University of Groningen

Urban living in healthy Tanzanians is associated with an inflammatory status driven by dietary and metabolic changes

Temba, Godfrey S; Kullaya, Vesla; Pecht, Tal; Mmbaga, Blandina T; Aschenbrenner, Anna C; Ulas, Thomas; Kibiki, Gibson; Lyamuya, Furaha; Boahen, Collins K; Kumar, Vinod

Published in:
Nature Immunology

DOI:
[10.1038/s41590-021-00867-8](https://doi.org/10.1038/s41590-021-00867-8)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2021

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Temba, G. S., Kullaya, V., Pecht, T., Mmbaga, B. T., Aschenbrenner, A. C., Ulas, T., Kibiki, G., Lyamuya, F., Boahen, C. K., Kumar, V., Joosten, L. A. B., Schultze, J. L., van der Ven, A. J., Netea, M. G., & de Mast, Q. (2021). Urban living in healthy Tanzanians is associated with an inflammatory status driven by dietary and metabolic changes. *Nature Immunology*. <https://doi.org/10.1038/s41590-021-00867-8>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Urban living in healthy Tanzanians is associated with an inflammatory status driven by dietary and metabolic changes

Godfrey S. Temba^{1,2,3,11}, Vesla Kullaya^{1,2,4,11}, Tal Pecht^{5,6,11}, Blandina T. Mmbaga⁴, Anna C. Aschenbrenner^{5,6}, Thomas Ulas^{6,7}, Gibson Kibiki^{4,8}, Furaha Lyamuya⁴, Collins K. Boahen^{1,3}, Vinod Kumar^{1,3,9}, Leo A. B. Joosten^{1,3}, Joachim L. Schultze^{5,6,7,12}, Andre J. van der Ven^{1,3,12}, Mihai G. Netea^{1,3,10,12}✉ and Quirijn de Mast^{1,3,12}✉

Sub-Saharan Africa currently experiences an unprecedented wave of urbanization, which has important consequences for health and disease patterns. This study aimed to investigate and integrate the immune and metabolic consequences of rural or urban lifestyles and the role of nutritional changes associated with urban living. In a cohort of 323 healthy Tanzanians, urban as compared to rural living was associated with a pro-inflammatory immune phenotype, both at the transcript and protein levels. We identified different food-derived and endogenous circulating metabolites accounting for these differences. Serum from urban dwellers induced reprogramming of innate immune cells with higher tumor necrosis factor production upon microbial re-stimulation in an in vitro model of trained immunity. These data demonstrate important shifts toward an inflammatory phenotype associated with an urban lifestyle and provide new insights into the underlying dietary and metabolic factors, which may affect disease epidemiology in sub-Saharan African countries.

Sub-Saharan Africa (SSA) is currently experiencing an unprecedented wave of urbanization¹, which is associated with a sharp increase in noncommunicable diseases (NCDs), alongside a continuing high burden of infectious diseases^{2,3}. Important shifts during this rural-to-urban transition are changes in physical activity and diet, often referred to as ‘nutritional transition’^{4,5}. Common dietary changes include replacement of traditional staple foods, rich in starch and dietary fiber and plant protein sources such as legumes, with energy-dense, processed foods. Studies have highlighted an important role of diet in inflammation and immunity^{6,7} and diet is increasingly recognized as an important determinant of both communicable diseases and NCDs. The impact of dietary and bacterial metabolites on the function of immune cells is increasingly recognized as a mechanism of these effects⁶.

To date, biological changes associated with rural-to-urban migration in SSA are only partly elucidated⁸. The current wave of urbanization in SSA provides unique opportunities to investigate the biological effects of urbanization and nutritional transition and how they interact to determine health and disease. Urbanization is associated with important changes in different biological systems that interact closely, including metabolic and endocrine pathways^{9,10}, the immune system¹¹, psychological functioning¹² and adipose and musculoskeletal systems¹³. In this study, we aimed to understand

and integrate the metabolic and immune consequences of rural-urban transition and their interaction with demographic, environmental and dietary factors in a cohort of healthy Tanzanian adults.

Our results reveal that the unstimulated whole blood transcriptome from urban dwellers demonstrates a distinctive pro-inflammatory transcriptional programming compared to that of rural dwellers, which is further validated by the finding that urban dwellers exhibit an increased pro-inflammatory cytokine production capacity. We identified different food-derived and endogenous metabolites that may account for these differences and highlight the importance of seasonality on the blood immune-metabolic system. Together, these findings provide important new insights into the underlying mechanisms through which urbanization and the nutritional transition affect the epidemiology of inflammatory diseases and how efficient use of natural resources may help to promote the overall health of populations.

Results

Study demographics. A total of 383 individuals were screened, of whom 323 were enrolled. Of these, 253 (78.3%) were from an urban area and 70 (21.7%) were from a rural area (Fig. 1). A summary of study participant characteristics and measured parameters are given in Table 1 and Extended Data Fig. 1. The rural cohort was older

¹Department of Internal Medicine, Radboud University Medical Center, Nijmegen, the Netherlands. ²Department of Medical Biochemistry and Molecular Biology, Kilimanjaro Christian Medical University College, (KCMUCo), Moshi, Tanzania. ³Radboud Centre for Infectious Diseases (RCI), Radboud University Medical Center, Nijmegen, the Netherlands. ⁴Kilimanjaro Clinical Research Institute, Kilimanjaro Christian Medical Center, Moshi, Tanzania.

⁵Department for Genomics and Immunoregulation, Life & Medical Sciences (LIMES) Institute, University of Bonn, Bonn, Germany. ⁶Systems Medicine, German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany. ⁷PRECISE Platform for Single Cell Genomics and Epigenomics, German Center for Neurodegenerative Diseases (DZNE) and University of Bonn, Bonn, Germany. ⁸East African Health and Research Commission (EAHRC), Bujumbura, Burundi. ⁹Department of Genetics, University Medical Centre Groningen, University of Groningen, Groningen, the Netherlands. ¹⁰Department for Immunology and Metabolism, Life & Medical Sciences (LIMES) Institute, University of Bonn, Bonn, Germany. ¹¹These authors contributed equally: Godfrey S. Temba, Vesla Kullaya, Tal Pecht. ¹²These authors jointly supervised this work: Joachim L. Schultze, Andre J. van der Ven, Mihai G. Netea, Quirijn de Mast. ✉e-mail: Mihai.Netea@radboudumc.nl; Quirijn.deMast@radboudumc.nl

Fig. 1 | Schematic depiction of the study population and distribution. a, Map of Kilimanjaro region showing sampled areas (red and blue dots) and sample size from both urban and rural populations with the distribution of sex, use of contraceptives in females and use of tobacco among males. **b**, Histograms showing distribution of age and BMI in urban- and rural-living participants and BMI in females and males.

than the urban cohort with a median (interquartile range (IQR)) age of 39.7 years (32.4–49.4 years) versus 27.6 years (22.6–38.2 years; $P < 0.001$). Females had a higher body mass index (BMI) compared to males, but there were no significant differences in BMI between females or males from the urban or rural cohort. Overall, 19% of women used oral contraceptives and none of them smoked, whereas 29% of men were smokers. A full list of characteristics and dietary information is given in Supplementary Table 1.

Increased interferon pathway transcriptional programming and pro-inflammatory cytokine responses in urban participants. To define major differences among urban- and rural-living individuals, we first examined unstimulated total whole blood transcriptomes analyzed by RNA sequencing (RNA-seq) (Supplementary Table 2). Using principal-component analysis (PCA), we observed a shift in the distribution in the PCA space across PC1 and PC2 (Fig. 2a), indicating that the global transcriptome of the urban and rural populations differs. Using gene set enrichment analysis (GSEA) for MSigDB hallmark gene sets, we identified interferon (IFN)- α and IFN- γ response pathways to be enriched in the urban versus rural dwellers (Fig. 2b and Supplementary Table 3). Within the genes that contributed most to the enrichment of the IFN- γ response pathway (core-enrichment genes), 20 genes also showed significant differential expression

between urban versus rural dwellers, albeit with low effect size, which is not surprising comparing healthy individuals under homeostatic conditions (Fig. 2c and Supplementary Table 4). Among these genes, we identified the main transcription factors of the IFN pathway such as *STAT1* and *STAT2* and downstream effectors such as transcription factor of the IFN regulatory factor 1 (*IRF1*) and transporter associated with antigen processing 1 (*TAP1*) (Fig. 2c).

Next, we tested the functional consequences of these transcriptional changes. Except for interleukin (IL)-1 β , plasma concentrations of IL-1 receptor antagonist (IL-1Ra), IL-6, IL-18 and IL-18 binding protein (IL-18BP) were similar in urban versus rural dwellers (Fig. 2d). However, ex vivo stimulation of whole blood with different bacterial, fungal and Toll-like receptor (TLR) agonists, yielded significantly higher concentrations (corrected for age and sex) of pro-inflammatory cytokines IFN- γ , tumor necrosis factor (TNF) and IL-6 in urban-living participants, whereas the anti-inflammatory cytokine IL-10 was higher in rural-living participants (Fig. 2e,f). An exception was stimulation with TLR3 ligand poly(I:C), which was associated with higher cytokine responses in rural dwellers.

Considering the higher median age of the rural cohort (Table 1), we also analyzed the blood transcriptome and cytokine responses in the subgroup of participants aged 30 years and above. The median

Table 1 | Sociodemographic characteristics of the study population

	All (n = 323)	Urban (n = 253)	Rural (n = 70)	P value
Sex, female (n, %)	164 (51)	137 (54.2)	27 (38.6)	0.042
Age, years (median, IQR)	30.3 (23.4–40.2)	27.6 (22.6–38.2)	39.7 (32.4–49.4)	<0.001
Age categories (n, %)				
18–30 years	165 (51.0)	152 (60.1)	13 (18.6)	<0.001
31–45 years	98 (30.7)	68 (26.9)	31 (44.3)	
46–80 years	59 (18.3)	33 (13.0)	26 (37.1)	
BMI (median, IQR)	23.8 (21.5–27.2)	24.2 (21.5–27.3)	23.6 (21.3–27.1)	NS
BMI by sex (median, IQR)				
Male	22.8 (20.8–24.9)	22.8 (20.7–24.9)	22.5 (21.0–25.1)	NS
Female	25.6 (22.6–30.0)	25.5 (22.6–29.9)	25.7 (21.9–33.2)	NS
BMI classification (n, %)				
<18.5	15 (4.6)	13 (5.1)	2 (2.9)	NS
18.5–24.9	181 (56.0)	138 (54.6)	43 (61.4)	
25.0–29.9	79 (24.5)	64 (25.3)	15 (21.4)	
>30.0	48 (14.9)	38 (15.0)	10 (14.3)	
Systolic blood pressure, mm Hg (median, IQR)	122 (113–136)	120 (110–135)	129 (120–138)	0.007
Diastolic blood pressure, mm Hg (median, IQR)	79 (70–81)	77 (70–80)	80 (72–86)	0.029
Random blood sugar, mmol l ^{−1} (median, IQR)	5.1 (4.6–5.7)	5.1 (4.6–5.8)	5.0 (4.7–5.4)	NS
Tobacco use (n, % of males) ^a	46 (28.9)	32 (27.1)	14 (34.1)	NS
Use of contraceptives (n, % females)	31 (18.9)	24 (17.5)	7 (25.9)	NS
Highest level of education (n, %)				
Primary	122 (37.8)	77 (33.2)	45 (66.2)	<0.001
Secondary	70 (21.7)	58 (22.7)	12 (17.6)	
College	61 (18.9)	56 (22.0)	5 (7.4)	
University	25 (7.7)	22 (8.6)	3 (4.4)	
In training	45 (13.9)	42 (16.5)	3 (4.4)	
Occupational status (n, %)				
Students	58 (18.0)	54 (21.2)	4 (5.9)	<0.001
Employed with qualification	67 (20.7)	60 (23.5)	7 (10.3)	
Service workers and shop sales workers (for example, cooks, market and shops salespersons)	134 (41.5)	107 (42.0)	27 (39.7)	
Elementary occupations (for example, unskilled farmhands and street vendors)	64 (19.8)	34 (13.3)	30 (44.1)	
Toilet facility (n, %)				
Pit latrine	77 (24.1)	52 (20.7)	25 (36.8)	0.006
Water closet	242 (75.9)	199 (79.3)	43 (63.2)	
Cooking fuel (n, %) ^b				
Smoky	115 (35.6)	69 (27.1)	46 (67.6)	<0.001
Non-smoky	208 (64.4)	186 (71.9)	22 (32.4)	
Exposure to animals, yes (n, %)	143 (44.3)	91 (35.7)	52 (76.5)	<0.001

Comparison between urban and rural populations was performed using Fisher's exact, chi-squared and Mann-Whitney *U*-test for categorical and continuous variables, respectively. ^aUse of tobacco was only analyzed in males as none of the women in this cohort used tobacco. ^bCategories of cooking fuel included firewood, charcoal or kerosene (smoky) or gas and electrical (non-smoky). NS, not significant.

(IQR) age of the urban- ($n = 106$) and rural-living ($n = 58$) participants in this subgroup was similar (39.76 yr (35.97–46.14 yr) versus 41.8 yr (36.30–50.31 yr); $P = 0.12$ by Mann-Whitney *U*-test). This analysis revealed that the major differences in the transcriptome and cytokine secretion across urban and rural dwellers were pre-served (Supplementary Fig. 1a–c).

Age and sex may influence cytokine responses^{14,15}. Previous studies have highlighted a gradual decline in cytokine production with advancing age, a phenomenon referred to as 'immunosenescence'.¹⁶

In both groups, IFN- γ , IL-6 and TNF responses declined with age, but this decline was more pronounced in the urban cohort (Fig. 3a). Regarding the effect of sex, in general, males had higher cytokine responses to the different stimuli, except for responses to *Staphylococcus aureus* and *Candida albicans* yeast, which were higher in females (Fig. 3b). We also tested whether genetic differences may explain the immune differences between urban and rural cohorts; however, clustering analysis of the genetic data from the same individuals showed no separate grouping of rural and

urban cohorts (Supplementary Fig. 2a,b). Overall, these results demonstrate that urban compared to rural living is associated with a pro-inflammatory transcriptional program of circulating blood immune cells, which is already present in a homeostatic state and may underlie increased immune responses upon various microbial-related stimulations.

Food-derived and endogenous metabolites explain differences in cytokine responses. Considering the observation that genetic variation, age and sex differences do not explain immune differences between rural and urban dwellers, we postulated that differences in diet and lifestyle explain part of this variation and that these are likely mediated by metabolites in the circulation. We therefore, assessed the plasma metabolome in all individuals using a method previously described¹⁷. Untargeted plasma metabolomics yielded a total of 2,101 annotated compounds, 795 of which had unique KEGG IDs (<https://www.genome.jp/kegg/>; Supplementary Table 5). To test whether rural versus urban living constitutes the different circulating metabolite abundance, we first performed univariate analysis, Student's *t*-tests as provided in MetaboAnalyst pipeline (<https://www.metaboanalyst.ca>; Methods) to identify compounds/metabolites that were significantly different between cohorts. A total of 348 metabolites were significantly elevated in the rural cohort and 157 metabolites were elevated in the urban cohort (Supplementary Table 6). Next, using a general linear model we analyzed independent effects of rural versus urban living on plasma metabolome after adjusting for age and sex differences. Biologically relevant metabolites with the highest significance were aggregated in exogenous (food-derived), exogenous/endogenous or endogenous metabolites. In the rural cohort, abundant metabolites with the highest significance were predominantly food-derived, including flavones (apigenin having the highest annotation, genistein and 2'-hydroxydaidzein having the same ion *m/z*) (Fig. 4a,b). These flavones are found in fruit and vegetables, but high concentrations are also found in the traditional Tanzanian staple cereals millet and sorghum¹⁸. Results from the food history showed that rural-living individuals consume more vegetables and ugali (a traditional thick porridge dish made from millet, sorghum, cassava or maize flour) and banana beer (a home-brewed alcoholic beverage made from banana juice and millet) (Supplementary Table 7). Other abundant metabolites in the rural cohort were ascorbate, the disaccharide sugar trehalose, the nucleoside inosine and metabolites of the tricarboxylic acid cycle (for example, succinate, malate and citrate). In contrast, the most abundant metabolites in the urban cohort were intermediate metabolites of the primary bile acid and cholesterol metabolisms, such as 7 α -hydroxyl-cholesterol, 3 β ,7 α -dihydroxy-

5-cholestane and 3 α ,7 α ,12 α -trihydroxy-5 β -cholestane and different steroid hormone metabolites (Fig. 4a,b). Unbiased metabolic pathway analysis, which combines functional enrichment analysis and network topology analysis, also demonstrated important differences in metabolic phenotype between the cohorts. The most significantly enriched pathways in the urban cohort were the 'primary bile acid biosynthesis', 'histidine metabolism' and 'steroid hormone biosynthesis', whereas the 'alanine, aspartate and glutamate metabolism' and the 'tricarboxylic acid cycle' were upregulated in the rural cohort (Fig. 4c and Supplementary Table 8).

Next, we sought to determine whether observed differences in metabolites in the blood influence molecular programming of circulating cells. As different circulating metabolites were shown to modulate the phenotype of immune cells¹⁹, we hypothesized that differences in circulating metabolite abundance between urban- and rural-living individuals may be translated to changes in whole blood transcriptome. To define the potential molecular effect by each metabolite, we used publicly available data for differential expression (DE) analysis to define up- and downregulated gene signatures for each metabolite (Fig. 5a and Supplementary Table 9; Methods). We utilized these specific metabolite gene signatures to test whether there is a potential molecular effect by each of the metabolites by applying a single-sample GSEA (ssGSEA)²⁰. This analysis determines an estimate for the activity level of a set of genes (GSEA) for each individual (Fig. 5a). For apigenin and dimethyl-itaconate (DMI), but not trehalose, we found a statistically significant dependency between residence status and metabolite-specific gene signatures, with a higher enrichment in rural-living individuals (Fig. 5a and Supplementary Fig. 3). Such dependency was not present for the metabolite upregulated gene signatures (Supplementary Fig. 3). Overall, these results indicate that increased circulating levels of apigenin and itaconate in the rural population associate with transcriptional changes in the whole blood transcriptome.

To validate metabolite-related functional differences, we assessed associations between specific metabolites and cytokine responses and validated these effects experimentally (Methods). Notably, most of the metabolites that were abundant in the rural cohort, including the flavones, inosine, amino acids (for example, pyruvate, citrate) and trehalose, correlated negatively with pro-inflammatory cytokine responses, whereas an opposite trend was found for metabolites that were abundant in the urban cohort (Fig. 5b). We validated the anti-inflammatory effects of three of the abundant metabolites in the rural cohort (apigenin, trehalose and itaconate) in an ex vivo model of cytokine induction, using peripheral blood mononuclear cells (PBMCs) from the blood of nine healthy Dutch volunteers. Pre-incubation for 2 h with apigenin or trehalose significantly

Fig. 2 | Associations of blood transcriptomes, ex vivo cytokine immune responses and plasma cytokines with urban or rural living. **a**, PCA performed on transformed corrected counts of all expressed genes in the urban (*n* = 247) and rural (*n* = 69) subcohorts. Distribution across PC1 and PC2 is presented in marginal histogram plots. The mean differences of the cohort's distribution in PC1 (*P* < 0.0001) or PC2 (*P* = 0.007) were calculated by a two-tailed Student's *t*-test. ***P* < 0.01, ****P* < 0.001. **b**, Dot plot presenting GSEA MSigDB hallmark of significantly enriched pathways (normalized enrichment score (NES) > 1.5 and *P*_{adj} < 0.1) in urban and rural areas. Dot size represents the number of core-enrichment genes. **c**, GSEA plot of the MSigDB hallmarks IFN- γ response pathway GSEA plot presents the distribution of the gene set according to their position in the pre-ranked gene list used for the GSEA analysis (*x* axis) (top). In the same plot, the purple line represents the running enrichment score (*y* axis) and the dashed line depicts the core-enrichment genes. Core-enrichment genes are plotted by their log₂ (fold change) (*x* axis) and *P*_{adj} value ($-\log_{10}(P_{adj})$) according to the DE analysis. Labeled genes (independent hypothesis weighting) had *P*_{adj} < 0.2 (value marked at dashed line) (bottom left). Gene expression of representative core-enriched genes with *P*_{adj} < 0.2 presented in combined violin-box-dot plot: STAT1 (*P*_{adj} = 0.049), STAT2 (*P*_{adj} = 0.069), IRF1 (*P*_{adj} = 0.007) and TAP1 (*P*_{adj} = 0.002) (bottom right). **d**, Combined dot-violin-box plots showing the plasma concentrations of the inflammatory cytokines IL-1 β (*P* = 0.041), IL-1Ra (*P* = 0.49), IL-6 (*P* = 0.17) and IL-18 (*P* = 0.22) comparing urban (*n* = 250) versus rural (*n* = 70) biologically independent samples using a two-tailed Mann-Whitney *U*-test. **e**, Heat map showing linear regression model β coefficients of ex vivo cytokine immune responses in urban or rural dwellers (corrected for age and sex). Purple and green colors indicate higher cytokine responses in urban versus rural areas, respectively. **P* < 0.05 and ***P* < 0.01; false discovery rate (FDR) corrected. **f**, Combined dot-violin-box plots depict cytokine production of IFN- γ (*n* = 230 and 66) and TNF (*n* = 230 and 67) by LPS stimulation and IFN- γ (*n* = 238 and 67) and TNF (*n* = 234 and 68) by *S. aureus* stimulation in residential cohorts (urban and rural, respectively). *P* values were calculated by two-tailed Mann-Whitney *U*-test; ****P* < 0.001 and *****P* < 0.0001. In all box plots, the in-box line defines the median value, hinges depict 25th and 75th percentiles and whiskers extend to ± 1.5 IQR. MTB, *Mycobacterium tuberculosis*; FC, fold change.

reduced TNF in the culture supernatant in response to lipopolysaccharide (LPS) stimulation for 24 h without affecting cell viability (Fig. 5c), whereas DMI significantly reduced IFN- γ production in response to heat-killed *C. albicans* yeast stimulation for 48 h (Fig. 5c). Different molecular mechanisms underlying the anti-inflammatory

effect of apigenin were previously reported, including inhibition of caspase-1 activation through the disruption of the NOD-like receptor pyrin domain containing 3 (NLRP3) inflammasome assembly, inhibition of ERK1/2 activation and inhibition of cytokine-induced nuclear factor (NF)- κ B activation²¹. In LPS-treated macrophages²²,

apigenin reduced TLR-4, MyD88 and p-I κ B- α expression levels via the NF- κ B p65 signaling pathway. Finally, apigenin was also shown to inhibit cyclooxygenase-2 expression and reduce monocyte adhesion.^{21,23}

Immune modulating effects of other metabolites identified here were previously shown, including anti-inflammatory effects exerted by the endogenous purine inosine²⁴ and pyruvate²⁵ and pro-inflammatory effects of the glycolipid sulfatide²⁶, leukotrienes B4 (ref. ²⁷), prostaglandin F2 (ref. ²⁸) and histamine (reviewed previously²⁹). Collectively, these findings provide evidence that urban compared to rural living has a clear impact on blood transcriptome and the plasma metabolome, with the latter holding an important secondary immune effect.

Serum from urban participants induces ‘trained immunity’. Considering that the abundance of plasma immune metabolites may be related to transcriptional programming of circulating immune cells, we hypothesized that metabolites in the circulation may induce long-term programming of the immune cells. Long-term adaptation of innate immune cells into an enhanced functional state has been referred to as ‘trained immunity’^{30,31}. Not only microbial triggers, such as Bacille Calmette–Guérin (BCG) but also sterile dietary triggers, including a ‘western-type’ diet, induce trained immunity⁷. To assess whether serum from urban or rural dwellers induces functional reprogramming of immune cells, we pooled serum from the urban or rural dwellers that exhibited the lowest (first quartile) or highest (fourth quartile) TNF responses. The number of samples pooled in both extremities was as follows: high responders (urban, $n=68$; rural, $n=7$) and low responders (urban, $n=51$; rural, $n=25$). Isolated monocytes from 12 healthy donors were subsequently exposed to 10% pooled serum from either these urban or rural dwellers for 24 h. Cells were then washed and rested in a standard culture medium for 5 d, followed by re-stimulation 5 d later with LPS (10 ng ml⁻¹) for 24 h. Serum from urban volunteers from the highest TNF production quartile primed monocytes to respond with significantly higher TNF production upon re-stimulation, which was comparable to the priming effect of BCG used as a positive control. In contrast, serum from either the urban participants from the lowest TNF quartile or rural participants did not induce a pro-inflammatory phenotype of monocytes (Fig. 5d). Next, we assessed whether apigenin prevents immune reprogramming by plasma from urban-living individuals. Oxidized low-density lipoprotein (oxLDL) was included as a positive control, as it is known to induce trained immunity in atherosclerosis³². Pre-exposure of monocytes for 1 h to apigenin indeed reduced TNF production upon re-stimulation with LPS, suggesting that food-derived apigenin prevents functional reprogramming of monocytes (Fig. 5d). Collectively, these results show that plasma from urban-living individuals may induce immune reprogramming of circulating monocytes, which may be prevented by dietary-derived metabolites such as apigenin.

Impact of annual seasonality on whole blood cytokine immune responses, transcriptome and plasma metabolome. To further

relate dietary changes to cytokine production, we studied the effect of seasonality on immune responses and the metabolome. The study area in Northern Tanzania is located at 3° south of the Equator but has a clear seasonal variation in precipitation with a wet and dry season (Fig. 6a). Daily diet during the dry season in the town of Moshi may more closely resemble the rural diet due to the availability of staples in the harvest period. We found a strong association of cytokine immune responses, with seasonality showing the highest cytokine production during the wet season, especially in the urban cohort (Fig. 6b and Supplementary Fig. 4). Therefore, we divided our cohort into three groups according to the season at which participants were recruited: urban wet, urban dry and rural. In GSEA analysis, regardless of seasonality, both IFN- α and IFN- γ response pathways emerged as a prominent enriched gene sets in the urban cohorts (urban wet and urban dry) compared to the rural cohort (Fig. 6c and Supplementary Table 10). Metabolic pathways such as cholesterol metabolism were enriched in urban-wet compared to urban-dry individuals (Fig. 6c). Moreover, oxidative phosphorylation and mTORC1 signaling were constantly and exclusively enriched in urban-wet versus urban-dry or rural individuals. Among the genes that contribute most to this enrichment and also identified in the DE analysis (adjusted P (P_{adj}) < 0.2, Supplementary Table 11), we identified genes related to mitochondria, electron transport chain, cell proliferation and/or DNA damage repair, stress response and T cell function (Fig. 6d and Supplementary Table 11). Exploring both the transcriptome and metabolome by PCA revealed that the urban-dry and rural subcohorts were grouped closer to one another, whereas the urban-wet subcohort were grouped separately in the PCA space (Fig. 6e). Furthermore, for the metabolite apigenin and its upregulated as well as downregulated gene signature, there was a notable season-dependent enrichment in urban-living individuals during the wet season compared to the dry season and also to the rural cohort (Fig. 6f and Supplementary Fig. 5a). This corresponded with the lower plasma concentration of apigenin specifically in the urban-wet compared to the rural and urban-dry individuals (Fig. 6f). For itaconate, though the concentration of this metabolite was increased in the rural and urban-dry cohort, the downregulated and upregulated gene signatures did not reach statistical significance in chi-squared or subsequent post hoc tests (Supplementary Fig. 5b,c). Overall, these results showed season-related changes in the immune profile, whereby food-derived metabolites may relate to changes in the transcriptional program of blood immune cells.

To further explore the contribution of dietary habits to immunofunctional changes, we performed additional analyses considering only food-derived metabolites (Supplementary Table 12). In a PCA of the food-derived metabolome, the urban-wet and urban-dry subpopulations varied in their distribution across PC1, with apigenin being among the top-ten most represented metabolites (Fig. 7a and Supplementary Fig. 6a). Hierarchical clustering of the food-derived metabolome within the urban cohort resulted in four different food-metabolome clusters: pink, green, turquoise and lilac. These four clusters were significantly different in the composition of urban-dry and urban-wet individuals (chi-squared test, $P < 0.001$),

Fig. 3 | Impact of age and sex on cytokine production capacity. **a**, Heat maps showing linear regression model β coefficient of ex vivo whole blood cytokine responses and age (corrected for sex) in the total cohort, urban or rural living (top). Red and blue colors indicate higher and lower cytokine responses, respectively, with increasing age. P values are FDR corrected, $^*P < 0.05$, $^{**}P < 0.01$ and $^{***}P < 0.001$. Cytokine production of IFN- γ ($n=233$), IL-6 ($n=225$) and TNF ($n=227$) in response to *Streptococcus pneumoniae* depicted in the scatter plots (bottom), presenting R -values calculated by Spearman correlation. **b**, Heat map showing linear regression model β coefficient (corrected for age) of ex vivo cytokine responses with sex in the total cohort, urban or rural living (top). Red and blue colors indicate higher cytokine responses in males versus females, respectively. P values are FDR corrected, $^*P < 0.05$, $^{**}P < 0.01$ and $^{***}P < 0.001$. Combined dot-violin-box plots presenting the cytokine production of IL-10 ($n=147$ and 142, $P < 0.001$) and IL-1 β ($n=153$ and 147, $P < 0.001$) in response to *Escherichia coli*, IL-6 in response to *S. aureus* ($n=147$ and 143, $P=0.0012$) and TNF in response to *C. albicans* yeast ($n=152$ and 147, $P=0.0067$) in females and males, respectively (bottom). In all box plots, in-box line defines the median value, hinges are 25th and 75th percentiles, whiskers extend to ± 1.5 IQR and notches depict 95% confidence interval (CI) around the median. log₂ fold change is listed for each comparison, colored by the respective group with upregulated secretion. P values were calculated by two-sided Mann–Whitney U-test. $^*P < 0.05$, $^{**}P < 0.01$ and $^{***}P < 0.001$.

with the pink cluster being exclusively represented by urban-dry individuals ($n=70$). The ratio of urban-dry to urban-wet individuals in the green, turquoise and lilac clusters were 9:1, 45:23

and 50:55, respectively (Fig. 7b). Apigenin concentration was statistically different among these clusters, with the pink cluster presenting the highest concentration (Fig. 7c). Sex, age and BMI

Fig. 4 | Differences in plasma metabolite abundance in rural- versus urban-living individuals. a, Heat map showing metabolite abundance in rural- (green) and urban-living individuals (purple). Metabolites are aggregated in exogenous (for example, food-derived), exogenous/endogenous or endogenous metabolites. Plotted are *t*-statistic values of the metabolites' mean difference in urban ($n=250$) versus rural ($n=70$) populations calculated by two-tailed Student's *t*-test with FDR correction; the level of significance is marked * $P < 0.05$, ** $P < 0.01$ and *** $P < 0.001$. **b**, Combined dot-violin-box plots presenting the abundance of example metabolites significantly enriched in the rural cohort (apigenin, $P < 0.0001$) or in the urban cohort ($3\alpha,7\alpha$ -dihydroxy- 5β -cholestane, $P < 0.0001$). Box plot in-box line defines the median value, hinges depict 25th and 75th percentiles, whiskers extend to ± 1.5 IQR, and notches depict 95% CI around the median. Presented are results of the Mann-Whitney *U*-test for the comparison between urban ($n=250$) and rural ($n=70$) areas. **c**, Scatter plot showing metabolic pathway analysis between urban and rural cohorts. Pathway enrichment was determined by the hypergeometric test, whereas topology analysis was performed using relative-betweenness centrality using *Homo sapiens* (KEGG) pathway library. Data analysis in **a,c** was performed using MetaboAnalyst v.4.0.

were not significantly different among these metabolic clusters (Supplementary Fig. 6b). Weekly food frequency consumption of ugali, rice, cooked banana (plantain), wheat and green vegetables differed significantly among metabolome clusters and consumption of green vegetables showed a trend (Fig. 7d). Individuals from the pink cluster more frequently consumed ugali, rice and green vegetables and less frequently consumed wheat, whereas those from the lilac cluster consumed less ugali and green vegetables and those from the green cluster largely refrained from eating rice, while eating wheat more frequently.

Both circulating cytokine concentrations and cytokine production significantly differed among metabolome clusters. Regarding circulating cytokines and adipokines, the pink cluster had significantly lower concentrations of resistin, α -1 antitrypsin and IL-6 than the turquoise cluster and lower concentrations of resistin and IL-1 β than the lilac cluster (Fig. 7e and Supplementary Fig. 6c). Furthermore, the pink cluster also had a lower cytokine production capacity to microbial stimulation (Fig. 7f,d and Supplementary

Fig. 6d). We then performed DE analysis of whole blood transcriptome, defining DE genes among the pink and the green, turquoise or lilac metabolome clusters ($P_{\text{adj}} < 0.05$ and absolute fold change > 1.2 ; Supplementary Fig. 6e and Supplementary Tables 13 and 14). Gene ontology enrichment analysis (GOEA) of DE genes demonstrated that the pink cluster was characterized by downregulation of gene signatures related to IFN and pathogenic response pathways. Downregulated DE genes in the pink versus turquoise cluster were enriched for gene ontology (GO) terms related to viral response, such as 'Defense response to virus', 'Type-1 IFN-signaling pathway' and 'Negative regulation of viral process' (Fig. 7h and Supplementary Table 15). Compared to the lilac cluster, downregulated DE genes in the pink cluster were enriched for GO terms linked to antimicrobial, response to fungus, IFN- γ and coagulation, including 'Defense response to Gram-negative bacterium', 'Defense response to fungus', 'Cellular response to IFN- γ ' and 'Platelet degranulation' (Fig. 7h, Supplementary Fig. 6e and Supplementary Tables 14 and 15). Downregulated DE genes in the pink versus green cluster were

enriched for GO terms related to humoral response and phagocytosis (Supplementary Fig. 6f and Supplementary Tables 14 and 15). Among the nine genes that were commonly upregulated in the turquoise or lilac clusters versus the pink cluster was the expression of 2'-5'-oligoadenylate synthetase 1 (*OAS1*), which is upregulated by type-1 IFN and promotes an antiviral response³³. Another commonly upregulated gene was radical S-adenosyl methionine domain containing 2 (*RSAD2*), also known as viperin, an IFN-inducible protein that is responsible for various antiviral responses to different viruses³⁴ (Fig. 7g). Collectively these results show that changes in food-derived metabolites in the urban cohort associate with seasonality and present a strong relationship with immunological changes in whole blood, including alterations in circulating cytokine levels, cytokine production and transcriptional reprogramming (Supplementary Fig. 6g).

Discussion

This study comprehensively assessed the effects of rural–urban residence on the plasma metabolome, immune cell transcriptome, inflammation and host defense. We show that the transcriptional profile of urban dwellers is different than that of rural dwellers at steady-state conditions, with an IFN-signaling program emerging as a transcriptional signature for urban residency. We further demonstrate that healthy Tanzanian urban dwellers have increased pro-inflammatory cytokine production compared to rural dwellers and identify different food-derived and endogenous metabolites underlying these differences in cytokine profiles. Among the rural-abundant metabolites investigated, we demonstrate that flavones such as apigenin play a substantial role in modulating immune responses, reversing long-term pro-inflammatory reprogramming of monocytes *ex vivo* as well as potentially affecting the transcriptional program of blood immune cells. Considering seasonality, enrollment in the wet season was related to differences in metabolic transcriptional profile. Finally, we explored in an unbiased manner, the role of food-derived metabolites as modulators of the innate immune response in the urban cohort, demonstrating that changes in food-derived metabolites associate with seasonality and dietary habits as well as inflammatory status. Collectively, these data provide evidence for a real-life and direct regulatory effect of diet and food-derived metabolites on the innate immune system, which has potentially important implications at different levels.

Our findings suggest that urban living and dietary changes may affect the pathophysiology and outcome of infections through modulation of important host defense mechanisms such as cytokine production. These changes are complex and could influence both positively the host defense through improved responsiveness, but can also potentially lead to increased tissue damage due to hyperinflammation. Previous studies have shown that mice fed a

western diet have increased sepsis severity and mortality^{35,36}. This is especially relevant as infectious diseases are still a dominant public health problem in many parts of SSA. Second, immune responses to vaccination may also be affected by diet and season of administration. So far, the literature on this topic is still very limited and this aspect warrants future studies. Third, the findings of an inflammatory bias between individuals living in a rural or urban area provide insight into the mechanisms contributing to changing disease patterns and the current epidemic of NCDs in SSA. Inflammation is a key process in the pathogenesis of NCDs³⁷ and studies show that a western diet induces systemic inflammation and a long-term epigenetic reprogramming of innate immune cells.⁷ Our study shows that serum from urban dwellers induces a 'trained immunity' phenotype, which, as recently reviewed,³⁸ is very likely to contribute to atherosclerosis and other inflammatory diseases. This finding is also relevant for populations outside SSA that suffer from the health consequences of unhealthy diets and lifestyle. Nonetheless, it is important to consider that the process of urbanization in Tanzania is relatively recent and still ongoing and that these may still be a long way from what is seen in some populations in high-income countries.

We have also identified several important plasma metabolites that have a modulatory effect on cytokine production. An important observation was that the flavones, such as apigenin, were among the most abundant plasma metabolites in rural-living individuals. Traditional diets in Tanzania contain high amounts of products containing flavones, including millet, sorghum and vegetables.¹⁸ These flavones have anti-inflammatory, antioxidants and anticancer properties,^{39,40} and we validated their anti-inflammatory properties in our study. Plant-based diets are generally considered beneficial for cardiovascular health⁴¹ and our current findings support the notion that promoting traditional diets is a valid public health intervention to combat the rapidly growing epidemic of NCDs in low-income countries. The disaccharide sugar trehalose was also abundant in rural dwellers and we validated its anti-inflammatory properties⁴² *in vitro*. However, this should be considered in the light of findings showing that trehalose enhances the virulence of epidemic *Clostridium difficile* strains⁴³. Short-chain fatty acids (SCFAs) that are produced during fermentation of dietary fiber and proteins such as acetate, propionate and butyrate, have emerged as important regulators of inflammation. These SCFA were not identified by our untargeted metabolomics platform, possibly because the liver clears the major part of SCFAs from the portal circulation.

An interesting observation was the fact that poly(I:C) stimulation showed an opposite effect on cytokine responses compared with the other stimuli (rural dwellers exhibited higher cytokine responses to poly(I:C) compared with urban dwellers). The TLR3 ligand poly(I:C) is known to induce a distinct inflammatory

Fig. 5 | Relation of plasma metabolites with *ex vivo* cytokine production capacity. **a**, Illustration describing the definition of metabolite down- or upregulated gene signatures by DE analysis of publicly available data as described in Methods (left). Histograms presenting the distribution of GSEA for downregulated genes of apigenin or itaconate calculated by ssGSEA for the complete cohort (gray) and the urban (purple) and rural (green) subcohorts (right). The vertical solid black line depicts the all-cohort median enrichment score (M-ES). The percentage of individuals with enrichment scores above or below the all-cohort M-ES is presented in a mosaic plot. *P* value and effect size for the dependency of residence status and enrichment score was calculated via the chi-squared test (*P* = 0.0004 and *P* = 0.0294 for apigenin and itaconate, respectively). **b**, Heat map showing associations between *ex vivo* cytokine immune responses and metabolites significantly enriched either in the rural or urban cohorts. Depicted are correlation coefficients (*r*). Red and blue colors indicate positive and negative correlations, respectively. Analysis was performed using Pearson partial correlation, corrected for age, sex and BMI. **c**, Combined bar-dot plot (mean ± s.e.m.) showing the anti-inflammatory effect of apigenin (*n* = 9) and trehalose (*n* = 6) on TNF and DMI (*n* = 6) on IFN-γ production by PBMCs following either 24 h stimulation for apigenin and trehalose or 48 h for DMI. **d**, Combined bar-dot plot (mean ± s.e.m.) presenting human monocytes trained with serum from either urban or rural participants from the highest (upper quartile) and lowest (lower quartile) responders of TNF production capacity for 24 h and re-stimulated with *E. coli* LPS at day 6. The concentration of TNF was measured in the supernatant (*n* = 12). BCG was included as a positive control of immune training. The addition of apigenin (10 μM) during the first 24 h inhibited trained immunity induced by either oxLDL or serum from participants from the highest responders of TNF. *P* values were determined with a two-sided Mann-Whitney *U*-test. (**P* < 0.05, ***P* < 0.01 and ****P* < 0.001). The *n* values in **c,d** represent biologically independent samples from three independent experiments. HK, heat-killed; NS, not significant.

Fig. 6 | Impact of annual seasonality on blood transcriptome and cytokine immune responses. **a**, Monthly average of total precipitation as recorded in 2017 in Moshi, Tanzania. **b**, Heat maps presenting general linear regression model β coefficients of whole blood ex vivo cytokine responses (corrected for age and sex) and seasonality, dry versus wet season. $^*P < 0.05$, $^{**}P < 0.01$ and $^{***}P < 0.001$, FDR-adjusted P values. Red and blue colors indicate higher cytokine responses in the wet versus dry season, respectively. **c**, Circular dot plot presenting GSEA results of MSigDB hallmark pathways enriched in urban-wet, urban-dry and rural subcohorts. The comparison is listed outside of the circle, dividing the circle into thirds (heavy lines). Dot positions: perimeter indicates NES, radius indicates $-\log_{10} P_{\text{adj}}$ value (Benjamini-Hochberg), dashed line indicates P_{adj} of 0.1. Dot color: light-gray, black, purple and red indicate enrichment: not significant, significant, significant exclusive and constant in the urban-wet cohort and significant and constant in all urban cohorts, respectively. Dot size represents the pathway's gene set core size. **d**, Heat map presenting the group average-normalized corrected expression of DE genes contributing to enrichment of oxidative phosphorylation and mTORC1 signaling hallmark pathways in urban-wet, urban-dry and rural cohorts. **e**, PCA performed on transformed corrected counts of all expressed genes (transcriptome; left) and plasma metabolite $\log_2$ scaled values (right) in the urban-wet, urban-dry and rural groups. Distribution across PC1 and PC2 is presented in marginal histogram plots. Differences between the groups' mean distribution was calculated by two-tailed Student's t -test ($^*P < 0.05$, $^{**}P < 0.01$ and $^{***}P < 0.001$). **f**, Histograms presenting the distribution of apigenin upregulated gene signature enrichment scores calculated by ssGSEA for all cohorts and the urban-wet, urban-dry or rural subcohorts (top). The vertical solid black line depicts the all-cohort median enrichment score (M-ES). The percentage of individuals with enrichment scores above or below the all-cohort M-ES is presented in a mosaic plot. P value and effect size for the dependency of residence status and enrichment score were calculated via chi-squared test ($P = 0.0115$) and post hoc chi-squared test ($P = 0.0168$ for urban wet). $^*P < 0.05$ in the post hoc test. **g**, Combined dot-violin-box plots for apigenin $\log_2$ scaled values in the urban-wet ($n = 79$), urban-dry ($n = 171$) and rural ($n = 70$) subcohorts. Differences between group median values were calculated by a two-sided Mann-Whitney U -test ($^{****}P < 0.0001$). Box plot in-box line defines the median value, hinges depict 25th and 75th percentiles, whiskers extend to ± 1.5 IQR, and notches depict 95% CI around the median.

signature in monocytes compared with LPS and other stimuli. Its actions are independent of the MyD88 pathway but mediated by TRAF signaling, which may explain why apigenin, which exerts its anti-inflammatory actions among others through the MyD88 pathway⁴⁴, did not inhibit cytokine responses to poly(I:C) in the rural population.

Another notable observation of our study is the association of seasonality, as well as age and sex, on innate immune responses in the setting of a low- and middle-income country. We previously showed a seasonal pattern in cytokine production in a cohort of healthy individuals residing in the Netherlands¹⁴. Moreover, our group and other researchers have also reported that gene expression in human immune cells presents annual seasonality patterns^{14,45}. This line of evidence highlights seasonal dietary changes as an environmental factor that may contribute to altered susceptibility and the manifestation of diseases in the urban setting. Regarding the effect of aging, it has been shown that there is a gradual decline in T cell-derived cytokines (IFN- γ and IL-22) with age in a Western European population.^{14,46} In our current study, not only T cell-derived IFN- γ but also the monocyte-derived cytokines IL-6 and TNF declined with age, which is relevant as populations in low-income countries are

also aging rapidly. Regarding sex differences, males had in general, higher cytokine responses than females, except for responses to *S. aureus* and *C. albicans* yeast, which were significantly higher in females. Sex-specific differences in immune responses have been reported elsewhere⁴⁷ (also reviewed previously⁴⁸).

Limitations of our study are that more urban than rural dwellers were enrolled and that the rural cohort had higher mean age, even though we corrected for age differences in our analyses and our findings were maintained even when limiting the analysis to individuals aged 30 years and above. Furthermore, the effect of physical inactivity that is related to an urban lifestyle cannot be excluded⁴⁹. However, the BMI of the rural and urban cohort was not different in our study.

In conclusion, our study demonstrates that urban dwellers in Tanzania have a pro-inflammatory phenotype compared to rural dwellers. Serum from urban dwellers induced functional reprogramming of innate immune cells into an inflammatory phenotype. This could be inhibited by the flavonoid apigenin, supporting the importance of traditional, plant-based diets, rich in flavonoids and other anti-inflammatory compounds, for modulating disease epidemiology in SSA and beyond.

Fig. 7 | Association of urban individuals' food-derived metabolome on cytokine immune responses and blood transcriptome. **a**, PCA performed on plasma food-derived metabolites ($n = 297$ metabolites) in urban-wet ($n = 79$) and urban-dry ($n = 174$) subcohorts. Distribution across PC1 and PC2 is presented in marginal histogram plots. Differences between group mean distribution was calculated by a two-tailed Student's t -test ($P < 0.0001$ and $P = 0.182$, for PC1 and PC2, respectively). **b**, Agglomerative hierarchical clustering of individuals from the urban cohort according to the food-derived metabolome. Presented are annotations for seasonality, sex, age and food-metabolome cluster. **c**, Combined dot-violin-box plots for apigenin $\log_2$ scaled values in the metabolome clusters pink, green, turquoise and lilac ($n = 70, 10, 68$ and 105 individuals, respectively). Dots are colored according to seasonality. Differences between clusters calculated by one-way ANOVA ($P < 0.0001$). **d**, Weekly food consumption of ugali ($P = 0.036$), rice ($P = 0.002$), cooked banana/plantain ($P = 0.002$), wheat ($P = 0.047$) and green vegetables ($P = 0.070$) in the metabolic clusters. Dependency among the clusters and weekly food consumption frequency categories (0–7 meals per week) was tested using the chi-squared test, and post hoc chi-squared test results are marked in plots ($^*P < 0.05$ and $^{\#}P < 0.1$). **e,f**, Heat maps showing general linear regression model β coefficients (corrected for age and sex) of circulating plasma inflammatory mediators (**e**) or whole blood ex vivo cytokine response values (**f**) and the metabolome clusters pink and turquoise or pink and lilac. FDR-corrected P values are presented inside the heat map ($^*P < 0.05$, $^{**}P < 0.01$ and $^{***}P < 0.001$). **g**, Combined dot-violin-box plots presenting cytokine production by stimulations: IL-6 by LPS ($n = 65, 9, 64$ and 91 individuals), IL-1 β by *C. albicans* yeast ($n = 66, 9, 66$ and 94 individuals), TNF by *Mycobacterium tuberculosis* ($n = 66, 3, 64$ and 92 individuals) and IFN- γ by *Salmonella typhi* ($n = 67, 9, 65$ and 95 individuals) according to metabolome clusters (pink, green, turquoise and lilac, respectively). Differences between group median values were calculated by a two-sided Mann-Whitney U -test ($^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$). **h**, GOEA of downregulated DE genes ($P_{\text{adj}} < 0.05$ and absolute fold change > 1.2) between pink and turquoise clusters or pink and lilac clusters. **i**, Combined dot-violin-box plots for corrected and variance stabilizing transformation (VST) values of DE genes ($P_{\text{adj}} < 0.05$ and absolute fold change compared to pink cluster > 1.2) *RSAD2* and *OAS1* gene expression in metabolic clusters pink, green, turquoise and lilac ($n = 66, 10, 68$ and 103 individuals, respectively). Dots are colored according to seasonality; pie charts present the percentage of individuals from the urban-dry/wet seasonality groups in each metabolic cluster. In all box plots, the in-box line defines the median value, hinges depict 25th and 75th percentiles, whiskers extend to ± 1.5 IQR, and notches depict 95% CI around the median.

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/s41590-021-00867-8>.

Received: 12 March 2020; Accepted: 29 December 2020;
Published online: 11 February 2021

References

- De Brauw, A., Mueller, V. & Lee, H. L. The role of rural–urban migration in the structural transformation of sub-Saharan Africa. *World Dev.* **63**, 33–42 (2014).
- Beaglehole, R. et al. Priority actions for the non-communicable disease crisis. *Lancet* **377**, 1438–1447 (2011).
- Unwin, N. et al. Rural to urban migration and changes in cardiovascular risk factors in Tanzania: a prospective cohort study. *BMC Public Health* **10**, 272 (2010).
- Popkin, B. M. The nutrition transition: an overview of world patterns of change. *Nutr. Rev.* **62**, S140–S143 (2004).
- Abrahams, Z., McHiza, Z. & Steyn, N. P. Diet and mortality rates in sub-Saharan Africa: stages in the nutrition transition. *BMC Public Health* **11**, 801 (2011).
- Thorburn, A. N., Macia, L. & Mackay, C. R. Diet, metabolites, and ‘western-lifestyle’ inflammatory diseases. *Immunity* **40**, 833–842 (2014).
- Christ, A. et al. Western diet triggers NLRP3-dependent innate immune reprogramming. *Cell* **172**, 162–175 (2018).
- Bickler, S. W. et al. Urbanization in sub-Saharan Africa: declining rates of chronic and recurrent infection and their possible role in the origins of non-communicable diseases. *World J. Surg.* **42**, 1617–1628 (2018).
- Kodaman, N. et al. Cardiovascular disease risk factors in Ghana during the rural-to-urban transition: a cross-sectional study. *PLoS ONE* **11**, e0162753 (2016).
- Kann, P. H. et al. Alterations of cortisol homeostasis may link changes of the sociocultural environment to an increased diabetes and metabolic risk in developing countries: a prospective diagnostic study performed in cooperation with the Ovahimba people of the Kunene region/northwestern. *J. Clin. Endocr. Metab.* **100**, E482–E486 (2015).
- Mbow, M. et al. Changes in immunological profile as a function of urbanization and lifestyle. *Immunology* **143**, 569–577 (2014).
- Gruebner, O. et al. Cities and mental health. *Dtsch Arztebl. Int.* **114**, 121–127 (2017).
- Misra, A. & Ganda, O. P. Migration and its impact on adiposity and type 2 diabetes. *Nutrition* **23**, 696–708 (2007).
- Ter Horst, R. et al. Host and environmental factors influencing individual human cytokine responses. *Cell* **167**, 1111–1124 (2016).
- Piasecka, B. et al. Distinctive roles of age, sex, and genetics in shaping transcriptional variation of human immune responses to microbial challenges. *Proc. Natl Acad. Sci. USA* **115**, E488–E497 (2018).
- Baylis, D., Bartlett, D. B., Patel, H. P. & Roberts, H. C. Understanding how we age: insights into inflammaging. *Longev. Healthspan* **2**, 8 (2013).
- Fuhrer, T., Heer, D., Begemann, B. & Zamboni, N. High-throughput, accurate mass metabolome profiling of cellular extracts by flow injection-time-of-flight mass spectrometry. *Anal. Chem.* **83**, 7074–7080 (2011).
- Hostetler, G. L., Ralston, R. A. & Schwartz, S. J. Flavones: food sources, bioavailability, metabolism, and bioactivity. *Adv. Nutr.* **8**, 423–435 (2017).
- Ratter, J. M., Tack, C. J., Netea, M. G. & Stienstra, R. Environmental signals influencing myeloid cell metabolism and function in diabetes. *Trends Endocrinol. Metab.* **29**, 468–480 (2018).
- Barbie, D. A. et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* **462**, 108–112 (2009).
- Lee, J. H. et al. Anti-inflammatory mechanisms of apigenin: inhibition of cyclooxygenase-2 expression, adhesion of monocytes to human umbilical vein endothelial cells, and expression of cellular adhesion molecules. *Arch. Pharm. Res.* **30**, 1318–1327 (2007).
- Ren, K., Jiang, T., Zhou, H. F., Liang, Y. & Zhao, G. J. Apigenin retards atherogenesis by promoting ABCA1-mediated cholesterol efflux and suppressing inflammation. *Cell. Physiol. Biochem.* **47**, 2170–2184 (2018).
- Wang, J. et al. Anti-inflammatory effects of apigenin in lipopolysaccharide-induced inflammatory in acute lung injury by suppressing COX-2 and NF- κ B pathway. *Inflammation* **37**, 2085–2090 (2014).
- Marton, A. et al. Anti-inflammatory effects of inosine in human monocytes, neutrophils and epithelial cells in vitro. *Int. J. Mol. Med.* **8**, 617–621 (2001).
- Wang, Q. et al. Pyruvate protects against experimental stroke via an anti-inflammatory mechanism. *Neurobiol. Dis.* **36**, 223–231 (2009).
- Constantin, G., Laudanna, C., Baron, P. & Berton, G. Sulfatides trigger cytokine gene expression and secretion in human monocytes. *FEBS Lett.* **350**, 66–70 (1994).
- Horii, Y. et al. Leukotriene B₄ receptor 1 exacerbates inflammation following myocardial infarction. *FASEB J.* **34**, 8749–8763 (2020).
- Basu, S. Bioactive eicosanoids: role of prostaglandin F_{2 α} and F₂-isoprostanes in inflammation and oxidative stress related pathology. *Mol. Cells* **30**, 383–391 (2010).
- Branco, A., Yoshikawa, F. S. Y., Pietrobon, A. J. & Sato, M. N. Role of histamine in modulating the immune response and inflammation. *Mediators Inflamm.* **2018**, 9524075 (2018).
- Netea, M. G., Quintin, J. & van der Meer, J. W. Trained immunity: a memory for innate host defense. *Cell Host Microbe* **9**, 355–361 (2011).
- Li, Y. et al. Inter-individual variability and genetic influences on cytokine responses to bacteria and fungi. *Nat. Med.* **22**, 952–960 (2016); erratum **22**, 1192 (2016).
- Bekkering, S. et al. Oxidized low-density lipoprotein induces long-term proinflammatory cytokine production and foam cell formation via epigenetic reprogramming of monocytes. *Arterioscler. Thromb. Vasc. Biol.* **34**, 1731–1738 (2014).
- Sadler, A. J. & Williams, B. R. Interferon-inducible antiviral effectors. *Nat. Rev. Immunol.* **8**, 559–568 (2008).
- Minton, K. Viperin breaks viral chains. *Nat. Rev. Immunol.* **18**, 480–481 (2018).
- Napier, B. A. et al. Western diet regulates immune status and the response to LPS-driven sepsis independent of diet-associated microbiome. *Proc. Natl Acad. Sci. USA* **116**, 3688–3694 (2019).
- Strandberg, L. et al. Mice chronically fed high-fat diet have increased mortality and disturbed immune response in sepsis. *PLoS ONE* <https://doi.org/10.1371/journal.pone.0007605> (2009).
- Christ, A., Lauterbach, M. & Latz, E. Western diet and the immune system: an inflammatory connection. *Immunity* **51**, 794–811 (2019).
- Leentjens, J. et al. Trained innate immunity as a novel mechanism linking infection and the development of atherosclerosis. *Circ. Res.* **122**, 664–669 (2018).
- García-Lafuente, A., Guillaumon, E., Villares, A., Rostagno, M. A. & Martínez, J. A. Flavonoids as anti-inflammatory agents: implications in cancer and cardiovascular disease. *Inflamm. Res.* **58**, 537–552 (2009).
- Zhang, X., Wang, G., Gurley, E. C. & Zhou, H. Flavonoid apigenin inhibits lipopolysaccharide-induced inflammatory response through multiple mechanisms in macrophages. *PLoS ONE* **9**, e107072 (2014).
- Hemler, E. C. & Hu, F. B. Plant-based diets for cardiovascular disease prevention: all plant foods are not created equal. *Curr. Atheroscler. Rep.* **21**, 18 (2019).
- Minutoli, L. et al. The disaccharide trehalose inhibits proinflammatory phenotype activation in macrophages and prevents mortality in experimental septic shock. *Shock* **27**, 91–96 (2007).
- Collins, J. et al. Dietary trehalose enhances virulence of epidemic *Clostridium difficile*. *Nature* **553**, 291–294 (2018).
- Choy, K. W., Murugan, D. D., Leong, X.-F., Abas, R. & Alias, A. Flavonoids as natural anti-inflammatory agents targeting nuclear factor-kappa B (NF κ B) signalling in cardiovascular diseases: a mini review. *Front. Pharmacol.* **10**, 1295 (2019).
- Dopico, X. C. et al. Widespread seasonal gene expression reveals annual differences in human immunity and physiology. *Nat. Commun.* **6**, 7000 (2015).
- Oosting, M. et al. Functional and genomic architecture of *Borrelia burgdorferi*-induced cytokine responses in humans. *Cell Host Microbe* **20**, 822–833 (2016).
- Koeken, V. A. et al. BCG vaccination in humans inhibits systemic inflammation in a sex-dependent manner. *J. Clin. Invest.* <https://doi.org/10.1172/JCI133935> (2020).
- Klein, S. L. & Flanagan, K. L. Sex differences in immune responses. *Nat. Rev. Immunol.* **16**, 626–638 (2016).
- Onyango, E. M. & Onyango, B. M. The rise of noncommunicable diseases in Kenya: an examination of the time trends and contribution of the changes in diet and physical inactivity. *J. Epidemiol. Glob. Health* **8**, 1–7 (2018).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2021

Methods

Study area and population description. This study, also referred to as the 300 Functional Genomics (300FG) study, is part of the Human Functional Genomics Project (www.humanfunctionalgenomics.org). A total of 323 Tanzanian healthy individuals aged between 18–65 years residing within the Kilimanjaro region in Northern Tanzania were enrolled between March and December 2017, at the Kilimanjaro Christian Medical Center and Lucy Lameck Research Center, in Moshi municipality. Moshi city is the administrative, commercial and educational center of the Kilimanjaro region with over 200,000 inhabitants and six higher education institutions. Due to this regional function, there is substantial diversity in ethnic groups, economic status and lifestyle, whereby most inhabitants have adopted a western lifestyle. People residing in the Moshi municipality (urban zone) have access to good sanitation with universal coverage of chlorinated tap water and toilet facilities. In contrast, people we designated as living in rural conditions represent the rural population of Tanzania; people live within large family units, whereby the economy depends mainly on subsistence farming and animal husbandry. Most individuals in the rural area belong to the Chagga tribe and follow a traditional lifestyle, whereby using a pit latrine and consumption of a starch and vegetable diet are common. They also have access to good sanitation and water primarily from the slope of Mount Kilimanjaro or from wells. We enrolled individuals from both urban and rural zones with a higher number from urban zones to correct for the higher genetic diversity.

Information on the study was given through leaflets or announced during a mass gathering. All volunteers were interviewed by a member of the study team using a guided pre-screening questionnaire before being invited to the study center. Exclusion criteria were pregnancy, a known acute or chronic disease, use of antibiotics or antimalarials in the previous 3 months or receiving treatment for tuberculosis infection in the past year. Following written informed consent, participants were screened for HIV infection (SD BIOLINE HIV-1/2 3.0 kit, Standard Diagnostics, 03FK16) and malaria (Malaria Pf/PAN (HRP2/pLDH) Ag Combo RDT kit, Accessbio, RMRM-02571) and random blood glucose (ACCU-CHECK glucose test strips, Roche Diabetes Care, 07124112036), blood pressure, body weight and height were measured. Individuals with a positive HIV or malaria rapid test, blood pressure $\leq 90/60$ mm Hg or $\geq 140/90$ mm Hg or with random blood glucose > 8.0 mmol $^{-1}$ were excluded. A questionnaire was used to collect demographic data, as well as data on lifestyle, diet, animal exposure and disease history. Data were collected using cloud-based electronic data capture (Castor EDC 2017.2 EU server).

Laboratory procedures. Whole blood RNA isolation and RNA-seq. Total RNA was isolated from whole blood samples stored in PAXgene tubes using a QIAGEN PAXgene RNA isolation kit according to the manufacturer's guidelines and eluted RNA was dissolved in RNase-free water. The quality of RNA was evaluated by visualization of 28S and 18S band integrity on a TapeStation 4200 (Agilent). Total RNA was converted into double-stranded complementary DNA molecule libraries using TruSeq Stranded Total RNA with Ribo-Zero Globin (Illumina). In brief, ribosomal and globin mRNA were depleted from 750 ng purified total RNA using biotinylated, target-specific oligonucleotides combined with Ribo-Zero rRNA removal beads, remaining RNA was fragmented using divalent cations under elevated temperature. First strand was generated using SuperScript 2 RT (Invitrogen) supplemented with actinomycin D, followed by second-strand synthesis with dUTP replacing dTTP. The 3' ends were adenylated and index adaptors were ligated before subsequent PCR amplification to yield the final library. Remaining overhangs were converted into blunt ends via exonuclease/polymerase activities and enzymes were removed. Selective enrichment of DNA fragments with ligated adaptor molecules was performed using Illumina PCR primers in 15 cycles of PCR reaction, followed by purification cDNA using SPRIbeads (Beckman-Coulter). High-throughput sequencing was carried out with the NovaSeq 6000 Sequencing System S2 (50-bp pair-end reads) and data were converted into fastq files using bcl2fastq v.2.20.

Gene expression profiling and bioinformatic analysis. Raw RNA-seq data were aligned using the pseudo-aligner Kallisto (v.0.44.0) and human reference transcriptome version HG38. Data were imported into DESeq2 (v.1.26.0) using the TXimport (v.1.2.0) package in R (v.3.6.1). Raw gene counts were normalized using the DESeq2 package (v.1.26.0). All low-expressed gene counts (sum of counts across samples $< 3,000$) were excluded, resulting in 21,224 present genes. Normalized data were further transformed using VST^{50,51}. Unwanted or hidden sources of variation, such as the technical variance introduced by batch and preparation date or donors, were modeled during the normalization or were removed using surrogate variable analysis⁵². Surrogate variables identified by surrogate variable analysis (15 variables) were included in the DESeq2 model design, and the VST-transformed expression values were adjusted for these variables using the function 'remove batch effect' from the limma package⁵³. PCA was performed on VST-transformed corrected counts of all present genes and the groups were further presented in PCA space by ellipses, calculated by the extension function `stat_ellipse` provided by `ggplot2` (v.3.3.2) setting the level at 0.95 and by density histograms. To evaluate overlap of the groups in PCA space, the mean difference of the group's distributions in PC1 and PC2 was compared

using a Student's *t*-test or one-way ANOVA. GSEA for hallmark MSigDB gene sets (h.all.v7.0) was performed using clusterProfiler package GSEA function⁵⁴, using all present genes ranked by fold change. Gene sets with a NES above or below 1.5 and $P_{adj} < 0.1$ (Benjamini–Hochberg correction)⁵⁵ were plotted in vertical dot plots or circular volcano plots. DE analysis was performed between the residential and seasonal groups using the FDR-controlling method for P_{adj} (independent hypothesis weighting). Heat maps of selected genes present each seasonal group average of the scaled expression of each gene across all individuals. Genes presented in the heat map are those defined as (1) genes with DE analysis $P_{adj} \leq 0.2$ and (2) genes that were defined as core-enrichment genes by the GSEA output (genes that contributed most to the enrichment of the selected gene set in the GSEA analysis)⁵⁶. To investigate the effect of metabolites on the transcriptome, we generated metabolites specific gene sets for upregulated (upregulated gene set) or downregulated (downregulated gene set) DE genes with an absolute fold change of 1.2. Apigenin gene sets relied on reported DE genes⁵⁷ and trehalose and itaconate gene sets were defined by calculating DE genes in respective publicly available datasets (GSE76643 and GSE110749) for control versus metabolite-treated cells using limma linear modeling as described in <https://github.com/schultzelab/Temba-Kullaya-Pecht-et-al.->. For each metabolite-specific gene set and each sample, an enrichment score was calculated using ssGSEA from the GSVA package^{58,59} and method ssGSEA. The dependency between residence and/or seasonality and the respective individual enrichment scores was evaluated by chi-squared test, building the contingency table by the residence or seasonal groups and the abundance of individuals with above or below all-cohort median enrichment score value. Power calculation to estimate effect size was performed using Cramer's *V* equation $V = \chi^2 / (n \times d.f.)$, where χ^2 is chi-squared of the test, *n* is number of observations and d.f. is degrees of freedom.

DE analysis among the urban metabolic clusters was performed as described above and DE genes were defined by $P_{adj} < 0.05$ and absolute fold change > 1.2 . GOEA of up- or downregulated genes was performed using enrichGO from the clusterProfiler package (v.3.14.3) with $P_{adj} < 0.05$ (Benjamini–Hochberg correction) defining all expressed genes as the universe.

Measurement of circulating inflammatory mediators. The circulating plasma cytokines IL-6, IL-1 β , IL-1Ra, IL-18 (Bio-Tech/R&D; SPCKC-PS-001559) and IL-18 binding protein (IL-18BP) (Bio-Tech/R&D; SPCKB-PS-000502) were measured in EDTA plasma using Simple Plex cartridges run on the Ella platform (Protein Simple) following the manufacturer's instructions.

Whole blood stimulation assays and cytokine quantification. Ex vivo cytokine stimulation experiments were performed at the biotechnology laboratory facility available at Kilimanjaro Clinical Research Institute in Moshi, Tanzania. Whole blood was stimulated with bacterial, fungal and TLR-3 and TLR-4 agonists as follows; 100 μ l of heparin blood was added to a 48-well culture plate and subsequently stimulated with 400 μ l of stimulus for 48 h at 37°C and 5% CO₂. Stimuli were prepared in RPMI culture medium (Dutch modified, Invitrogen) supplemented with 50 μ g ml $^{-1}$ gentamicin, 2 mM Glutamax and 1 mM pyruvate. Supernatants were collected and stored at -80° C until used for ELISA. Concentrations of cytokines IL-6, IL-1 β , IFN- γ , TNF and IL-10 were measured in the stored supernatants using ELISA according to the instructions (given IL-6, IL-1 β , IL-10 and TNF: R&D systems; IFN- γ : Sanquin). To minimize variation between measurements, all samples were measured using kits of the same lot number. Stimuli used and their respective concentrations are listed in Supplementary Table 16. The distribution of all the measured cytokines is summarized by the histograms shown in Supplementary Fig. 7.

Plasma metabolome. Plasma metabolome measurement was performed as previously described¹⁷. Briefly, using an untargeted metabolomics workflow, metabolites were first measured from plasma samples by high-throughput mass spectrometry technique using a platform consisting of an Agilent Series 1100 LC pump coupled to a Gerstel MPS2 autosampler and an Agilent 6520 Series Quadrupole Time-of-flight mass spectrometer (Agilent) equipped with an electrospray source operated in a negative and positive mode. The flow rate was 150 μ l min $^{-1}$ and able to detect 50–1,000 ion m/z (mass-to-charge ratio) at 1.4 Hz in a 1-min cycle time. Data were further analyzed by performing centroiding on a high-performance computing cluster using the bioinformatics function in Matlab R2018a (The Mathworks, Natick) with the final result of matrix construction from the centroid lists. A list of putative metabolites was annotated with a series of analysis strategy, including deisotoping, decluttering, adduct detection and library matching in KEGG (<https://www.genome.jp/kegg/>), human metabolic database (HMDB) (<https://www.hmdb.ca/>) and ChEBI (<https://www.ebi.ac.uk/chebi/>). The resulting peak intensity data table was then uploaded in MetaboAnalyst v.4.0 and comprehensive metabolomics data analysis was performed as per existing protocols.⁵⁹ In brief, data were uploaded as peak intensity table (MS peak list data) in comma-separated values (.csv) format. A data matrix of 323 individual samples and 1,601 compounds/metabolites (peaks (mz/rt)) were uploaded in a column and a row respectively. Data were normalized by generalized log transformation (`glog2`) and Pareto scaling. PCA was applied to check for outliers and three potential outliers were excluded from further analysis. Differences in compounds/

metabolites between rural and urban dwellers were analyzed using Student's *t*-tests. Metabolic pathway analysis, which combines functional enrichment analysis and network topology analysis was performed on the list of compound labels (common names) to help identify metabolic pathways that were most likely associated with the different inflammatory profiles among rural and urban-living individuals. Over-representation analysis was performed using a hypergeometric test, whereas pathway topology analysis was performed using relative-betweenness centrality. Associated pathways were visualized and explored via a Google-map-style interactive visualization framework. PCA was performed and groups were further presented in PCA space by ellipses, calculated by the extension function `stat_ellipse` provided by `ggplot2` (v.3.3.2) setting the level at 0.95 and by adjacent density histograms.

To further investigate the contribution of dietary habits to the immunofunctional changes observed among urban individuals recruited during the dry versus wet seasons, we performed a selection of only food-derived metabolites based on the ontology given in the HMDB (<https://hmdb.ca/>). A total of 293 food-derived metabolites were selected and further explored for their contribution to the immune functional changes. Data were normalized by generalized log transformation ($\log_2$) and Pareto scaling using `MetaboAnalyst` and further analyses were performed in R. PCA of urban cohort food-derived metabolome was performed and the quality of representation for metabolites on the factor map was presented by $\cos^2$ value, using the `get_pca_var` function from the `factoextra` package (v.1.0.7). Definition and visualization of metabolic clusters in the urban cohort was conducted by agglomerative hierarchical clustering on the food-derived metabolome by setting the Euclidean distance calculation and using the ward.D2 linkage algorithm in the `heatmap` function (`heatmap` v.1.0.12).

In vitro validation experiments. PBMCs were isolated by density centrifugation of buffy coats collected from healthy Dutch donors after obtaining written informed consent (Sanquin Blood Bank). In short, 5×10^5 cells in a total volume of 200 μ l were co-cultured for 2 h at 37°C and 5% CO₂ in the presence of one of the following stimuli: apigenin (4',5,7-trihydroxyflavone) (Sigma-Aldrich; SMB00702-5MG) in a concentration of (3.125, 6.25, 12.5 and 25 μ M), trehalose dihydrate (Sigma-Aldrich; T0167-10G) in a concentration of (0.01, 0.1, 1 and 10 mM) or with 250 μ M DMI (Sigma-Aldrich; 592498-25 G). After the 2 h incubation, cells were stimulated for 24 h with LPS (10 ng ml⁻¹ *E. coli* serotype 055: B5; Sigma-Aldrich). Concentrations of TNF and IL-6 were measured in the culture supernatant using ELISA according to the manufacturer's instructions.

Induction of trained immunity in monocytes. Buffy coats from healthy donors were obtained after written informed consent (Sanquin Blood Bank). Percoll isolation of monocytes was performed as previously described⁶⁰. Briefly, 150–200 $\times 10^6$ PBMCs were layered on top of a hyperosmotic Percoll solution (48.5% Percoll; Sigma-Aldrich), 41.5% sterile H₂O and 0.16 M filter-sterilized NaCl) and centrifuged for 15 min at 580g. The interphase layer was isolated and cells were washed with cold PBS. Cells were re-suspended in RPMI culture medium (RPMI medium, Dutch modified, Invitrogen) supplemented with 50 mg ml⁻¹ gentamicin, 2 mM Glutamax and 1 mM pyruvate and counted. An extra purification step was added by adhering Percoll-isolated monocytes to polystyrene flat-bottom plates (Corning) for 1 h at 37°C; a washing step with warm PBS was then performed to yield maximal purity.

Monocytes were trained as described before⁶¹. Briefly, 1×10^5 monocytes in flat-bottom 96-well plates were trained for 24 h either with only culture medium as a negative control, 1 μ g ml⁻¹ BCG positive control or 10% pooled serum from either urban or rural participants from the highest or lowest quartiles of TNF production capacity. After 24 h of incubation, cells were washed once with 200 μ l of warm PBS and incubated further for 5 d in a culture medium with a normal 10% pooled human serum and the medium was changed once. Cells were re-stimulated with either 200 μ l RPMI or 10 ng ml⁻¹ LPS for 24 h and supernatants were collected and stored at –20°C. In some experiments, cells were pre-incubated for 1 h with 10 μ M apigenin (4',5,7-trihydroxyflavone) (Sigma-Aldrich; SMB00702-5MG) followed by training with 1 μ g ml⁻¹ BCG positive control or 10% pooled serum from either urban or rural participants or oxLDL. After 24 h, stimuli and inhibitors were washed away, cells were rested for 5 d and re-stimulated with LPS for 24 h as described above. Supernatants were collected and stored at –20°C until used for ELISA.

Genotyping, quality control and imputation of genetic data. DNA was extracted from whole blood using DNeasy Blood and Tissue kit (cat. nos. 69504 and 69506, QIAGEN). Genotyping was performed using the Global Screening Arrays SNP chip. Default settings of optical 0.70 were used to perform genotype calling.⁶²

Before imputation of genetic data, filtering was performed by excluding variants with call rates >0.1, low minor allele frequencies <0.001 and single-nucleotide polymorphisms deviating from Hardy–Weinberg equilibrium with $P < 1 \times 10^{-4}$.

Finally, we removed individuals with genetic outliers that were identified by using multidimensional scaling plots. A total of 409,261 variants and 308 samples passed quality control procedures. Strand alignment to a reference panel, 1000 Genomes Project reference panel dataset was performed using Genotype harmonizer⁶³.

To improve genome coverage, genotype imputation was performed for all chromosomes except for sex chromosomes using Minimac4 software through the publicly available Michigan Imputation Server⁶⁴. The Human Reference Consortium (HCR r.1.1 2016) was used as a reference panel and the dataset was phased using Eagle v.2.3. We filtered out variants with imputation quality score (R^2) < 0.3. Individuals with extreme heterozygosity and cryptic relatedness were excluded. Genotyping and imputation generated a total of 5,271,779 variants from 285 individuals.

Clustering analysis using genetic and cytokine data. We performed unsupervised hierarchical clustering analysis on both cytokine and genetic data. With the genetic data, multidimensional scaling analysis was performed with PLINK. Briefly, PLINK computes a matrix of genome-wide identity by state pairwise distances using whole-genome single-nucleotide polymorphism data. This matrix serves as input for multidimensional scaling analysis, to compute genetic distances of urban and rural samples. To explore clustering patterns between urban and rural residents, the first six dimensions were extracted and used as variables to construct a heat map.

Also, to cluster ex vivo cytokine data consisting of urban and rural samples, the raw cytokine data was first $\log_2$ transformed and the resulting numeric matrix was used as input to the `heatmap` function. A Euclidean distance metric was employed to evaluate the distance between pairs of samples.

All statistical analyses were performed with R, a free software environment for statistical computing and graphics.

Weekly food dietary intake. A questionnaire was used to collect the frequency of weekly food intake from all study participants on a scale of 0x–7x (meals/week). Analysis of differences in the weekly food intake among the urban food-derived clusters was conducted in R applying a chi-squared test using the function `chisq.test` (stats package, v.3.6.2) following post hoc tests using the `chisq.posthoc.test` package (v.0.1.3), adjusting *P* value for multiple comparisons using the Benjamini–Hochberg method.

Ethical statement. This study was approved by the Ethical Committees of the Kilimanjaro Christian Medical University College (CREC) (no. 2443) and the National Institute for Medical Research in Tanzania (NIMR/HQ/R.8a/Vol. IX/2290 and NIMR/HQ/R.8a/Vol.IX/3318). In the Netherlands, the study was approved by the Ethical Committee of the Radboud University Medical Centre Nijmegen (CMO Arnhem-Nijmegen; 2016-2657). Participant recruitment and experimental procedures were conducted according to the principles mentioned in the Declaration of Helsinki. For the validation experiments, the use of the buffy coat samples from healthy donors received institutional review board approval. All participants gave written informed consent before the recruitment procedure.

Statistical analysis. Data of the ex vivo cytokine and circulating inflammatory mediators were $\log_2$ transformed. For the correlation analysis, correlation coefficients were obtained using Pearson partial correlation, which enables controlling for confounding variables. The linear regression model was used to explore factors associated with cytokine responses. With this method, the mean differences in cytokine production capacity between rural and urban cohorts after correcting for age, sex and BMI were obtained. A similar model was used to determine the effect of age, sex and seasonality on cytokine immune responses. Differences between groups were analyzed using chi-squared or Mann–Whitney *U*-tests for categorical and continuous variables, respectively. For each comparison, missing values were handled on a pairwise basis. Statistical significance was based on a *P* value <0.05. Analyses and visualization were performed using R packages, SPSS v.25 and GraphPad prism v.8 for the cytokines. Plots for individual genes or cytokines were plotted in R using `ggplot2` package (v.3.3.2.). Data were represented by three layers of information: (1) dot plot to show single-sample behavior; (2) violin-plot to define group distribution behavior; and (3) box-notch plot in which the notch displays a CI around the median, box boundary defines the IQR and whiskers show 1.5 $\times$ IQR. Metabolome data were normalized, transformed and further analyzed using `MetaboAnalyst` v.4.0 using an existing protocol⁵⁹ (accessed in 2019–2020 from <https://www.metaboanalyst.ca/>) and additional analyses and visualization were conducted in R, as described above.

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

Data availability

Data that support the findings of this study are available from the corresponding author upon request. Sequence data have been deposited at the European Genome–phenome Archive, which is hosted by the EBI and the CRG, under accession number EGAS00001004284. In addition to the deposition of the raw sequencing data on the European Genome–phenome Archive, we provide an interactive platform for data inspection and analysis via FASTGenomics (https://beta.fastgenomics.org/p/Temba_300FG_NatureImmun). In this platform, we provide processed count tables of the datasets generated in this study as well as key analytical results and the code written to analyze the respective data. Metabolomics

data have been deposited to the EMBL-EBI MetaboLights database⁶⁵ with study identifier MTBLS2267. Code analysis scripts are available at: <https://github.com/schultzelab/Temba-Kullaya-Pecht-et-al.-> and https://beta.fastgenomics.org/p/Temba_300FG_NatureImmun. Publicly available databases used for this study include NCBI's Gene Expression Omnibus G3 under accession code GSE110749 and differential gene analysis results from the study of Arango et al. (*Biochem. Pharmacol.*, <https://doi.org/10.1016/j.bcp.2012.09.005>, 2012). Other databases are KEGG (<https://www.genome.jp/kegg/>), HMDB (<https://www.hmdb.ca/>) and ChEBI (<https://www.ebi.ac.uk/chebi/>). All other data are available in the main text, supplementary materials and auxiliary supplemental tables.

References

50. Tibshirani, R. Estimating transformations for regression via additivity and variance stabilization. *J. Am. Stat. Assoc.* **83**, 394–405 (1988).
51. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Nat. Preced.* <https://doi.org/10.1038/npre.2010.4282.2> (2010).
52. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883 (2012).
53. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
54. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics* **16**, 284–287 (2012).
55. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).
56. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
57. Arango, D. et al. Apigenin induces DNA damage through the PKC δ -dependent activation of ATM and H2AX causing down-regulation of genes involved in cell cycle control and DNA repair. *Biochem. Pharmacol.* **84**, 1571–1580 (2012).
58. Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* **14**, 7 (2013).
59. Chong, J., Wishart, D. S. & Xia, J. Using Metaboanalyst 4.0 for comprehensive and integrative metabolomics data analysis. *Curr. Protoc. Bioinformatics* **68**, e86 (2019).
60. Repnik, U., Knezevic, M. & Jeras, M. Simple and cost-effective isolation of monocytes from buffy coats. *J. Immunol. Methods* **278**, 283–292 (2003).
61. Bekkering, S. et al. In vitro experimental model of trained innate immunity in human primary monocytes. *Clin. Vaccine Immunol.* **23**, 926–933 (2016).
62. Shah, T. S. et al. optiCall: a robust genotype-calling algorithm for rare, low-frequency and common variants. *Bioinformatics* **28**, 1598–1603 (2012).
63. Deelen, P. et al. Genotype harmonizer: automatic strand alignment and format conversion for genotype data integration. *BMC Res. Notes* **7**, 901 (2014).
64. Das, S. et al. Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).
65. Haug, K. et al. MetaboLights: a resource evolving in response to the needs of its scientific community. *Nucleic Acids Res.* **48**, D440–D444 (2020).

Acknowledgements

The authors thank all volunteers in the Human Functional Genomics in Healthy Tanzanian Individuals study for their participation. We thank J. Njau, J. Kwayu and E. Kimaro for help in sample collection; H. Lemmers and H. Toenhake-Dijkstra for help in laboratory analysis; Y. Li for help in statistics; and M. Mićlaus for help with metabolome data processing. We also thank M. Kraut and K. Händler for their great contribution to RNA sequencing. This study was funded by the following grants: the European Union's Horizon 2020 Research and Innovation Program under the ERA-Net Cofund action no. 727565; the Joint Programming Initiative, A Healthy Diet for a Healthy Life (JPI-HDHL; project 529051018) awarded to M.G.N., Q.d.M., A.V. and J.L.S.; ZonMw (the Netherlands Organisation for Health Research and Development) awarded to M.G.N., Q.d.M. and A.V.; Radboud Revolving Research Funds (3R-Fund) awarded to G.S.T.; Spinoza Prize (NWO SPI94-212) and ERC Advanced grant (no. 833247) awarded to M.G.N.; and the Deutsche Forschungsgemeinschaft (German Research Foundation) under Germany's Excellence Strategy (EXC2151) 390873048 awarded to M.G.N. and J.L.S.

Author contributions

Q.d.M., A.V., M.G.N., L.A.B.J., G.K., J.L.S. and B.T.M. contributed to the conceptualization, study design and data interpretation and led the project; G.S.T., V.K., B.T.M. and F.L. contributed to participant recruitment, data collection and laboratory analyses; G.S.T. designed and performed functional validation experiments and analysis of immunological and metabolome data; T.P., T.U., A.C.A. and J.L.S. contributed to RNA-seq analysis and analytical integration with metabolome data and interpretation; C.K.B. and V.K. contributed to genetics analysis and interpretation; G.S.T., T.P. and Q.d.M. wrote the original draft of the manuscript; and G.S.T., T.P., V.K., C.K.B., B.T.M., A.C.A., T.U., G.K., F.L., V.K., L.A.B.J., J.L.S., A.V., M.G.N. and Q.d.M. contributed to writing and editing the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41590-021-00867-8>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41590-021-00867-8>.

Correspondence and requests for materials should be addressed to M.G.N. or Q.d.M.

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

Peer reviewer 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.

Extended Data Fig. 1 | Schematic depiction of the study parameters. Schematic depiction of the study recruitment procedure. **b**, collected samples for blood transcriptome measured in unstimulated blood and circulating inflammatory mediators and metabolome were measured in EDTA plasma. **c**, Cytokine production capacity of the circulating immune cells in the ex vivo whole blood stimulation experimental setup.

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 Data were collected using cloud-based electronic data capture (Castor EDC 2017.2 EU server)

Data analysis Processing, analyses and visualization of the cytokine data were performed using R-packages, SPSS version 25 and Graphpad prism version 8.2. Metabolome data was processed using Matlab R2018a. Data normalization, transformation and further analyses were performed using MetaboAnalyst version 4.0 (accessed in 2019-2020 at <https://www.metaboanalyst.ca/>), and additional analyses and visualization were conducted in R using the following packages ggplot2 (v 3.3.2), factoextra (v 1.0.7), pheatmap (v 1.0.12). Sequencing data were converted into fastq files using bcl2fastq2 (v 2.20). Raw RNA-seq data were aligned using the pseudo-aligner Kallisto (v 0.44.0). RNAseq data analysis was done using R (v 3.6.1) and the following packages: TXimport (v 1.14.0), DESeq2 package (v 1.25.13), clusterProfiler (v3.14.3), limma (3.41.17), GSVA (1.34.0).

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 data that support the findings of this study are available from the corresponding author upon request. Sequence data have been deposited at the European Genome-phenome Archive (EGA), which is hosted by the EBI and the CRG, under accession number EGAS00001004284. In addition to the deposition of the raw sequencing data on EGA, we provide an interactive platform for data inspection and analysis via FASTGenomics (https://beta.fastgenomics.org/p/Temba_300FG_NatureImmune). In this platform, we provide processed count tables of the datasets generated in this study as well as key analytical results and the

code written to analyze the respective data. Metabolomics data has been deposited to the EMBL-EBI MetaboLights database 70 with study identifier MTBLS2267. The complete data can be accessed here www.ebi.ac.uk/metabolights/MTBLS2267. Code analysis scripts are available at https://github.com/schultzelab/Temba-Kullaya-Pecht-et-al.-and-https://beta.fastgenomics.org/p/Temba_300FG_NatureImmun. Publicly available databases that were used for this study are from NCBI's Gene Expression Omnibus (GEO) : GSE76643 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE76643>), GSE110749 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE110749>) and differential genes analysis results from the study Arango, D. et al. 2012 (DOI: 10.1016/j.bcp.2012.09.005). Other publicly used databases includes KEGG (<https://www.genome.jp/kegg/>), HMDB (<https://www.hmdb.ca/>) and ChEBI (<https://ebi.ac.uk/chebi/>). All other data is available in the main text, supplementary materials and auxiliary supplemental tables.

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](https://www.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	In earlier studies by our collaboration group, the associations between genome and immunity have been reliably obtained in a group of approximately 100 volunteers (Smeekens et al, Nature Communications 2013). Because we intend to add an additional aspect to this analysis, and assess the interaction between genome, metabolome and immunity we decided to increase the number of individuals tested to 300.
Data exclusions	Owing to the samples size of the cohort (n=323) we could determine outliers that exhibit extreme variability, which is beyond the heterogeneity of the population. As a standard procedure, the quality of RNAseq samples was evaluated by examining the expression pattern and expression density using nowclean R package (Bøvelstad et al. 2017 BioRxiv). Samples were excluded (n=7) in case they were defined as outliers by at least 2 out of 3 calculation methods offered by the package, using a cut-off of 2.7 standard deviation.
Replication	All validation and monocytes training experiments were performed in three independent sets of experiments and in all sets were reproducible.
Randomization	No randomization was required since this study was designed as a cross sectional study in which healthy volunteers from rural and urban areas were sampled once without a follow up. All participants were coded anonymously and all downstream measurements and statistical analyses used only these codes. Covariates like age and sex were controlled for during analyses.
Blinding	Blinding was not relevant in this type of study design but all participants were coded anonymously and all downstream measurements and statistical analyses used only these codes.

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

Human research participants

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

Population characteristics	A total of 323 Tanzanian healthy individuals aged between 18-65 years residing within the Kilimanjaro region in Northern Tanzania were enrolled between March and December 2017, at the Kilimanjaro Christian Medical Center and Lucy Lameck Research Center, in Moshi municipal. Participants from both rural and urban areas were enrolled. Approx. 50% males and females were recruited. Exclusion criteria were pregnancy, a known acute or chronic disease, use of antibiotics or anti-malarials in the previous three months, or receiving treatment for tuberculosis infection in the past year.
Recruitment	Information on the study was given through leaflets or announced during the mass gathering. All volunteers were interviewed by a member of the study team using a guided pre-screening questionnaire prior to being invited to the study center

Ethics oversight

Ethical Committees of the Kilimanjaro Christian Medical University College (CREC) (No 2443) and the National Institute for Medical Research in Tanzania (NIMR/HQ/R.8a/Vol. IX/2290 and NIMR/HQ/R.8a/Vol.IX/3318). In the Netherlands, the study was approved by the Ethical Committee of the Radboud University Medical Centre Nijmegen (CMO Arnhem-Nijmegen; 2016-2657).

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