 fruit bats and three pathogenic stimuli and conducted single-cell RNA sequencing  
112 (scRNA-seq) analysis to elucidate the differences in innate immune responses against  
113 pathogenic stimuli.

114

## 115 **Results**

116

### 117 **Experimental design**

118 To illuminate the differences in immune responses to infectious pathogens among  
119 mammalian species, we isolated PBMCs from four mammals including humans (*Homo*  
120 *sapiens*, Hs), chimpanzees (*Pan troglodytes*, Pt), rhesus macaques (*Macaca mulatta*,  
121 Mm), and Egyptian fruit bats (*Rousettus aegyptiacus*, Ra) (**Fig. 1A**). In this study, the  
122 Egyptian fruit bat was used as a representative model organism for bat species  
123 because this bat species is bred and available in captivity and is known to be a natural  
124 host of human pathogenic viruses, such as Marburg virus [3]. These PBMCs were  
125 inoculated with herpes simplex virus type 1 (HSV-1; a DNA virus), Sendai virus (SeV;  
126 an RNA virus), or lipopolysaccharide (LPS; a proxy for bacterial infection). We verified  
127 that these PBMCs could be infected with and/or respond to these viruses and LPS  
128 stimulation by quantifying viral RNAs and the upregulation of proinflammatory  
129 cytokines (e.g., IL1B and IL6), ISGs (e.g., EIF2AK2 and DDX58) and IFNB1 (**Fig.**  
130 **S1A–C**) at the level of mRNA transcripts.

131 To analyze immune responses to stimuli at single-cell resolution, we performed  
132 scRNA-seq analysis of 16 types of PBMC samples: four mammalian species (Hs, Pt,  
133 Mm, and Ra) versus four conditions (mock infection/stimulation, HSV-1 infection, SeV  
134 infection, and LPS stimulation) using the 10x Genomics Chromium platform at one day  
135 post-infection. Next, quality control (QC) was performed to exclude both cells with  
136 lower data quality and cells not targeted in this study (**Fig. S1D–G**) (See **Methods**).  
137 Before QC, there was a group of cells with low genes-per-cell and counts-per-cell in  
138 PBMCs of SeV-infected bats (**Fig. S1D–E**). Although one possible interpretation of  
139 this could be that SeV infection may have suppressed the gene expression in these  
140 cells, these cells were excluded to ensure the integrity of the downstream quantitative  
141 analysis. After filtering low-quality cells, a total of 40,717 cells from the 16 samples  
142 were used in the following analysis.

143

#### 144 **The cellular composition of PBMCs from primates and bats**

145 We characterized the cellular composition of PBMCs from each mammalian species  
146 by annotating the cell type of individual single cells using tools available in Seurat [23,  
147 24] and Azimuth [25] (See **Methods**). To establish a common classification system for  
148 the cells from the different mammalian species, we first identified cell types present in  
149 multiple species (**Fig. 1B and 1C**). As cell types detected in multiple species, naïve B  
150 cells, non-naïve B cells (including memory B cells and intermediate B cells), naïve  
151 CD4<sup>+</sup> T cells, non-naïve CD4<sup>+</sup> T cells (including central memory CD4<sup>+</sup> T cells, effector  
152 memory CD4<sup>+</sup> T cells, proliferating CD4<sup>+</sup> T cells, and regulatory T cells), naïve CD8<sup>+</sup>  
153 T cells, non-naïve CD8<sup>+</sup> T cells (including central memory CD8<sup>+</sup> T cells, effector  
154 memory CD8<sup>+</sup> T cells, and proliferating CD8<sup>+</sup> T cells), natural killer (NK) cells,  
155 mucosal-associated invariant T cells (MAITs), monocytes (Monos), conventional  
156 dendritic cells (cDCs), and plasmacytoid DCs (pDCs) were identified (**Fig. 1C**). Known  
157 marker genes for each cell type in humans were detected in the corresponding cell  
158 type in the unstimulated samples from the other animal species (**Fig. S2G**). Although  
159 most cell types were detected in all four species investigated, naïve CD8<sup>+</sup> T cells and  
160 MAITs were undetectable in bat PBMCs, presumably because the cell numbers of  
161 these populations were relatively low in bats and/or the transcriptomic signatures of  
162 naïve CD4<sup>+</sup> T cells and non-naïve CD8<sup>+</sup> T cells were too similar in bats (hereafter we  
163 simply referred to Egyptian fruit bats as “bats”) (**Fig. 1C**). This result was consistent  
164 with a previous study, in which clear clusters of naïve CD8<sup>+</sup> T cells and MAITs were

165 not detected [26]. To establish a cellular classification system for the comparative  
166 transcriptome analysis, we defined six species-common cell types, namely, B cells,  
167 naïve T cells, killer TNK cells, Monos, cDCs, and pDCs, according to similarities in  
168 expression patterns (**Fig. S2H**).

169 The ratio of the six cell types exhibited different changes upon exposure to the  
170 stimuli in the different species (**Fig. 1D**). The frequency of monocytes decreased after  
171 stimulation in all four species, whereas the frequency of B cells and killer TNK cells  
172 changed differently within and across the animal species. Upon stimulation, there was  
173 generally a notable increase in B cells and a decrease of killer TNK cells in the bat  
174 (and non-human primates) samples, but not in the human samples.

175

### 176 **Immune response differs largely among animal species**

177 To describe the differences in immune responses to various stimuli in specific cell  
178 types among animal species, we first calculated the average expression levels of  
179 appropriate genes in each condition (4 animal species × 4 stimuli × 6 cell types = 96  
180 conditions). Using this “pseudobulk” transcriptome dataset, we first investigated which  
181 axis (i.e., animal species, stimulus, and cell type) was the most impactful element in  
182 shaping the expression patterns of immune cells. Thereby, we calculated the fold-  
183 change (FC) values of gene expression levels between unstimulated and  
184 corresponding stimulated conditions and performed principal component analysis  
185 (PCA) on the FC values. Hierarchical clustering analysis was subsequently performed  
186 according to principal components (PCs) 1–30. The transcriptome data branch  
187 according to the animal species and then branch according to the cell type followed  
188 finally by the stimulus (**Fig. 1E**). This suggests the difference in host species is the  
189 more impactful element in shaping the immune system, having a greater impact than  
190 the type of stimulus and cell type. In particular, bat PBMCs exhibited different  
191 transcriptomic patterns irrespective of the type of stimulus and cell type compared to  
192 the PBMCs from the other three species used. These results suggest that bats  
193 respond to pathogens in a different manner than primates.

194

### 195 **Extraction of species-specific immune responses**

196 We next characterized the differences in the immune responses to pathogenic stimuli  
197 among animal species. The FC values of our pseudobulk transcriptome dataset were  
198 represented by a four-mode tensor (4 animal species × 3 stimuli × 6 cell types × 7557



199 orthologous genes). To characterize this extraordinary high-dimensionality  
200 transcriptome dataset, we utilized Tucker decomposition, a method of tensor  
201 decomposition (**Fig. 2A**). In this analysis, we excluded cDC and pDC data due to many  
202 missing values. Tucker decomposition generated a core tensor and four-factor  
203 matrices (A1–A4) related to the four axes (animal species, stimulus, cell type, and  
204 gene). For example, the factor matrix A1 (for host species) included three latent factors  
205 (L1\_1, L1\_2, and L1\_3), which could be regarded to represent common, bat-specific,  
206 and macaque-specific expression patterns, respectively (**Fig. 2B**).

207 To characterize species-specific immune responses, we developed a gene  
208 classification system according to the pattern of the species-associated latent factor  
209 in the tensor decomposition framework. First, we calculated the product of a core  
210 tensor and the three-factor matrices A2 (for stimulus), A3 (for cell type), and A4 (for  
211 gene) (**Fig. 2C and Fig. S3A–B**). Consequently, we obtained three cubic datasets  
212 with three axes—stimulus, cell type, and gene. These cubic data were related to L1\_1  
213 (for the common factor), L1\_2 (for the bat-specific factor), or L1\_3 (for the macaque-  
214 specific factor). Subsequently, we classified the genes into 10 categories according to  
215 their expression patterns in each cubic dataset (the results for the bat-specific (L1\_2)  
216 and other factors (L1\_1 and L1\_3) are shown in **Fig. 2D, Fig. S3G, and Fig. S3I**,  
217 respectively). In the factor matrix A2 (for stimulus), the values for the latent factors  
218 related to HSV-1 and SeV were similar (**Fig. S3A**). Therefore, these two categories  
219 were integrated into the category “Virus” in the gene classification. Additionally, two  
220 cell type categories, NaiveT and KillerTNK, were integrated into the category “TNK”  
221 (**Fig. S3B**). The pattern for raw FC values supported that the gene classification by  
222 the tensor decomposition framework succeeded in extracting the characteristic  
223 patterns of gene expression alterations upon pathogenic stimuli (**Fig. S3J–L**).

224

### 225 **Differential dynamics of pathogen sensing and immune responses**

226 To highlight the uniqueness of immunity in bats compared to that in primates, we  
227 focused on the expression pattern represented by the bat-specific factor (L1\_2) and  
228 performed Gene Ontology (GO) analysis on the 10 gene categories (**Fig. 2E**). In the  
229 gene category “ALL\_high”, which included genes upregulated particularly in bats  
230 regardless of the stimulus and cell type, GO terms related to innate immune responses,  
231 such as IFN signaling, DDX58/IFIH1-mediated induction of IFN, RIG-I like receptors

232 (RLRs) signaling pathways, and the antiviral mechanism by ISGs, were over-  
233 represented.

234 To dissect the “ALL\_high” genes in the bat-specific factor, we further extracted  
235 the genes that belonged not only to the “ALL\_high” category in the bat-specific factor  
236 but also to that in the common factor (L1\_1). This fraction represented genes that were  
237 upregulated by stimuli in all species but whose induction levels were highest in bats.  
238 These genes included various PPRs, such as RIG-I-like receptors (RLRs) (RIG-I,  
239 LGP2, and MDA5) and cGAS, a DNA sensor, suggesting that these genes were  
240 upregulated to higher levels in bats than in the other species across the cell types and  
241 stimuli (**Fig. 2F**). There are two possible scenarios that could potentially explain these  
242 higher FC values observed in bats. One possibility is that expression levels of these  
243 genes after stimulation are higher than in primates. The second possibility is that basal  
244 expression levels of these genes in bats are lower than those in primates. Therefore,  
245 we calculated the relative expression levels of these genes in bats compared to  
246 humans and showed that the basal expression levels of these genes were lower in  
247 bats than in humans (**Fig. 2G**). These results suggest that the induction dynamics of  
248 these PRRs in bats are likely different from those in primates, possibly leading to the  
249 differences in the induction of immune responses.

250

### 251 **Robust immune responses to a DNA virus in bats**

252 As critical DNA sensors, such as cGAS, AIM2, IFI16, and TLR9, are dampened or  
253 genetically lost in bat species [16, 17, 27], it has been hypothesized that bats, including  
254 Egyptian fruit bats, cannot efficiently activate innate immune responses against DNA  
255 viruses. To test this hypothesis, we analyzed the IFN response upon HSV-1 (a DNA  
256 virus) infection. However, the expression levels of IFN- $\alpha$  genes were not examined  
257 because they were not annotated in the transcript model for the Egyptian fruit bat used  
258 in this study, and the expression of IFN- $\beta$  genes were too low. Thus, even though the  
259 expression levels of IFN-I is the primary factor to examine the activity of the IFN  
260 response, we instead analyzed the induced levels of “core<sup>mamm</sup> ISGs”—a set of genes  
261 that are commonly induced by type I IFNs across mammals that were defined in a  
262 previous study [28]. Intriguingly, we found that the core<sup>mamm</sup> ISGs were upregulated  
263 upon HSV-1 infection in most cell types in bats (**Fig. 3A**). The induced levels were  
264 comparable to those induced by SeV (an RNA virus) infection and higher than those  
265 induced by LPS stimulation. Furthermore, the induced levels in bats were comparable

266 to those in primates. This suggests that immune cells in bats can sense and respond  
267 to HSV-1 infection even though critical DNA sensors are dampened.

268 To address the possibility that pathogen sensors other than DNA sensors  
269 contribute to the sensing of HSV-1 infection in bats, we examined the expression  
270 levels of various PRRs (**Fig. 3B**). The expression of some PRRs, including TLR3, a  
271 dsRNA sensor associated with HSV-1 sensing in humans and mice [29], was detected  
272 not only in primates but also in bats, suggesting the possibility that these PRRs  
273 compensate in the response to HSV-1 infection in bats (see **Discussion**).

274

275

### 276 **Identification of bat-specific subsets of monocytes**

277 Next, we investigated cellular subsets within the cell types that are characteristic in  
278 bats to explain the differences in immune responses among the species. We  
279 particularly searched for cellular subsets that specifically appeared after pathogenic  
280 stimulus exposure in each species according to the dimensionality reduction analysis  
281 of transcriptome data. In humans, chimpanzees, and macaques, no subset appeared  
282 in any cell type after stimulation (**Fig. S4A**). Similarly, such subsets were not identified  
283 in T/NK or B cells in bats. In contrast, we found that two subsets of bat monocytes  
284 (referred to as Clusters 5 and 7) specifically appeared after stimulation (**Fig. 4A**). To  
285 validate whether these subsets (Clusters 5 and 7) are unique in bats, we identified  
286 marker genes for these clusters and subsequently examined whether the marker  
287 genes were expressed in monocytes from the other animal species. The marker genes  
288 for Cluster 5 (referred to as C5 markers) were not highly expressed in any cluster of  
289 monocytes from primates (**Fig. 4B**). Furthermore, high expression levels of C5  
290 markers in bat monocytes were found only after stimulation. This suggested that  
291 Cluster 5 was not only bat-specific but also specifically induced by pathogenic stimuli.  
292 Unlike the C5 markers, the marker genes for Cluster 7 (C7 markers) were highly  
293 expressed not only in bat Cluster 7 but also in some monocytes in primates (**Fig. 4C**).  
294 Although cells with higher expression of C7 markers were induced upon stimulation in  
295 both bats and primates, these cells in primates did not form a separate cluster similar  
296 to Cluster 7 in bats (**Fig. S4B**). Furthermore, the proportions of Clusters 5 and 7  
297 differed depending on the stimulus: HSV-1-infected and LPS-stimulated samples  
298 showed the highest frequencies of Clusters 5 and 7, respectively (**Fig. 4D**).

299 To characterize these two clusters, we identified differentially expressed genes  
300 (DEGs) in Clusters 5 and 7 compared to the other clusters of bat monocytes.  
301 According to GO analysis, Cluster 5 is characterized by lower expression of ISGs (**Fig.**  
302 **4E, 4F**). Additionally, Cluster 5 highly expresses known suppressors of the  
303 inflammatory response, such as DUSP1, DUSP5, and SOCS2 [30–32]. On the other  
304 hand, Cluster 7 can be characterized by a higher expression of various cytokines  
305 related to chemotaxis (**Fig. 4G**), including CXCL6, IL18BP, CXCL8, CCL2, CCL8,  
306 CCL13, CCL5, CXCL10, IL15, and IL411 (MSigDB [33]: GO:0060326) (**Fig. 4G, 4H**).  
307 Overall, we established that there are two unique subsets of bat monocytes with  
308 different characteristics (see **Discussion**).

## 309 Discussion

310 Differences in viral pathogenicity among host species are thought to be attributed to  
311 differences in immune responses against viral infections among the species [34].  
312 However, it remains unclear how immune responses, particularly innate immunity  
313 against viral infections, differ among host species. In the present study, we performed  
314 scRNA-seq on 16 types of PBMC samples, derived from a combination of four host  
315 species and four infection conditions (**Fig. 1A**), and showed that the differences in the  
316 immune responses among the host species were more impactful than those among  
317 both the stimuli and the cell types (**Fig. 1E**). In particular, the transcriptomic changes  
318 resulting from pathogenic stimulation in bats differed from those in primates. It is also  
319 noteworthy that post-stimuli changes in the ratio of cell types differed between humans  
320 and bats (**Fig. 1D**). For further analysis, we established a bioinformatic pipeline to  
321 characterize species-specific immune responses from transcriptome profiles with  
322 extraordinarily high dimensions (4 animal species × 3 stimuli × 4 cell types × 7,557  
323 orthologous genes) (**Fig. 2A**). We illuminate differences in innate immune systems  
324 among mammalian species that partly explain the differences in viral pathogenicity  
325 among host species.

326 It is known that two DNA sensing pathways mediated by cGAS-STING pathway  
327 [16] and PYHIN proteins, including AIM2 and IFI16 [17], are dampened in bats,  
328 including Egyptian fruit bats. In addition, a previous study using a cell line derived from  
329 big brown bats (*Eptesicus fuscus*) suggested that the TLR9-mediated DNA sensing  
330 pathway is also weakened in bats [27]. Based on these observations, it was  
331 hypothesized that the ability to sense DNA virus infection is weakened in bats [12, 13].  
332 However, we showed that bat PBMCs robustly induced IFN responses upon infection  
333 with the DNA virus HSV-1 (**Fig. 3A**). This suggests that bats can initiate an innate  
334 immune response after infection with DNA viruses (at least HSV-1) and that bats have  
335 another pathway to sense DNA viruses. An alternative possibility is that the IFN  
336 response in response to HSV-1 infection was triggered by sensing viral molecules  
337 other than DNAs. It is known that, in humans and mice, dsRNA sensing by TLR3 plays  
338 an important role in responding to HSV-1 infection [29, 35]. Additionally, the Egyptian  
339 fruit bat genome encodes an intact TLR3 gene (NCBI Gene ID: 107510436), and bat  
340 immune cells express TLR3 (**Fig. 3B**). Furthermore, other RNA sensors, such as RIG-  
341 I, LGP2, and MDA5, were upregulated in bat cells similarly as in primate cells upon  
342 HSV-1 infection (**Fig. 3B**). These data suggest that TLR3 or other RNA sensors in bats

343 may compensate for weakened DNA sensing pathways, leading to IFN responses to  
344 HSV-1 infection.

345 To characterize the bat-specific innate immune responses based on ultrahigh-  
346 dimensionality transcriptome data (4 animal species × 4 stimuli × 6 cell types × 7,557  
347 orthologous genes), we established an analytical framework utilizing tensor  
348 deconvolution (**Fig. 2A**). This framework could i) extract a species-specific effect on  
349 gene expression changes, ii) compare the effects among the cell types and the stimuli,  
350 and iii) classify genes according to the differential pattern of a species-specific effect  
351 among the cell types and the stimuli. Using this framework, we found that the  
352 expression levels of key DNA and RNA sensors, including cGAS, RIG-I, MDA5, and  
353 LGP2, were highly induced in bats compared with primates, regardless of the cell type  
354 or stimulus (**Fig. 2F**). Furthermore, the basal expression levels of these PRRs in bats  
355 were lower than those in humans (**Fig. 2G**). On the other hand, after stimulation, the  
356 expression levels of these PRRs in bats were comparable to those in humans. These  
357 results suggest that the induction dynamics of these PRRs in bats are likely different  
358 from those in primates, leading to the differences in the induction of immune responses.  
359 Indeed, several antiviral ISGs, such as IFI6 and IFIT3, exhibited expression dynamics  
360 similar to those of these PRRs (**Fig. 2F, 2G**). These differences could be one of the  
361 reasons why immune responses differ between bats and primates.

362 Another factor that can explain the differences in immune responses among  
363 host species is the presence of species-specific cellular subsets. In bat monocytes,  
364 we identified two subsets that were specifically induced by stimuli (i.e., Clusters 5 and  
365 7) (**Fig. 4A**). Cluster 5 was a bat-specific subset induced preferentially by HSV-1  
366 infection (**Fig. 4B, 4D**). Interestingly, even though Cluster 5 was induced after  
367 stimulation, Cluster 5 exhibited lower expression of ISGs and higher expression of  
368 immunosuppressive genes (DUSP1, DUSP5, and SOCS2) [30–32] (**Fig. 4E, 4F**). This  
369 observation suggests that the immune responses in Cluster 5 are downregulated  
370 presumably by negative feedback signaling and that Cluster 5 may contribute to  
371 controlling excessive immune activation in bats. On the other hand, Cluster 7 was  
372 identified as a monocyte subset that was mainly induced by LPS stimulation (**Fig. 4C,**  
373 **4D**). Cluster 7 highly expressed several proinflammatory cytokines and chemokines  
374 (CXCL6, IL18BP, CXCL8, CCL2, CCL8, CCL13, CCL5, CXCL10, IL15, and IL4I1) (**Fig.**  
375 **4G, 4H**). Cluster 7 may contribute to the recruitment of leukocytes since these  
376 cytokines are associated with the chemotaxis of neutrophils (CCL8, CXCL6, and

377 CXCL8), basophils (CXCL8, CCL2, CCL5, CCL8, and CCL13), eosinophils (CCL5,  
378 CCL8, and CCL13), monocytes (CCL5, CCL8, and CCL13), T cells (CCL5, CCL8,  
379 CCL13, CXCL8, and CXCL10), and NK cells (CCL5 and CCL8) in humans and mice  
380 [36, 37]. Based on the expression pattern of the marker genes for Cluster 7 (**Fig. 4C,**  
381 **S4B**), cellular subsets corresponding to Cluster 7 were also present in primate  
382 monocytes. However, these primate cells did not form a separate cluster in the  
383 dimensionality reduction analysis based on the transcriptome profile (**Fig. 4A**). These  
384 results suggest that the monocyte subset represented by Cluster 7 exhibits unique  
385 gene expression and thus may exert unique functions in bats. Although the specific  
386 functions of these monocyte subsets (Clusters 5 and 7) in immune responses in bats  
387 are still unclear, these unique subsets may contribute to bat-specific host immune  
388 responses.

389

390

#### 391 **Limitations of the study**

392 In the present study, we elucidated differences in innate immune responses among  
393 host species from various aspects. However, we did not address differences in the  
394 outcomes of the innate immune responses, such as differences in viral pathogenicity.  
395 Another limitation is that the bioinformatic resources we used, such as gene annotation,  
396 gene ontology, and cellular annotation, have been developed in a human-centric way.  
397 Therefore, there is the possibility that immune responses induced by species-specific  
398 genes and cell types were overlooked. Moreover, because the results of this study rely  
399 on an analysis using a single bat species, the Egyptian fruit bat, it is unclear whether  
400 the observed bat-specific characteristics are conserved across bat species.  
401 Furthermore, we did not perform biological replicates of scRNA-seq in this study.  
402 Despite these limitations, we present valuable resources to illuminate differences in  
403 immune responses among host species, including Egyptian fruit bats, and clues to  
404 elucidate differences in viral pathogenicity among species. Further study to elucidate  
405 the functional consequences of these differences is needed to reveal the mechanisms  
406 by which bats can tolerate infections with various viruses.

407

408 **Figure legends**

409

410 **Figure 1. scRNA-seq analysis of PBMCs from four animal species inoculated**  
411 **with pathogenic stimuli**

412 (A) Schematic of the experimental design. See also **Fig. S1**.

413 (B) Uniform manifold approximation and projection (UMAP) plots representing the  
414 gene expression patterns of the cells from the four species. Each dot is colored  
415 according to the cell type. Gray dots indicate cells unassigned into any cell type. See  
416 also **Fig. S2**.

417 (C) Comparison of identified cell types among the species. Dot: detected, question  
418 mark: undetected. The definitions of six species-common cell types are shown on the  
419 right side. See also **Fig. S2H**.

420 (D) The cellular compositions of PBMC samples. The compositions according to the  
421 six common cell types are shown.

422 (E) Hierarchical clustering analysis of 48 pseudobulk datapoints (4 animal species x 3  
423 stimuli x 4 cell types = 48 conditions) based on PC1-30 calculated from the fold-change  
424 values (respective stimulus versus unstimulated) for gene expression.

425

426 **Figure 2. Characterization of species-specific immune responses using a tensor**  
427 **decomposition framework**

428 (A) Tensor decomposition of the fold-change values for pseudobulk transcriptome data.

429 (B) Heatmap representing a latent factor matrix relating to species. Columns indicate  
430 the animal species, and rows indicate the latent factors representing species-common  
431 (L1\_1), bat-specific (L1\_2), and macaque-specific (L1\_3) factors. See also **Fig. S3A–**  
432 **B**.

433 (C) Classification of genes according to the differential patterns of the latent factors  
434 related to species. For each of the species-common (L1\_1), bat-specific (L1\_2), and  
435 macaque-specific (L1\_3) factors, the product of the core tensor and three latent factor  
436 matrices related to stimulus, cell type, and gene was calculated (left), and the genes  
437 were classified into 11 categories according to the binary patterns for each calculated  
438 product (right). See also **Fig. S3C–F**.

439 (D) Heatmap representing the values of the products calculated in **Figure 2C**. From  
440 the three products, the data related to the bat-specific factor (L1\_2) are shown. Each



441 row indicates the respective gene. The color keys shown on the right of the heatmap  
442 indicate gene categories. See also **Fig. S3G–L**.

443 (E) GO terms enriched in each gene category relating to the bat-specific factor. GO  
444 terms with a false discovery rate (FDR)  $\leq 0.1$  and an odds ratio  $\geq 1$  are shown.

445 (F) Heatmap representing the induction levels of ALL\_high genes for the bat-specific  
446 factor. Additional classification according to the gene classification of the species-  
447 common factors is shown to the right of the heatmap. Genes categorized as ALL\_high  
448 in both the species-common factor and the bat-specific factor are shown on the right  
449 side. The colored circle indicates the functional category of the gene.

450 (G) Heatmap representing the relative expression levels (bats versus humans) of the  
451 genes shown in **Figure 2F**.

452

### 453 **Figure 3. Robust immune responses to a DNA virus in bats**

454 (A) Boxplot of the expression levels of core<sup>mamm</sup> ISGs in every single cell. The Y-axis  
455 indicates the global expression level (GSVA score) of the core<sup>mamm</sup> ISGs.

456 (B) Heatmap representing the mean expression levels of sensor genes. The mean  
457 values were calculated without using the information for the stimulus.

458

### 459 **Figure 4. Identification of bat-specific subsets of monocytes**

460 (A) UMAP plots representing the gene expression patterns of monocytes from the four  
461 species. The dots are colored according to the cell cluster defined for each animal  
462 species. See also **Fig. S4A**.

463 (B, C) UMAP plots representing the average expression levels of marker genes for  
464 Cluster 5 [C5markers] (B) and Cluster 7 [C7markers] (C). See also **Fig. S4B**.

465 (D) The cellular composition of bat monocytes. The composition is shown according  
466 to the cluster. The black frame indicates Clusters 5 and 7 in stimulated samples.

467 (E) Heatmap representing the mean expression levels of differentially expressed  
468 genes (DEGs) in Cluster 5 of bat monocytes.

469 (F) Summary of the GO terms enriched in DEGs in Cluster 5. GO terms enriched in  
470 up- and downregulated genes are shown in red and blue, respectively.

471 (G) Heatmap representing the mean expression levels of differentially expressed  
472 genes (DEGs) in Cluster 7 of bat monocytes.

473 (H) Summary of the GO terms enriched in DEGs in Cluster 7. GO terms enriched in  
474 up- and downregulated genes are shown in red and blue, respectively.

475 **Methods**

476 **Cells**

477 Vero cells (obtained from the Laboratory of Bernard Roizman, University of Chicago,  
478 USA)

479 LLC-MK2 cells (rhesus macaque kidney epithelial cells) (CCL-7, ATCC)

480

481 **PBMC collection**

482 Human peripheral blood was obtained from the arm vein. To obtain chimpanzee  
483 peripheral blood, a chimpanzee was anesthetized for a regular health examination.  
484 Anesthesia was induced with intramuscular administration of the combination of 0.012  
485 mg/kg medetomidine (Meiji Seika Pharma Co., Ltd. ), 0.12 mg/kg midazolam (Sand  
486 Co., Ltd. ), and 3.5 mg/kg ketamine (Fujita Pharm, Tokyo) and maintained with  
487 constant rate infusion (4-10 mg/kg/h) of propofol (1% Diprivan, Sand Co., Ltd. ).  
488 Peripheral blood was obtained from the femoral vein. To obtain rhesus macaque  
489 peripheral blood, a rhesus macaque was anesthetized. Anesthesia was induced with  
490 intramuscular administration of 8 mg/kg ketamine followed by deep anesthetization  
491 using an intravenous injection of sodium pentobarbital (30 mg/kg) (Kyoritsu Seiyaku).  
492 Peripheral blood was obtained by cardiac puncture before exsanguination and  
493 perfusion. Bat peripheral blood was obtained from the cephalic vein in the patagium.  
494 PBMCs were isolated from peripheral blood by density gradient centrifugation using  
495 Ficoll-Paque™ Plus (Cytiva, Cat# 17144003).

496

497 **HSV-1 preparation and titration**

498 HSV-1 (strain F; GenBank accession number: GU734771) [38] was prepared as  
499 previously described [29] and kindly provided by Dr. Yasushi Kawaguchi (The Institute  
500 of Medical Science, The University of Toyo, Japan). Briefly, Vero cells were infected  
501 with HSV-1 and the supernatant was collected and used without purification. To titrate  
502 viral infectivity, prepared virus was diluted 10-fold in Medium 199 (Thermo Fisher  
503 Scientific, Cat# 11825015) containing 1% fetal calf serum (FCS) (Nichirei Biosciences,  
504 Cat# 175012), and Vero cells were infected with dilutions of the virus at 37 °C. At one  
505 hour postinfection, the culture medium was replaced with Medium 199 containing 160  
506 µg/ml human γ-globulin (Sigma Aldrich, G4386-25G), and the cells were cultured at  
507 37 °C for 2–3 days. To calculate the viral titer [plaque forming unit (PFU)], the number  
508 of plaques per well was counted.

509

### 510 **SeV preparation and titration**

511 SeV (strain Cantrell, clone cCdi; GenBank accession number: AB855654) was  
512 prepared as previously described [39] and kindly provided by Dr. Takashi Irie  
513 (Hiroshima University, Japan). Briefly, LLC-MK2 cells were infected with SeV and the  
514 supernatant was collected and used without purification. To titrate viral infectivity,  
515 prepared virus was diluted 10-fold in Dulbecco's modified Eagle's medium (DMEM)  
516 (Sigma–Aldrich, Cat# D6046-500ML) containing 10% FCS, and LLC-MK2 cells were  
517 infected with dilutions of the virus at 37 °C. At one hour postinfection, the cells were  
518 washed with PBS and cultured with DMEM containing 10% FCS at 37 °C. At one day  
519 postinfection, the infected cells were fixed with acetone (Nacalai Tesque, Cat# 21914-  
520 03)/methanol (Nacalai Tesque, Cat# 00310-95). To calculate the viral titer [cell  
521 infectious unit (CIU)], the fixed cells were stained with a rabbit anti-SeV polyclonal  
522 antibody [40] as the primary antibody and an Alexa 488-conjugated goat anti-rabbit IgG  
523 antibody (Thermo Fisher Scientific, Cat# A-11008) as the secondary antibody, and the  
524 number of fluorescent foci per well was counted.

525

### 526 **Infection and stimulation**

527 One million PBMCs were maintained in 500 µl RPMI 1640 medium (Sigma–Aldrich,  
528 Cat# R8758-500ML) and infected with HSV-1 or SeV at a multiplicity of infection of  
529 0.1. To mimic microbial infection, LPS (Sigma–Aldrich, Cat# L5024-10MG) was added  
530 at a final concentration of 200 ng/ml. At one day post infection, all types of  
531 infected/stimulated PBMCs were centrifuged, resuspended in PBS, and used for bulk  
532 RT–qPCR and scRNA-seq (see below).

533

### 534 **RT–qPCR**

535 RT–qPCR was performed as previously described [41]. Briefly, cellular RNA was  
536 extracted using the QIAamp RNA Blood Mini Kit (Qiagen, Cat# 52304) and then  
537 treated with an RNase-free DNase set (Qiagen, Cat# 79254). cDNA was synthesized  
538 using SuperScript III reverse transcriptase (Thermo Fisher Scientific, Cat# 18080044)  
539 and random primers (Thermo Fisher Scientific, Cat# 48190011). RT–qPCR was  
540 performed using Power SYBR Green PCR Master Mix (Thermo Fisher Scientific, Cat#  
541 4367659) and the primers listed in **Table S1**. For RT–qPCR, the CFX Connect Real-  
542 Time PCR Detection System (Bio-Rad) was used.

543

#### 544 **Sequencing of scRNA-seq libraries**

545 scRNA-seq libraries were constructed using the Chromium Next GEM Single Cell 3'  
546 Kit according to the manufacturer's instructions (10x Genomics). Briefly, cells, gel  
547 beads, and oil were loaded onto the Chromium platform to generate single-cell gel  
548 beads-in-emulsion (GEMs). Before loading, cell numbers and viability were confirmed.  
549 To acquire 5,000 cells recovery, 8,000 cells were loaded. Barcoded cDNAs were  
550 pooled for amplification, and adaptors and indices for sequencing were added. The  
551 evaluation was conducted using a BioAnalyzer (Agilent Technologies). The libraries  
552 were sequenced with paired-end reads using the Illumina NovaSeq6000 platform  
553 (RRID:SCR\_016387) .

554

#### 555 **Genome sequence dataset**

556 Genome sequences of the animal species including humans (GRCh38.p13, RefSeq  
557 accession: GCF\_000001405.39), chimpanzees (Clint\_PTRv2, RefSeq accession:  
558 GCF\_002880755.1), rhesus macaques (Mmul\_10, RefSeq accession:  
559 GCF\_003339765.1), and Egyptian fruit bats (mRouAeg1.p, RefSeq accession:  
560 GCF\_014176215.1) were obtained from NCBI RefSeq [42]. From the genome  
561 sequences, ALT contig sequences were excluded. The genome sequences of viruses  
562 including HSV-1 (strain: F, accession: GU734771.1) and SeV (strain: Cantell clone  
563 cCdi, accession: AB855654.1) were also obtained from NCBI RefSeq. A custom  
564 reference genome sequence for each animal species was generated by adding the  
565 genome sequences of HSV-1 and SeV to the genome sequence of the animal species.

566

#### 567 **Gene annotation and ortholog information**

568 Gene annotations of humans (GRCh38.p13, Release 109.20200228), chimpanzees  
569 (Clint\_PTRv2, Release 105), rhesus macaques (Mmul\_10, Release 103), and  
570 Egyptian fruit bats (mRouAeg1.p, Release 101) were obtained from NCBI RefSeq.  
571 From the gene annotations, only the records for protein\_coding,  
572 transcribed\_pseudogene, lncRNA, pseudogene, antisense\_RNA,  
573 ncRNA\_pseudogene, V\_segment, V\_segment\_pseudogene, C\_region,  
574 C\_region\_pseudogene, J\_segment, J\_segment\_pseudogene, and D\_segment were  
575 extracted according to the CellRanger tutorial [43]. In addition, to quantify viral RNA  
576 abundance, the records for viruses were added. The whole viral genome was treated

577 as a single exon, and a total of four lines (the positive and negative strands of HSV-1  
578 and SeV) were added.

579 A list of orthologous genes between humans and the other animal species  
580 (chimpanzees, rhesus macaques, and Egyptian fruit bats) was obtained from NCBI on  
581 July 26th, 2021 [44]. From the file, the records for orthologs between humans  
582 (taxonomy ID: 9606) and chimpanzees (taxonomy ID: 9598), rhesus macaques  
583 (taxonomy ID: 9544), or Egyptian fruit bats (taxonomy ID: 9407) were extracted.

584 The ortholog list from NCBI lacked information on some critical immune-related  
585 genes of Egyptian fruit bats, such as CD4 and IRF1. Therefore, we retrieved  
586 information from the Bat1K gene annotation [45, 46] downloaded from UCSC genome  
587 browser [47]: First, we made a custom gene annotation for Egyptian fruit bats by  
588 adding information from the Bat1K gene annotation to the RefSeq gene annotation.  
589 Second, we extracted exons in the Bat1K gene annotation that overlapped with exons  
590 in the RefSeq gene annotation by using the bedtools intersect command with the wao  
591 option (v2.30.0) [48]. In this step, the exons in the Bat1K gene annotation that did not  
592 overlap with the exons in the RefSeq gene annotation were also extracted and added  
593 to custom gene annotations as additional genes. Next, the exons that contained  
594 overlaps and had the same gene name (the same symbol or known to be an ortholog)  
595 were added to custom gene annotations as an alternative splicing variant of the gene.  
596 Then, the remaining overlapping exons were processed by determining which  
597 information (RefSeq or Bat1K) should be used preferentially. The criteria were as  
598 follows: i) genes whose symbols are not prefixed with “LOC” were given priority, ii)  
599 genes whose symbols are included in the human gene list were given priority, and iii)  
600 information from RefSeq was given priority otherwise. According to these criteria, the  
601 annotation with the higher priority (RefSeq or Bat1K) was selected and used in the  
602 custom gene annotation.

603 As a result of the integration of gene annotations, the number of orthologous  
604 genes in the custom gene annotation of bats increased from 16374 to 16903.  
605 Importantly, immune-related genes that were not defined in the RefSeq gene  
606 annotation, such as TLR1, IRF1, and CD4, were added to the custom gene annotation.

607 Considering the orthologous relationships, we prepared three types of gene  
608 sets for each animal species: i) “all genes”, including all genes in the animal species;  
609 ii) “genes shared with humans”, including genes with orthologs in humans; and iii)  
610 “common genes”, genes shared among the four analyzed animal species. Unless

611 otherwise noted, “all genes” were used up to cell annotation, and “common genes”  
612 were used after cell annotation.

613

#### 614 **Processing scRNA-seq data for generating count matrices**

615 Gene expression count matrices for scRNA-Seq data were generated using  
616 Cell Ranger (RRID:SCR\_023221) (v6.0.1) (10x Genomics) [49, 50]. First, we built a  
617 custom reference for each animal species from the custom reference genome  
618 sequence and custom gene annotation using the “cellranger mkref” command.  
619 Subsequently, we generated unique molecular identifier (UMI)-based count matrices  
620 from the raw scRNA-seq data and custom references using the “cellranger count”  
621 command with default settings.

622

#### 623 **Quality control (QC) of scRNA-seq data**

624 First, we removed cells with abnormal genes per cell (genes/cell) and counts per cell  
625 (counts/cell) values using the Seurat package (RRID:SCR\_016341) (v4.0.4) [23, 24]:  
626 Cells with 800–5,000 genes/cell or 1,200–25,000 counts/cell were extracted. The  
627 thresholds were determined based on the distributions of genes/cell and counts/cell  
628 before QC (**Fig. S1D–E**). Second, we annotated the cell type of individual cells using  
629 Azimuth (v0.4.3) [25], a reference-based cell annotation prediction program, and then,  
630 cells annotated as erythrocytes, platelets, hematopoietic stem cells, or innate lymphoid  
631 cells were excluded as nontargeted cells in the present study. This is because  
632 erythrocytes and platelets are probably residuals after experimental PBMC extraction,  
633 and hematopoietic stem cells and innate lymphoid cells are not the major cell types in  
634 the analysis of innate immune responses using PBMCs. In this step, the gene  
635 annotation “genes shared with humans” (see **Gene annotation and ortholog  
636 information**) for each animal species was used. Finally, regarding genes/cell and  
637 counts/cell values, cells with  $>3$  |Z score| were excluded as outliers.

638

#### 639 **Data integration, visualization, and cell clustering**

640 Data integration, visualization, and cell clustering for each animal species were  
641 performed using the Seurat package. In these processes, the expression levels of  
642 HSV-1 and SeV were not used.

643 Data integration is a method merging the gene expression count matrices  
644 obtained from different experimental conditions while removing batch effects. We

645 integrated the count matrices from the four different conditions for each animal species.  
646 In the data integration, SCTransform (RRID:SCR\_022146) (a modeling framework for  
647 the normalization and variance stabilization of molecular count data from scRNA-seq  
648 data) was performed using the SCTransform function for each count matrix. Next, to  
649 extract 2000 genes with higher variance and thus greater information for integration,  
650 the four count matrices were processed using the SelectIntegrationFeatures function.  
651 Next, we used the PrepSCTIntegration function to transform normalized counts into  
652 counts per 10,000 counts in the cell (CP10k). After that, we used the  
653 FindIntegrationAnchors function with the setting Mock as a reference to find  
654 “Integration anchors”. Finally, we integrated the four normalized count matrices using  
655 the IntegrateData function with the option ‘normalization.method=“SCT”’.

656 For visualization, we first performed principal component analysis (PCA) using  
657 the RunPCA function. Then, UMAP (RRID:SCR\_018217) [51] was performed with the  
658 RunUMAP function. In this step, principal components (PC) 1-50 were used, and the  
659 parameter “n.neighbors” was set individually for each animal species (Hs: 20, Pt: 20,  
660 Mm: 50, and Ra: 40).

661 To define cell clusters in each animal species, we performed graph-based  
662 unsupervised clustering (**Fig. S2A**). First, the FindNeighbors function was used, and  
663 then, the FindClusters function was used. In these steps, the parameter ‘k.param’ for  
664 FindNeighbors was set individually for each animal species (Hs: 12, Pt: 10, Mm: 10,  
665 and Ra: 20). The parameter ‘resolution’ for FindClusters was also set individually for  
666 each animal species (Hs: 2.0, Pt: 2.2, Mm: 1.7, Ra: 1.2).

667

## 668 **Cell annotation**

669 Regarding each cluster identified by graph-based unsupervised clustering in the  
670 section “**Data integration, visualization, and cell clustering**” (**Fig. S2A**), 11 cell  
671 types were manually annotated according to i) the predicted cell type by Azimuth (**Fig.**  
672 **S2B**), ii) the distances between each cluster (**Fig. S2C**), and iii) the correspondence  
673 of clusters between animal species (**Fig. S2D–F**). First, reference-based cell type  
674 prediction was performed using Azimuth for the mock data from each animal species  
675 (**Fig. S2B**). In this step, the gene annotation “genes shared with humans” (see **Gene**  
676 **annotation and ortholog information**) for each animal species was used. We  
677 checked the enrichment of each predicted cell type in each cluster by Azimuth. Second,  
678 we checked the similarities between clusters by hierarchical clustering (**Fig. S2C**)

679 using the mean values of PCs 1-50 among the individual cells (see **Data integration,**  
680 **visualization, and cell clustering**) in each cluster. Notably, PCA was performed  
681 using the expression levels of “all genes” (see **Gene annotation and ortholog**  
682 **information**). The Euclidian distance was used for clustering by Ward’s method. Third,  
683 to check the correspondence between clusters in each animal species, we performed  
684 data integration, clustering, and visualization for mock data from all four animal  
685 species (**Fig. S2D–F**). In the integration, the mock data from humans were used as  
686 reference data. In this step, the gene annotation “common genes” (see **Gene**  
687 **annotation and ortholog information**) was used.

688 After categorizing cells into 11 cell types, the 11 cell types were coarse-grained  
689 into 6 cell types based on the results of hierarchical clustering analysis (see  
690 **Hierarchical clustering**). The six cell types were used in the subsequent analysis.

691

692

### 693 **Hierarchical clustering**

694 To examine the similarities in expression patterns among the conditions (4 animal  
695 species × 4 stimuli × 6 cell types = 96 conditions), hierarchical clustering analysis was  
696 performed. In this analysis, the 5,000 genes with the highest median absolute  
697 deviation (mad) values were used (**Fig. S2H**). First, the average expression levels of  
698 the respective genes in each condition were calculated. Next, PCA was performed  
699 using the average expression profiles. Third, using PCs 1-30, the distance matrix for  
700 the 96 conditions was generated using 1–Pearson’s correlation coefficient. Finally,  
701 hierarchical clustering by Ward’s method was performed using the distance matrix.

702 To determine which factor (e.g., animal species, stimulus, or cell type) was the  
703 most impactful on the gene expression in immune cells, hierarchical clustering was  
704 performed using induction patterns upon stimulation (**Fig. 1E**). Unlike for the results  
705 shown in **Fig. S2H**, FC values were used to perform PCA. This analysis used 7557  
706 genes, the union of the top 6000 genes related to total expression levels in the  
707 expression profiles of each animal species. The FC expression values (stimulated vs.  
708 unstimulated conditions) of those genes were calculated for each cell type in each  
709 animal species. To avoid generating infinite FC values, the data for genes with zero  
710 expression in mock data were set at the minimum nonzero expression level in the  
711 mock data. Finally, hierarchical clustering was performed using the method described  
712 above.



713

## 714 **Tensor decomposition**

715 To extract species-specific/common induction patterns upon stimulation from  
716 transcriptome data with complex structures (4 animal species × 3 stimuli × 4 cell types  
717 × 7557 orthologous genes), we used tensor decomposition (**Fig. 2A**). As the input data  
718 for tensor decomposition, the FC values of 7557 genes, the union of the top 6000  
719 genes related to total expression levels in the expression profiles of each animal, were  
720 used. The calculation method for FC values is described in the section “**Hierarchical**  
721 **clustering**”. The standardized FC values for each condition were represented as a 4-  
722 mode tensor (animal species × stimulus × cell type × orthologous gene). To  
723 perform Tucker decomposition (TD), a method of tensor decomposition, we used  
724 TensorLy (v0.6.0) [52]. We performed TD via higher-order orthogonal iteration (HOI)  
725 with the parameter ‘init=“svd”’. In HOI, the size of the core tensor (ranks) was set as  
726 [animal species: 3, stimulus: 2, cell type: 3, gene: 15]. The number of iterations was  
727 set as 100.

728

## 729 **Gene classification using the tensor decomposition results**

730 A schematic of the gene classification using tensor decomposition is shown in **Fig. 2C**  
731 **and Fig. S3C–F**. Briefly, we selected the candidate gene categories that had patterns  
732 of values (high, mid, or low) (**Fig. S3C**) that matched the ideal pattern (**Fig. S3D**) and  
733 then selected the gene category with the best “similarity score” (**Fig. S3E**) from the  
734 candidates as the gene category for that gene (**Fig. S3F**).

735 Initially, the product of the core tensor and the three factor-matrices, A2 (for  
736 stimulus), A3 (for cell type), and A4 (for gene), was calculated to obtain three cubic  
737 data with three axes, stimulus, cell type, and gene, using the ttl function of rTensor  
738 (v1.4.8) [53]. Each cubic data point indicated information related to species-common,  
739 bat-specific, and macaque-specific factors (**Fig. 2B**). Next, since the values of latent  
740 factors related to HSV-1 and SeV were similar (**Fig. S3A**), these two categories were  
741 integrated into the category “Virus” by calculating mean values. Additionally, since the  
742 values of latent factors related to NaiveT and KillerTNK were similar (**Fig. S3B**), these  
743 two categories of cell types were integrated into the category “TNK” by calculating  
744 mean values. Thus, hereafter, the category of stimuli included virus and LPS, and the  
745 category of cell types included B cells, TNK cells and Monos.

746 Then, in each cubic data, genes were classified into 11 categories (**Fig. 2C**)  
747 through the following three steps. Briefly, from the candidate gene categories that had  
748 patterns of values (high, mid, or low) (**Fig. S3C**) that matched the ideal pattern (**Fig.**  
749 **S3D**), the gene category with the lowest “similarity score” (**Fig. S3E**) was selected as  
750 the gene category for that gene (**Fig. S3F**).

751 In the first step (**Fig. S3C**), the values in each cubic data were normalized, and  
752 the genes were classified into three classes (high, mid, and low) according to the  
753 ranking of values in each condition (stimulus × cell type). First, six column vectors in  
754 the TD results for the 6 conditions (2 stimuli × 3 cell types) were normalized by dividing  
755 them by the 90th percentile for the individual vectors. After the division step, to  
756 suppress the effect of abnormally high or low values, data with  $> 1$  or  $< -1$  were  
757 assigned as 1 and -1, respectively. Next, the genes were categorized into three  
758 classes based on the rule that if the rank of a value was greater than the 80th  
759 percentile or smaller than the 20th percentile, it was categorized as “high” or “low”,  
760 respectively; otherwise, it was categorized as “mid”.

761 In the second step (**Fig. S3E**), a “similarity score” was calculated to represent  
762 the similarity between the genewise pattern of the TD results and the “ideal patterns”  
763 for each gene category. The “ideal patterns” were defined as vectors composed of 1,  
764 0, and -1 for 16 gene categories (Virus\_high, LPS\_low, Virus\_low, LPS\_high, B\_high,  
765 TNKM\_low, B\_low, TNKM\_high, TNK\_high, BM\_low, TNK\_low, BM\_high, M\_high,  
766 BTNK\_low, M\_low, and BTNK\_high) (**Fig. S3D**). The “similarity score” was defined as  
767 the sum of the residual squares between the two vectors, the genewise vector of  
768 normalized values from the TD results (**Fig. S3C**) and the “ideal patterns” (**Fig. S3D**).  
769 According to the definition, the “similarity scores” for every combination of genes and  
770 gene categories were calculated. After calculating all similarity scores, to obtain the  
771 threshold for checking if a gene should be recognized as a gene in that category, the  
772 20th percentile of the similarity score in the vector for each gene category was  
773 calculated.

774 In the third step (**Fig. S3F**), the gene category for each gene was determined.  
775 First, the candidate gene categories for each gene were filtered according to the  
776 pattern assigned in the first step (**Fig. S3C**). If the pattern (high/mid/low) of all 6  
777 conditions was high or low, the gene was categorized as ALL\_high or ALL\_low,  
778 respectively. If the pattern of a gene matched the “ideal pattern” of a gene category,  
779 the gene category was added as a candidate gene category for the gene. For example,

780 if the pattern of gene A was (Virus\_B: high, Virus\_TNK: high, Virus\_M: high, LPS\_B:  
781 high, LPS\_TNK: low, LPS\_M: mid), the candidate gene category for gene A was  
782 “Virus\_high” and “B\_high” because all virus-infected data were assigned as “high” and  
783 all B-cell data were assigned as “high” (**Fig. S3D**). Second, the gene category with the  
784 lowest “similarity score” among the candidate gene categories was selected as the  
785 tentative gene category. In this selection, if the “similarity score” was higher than the  
786 threshold of the gene category (**Fig. S3E**), the gene was categorized as “Others” (See  
787 gene B in **Fig. S3F**) because the pattern for the gene was recognized as being too  
788 different from the “ideal pattern”. If no candidate gene category was available, the gene  
789 was also classified as “Others” (See gene C in **Fig. S3F**). Finally, the final gene  
790 category was determined by integrating similar gene categories (**Fig. S3F**). For  
791 instance, the categories Virus\_high and LPS\_low were integrated into the category  
792 Virus\_high because both categories indicated that virus-infected data were higher than  
793 LPS-stimulated data (See gene D in **Fig. S3F**). As a result of the gene classification  
794 process, genes were categorized into one of 11 categories (**Fig. 2C, S3D**).

795

#### 796 **GO term enrichment analysis**

797 Gene Ontology (GO) analysis was performed with Fisher’s exact test. This analysis  
798 used the GO canonical pathways and GO biological processes defined by MSigDB  
799 (RRID:SCR\_022870) (v7.3) [30]. Adjusted P values were calculated using the  
800 Benjamini–Hochberg (BH) method.

801

#### 802 **Calculation of gene set variation analysis (GSVA) scores**

803 The gene set-wise expression scores used in **Figs. 3A, 4B, 4C, and S4B** were  
804 calculated using GSVA (RRID:SCR\_021058) (v1.38.2) [54, 55] with the algorithm  
805 “ssgsea”.

806

#### 807 **Identification of differentially expressed genes (DEGs) and marker genes**

808 In bat monocytes, DEGs were identified in Cluster 5 or Cluster 7 compared to the other  
809 clusters using the FindMarkers function of Seurat packages. A gene that met the  
810 following three criteria was considered a DEG: 1) the false discovery rate (FDR)  
811 calculated using the BH method was less than 0.05, 2) the average log<sub>2</sub>FC was  
812 greater than 1 or less than -1, and 3) the proportion of expressing cells was greater  
813 than 0.2.

814           The marker genes of Cluster 5 and Cluster 7 of bat monocytes (RaC5marker  
815 and RaC7marker, respectively) were defined as upregulated DEGs in Cluster 5 (**Fig.**  
816 **4E**) and Cluster 7 (**Fig. 4G**), respectively.

817 **Availability of source code and requirements**

818 Project name: scRNA-seq\_PBMC\_Animals\_Aso\_et\_al

819 Project homepage: [https://github.com/TheSatoLab/scRNA-](https://github.com/TheSatoLab/scRNA-seq_PBMC_Animals_Aso_et_al)  
820 [seq\\_PBMC\\_Animals\\_Aso\\_et\\_al](https://github.com/TheSatoLab/scRNA-seq_PBMC_Animals_Aso_et_al) [56]

821 Operating system: Linux

822 Programming languages: bash, R, Python

823 License: CC0-1.0

824

825 **Data availability**

826 The raw and processed single-cell RNA-seq data have been deposited in the Gene  
827 Expression Omnibus (GEO) database (GSE218199) and are publicly available.

828 Original data to describe figures in this paper have been deposited at Mendeley [57]  
829 and are publicly available. All additional supporting data are available in the  
830 *GigaScience* database [59].

831

832 **Declarations**

833 **List of abbreviations**

834 cDCs: conventional dendritic cells; CIU: cell infectious unit; CP10k: counts per 10,000  
835 counts in the cell; DEGs: differentially expressed genes; DMEM: Dulbecco's modified  
836 Eagle's medium; dsRNAs: double-stranded RNAs; FC: fold-change; FCS fetal calf  
837 serum; FDR: false discovery rate; GEMs: gel beads-in-emulsion; GEO: Gene  
838 Expression Omnibus; GSVA gene set variation analysis; GO: Gene Ontology; HOI:  
839 higher-order orthogonal iteration; Hs: *Homo sapiens*; HSV-1: herpes simplex virus  
840 type 1; IFNs: interferons; ISGs: IFN-stimulated genes; LPS: Lipopolysaccharide; mad:  
841 median absolute deviation; MAITs: mucosal-associated invariant T cells; Mm: *Macaca*  
842 *mulatta*; Monos: monocytes; NK: natural killer; PAMPs: pathogen-associated  
843 molecular patterns; PBMCs: peripheral blood mononuclear cells; PCs: principal  
844 components; PCA: principal component analysis; pDCs: plasmacytoid dendritic cells;  
845 PFU plaque forming unit; PRRs: pattern recognition receptors; Pt: *Pan troglodytes*;  
846 Ra: *Rousettus aegyptiacus*; RLRs: RIG-I-like receptors; scRNA-seq: single-cell RNA

847 sequencing; SeV: Sendai virus; TD: Tucker decomposition; UMAP: uniform manifold  
848 approximation and projection; UMI: unique molecular identifier; QC: quality control

849

### 850 **Ethics statement**

851 All protocols involving specimens from animals were performed in accordance with the  
852 Science Council of Japan's Guidelines for the Proper Conduct of Animal Experiments.  
853 The protocols were approved by the Institutional Animal Care and Use Committee of  
854 Kyoto University (approval IDs: 2017-B-5, 2019-C-9, 2019-162, 2019-177, and 2020-  
855 C-5). All protocols involving specimens from humans recruited at Kyoto University  
856 were reviewed and approved by the Institutional Review Boards of Kyoto University  
857 (approval ID: G1089). All protocols for the use of human specimens were reviewed  
858 and approved by the Institutional Review Boards of The Institute of Medical Science,  
859 The University of Tokyo (approval ID: 2019-55) and Kyoto University (approval ID:  
860 G1089).

861

### 862 **Consent for publication**

863 All human subjects provided written informed consent.

864

### 865 **Competing interests**

866 The authors declare no competing interests.

867

### 868 **Funding**

869 This study was supported in part by AMED SCARDA Japan Initiative for World-leading  
870 Vaccine Research and Development Centers "UTOPIA" (JP223fa627001, to K.S.),  
871 AMED SCARDA Program on R&D of new generation vaccine including new modality  
872 application (JP223fa727002, to K.S.); AMED Research Program on Emerging and Re-  
873 emerging Infectious Diseases (JP22fk0108146, to Y.Kashima and K.S.;  
874 JP21fk0108494 to K.S.; 21fk0108425, to K.S.; 21fk0108432, to K.S.); AMED  
875 Research Program on HIV/AIDS (JP22fk0410039, to K.S.); JST PRESTO  
876 (JPMJPR22R1, to J.I.); AMED Moonshot Research and Development Program  
877 (JP21zf0127005, to H.O.); JST CREST (JPMJCR20H4, to K.S.); JSPS KAKENHI  
878 Grant-in-Aid for Early-Career Scientists (20K15767, to J.I.; 19K20394, to H.O.); JSPS  
879 Core-to-Core Program (A. Advanced Research Networks) (JPJSCCA20190008, to  
880 K.S.); JSPS Research Fellow DC1 (20J23299, to H.A.).

881

### 882 **Author's contributions**

883 H.A. mainly performed bioinformatics analysis. J.I. and H.O. supervised the  
884 bioinformatics analysis. Y.Kashima mainly performed the experiments. Y.S.,  
885 Y.Koyanagi, and K.S. supervised the experiments. K.S. and Y.Koyanagi provided  
886 reagents. K.S. conceived and designed the experiments. H.A. and J.I. wrote the initial  
887 manuscript. All authors reviewed and edited the manuscript.

888

### 889 **Acknowledgements**

890 We would like to thank Naoko Misawa, Akiko Oide, Mai Suganami, and Kazumi Abe  
891 (The University of Tokyo), for technical support, Hiroo Imai (Kyoto University) for  
892 providing primate PBMCs, Ayuko Morita (Kyoto City Institute of Health and  
893 Environmental Sciences) for providing bat PBMCs, Yasushi Kawaguchi (The  
894 University of Tokyo) for providing HSV-1, Takashi Irie (Hiroshima University) for  
895 providing SeV, and Human Genome Center (the Institute of Medical Science, the  
896 University of Tokyo) for providing the super-computing resource SHIROKANE  
897 (<http://sc.hgc.jp/shirokane.html>). This work was supported by the Cooperative  
898 Research Program of the Primate Research Institute, Kyoto University (2019-c9).

899

900

### 901 **Author's information**

902 H.A.'s current affiliation: Department of AI Systems Medicine, M&D Data Science  
903 Center, Tokyo Medical and Dental University, Tokyo 113-8510, Japan

904

### 905 **Additional Files**

906

### 907 **Figure S1. Validation of viral infectivity and the innate immune response (related 908 to Figure 1)**

909 (A) Heatmap of the induction levels of genes related to the IFN response and  
910 inflammation. The rows indicate genes, and the columns indicate combinations of  
911 species, stimulus, and dose. The color represents the log<sub>2</sub> Fold Change of ddCt upon  
912 stimulation measured by qRT-PCR. "rep. 1" and "rep. 2" indicate biological replicates.

913 (B-C) Heatmap of the expression levels of viral genes (B: HSV-1; C: SeV) measured  
914 by qRT-PCR. The rows indicate viral genes, and the columns indicate combinations  
915 of species and doses. "rep. 1" and "rep. 2" indicate biological replicates. The color  
916 represents the ddCt values based on the expression levels of GAPDH.

917 (D-G) Violin plots of (D) the numbers of detected genes per cell before QC, (E)  
918 numbers of counted reads per cell before QC, (F) numbers of detected genes per cell  
919 after QC, and (G) numbers of counted reads per cell after QC.

920

921 **Figure S2. Heterogeneous expression patterns in the four animal species**  
922 **(related to Figure 1)**

923 (A-B) UMAP plots representing the gene expression patterns of PBMCs from the four  
924 species. Each dot is colored according to the results of unsupervised clustering (A)  
925 and reference-based label transfer (B).

926 (C) Heatmaps showing pairwise Euclid distances representing the gene expression  
927 differences among clusters. The distances were calculated using PCs 1-50 of the gene  
928 expression data.

929 (D-E) UMAP plots representing the gene expression patterns of PBMCs from the mock  
930 samples for the four species. Each dot is colored according to the results of  
931 unsupervised clustering using the integrated data for the four mock samples (D) or the  
932 four samples from each animal shown in **Figure S2A** (E).

933 (F) Heatmaps showing pairwise Euclid distances representing the gene expression  
934 differences among clusters shown in **Figure S2D**. The distances were calculated  
935 using PCs 1-30 of the gene expression data.

936 (G) Dot plots representing the expression patterns of marker genes for each cell type  
937 defined by Azimuth [58]

938 (H) Hierarchical clustering analysis of 48 pseudobulked FC gene expression  
939 datapoints (4 animal species x 4 stimuli x 11 cell types = 176 conditions).

940

941 **Figure S3. Classification of genes according to species-specific expression**  
942 **patterns (related to Figure 2)**

943 (A) Heatmap representing a latent factor matrix related to stimuli. The columns indicate  
944 stimuli, and the rows indicate latent factors representing stimulus-common (L2\_1) and  
945 virus vs. LPS (L2\_2) factors.

946 (B) Heatmap representing a latent factor matrix related to cell types. The columns  
947 indicate cell types, and the rows indicate latent factors representing cell type-common  
948 (L3\_1), monocyte-specific (L3\_2), and B-cell-specific (L1\_3) factors.

949 (C) Summary of the normalization of values and patterning according to the ranking of  
950 the values. First, six column vectors (2 stimuli x 3 cell types) in the TD results were



951 normalized by dividing them by the 90th percentile of the individual vectors. Then, data  
952 with  $> 1$  or  $< -1$  were assigned as 1 and -1, respectively. Next, the genes were  
953 categorized into three classes (high, mid, and low) based on the rule that if the rank of  
954 a value was greater than the 80th percentile or smaller than the 20th percentile, it was  
955 categorized as “high” or “low”, respectively; otherwise, it was categorized as “mid”.

956 (D) Summary of the ideal patterns for each gene category used in the gene  
957 classification in **Figure 2C**.

958 (E) Summary of the calculation of the similarity score and establishment of the  
959 threshold for the gene classification in **Fig. S3F**. The sum of the residual squares  
960 between two vectors, the genewise vector of normalized values from the TD results  
961 (**Fig. S3C**) and the “ideal patterns” (**Fig. S3D**) were calculated. Then, the threshold  
962 used in **Fig. S3F** was obtained by calculating the 20th percentile of the similarity score  
963 for the vector for each gene category.

964 (F) Summary of gene classification. By comparing patterns from the TD results (**Fig.**  
965 **S3C**) and the ideal patterns (**Fig. S3D**), candidate gene categories were selected.  
966 Next, the gene category with the lowest “similarity score” among the candidate gene  
967 categories was selected as the tentative gene category. In this selection, if the  
968 “similarity score” was higher than the threshold of the gene category (**Fig. S3E**), the  
969 gene was categorized as “Others” (gene B). If no candidate gene category was  
970 available, the gene was also classified as “Others” (gene C). Finally, the final gene  
971 category was determined by integrating similar gene categories (genes A and D).

972 (G-I) Heatmap representing the values of the products calculated in **Figure 2C**. The  
973 data relating to (G) the species-common factor (L1\_1), (H) the bat-specific factor  
974 (L1\_2), and (I) the macaque-specific factor (L1\_3) are shown. Each row indicates the  
975 respective gene. The color keys shown on the right of the heatmap indicate gene  
976 categories.

977 (J-L) Heatmap representing the FC values in the input tensor. The orders of the rows  
978 are the same as in (J) **Figure S3G**, (K) **Figure S3H**, and (L) **Figure S3I**. Each row  
979 indicates the respective gene. The color keys shown on the right of the heatmap  
980 indicate gene categories.

981

982 **Figure S4. Identification of species-specific cell types (related to Figure 4)**

983 (A) UMAP plots representing the expression patterns of every single cell.  
984 Dimensionality reduction was performed for each combination of the four species and  
985 three cell types.

986 (B) UMAP plots representing the average expression levels of marker genes for  
987 Cluster 7 [C7markers].

988

989 **Table. S1. Primers used for RT-qPCR (related to the Methods)**

990 The sequences of the primers used for RT-qPCR are listed.

991

992 **References**

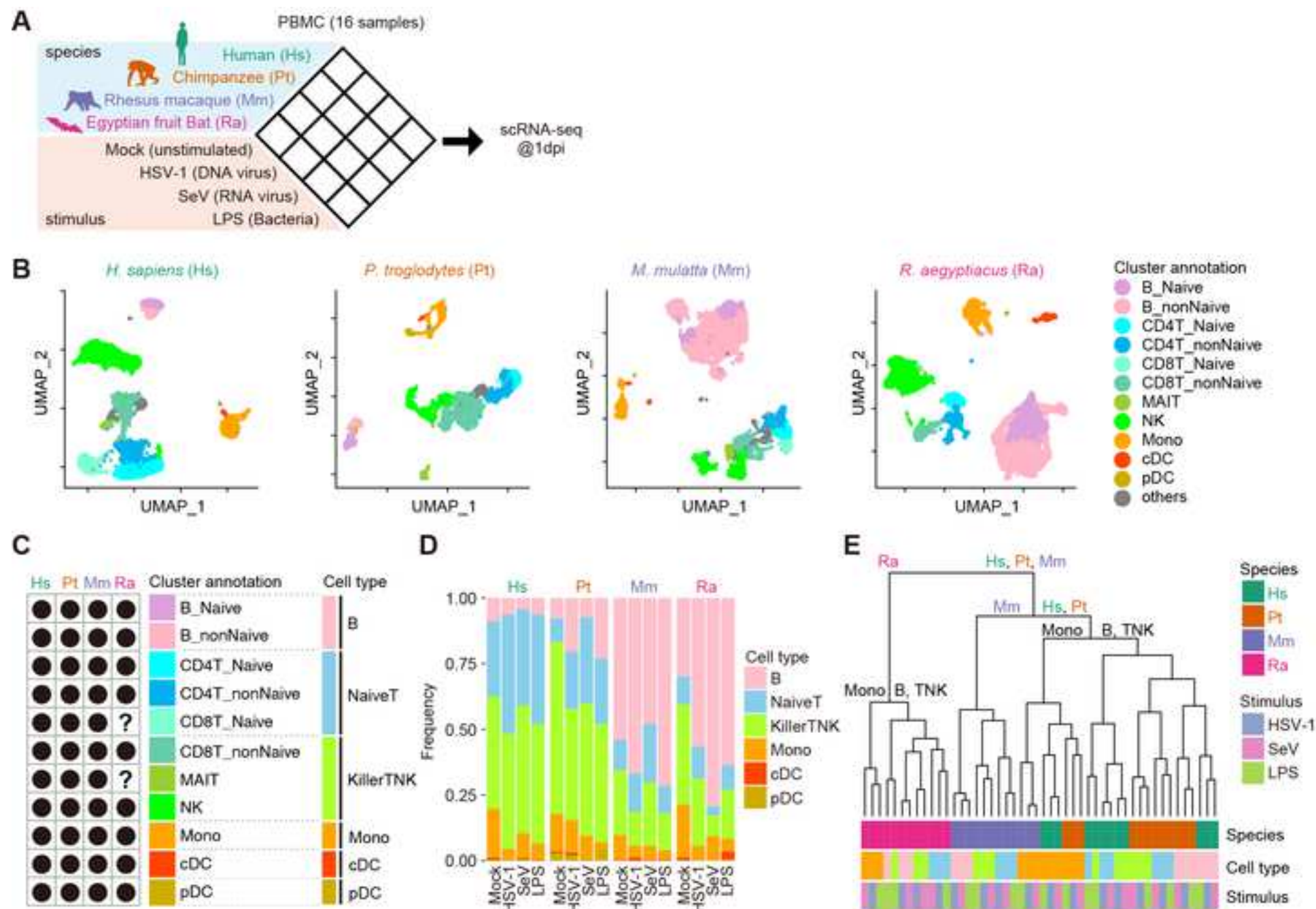
- 993 1. Huff JL, Barry PA. B-virus (Cercopithecine herpesvirus 1) infection in humans  
994 and macaques: potential for zoonotic disease. *Emerg Infect Dis.* 2003;9(2):246-  
995 250.
- 996 2. Letko M, Seifert SN, Olival KJ, et al. Bat-borne virus diversity, spillover and  
997 emergence. *Nat Rev Microbiol.* 2020;18(8):461-471.
- 998 3. Towner JS, Amman BR, Sealy TK, et al. Isolation of genetically diverse Marburg  
999 viruses from Egyptian fruit bats. *PLoS Pathog.* 2009;5(7):e1000536.
- 1000 4. Guito JC, Prescott JB, Arnold CE, et al. Asymptomatic infection of Marburg  
1001 virus reservoir bats is explained by a strategy of immunoprotective disease  
1002 tolerance. *Curr Biol.* 2021;31(2):257-270.e5.
- 1003 5. Akira S, Uematsu S, Takeuchi O. Pathogen recognition and innate immunity.  
1004 *Cell.* 2006;124(4):783-801.
- 1005 6. Takeuchi O, Akira S. Pattern recognition receptors and inflammation. *Cell.*  
1006 2010;140(6):805-820.
- 1007 7. Hansen JD, Vojtech LN, Laing KJ. Sensing disease and danger: a survey of  
1008 vertebrate PRRs and their origins. *Dev Comp Immunol.* 2011;35(9):886-897.
- 1009 8. Lu Y, Su F, Li Q, et al. Pattern recognition receptors in Drosophila immune  
1010 responses. *Dev Comp Immunol.* 2020;102:103468.
- 1011 9. Dempsey A, Bowie AG. Innate immune recognition of DNA: A recent history.  
1012 *Virology.* 2015;479-480:146-152.
- 1013 10. Poltorak A, He X, Smirnova I, et al. Defective LPS signaling in C3H/HeJ and  
1014 C57BL/10ScCr mice: mutations in Tlr4 gene. *Science.* 1998;282:2085-2088.
- 1015 11. Schountz T, Baker ML, Butler J, et al. Immunological control of viral infections  
1016 in bats and the emergence of viruses highly pathogenic to humans. *Front*  
1017 *Immunol.* 2017;8:1098.
- 1018 12. Gorbunova V, Seluanov A, Kennedy BK. The world goes bats: Living longer  
1019 and tolerating viruses. *Cell Metab.* 2020;32(1):31-43.
- 1020 13. Banerjee A, Baker ML, Kulcsar K, et al. Novel insights into immune systems of  
1021 bats. *Front Immunol.* 2020;11:26.
- 1022 14. Pavlovich SS, Lovett SP, Koroleva G, et al. The Egyptian rousette genome  
1023 reveals unexpected features of bat antiviral immunity. *Cell.* 2018;173(5):1098-  
1024 1110.e18.

- 1025 15. Zhou P, Tachedjian M, Wynne JW, et al. Contraction of the type I IFN locus and  
1026 unusual constitutive expression of IFN- $\alpha$  in bats. *Proc Natl Acad Sci USA*.  
1027 2016;113(10):2696-2701.
- 1028 16. Xie J, Li Y, Shen X, et al. Dampened STING-dependent interferon activation in  
1029 bats. *Cell Host Microbe*. 2018;23(3):297-301.e4.
- 1030 17. Ahn M, Cui J, Irving AT, et al. Unique loss of the PYHIN gene family in bats  
1031 amongst mammals: Implications for inflammasome sensing. *Sci Rep*.  
1032 2016;6:21722.
- 1033 18. Irving AT, Zhang Q, Kong PS, et al. Interferon regulatory factors IRF1 and IRF7  
1034 directly regulate gene expression in bats in response to viral infection. *Cell Rep*.  
1035 2020;33(5):108345.
- 1036 19. De La Cruz-Rivera PC, Kanchwala M, Liang H, et al. The IFN response in bats  
1037 displays distinctive IFN-stimulated gene expression kinetics with atypical  
1038 RNASEL induction. *J Immunol*. 2018;200(1):209-217.
- 1039 20. Lin HH, Horie M, Tomonaga K. A comprehensive profiling of innate immune  
1040 responses in *Eptesicus* bat cells. *Microbiol Immunol*. 2022;66(3):97-112.
- 1041 21. Ahn M, Anderson DE, Zhang Q, et al. Dampened NLRP3-mediated  
1042 inflammation in bats and implications for a special viral reservoir host. *Nat*  
1043 *Microbiol*. 2019;4(5):789-799.
- 1044 22. Goh G, Ahn M, Zhu F, et al. Complementary regulation of caspase-1 and IL-1 $\beta$   
1045 reveals additional mechanisms of dampened inflammation in bats. *Proc Natl*  
1046 *Acad Sci USA*. 2020;117(46):28939-28949.
- 1047 23. Seurat (2021). Seurat (version 4.0.4).  
1048 <https://github.com/satijalab/seurat/releases/tag/v4.0.4>.
- 1049 24. Hao Y, Hao S, Andersen-Nissen E, et al. Integrated analysis of multimodal  
1050 single-cell data. *Cell*. 2021;184(13):3573-3587.e29.
- 1051 25. Azimuth (2020). Azimuth (version 0.4.3).  
1052 <https://github.com/satijalab/azimuth/releases/tag/v0.4.3>.
- 1053 26. Friedrichs V, Toussaint C, Schäfer A, et al. Landscape and age dynamics of  
1054 immune cells in the Egyptian rousette bat. *Cell Rep*. 2022;40(10):111305.
- 1055 27. Banerjee A, Rapin N, Bollinger T, et al. Lack of inflammatory gene expression  
1056 in bats: a unique role for a transcription repressor. *Sci Rep*. 2017;7(1):2232.

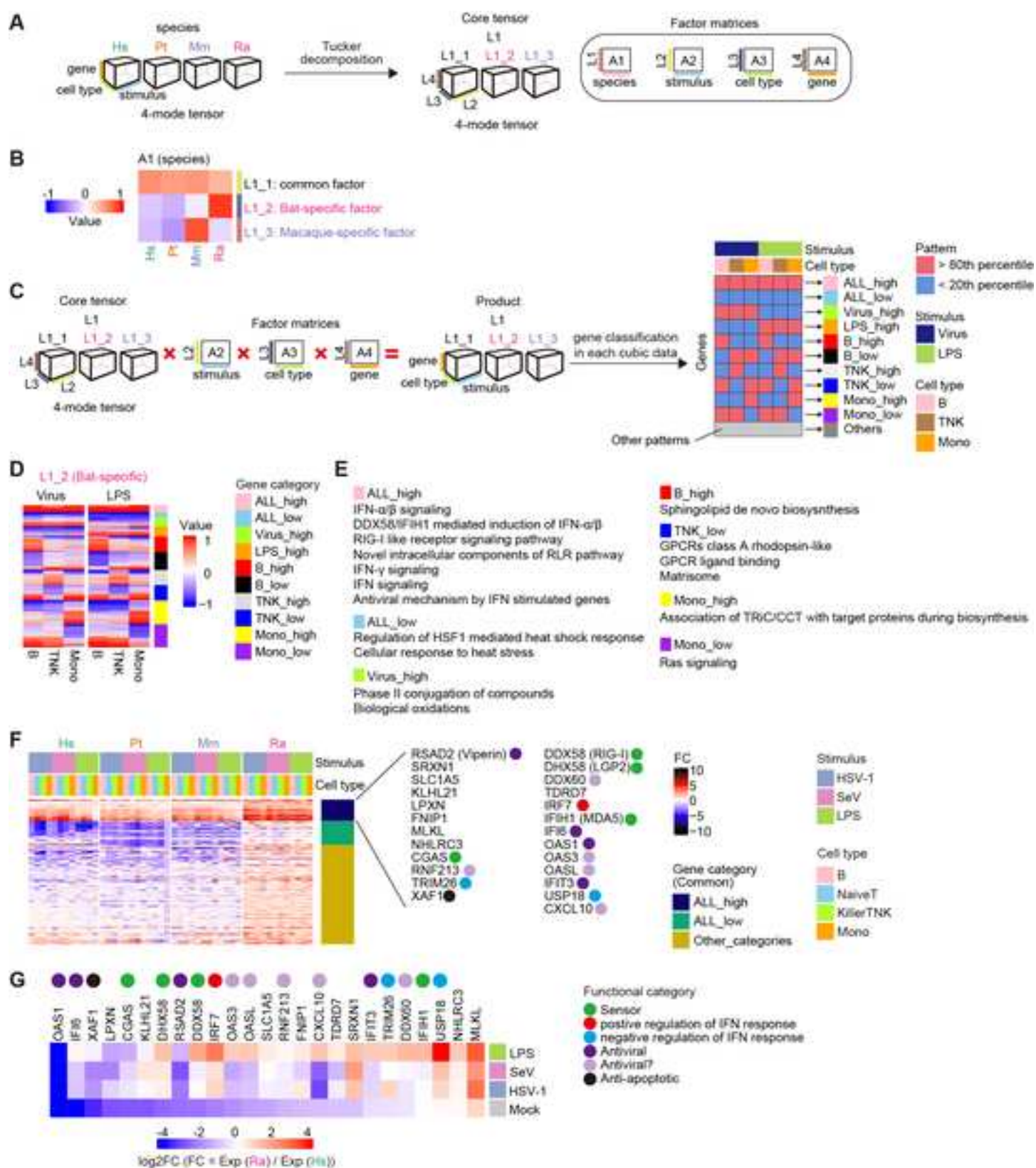
- 1057 28. Shaw AE, Hughes J, Gu Q, et al. Fundamental properties of the mammalian  
1058 innate immune system revealed by multispecies comparison of type I interferon  
1059 responses. *PLoS Biol.* 2017;15(12):e2004086.
- 1060 29. Sato R, Kato A, Chimura T, et al. Combating herpesvirus encephalitis by  
1061 potentiating a TLR3-mTORC2 axis. *Nat Immunol.* 2018;19(10):1071-1082.
- 1062 30. Shah S, King EM, Chandrasekhar A, et al. Roles for the mitogen-activated  
1063 protein kinase (MAPK) phosphatase, DUSP1, in feedback control of  
1064 inflammatory gene expression and repression by dexamethasone. *J Biol Chem.*  
1065 2014;289(19):13667-13679.
- 1066 31. Li QJ, Chau J, Ebert PJ, et al. miR-181a is an intrinsic modulator of T cell  
1067 sensitivity and selection. *Cell.* 2007;129(1):147-161.
- 1068 32. Posselt G, Schwarz H, Duschl A, et al. Suppressor of cytokine signaling 2 is a  
1069 feedback inhibitor of TLR-induced activation in human monocyte-derived  
1070 dendritic cells. *J Immunol.* 2011;187(6):2875-2884.
- 1071 33. MSigDB (version 7.3). GSEA and MSigDB team, Broad Institute,  
1072 Massachusetts. 2022. <https://www.gsea-msigdb.org/gsea/msigdb/>. Accessed 1  
1073 Sept 2022.
- 1074 34. Seal S, Dharmarajan G, Khan I. Evolution of pathogen tolerance and emerging  
1075 infections: A missing experimental paradigm. *Elife.* 2021;10:e68874.
- 1076 35. Lafaille FG, Pessach IM, Zhang SY, et al. Impaired intrinsic immunity to HSV-1  
1077 in human iPSC-derived TLR3-deficient CNS cells. *Nature.* 2012;491:769-773.
- 1078 36. López-Cotarelo P, Gómez-Moreira C, Criado-García O, et al. Beyond  
1079 chemoattraction: Multifunctionality of chemokine receptors in leukocytes.  
1080 *Trends Immunol.* 2017;38(12):927-941.
- 1081 37. Abcam. Chemokines and their receptors. Abcam, UK. 2021.  
1082 [https://docs.abcam.com/pdf/immunology/chemokines\\_poster.pdf](https://docs.abcam.com/pdf/immunology/chemokines_poster.pdf). Accessed 1  
1083 Sept 2022.
- 1084 38. Ejercito PM, Kieff ED, Roizman B. Characterization of herpes simplex virus  
1085 strains differing in their effects on social behaviour of infected cells. *J Gen Virol.*  
1086 1968;2(3):357-364.
- 1087 39. Yoshida A, Kawabata R, Honda T, et al. A single amino acid substitution within  
1088 the paramyxovirus Sendai virus nucleoprotein is a critical determinant for  
1089 production of interferon-beta-inducing copyback-type defective interfering  
1090 genomes. *J Virol.* 2018;92(5):e02094-17.

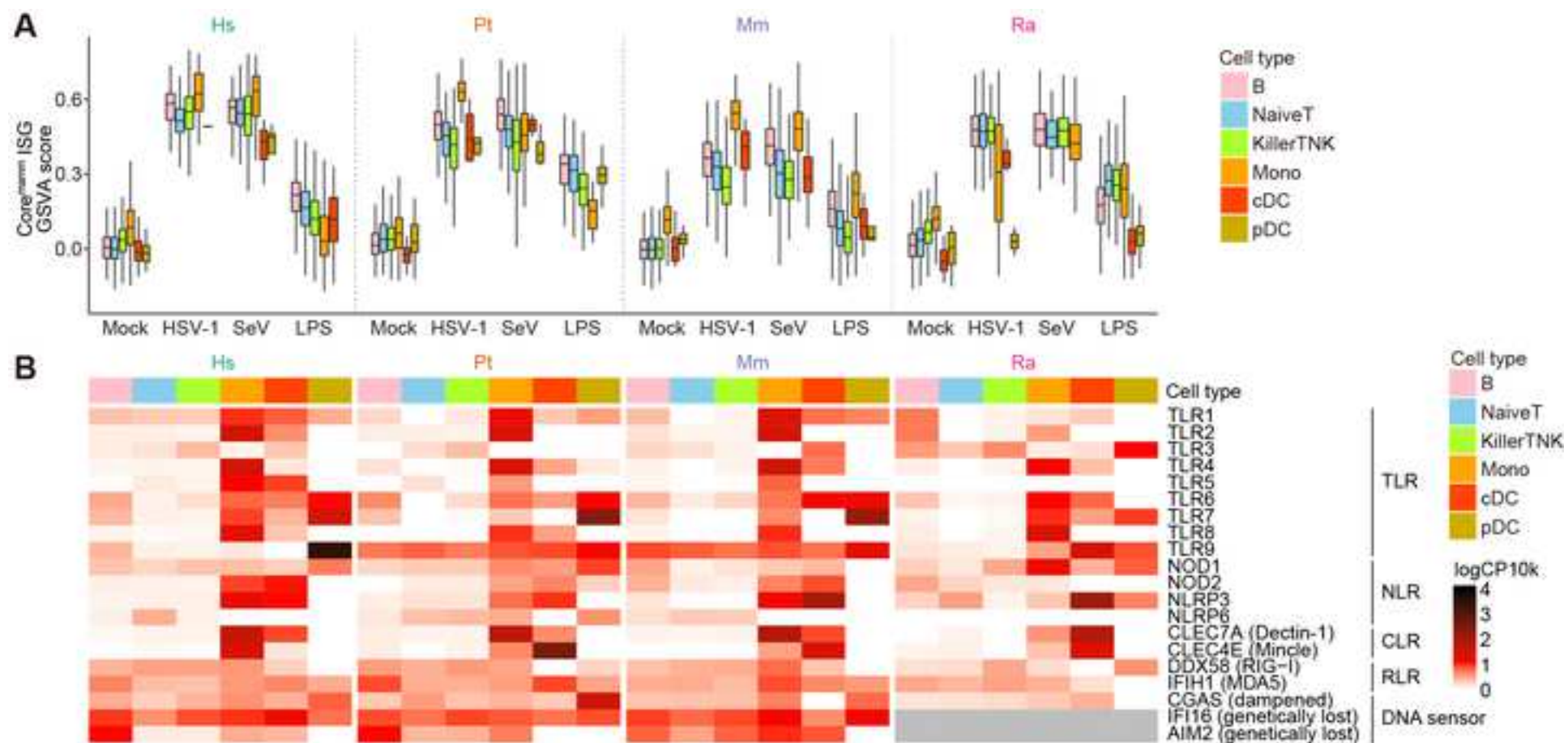
- 1091 40. Kiyotani K, Sakaguchi T, Kato A, et al. Paramyxovirus Sendai virus V protein  
1092 counteracts innate virus clearance through IRF-3 activation, but not via  
1093 interferon, in mice. *Virology*. 2007;359(1):82-91.
- 1094 41. Yamada E, Nakaoka S, Klein L, et al. Human-specific adaptations in Vpu  
1095 conferring anti-tetherin activity are critical for efficient early HIV-1 replication in  
1096 vivo. *Cell Host Microbe*. 2018;23(1):110-120.e7.
- 1097 42. RefSeq. National Center for Biotechnology Information, Maryland. 2020.  
1098 <https://www.ncbi.nlm.nih.gov/refseq/>. Accessed 6 Nov 2020.
- 1099 43. 10X Genomics. Build a custom reference (cellranger mkref). 10X Genomics,  
1100 California. 2020. [https://support.10xgenomics.com/single-cell-gene-](https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/using/tutorial_mr)  
1101 [expression/software/pipelines/latest/using/tutorial\\_mr](https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/using/tutorial_mr). Accessed 6 Nov 2020.
- 1102 44. The report of orthologous genes. National Center for Biotechnology Information,  
1103 Maryland. 2021. [https://ftp.ncbi.nih.gov/gene/DATA/gene\\_orthologs.gz](https://ftp.ncbi.nih.gov/gene/DATA/gene_orthologs.gz).  
1104 Accessed 26 July 2021.
- 1105 45. Jebb D, Huang Z, Pippel M, et al. Six reference-quality genomes reveal  
1106 evolution of bat adaptations. *Nature*. 2020;583:578-584.
- 1107 46. Bat1K Project. <https://bat1k.com> (2018). Accessed 14 July 2021.
- 1108 47. UCSC genome browser. University of California, California. 2000.  
1109 <https://genome.ucsc.edu>. Accessed 14 July 2021.
- 1110 48. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing  
1111 genomic features. *Bioinformatics*. 2010;26(6):841-842.
- 1112 49. Cell Ranger (2021). Cell Ranger (version 6.0.1).  
1113 [https://support.10xgenomics.com/single-cell-gene-](https://support.10xgenomics.com/single-cell-gene-expression/software/downloads/6.0)  
1114 [expression/software/downloads/6.0](https://support.10xgenomics.com/single-cell-gene-expression/software/downloads/6.0).
- 1115 50. Zheng GXY, Terry JM, Belgrader P, et al. Massively parallel digital  
1116 transcriptional profiling of single cells. *Nat Commun*. 2017;8:14049.
- 1117 51. McInnes L, Healy J, Melville J. UMAP: Uniform manifold approximation and  
1118 projection for dimension reduction. 2018. arXiv:1802.03426.  
1119 <https://doi.org/10.48550/arXiv.1802.03426>.
- 1120 52. TensorLy (2021). TensorLy (version 0.6.0).  
1121 <https://github.com/tensorly/tensorly/releases/tag/0.6.0>.
- 1122 53. rTensor (2022). rTensor (version 1.4.8).  
1123 <https://github.com/rikenbit/rTensor/releases/tag/v1.4.8>.

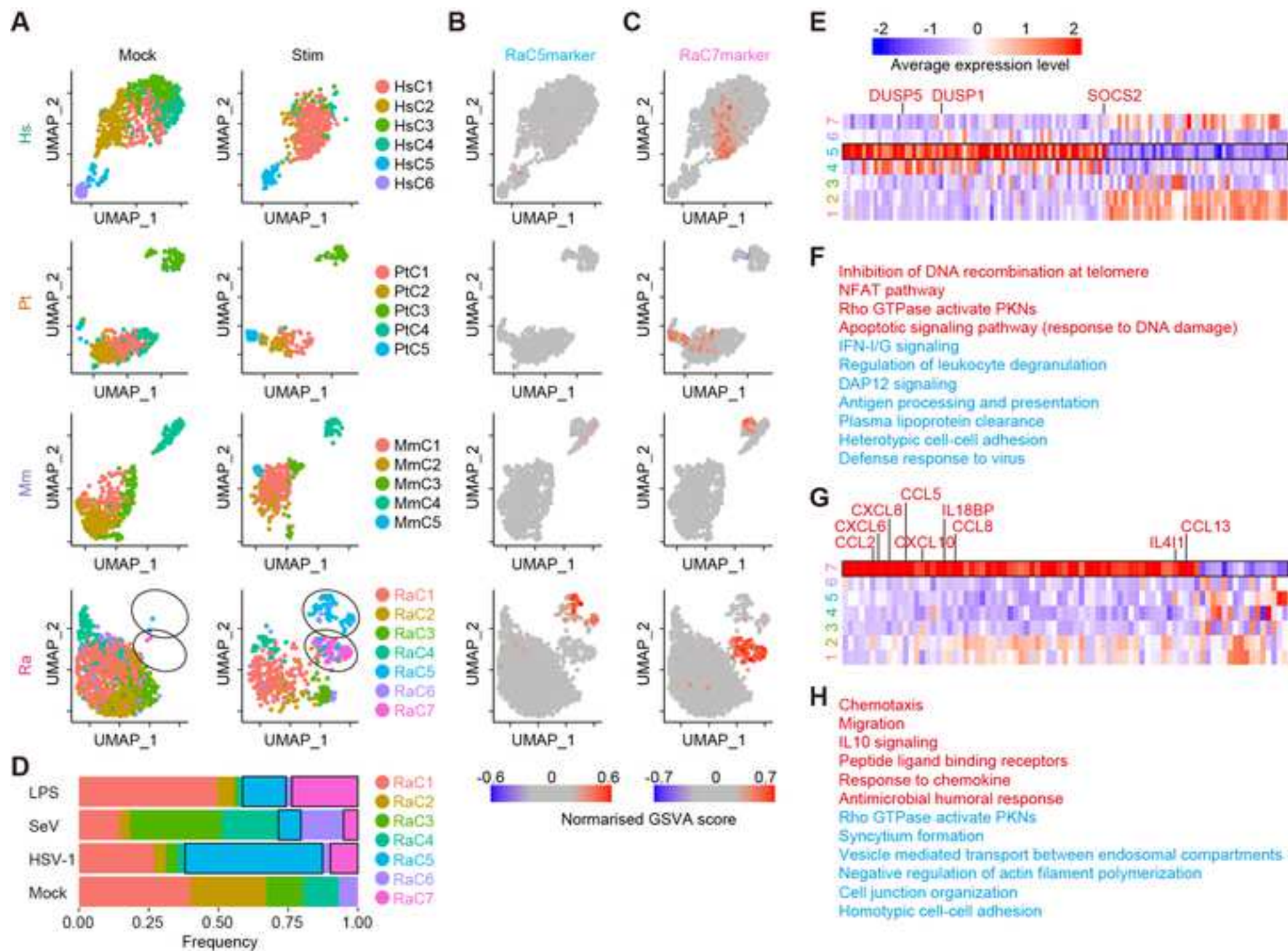
- 1124 54. GSVA (2021). GSVA (version 1.38.2).  
1125 <https://www.bioconductor.org/packages/release/bioc/html/GSVA.html>.
- 1126 55. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for  
1127 microarray and RNA-seq data. *BMC Bioinform.* 2013;14:7.
- 1128 56. Aso H, Ito J, Ozaki H, et al. scRNA-seq\_PBMC\_Animals\_Aso\_et\_al. *GitHub*.  
1129 2023. [https://github.com/TheSatoLab/scRNA-seq\\_PBMC\\_Animals\\_Aso\\_et\\_al](https://github.com/TheSatoLab/scRNA-seq_PBMC_Animals_Aso_et_al).
- 1130 57. 57. Single-  
1131 cell\_transcriptome\_analysis\_illuminating\_the\_characteristics\_of\_species-  
1132 specific\_innate\_immune\_responses\_against\_viral\_infections\_Aso\_et\_al.  
1133 Mendeley Data. 2023. DOI: 10.17632/kg3dfkyjv5.1
- 1134 58. Azimuth references (Human - PBMC). The Human BioMolecular Atlas Program  
1135 (HuBMAP), National Institutes of Health, Maryland. 2020.  
1136 <https://azimuth.hubmapconsortium.org/references/#Human%20-%20PBMC>.  
1137 Accessed 26 July 2021.
- 1138 59. Aso H, Ito J, Ozaki H, et al. Supporting data for "Single-cell transcriptome  
1139 analysis illuminating the characteristics of species-specific innate immune  
1140 responses against viral infections" GigaScience Database. 2023.  
1141 <http://dx.doi.org/10.5524/102448>.












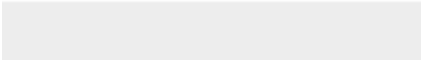



Click here to access/download  
**Supplementary Material**  
230812ResponseLetter\_v12.docx






Click here to access/download  
**Supplementary Material**  
FigS1.tif

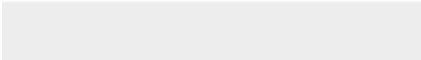






Click here to access/download  
**Supplementary Material**  
FigS2.tif

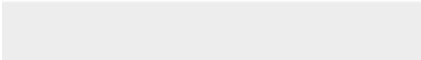



Click here to access/download  
**Supplementary Material**  
FigS3.tif

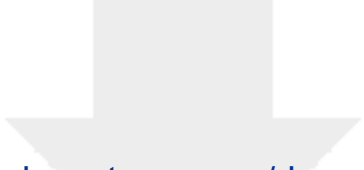




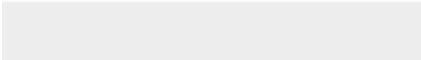

Click here to access/download  
**Supplementary Material**  
FigS4.tif








Click here to access/download  
**Supplementary Material**  
TableS1.xlsx





Click here to access/download  
**Supplementary Material**  
230812draft\_v32.docx



August 13, 2023



Editor,  
**GigaScience**

Please find enclosed our revised manuscript entitled "**Single-cell transcriptome analysis illuminating the characteristics of species-specific innate immune responses against viral infections**", by Aso et al., for the consideration of publication in **GigaScience**.

According to the comments raised by the two referees, we modified the manuscript. We hope that the editor will find our study important, and consider it suitable for publication in **GigaScience**.

Sincerely,

Kei

\*\*\*\*\*

Kei Sato, Ph.D.  
Professor, The Institute of Medical Science, The University of Tokyo, Japan.  
Email: [KeiSato@g.ecc.u-tokyo.ac.jp](mailto:KeiSato@g.ecc.u-tokyo.ac.jp)

\*\*\*\*\*