nature immunology

RESOURCE

https://doi.org/10.1038/s41590-021-00911-7

Check for updates

Systems analysis and controlled malaria infection in Europeans and Africans elucidate naturally acquired immunity

Sanne E. de Jong[1], Vincent van Unen[1,2,15,17], Mikhael D. Manurung[1,17], Koen A. Stam[1,17], Jelle J. Goeman[3], Simon P. Jochems[1], Thomas Höllt[4,5], Nicola Pezzotti[4], Yoanne D. Mouwenda[1,6], Madeleine Eunice Betouke Ongwe[1,6,7], Freia-Raphaella Lorenz[8,9], Yvonne C. M. Kruize[1], Shohreh Azimi[1], Marion H. König[1], Anna Vilanova[4], Elmar Eisemann[4], Boudewijn P. F. Lelieveldt[4,10], Meta Roestenberg[1], B. Kim Lee Sim[11], Marcel J. T. Reinders[12,13], Rolf Fendel[8,9], Stephen L. Hoffman[11], Peter G. Kremsner[6,8,9], Frits Koning[2], Benjamin Mordmüller[6,8,9,16], Bertrand Lell[6,14] and Maria Yazdanbakhsh[1]✉

Controlled human infections provide opportunities to study the interaction between the immune system and malaria parasites, which is essential for vaccine development. Here, we compared immune signatures of malaria-naive Europeans and of Africans with lifelong malaria exposure using mass cytometry, RNA sequencing and data integration, before and 5 and 11 days after venous inoculation with *Plasmodium falciparum* sporozoites. We observed differences in immune cell populations, antigen-specific responses and gene expression profiles between Europeans and Africans and among Africans with differing degrees of immunity. Before inoculation, an activated/differentiated state of both innate and adaptive cells, including elevated CD161^{+}CD4^{+} T cells and interferon-γ production, predicted Africans capable of controlling parasitemia. After inoculation, the rapidity of the transcriptional response and clusters of CD4^{+} T cells, plasmacytoid dendritic cells and innate T cells were among the features distinguishing Africans capable of controlling parasitemia from susceptible individuals. These findings can guide the development of a vaccine effective in malaria-endemic regions.

A highly effective malaria vaccine is needed to limit malaria morbidity and mortality worldwide[1]. However, this has been hampered by poor understanding of naturally acquired or vaccine-induced immunity[2]. Malaria vaccines that have entered clinical testing show great variation in efficacy in different geographical locations[3,4]. Increasing evidence supports the notion that human immune responses are not only shaped by genetics, but are also markedly influenced by the environment[5]. Indeed, vaccine responses can be affected by increased burden of exposure to micro-organisms and parasites in the environment[6,7] as well as by pre-exposure to the specific target pathogen. For example, a malaria vaccine that had shown good efficacy when tested in malaria-naive North Americans, induced weaker responses in endemic regions[8,9], indicating that an in-depth understanding of the interaction between our immune system and *Plasmodium* parasites is needed, not only in malaria-naive individuals but, more notably, also in those living in areas where malaria is endemic.

Immunity to malaria can develop naturally, as indicated by the lower parasite carriage and disease episodes with increasing age in endemic areas[10]. Immune responses to malaria parasites are complex, as antibodies[11,12] and a range of immune cells are thought to be involved in protection[13], but the exact contribution of cellular immunity requires further characterization[10]. Controlled human malaria infection (CHMI) trials provide exceptional opportunities to study immune responses and vaccine efficacy, as there is a clear onset of infection allowing the tracking of cause and effect[13,14]. Most CHMIs have been conducted in Europe or in the USA, where malaria-naive Europeans or European Americans and African Americans have been inoculated with *Plasmodium falciparum* sporozoites (PfSPZs) and have shown reproducible emergence, from the liver, of blood-stage parasites that can be detected by microscopy within a precise

[1]Department of Parasitology, Leiden University Center for Infectious Diseases (LU-CID), Leiden University Medical Center, Leiden, the Netherlands. [2]Department of Immunology, Leiden University Medical Center, Leiden, the Netherlands. [3]Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, the Netherlands. [4]Department of Intelligent Systems, Delft University of Technology, Delft, the Netherlands. [5]Computational Biology Center (CBC), Leiden University Medical Center, Leiden, the Netherlands. [6]Center of Medical Research Lambaréné (CERMEL), Lambaréné, Gabon. [7]Research Institute for Tropical Ecology (IRET), National Center for Scientific and Technological Research (CENAREST), Libreville, Gabon. [8]Institute of Tropical Medicine, University of Tübingen, Tübingen, Germany. [9]German Center for Infection Research, Partner Site Tübingen, Tübingen, Germany. [10]Division of Image Processing (LKEB), Department of Radiology, Leiden University Medical Center, Leiden, the Netherlands. [11]Sanaria Inc., Rockville, MD, USA. [12]Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, the Netherlands. [13]The Delft Bioinformatics Lab, Delft University of Technology, Delft, the Netherlands. [14]Division of Infectious Diseases and Tropical Medicine, Department of Medicine I, Medical University of Vienna, Vienna, Austria. [15]Present address: Institute for Immunity, Transplantation and Infection, Stanford University, Stanford, CA, USA. [16]Present address: Department of Medical Microbiology, Radboud University Medical Center, Nijmegen, the Netherlands. [17]These authors contributed equally: Vincent van Unen, Mikhael D. Manurung, Koen A. Stam. ✉e-mail: M.Yazdanbakhsh@lumc.nl

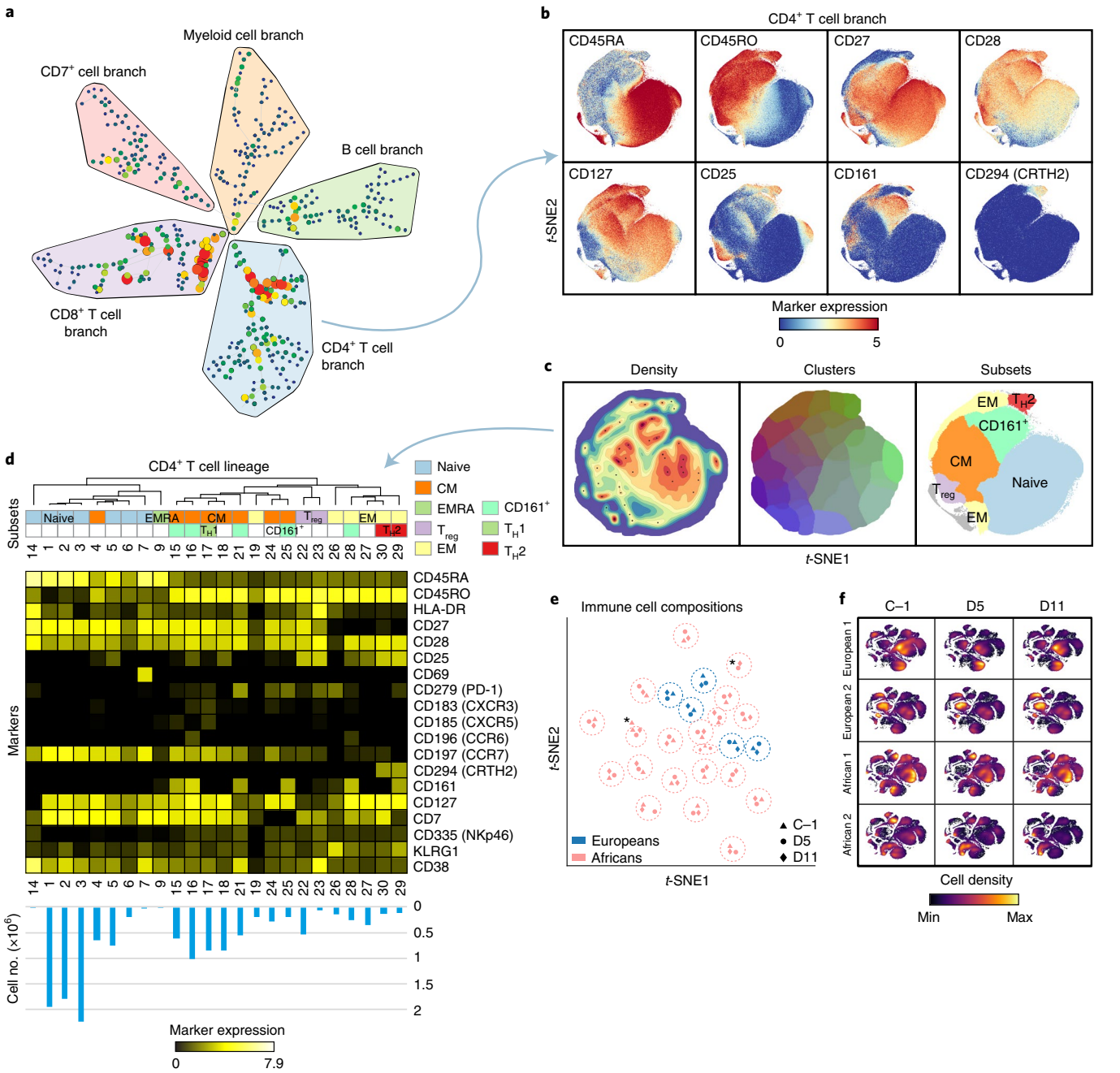


Fig. 1 | Individually unique and stable immune fingerprints revealed by mass cytometry. a, A SPADE tree of a PBMC sample after analysis of the combined sample ($n=75$) dataset containing 33.3 million cells. Size and color of the nodes are proportional to the number of clustered cells. The major immune lineages are annotated based on lineage marker expression profiles. **b**, *t*-SNE embeddings showing the collective CD4+ T cells (975,000 cells). Colors of the cells represent arcsinh-transformed (cofactor 5) expression values of indicated markers. **c**, A density map describing the local probability density of *t*-SNE-embedded CD4+ T cells, where black dots represent centroids of identified clusters using Gaussian mean-shift clustering (left). A *t*-SNE plot depicting the cluster partitions in different colors (middle) and the CD4+ T cell subset borders (right). **d**, Clusters and subsets within the CD4+ T cell lineage. A heat map summary of median expression values of markers expressed by 24 CD4+ T cell clusters and hierarchical clustering of clusters with labeling by subset name. Cell counts per cluster (bottom). **e**, *t*-SNE map showing samples per individual and time points clustered for cell frequencies. Samples with similar percentages of cell clusters relative to total cells end up close together. All 5 Europeans (blue) and 20 Africans (pink) are shown. The groups of three samples for each individual representing the three time points, 1 day before DVI with PfSPZs (C–1, triangle) and 5 days (D5, circle) and 11 days (D11, diamond) after DVI cluster together. For one individual, the C–1 sample (asterisk) is not clustering with the other two time points (D5, D11) of the same individual (also asterisk). **f**, *t*-SNE maps of PBMCs showing that each individual has their own immune fingerprint. Two European and two African individuals over time are visualized.

window of time. Currently, CHMI is also being established in malaria pre-exposed individuals in endemic areas, which will be invaluable for dissecting the relationship between baseline immune profiles, naturally acquired immunity[15] and malaria vaccine efficacy.

To date, a limited number of CHMI studies have examined cellular immune profiles during infection using flow cytometry and each of these studies could focus only on a limited number of cell subsets[9,13,16,17]. Developments in single-cell analysis by mass

cytometry[5,18], however, provide an opportunity for in-depth and broad immune profiling of responses to malaria parasites over time.

Here, we examined the immunological reactivity of malaria-naive Europeans as well as Africans with lifelong residence in a malaria-endemic area, known to exhibit naturally acquired immunity. Both Europeans and Africans were experimentally infected with *P. falciparum* in Gabon and mass cytometry was used to show, at an unprecedented depth, the detailed cellular immunological profiles at baseline, as well as the dynamics of immune responses to malaria parasites. This was complemented by determining cellular functionality through cytokine production, RNA-sequencing (RNA-seq) transcriptome analyses and by application of machine learning to identify key markers associated with naturally acquired immunity.

Results

Individually unique and stable immune fingerprints revealed by mass cytometry. We enrolled European ($n=5$) and African ($n=20$) adult volunteers[15] (Supplementary Table 1) and collected peripheral blood mononuclear cells (PBMCs) 1 day before direct venous inoculation (DVI) with nonattenuated PfSPZs and 5 days and 11 days after DVI. Through mass cytometry (Supplementary Table 2), we profiled a total of 33.3 million immune cells from 75 blood samples. As summarized in Fig. 1a, unsupervised hierarchical clustering with SPADE[19] and *t*-distributed stochastic neighbor embedding (*t*-SNE) analysis in Cytosplore (Fig. 1b)[18,20] identified distinct cell clusters (Fig. 1c), with unique marker expression profiles. Clusters were labeled with immune lineages and subsets (Fig. 1d, Extended Data Fig. 1 and Supplementary Table 3). Collectively, we were able to distinguish 198 cell clusters belonging to 45 immune subsets and 9 lineages in a data-driven fashion.

The immune cell composition of samples from the same individuals at three different time points were remarkably similar and we observed striking interindividual variation (Fig. 1e). Thus, each individual preserved their unique immune fingerprint over time (Fig. 1f).

Distinct immune signatures in Europeans and Africans at baseline. A very distinct European and African immune signature was identified at baseline. At the lineage level, (Extended Data Fig. 2), but in particular, when visualizing single cells within lineages by *t*-SNE, several unconventional αβ T cell, γδ T cell and innate lymphoid cell (ILC) subsets/clusters were present in Africans that were largely absent in Europeans (Fig. 2a). For example, subsets of $CD8^+$ natural killer (NK) T cells, γδ T cells ($CD8^+$, $CD45RA^+$, $CD127^-$) and type 2 ILCs (ILC2s) were strikingly lower in Europeans compared to Africans (Extended Data Figs. 1 and 2 and Supplementary Table 4). The *t*-SNE plots of adaptive cells reflect expansions of memory $CD4^+$ and $CD8^+$ T cell subsets and increased numbers of $CD11c^+$ B cells in Africans (Extended Data Fig. 2). The type 2 helper T (T_H2) cell (CRT_H2^+) subset, as well as $CD161^+CD4^+$ T cell subset, described to express more effector and pathogenic functions[21], are enriched in Africans. The expanded $CD11c^+$ B cells are also known as atypical memory B cells, reported to increase in settings of chronic stimulation, including malaria exposure[22].

In total, the frequencies of 45 clusters were significantly different between Africans and Europeans at baseline (Extended Data Fig. 1 and Supplementary Table 4). Overall, clusters expressing markers known to associate with a more differentiated state or with inflammation were expanded in Africans compared to Europeans, explaining the differences at the subset level. Examples for innate cells are CD56-expressing $CD8^+CD45RA^+CD127^-$ γδ T cells (clusters 60 and 61), which could represent cells with cytotoxic potential[23]; NK cells expressing CD16, but no or dim CD56 (clusters 93, 94, 96 and 99), described to be terminally differentiated NK cells[24]; and CD56-expressing monocytes (cluster 146), known to be increased during inflammatory disorders and to decrease after anti-tumor necrosis factor (TNF) therapy[25]. Regarding adaptive cells, the expanded memory $CD4^+$ T cell and B cell subsets in Africans could be accounted for by CD25 (cluster 28 and 29), CD38 and programmed cell death (PD)-1 (cluster 21)-expressing T cell clusters, as well as $CD185^+$ (CXCR5) (cluster 116) and $CD185^-CD11c^+$ B cells (cluster 136) (Extended Data Fig. 1 and Supplementary Table 4). Notably, a cluster of naive $CD8^+$ T cells associated with Africans was positive for CD161 and KLRG1 (cluster 38)[26,27]. Altogether, the higher frequency of clusters expressing activation/differentiation/pro-inflammatory markers in Africans reflects the exposure to micro-organisms and parasites that can lead to an inflammatory environment and chronic stimulation state[7,28,29].

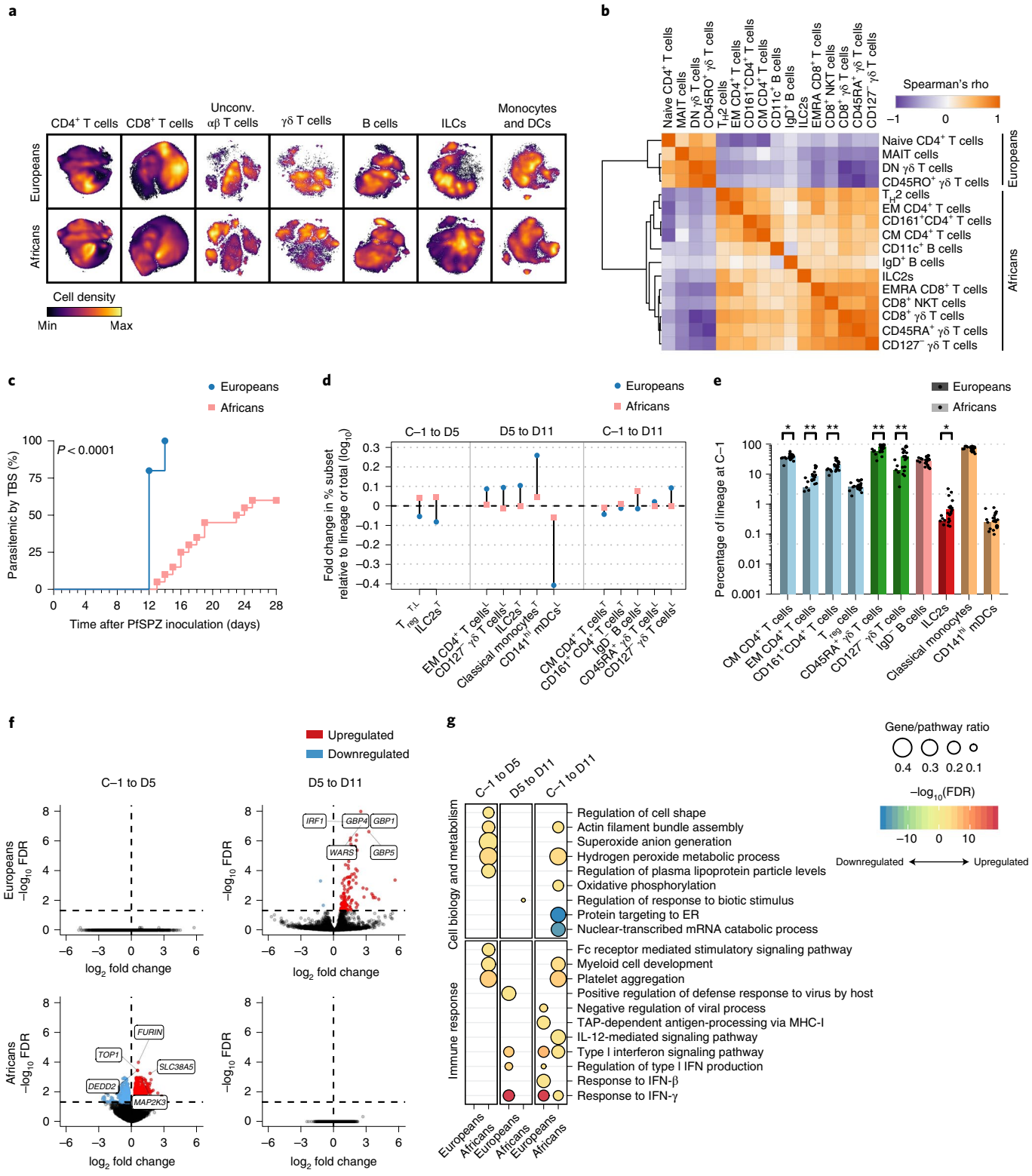
Additionally, correlations were performed on percentages of cell subsets at baseline within individuals (Fig. 2b), showing clear African and European immune profiles that reflect the large

Fig. 2 | Immune signatures of Europeans and Africans before and after *P. falciparum* sporozoite direct venous inoculation. a, *t*-SNE maps illustrating differences between Europeans ($n=5$) and Africans ($n=20$) at the single-cell level, per immune lineage. Cell density per individual map is indicated by color. **b**, Heat map showing Spearman's rank correlation coefficients (rho) for relationships between immune cell subsets that were different between Europeans ($n=5$) and Africans ($n=20$) at baseline, based on Global test within-test *P* values of ≤0.05, of all individuals. The orange block of positively correlated subsets was found in higher numbers in Europeans than Africans (top left), while high frequencies of subsets in the lower right block were associated with Africans. Orange indicates a positive correlation, whereas purple indicates a negative correlation. MAIT, mucosal-associated invariant T cell. **c**, Comparison of time to parasitemia as determined by TBS after PfSPZ DVI between Europeans ($n=5$) and Africans ($n=20$). *P* value is from log-rank (Mantel–Cox) test for survival analysis with chi-squared of 23.62 and d.f. = 1. **d**, Relative changes in response to PfSPZ DVI in frequencies of cell subsets that were different between Europeans ($n=5$) and Africans ($n=20$), between 1 day before (C −1) and 5 days (D5) or 11 days (D11) after DVI. Subset frequencies were calculated relative to total cells, as indicated with ^{T}, or relative to its lineage, as indicated with ^{L}. When the difference was significant relative to both total cells and lineage, this is indicated with ^{T,L}, and only the change relative to total was shown. Fold changes in these frequencies were calculated as 'D5/C−1', 'D11/D5' and 'D11/C−1'. Fold increase is indicated in red, whereas blue shows a decrease. All subsets with a within-test $P \leq 0.05$ for the Global test are depicted (see Supplementary Table 4 for more statistics). mDC, myeloid dendritic cell. **e**, Percentage of cell subsets relative to their lineage at baseline (C−1). Cell subsets are shown that responded differentially in Europeans ($n=5$) and Africans ($n=20$) after PfSPZ DVI as indicated in Fig. 2d. Significance is based on the within-test *P* values (Supplementary Table 4). $^*P \leq 0.05$, $^{**}P \leq 0.01$ (see also Extended Data Fig. 2c and Supplementary Table 4 for more statistics). **f**, Volcano plots of gene expression (11,659 genes) ($\log_2$ fold change versus $-\log_{10}$ adjusted *P* value) after PfSPZ DVI in Africans ($n=20$) and Europeans ($n=5$) between two time points. Top five DEGs are annotated. A false discovery rate (FDR) cutoff of 0.05 with no fold change threshold was used to define DEGs. **g**, Overrepresented Gene Ontology (GO) Biological Process pathways in Europeans ($n=5$) and Africans ($n=20$). GO testing was performed with adjustment for length bias. Pathways with Benjamini–Hochberg (BH)-adjusted *P* value <0.05 were uploaded to REViGO to reduce redundancy of pathway terms. SimRel semantic similarity measure was used with small (0.5) allowed similarity. The categorization of pathways into either immune response or cell biology and metabolism categories was performed manually. Gene/pathway ratio was defined as the proportion of DEGs present in a pathway divided by the total number of genes in a pathway. Overrepresentation analysis was equivalent to a one-sided Fisher's exact test. ER, endoplasmic reticulum; MHC, major histocompatibility complex.

phenotypic differences across both adaptive and innate immune compartments.

Differential dynamic changes in immune responses of Europeans and Africans upon PfSPZ DVI. Upon PfSPZ DVI, all Europeans and 12 of 20 Africans (60%) developed parasitemia detectable by microscopy of thick blood smear (TBS) within the 28-d study period (Fig. 2c, Extended Data Fig. 3 and Supplementary Table 1)[15]. Europeans exhibited blood-stage parasitemia within 12–14 d, whereas Africans who became TBS$^+$ did so at later time points (geo-mean 17.9 d, range 13–25 d), indicating varying degrees of immunity to liver and blood-stage parasites.

Differences between Europeans and Africans in response to PfSPZs over time was marked by an increase in Africans, in regulatory T (T_{reg}) cells and ILC2s seen on day 5 (D5), while in Europeans, an increase occurred later, between D5 and day 11 (D11) after DVI (Fig. 2d). The ILC2, effector memory (EM) CD4$^+$ T cell and CD127$^-$ (interleukin (IL)-7Rα) γδ T cell subsets, which were of



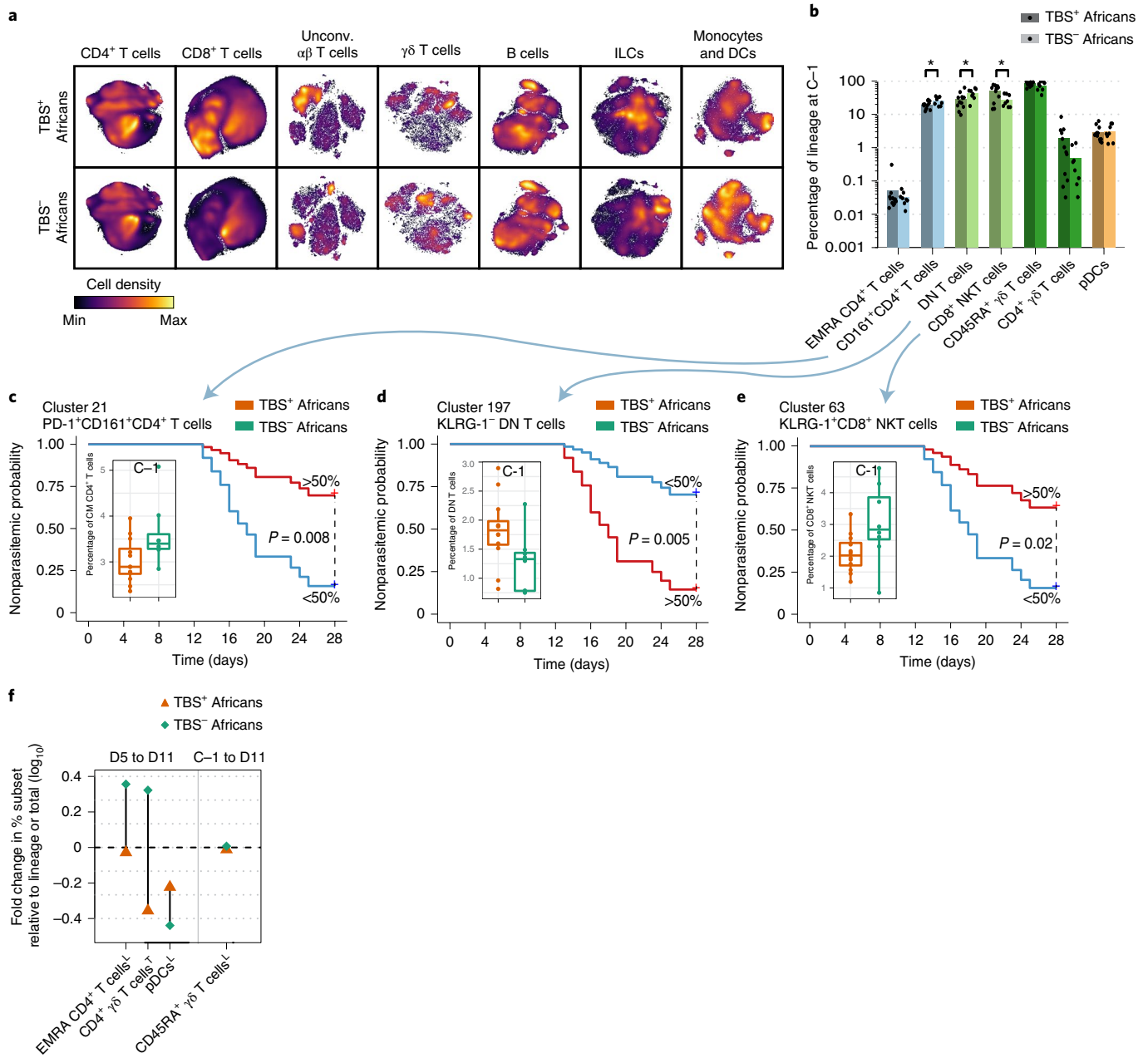


Fig. 3 | High-resolution immune signatures at baseline and following PfSPZ DVI in Africans related to parasitemia control. a, *t*-SNE maps illustrating the phenotypic differences between TBS$^+$ ($n=12$) and TBS$^-$ ($n=8$) Africans at the single-cell level, per lineage. Cell density per individual map is indicated by color. **b**, Percentage of cell subsets relative to their lineage at baseline (C−1). Cell subsets are shown that responded differentially in TBS$^+$ ($n=12$) and TBS$^-$ ($n=8$) Africans before or after (Fig. 3f) PfSPZ DVI and cell subsets that differ at baseline. Significance is based on Global test within-test *P* values (see Supplementary Table 4 for more statistics). *$P \leq 0.05$. **c**, Survival graph showing time until Africans ($n=20$) develop parasitemia (TBS) according to the frequency of PD-1$^+$CD161$^+$ cluster 21 relative to central memory (CM) CD4$^+$ T cells at baseline (C−1). The frequency was split on the median to create two groups of Africans, a top >50% (red), with the Africans who had a high frequency of cluster 21, and bottom <50% (blue). This grouping was used as the coefficient in a univariate Cox regression; the shown *P* value is based on the score (log-rank) test, with likelihood ratio test of 6.94, d.f. = 1, $n=20$, 12 events. The box plot shows the median, first and third quartiles of the respective cluster for the TBS$^+$ Africans ($n=12$) and TBS$^-$ Africans ($n=8$). Whiskers extend to the maximum/minimum of the respective groups, no further than 1.5 × interquartile range (IQR). All points are added to the box plot. **d**, Survival graph similar to **c**, but showing the time until Africans develop parasitemia according to the frequency of KLRG1$^+$ cluster 197 relative to DN T cells at baseline (C−1). Likelihood ratio test of 7.58, d.f. = 1, $n=20$, 12 events. **e**, Survival graph similar to **c**, but showing the time until Africans develop parasitemia according to the frequency of KLRG1$^+$ cluster 63 relative to CD8$^+$ NKT cells at baseline (C−1). Likelihood ratio test of 5.39, d.f. = 1, $n=20$, 12 events. **f**, Relative changes in response to PfSPZ DVI in frequencies of cell subsets that were different between TBS$^+$ ($n=12$) and TBS$^-$ ($n=8$) Africans. Comparisons between baseline (C−1) and D5 or D11 after DVI. All subsets with a within-test *P* value ≤0.05 for the Global test are shown (Supplementary Table 4). See Extended Data Fig. 4 for individual data points.

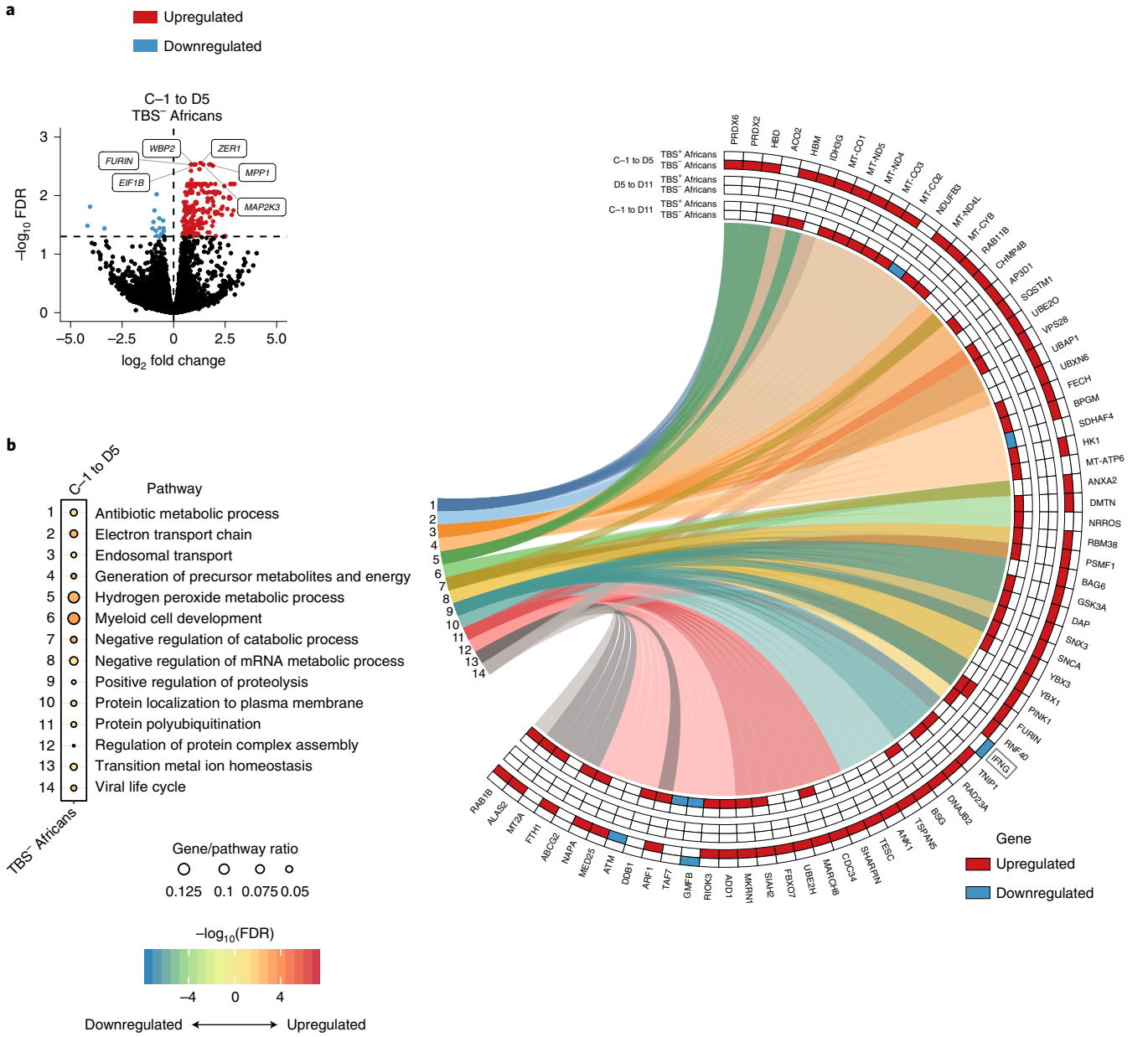


Fig. 4 | Gene expression in blood of Africans after PfSPZ DVI. a, Volcano plots of gene expression (log_2 fold change versus $-log_{10}$ adjusted *P* value) after PfSPZ DVI in TBS⁻ Africans ($n=8$) between baseline (C−1) and D5. Top five DEGs are annotated. BH-adjusted and two-sided *P* value <0.05 without fold change. **b**, Overrepresented GO Biological Process pathways in TBS⁺ ($n=12$) and TBS⁻ ($n=8$) Africans together with a circular graph depicting the genes that were differentially up- or downregulated between the TBS⁺ and TBS⁻ Africans between two time points.

lower frequency in Europeans at baseline (C−1, 1 day before DVI) (Fig. 2e and Extended Data Fig. 2), increased at this later time point. In the same period, classical monocytes showed a stronger increase, whereas CD141^{hi} (BDCA3) myeloid dendritic cells showed a much stronger decrease in frequency in Europeans than in Africans (Fig. 2d and Supplementary Table 4).

RNA-seq analysis of the whole blood transcriptome from parallel samples showed that PfSPZ DVI resulted in 473 differentially expressed genes (DEGs) over time unique to Africans and 108 DEGs unique to Europeans, with only 15 DEGs overlapping between the groups. The changes in gene expression occurred predominantly between C−1 and D5 in Africans and between D5 and D11 in Europeans (Fig. 2f), in line with mass cytometry data. Pathway overrepresentation analysis emphasized the distinctiveness of responses to PfSPZ DVI between Africans and Europeans (Extended Data Fig. 2). In Africans, mostly pathways associated with cell biology and metabolism (cellular shape regulation, actin filaments, oxidative phosphorylation) or with immune pathways (Fc receptor and platelet activation) were upregulated, whereas in Europeans, the immune pathways that were highly upregulated involved responses to interferons (Fig. 2g).

Further insight into how the immune system responds to *P. falciparum* was provided by identification of 23 clusters, through mass cytometry, which changed differentially in Europeans and Africans after PfSPZ DVI (Extended Data Fig. 1 and Supplementary Table 4). For example, malaria infection leads to the activation of CD161$^+$CD4$^+$ T cells evident from the increase in CD25$^+$ cells (cluster 16) in Europeans, in line with the higher frequencies of CD161$^+$CD4$^+$ T cells in pre-exposed Africans at baseline (Fig. 2e). Similarly, the increase in CD45RA$^+$ and CD127$^-$ γδ T cell subsets

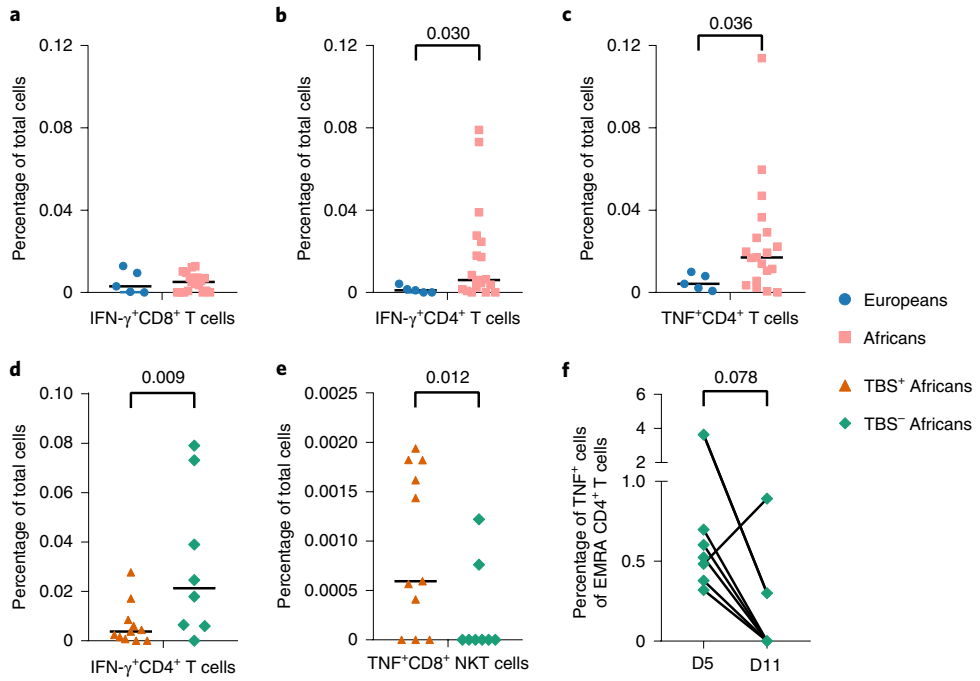


Fig. 5 | Cytokine production in response to *Plasmodium falciparum*–infected red blood cells. a, Baseline frequencies of PfRBC-specific IFN-γ response by CD8+ T cells to PfRBC stimulation in Europeans ($n=5$) and Africans ($n=20$). All cytokine response data have been subtracted with background response to uninfected red blood cells. Median of each group is displayed as horizontal line. **b**, Baseline frequencies of the PfRBC-specific IFN-γ response by CD4+ T cells in Europeans and Africans. *P* value from two-sided Wilcoxon rank-sum test with $W=17$. **c**, Baseline frequencies of PfRBC-specific TNF response by CD4+ T cells in Europeans and Africans. *P* value from two-sided Wilcoxon rank-sum test with $W=18$. **d**, Baseline frequencies of PfRBC-specific IFN-γ response by CD4+ T cells in TBS+ and TBS− Africans. *P* value from two-sided Wilcoxon rank-sum test with $W=106$. **e**, Baseline frequencies of PfRBC-specific TNF response by CD8+ NKT cells in TBS+ and TBS− Africans. *P* value from two-sided Wilcoxon rank-sum test with $W=23$. **f**, Changes in frequency of TNF-producing terminally differentiated EMRA CD4+ T cells in TBS− Africans from D5 to D11 after PfSPZ DVI. *P* value from two-sided Wilcoxon signed-rank test with $W=-22$.

in Europeans could be accounted for by an increase in the CD16+CD161+ cluster 70, which at baseline was higher in Africans. Some changes at the cluster level, not reflected at the subset level, were striking, exemplified by a strong decrease in NK cell cluster 107 (CD16−KLRG1−) in Europeans. This could reflect the relocation of these cells to tissues or their differentiation, in line with NK cells being one of the first innate cells that respond to malaria parasites and help activate adaptive responses[30]. It was also interesting that NK cell cluster 107 was already significantly lower in peripheral blood of pre-exposed Africans. Therefore, in malaria-naive individuals, *P. falciparum* infection seems to drive an immune profile that resembles the one seen at baseline in pre-exposed Africans.

High-resolution immune signatures at baseline and following PfSPZ DVI in Africans related to control of parasitemia. In contrast to Europeans, some Africans were able to control parasitemia following PfSPZ DVI, remaining negative by microscopy (TBS− Africans), demonstrating strong naturally acquired immunity. Although the Africans who became parasitemic (TBS+ Africans) did so largely at later time points than naive Europeans, which indicates some degree of immunity (Fig. 2c), their inability to control their infection distinguishes them from TBS− individuals. The visualization of immune cells by *t*-SNE shows distinct patterns between TBS− and TBS+ Africans (Fig. 3a). Indeed, at the subset level, TBS− Africans were characterized by fewer CD8+ NKT cells, but more CD161+CD4+ T cells and double-negative (DN) T cells than TBS+ Africans (Fig. 3b and Supplementary Table 4).

More in-depth profiling of cell populations revealed differences in 18 clusters (Extended Data Fig. 1 and Supplementary Table 4), 6 of which explained the observations at the subset level. Thus, in TBS− Africans, the higher percentage of CD161+CD4+ T cells could be explained by expansion of cells expressing PD-1 (CD279) (cluster 21), whereas the expanded DN T cells seem depleted of clusters that show little expression of differentiation/activation markers (clusters 10 and 197). Moreover, the CD8+ NKT cell subset that was lower in peripheral blood of TBS− Africans and thus might be residing in tissues or secondary lymphoid organs, consists of an increase in clusters expressing KLRG1 (cluster 63, 75 and 103), which has been reported to be a marker for long-lived invariant NKT cells[31]. Together, these results indicate that cells in a more activated/differentiated state associate with parasite control. Kaplan–Meier plots show that higher frequencies of the PD-1+CD161+ cluster 21 (Fig. 3c), lower frequencies of cluster 197 (with virtually no other activation/differentiation markers) (Fig. 3d) and higher frequencies of the KLRG1+ cluster 63 (Fig. 3e) at baseline were associated with a higher probability of being TBS−.

Following PfSPZ DVI, four cell subsets changed differentially in Africans who controlled their parasites (TBS−) and those who developed parasitemia (TBS+) (Fig. 3f, Extended Data Fig. 4 and Supplementary Table 4). In TBS− Africans, an increase was seen in terminally differentiated effector memory (EMRA) CD4+ T cells as well as in γδ T cells, which would be in line with the possibility that these cells reside in peripheral organs, poised to respond rapidly to infection and thereafter move out into peripheral blood. In parallel, a marked decrease was seen in the percentage of plasmacytoid dendritic cells (pDCs), which, by moving out of the blood into the tissue and/or secondary lymphoid organs, can elicit further T cell responses.

Analyzed at the cluster level, changes in 12 clusters were seen (Supplementary Table 4). The differential change in γδ T cells,

when comparing before and 11 d after DVI, might be due to cluster 12 (of CD127$^+$ γδ T cells), which notably expresses KLRG1 and increase in TBS$^-$ Africans compared to a decrease in TBS$^+$ Africans (Extended Data Fig. 1 and Supplementary Table 4). Additional analysis of this cluster with a panel containing Vδ2, showed that this cluster, as well as γδ T cell cluster 79, which was associated with protection in TBS$^-$ Africans at baseline, seem to encompass Vδ2$^+$ γδ T cells (Extended Data Fig. 5). Vδ2$^+$ γδ T cells have been associated with protection, but can also become dysfunctional with repeated exposure resulting in clinical tolerance to malaria[32]. Furthermore, in TBS$^-$ Africans, cluster 187, which comprises CCR7-expressing (CD197) cells within pDCs, decreased between C−1 and D5, whereas cluster 26, expressing PD-1 and KLRG1 within EM CD4$^+$ T cells, increased later, between D5 and D11, which could indicate activation of pDCs that leave the peripheral blood to contribute to stimulation of EM CD4$^+$ T cells to expand and appear in peripheral blood.

The CD8$^+$ NKT cell and DN T cell subsets, which at baseline distinguished TBS$^-$ from TBS$^+$ Africans, changed in frequency. CD8$^+$ NKT cells, lower in frequency at baseline, increased more strongly ($P=0.091$) in peripheral blood of TBS$^-$ Africans between day 5 and 11, suggesting that these cells reside in the liver[33] or secondary lymphoid organs and respond rapidly to malaria infection and thereafter leave to enter peripheral blood. Within the DN T cell subset, which was higher at baseline, the KLRG1-expressing cluster 66 (KLRG1$^+$ DN T cells) increased in TBS$^-$ Africans (Supplementary Table 4) after PfSPZ DVI within the first 5 d, indicating an early response in those able to control parasitemia.

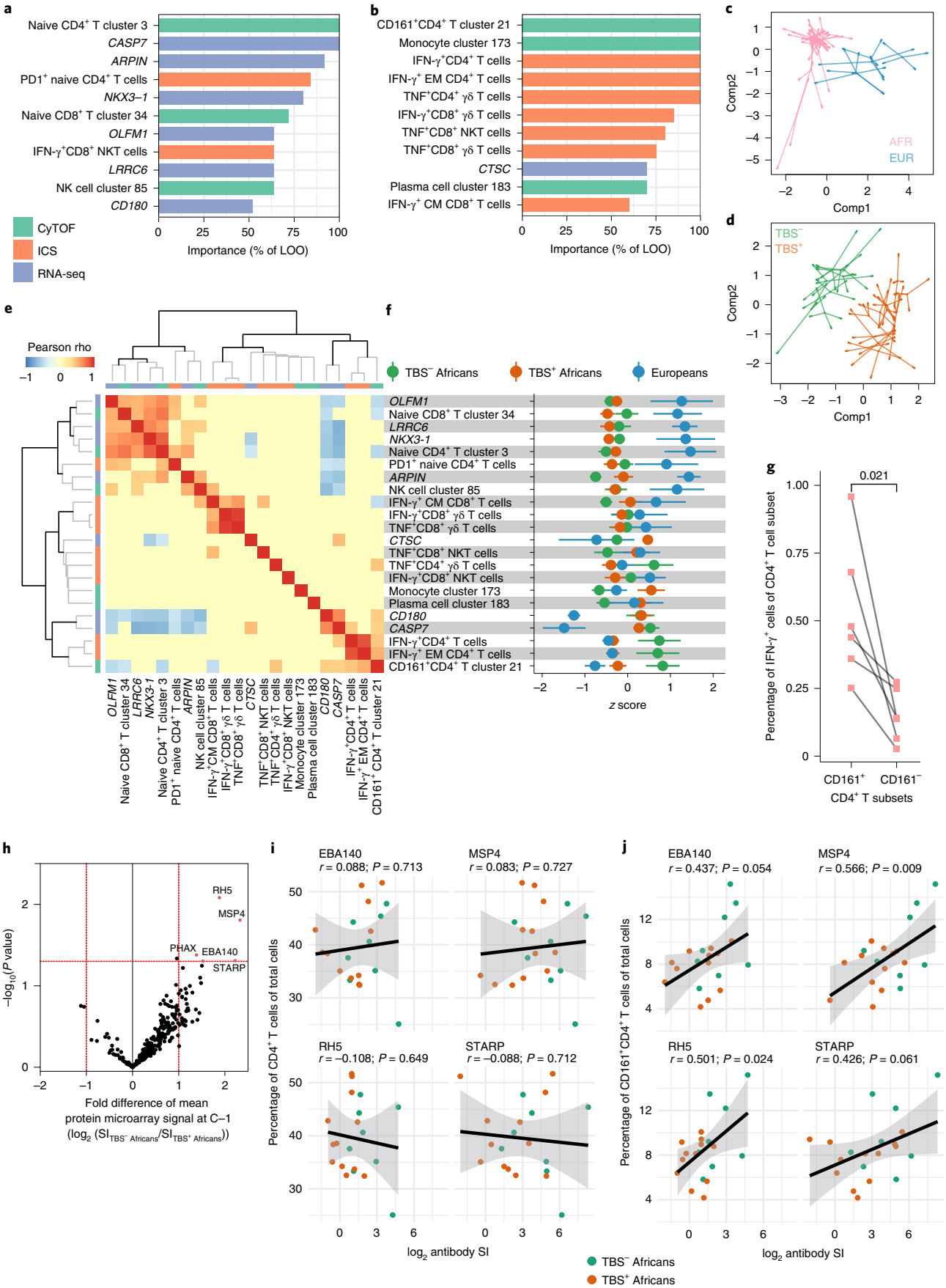
The RNA-seq data showed that PfSPZ DVI resulted in 236 DEGs over time unique to TBS− Africans and only 42 DEGs unique to TBS$^+$ Africans; thus, changes in gene expression occurred predominantly in TBS$^-$ Africans (Fig. 4a). The pathways enriched in TBS$^-$ Africans (Fig. 4b) reflected activation (cellular metabolism or response to intracellular organisms), but also those that could explain the strong changes seen in pDCs relative to the monocyte and DC lineage (myeloid cell development) (Fig. 3f at subset level and Supplementary Table 4 for cluster 187). Gene set enrichment analysis identified two out of eight interferon (IFN)-γ-related pathways (GO 0060334, Bonferroni-adjusted $P=0.015$; GO 0060330, Bonferroni-adjusted $P=0.015$) in TBS$^-$ Africans at day 5. The few DEGs in TBS$^+$ Africans were not enriched for any identifiable pathway but a decrease in *IFNG* was seen in this group from C−1 to D5 (Fig. 4b).

Taken together, by using CHMI, it has been possible to associate subsets and clusters identified as CD4$^+$ T cells, in particular, EM and CD161$^+$CD4$^+$ T cells, NKT cells, DN T cells and γδ T cells as well as pDCs, with naturally acquired immunity. Notably, these cells often exhibited differentiation/activation markers that might portray stronger effector responses for control of parasitemia in Africans with lifelong residence in malaria-endemic regions. In line with this, the analysis of the transcriptome revealed changes in a set of genes indicative of early cellular activation in TBS$^-$ Africans only.

Cytokine production in response to *Plasmodium falciparum*–infected red blood cells. To test immune cell functionality, PBMCs were stimulated with *P. falciparum*-infected red blood cells (PfRBCs) and cytokine-producing CD4$^+$ T cells, CD8$^+$ T cells, DN T cells, NKT cells and γδ T cells were analyzed by flow cytometry (Extended Data Fig. 6 and Supplementary Table 5). IL-17 and IL-2 responses, while detectable in response to Staphylococcal enterotoxin B, were negligible in response to PfRBCs. PfRBC-specific IFN-γ and TNF responses were readily observed (Extended Data Fig. 6). The highest proportion of cells that produced TNF and IFN-γ were γδ T cells and NKT cells in both European and Africans, indicating that there can be functional consequences to the associations seen with these cells and varying degrees of immunity to *P. falciparum*. Even though innate immune cells showed the largest cytokine responses to PfRBCs, an adaptive immune response by CD4$^+$ T cells was clearly measurable, whereas we were unable to detect antigen-specific cytokine-producing CD8$^+$ T cells (Fig. 5a). The CD4$^+$ T cell response to *P. falciparum* antigen was characterized by higher IFN-γ (Fig. 5b) and TNF (Fig. 5c) responses in Africans compared to Europeans, with highest responses seen in TBS$^-$ Africans, who control their infection (Fig. 5d). It was also interesting to note that in line with the mass cytometry data, lower frequencies of TNF$^+$CD8$^+$ NKT cells were found in TBS$^-$ Africans at baseline (Fig. 5e).

Considering responses over time, EMRA CD4$^+$ T cells, which increased in frequency between D5 and D11 in TBS$^-$ Africans, showed no increase, but a decrease ($P=0.078$) in TNF response

Fig. 6 | Integrative data analysis. a, Bar charts showing consensus features included in DIABLO machine learning to classify Africans versus Europeans. The most important CyTOF clusters (green), genes (RNA-seq, blue) and cellular responses (intracellular cytokine staining (ICS), red) for discriminating between Africans and Europeans at baseline are shown. Features were selected for every fold in cross-validation, and only features retained in at least 50% of folds are depicted. **b**, Bar charts showing the consensus features included in the DIABLO machine learning to classify TBS$^+$ and TBS$^-$ Africans. The most important CyTOF clusters (green), genes (RNA-seq, blue) and cellular responses (ICS, red) for discriminating between TBS$^+$ and TBS$^-$ Africans at baseline are shown. Features were selected for every fold in cross-validation, and only features retained in at least 50% of folds are depicted. **c**, Arrow plots showing projection onto the latent space of the full model for all datasets, including outcome. For each individual, a centroid is shown, with arrows pointing to where the CyTOF, RNA, ICS and outcome are projected for that individual. Shorter arrows indicate better alignment between datasets. Africans and Europeans are depicted in pink and blue, respectively. Two components were used for the latent space. **d**, Arrow plots showing the projection onto the latent space of the full model for all datasets, including outcome. For each individual, a centroid is shown, with arrows pointing to where the CyTOF, RNA, ICS and outcome are projected for that individual. TBS$^+$ and TBS$^-$ Africans are depicted in green and orange, respectively. **e**, Correlation matrix of baseline levels between all 22 scaled features that were most strongly associated with either ethnicity or TBS outcome. Nonsignificant correlations (Pearson test) were set to 0. Hierarchical clustering was performed on the Euclidean distance using complete linkage. **f**, Baseline levels of important genes, CyTOF clusters and cellular responses per group. Mean and s.e.m. are shown for *z*-scaled levels for Europeans (blue), TBS$^+$ Africans (orange) and TBS$^-$ Africans (green). **g**, Differences in frequency of the PfRBC-specific IFN-γ response between CD161$^-$CD4$^+$ T cells and CD161$^+$CD4$^+$ T cells of Africans ($n=6$) from this study. Cytokine response data have been subtracted with background response to uninfected red blood cells. Statistical analysis was performed with a Student's *t*-test for paired samples. *P* value from two-sided paired Student's *t*-test, with $T=3.316$ and d.f. $=5$. **h**, Antibody reactivity of TBS$^+$ and TBS$^-$ African individuals to *Plasmodium* protein microarray, showing the reactivities at baseline (C−1) to five antigens associated with TBS$^-$ Africans (*P* value <0.05 and a fold change >2 of mean signal intensity (SI), annotated). Statistics are based on two-sided Welch-corrected Student's *t*-test. **i**, Correlation between antibody reactivity and abundance of total CD4$^+$ T cells at C−1 for TBS$^+$ ($n=12$) and TBS$^-$ ($n=8$) Africans. The *y* axis represents cell subset at baseline as a percentage of total cells. The *x* axis represents relative binding antibody value to the measured antigens. Both the Pearson correlation and its corresponding two-sided *P* value are reported for each antigen, without adjustments for multiple comparisons. The black line represents a fitted linear model ($y \sim x$), and the shaded gray area is the 95% confidence interval. **j**, Correlation as in **i**, between antibody reactivity and abundance of CD161$^+$CD4$^+$ T cells.


a
Naive CD4+ T cluster 3
CASP7
ARPIN
PD1+ naive CD4+ T cells
NKX3–1
Naive CD8+ T cluster 34
OLFM1
IFN-γ+CD8+ NKT cells
LRRC6
NK cell cluster 85
CD180
CyTOF
ICS
RNA-seq
0 25 50 75 100
Importance (% of LOO)
b
CD161+CD4+ T cluster 21
Monocyte cluster 173
IFN-γ+CD4+ T cells
IFN-γ+ EM CD4+ T cells
TNF+CD4+ γδ T cells
IFN-γ+CD8+ γδ T cells
TNF+CD8+ NKT cells
TNF+CD8+ γδ T cells
CTSC
Plasma cell cluster 183
IFN-γ+ CM CD8+ T cells
0 25 50 75 100
Importance (% of LOO)
c
Comp2
Comp1
AFR
EUR
d
TBS−
TBS+
Comp2
Comp1
e
Pearson rho
−1 0 1
OLFM1
Naive CD8+ T cluster 34
LRRC6
NKX3-1
Naive CD4+ T cluster 3
PD1+ naive CD4+ T cells
ARPIN
NK cell cluster 85
IFN-γ+CM CD8+ T cells
IFN-γ+CD8+ γδ T cells
TNF+CD8+ γδ T cells
CTSC
TNF+CD8+ NKT cells
TNF+CD4+ γδ T cells
IFN-γ+CD8+ NKT cells
Monocyte cluster 173
Plasma cell cluster 183
CD180
CASP7
IFN-γ+CD4+ T cells
IFN-γ+ EM CD4+ T cells
CD161+CD4+ T cluster 21
f
TBS− Africans
TBS+ Africans
Europeans
z score
g
0.021
Percentage of IFN-γ+ cells of CD4+ T cell subset
CD161+
CD161−
CD4+ T subsets
h
RH5
MSP4
PHAX
EBA140
STARP
−log10(P value)
Fold difference of mean protein microarray signal at C–1 (log2 (SITBS− Africans/SITBS+ Africans))
i
EBA140
r = 0.088; P = 0.713
MSP4
r = 0.083; P = 0.727
RH5
r = −0.108; P = 0.649
STARP
r = −0.088; P = 0.712
Percentage of CD4+ T cells of total cells
log2 antibody SI
j
EBA140
r = 0.437; P = 0.054
MSP4
r = 0.566; P = 0.009
RH5
r = 0.501; P = 0.024
STARP
r = 0.426; P = 0.061
Percentage of CD161+CD4+ T cells of total cells
log2 antibody SI
TBS− Africans
TBS+ Africans

from D5 to D11 (Fig. 5f). This could reflect a level of hyporesponsiveness following their earlier activation upon infection. A similar dysfunction has been described upon repeated malaria infection in children[34,35].

Altogether, through examining the cytokine production in response to *P. falciparum*-infected red blood cells, we were able to show the functionality of cell types, identified by mass cytometry, to be associated with naturally acquired immunity to *P. falciparum*.

Multi-omics implicates IFN-γ-producing CD161$^+$CD4$^+$ T cells in natural protection against malaria infection. To understand which cell clusters, genes and antigen-specific responses were most strongly associated with naturally acquired resistance to malaria, the machine-learning algorithm DIABLO was used to classify individuals and identify discriminant features (see Extended Data Fig. 7 and Methods for details on model selection and performance)[36]. Feature selection was performed using Lasso-like regularization for each fold in the cross-validation and consensus features present in at least half the folds were retained. At baseline, comparing Europeans and Africans, three CyTOF clusters, two cellular responses and six genes were consistently included in the model (Fig. 6a). This encompassed the naive CD4$^+$ T cell cluster 3, the naive CD8$^+$ T cell cluster 34, the NKT cell cluster 85 and the genes *CASP7*, encoding for caspase-7, and *CD180*. We then assessed which features were most important for predicting naturally acquired resistance to malaria by comparing TBS$^+$ and TBS$^-$ Africans at baseline (Fig. 6b). The CD161$^+$CD4$^+$ T cell cluster 21 and IFN-γ production by total CD4$^+$ T cells and EM CD4$^+$ T cells were retained in all folds. In addition, cluster 173 of nonclassical monocytes, the gene encoding cathepsin-C (*CTSC*) and cytokine production by CD8$^+$ γδ T cells and NKT cells were also consistently selected.

The three datasets and outcomes aligned well in the latent space (Fig. 6c,d) and we therefore correlated baseline levels of the predictive genes, CyTOF clusters and cellular responses (Fig. 6e). This revealed three groups of features, one of which was associated with Europeans and consisted of naive CD4$^+$ and CD8$^+$ T cell clusters and the NKT cell cluster 85, as well as genes encoding for Homeobox protein Nkx-3.1 (*NKX3-1*), actin-related protein 2/3 complex inhibitor (*ARPIN*), olfactomedin 1 (*OLFM1*) and leucine-rich repeat containing 6 (*LRRC6*) (Fig. 6e,f).

A second cluster associated with resistance to parasitemia consisted of the protective CD161$^+$CD4$^+$ T cell cluster 21 and IFN-γ producing (EM) CD4$^+$ T cells (Fig. 6e,f). Indeed, CD161$^+$CD4$^+$ T cells showed an increased capacity to produce IFN-γ upon PfRBC stimulation compared to CD161$^-$CD4$^+$ T cells (Fig. 6g). Although the genes *CASP7* and *CD180* clustered together with CD161$^+$CD4$^+$ T cells and IFN-γ production, these genes were increased in Africans over Europeans but not associated with resistance to parasitemia (Fig. 6e,f).

The identification of CD161$^+$CD4$^+$ T cells as a prominent feature associated with control of parasitemia, prompted us to study its further functional relevance. To this end, a protein array with 228 unique antigens of *P. falciparum* was probed by sera from study participants and antibody reactivity, at baseline, to five proteins was identified to associate with parasite control in TBS$^-$ group (Fig. 6h). While CD4$^+$ T cells did not correlate (Fig. 6i), frequencies of CD161$^+$CD4$^+$ T cells at baseline showed a correlation with two of these antibody reactivities and a trend for two more (Fig. 6j). This indicates a possible mechanism through which CD161$^+$CD4$^+$ T cells can play a role in controlling parasitemia.

Taken together, a machine-learning approach was able to identify a minimal signature of features before inoculation that was predictive of naturally acquired protection against *P. falciparum* upon CHMI, which included CD161$^+$CD4$^+$ T cells, as well as CD4$^+$ T cell IFN-γ production upon stimulation. Although validation in an independent cohort is required, this signature could help toward the development of vaccines with high protective potential in endemic areas.

Discussion

Mass cytometry revealed distinct European and African immune signatures marked by enrichment of memory cells and expression of activation/differentiation markers such as CD25, CD161, KLRG1 and PD-1, not only on adaptive but also on innate immune cells of Africans. These distinct immunological patterns are likely a reflection of variation in the burden of exposure to micro-organisms and parasites in Europe and Africa.

By collecting samples 5 days after PfSPZ DVI, we were able to observe a rapid increase in T_{reg} cells as well as ILC2s in Africans, whereas in Europeans a later increase was seen in a number of adaptive and innate cells in peripheral blood. Transcriptomics of parallel whole blood samples supported the rapid response to malaria in Africans and a delayed response in Europeans, but also highlighted the very different pathways that were activated in these two groups. For example, 'myeloid development' or 'platelet aggregation' typified the immunological pathways activated early in Africans. An early increase in platelet-activation-pathway genes has been reported recently in Tanzanian adults undergoing malaria challenge[37] and might represent a response to antibodies present in pre-exposed individuals. Interestingly, a delayed but strong type-I IFN signature was seen in naive Europeans. The stronger increase between day 5 and 11 in γδ T cells in Europeans, in line with earlier studies of CHMI in malaria-naive individuals[14], as well as the stronger increase in EM CD4$^+$ T cells and classical monocytes in these volunteers is consistent with the ability of type-I IFN to lead to activation and recruitment of a network of cells that can mediate malaria-induced inflammation[38,39]. In this regard, the early increase in T_{reg} cells or ILC2s seen in Africans might represent the initiation of a distinct response; one that is anti-inflammatory, not present in Europeans.

Comparison of African volunteers that were susceptible to infection (TBS$^+$) with those that controlled parasitemia (TBS$^-$), demonstrated that naturally acquired immunity was associated with a cellular profile involving subsets and clusters within CD4$^+$ T cells, NKT cells, DN T cells and γδ T cells as well as pDCs. These cells often exhibited differentiation/activation markers, which might indicate stronger effector responses capable of controlling parasitemia. In particular, a high frequency of CD161$^+$CD4$^+$ T cells was a notable feature of the baseline immune profile associated with the control of *P. falciparum*. This parallels a stronger cytokine response to *Plasmodium* antigen by CD4$^+$ T cells of the TBS$^-$ group, reflected in the higher frequency of antigen stimulated CD161$^+$CD4$^+$ rather than CD161$^-$CD4$^+$ T cells that produced IFN-γ. RNA-seq analysis of samples from TBS$^-$ Africans showed a rapid alteration of transcriptional profiles upon DVI, mirroring the strong changes seen in pDCs and CD4$^+$ T cells by CyTOF. In addition, a rapid increase in IFN-γ-related pathways was observed early after PfSPZ DVI in protected participants only. CD4$^+$ T cell responses have been associated with protection when chemo-attenuated malaria parasites were used to vaccinate European volunteers[14]. Our study identifies the CD161$^+$ subset of CD4$^+$ T cells as a correlate of protection in individuals with lifelong malaria exposure. One mechanism through which CD161$^+$CD4$^+$ T cells might contribute to parasite control, could be through promoting antibody responses. Using malaria parasite protein arrays, we identified antigen-specific antibody reactivities associated with parasite control; these antibodies were correlated with CD161$^+$CD4$^+$ T cells.

The role of innate immune cells in immune memory is of increasing interest now that there is considerable evidence for trained immunity[31,40]. Although there are reports of NKT cells that enhance immunity to malaria infection in animal models[41], we identified CD8$^+$ NKT cells expressing KLRG1, as players in naturally acquired

immunity in humans. In addition and vδ2 γδ T cells were found to be associated with stronger parasite control. Finally, γδ T cell subset and clusters expressing KLRG1 increased 11 d after PfSPZ DVI only in the TBS⁻ group. Given the residence of γδ T cells in the liver[42] and their importance in parasite clearance[42,43], our data support the hypothesis that γδ T cells activated by *P. falciparum* traffic from the liver as part of effective immune responses in individuals with naturally acquired immunity.

Integrative machine learning identified genes, cell types and responses strongly associated with immunity to malaria. It revealed that expression of the genes *CASP7* and *NKX3-1* were altered in Africans relative to Europeans. These genes are involved in signaling of p53, which has recently been identified to play a role in malaria-induced inflammation[44–46]. Integrative machine learning also provided further evidence that CD161⁺CD4⁺ T cells producing IFN-γ are a correlate of naturally acquired protection against malaria infection. To date, malaria vaccines aim to induce a strong memory T cell response, based on compelling evidence for the importance of CD8⁺ T cells, which can be helped by CD4⁺ T cells[47]. Our findings regarding the role of CD161⁺CD4⁺ T cells, along with emerging evidence for their strong effector functions in infection and inflammatory diseases[21], could be a starting point for the development of vaccines inducing strong protective responses in individuals residing in malaria-endemic areas.

In conclusion, by combining CHMI with high-dimensional single-cell technology, RNA-seq, functional analysis and data integration, we created a detailed map of responses in naive European and pre-exposed African volunteers with varying degrees of naturally acquired immunity and identified rapid responses that associate with control of parasitemia. We also provide a dataset repository that can help design independent studies to confirm the cellular responses that can be harnessed against malaria.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41590-021-00911-7.

Received: 12 February 2020; Accepted: 2 March 2021;
Published online: 22 April 2021

References

1. WHO. *World Malaria Report 2019*, report no. 978-92-4-156572-1 (WHO, 2019).
2. McCall, M. B. B., Kremsner, P. G. & Mordmuller, B. Correlating efficacy and immunogenicity in malaria vaccine trials. *Semin. Immunol.* **39**, 52–64 (2018).
3. Duffy, P. E. & Patrick Gorres, J. Malaria vaccines since 2000: progress, priorities, products. *NPJ Vaccines* **5**, 48 (2020).
4. Jongo, S. A. et al. Immunogenicity and protective efficacy of radiation-attenuated and chemo-attenuated PfSPZ vaccines in Equatoguinean adults. *Am. J. Trop. Med. Hyg.* **104**, 283–293 (2021).
5. Brodin, P. et al. Variation in the human immune system is largely driven by non-heritable influences. *Cell* **160**, 37–47 (2015).
6. Hur, Y. G. et al. Factors affecting immunogenicity of BCG in infants, a study in Malawi, The Gambia and the UK. *BMC Infect. Dis.* **14**, 184 (2014).
7. Muyanja, E. et al. Immune activation alters cellular and humoral responses to yellow fever 17D vaccine. *J. Clin. Invest.* **124**, 3147–3158 (2014).
8. Jongo, S. A. et al. Safety, immunogenicity, and protective efficacy against controlled human malaria infection of *Plasmodium falciparum* sporozoite vaccine in Tanzanian adults. *Am. J. Trop. Med. Hyg.* **99**, 338–349 (2018).
9. Obiero, J. M. et al. Impact of malaria preexposure on antiparasite cellular and humoral immune responses after controlled human malaria infection. *Infect. Immun.* **83**, 2185–2196 (2015).
10. Langhorne, J., Ndungu, F. M., Sponaas, A. M. & Marsh, K. Immunity to malaria: more questions than answers. *Nat. Immunol.* **9**, 725–732 (2008).
11. Cohen, S., McGregor, I. A. & Carrington, S. γ-globulin and acquired immunity to human malaria. *Nature* **192**, 733–737 (1961).
12. Triller, G. et al. Natural parasite exposure induces protective human anti-malarial antibodies. *Immunity* **47**, 1197–1209 (2017).
13. Seder, R. A. et al. Protection against malaria by intravenous immunization with a nonreplicating sporozoite vaccine. *Science* **341**, 1359–1365 (2013).
14. Mordmüller, B. et al. Sterile protection against human malaria by chemoattenuated PfSPZ vaccine. *Nature* **542**, 445–449 (2017).
15. Lell, B. et al. Impact of sickle cell trait and naturally acquired immunity on uncomplicated malaria after controlled human malaria infection in adults in Gabon. *Am. J. Trop. Med. Hyg.* **98**, 508–515 (2018).
16. Ishizuka, A. S. et al. Protection against malaria at 1 year and immune correlates following PfSPZ vaccination. *Nat. Med.* **22**, 614–623 (2016).
17. Teirlinck, A. C. et al. *Plasmodium falciparum* infection of human volunteers activates monocytes and CD16⁺ dendritic cells and induces upregulation of CD16 and CD1c expression. *Infect. Immun.* **83**, 3732–3739 (2015).
18. Höllt, T. et al. Cytosplore: interactive immune cell phenotyping for large single-cell datasets. *Comput. Graph. Forum* **35**, 171–180 (2016).
19. Qiu, P. et al. Extracting a cellular hierarchy from high-dimensional cytometry data with SPADE. *Nat. Biotechnol.* **29**, 886–891 (2011).
20. van der Maaten, L. & Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
21. Fergusson, J. R. et al. CD161 defines a transcriptional and functional phenotype across distinct human T cell lineages. *Cell Rep.* **9**, 1075–1088 (2014).
22. Sundling, C. et al. B cell profiling in malaria reveals expansion and remodelling of CD11c⁺ B cell subsets. *JCI Insight* https://doi.org/10.1172/jci.insight.126492 (2019).
23. Urban, E. M. et al. Control of CD56 expression and tumor cell cytotoxicity in human Vγ2Vδ2 T cells. *BMC Immunol.* **10**, 50 (2009).
24. Forconi, C. S. et al. Poorly cytotoxic terminally differentiated CD56^{neg}CD16^{pos} NK cells accumulate in Kenyan children with Burkitt lymphomas. *Blood Adv.* **2**, 1101–1114 (2018).
25. Krasselt, M., Baerwald, C., Wagner, U. & Rossol, M. CD56⁺ monocytes have a dysregulated cytokine response to lipopolysaccharide and accumulate in rheumatoid arthritis and immunosenescence. *Arthritis Res. Ther.* **15**, R139 (2013).
26. Fergusson, J. R. et al. CD161^{int}CD8⁺ T cells: a novel population of highly functional, memory CD8⁺ T cells enriched within the gut. *Mucosal Immunol.* **9**, 401–413 (2016).
27. Huntington, N. D. et al. NK cell maturation and peripheral homeostasis is associated with KLRG1 up-regulation. *J. Immunol.* **178**, 4764–4770 (2007).
28. Bowyer, G. et al. Reduced Ebola vaccine responses in CMV⁺ young adults is associated with expansion of CD57⁺KLRG1⁺ T cells. *J. Exp. Med.* https://doi.org/10.1084/jem.20200004 (2020).
29. Wagar, L. E. et al. Increased T cell differentiation and cytolytic function in Bangladeshi compared to American children. *Front. Immunol.* **10**, 2239 (2019).
30. McCall, M. B. et al. Memory-like IFN-γ response by NK cells following malaria infection reveals the crucial role of T cells in NK cell activation by *P. falciparum*. *Eur. J. Immunol.* **40**, 3472–3477 (2010).
31. Shimizu, K. et al. KLRG⁺ invariant natural killer T cells are long-lived effectors. *Proc. Natl Acad. Sci. USA* **111**, 12474–12479 (2014).
32. Jagannathan, P. et al. Vδ2⁺ T cell response to malaria correlates with protection from infection but is attenuated with repeated exposure. *Sci. Rep.* **7**, 11487 (2017).
33. Bandyopadhyay, K., Marrero, I. & Kumar, V. NKT cell subsets as key participants in liver physiology and pathology. *Cell Mol. Immunol.* **13**, 337–346 (2016).
34. de Jong, S. E. et al. Longitudinal study of changes in γδ T cells and CD4⁺ T cells upon asymptomatic malaria infection in Indonesian children. *Sci. Rep.* **7**, 8844 (2017).
35. Illingworth, J. et al. Chronic exposure to *Plasmodium falciparum* is associated with phenotypic evidence of B and T cell exhaustion. *J. Immunol.* **190**, 1038–1047 (2013).
36. Singh, A. et al. DIABLO: an integrative approach for identifying key molecular drivers from multi-omics assays. *Bioinformatics* **35**, 3055–3062 (2019).
37. Rothen, J. et al. Whole blood transcriptome changes following controlled human malaria infection in malaria pre-exposed volunteers correlate with parasite prepatent period. *PLoS ONE* **13**, e0199392 (2018).
38. Tran, T. M. et al. Transcriptomic evidence for modulation of host inflammatory responses during febrile *Plasmodium falciparum* malaria. *Sci. Rep.* **6**, 31291 (2016).
39. Yu, X. et al. Cross-regulation of two type I interferon signaling pathways in plasmacytoid dendritic cells controls anti-malaria immunity and host mortality. *Immunity* **45**, 1093–1107 (2016).
40. Netea, M. G. et al. Trained immunity: a program of innate immune memory in health and disease. *Science* **352**, aaf1098 (2016).
41. Gonzalez-Aseguinolaza, G. et al. Natural killer T cell ligand α-galactosylceramide enhances protective immunity induced by malaria vaccines. *J. Exp. Med.* **195**, 617–624 (2002).

42. Mamedov, M. R. et al. A macrophage colony-stimulating-factor-producing γδ T cell subset prevents malarial parasitemic recurrence. *Immunity* **48**, 350–363 (2018).
43. Zaidi, I. et al. γδ T cells are required for the induction of sterile immunity during irradiated sporozoite vaccinations. *J. Immunol.* **199**, 3781–3788 (2017).
44. Tran, T. M. et al. A molecular signature in blood reveals a role for p53 in regulating malaria-induced inflammation. *Immunity* **51**, 750–765 (2019).
45. Lei, Q. Y. et al. NKX3.1 stabilizes p53, inhibits AKT activation, and blocks prostate cancer initiation caused by PTEN loss. *Cancer Cell* **9**, 367–378 (2006).
46. Yang, C. et al. Transcriptional activation of caspase-6 and -7 genes by cisplatin-induced p53 and its functional significance in cisplatin nephrotoxicity. *Cell Death Differ.* **15**, 530–544 (2008).
47. Kurup, S. P., Butler, N. S. & Harty, J. T. T cell-mediated immunity to malaria. *Nat. Rev. Immunol.* **19**, 457–471 (2019).

Methods

Clinical trial details. Samples in this study were part of the LaCHMI-1 trial[15], which studied immunity to *P. falciparum* malaria in a controlled infection setting in Lambaréné, Gabon, in August 2014. The trial followed International Council for Harmonization of Technical Requirements for Pharmaceuticals for Human Use Good Clinical Practice guidelines and the principles of the Declaration of Helsinki and was performed in accordance with guidelines approved by the national ethics committee, le Comité Nationale d'Ethique pour la Recherche, Gabon's regulatory authority and under a US Food and Drug Administration Investigational New Drug application. Written informed consent was obtained from all participants. The trial is registered at ClinicalTrials.gov under NCT02237586, which contains a full list of eligibility criteria. Exclusion criteria included positivity for HIV, hepatitis B or hepatitis C. Included in the the trial were 5 malaria-naive European adults and 20 malaria-exposed African adults, of which 11 had a normal hemoglobin genotype (HbAA) and 9 had sickle cell trait (HbAS) (Supplementary Table 1). Hemoglobin genotype and sex had no measurable effect on the parasitological and immunological outcomes of this study (Supplementary Table 1 and Extended Data Fig. 8). Furthermore, eosinophil counts did not differ significantly between TBS^+ and TBS^- Africans, as a proxy for helminth infections.

Before the start of the trial, four Africans had low, asymptomatic parasitemia by TBS (45 to 105 parasites per µl). As per protocol, all volunteers received a treatment to radically cure possible *Plasmodium* infection with a 5-d course of 12-h 5 mg kg^{-1} clindamycin. After clindamycin treatment, all volunteers were free of asexual parasites at C−1, whereas one African individual had gametocytemia by PCR and two Africans had gametocytemia by both PCR and TBS.

At least 3 d after clindamycin treatment, individuals received 3,200 aseptic, purified, fully infectious, nonattenuated PfSPZs (Sanaria PfSPZ Challenge), strain NF54, by intravenous injection[48–50]. Blood was drawn on C−1, D5 and D11 in sodium heparin blood collection tubes for immunophenotyping. From D5 onwards, TBS and PCR tests were performed daily to detect malaria parasites. Effective treatment was initiated upon development of TBS parasitemia for Europeans. For Africans, treatment was initiated if parasitemia was accompanied by symptoms consistent with malaria, if there was parasitemia of >1,000 parasites per µl, or at the end of the study at day 28 in those not previously treated.

Cryopreservation of PBMCs. Heparinized blood was diluted 2× with HBSS, followed by PBMC isolation by 1.077 Ficoll density centrifugation. The HBSS contained 100 U ml^{-1} penicillin G sodium and 100 µg ml^{-1} streptomycin. After washing twice with HBSS, the PBMCs were cryopreserved in 20% fetal bovine serum (FBS; Bodinco; lot. BDC 1886)/10% dimethyl sulfoxide (DMSO)/RPMI-1640 medium. RPMI medium contained 1 mM pyruvate, 2 mM L-glutamine, penicillin G and streptomycin. Cryovials were placed overnight in a Nalgene Mr. Frosty Freezing Container (Thermo Fisher Scientific) at −80 °C before transfer to liquid nitrogen for long-term storage. Cryopreserved PBMCs were shipped in a liquid nitrogen dry vapor shipper from Lambaréné, Gabon, to Leiden, the Netherlands, for analysis.

On the day of mass cytometry staining and PfRBC stimulation, cryopreserved PBMCs were thawed with 50% FBS/RPMI medium at 37 °C, washed twice with 10% FBS/RPMI medium and 2×10^6 cells per sample were stored on ice temporarily. Thawing was performed in five batches in subsequent weeks in 2015, with a median recovery of 75% and viability of 98.5% and without significant differences between the batches.

Mass cytometry staining. The mass cytometry antibody panel consisted of preconjugated antibodies as well as self-conjugated antibodies (Supplementary Table 2). Staining for $V\delta2^+$ γδ T cells was conducted with the same panel, with the following adjustments: TCR Vd2-^{157}Gd (clone B6, BioLegend, cat. 331402, RRID AB_1089226, 1:100 dilution) was added, CD86-^{144}Nd was replaced by CD86-^{198}Pt (clone IT2.2, BioLegend, cat. 305435, RRID AB_2563764, 1:200 dilution), CD14-^{160}Gd was replaced by CD14-^{148}Nd (clone M5E2, BioLegend, cat. 301843, RRID AB_2562813, 1:100 dilution), CD16-^{148}Nd was replaced by CD16-^{209}Bi (clone 3G8, Fluidigm, cat. 3209002B, 1:200 dilution), CD185 (CXCR5)-^{150}Nd was replaced by CD185-150^{Nd} (clone J252D4, BioLegend, cat. 356902, RRID AB_2561811, 1:100 dilution), barcoding with β2-microglobulin (BioLegend, cat. 316302, RRID AB_492835, 1:50 dilution) coupled to ^{106}Cd, ^{110}Cd, ^{111}Cd, ^{112}Cd, ^{114}Cd and ^{116}Cd was added and most Fluidigm preconjugated antibodies were self-conjugated with the same clone antibody from BioLegend. For self-conjugation, 100 µg of antibody and the MaxPar X8 Antibody Labeling kit were used according to manufacturer's protocol V2. The conjugated antibody was stored in 200 µl Antibody Stabilizer PBS (Candor Bioscience) at 4 °C. All antibodies were titrated on study samples.

Staining for mass cytometry was based on MaxPar Cell Surface Staining Protocol V2 (Fluidigm). First, cells were washed with MaxPar staining buffer (Fluidigm) and 5 min centrifugation at 300*g* in 5 ml polystyrene round-bottom tubes. Then, the cells were incubated with 1 ml 500× diluted 500 µM Cell-ID Intercalator-103Rh (Fluidigm) in staining buffer at room temperature for 15 min. After washing with staining buffer, cells were incubated with 5 µl Human TruStain FcX Fc-receptor blocking solution (BioLegend) and 40 µl staining buffer at room temperature for 10 min. Then, 50 µl of freshly prepared antibody cocktail was added and incubated at room temperature for another 45 min. Subsequently, the cells were washed 3× with staining buffer and incubated with 1 ml 1,000× diluted 125 µM Cell-ID Intercalator-Ir (Fluidigm) in MaxPar Fix and Perm buffer (Fluidigm) at 4 °C overnight. After three washes with staining buffer and centrifugation at 800*g*, cells were stored as a pellet at 4 °C and measured within a week. The median recovery after staining was 80.5%.

Measurement by mass cytometry. Measurement of samples by mass cytometry was randomized per individual to avoid bias by staining and measurement day. Samples belonging to the same individual were stained and measured together. Furthermore, every staining day, PBMCs from the same cryopreserved control source were thawed and stained, after which the sample was split over two tubes. These tubes were the first and the last samples measured within a measurement week for quality control and were found to be highly comparable.

Samples were measured with a CyTOF2 mass cytometer (Fluidigm), which was automatically tuned according to Fluidigm's recommendations. Before measurement, cells were counted, washed with Milli-Q water, passed over a cell strainer and brought to a concentration of 0.5×10^6 cells ml^{-1} with 10% EQ Four Element Calibration Beads (Fluidigm) in Milli-Q water. Samples were measured and analyzed on-the-fly, using dual-count mode, acquisition delay set to 30 s and detector stability delay set to 10 s. Noise-reduction was applied, with a lower convolution threshold of 200, no event subtraction, minimum event duration of 10 s, maximum event duration of 150 s, sigma of 3 and no event limit. Threshold filtering was set to default. Next to channels used to detect antibodies, channels for intercalators (^{103}Rh, ^{191}Ir and ^{193}Ir), calibration beads (^{140}Ce, ^{151}Eu, ^{153}Eu, ^{165}Ho and ^{175}Lu) and background/contamination (^{133}Cs, ^{138}Ba, ^{206}Pb and background) were acquired. FCS files were normalized and concatenated in CyTOF2 software, without removing beads. The median sampling efficiency (or recovery), which is the percentage of cells in a sample that is saved in the FCS file, was 36.0%, which falls within the expected range for CyTOF2 mass cytometers.

Stimulation with PfRBCs. After thawing, cells were rested overnight at 1×10^6 cells ml^{-1} in 10% FBS/RPMI in an upright-standing 25 cm^2 cell culture flask placed in an incubator at 37 °C and with 5% CO_2. Subsequently, cells were stimulated in 96-well round-bottom plates with 0.5×10^6 cells per well in 200 µl medium with 0.5×10^6 intact PfRBCs or uninfected red blood cells (uRBCs) or Staphylococcal enterotoxin B (200 ng ml^{-1}; Sigma) in the incubator for 24 h. After 2 h of incubation, brefeldin A (10 µg ml^{-1}; Sigma) was added to all conditions. After 24 h, cells were stained with LIVE/DEAD Fixable Aqua Dead Cell Stain (50 µl per well, 1:400 diluted) according to manufacturer's instructions (Thermo Fisher Scientific), fixed with 1.9% formaldehyde solution for 15 min at room temperature and frozen at −20 °C in 10% FBS/10% DMSO/RPMI and stored at −80 °C.

Measurement by flow cytometry. Measurement of samples by flow cytometry was also randomized per individual to avoid bias by staining and measurement day. Samples belonging to the same individual were stained and measured together. After thawing, fixed cells were stained in 96-well V-bottom microplates with 50 µl of antibody mixture (see Supplementary Table 5 for panel) diluted in eBioscience permeabilization buffer (Thermo Fisher Scientific) with 1% human Fc-receptor binding inhibitor (Thermo Fisher Scientific; cat. 14-9161) at 4 °C for 30 min. Leftover cells were pooled and then split for use as fluorescence-minus-one (FMO) controls. Compensation beads (BD CompBead, BD Biosciences) were made fresh every day. PBMCs were acquired with a BD LSR-Fortessa X-20 SORP (Supplementary Table 5). BD FACSDiva 8.0.1. CST setup beads (BD Biosciences, lot 83654) were run before every measurement. A median of 141,800 live, single cells was obtained per sample.

RNA-seq. Blood for RNA-seq was collected in PAXgene blood RNA tubes (QIAGEN; cat. 762125) and kept at room temperature for 2 h, frozen overnight at −20 °C and then brought to −80 °C for long-term storage. The samples were shipped from Gabon to the Netherlands on dry ice and stored again at −80 °C. RNA was purified from PAXgene tubes, following the manufacturer's protocol. Library preparation and sequencing were contracted to GenomeScan B.V. The NEBNext Ultra Directional RNA Library Prep kit for Illumina was used to process the samples. Messenger RNA was isolated from total RNA using oligo-dT magnetic beads and used for library preparation. DNA sequencing was performed with the Illumina NextSeq 500 according to the manufacturer's protocol.

Mass cytometry data processing. The workflow for the mass cytometry data processing is summarized in Fig. 1. At first, FlowJo v.10.2-10.7.1 for Mac (FlowJo) was used to gate out beads and to subsequently select single, live, $CD45^+$ cells (Extended Data Fig. 3). A total of 95.1% of the collected events were cells, of which 77% were single cells, 96.7% were viable and 99.8% were $CD45^+$ cells (medians). The gating steps resulted in a median number of 432,000 single, live $CD45^+$ cells per sample and a total number of 33.3 million cells for all samples combined.

Next, FCS files were analyzed using Cytobank (2015, Cytobank) with arcsinh transformation with cofactor 5. SPADE[19] was applied on all 75 samples, with 35 clustering channels (all antibodies, except CD197 (CCR7)), a target number of 500 nodes and 70,000 downsampled events and without compensation. Nodes (clusters of cells) comprising artifacts such as dead cells and doublets were identified by

combinations of high levels of the dead cell stain ^{103}Rh, contaminating ^{138}Ba and ^{206}Pb, bead marker ^{140}Ce, long event length and/or expression of mutually exclusive markers such as CD3, CD14 and CD19. The remaining nodes contained 33.1 million cells and formed five branches: the CD4$^+$ T cell, CD8$^+$ T cell, B cell, CD7$^+$ T cell and myeloid cell branch (Fig. 1 and Extended Data Fig. 6). These branches were subsequently analyzed separately to reduce the size of the individual analyses and to improve the discriminatory power of these analyses but were not intended to reflect pure hematopoietic lineages.

Cytosplore (www.cytosplore.org, 2015–2021) was used to determine cell clusters within the branches. Cytosplore makes use of approximated *t*-SNE[51], followed by interactive determination of cell clusters by a Gaussian mean-shift method. This results in more accurate and reproducible cell cluster classification than with SPADE or ACCENSE[52]. In Cytosplore, the number of output clusters is determined by a manually adjustable kernel bandwidth, which could be interactively optimized thanks to the visualization of clusters on the *t*-SNE map and was further guided by visualization of marker expression variations of clusters in the corresponding heat map. Before *t*-SNE was run, FCS files were arcsinh-transformed with cofactor 5 and an extra channel called SampleTag was added to the FCS files to be able to identify from which sample an event originated after *t*-SNE. Currently, Cytosplore can perform these tasks, but for this study, MATLAB R2015b (The MathWorks) was used for the transformation with CYT[53] as well as the addition of the SampleTag. Furthermore, FCS files were downsampled before *t*-SNE and upsampled after *t*-SNE. Downsampling involved random selection of events from each sample, with a maximum of 13,000 events per sample to reach a maximum of 975,000 total events per t-SNE, as *t*-SNE does not perform well with more events. After downsampling, *t*-SNE analyses and clustering was carried out separately for the five SPADE branches with Cytosplore, with perplexity set at 30 and 1,000 iterations and resulted in a total of 235 clusters. Afterward, upsampling was performed, which was based on matching of events to the median signal intensity of the 235 previously identified clusters and the maximum arcsinh-transformed single marker distance was set to 3.5 and the maximum total distance to 25 in Cytosplore. Per SPADE branch, 97.0% to 98.2% of cells were matched to these clusters after upsampling. The remaining unassigned cells were analyzed in a second round of *t*-SNE runs, which resulted in an additional 100 clusters. Many of these clusters were later of significance in the comparative analyses. After upsampling, the 335 cell clusters were manually merged to a final number of 198 immunological distinguishable clusters (Extended Data Fig. 1), based on a heat map of the median signal intensity of all channels and all clusters.

Next, the 198 clusters were labeled manually at the subset level and a higher lineage level (Fig. 1 and Supplementary Table 3). The lineage level consists out of CD4$^+$ T cells, CD8$^+$ T cells, unconventional αβ T cells, γδ T cells, B cells, ILCs, monocytes and dendritic cells, basophils and undefined cells. Multiple subsets were identified within each of these lineages. To illustrate, two clusters were found to be CD3$^+$CD4$^+$TCRγδ$^-$CD56$^-$CRTH2$^+$ and were therefore classified into the T_H2 cell subset, which was part of the CD4$^+$ T cell lineage. These clusters were also CD45RA$^{-/lo}$CD45RO$^{+/hi}$CCR7$^-$ and could therefore also be analyzed as part of the EM CD4$^+$ T cell subset. These lineages, subsets and clusters were used for the subsequent statistical analysis.

Regarding the analysis for Vδ2$^+$ γδ T cells (Extended Data Fig. 5), batch correction was performed using the fastMNN function[54] and uniform manifold approximation and projection dimensionality reduction was run on combined data.

Flow cytometry data processing. Flow cytometry data analysis was performed with FlowJo software (v.10.4.1; TreeStar). Gating for time, singlets and live cells was performed before gating of the subsets of interest (Extended Data Fig. 6). Gates were initially set according to FMO control and adjusted according to negative controls for gating of cytokine-positive cells. The percentage of PfRBC-specific cytokine-producing cells was corrected by subtracting background levels as observed in uRBC conditions.

RNA-seq data processing. All RNA sequence files were processed using the BIOPET Gentrap pipeline v.0.8 developed at the LUMC (https://biopet-docs.readthedocs.io/en/latest/pipelines/gentrap/). The pipeline consists of FASTQ pre-processing (including quality control, quality trimming and adaptor clipping), RNA-seq alignment, read and base quantification and optionally transcript assembly. FastQC v.0.11.2 was used for raw read quality control. Low-quality read trimming was conducted using sickle v.1.33 with default settings. Cutadapt v.1.10 with default settings was used for adaptor clipping on the basis of detected adaptor sequences by the FastQC toolkit. RNA-seq reads were aligned against human reference genome GRCh38 using RNAseq aligner GSNAP v.2014-12-23 with settings '–npaths 1–quiet-if-excessive'. Ensembl human genome annotation v.87 was used for raw read counting. The gene read quantification step was performed using htseq-count v.0.6.1p1 with the setting '–stranded=reverse'. FASTQ files and the gene count matrix will be made available upon request.

Statistics. Statistical testing of the lineages, subsets and clusters was performed using RStudio v.0.99.902 for Windows (RStudio; www.rstudio.com), R x64 v.3.3.1 for Windows (R Foundation for Statistical Computing; http://www.r-project.org/) and the Global test R package, v.5.24.0 (http://bioconductor.org/packages/globaltest/)[55]. Survival analysis at the cluster level was performed with the survival R package, v.2.38 (https://CRAN.R-project.org/package=survival), using the Cox proportional hazards regression model. Other statistics were performed using SPSS Statistics v.23 for Windows (IBM). The Fisher's exact test for contingency tables and the nonparametric Kruskal–Wallis H test were used to compare population characteristics. Log-rank (Mantel–Cox) test was used for survival/time-to-event analysis. Complete blood counts were compared by Kruskal–Wallis H test followed by Dunn's post hoc test. Hierarchical clustering was performed based on one minus Spearman's rank correlation with average linkage. Measurements were taken from distinct samples, rather than repeated measurements, unless stated otherwise. *P* values ≤0.05 were considered statistically significant. Tests were two-tailed, where applicable. Results presented come from single experiments, unless otherwise specified.

Percentages of lineages were expressed relative to total cells, percentages of subsets relative to either total cells or to its respective lineage and percentages of clusters relative to its respective subset. Percentages of lineages and subsets at baseline were $\log_2$-transformed for the Global test; percentages of clusters were square-root transformed. Responses of cell lineages, subsets and clusters to the PfSPZ DVI were assessed as relative changes in percentages over time, for example percentage at D5/percentage at C−1 and then $\log_2$-transformed. Percentage of cytokine-producing cells as measured by flow cytometry were determined after subtraction of background (uRBC control) from the PfRBC-stimulated samples. Negative values were set to zero. Percentages were calculated relative to the total live, singlet cells. Wilcoxon signed-rank or rank-sum test was used to compare the difference between two groups of samples, paired or unpaired, respectively.

Differential gene expression analysis. For downstream analysis of the transcriptomics data, we included 19,331 genes after excluding genes of the Y chromosome and genes without known Entrez ID or HGNC symbols. Exclusion of low-expressed genes was performed with the filterByExpr function as implemented in the edgeR R package, v.3.24.0, (http://bioconductor.org/packages/edgeR/) resulting in 11,659 genes. This resulted in gene expression profiles sufficient for our exploratory analysis, although it is possible that more lowly expressed genes could have been found with deeper sequencing. The trimmed means of M values method was used to produce normalization factors correcting raw counts for different library sizes.

Differential expression testing was conducted with the limma-voom workflow as implemented in the limma R package, v.3.38.3 (http://bioconductor.org/packages/limma/)[56]. Correlations between paired measurements were validated by the duplicateCorrelation function in limma. For linear modeling of differential gene expression, a factorial design matrix for a model with no intercept and six groups was constructed (three time points for both Europeans and Africans). Moderated *F*-statistics combining comparisons for all time points was performed separately for Europeans and Africans to identify DEGs on any contrast. For pairwise comparison between time points, Student's *t*-tests with a moderated Bayesian variance estimator were applied. An FDR cutoff of the BH-adjusted *P* values <0.05 was used to select DEGs.

Hypergeometric tests to test for GO term enrichment with adjustment for gene length were performed with the goana function as implemented in limma. The universe for the enrichment test was restricted to genes included for differential expression analysis. REViGO was used to summarize significant GO Biological Process terms ($P < 0.05$ after BH adjustment)[57]. *Homo sapiens* GO database (GO Biological Process from MSigDB, v.6.2, July 2018; https://www.gsea-msigdb.org/gsea/downloads_archive.jsp) with SimRel semantic similarity was used with medium-allowed similarity. Annotation of GO terms into categories was conducted manually.

Gene set enrichment analysis with CAMERA was performed to compare pathway enrichment between the TBS$^+$ and TBS$^-$ group. For this analysis, we focused our test on eight GO Biological Process pathway terms that contained 'interferon-gamma' in their names. The resulting *P* values were corrected with the Bonferroni procedure.

Machine learning and data integration. For machine learning, the 4,000 most variable genes were selected and all 176 CyTOF clusters and 49 cellular responses identified by ICS were included. Baseline features were *z*-scaled before inclusion in machine-learning models. Two machine-learning approaches were tested: Extreme Gradient Boosting, an adaptation of random forests, and DIABLO, which combines partial least-squares discriminant analysis with canonical correlation analysis by aligning individual datasets and outcomes on the same latent space[36,58]. XGB modeling and tuning of parameters was performed using the Caret package (v.6.0-84) in R, with a tune length of 3 and a leave-one-out (LOO) cross-validation scheme. For each fold, important parameters were extracted using the varImp function. DIABLO was performed using mixOmics (v.6.8.0) with a two-component latent space and tuning was performed using the tune.block.splsda function to determine the minimum number of features that afforded maximum performance in the LOO cross-validation[59]. The weight on aligning between the different datasets (RNA-seq, CyTOF and ICS), relative to the outcome, was manually

tuned with 0, 0.1, 0.25, 0.5, 0.75 and 1 included in the design matrix. At baseline, Extreme Gradient Boosting had an 84% accuracy in classifying Europeans versus Africans and 70% accuracy to classify TBS$^+$ versus TBS$^-$ Africans, using LOO cross-validation (Extended Data Fig. 7). DIABLO performed better with accuracies of 96% and 85%, respectively. These accuracies were significantly increased over random permutations ($n=1{,}000$), indicating that at baseline both ethnicity and susceptibility to malaria infection could be accurately predicted. Moreover, Extreme Gradient Boosting identified only genes as important features, due to the very large number of genes relative to CyTOF clusters and ICS responses, whereas DIABLO inherently selected features from all datasets. The latter method was thus used to analyze the most discriminating combination of genes, cell clusters and responses between groups. For comparison of Africans versus Europeans, a weight of 0.5 between datasets and ten genes, four CyTOF clusters and four ICS features were used for the final model. For comparison between TBS$^+$ and TBS$^-$ Africans, a weight of 0.1 between datasets and 4 genes, 6 CyTOF clusters and 12 ICS features were selected for the final model. For each of the rounds in the LOO cross-validation, selected features and correlation between datasets on the first component of the latent space were extracted for analysis.

Of note, the aim of this analysis was to identify features across the datasets that were most strongly associated with naturally acquired resistance to malaria infection rather than to build a predictive machine-learning model, although we used LOO cross-validation to assess model performance in our study.

Protein microarray. Protein microarray experiments and analyses were performed as previously described[60,61]. Microarray slides were spotted with malaria proteins at the University of California Irvine[62]. In total, 262 *P. falciparum* proteins representing 228 unique antigens were expressed using an *Escherichia coli* lysate in vitro expression system and spotted on a 16-pad ONCYTE AVID slide, representing 228 important *P. falciparum* antigens known to frequently provide a positive signal when tested with serum from those with sterile and naturally acquired immunity against the parasite. For the detection of binding antibodies, secondary IgG antibody (goat anti-human IgG QDot800, Grace Bio-Labs) was used[62–65].

Study serum samples as well as European control serum were diluted 1:50 in 0.05× Super G Blocking Buffer (Grace Bio-Labs) containing 10% *E. coli* lysate (GenScript) and incubated for 30 min on a shaker at room temperature. Meanwhile, microarray slides were rehydrated using 0.05× Super G Blocking buffer at room temperature. Rehydration buffer was subsequently removed and samples added onto the slides. Arrays were incubated overnight at 4 °C on a shaker (180 r.p.m.). Serum samples were removed the following day and microarrays were washed using 1× TBST buffer (Grace Bio-Labs). Secondary antibodies were then applied at a dilution of 1:200 and incubated for 2 h at room temperature on the shaker, followed by another washing step and a 1-h incubation in a 1:250 dilution of Qdot585 Streptavidin Conjugate. After a final washing step, slides were dried by centrifugation at 500*g* for 10 min. Slide images were taken using the ArrayCAM Imaging System (Grace Bio-Labs) and the ArrayCAM 400-S Microarray Imager Software v.2.2.

Microarray data were analyzed in R statistical software package v.3.6.2. All images were manually checked for any noise signal. Each antigen spot signal was corrected for local background reactivity by applying a normal-exponential convolution model[66] using the RMA-75 algorithm for parameter estimation (available in the limma package v.3.28.14, https://bioconductor.org/packages/limma/)[67]. Data were $\log_2$-transformed and further normalized by subtraction of the median signal intensity of mock expression spots on the particular array to correct for background activity of antibodies binding to *E. coli* lysate. After $\log_2$ transformation, data were normally distributed. Differential antibody levels (protein array signal) in the different allocated study outcomes (protected participants, nonprotected participants developing parasitemia) were determined by Welch-corrected Student's *t*-test. Antigens with $P<0.05$ and a fold change >2 of mean signal intensities were defined as differentially recognized between the tested sample groups. Volcano plots were generated using GraphPad Prism v.9.0.0.

Graphing. Graphs were made using R, GraphPad Prism v.7 to 9 for Windows (GraphPad Software), Morpheus heat map software v.2016–2018 (https://software.broadinstitute.org/morpheus) and ColorBrewer 2.0 color schemes (www.colorbrewer2.org). *t*-SNE plots were made with MATLAB, the dscatter function, v.1.1.0.1 (https://www.mathworks.com/matlabcentral/fileexchange/8430-flow-cytometry-data-reader-and-visualization) and inferno color scheme (https://bids.github.io/colormap/). Arrow plots were created with mixOmics (v.6.8.0). Adobe Illustrator CC version 2015–2021 (Adobe) was used to combine multiple graphs and create the figures.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The mass cytometry, flow cytometry and RNA-seq data that support the findings of this study are available at ImmPort under study accession SDY1734. The *Homo sapiens* GO database that was used (GO Biological Process from MSigDB, v.6.2, July 2018) is available at https://www.gsea-msigdb.org/gsea/downloads_archive.jsp. Further information is available from the corresponding author upon reasonable request.

Code availability

Code is available from the corresponding author upon reasonable request.

References

48. Roestenberg, M. et al. Controlled human malaria infections by intradermal injection of cryopreserved *Plasmodium falciparum* sporozoites. *Am. J. Trop. Med. Hyg.* **88**, 5–13 (2013).
49. Gómez-Pérez, G. P. et al. Controlled human malaria infection by intramuscular and direct venous inoculation of cryopreserved *Plasmodium falciparum* sporozoites in malaria-naive volunteers: effect of injection volume and dose on infectivity rates. *Malar. J.* **14**, 306 (2015).
50. Mordmüller, B. et al. Direct venous inoculation of *Plasmodium falciparum* sporozoites for controlled human malaria infection: a dose-finding trial in two centres. *Malar. J.* **14**, 117 (2015).
51. Pezzotti, N. et al. Approximated and user steerable tSNE for progressive visual analytics. *IEEE Trans. Vis. Comput. Graph.* **23**, 1739–1752 (2017).
52. Shekhar, K., Brodin, P., Davis, M. M. & Chakraborty, A. K. Automatic classification of cellular expression by nonlinear stochastic embedding (ACCENSE). *Proc. Natl Acad. Sci. USA* **111**, 202–207 (2014).
53. Amir el, A. D. et al. viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. *Nat. Biotechnol.* **31**, 545–552 (2013).
54. Haghverdi, L., Lun, A. T. L., Morgan, M. D. & Marioni, J. C. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat. Biotechnol.* **36**, 421–427 (2018).
55. Goeman, J. J., van de Geer, S. A., de Kort, F. & van Houwelingen, H. C. A global test for groups of genes: testing association with a clinical outcome. *Bioinformatics* **20**, 93–99 (2004).
56. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
57. Supek, F., Bosnjak, M., Skunca, N. & Smuc, T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS ONE* **6**, e21800 (2011).
58. Chen, T. & Guestrin, C. XGBoost: a scalable tree boosting system. In *Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Francisco, California, USA, 785–794 (ACM, 2016).
59. Rohart, F., Gautier, B., Singh, A. & Le Cao, K. A. mixOmics: an R package for 'omics feature selection and multiple data integration. *PLoS Comput. Biol.* **13**, e1005752 (2017).
60. Borrmann, S. et al. Mapping of safe and early chemo-attenuated live *Plasmodium falciparum* immunization identifies immune signature of vaccine efficacy. Preprint at *bioRxiv* https://doi.org/10.1101/2020.09.14.296152 (2020).
61. Wichers, J. S. et al. Common virulence gene expression in naive and severe malaria cases. Preprint at *bioRxiv* https://doi.org/10.1101/2020.11.13.381137 (2020).
62. Doolan, D. L. et al. Profiling humoral immune responses to *P. falciparum* infection with protein microarrays. *Proteomics* **8**, 4680–4694 (2008).
63. Obiero, J. M. et al. Antibody biomarkers associated with sterile protection induced by controlled human malaria infection under chloroquine prophylaxis. *Msphere* https://doi.org/10.1128/mSphereDirect.00027-19 (2019).
64. Felgner, P. L. et al. Pre-erythrocytic antibody profiles induced by controlled human malaria infections in healthy volunteers under chloroquine prophylaxis. *Sci. Rep.* **3**, 3549 (2013).
65. Dent, A. E. et al. *Plasmodium falciparum* protein microarray antibody profiles correlate with protection from symptomatic malaria in Kenya. *J. Infect. Dis.* **212**, 1429–1438 (2015).
66. McGee, M. & Chen, Z. Parameter estimation for the exponential-normal convolution model for background correction of affymetrix GeneChip data. *Stat. Appl. Genet. Mol. Biol.* **5**, 24 (2006).
67. Silver, J. D., Ritchie, M. E. & Smyth, G. K. Microarray background correction: maximum likelihood estimation for the normal–exponential convolution. *Biostatistics* **10**, 352–363 (2009).

Acknowledgements

We thank the volunteers who participated in the trial; the trial staff, in particular A. A. Adegnika, J. Honkpehedji, J. Zinsou and C. Dejon-Agobe; and the LUMC Flow Cytometry Core Facility staff. We also thank the manufacturing, quality systems, regulatory, logistics, pharmaceutical operations, clinical, operations and legal teams at Sanaria for their contributions to the clinical trial. Furthermore, we thank H. Lima Mbenkep and N. Li for their help with the study and B. Everts, S. Khan, A. Luty and H. Smits for critically reading the manuscript. This work was supported by the Leiden University Medical Center excellent student funding (S.E.J.); the German Ministry of Education and Research in the framework of the German Center for Infection Research (P.G.K., B.M. and B.L.); the Indonesian Endowment Fund for Education

(Lembaga Pengelola Dana Pendidikan) of the Ministry of Finance, Republic of Indonesia (reference no. S-1598/LPDP.3/2016) (M.D.M.); the Bontius Stichting (M.Y.); Stichting Tabernaleporis (M.Y.); Leiden University Medical Center Strategic Fund for Leiden Controlled Human Infection Center (M.Y. and M.R.); and Sanaria Inc. (B.K.L.S. and S.L.H.). Production of PfSPZ Challenge (NF54) by Sanaria was supported in part by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under SBIR award no. R01GM9987654 (B.K.L.S. and S.L.H.). Slide reading training was provided through the EDCTP strategic primer grant SP.2011.41304.062 (S.L.H.). This project also received funding from LEaDing Fellows and the European Union's Horizon 2020 Research and Innovation Program under the Marie Skłodowska-Curie grant agreement no. 707404 (S.P.J.). The opinions expressed in this document reflect only the authors' views. The European Commission is not responsible for any use that may be made of the information it contains. B.K.L.S. and S.L.H. have a financial interest in Sanaria Inc., which is the developer and owner of PfSPZ Challenge and the sponsor of the clinical trial. All other authors declare no competing interests.

Author contributions

B.M., B.K.L.S., S.L.H., P.G.K. and B.L. designed and performed the trial. Y.D.M. and M.E.B.O. were involved in the logistics and sample collection of the trial. S.E.J. and M.Y. conceived and designed the immunological study and specific experiments. S.E.J. and Y.C.M.K. performed the sample preparation. S.E.J. and M.H.K. performed the mass cytometry experiments. S.E.J. analyzed the mass cytometry data. M.D.M. and S.A. performed the flow cytometry experiments and analysis. M.D.M. analyzed the RNA-seq data. F.R.L. and R.F. performed the antibody experiments and analysis. V.U., T.H., N.P., A.V., E.E., B.P.F.L. and F.K. developed Cytosplore. S.E.J., J.J.G., K.A.S., S.P.J. and M.J.T.R. performed the statistical analysis and data integration. V.U. and F.K. advised on the mass cytometry work. M.R. advised on interpretation of data and revised the manuscript. S.E.J. and M.Y. supervised the experiments and analyses. S.E.J. and M.Y. wrote the initial manuscript. All authors reviewed the manuscript.

Competing interests

B.K.L.S. and S.L.H. have a financial interest in Sanaria Inc., which is the developer and owner of PfSPZ Challenge and the sponsor of the clinical trial. All other authors declare no competing interests.

Additional information

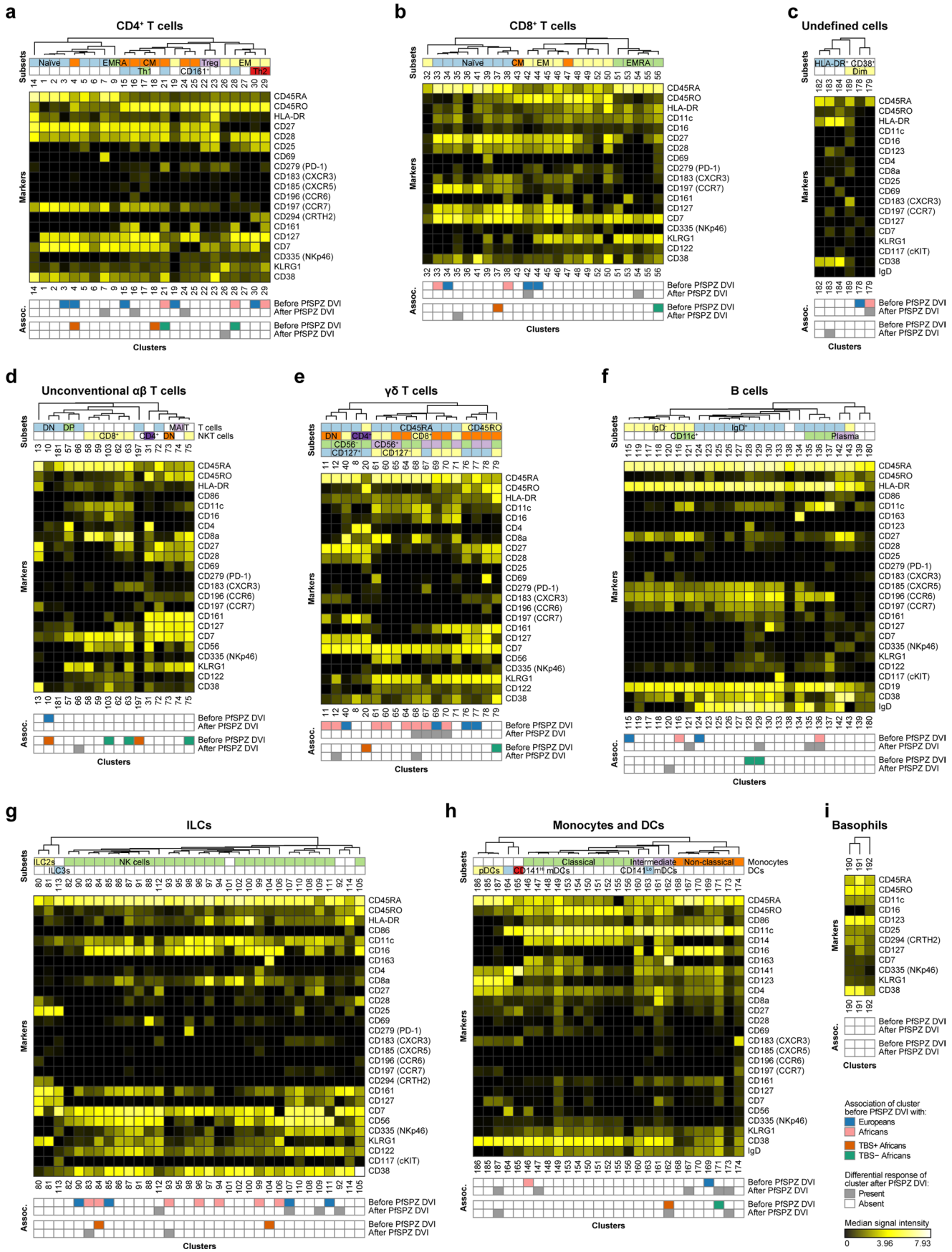
Extended data is available for this paper at https://doi.org/10.1038/s41590-021-00911-7.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41590-021-00911-7.

Correspondence and requests for materials should be addressed to M.Y.

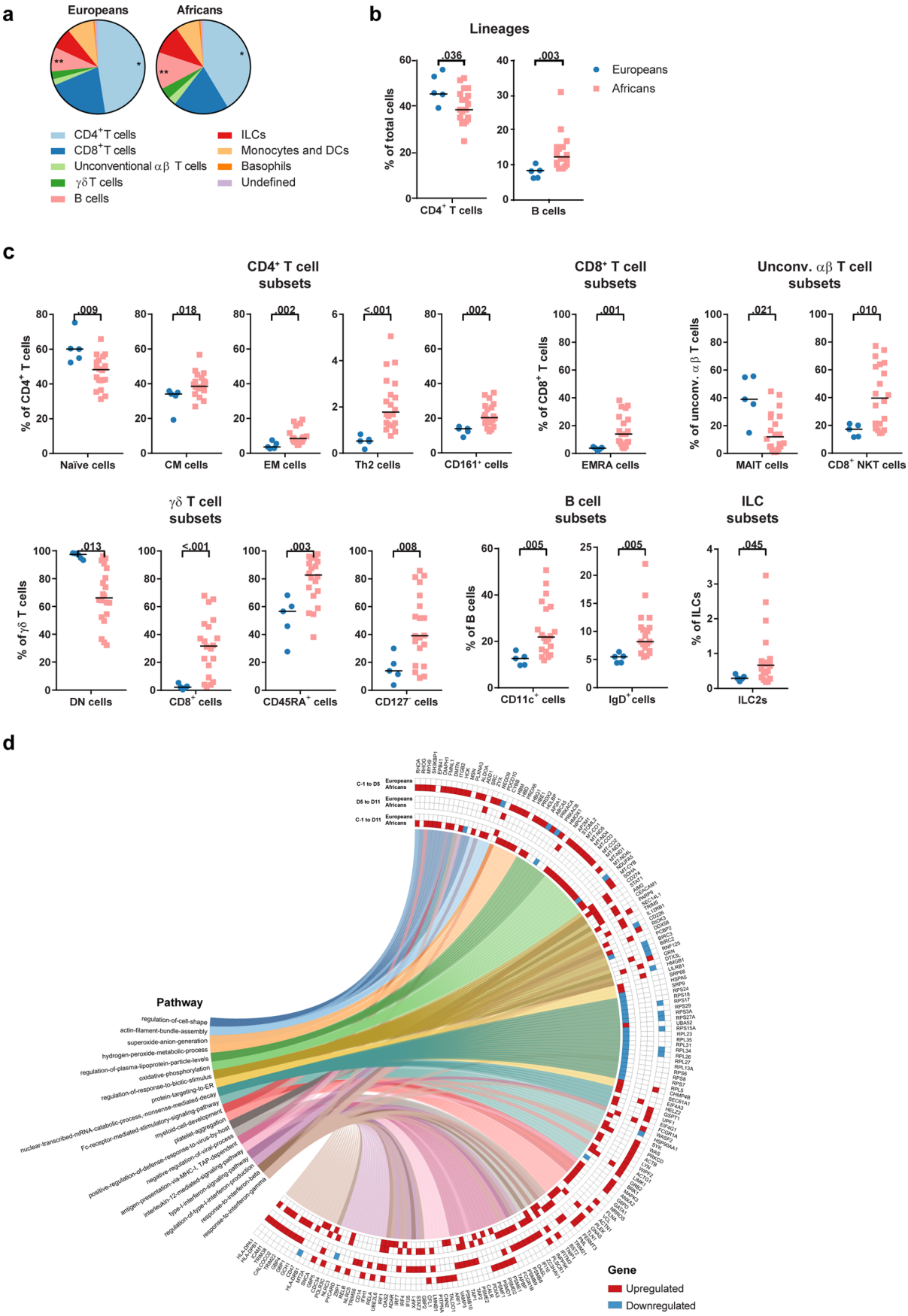
Peer review information *Nature Immunology* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available. Zoltan Fehervari was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Reprints and permissions information is available at www.nature.com/reprints.



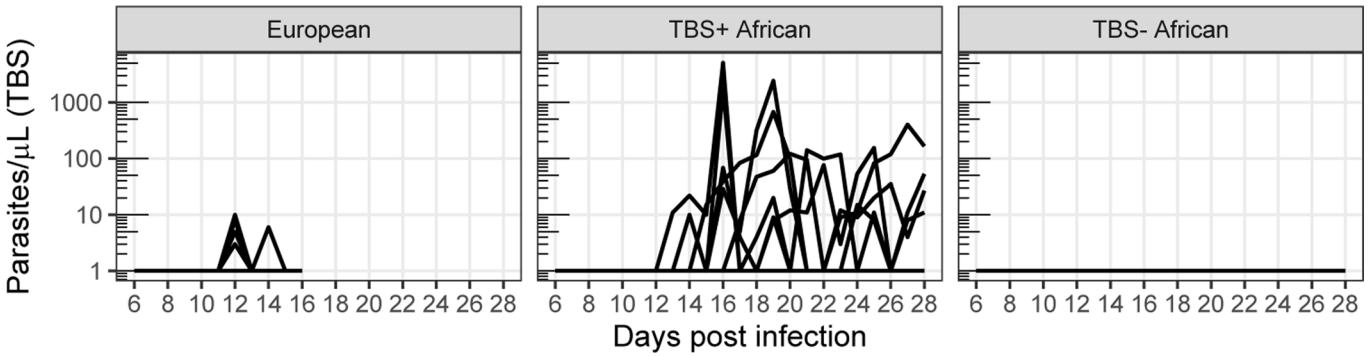
Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Immune signatures at the cluster level before and after *Plasmodium falciparum* sporozoite direct venous inoculation. Heat maps showing expression of markers as median signal intensity (MSI) after arcsinh transformation for the 198 clusters identified. Each cluster has a unique cluster number. Clusters are shown within $CD4^+$ T cells (**a**), $CD8^+$ T cells (**b**), undefined (**c**), unconventional αβ T cells (**d**), γδ T cells (**e**), B cells (**f**), ILCs (**g**), monocytes and DCs (**h**), and basophils (**i**) lineages. A dendrogram depicts the hierarchical clustering within the lineages. The colors in the upper horizontal panels (Subset panels) under the dendrograms show to which subset(s) a cluster belongs. Note that these colors are specific per row and lineage and are independent of colors in other rows and lineages. In the bottom panels (Assoc.), the associations of clusters with Europeans (dark blue) (n = 5) versus Africans (pink) (n = 20) are given in row 1 from top. The associations of clusters with those Africans that developed parasitemia as detected by thick blood smear microscopy (TBS+) (light orange) (n = 12) versus those who did not (TBS−) (dark green) (n = 8) are given in row 3. When a change in cluster frequency after DVI with PfSPZ was associated with either Europeans or Africans and TBS+ or TBS− Africans, this is indicated in grey in rows 2 and 4, respectively. See Supplementary Table 4 for the direction of those responses and P-values. Markers that were throughout negative or positive for a lineage are not shown. All cell clusters were positive for CD45, and therefore CD45 expression is not shown. Similarly, all $CD4^+$ T cells were positive for CD3 and CD4, all $CD8^+$ T cells for CD3 and CD8, all unconventional αβ T cells for CD3, all γδ T cells for CD3 and TCRγδ, and all monocytes and dendritic cells (DCs) for HLA-DR, and therefore these markers were also not shown in the heat maps. For abbreviations and subset marker combinations, see Supplementary Table 3.

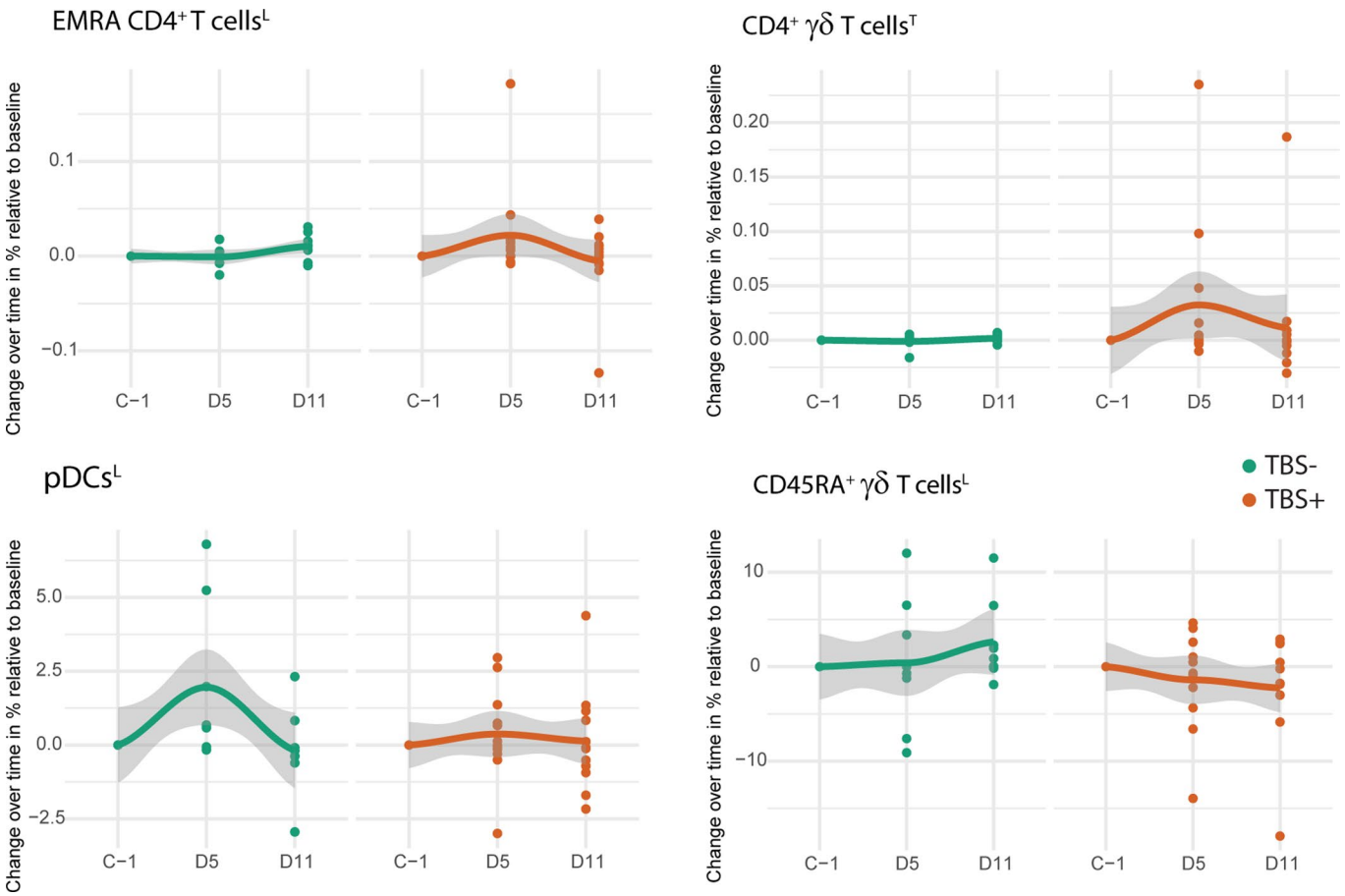


Extended Data Fig. 2 | See next page for caption.

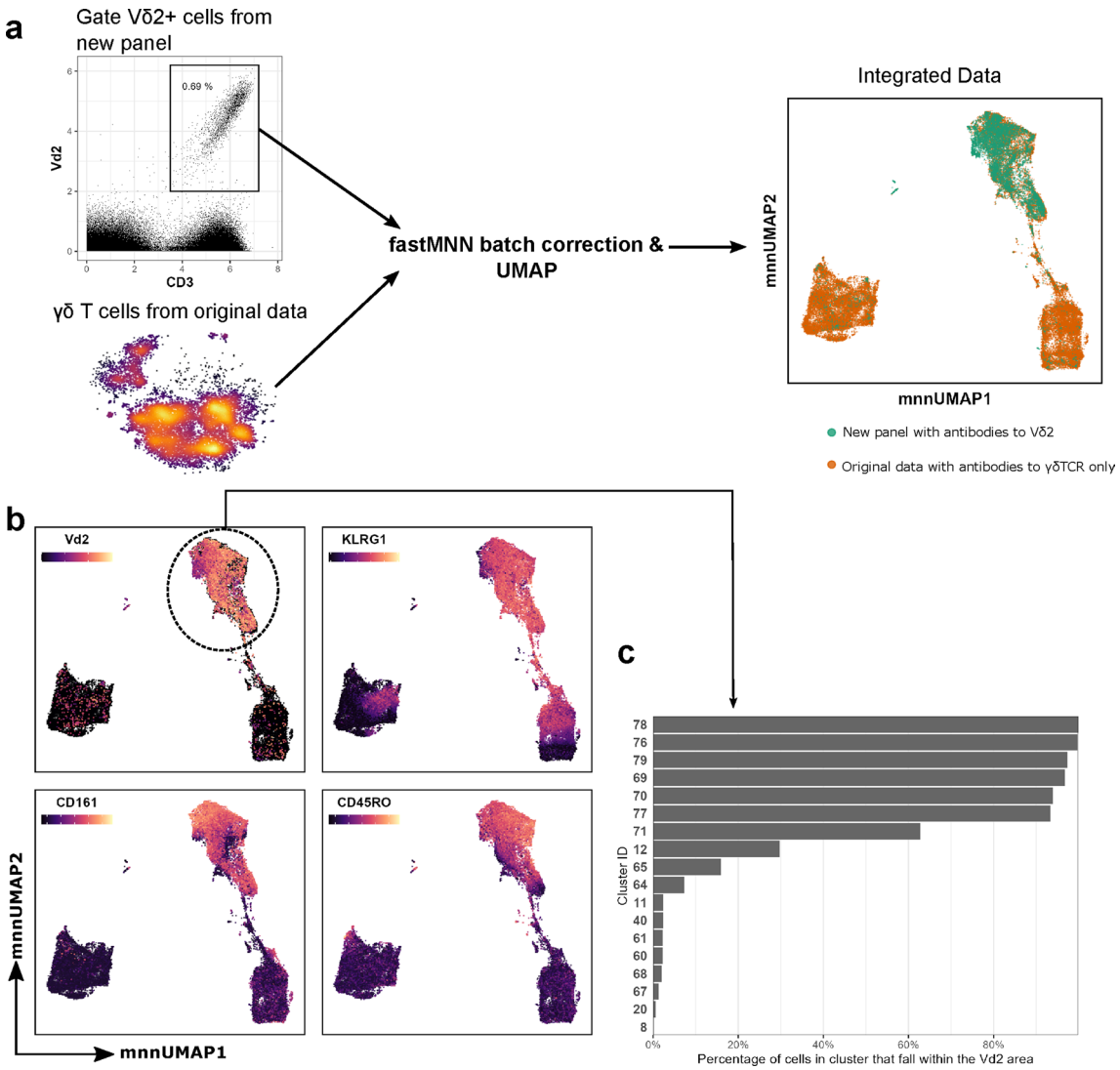
Extended Data Fig. 2 | Dot plots of the baseline differences at the lineage and subset levels. a) Pie charts displaying the baseline (C−1) median frequencies of immune cell populations at the lineage level for Europeans ($n = 5$) and Africans ($n = 20$). Global test within-test *P ≤ 0.05 and **P ≤ 0.01 (see Supplementary Table 4). **b**) Lineage level (relative to total cells) differences between Europeans and Africans at C−1, as tested by global test (see Supplementary Table 4). **c**) Subset level (relative to lineage) differences between Europeans and Africans at C−1, as tested by global test (see Supplementary Table 4). **d**) Overrepresented gene ontology biological process pathways in Europeans and Africans together with a circular graph depicting the genes that were differentially up- or downregulated between the Europeans and Africans between two time points.



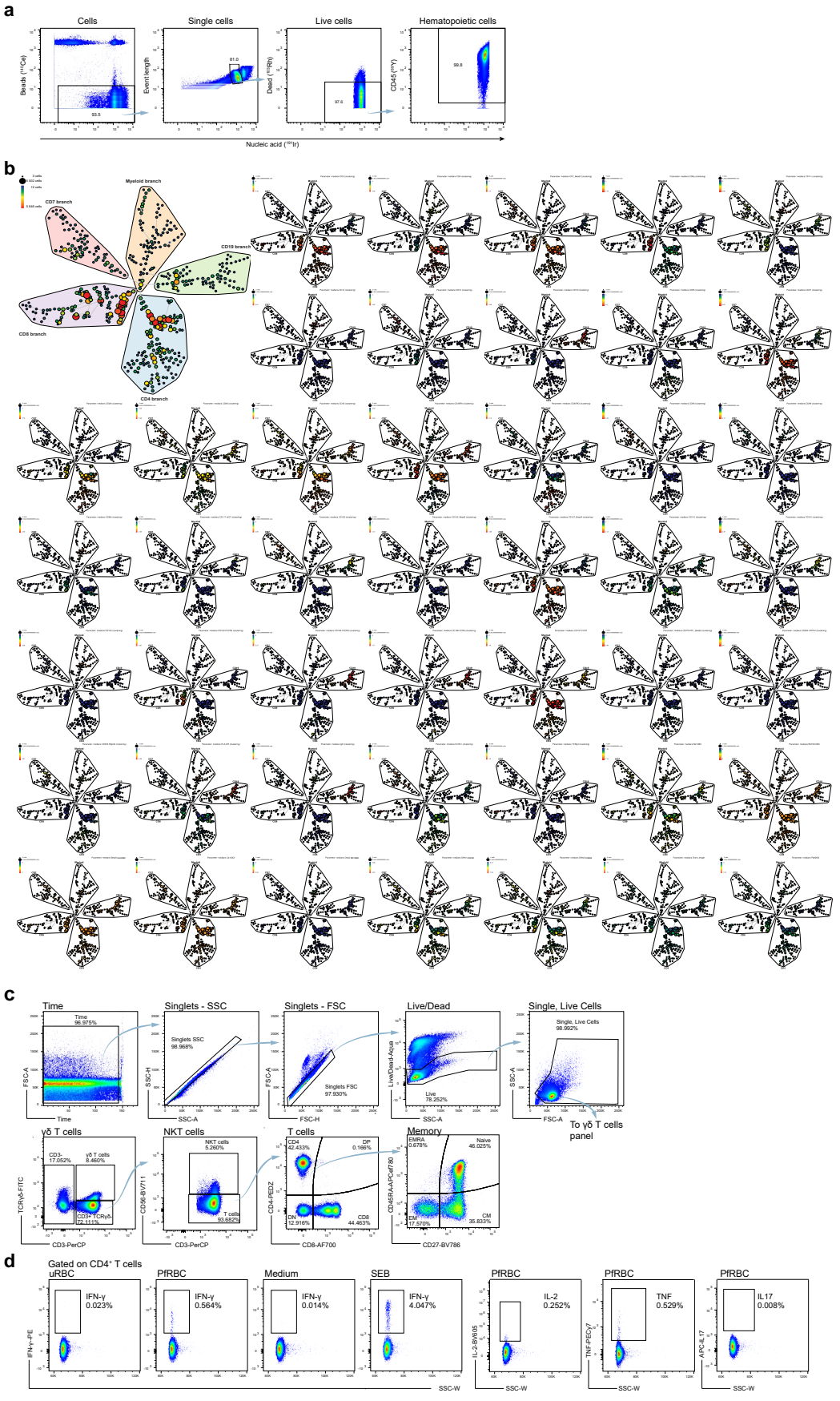
Extended Data Fig. 3 | The parasite density of individual study participants as determined by TBS per day after SPZ inoculation.



Extended Data Fig. 4 | Changes in abundance of cell subsets over time, from baseline (C-1) to 5 (D5) and 11 (D11) days after PfSPZ inoculation for TBS+ (n = 12) and TBS− (n = 8) Africans. Data is normalized relative to baseline. Change in percentage of EMRA CD4+ T cells relative to lineage, CD4+ γδ T cells relative to total, and pDCs and CD45RA+ γδ T cells relative to lineage. Shown are the individual datapoints, a Loess regression line, and its 95% confidence interval.

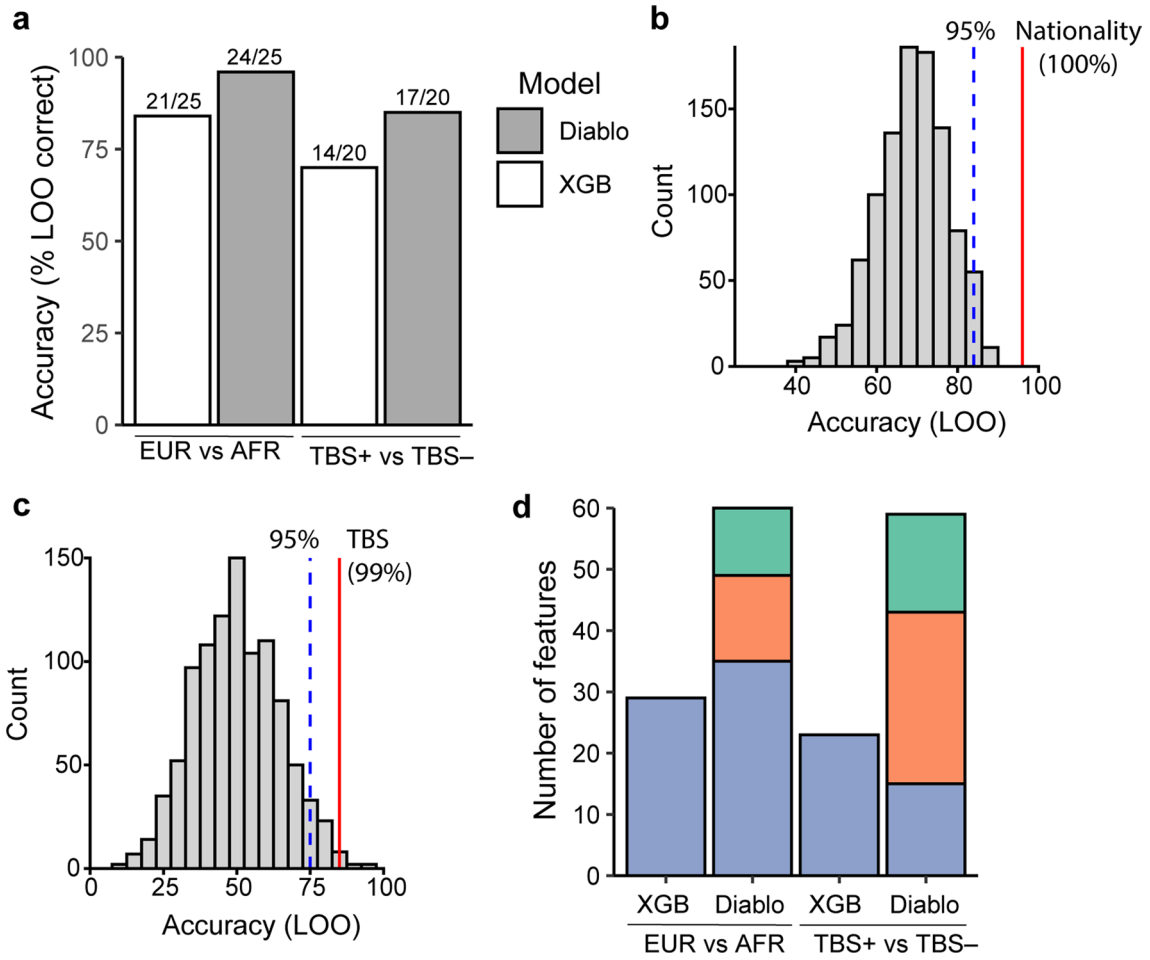


Extended Data Fig. 5 | Identification of Vδ2⁺ γδ T cells among cells stained using the old panel through data integration. a) Vδ2⁺ cells, which were stained using the Vδ2 panel, were gated and then integrated with the γδ T cells from the old panel. Batch effect correction was performed using the MNN method and then UMAP was performed on cells from both panels while excluding γδTCR and Vδ2 as a clustering parameter. **b**) Vδ2⁺ cells were identified on the UMAP embeddings and the expression of other markers were overlayed. **c**) The proportion of γδ T cell clusters within the projected Vδ2 cells.

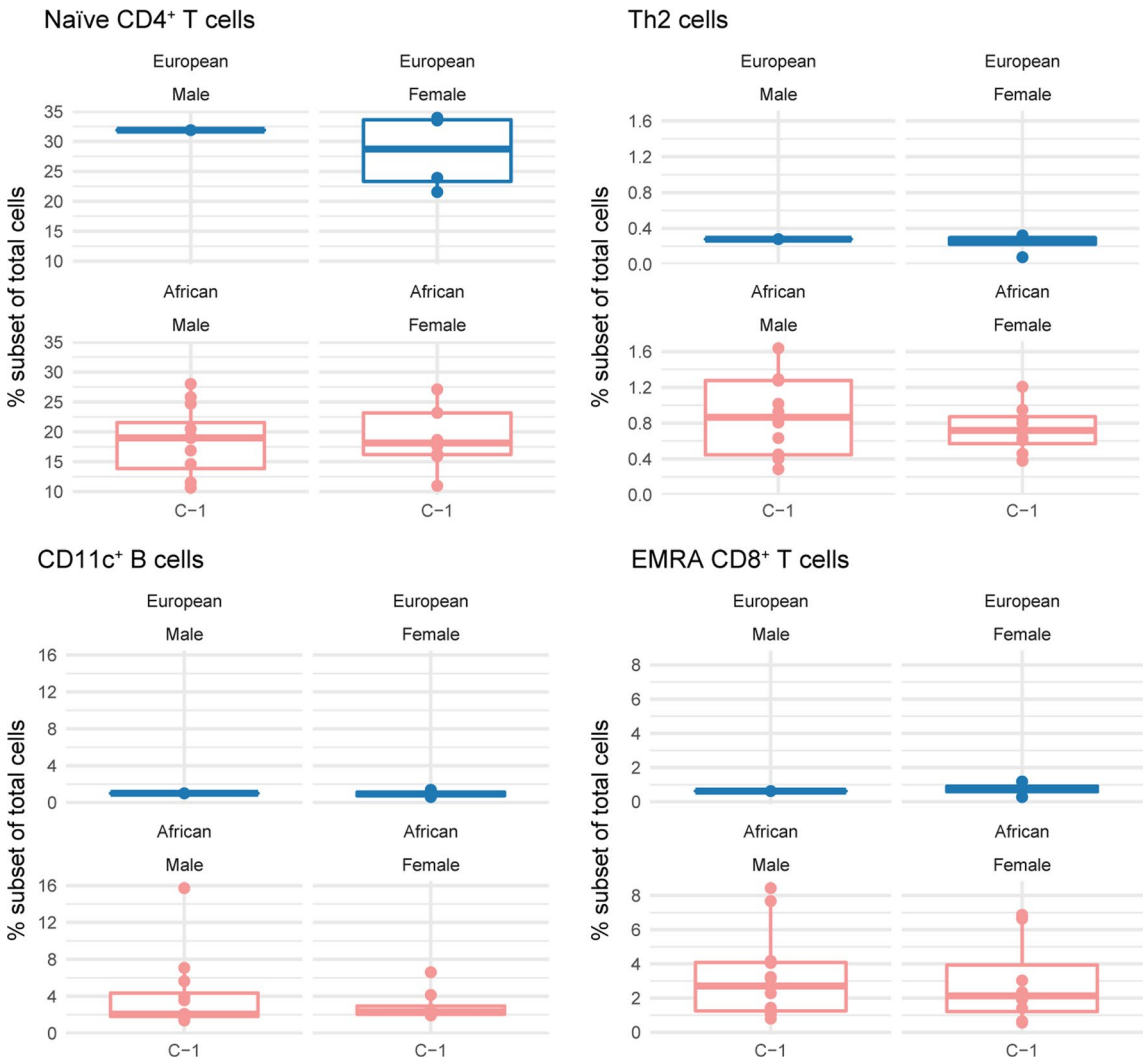


Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Mass cytometry and flow cytometry gating steps. a) First, mass cytometry-acquired cells were selected in FlowJo based on presence of DNA and absence of the metal isotope ^{140}Ce, which is only found on beads. Subsequently, single cells were selected, by excluding doublets with higher DNA content and event length. Then, live cells were selected, which were not stained by ^{103}Rh due to their impermeable cell membrane. Lastly, hematopoietic cells were selected based on their CD45 expression. Supplementary Table 2 lists the panel used. **b**) SPADE analysis of the single, live, CD45$^+$ cells, which resulted in a tree with 5 branches. The tree was based on 75 samples containing 33.3 million events, downsampled to 70,000 events. The size of the nodes is proportional to the number of clustered cells. The colors represent arcsinh-transformed expression levels of the different markers. **c**) Gating strategy of flow cytometry data is shown using time, singlets, and live cells before the gating of the subsets of interest, using the panel in Supplementary Table 5. **d**) Cytokine gates were set initially according to FMO control and adjusted according to negative controls for gating of cytokine-positive cells. The percentage of PfRBC-specific cytokine-producing cells was corrected by subtracting background levels as observed in uRBC conditions.



Extended Data Fig. 7 | Machine learning algorithm performance. a) Bar charts showing accuracy of DIABLO (grey) and Extreme Gradient Boosting (XGB, white) machine learning classifiers for prediction of Nationality or TBS using baseline data with leave-one out cross-validation (LOO). The number of correct predictions per model is indicated above the bars. **b**) Histogram showing accuracy of 1000 random permutations comparing all 25 subjects (with 20-5 split to be able to compare with the split Africans versus Europeans). Performance of each permutation was analyzed with LOO. The 95th percentile of accuracies (dotted blue lines) and the percentile of accuracy of the African-European split versus the random permutations (red line) are indicated. All random permutations had lower accuracy than comparing Europeans with Africans (95% accuracy). **c**) Histogram showing accuracy of 1000 random permutations comparing all 20 African subjects (with 12-8 split to be able to compare with the split TBS+ versus TBS– Africans). Performance of each permutation was analyzed with LOO. The 95th percentile of accuracies (dotted blue lines) and the percentile of accuracy of the TBS+ versus TBS– split versus the random permutations (red line) are indicated. There were 4 permutations with higher accuracy, 8 permutations with equal accuracy and 988 permutations with lower accuracy than observed when comparing susceptible and resistant Africans (85% accuracy). **d**) Stacked bar charts showing the number of identified cell clusters (CyTOF, green), genes (RNA-seq, blue) and responses (ICS, orange) that were observed in at least two folds of the LOO for both models and both comparisons.



Extended Data Fig. 8 | Percentage of naïve CD4⁺ T cell, T_H2 cell, CD11c⁺ B cell, or EMRA CD8⁺ T cell subsets relative to total cells at baseline (C-1), compared between male (n = 1) and female (n = 4) Europeans and male (n = 12) and female (n = 8) Africans. The boxplots show the median and 1st and 3rd quartiles. Whiskers are extending to the maximum/minimum of the respective groups, no further than 1.5x the IQR. All datapoints are overlaid on the boxplot.

nature research

Corresponding author(s): Maria Yazdanbakhsh

Last updated by author(s): Feb 17, 2021

Reporting Summary

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

Statistics

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

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

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

Software and code

Policy information about availability of computer code

Data collection	Software used include: CyTOF2 for mass cytometry and BD FACSDiva 8.0.1 for flow cytometry. See Methods section for complete descriptions.
Data analysis	Software and packages used include: FlowJo V10.2-10.7.1 for Mac (FlowJo LLC, Ashland, OR, USA), Cytobank 2015 (Cytobank Inc., Mountain View, CA, USA) , Cytosplore 2016-2021 (www.cytosplore.org), Matlab R2015b (The MathWorks, Inc., Natick, MA, USA), BIOPET Gentrap pipeline version 0.8 (https://biopet-docs.readthedocs.io/en/latest/pipelines/gentrap/), FastQC version 0.11.2, sickle version 1.33, Cutadapt version 1.10, Ensembl human genome annotation version 87, htseq-count version 0.6.1p1, RStudio version 0.99.902 for Windows (RStudio, Inc., Boston, MA, USA, www.rstudio.com), R x64 version 3.3.1 and 3.6.2 for Windows (R Foundation for Statistical Computing, Vienna, Austria, http://www.r-project.org/), the Globaltest R package, version 5.24.0 (http://bioconductor.org/packages/globaltest/), survival R package, version 2.38 (https://CRAN.R-project.org/package=survival), LIMMA package v3.28.14 (https://bioconductor.org/packages/limma/), IBM SPSS Statistics version 23 for Windows (IBM Corp., Armonk, NY, USA), edgeR R package, version 3.24.0, (http://bioconductor.org/packages/edgeR/), the limma R package, version 3.38.3 (http://bioconductor.org/packages/limma/), the Caret R package (v. 6.0-84), the R package mixOmics (v. 6.8.0), GraphPad Prism version 7 to 9 for Windows (GraphPad Software, San Diego, CA, USA), Morpheus heat map software version 2016-2018 (https://software.broadinstitute.org/morpheus), the Matlab dscatter function, version 1.1.0.1 (https://www.mathworks.com/matlabcentral/fileexchange/8430-flow-cytometry-data-reader-and-visualization), mixOmics (v. 6.8.0), the batchelor R package (v. 1.6.2), Adobe Illustrator CC version 2015-2021 (Adobe Inc, San Jose, CA, USA), and ArrayCAM 400-S Microarray Imager Software version 2.2. See Methods section for complete descriptions.

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

Data

Policy information about availability of data

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

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

The mass cytometry, flow cytometry, and RNAseq data that support the findings of this study are available in ImmPort (https://www.immport.org) under study accession SDY1734. The Homo sapiens GO database that was used (Gene Ontology Biological Process from MSigDB, version 6.2, July 2018) is available at https://www.gsea-msigdb.org/gsea/downloads_archive.jsp. Further information is available from the corresponding author upon reasonable request.

Field-specific reporting

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

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

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

Life sciences study design

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

Sample size	The samples used in the current study came from a predesigned clinical trial. This clinical trial was a controlled human infection study, which are generally relatively small due to the character of the study and novel nature, especially at the time. No statistical methods were used to predetermine sample size, but we followed standards in the field. See for reviews on controlled infection studies and sample size: https://doi.org/10.1016/S1473-3099(18)30177-4 and https://doi.org/10.1038/nri2902. A selection of CHMI papers looking at immunology: https://doi.org/10.1371/journal.ppat.1002389 https://doi.org/10.4049/jimmunol.1302960, and http://doi.org/10.1126/science.1241800.
Data exclusions	Only single (to exclude doublets), live (to exclude dead cells), CD45+ cells (to select for leukocytes) were analyzed from the mass cytometry data, as these are cells of interested and to avoid artifacts. During SPADE analysis, nodes (clusters of cells) comprising artifacts such as dead cells and doublets were further identified by combinations of high levels of the dead cell stain 103Rh, contaminating 138Ba and 206Pb, bead marker 140Ce, long event length, and/or expression of mutually exclusive markers such as CD3, CD14, and CD19. The exclusion of doublets, dead cells, and beads are common practice in the field. These exclusion criteria were pre-established.
Replication	The immune profiles and responses were analyzed by mass cytometry, flow cytometry, and RNA seq, as well as by multi-omics data integration, which confirmed and supplemented each other. The study involved 5 European subjects and 20 African subjects. The study was not replicated.
Randomization	The study subjects were not randomized, as we compared healthy Europeans and Africans that all received the same treatment. The samples were randomized as much as possible for all assays.
Blinding	Blinding during the clinical part of the study was not needed as all study participants received the same PfSPZ inoculation (no placebo was used) and it was not possible to hide whether participants belonged to the European or African group. The investigators were partially blinded during data collection or analysis. For example, during experiments, the samples were transferred to numbered tubes, so that the samples could not be directly identified during handling. The samples were also randomized during these experiments and measurements. Most phases of the analysis were data driven, so that blinding in those stages was not necessary. Samples were only analyzed in the final stages on a level that made comparison of groups possible.

Reporting for specific materials, systems and methods

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

Materials & experimental systems

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

Methods

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

Antibodies

Antibodies used

See Supplementary Table 2 and 5 for the mass and flow cytometry panels.

Antibodies used for mass cytometry were: 89Y-CD45 (clone HI30, Fluidigm, cat. 3089003B, RRID AB_2661851, dilution 1/100), 141Pr-CD196 (CCR6) (clone G034E3, Fluidigm, cat. 3141003A, RRID AB_2687639, dilution 1/100), 142Nd-CD19 (clone HIB19, Fluidigm, cat. 3142001B, RRID AB_2651155, dilution 1/100), 143Nd-CD117 (c-Kit) (clone 104D2, Fluidigm, cat. 3143001B, RRID AB_2687630, dilution 1/100), 144Nd-CD86 (clone IT2.2, BioLegend, cat. 305435, RRID AB_2563764, dilution 1/200), 145Nd-CD4 (clone RPA-T4, Fluidigm, cat. 3145001B, RRID AB_2661789, dilution 1/100), 146Nd-CD8a (clone RPA-T8, Fluidigm, cat. 3146001B, RRID AB_2687641, dilution 1/200), 147Sm-CD183 (CXCR3) (clone G025H7, BioLegend, cat. 353733, RRID AB_2563724, dilution 1/50), 148Nd-CD16 (clone 3G8, Fluidigm, cat. 3148004B, RRID AB_2661791, dilution 1/100), 149Sm-CD25 (IL-2Ra) (clone 2A3, Fluidigm, cat. 3149010B, RRID AB_2756416, dilution 1/100), 150Nd-CD185 (CXCR5) (clone 51505, R&D Systems, cat. MAB190-100, RRID AB_2292654, dilution 1/100), 151Eu-CD123 (clone 6H6, Fluidigm, cat. 3151001B, RRID AB_2661794, dilution 1/100), 152Sm-TCRγδ (clone 11F2, Fluidigm, cat. 3152008B, RRID AB_2687643, dilution 1/50), 153Eu-CD7 (clone CD7-6B7, Fluidigm, cat. 3153014B, dilution 1/100), 154Sm-CD163 (clone GHI/61, Fluidigm, cat. 3154007B, RRID AB_2661797, dilution 1/100), 155Gd-CD69 (clone FN50, BioLegend, cat. 310939, RRID AB_2562827, dilution 1/200), 156Gd-CD294 (CRTH2) (clone BM16, BioLegend, cat. 350102, RRID AB_10639863, dilution 1/100), 158Gd-CD122 (IL-2Rb) (clone TU27, BioLegend, cat. 339015, RRID AB_2563712, dilution 18264), 159Tb-CD197 (CCR7) (clone G043H7, Fluidigm, cat. 3159003A, RRID AB_2714155, dilution 1/100), 160Gd-CD14 (clone M5E2, Fluidigm, cat. 3160001B, RRID AB_2687634, dilution 1/100), 161Dy-KLRG1 (MAFA) (clone REA261, Miltenyi Biotech, special order, dilution 1/50), 162Dy-CD11c (clone Bu15, Fluidigm, cat. 3162005B, RRID AB_2687635, dilution 1/200), 163Dy-IgD (clone IA6-2, BioLegend, cat. 348235, RRID AB_2563775, dilution 1/200), 164Dy-CD161 (clone HP-3G10, Fluidigm, cat. 3164009B, RRID AB_2687651, dilution 1/100), 165Ho-CD127 (IL-7Rα) (clone AO19D5, Fluidigm, cat. 3165008B, dilution 1/200), 166Er-CD141 (BDCA3) (clone 1A4, BD Biosciences, cat. 559780, RRID AB_397321, dilution 1/200), 167Er-CD27 (clone O323, Fluidigm, cat. 3167002B, dilution 1/100), 168Er-HLA-DR (clone L243, BioLegend, cat. 307651, RRID AB_2562826, dilution 1/200), 169Tm-CD45RA (clone HI100, Fluidigm, cat. 3169008B, dilution 1/100), 170Er-CD3 (clone UCHT1, Fluidigm, cat. 3170001B, RRID AB_2661807, dilution 1/100), 171Yb-CD28 (clone CD28.2, BioLegend, cat. 302937, RRID AB_2563737, dilution 1/100), 172Yb-CD38 (clone HIT2, Fluidigm, cat. 3172007B, RRID AB_2756288, dilution 1/200), 173Yb-CD45RO (clone UCHL1, BioLegend, cat. 304239, RRID AB_2563752, dilution 1/100), 174Yb-CD335 (NKp46) (clone 92E, BioLegend, cat. 331902, RRID AB_1027637, dilution 1/50), 175Lu-CD279 (PD-1) (clone EH 12.2H7, Fluidigm, cat. 3175008B, RRID AB_2687629, dilution 1/100), and 176Yb-CD56 (clone NCAM16.2, Fluidigm, cat. 3176008B, RRID AB_2661813, dilution 1/100). For the Vd2 panel, the following antibodies were also used: TCR Vd2-157Gd (clone B6, BioLegend, cat. 331402, RRID AB_1089226, 1/100 dilution), CD86-198Pt (clone IT2.2, BioLegend, cat. 305435, RRID AB_2563764, 1/200 dilution), CD14-148Nd (clone M5E2, BioLegend, cat. 301843, RRID AB_2562813, 1/100 dilution), CD16-209Bi (clone 3G8, Fluidigm, cat. 3209002B, 1/200 dilution), and CD185-150Nd (clone J252D4, BioLegend, cat. 356902, RRID AB_2561811, 1/100 dilution), as well as barcoding with β2-microglobulin (BioLegend, cat. 316302, RRID AB_492835, 1/50 dilution) coupled to 106Cd, 110Cd, 111Cd, 112Cd, 114Cd, and 116Cd.

Antibodies used for flow cytometry were: PerCP-CD3 (clone SK7, BioLegend, cat. 344813, RRID AB_10641841, dilution 1/80), FITC-TCRγδ (clone 11F2, BD Biosciences, cat. 347903, RRID AB_400358, dilution 1/40), APC-eFluor780-CD45RA (clone HI100, BioLegend, cat. 304127, RRID AB_10708419, dilution 1/640), AlexaFluor700-CD8 (clone RPA-T8, BioLegend, cat. 301027, RRID AB_493744, dilution 1/1280), APC-IL-17 (clone BL168, BioLegend, cat. 512334, RRID AB_2563986, dilution 1/80), BV786-CD27 (clone L128, BioLegend, cat. 563327, dilution 1/40), BV711-CD56 (clone NCAM 16.2, BD Biosciences, cat. 563169, RRID AB_2738043, dilution 1/80), BV650-CD14 (clone M5E2, BioLegend, cat. 301836, RRID AB_2563799, dilution 1/20), BV605-IL-2 (clone MQ1-17H12, BioLegend, cat. 500332, RRID AB_2563877, dilution 1/160), LIVE/DEAD Fixable Aqua-Viability (ThermoFisher Scientific, cat. L34957, dilution 1/400), BV421-CD279 (PD-1) (clone EH12.2H7, BioLegend, cat. 329920, RRID AB_10960742, dilution 1/20), PE-Cy7-TNF (clone MAb11, BioLegend, cat. 502929, RRID AB_2204080, dilution 1/1280), PE-Dazzle594-CD4 (clone 3G8, BioLegend, cat. 302054, RRID AB_2563639, dilution 1/640), and PE-IFN-γ (clone 4S.B3, BioLegend, cat. 502509, RRID AB_315234, dilution 1/80).

Validation

Well know antibody clones were chosen, titrated, and validated on human peripheral blood mononuclear cell (PBMC) samples comparable to or from this study. Fluorescence-minus-one (FMO) controls, (unstimulated) culture control conditions, and comparison to expression patterns described in literature and on the manufacturer's website were used for validation.

Fluidigm states that "each lot of conjugated antibody is quality control tested by CyTOF® analysis of stained cells using the appropriate positive and negative cell staining and/or activation controls". BioLegend states that "all newly developed clones undergo validation testing for multiple applications. This serves as a cross-check for specificity and provides clarity for research uses. Typically, antibodies are tested by two or more of these methods: flow cytometry, Western blot, chromatin immunoprecipitation, immunofluorescence, immunohistochemistry, biofunctional assay". As well as these controls: "endogenous positive and negative expressing cells, knockdown/knockout cell lines, isotype controls, biological testing, i.e. treatments, activitation etc., multiple donor samples from human cells, and side-by-side comparison with currently available

clones." Clone 51505 for CD185 was used d by HLDA to establish the CD designation. BD Biosciences states that "Our product development process includes testing on a combination of primary cells, cell lines and/or transfectant cell models with relevant controls using multiple immunoassays to ensure biological accuracy. We also perform multiplexing with additional antibodies to interrogate antibody staining in multiple cell populations." Miltenyi Biotec states for the KLRG1 antibody, clone REA261: "In order to compare the epitope specificity of an antibody, the clone being used is compared with other known clones recognizing the same antigen in a competition assay. And "selected fluorochrome conjugated antibodies from Miltenyi Biotec were compared to commercially available hybridoma clones in flow cytometry analysis".

The given RRID codes for the antibodies provide more information at for example www.scicrunch.org and www.antibodyregistry.org.

Human research participants

Policy information about studies involving human research participants

Population characteristics	The trial consisted of 5 malaria-naïve European adults and 20 malaria-exposed African adults. Of the study participants, 13 were male and 12 female, 16 had the HbAA haemoglobin genotype and 9 had sickle cell trait (HbAS), age varied from 18.6 to 28.8 years old, and BMI from 16.2 to 27.0. See Supplementary Table 1 for the full characteristics of the study population.
Recruitment	Volunteers recruited were healthy adults between 18 and 30 years of age, without a history of severe chronic diseases and without a history of intake of an antimalarial drug in the preceding weeks. Pregnancy and the intention to become pregnant during the study were exclusion criteria. Volunteers were screened and allocated if applicable to either the Europeans group, who were malaria naive or minimally exposed without sickle cell trait (Europeans visiting or living in Gabon) or the Africans group, with lifelong malaria exposure. Malaria naivety was defined as no history of Pf malaria and no long-term (> 5 years) residence in a malaria-endemic area. Lifelong malaria exposure was defined as long-term (> 10 years) residence in a highly malaria-endemic area. See Methods section and Lell et al. (Am. J. Trop. Med. Hyg., 2018, DOI: 10.4269/ajtmh.17-0343) for more information.
Ethics oversight	The trial followed International Council for Harmonization of Technical Requirements for Pharmaceuticals for Human Use Good Clinical Practice guidelines and the principles of the Declaration of Helsinki and was performed in accordance with guidelines approved by the national ethics committee, le Comité Nationale d'Ethique pour la Recherche, Gabon's regulatory authority, and under a United States Food and Drug Administration Investigational New Drug application. Written informed consent was obtained from all participants.

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

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	NCT02237586
Study protocol	ClinicalTrials.gov, identifier NCT02237586
Data collection	The clinical trial, including sample collection, is described in the dedicated paper on https://doi.org/10.4269/ajtmh.17-0343
Outcomes	The clinical trial is described in the dedicated paper on https://doi.org/10.4269/ajtmh.17-0343. The original study outcomes are also described on ClinicalTrials.gov. These are: Primary Outcome Measures : - Days from inoculation to start of malaria episode [Time Frame: From day of injection until day 28]: The time from parasite inoculation to first detection of malaria will be assessed by thick blood film microscopy and a clinical questionnaire. Malaria is defined as both parasitemia and clinical symptoms suggestive of malaria. - Frequency, incidence and nature of adverse events [Time Frame: From day of injection until day 28]: The safety of PfSPZ Challenge administered IV and the resultant P. falciparum infection will be assessed by analysing actively and passively collected data from clinical review of volunteers and laboratory measurements. Secondary Outcome Measures : - Dynamics of P. falciparum parasite growth [Time Frame: From day 6 after injection until approximately day 28]: The dynamics of P. falciparum parasite growth following administration of PfSPZ Challenge is assessed by analyzing parasite multiplication rates using highly sensitive qPCR for P. falciparum DNA. Measurements will be used to model parasite kinetics and to estimate the number of infected liver cells. Other Outcome Measures: - Cellular and humoral immune responses against parasites [Time Frame: From screening until 6 months]: Cellular and humoral immune responses against parasites will be assessed by individual and collective results of cytometry, enzyme-linked immunospot analysis (ELISpot), enzyme-linked immunosorbent assay (ELISA), immunofluorescence assay (IFA), inhibition of sporozoite invasion of hepatocytes assay (ISI), quantification of erythrophagocytosis, B-cell typing and isolation as well as protein microarray. - Stage specific expression patterns of parasite genes [Time Frame: From screening until 6 months]: Stage specific expression patterns of parasite genes will be assessed by RNA quantification using reverse transcriptase PCR (rtPCR) and transcriptional

profiling on microarray and sequencing platforms. Differences in miRNA expression between AA and AS erythrocytes will be determined and related to multiplication rates.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Measurement of samples by flow cytometry was also randomized per subject to avoid bias by staining and measurement day. Samples belonging to the same subject were stained and measured together. After thawing, the fixed cells were stained in 96-wells V-bottom microplates with 50 µL of antibody mixture (see Supplementary Table 5 for panel) diluted in eBioscience permeabilization buffer (Thermo Fisher Scientific) with 1% human Fc-receptor binding inhibitor (Thermo Fisher Scientific; cat. 14-9161) at 4 °C for 30 minutes. Leftover cells were pooled and then split for use as FMO controls. Compensation beads (BD CompBead, BD Biosciences) were made fresh every day. BD FACSDiva 8.0.1. CST setup beads (BD Biosciences, lot 83654) were run before every measurement. A median of 141.800 live, single cells was obtained per sample.
Instrument	BD LSR-Fortessa X-20 SORP. See Supplementary Table 5 for specifications.
Software	FlowJo software (version 10.4.1; TreeStar).
Cell population abundance	An example of typical cell abundances can be found in Supplementary Figure S3.
Gating strategy	FMO controls and medium, uRBC, and SEB control conditions were used to determine the cutoff between negative and positive populations. An example of the gating strategy can be found in Supplementary Figure S3.

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