

---

# A functional genomics approach in Tanzanian population identifies distinct genetic regulators of cytokine production compared to European population

## Authors

Collins K. Boahen, Godfrey S. Temba,  
Vesla I. Kullaya, ..., Quirijn de Mast,  
Mihai G. Netea, Vinod Kumar

## Correspondence

[v.kumar@radboudumc.nl](mailto:v.kumar@radboudumc.nl)



# A functional genomics approach in Tanzanian population identifies distinct genetic regulators of cytokine production compared to European population

Collins K. Boahen,<sup>1,2</sup> Godfrey S. Temba,<sup>2,3</sup> Vesla I. Kullaya,<sup>3,4</sup> Vasiliki Matzaraki,<sup>1,2</sup> Leo A.B. Joosten,<sup>1,2</sup> Gibson Kibiki,<sup>4,5</sup> Blandina T. Mmbaga,<sup>4</sup> Andre van der Ven,<sup>2,3</sup> Quirijn de Mast,<sup>2</sup> Mihai G. Netea,<sup>1,2,6</sup> and Vinod Kumar<sup>1,2,7,8,\*</sup>

## Summary

Humans exhibit remarkable interindividual and interpopulation immune response variability upon microbial challenges. Cytokines play a vital role in regulating inflammation and immune responses, but dysregulation of cytokine responses has been implicated in different disease states. Host genetic factors were previously shown to significantly impact cytokine response heterogeneity mainly in European-based studies, but it is unclear whether these findings are transferable to non-European individuals. Here, we aimed to identify genetic variants modulating cytokine responses in healthy adults of East African ancestry from Tanzania. We leveraged both cytokine and genetic data and performed genome-wide cytokine quantitative trait loci (cQTLs) mapping. The results were compared with another cohort of healthy adults of Western European ancestry via direct overlap and functional enrichment analyses. We also performed meta-analyses to identify cQTLs with congruent effect direction in both populations. In the Tanzanians, cQTL mapping identified 80 independent suggestive loci and one genome-wide significant locus (*TBC1D22A*) at chromosome 22; SNP rs12169244 was associated with IL-1b release after *Salmonella enteritidis* stimulation. Remarkably, the identified cQTLs varied significantly when compared to the European cohort, and there was a very limited percentage of overlap (1.6% to 1.9%). We further observed ancestry-specific pathways regulating induced cytokine responses, and there was significant enrichment of the interferon pathway specifically in the Tanzanians. Furthermore, contrary to the Europeans, genetic variants in the *TLR10-TLR1-TLR6* locus showed no effect on cytokine response. Our data reveal both ancestry-specific effects of genetic variants and pathways on cytokine response heterogeneity, hence arguing for the importance of initiatives to include diverse populations into genomics research.

## Introduction

Cytokines are key regulators of immune response to invading pathogens.<sup>1</sup> Variations in cytokine responses determine susceptibility to infectious diseases as well as autoimmune and inflammatory diseases.<sup>2</sup> Cytokine responses elicited by different microbial stimuli are known to be highly heterogeneous across individuals and populations.<sup>3</sup> Thus, identifying the major factors that determine cytokine responses to microbial and other environmental triggers has the potential to refine our understanding of immune system variation and understand the source of the variability in disease susceptibility between individuals.

The Human Functional Genomics Project (HFGP) was established in 2013 with the aim of identifying genetic as well as non-genetic host factors that determine interindividual variability in immune responses. In cohorts of healthy individuals of Western European ancestry, these functional genomics studies have demonstrated the significant role of genetic as well as environmental and non-genetic (such as age, sex, and seasonality) factors for the

interindividual differences in cytokine production capacity upon stimulation.<sup>4–6</sup> Undoubtedly, these studies have broadened our knowledge on the major factors that contribute to cytokine response variability. For example, SNPs at the *TLR10-TLR1-TLR6* (MIM: 606270, 601194, 605403) locus showed the strongest effect on production of multiple cytokines.

Prior to this, Wurfel et al. (2008), through candidate gene approach study in individuals recruited from Seattle metropolitan area, had reported a genetic variant, rs5743551, within the TLR1 locus to be strongly associated with induced cytokine production and sepsis outcome.<sup>7</sup> Similarly, a previous work on individuals of Caucasian origin demonstrated that variations in the *TLR10-TLR1-TLR6* locus drives the most interindividual variation in TLR2-mediated cytokine responses.<sup>8</sup> In the same genomic region, previous work on transcriptional response of primary monocytes to viral and bacterial stimuli reported European-specific strong *trans*-eQTL (expression quantitative trait locus) effect.<sup>9</sup> Given that eQTLs effects are context or tissue dependent, this study also focused on self-reported Africans residing in Belgium

<sup>1</sup>Department of Internal Medicine and Radboud Institute of Molecular Life Sciences, Radboud University Medical Center, Nijmegen 6525, the Netherlands;

<sup>2</sup>Department of Internal Medicine and Radboud Center for Infectious Diseases, Radboud University Medical Center, Nijmegen 6525, the Netherlands;

<sup>3</sup>Department of Medical Biochemistry and Molecular Biology, Kilimanjaro Christian Medical University College, Moshi 251, Tanzania; <sup>4</sup>Kilimanjaro Clinical Research Institute, Kilimanjaro Christian Medical Center, Moshi 251, Tanzania; <sup>5</sup>Department of Paediatrics, Kilimanjaro Christian Medical University College, Moshi 251, Tanzania; <sup>6</sup>Department for Genomics and Immunoregulation, Life and Medical Sciences Institute, University of Bonn, Bonn 53115, Germany; <sup>7</sup>University of Groningen, University Medical Center Groningen, Department of Genetics, Groningen 9700, the Netherlands; <sup>8</sup>Nitte (Deemed to be University), Nitte University Centre for Science Education and Research, Medical Sciences Complex, Deralakatte, Mangalore 575018, India

\*Correspondence: [v.kumar@radboudumc.nl](mailto:v.kumar@radboudumc.nl)

<https://doi.org/10.1016/j.ajhg.2022.01.014>

© 2022 American Society of Human Genetics.



only. Arguably, the genetic architecture of polymorphisms in the *TLR10-TLR1-TLR6* region might have differential effect on various human phenotypes and also varies substantially among populations. Therefore, further exploration of this genomic region by leveraging different molecular phenotypes and individuals from other ethnicities is warranted to draw general conclusions.

Africa is the continent where modern humans evolved and from where they later migrated all around the world. Africa is also extremely diverse culturally, ethnically, and genetically. Burgeoning evidence in the scientific literature shows that Africans have the most genetic variation among all other populations.<sup>10,11</sup> Despite this genetic richness, African populations are poorly represented in genetic and functional genomic research, especially compared to individuals of European ancestry.<sup>12,13</sup> In the context of induced cytokines responses for example, to the best of our knowledge, only the Hi-Host Phenome Project (H2P2) has incorporated individuals of African ancestry in a study that aimed to identify human genetic variation in pathogen-induced cellular traits. However, this study was conducted in lymphoblast cell lines and also not in Tanzanians.<sup>14</sup> The Tanzanian population, among the sub-Saharan African regions, has high prevalence of infectious diseases<sup>15</sup> with diverse patterns of causative agents compared to high-income countries.<sup>16</sup> Exposure to different environmental conditions including infectious or microorganisms may affect certain genetic markers through positive selection<sup>17</sup> with differential effect on modern human populations.<sup>18</sup>

Therefore, the primary aim of this study was to investigate genetic variants modulating *ex vivo* microbial-induced cytokine responses in healthy adults of East African ancestry. The other important aim was to assess whether distinct genetic variants and pathways contribute to inter-population heterogeneity to cytokine responses upon stimulation.

To achieve these aims, we extended the HFGP to include a population of Sub-Saharan Africa: we performed a functional genomics study in healthy adult individuals in the Kilimanjaro region in Northern Tanzania.<sup>19</sup> Here, we not only identify significant genetic loci that affect cytokine production in Tanzanian individuals, but we also show considerable differences in the genetic basis for cytokine production between the East Africans from Tanzania and Western European individuals. Finally, we provide evidence for ancestry-associated differences in inflammatory pathways that regulate cytokine production in response to infection.

## Subjects and methods

### Study design and participants

Data from the 300 Tanzania functional genomics (FG) (300TZFG) were used in this study. Details of this cohort have been described recently.<sup>19</sup> In summary, the 300TZFG cohort consists of apparently healthy Tanzanian individuals, aged between 18 and 65 years, living in the Kilimanjaro region in Northern Tanzania. All

participants had a negative rapid diagnostic test for malaria and HIV. Exclusion criteria were pregnancy, any acute or chronic disease, use of antibiotics or anti-malaria medication in the 3 months before sampling, tuberculosis in the past year, a blood pressure  $\leq 90/60$  mmHg or  $\geq 140/90$  mmHg, or a random blood glucose  $> 8.0$  mmol/L. In this study, a total of 323 and 307 participants had genotype and cytokine data, respectively. After quality control procedures, matched genotype data (SNPs dosages) and cytokine data were available for 271 individuals, consisting of 137 males and 134 females. A general overview of the study samples at various quality control steps are depicted in [Figure S1](#). The general 500FG cohort of individuals of Western European origin consists of 237 males and 296 females with age range of 18 to 75 years.

### Ethics approval and consent to participate

The 300TZFG study was approved by the Ethical Committees of the Kilimanjaro Christian Medical University College (CREC) (no. 2443) and the National Institute for Medical Research (NIMR/HQ/R.8a/Vol. IX/2290 and NIMR/HQ/R.8a/Vol.IX/3318) in Tanzania. Written informed consent was obtained from all subjects. The 500FG cohort study was approved by the Ethical Committee of the Radboud University Medical Centre Nijmegen, the Netherlands (NL42561.091.12, 2012/550). Informed consent was provided by all the participants.

### Whole blood stimulation experiments

As previously described,<sup>19</sup> *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 ten stimuli (see Table S16 from Temba et al.<sup>19</sup> for details) including bacterial and fungal pathogens as well as TLR3 and TLR4 agonists. 100 mL of heparin blood was added to a 48-wells culture plate and subsequently stimulated with 400 mL of stimulus for 48 h at 37°C and 5% CO<sub>2</sub>. Stimuli were prepared in RPMI culture medium (Dutch modified, Invitrogen) supplemented with 50 µg/mL gentamicin, 2 mM Glutamax, and 1 mM pyruvate. Supernatants were collected and stored at -80°C until used for ELISA. To minimize variation between measurements, we measured all samples by using kits of the same lot number.

### Cytokine measurements

Concentrations of cytokines interleukin (IL)-6, IL-1b, interferon (IFN)-g, tumor necrosis factor (TNF)- $\alpha$ , and IL-10 were measured in the stored supernatants via ELISA according to the instructions (given (IL)-6, IL-1b, IL-10, and TNF- $\alpha$ : R&D Systems; IFN-g: Sanguin). We excluded IL-10 cytokine responses to *S. aureus*, *C. albicans*, and PolyIC, as over 75% of the individuals had values below the detection limit, resulting in 47 cytokine-stimulation combinations used in this analysis.

### Quality control of cytokine data

Preprocessing of cytokine data was performed before statistical analysis. Raw cytokine concentrations were first log<sub>2</sub>-transformed followed by normalization for approximation of standard Gaussian distribution ([Figure S2A](#)). Inverse-ranked normalization was performed with the “*rnttransform*” function implemented in the GenABEL R package.<sup>20</sup> To examine the presence of outliers, we performed unsupervised hierarchical clustering analysis by using Pearson’s correlation as a measure of similarity. All the samples in the dataset clustered uniformly ([Figure S2B](#)).

## SNP genotyping, quality control, and imputation

DNA was extracted from whole blood with the DNeasy kit. Genotyping was performed with the Global Screening Array (GSA) SNP chip. We used default settings of OptiCall 0.70<sup>21</sup> to perform genotype calling. Quality control filters prior to imputation include excluding variants with call rate exceeding 0.1, low minor allele frequencies (MAFs < 0.001), and SNPs deviating from Hardy-Weinberg equilibrium (HWE) with a  $p$  value <  $1 \times 10^{-4}$ . Next, we excluded 15 samples that were potential genetic outliers through identification by using multi-dimensional scaling plots. A total of 409,261 variants and 308 samples passed the quality control procedures. Strand alignment to a reference panel, 1000 Genomes reference panel dataset, was performed via Genotype harmonizer.<sup>22</sup> To improve genome coverage, we performed genotype imputation for all autosomal chromosomes by using the Minimac4 software through the publicly available Michigan Imputation Server.<sup>23</sup> The Human Reference Consortium (HRC r.1.1 2016) was used as a reference panel and the dataset was phased with Eagle v2.3. Variants with imputation quality score ( $R^2$ ) < 0.3 were excluded from further analysis. Genotyping and imputation generated a total of 5,271,779 variants from 308 individuals. We further excluded samples because of extreme heterozygosity rates and cryptic relatedness. No sample was removed as a result of incorrect or ambiguous sex information when compared with self-reported sex in the phenotype data. We used a multidimensional scaling approach to examine potential population structure by merging our data with the 1000 Genomes Project data. All individuals in our data clustered homogeneously with individuals of African origin included in the 1000 Genomes dataset (Figure S2C). However, by excluding non-Africans, the Tanzanian data clustered differently from the Africans in the 1000 Genomes Project data (Figure S2D). We considered SNPs with MAFs  $\geq 5\%$  and not deviating from HWE with  $p$  value >  $1 \times 10^{-6}$ , yielding a final dataset of 5,269,992 SNPs that were subsequently used for cytokine QTL analysis.

## Cytokine quantitative trait loci (cQTLs) mapping

The cQTL analysis was performed with a linear model implemented in the Matrix-eQTL R package.<sup>24</sup> We adjusted the linear model for age, sex, and residency (rural and urban samples) on the inverse ranked normalized cytokine concentrations. We considered the conventional genome-wide significance threshold for statistical significance of  $p$  value  $\leq 5 \times 10^{-8}$  to account for multiple testing.

For genome-wide QTL mapping, multiple SNPs are tested throughout the genome and based on linkage disequilibrium (LD) between SNPs, a particular locus might contain multiple cytokine-associated genetic variants. Therefore, we performed LD clumping with the greedy algorithm (SNPs in LD are removed in ascending order of  $p$  values) implemented in PLINK<sup>25</sup> to identify independent associations. Here, our imputed genotype data were used as a reference dataset for estimating LD, clumps around lead SNPs were formed with a default window size of 250 kb, and  $r^2$  threshold greater than 0.1 was applied.

## Meta-analysis

We conducted a meta-analysis by using a weighted sum fixed-effect model approach as implemented in METAL software program.<sup>26</sup> Meta-analysis was performed by integrating summary statistics of cytokine QTL association results from this study and the 500FG

study. To explain briefly, the 500FG is a population-based cohort made up of healthy adult volunteers of Western European origin from the Human Functional Genomics Projects. This cohort consists of 237 males and 296 females with age range of 18 to 75 years. Previously, researchers from our group used this cohort to perform cytokine QTL analyses on pro-inflammatory cytokine responses after bacteria, fungal, and viral stimulations.<sup>6</sup> The meta-analysis was restricted to only cytokine-stimulation pairs common in both studies. We calculated heterogeneity statistics based on the chi-square test for all SNPs to estimate heterogeneity of effect sizes in both cohorts. In general, associations with meta  $p$  value  $\leq 5 \times 10^{-8}$  were considered genome-wide significant.

## Pathway enrichment analysis

The potential biological significance of cQTL was assessed via over-representation analysis (ORA), which performs a hypergeometric test to identify causal pathways. WebGestalt, a freely available online tool and one of the most widely used gene set enrichment analysis tools, was used. We determined statistically significant pathways after correcting for multiple testing (false discovery rate  $p < 0.05$ ) by using the Benjamini-Hochberg method.

## Haplotype and LD block analysis

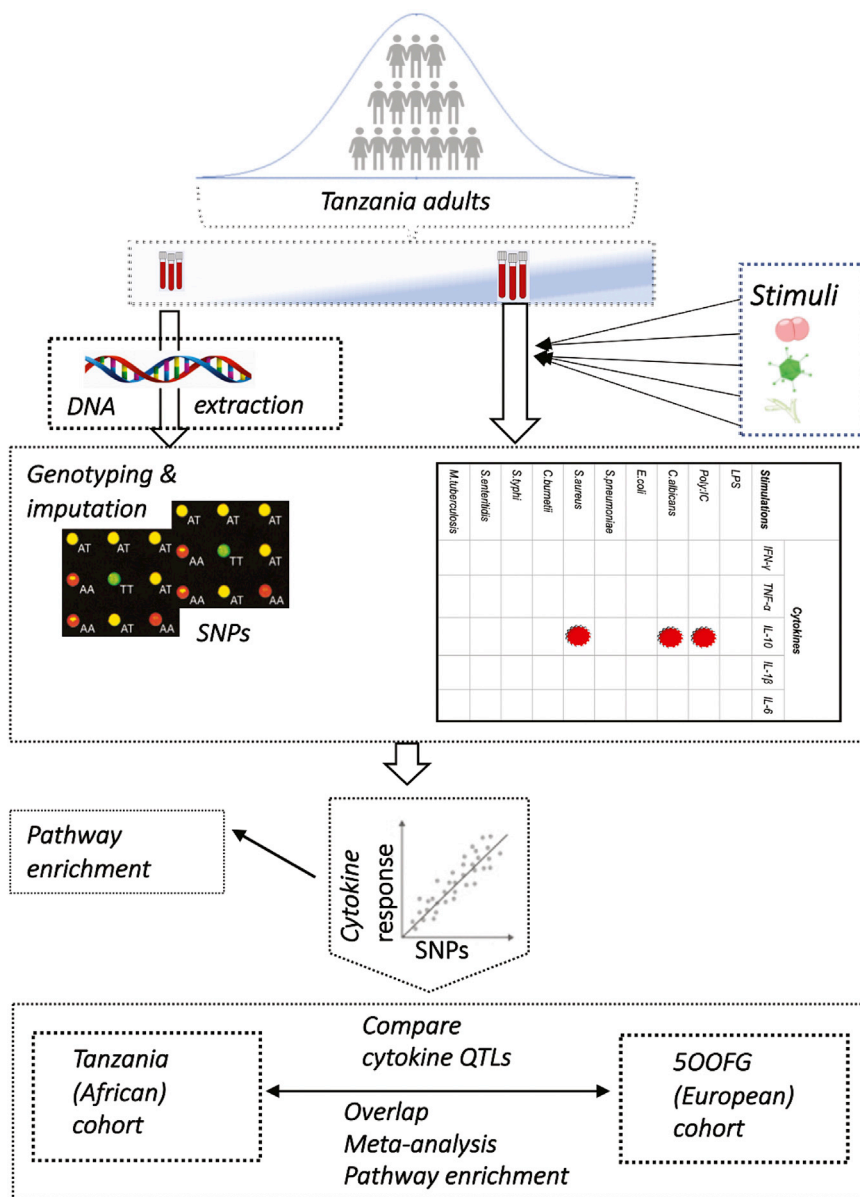
We extracted common genetic markers that passed quality control after imputation between the 500FG and 300TZFG cohorts to determine haplotype blocks separately in both cohorts by using PLINK (v1.09) software. Default parameters such as pairwise LD calculation based on SNPs within 200 kb window were applied. For each block that was identified per chromosome, we analyzed the number of haplotypes and median haplotype block length between the two populations and carried out Wilcoxon signed-rank non-parametric test to assess statistical significance. Pairwise LD measured as  $R^2$  for genetic markers within genomic regions were estimated and visualized with the R package LDheatmap.<sup>27</sup>

## Genetic differentiation or divergence analysis

Interpopulation genetic differentiation analysis was conducted between the two populations with the fixation index ( $F_{st}$ ) metric as implemented in PLINK (v1.09) software. Average  $F_{st}$  values were calculated with a step size of 10 kb. The analysis was performed on common SNPs remaining after quality control in each cohort. We used empirical approach by selecting  $F_{st}$  values exceeding 99<sup>th</sup> percentile per SNP to determine significance.

## Statistical analyses

All statistical analyses were performed in R: a free software environment for statistical computing and graphics.<sup>28</sup> We used Fligner-Killeen non-parametric test for homogeneity of group variances and robust to dataset with outliers to compared cell culture medium with stimulated cytokine levels. Quality control, preprocessing analysis of genetic data, and LD clumping procedures around any lead SNPs analysis to determine independent loci were performed with the free, open-source whole-genome analysis toolset PLINK 1.70.<sup>25</sup> The cytokines association with imputed genome-wide data analyses were performed with a linear model as implemented in the Matrix-eQTL R package.<sup>24</sup> The meta-analyses on cQTL summary statistic ( $p$  values) were performed with the fixed effects sample-size-weighted analysis method implemented in the METAL software.<sup>26</sup>



**Figure 1. Schematic diagram of study design and bioinformatics analyses**

We utilized samples of a cohort of healthy Tanzanian adults. *In vitro* stimulation of their blood was performed with ten different stimuli, including bacterial, fungal, and TLR3 and TLR4 agonists followed by measuring five different cytokines in the supernatant. Three (indicated by red ovals) out of the resulting 50 cytokine-stimulation pairs were excluded from downstream analysis after quality control. Genotyping was performed using global screening arrays (GSAs) SNP chip. We performed genome-wide SNP cytokine quantitative trait loci (cQTL) mapping by correlating cytokine abundances and imputed genotype data and subsequently performed pathway enrichment analyses. Finally, we compared the results with another cohort of Western European ancestry individuals by using three different approaches: direct overlap of genetic variants, meta-analysis, and functional enrichment analysis.

with *S. enteritidis* produced the most abundant cytokine response, whereas the least abundance was observed in *C. burnetii*-stimulated cells. Also, we observed that there is variation in cytokine abundances per stimulation and this is not because of the presence of multiple outliers as depicted in (Figure 2A). Boxplots supporting this observation are presented in (Figure S3A). Moreover, we observed a high level of interindividual variations in cytokine concentrations of samples after stimulation (Figure 2B). To test whether there is a statistically significant difference in the interindividual variation in the production of cytokines upon stimulation compared to

unstimulated state, we applied the Fligner-Killeen test for group comparison. We observed more significant increase in interindividual variations in stimulated conditions than in unstimulated states ( $p$  value  $< 2 \times 10^{-16}$ ). Computing paired Wilcoxon tests resulted to the same level of significance (Figure S3B). Interestingly, this interindividual variation was highest for IFN-g production under most stimulations, suggesting the major role of host factors (genetic and non-genetic) in determining IFN-g production capacity.

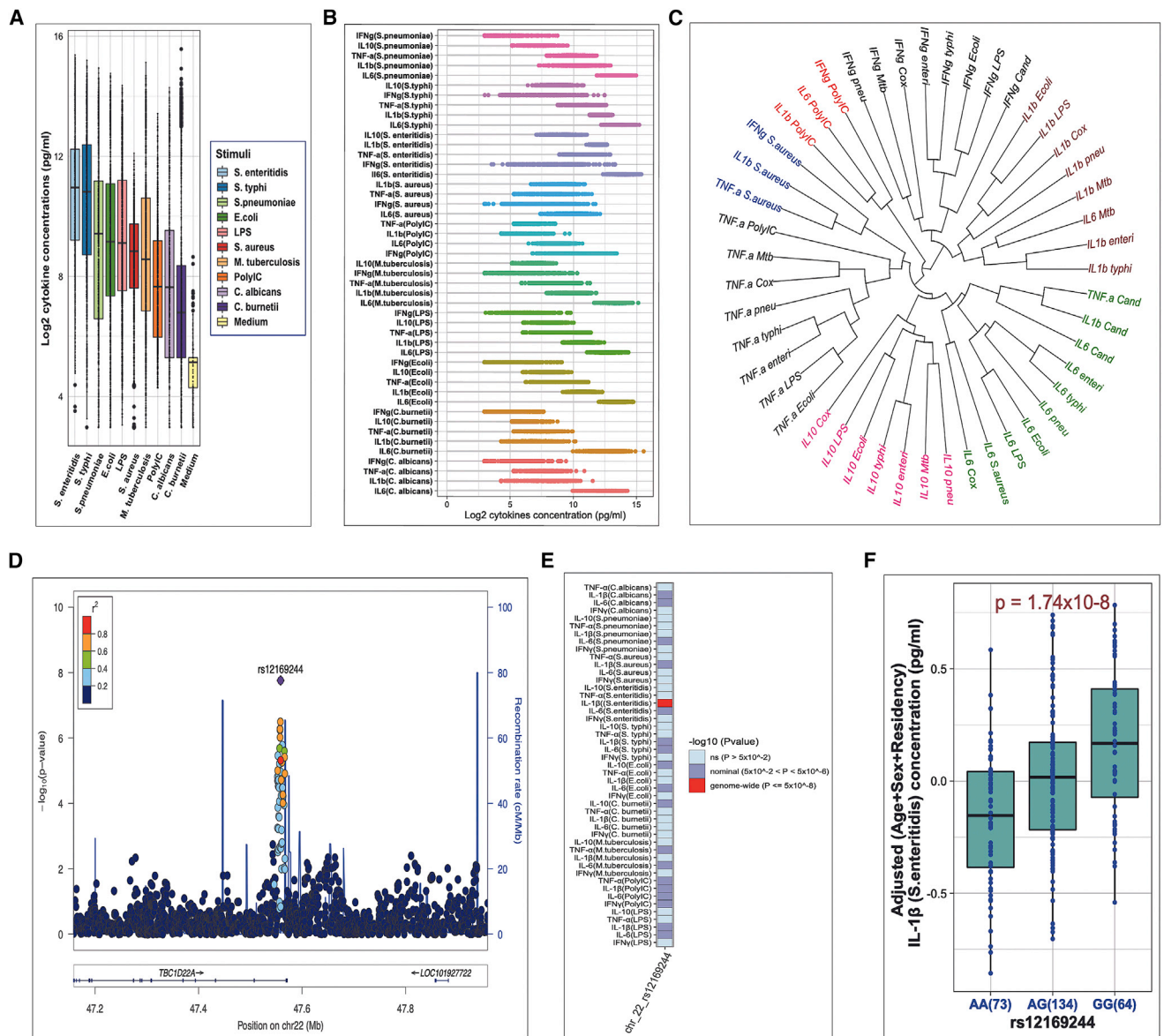
#### Induced cytokine responses are clustered in cytokine-dependent manner

The architectures of cytokine response in a population can be dependent on the type of cytokine or the microbial ligands used for the stimulation. Thus, to mine

## Results

### Significant increase in cytokine abundance and interindividual variations upon stimulation

To investigate the contribution of common genetic variations on human induced cytokine responses, we restricted downstream analyses to the 47 cytokine-stimulation combinations passing quality control procedures. Details of these cytokine-stimulation pairs are illustrated in the general overview of the study design and bioinformatics analysis represented in Figure 1. First, we compared the abundance of cytokines from unstimulated state (cell culture medium) and stimulated conditions to establish their potential relationship. Compared to the spontaneous release of cytokines from cells in culture medium and as expected, we observed higher production of cytokines upon stimulation with microbial ligands (Figure 2A). On average, stimulation-induced



**Figure 2. General overview of cytokine responses architecture and summary of genome-wide significant cQTL results**  
 (A) Boxplot comparing cytokine production capacity of unstimulated state (cell culture medium) and ten different stimuli. Profiles of all cytokines (IFN- $\gamma$ , TNF- $\alpha$ , IL-1 $\beta$ , IL-10, and IL-6) were merged per stimuli to compare with unstimulated state.  
 (B) Dot plot depicting interindividual variability in cytokine response upon stimulation. While the x axis denotes cytokine concentrations after log<sub>2</sub> transformation, the y axis represents the total 47 cytokine-stimulation combinations used in the current study. The dot plots have been arranged in ascending order of median values per stimuli.  
 (C) Dendrogram visualization of unsupervised hierarchical clustering of cytokine-stimulation pairs based on cytokine abundances. The “Euclidean” method in the “hclust” function was used for computing distances.  
 (D) Regional association plot at the *TBC1D22A* locus (top SNP rs12169244) associated with IL-1 $\beta$  upon *Salmonella enteritidis* stimulation. Other SNPs flanking a genomic window of 400 kb are color-coded according to their linkage disequilibrium ( $r^2$ ) with the top SNP (purple color). The horizontal axis indicates chromosomal positions (NCBI human genome build 37) and the vertical axes represent  $-\log_{10}$  p values and recombination rates (cM/Mb) estimated from 1000G (African population) version 3.3.  
 (E) Heatmap showing the association between the genome-wide significant SNP with all cytokine-stimulation pairs (on the vertical axis). SNP-cytokine associations with p values greater than  $5 \times 10^{-2}$  were considered statistically non-significant (ns).  
 (F) Boxplot showing distribution of cytokine concentrations (pg/mL) stratified by genotypes of the top SNP (rs12169244).

patterns underlying induced cytokine responses, we performed unsupervised hierarchical clustering on the cytokine stimulation combinations. In general, we observed distinct correlations at cytokine level irrespective of pathogens (Figure 2C). Interestingly, all cytokines, including anti-inflammatory cytokine IL-10, showed a

very strong cytokine-dependent clustering except for PolyIC stimulation. Many of these stimuli are known to activate TLR-signaling to affect cytokine production. Why PolyIC-induced cytokine production is dependent on the pathogen compared to other stimuli needs to be investigated further.

**Table 1. Summary of top six independent SNP-cytokine loci**

Loci	SNPs	Chr	Base pair	Cytokines	Stimulation	p value	Causal genes
1	rs12169244	22	47,558,596	IL-1b	<i>S. enteritidis</i>	$1.74 \times 10^{-8}$	<i>TBC1D22A</i> <sup>a</sup>
2	rs9563018	13	51,656,046	IL-1b	<i>M. tuberculosis</i>	$7.81 \times 10^{-8}$	<i>LINC0037</i> , <sup>a</sup> <i>GUCY1B2</i> <sup>b</sup>
3	rs4474665	16	17,498,563	IFN-g	<i>S. enteritidis</i>	$7.99 \times 10^{-8}$	<i>XYLT1</i> <sup>a</sup>
4	rs74115411	1	153,331,776	IL-6	<i>S. aureus</i>	$8.90 \times 10^{-8}$	<i>S100A9</i> <sup>a</sup>
5	rs11829172	12	98,579,287	IL-1b	<i>S. aureus</i>	$9.63 \times 10^{-8}$	<i>MIR4303</i> <sup>a</sup>
6	rs10483241	12	48,630,139	IFN-g	<i>C. albicans</i>	$9.71 \times 10^{-8}$	<i>MIR4303</i> <sup>a</sup>

<sup>a</sup>The gene in closest proximity to the cytokine QTL SNPs.

<sup>b</sup>Expression of this gene shows association with cytokine QTL SNP in whole blood.

### Identification of genome-wide genetic variants affecting induced cytokine responses

Next, to identify genetic variants significantly associated with cytokine production, we mapped cQTLs at genome-wide scale. We identified one genome-wide significant locus on the basis of the stringent p value threshold of  $5 \times 10^{-8}$ , while 80 other loci showed strong suggestive independent associations (p value  $> 5 \times 10^{-8}$  to  $1 \times 10^{-6}$ ) after adjusting for covariates (Table S1). The genome-wide significant hit was SNP rs12169244 on chromosome 22, correlating with IL-1b production upon *S. enteritidis* stimulation (Figure 2D). SNP rs12169244 is an intronic variant mapping near *TBC1D22A* (MIM: 616879) gene (TBC1 domain family member 22A), which acts as a GTPase-activating protein for Rab family protein(s). Interestingly, the genome-wide significant SNP also showed nominal association with 18 other cytokine-stimulation pairs, thus suggesting a pleiotropic effect (Figure 2E). At this locus, relative to individuals carrying the AA genotypes, we observed elevated cytokine production in individuals with GG genotypes (Figure 2F). Quantile-quantile (QQ) plot of the association results at this locus is presented in Figure S2E. The genomic inflation factor or lambda(l) was observed to be 1.01, suggesting no evidence of systematic bias in the analysis. We report more detailed information of the top six independent associations in Table 1.

### Functional enrichment analysis of cQTLs identified in Tanzania cohort

Based on the observation that cytokine-stimulation pairs clustered in a cytokine-dependent manner (Figure 2C), we hypothesized that there are common pathways affecting cytokine production, irrespective of the pathogen that induces the stimulation. Therefore, to identify these pathways, we first prioritized suggestive independent cQTLs (p value  $< 1 \times 10^{-5}$ ) for each of the five cytokines (IL-6, IL-1b, TNF- $\alpha$ , IFN-g, and IL-10) and extracted gene sets mapping near these genetic variants by using a window size of 250 kb upstream and downstream of each SNP. Henceforth, we refer to these groups as TNF- $\alpha$ -based, IL-6-based, IFN-g-based, IL-1b-based, and IL-10-based pathways. As expected, we found significant enrichment

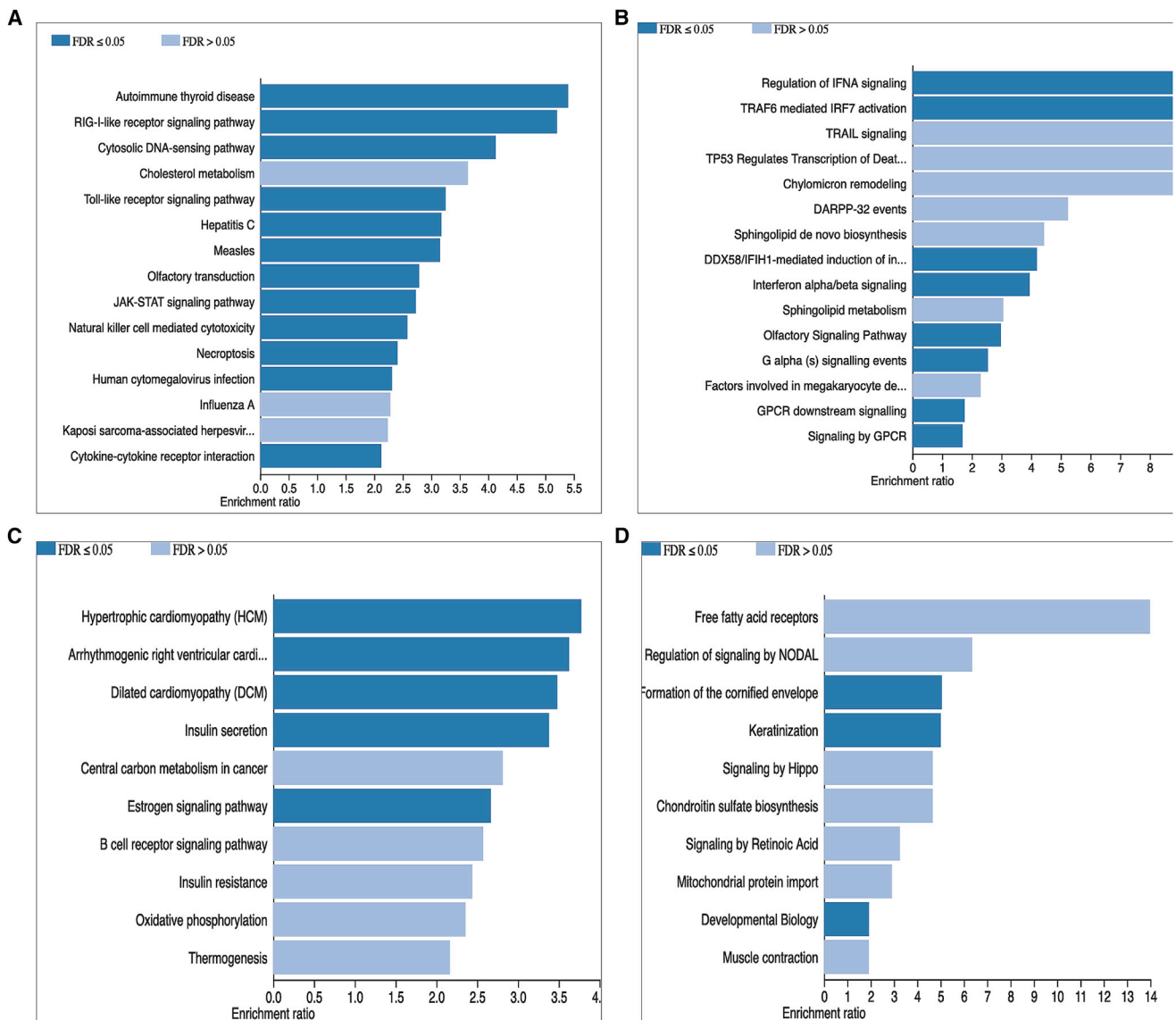
of the genes at IFN-g-based cQTLs for various diseases (Figure 3A), and regulation of interferon signaling (Figure 3B), suggesting that *cis*-genes around IFN-g cQTLs determine IFN-g production by mainly regulating pathogen recognition receptors such as TLRs and DDX58/IFIH1-mediated signaling. Moreover, we observed significantly enriched pathways; for example, estrogen signaling pathway influences the TNF- $\alpha$ -based category (Figures 3C and 3D). Estrogen is known to regulate cytokine gene expressions in different cell types through estrogen receptor-mediated pathways.<sup>29</sup> Bar chart plots showing enrichment results of IL-6-, IL-1b-, IL-10-based pathways, with less significant enriched pathways, are presented in Figures S4A–S4F.

### Limited overlap between cQTLs identified in European and African samples

We next assessed the presence of any systematic differences between cQTLs identified via two population-based datasets of African (300TZFG) and European (500FG) individuals. To estimate the degree of shared genetic variants between these two populations, we directly compared the four cytokines (IFN-g, TNF- $\alpha$ , IL-1b, and IL-6) and three stimulations (LPS, *Candida albicans*, and *Staphylococcus aureus*) that were common to both cohorts (Figure 4A). Across all the 12 cytokine-stimulation pairs considered, we found only 1.6% to 1.9% of SNPs overlapping between both populations with nominal association threshold (p value  $< 5 \times 10^{-2}$ ). For example, a Venn diagram depicting this observation for *S. aureus*-cytokines pairs is shown in Figure 4B, while results for LPS- and *C. albicans*-cytokines are represented in Figure 4C. In addition, no evidence of variant sharing was detected by restricting the analysis to variants showing suggestive association with cytokines (p value  $< 5 \times 10^{-6}$ ), indicating that the identified genetic variants with strong cQTL effect are mostly population-specific variants.

### Meta-analysis revealed shared genetic locus related to cytokine response in African and European cohorts

We further sought to identify cQTLs that show effect on cytokines in the same allelic direction in both populations by meta-analyzing summary statistics of each marker



**Figure 3. Bar graphs summarizing IFN-g- and TNF- $\alpha$ -based pathway enrichment results**

(A and B) Top 15 enriched KEGG (A) and reactome (B) pathways for gene sets curated from SNPs associated with IFN-g. (C and D) Top 10 enriched KEGG (C) and reactome (D) pathways for gene sets curated from SNPs associated with TNF- $\alpha$ . Darker blue bars represent statistically significant pathways and lighter blue bars represent non-significant pathways after multiple testing. We set the false discovery rate (FDR)-adjusted p value to 0.05 as significance level.

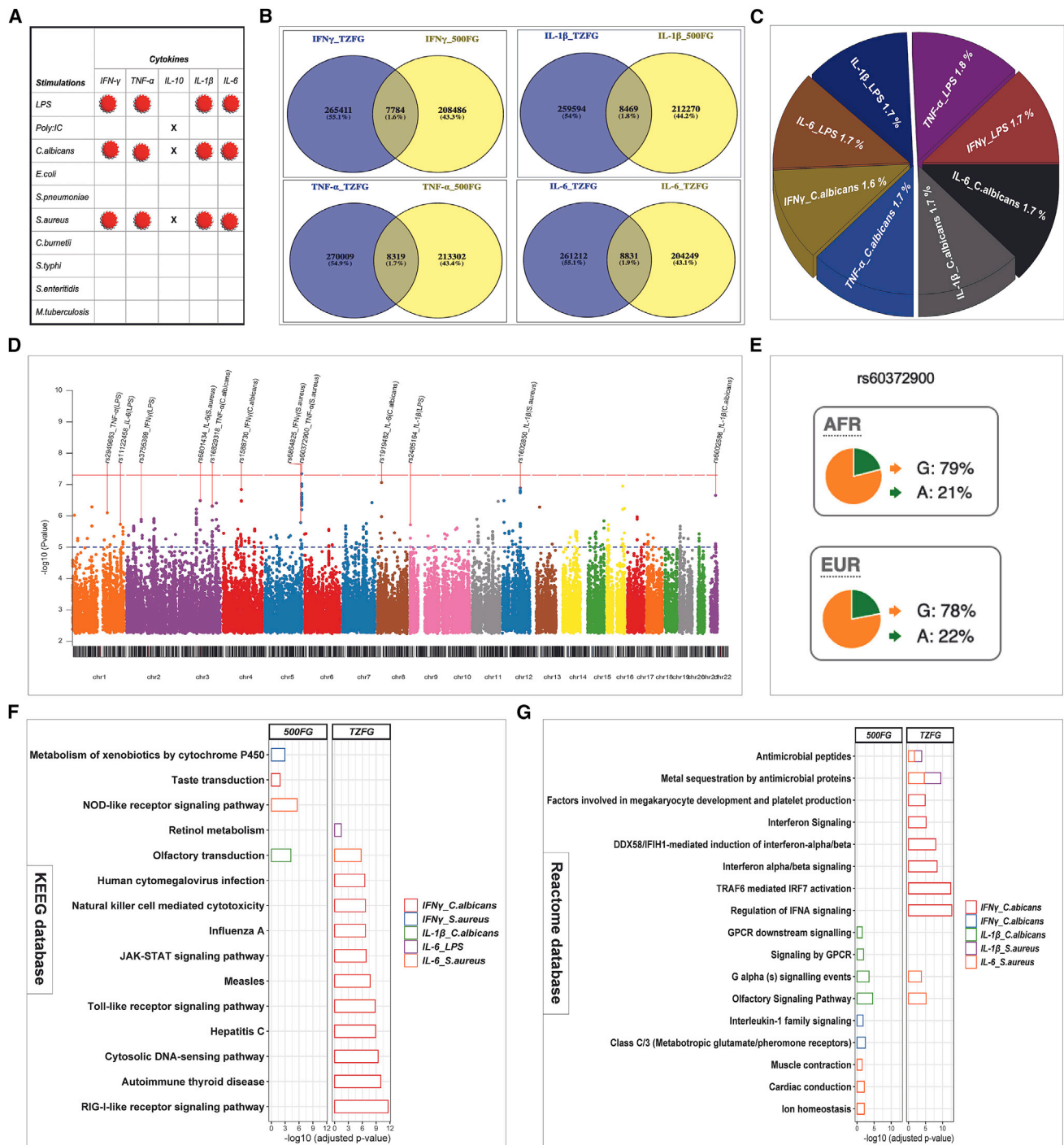
across the two cohorts. To accomplish this, we extracted the common markers showing nominally significant association ( $p$  values  $< 5 \times 10^{-2}$ ) in each independent study. Interestingly, among all the 12 cytokine-stimulation pairs analyzed separately, we identified one genome-wide significant locus at chromosome 5, the lead SNP rs60372900 correlating with TNF- $\alpha$  levels after *S. aureus* stimulation. This intronic variant is mapped near a protein-coding gene *ZNF354A*. Remarkably, data from 1000 Genomes Project Phase 3 shows nearly equal minor allele frequencies of 21% and 22% in Europeans and Africans, respectively, of rs60372900. (Figure 4E). This observation underscores the regulatory potential of this variant on induced cytokine levels in both populations. Meta-analysis performed by integrating the African and the European

data also identified several suggestive ( $p$  value  $< 1 \times 10^{-5}$ ) associations (Figure 4D). Summary statistics of meta-analysis results are presented in Table 2.

#### Population-specific cQTLs converge on distinct pathways to regulate cytokine levels

As we observed a few overlaps of cQTLs between the African and the European cohorts (Figures 4B and 4C), we tested whether these cQTLs affect cytokine production capacity through distinct pathways. We first identified independent suggestive associated cQTLs ( $p$  value  $< 1 \times 10^{-5}$ ) for all cytokine-stimulation pairs separately for each population, and then we extracted genes around these cQTLs and performed overrepresentation analysis as described in the subjects and methods section. Intriguingly, we





**Figure 4. Comparison of cQTLs between Tanzanian (TZFG) and Western European dataset (500FG)**

(A) Illustration of cytokine-stimulation combinations common in both African and European datasets (indicated by red circles). Cells with “x” labels denote cytokine-stimulation pairs excluded from downstream analysis, as more than 75% of samples had values below the detection limit of the assay.

(B) Venn diagrams illustrating *S. aureus* induced cytokine QTLs overlapping between individuals of African (TZFG) and European (500FG) ancestry.

(C) Pie chart showing the degree of *LPS*- and *C. albicans*-induced cytokine QTLs overlapping between individuals of African (TZFG) and European (500FG) ancestry.

(D) Manhattan plot of SNP-cytokines association test results indicating shared cQTL locus between individuals of African and European ancestry. The red horizontal dashed line represents the genome-wide significant threshold ( $p$  value  $< 1 \times 10^{-8}$ ) and the blue dashed line denotes the suggestive evidence of association threshold. Nominally significant ( $p$  value  $< 5 \times 10^{-2}$ ) cQTLs are plotted.

(E) Bar graphs representation of significantly enriched KEGG pathways after FDR correction.

(F) Bar graphs representation of significantly enriched reactome pathways after FDR correction. The colors of the bars associated with each pathway correspond to cytokine-stimulation pairs as shown by the color legend. Pathway analysis was performed with WebGestalt tool with subsequent visualization with R statistical software.

**Table 2. Summary of top loci identified by meta-analysis of TZFG and 500FG dataset**

Loc	SNPs	Chr	Base pair	Cytokines	Stimulation	Z score	Meta p value	TZFG p value	500FG p value	Allelic direction	Causal genes
1	rs60372900	5	178,220,798	TNF- $\alpha$	<i>S. aureus</i>	5.47	$4.50 \times 10^{-8}$	$6.56 \times 10^{-3}$	$1.24 \times 10^{-6}$	++	<i>AACSP1</i> <sup>a</sup>
2	rs1919482	8	19,754,718	IL-6	<i>C. albicans</i>	5.353	$8.69 \times 10^{-8}$	$2.31 \times 10^{-5}$	$5.78 \times 10^{-4}$	++	<i>LPL</i> , <sup>a,b</sup> <i>INT</i> , <sup>b</sup> <i>CSGALNC</i> <sup>b</sup>
3	rs1602850	12	86,657,959	IL-1b	<i>S. aureus</i>	5.281	$1.29 \times 10^{-7}$	$1.13 \times 10^{-4}$	$2.54 \times 10^{-4}$	++	<i>MGAT4C</i> <sup>a</sup>
4	rs1588730	4	89,955,623	IFN- $\gamma$	<i>C. albicans</i>	5.26	$1.44 \times 10^{-7}$	$2.70 \times 10^{-5}$	$8.05 \times 10^{-4}$	++	<i>FAM13A</i> , <sup>a,b</sup> <i>TIGD2</i> <sup>b</sup>
5	rs6002586	22	42,441,412	IL-1b	<i>C. albicans</i>	5.18	$2.22 \times 10^{-7}$	$9.63 \times 10^{-4}$	$6.57 \times 10^{-5}$	++	<i>WBP2NL</i> , <sup>a</sup> <i>SMDT1</i> , <sup>b</sup> <i>CYP2D6</i> , <sup>b</sup> <i>MEI1</i> , <sup>b</sup> <i>NAGA</i> , <sup>b</sup> <i>PHETA2</i> <sup>b</sup>
6	rs6801434	3	100,809,977	IL-6	<i>S. aureus</i>	5.108	$3.25 \times 10^{-7}$	$2.22 \times 10^{-4}$	$3.51 \times 10^{-4}$	++	<i>ABI3BP</i> , <sup>a</sup> <i>TFG</i> <sup>b</sup>
7	rs16829318	3	158,418,720	TNF- $\alpha$	<i>C. albicans</i>	5.032	$4.87 \times 10^{-7}$	$4.08 \times 10^{-4}$	$3.12 \times 10^{-4}$	++	<i>RARRES1</i> , <sup>a</sup> <i>LXN</i> , <sup>b</sup> <i>GFM1</i> , <sup>b</sup> <i>MLF1</i> , <sup>b</sup> <i>RSRC1</i> <sup>b</sup>
8	rs2949663	1	167,391,498	TNF- $\alpha$	<i>LPS</i>	-4.935	$8.00 \times 10^{-7}$	$6.17 \times 10^{-4}$	$3.55 \times 10^{-4}$	-	<i>POU2F1</i> , <sup>a</sup> <i>CD247</i> , <sup>b</sup> <i>AL359962.1</i> <sup>b</sup>
9	rs3755369	2	70,942,964	IFN- $\gamma$	<i>LPS</i>	4.84	$1.30 \times 10^{-6}$	$1.30 \times 10^{-3}$	$2.93 \times 10^{-4}$	++	<i>ADD2</i> , <sup>a,b</sup> <i>CLEC4E</i> <sup>b</sup>
10	rs6864825	5	173,659,256	IFN- $\gamma$	<i>S. aureus</i>	-4.794	$1.64 \times 10^{-6}$	$1.40 \times 10^{-4}$	$2.16 \times 10^{-3}$	-	<i>HMP19</i> <sup>a</sup>
11	rs11122458	1	230,308,855	IL-6	<i>LPS</i>	4.768	$1.86 \times 10^{-6}$	$2.51 \times 10^{-3}$	$2.25 \times 10^{-4}$	++	<i>GALNT2</i> <sup>a</sup>
12	rs2485164	9	1,706,565	IL-1b	<i>LPS</i>	4.761	$1.93 \times 10^{-6}$	$1.00 \times 10^{-3}$	$5.48 \times 10^{-4}$	++	<i>SMARCA2</i> <sup>a</sup>

<sup>a</sup>The gene in closest proximity to the cytokine QTL SNPs.

<sup>b</sup>Expression of this gene shows association with cytokine QTL SNP in whole blood.

observed significant enrichment of genes for different pathways in different populations (Figure 4F). For example, in the Tanzanian cohort, the top-ranked KEGG pathways were mostly related to human viral infectious diseases such as hepatitis C, measles, influenza A, and human cytomegalovirus infection. Notably, these pathways were distinct for both populations. In support of this finding, using Reactome database, we found enrichment for regulation of type I interferon signaling pathways (Figure 4G). In contrast, we observed moderate enrichment for IL-1 signaling pathway and olfactory signaling pathway in the European cohort.

### Exploring the Toll-like receptor (TLR) locus in the Tanzania cohort

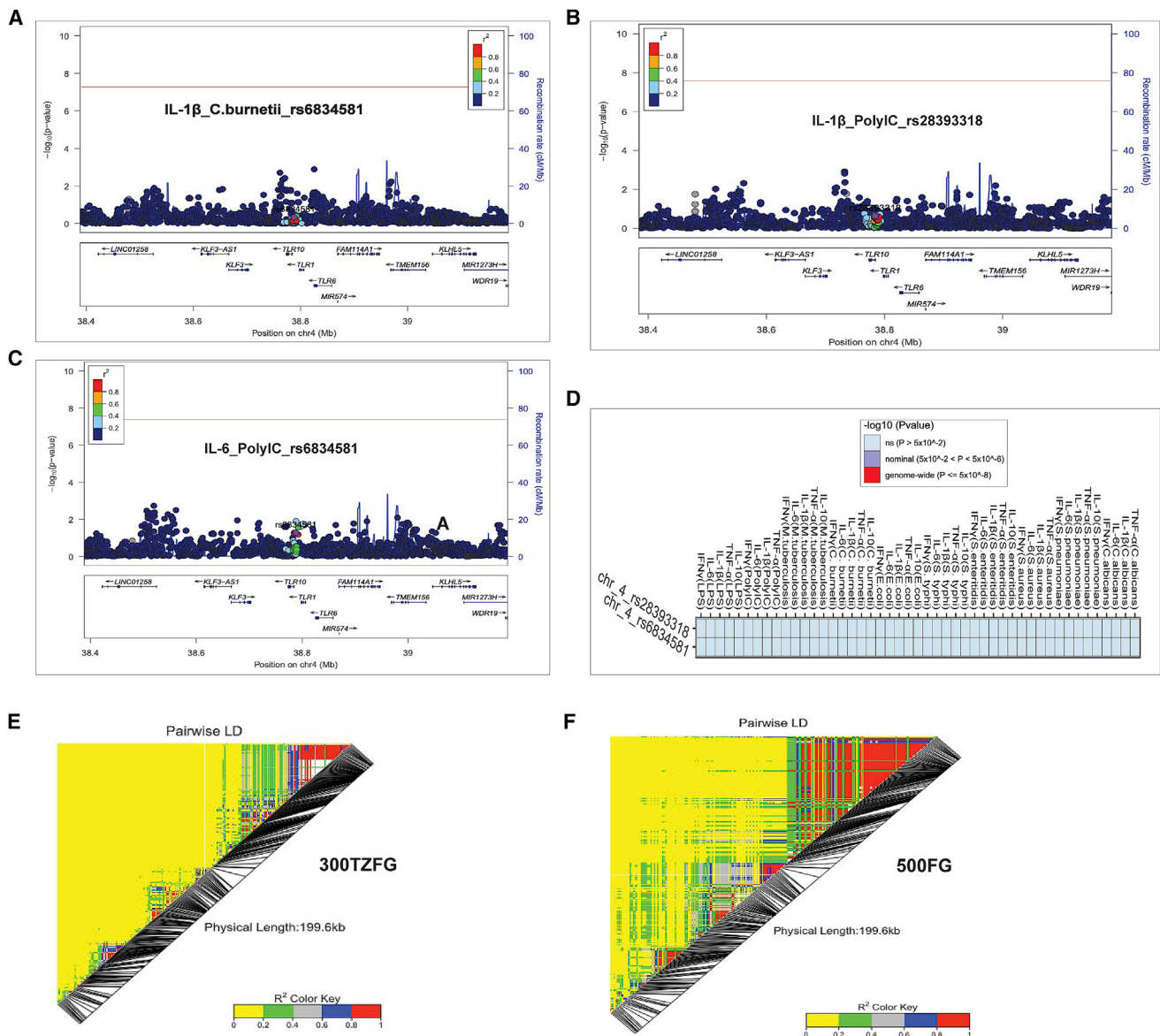
TLRs are pathogen recognition receptors through which our innate immune system “senses” foreign antigens and activates host protective inflammatory responses.<sup>30</sup> It has already been shown that the *TLR10-TLR1-TLR6* locus is under strong positive selection in European populations.<sup>31</sup> Our previous cQTL mapping study in the European population identified 17 independent loci, and among them, the SNPs at the *TLR10-TLR1-TLR6* locus on chromosome 4 showed the strongest association ( $p = 3.9 \times 10^{-25}$ ) with cytokine production.<sup>6</sup> Notably, a SNP, rs28393318, was strongly associated with PolyIC-induced IL-1b production and another SNP, rs6834581, was associated with both PolyIC-induced IL-6 and *C. burnetii*-induced IL-1b levels. Given the availability of these cytokine-stimulation pairs in this study cohort, we attempted to interrogate the transferability of the *TLR10-TLR1-TLR6* locus findings in the

context of cytokine responses to an African population. Mapping cQTLs revealed that, while these polymorphisms (rs6834581 and rs28393318) significantly affect cytokines (IL-6 and IL-1b) in European samples, we found no association ( $p$  value > 0.05) in the African cohort (Figures 5A–5C). Testing the effect of these SNPs on other cytokine-stimulation pairs also showed no association (Figure 5D). Interestingly, the lead cQTL SNPs (with smallest  $p$  values) at these loci in the African cohort are not in LD with the lead SNPs in the European cohort (see Figures S5A–S5C for regional association plots and Figures S5D–S5F for heatmaps depicting the lack of correlation).

In addition, through publicly available eQTL database searches, there was no evidence of the lead cQTL SNPs from the African cohort showing effect on expression levels of *TLR10-TLR1-TLR6* genes (data not shown). Furthermore, we observed varying LD patterns between the populations in this genomic region. While we observed some isolated markers in smaller LD blocks in the Tanzanians (Figure 5E), areas of stronger LD corresponding to uniform haplotype blocks were seen in the Europeans (Figure 5F). This observation points to the role of different haplotypes at this locus in the African population than in the European population contributing to the cytokine regulation in response to pathogens.

### Haplotype block lengths varied between the African and European cohorts

To get some insights into the substantial lack of cQTLs shared between both populations, we tested whether average length of haplotype blocks vary significantly



**Figure 5. Regional association plots and heatmaps of variants in the Toll-like receptor (TLR) locus**

(A–C) Regional association plots of SNPs flanking genomic window of 400 kb around SNPs rs28393318 and rs6834581 associated with IL-6 and IL-1 $\beta$  concentrations at the *TLR10-TLR1-TLR6* locus. These polymorphisms reached the genome-wide significant threshold (red lines) but showed non-significant association in the Tanzania cohort.

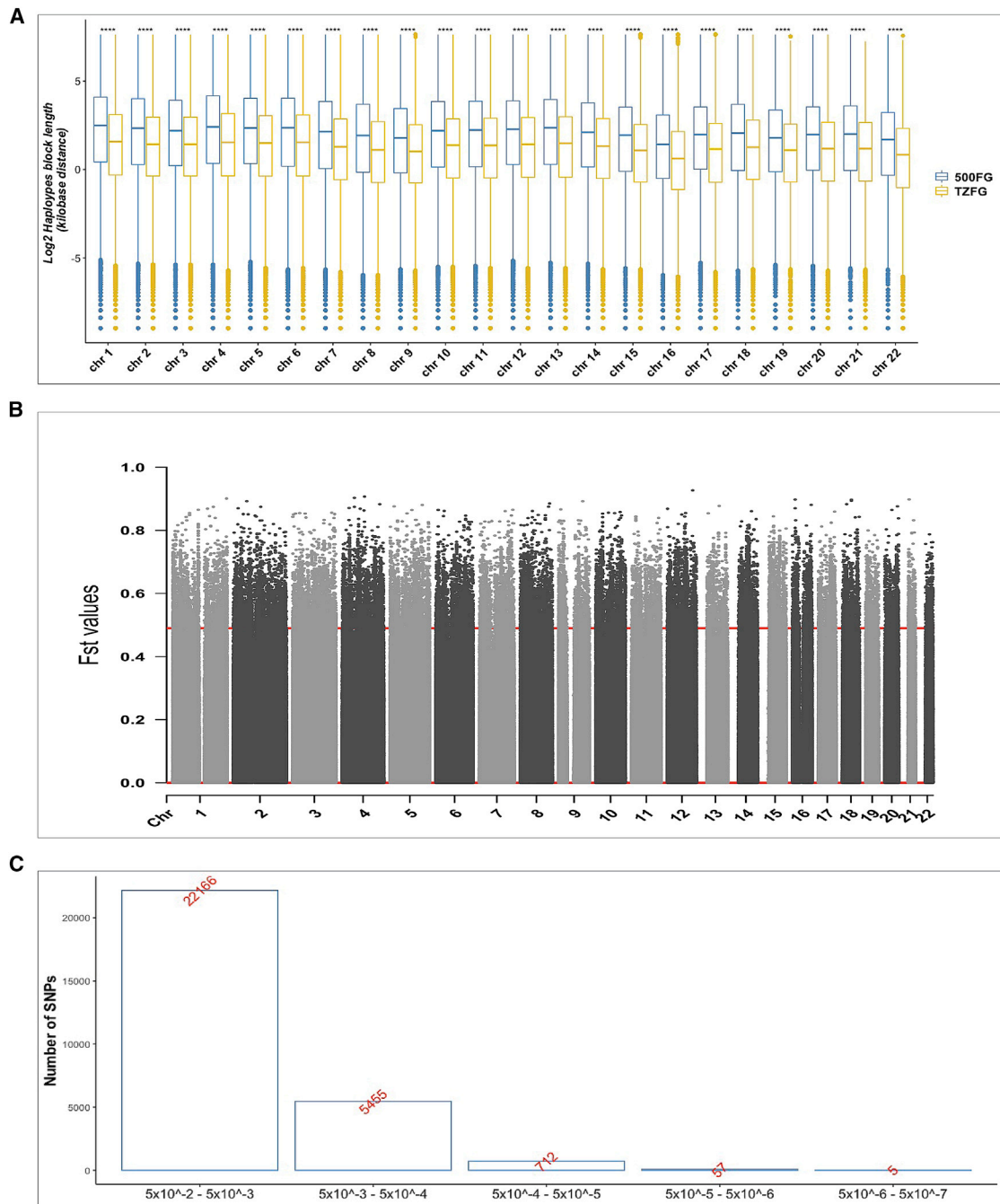
(D) Heatmap showing the association between SNPs (rs28393318 and rs6834581) with all cytokine-stimulation pairs (on the vertical axis). SNP-cytokine associations with p values greater than  $5 \times 10^{-2}$  were considered non-significant (ns).

(E and F) Linkage disequilibrium (LD) patterns of genetic variants in the TLR locus for the Tanzanians (E) and Europeans (F). Red color depicts areas of strong LD and yellow represents sections of weak LD. The black lines show the location of each marker in this genomic region.

between two populations. Across the entire autosomes, we observed significantly higher median haplotype block length in the 500FG samples contrary to the observed haplotype block length in the Tanzanians (Figure 6A). Also, compared to the 500FG samples, we observed higher number of blocks in the Tanzanians. Details of the results supporting Figure 6A are depicted in Table S2. These observations indicate differential genetic architecture underlying both populations, which could partly explain the limited concordance we observed with cQTLs between two populations.

#### Limited evidence for cQTLs in Tanzanians to be under positive selection

To test whether some of cQTLs identified in the Tanzanians are under positive selection, we first computed genome-wide  $F_{st}$  values for each SNP to uncover differential markers or signals of positive selection between both populations. The genome-wide  $F_{st}$  averaged 0.10. The top 1%  $F_{st}$  values, corresponding to a threshold of 0.535, were identified as markers ( $n = 37,867$ ) showing significant differentiation (Figure 6B). We then tested how many of the nominally associated cQTLs ( $p < 0.05$ ) were also showing a very high  $F_{st}$



**Figure 6. Haplotype analysis and candidate markers under positive selection**

(A) Boxplot of median haplotype blocks length (y axis) per chromosome (x axis) for the Tanzanians (yellow) and Europeans (steel blue). The asterisks (\*) show the degree of significance difference between both populations based on Wilcoxon signed-rank non-parametric test.

(B) Manhattan F<sub>st</sub> plot of SNPs (in grays) across the genome. The red horizontal line represents the 99<sup>th</sup> percentile, threshold of significance.

(C) Bar plot of the distribution of significant genetic markers across different p value categories, ranging from nominal to suggestive associations.

value. We found less than 0.75% of the cQTLs ( $p < 0.05$ ) showed  $>0.535$  of F<sub>st</sub> value, indicating the possibility that only a few cQTLs were under positive selection (Table S3). Interestingly, there were five strongly associated cQTLs ( $p < 10^{-6}$ ) with high F<sub>st</sub> value, while the majority of the cQTLs

were nominal associations (Figure 6C). Although this observation suggests lack of strong enrichment of cQTLs to be under natural selection, some of these loci could be interesting candidates to understand their function in the context of evolutionary genetics.

## Discussion

This study identified genetic loci in a Tanzanian population that affect cytokine responses of primary immune cells to diverse pathogenic stimuli. We also show that the genetic loci that regulate cytokine responses differ between populations of African ancestry and Western European ancestry. Notably, there is less than 2% overlap among cQTLs between the two populations, including lack of association at the *TLR10-TLR1-TLR6* locus in the Tanzanian population.

One of the important observations of our study is the pathogen-independent clustering of cytokine responses. The pattern with which an individual elicits cytokine response to pathogens could be either cytokine and/or stimulus dependent. While previous studies in the European population reported pathogen-dependent cytokine responses,<sup>5,6</sup> we observed the contrary in our Tanzanian cohort in which certain cytokine-stimulation pairs, particularly IL-10, clustered at the cytokine level, which suggests that anti-inflammatory responses are organized to respond to different pathogens through a specific cytokine pathway. Unfortunately, IL-10 measurement was not available from the previous 500FG study for exploration of whether this pattern is specific to anti-inflammatory cytokines. Nevertheless, it will be interesting to explore further through future studies whether ancestry could drive the differences at the patterns of cytokine clusters.

Our study also identified a genome-wide significant locus exerting a strong effect on cytokine levels in response to pathogens. The most significant SNP, rs12169244, associated with IL-1b levels resides near the *cis*-region of *TBC1D22A* rather than *IL1B*. In addition, by conducting a meta-analysis between the 500FG and 300TZFG cohorts, we found another genome-wide significant cQTL for TNF- $\alpha$  with the top SNP, rs60372900, residing near *ZNF354A* (MIM: 602444). We made a similar observation in our previous studies using European-based cohorts,<sup>5,6</sup> where all the significant cQTLs were located in *trans*, suggesting the significant role of *trans*-regulatory pathways in modulating cytokine responses upon stimulations. We were unable to establish any correlation between these top SNPs and potential causal genes by using available eQTL datasets. Thus, it is difficult to speculate the mechanistic role of any *cis*-genes, and we therefore emphasize the need for relevant eQTL datasets from the African population.

Comparative analysis of genetic variants associated with cytokine response upon stimulation revealed that most of the genetic variants exert their effect on cytokines response in a population-specific manner, as a very limited percentage (1.6% to 1.9%) of shared effect was observed. This result may not come as a surprise considering the distinct genetic architecture that is known to exist between populations on different continents, most likely due to different infectious pressure.<sup>32</sup> These findings highlight the potential role of common variants to the remarkable interpopulation diver-

sity in immune responses and susceptibility to infectious diseases.<sup>33,34</sup> In this context, by performing simulation analyses, studies have shown that the accuracy of Fst estimation directly depends on the number of genetic polymorphisms,<sup>35</sup> suggesting the intrinsic role of genetic markers on population differentiation estimates. Although our Fst analysis provided some indication for genetic differentiation between the two populations, including a large number of samples and whole-genome sequence data to represent the true population diversity, particularly from the African population, will help in accurately estimating the genetic differentiation. Evidently, non-genetic factors such as gut microbiome,<sup>36</sup> seasonality,<sup>4</sup> age and sex,<sup>37,38</sup> and urban or rural living also contribute to interindividual differences in immune responses. This observation again highlights the fact that host genetics is shaped by strong selective pressure from environmental factors, mainly pathogens.

Several lines of evidence have demonstrated that TLRs have undergone different selection pressures across populations and have identified the *TLR10-TLR1-TLR6* locus as a target of positive selection in contemporary European populations.<sup>39,40</sup> Indeed, it has been hypothesized that this locus may have been selected by the severe ancient and medieval plague epidemics in Europe, as this locus modulates the immune response to *Yersinia pestis*.<sup>41</sup> Our analysis evaluating LD patterns in this locus is consistent with these previous findings. We observed much longer blocks of LD in the Europeans as compared to the Africans, a typical characteristic of a region undergoing natural selection.<sup>42</sup> Another typical example is *TLR6*, and variants in this gene have been shown to exhibit differential allele frequencies between European populations and other populations.<sup>43</sup> Interestingly, the variation in the *TLR10-TLR1-TLR6* locus has been reported to be strongly influenced in Europeans by introgression from Neanderthals, thus showing the impact that archaic humans had on immune responses of modern human populations.<sup>40</sup> As practically no introgression of Neanderthals and Denisovans is found in African populations, it becomes less surprising the lack of impact of variation in this locus on the immune response of African populations. Our observation of lack of impact of the *TLR10-TLR1-TLR6* locus on cytokine responses in the African population strengthens the possibility that this locus has putatively undergone adaptive evolution exclusively in Europe.

In the same line of comparison of molecular phenotypes across diverse ancestries, a study systematically assessing allelic variants in cytokine genes showed that African American women were more likely to harbor allelic variants known to upregulate pro-inflammatory cytokines compared to women of European ancestry.<sup>44</sup> Also, on the basis of genotype-expression data, 9.3% of genes expressed in macrophages exhibit ancestry-related variances in the regulatory response to infection, and African ancestry typically predicts stronger inflammation than European ancestry.<sup>9</sup> Our pathway enrichment analyses based on

genes mapped to cQTLs also showed similar results: contrary to European samples, cQTL genes from the Tanzanian cohort were significantly enriched in the interferon pathway. This is interesting because, using the blood transcriptional data from the same cohort, we recently observed significant upregulation of interferon genes among urban residents of Tanzania.<sup>19</sup> In line with this, Manry et al. (2011) have also shown significant differences in the number of polymorphisms within the population between the genes encoding the various IFNs and their receptors, in which the highest levels of nucleotide diversity within interferon genes were observed in African populations.<sup>45</sup> Interestingly, Quach et al. (2016) have shown that virus-specific eQTLs in Europeans are mostly population specific and Neanderthals introduced regulatory variants into European genomes.<sup>33</sup> Given the lack of Neanderthal variants in African genomes and the possibility that people in Tanzania are much more exposed to *Mycobacterium tuberculosis*<sup>46</sup> than the people living in the Netherlands, it is tempting to speculate that the natural selection had independently led to an increased interferon response in populations of Africa and is partly driven by host genetic variations. Whether this enhanced interferon response in the Tanzanian population influences risk for COVID-19 remains to be tested.

We also acknowledge some limitations with our study. First, using only 300 samples in our discovery cohort, we are unable to recruit a sufficiently large independent cohort of Tanzanian volunteers for replication analysis. Second, contrary to the Tanzania cohort, immune cell counts data used for correcting cytokines abundance prior to cQTL mapping were only available for the European cohort (500FG). Therefore, it is possible that some of the cQTLs are false positive associations and further validation is required. Third, a number of recent studies have identified the inability of the currently available SNP arrays to capture or assign non-European variants and haplotypes accurately.<sup>47,48</sup> We therefore reasoned that some critical variants in the Tanzanians might be missed, especially untagged variants with different frequencies and rare variants. Therefore, whole-genome sequence data from diverse populations are very much needed to capture population-specific and rare variants to study their impact of immune function.

Nevertheless, this study offers unique contribution to studies comprehensively investigating the genetics of cytokines response variability in non-European ancestry, most importantly leveraging a large variety of pathogens for inducing cytokine responses. Follow-up studies in larger cohorts are required to validate our findings and also to discover more genetic variants with less strong effects.

## Conclusions

Our study shows that at both SNPs and immunological pathway levels, the cytokine production capacity upon diverse stimulations is modulated in a population-specific manner. We also demonstrate that the *TLR1-TLR6-TLR10* locus significantly affects cytokine production capacity

only in individuals of European descent. Given the dissimilarity of genetic determinants regulating induced cytokine responses in both populations, we argue for the inclusion of under-represented populations in genomics research to make new discoveries in health and disease disparities seen among individuals and populations.

## Data and code availability

Summary statistics of the association results have been deposited to the EBI GWAS Catalog under the accession number GCST90093065. All other data and code underlying the findings are also available on reasonable request from the corresponding author.

## Supplemental information

Supplemental information can be found online at <https://doi.org/10.1016/j.ajhg.2022.01.014>.

## Acknowledgments

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., and A.V.; ZonMw (the Netherlands Organization for Health Research and Development) awarded to M.G.N., Q.d.M., and A.V.; and 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. Hypatia tenure track grant from Radboud UMC to V.K. The authors are grateful to all volunteers from the 300TZFG and 500FG cohorts without whom this research would not have been possible for participation in the studies.

## Declaration of interests

The authors declare no competing interests.

Received: July 14, 2021

Accepted: January 24, 2022

Published: February 14, 2022

## Web resources

Ensembl, <https://www.ensembl.org>

HaploReg, <http://pubs.broadinstitute.org/>

Human Functional Genomics Project, <http://www.humanfunctionalgenomics.org/site/>

LDlink, <https://ldlink.nci.nih.gov>

LocusZoom, <http://locuszoom.org>

Online Mendelian Inheritance in Man, <https://www.omim.org/>

## References

1. Akdis, M., Burgler, S., Cramer, R., Eiwegger, T., Fujita, H., Gomez, E., Klunker, S., Meyer, N., O'Mahony, L., Palomares, O., et al. (2011). Interleukins, from 1 to 37, and interferon- $\gamma$ : receptors, functions, and roles in diseases. *J. Allergy Clin.*

- Immunol. 127, 701–21.e1, 70. <https://doi.org/10.1016/j.jaci.2010.11.050>.
2. Vignali, D.A.A., and Kuchroo, V.K. (2012). IL-12 family cytokines: immunological playmakers. *Nat. Immunol.* 13, 722–728. <https://doi.org/10.1038/ni.2366>.
  3. Labuda, L.A., de Jong, S.E., Meurs, L., Amoah, A.S., Mbow, M., Ateba-Ngoa, U., van der Ham, A.J., Knulst, A.C., Yazdanbakhsh, M., and Adegnik, A.A. (2014). Differences in innate cytokine responses between European and African children. *PLoS ONE* 9, e95241. <https://doi.org/10.1371/journal.pone.0095241>.
  4. Ter Horst, R., Jaeger, M., Smeekens, S.P., Oosting, M., Swertz, M.A., Li, Y., Kumar, V., Diavatopoulos, D.A., Jansen, A.F.M., Lemmers, H., et al. (2016). Host and Environmental Factors Influencing Individual Human Cytokine Responses. *Cell* 167, 1111–1124.e13. <https://doi.org/10.1016/j.cell.2016.10.018>.
  5. Li, Y., Oosting, M., Deelen, P., Ricaño-Ponce, I., Smeekens, S., Jaeger, M., Matzaraki, V., Swertz, M.A., Xavier, R.J., Franke, L., et al. (2016). Inter-individual variability and genetic influences on cytokine responses to bacteria and fungi. *Nat. Med.* 22, 952–960. <https://doi.org/10.1038/nm.4139>.
  6. Li, Y., Oosting, M., Smeekens, S.P., Jaeger, M., Aguirre-Gamboa, R., Le, K.T.T., Deelen, P., Ricaño-Ponce, I., Schoffelen, T., Jansen, A.F.M., et al. (2016). A Functional Genomics Approach to Understand Variation in Cytokine Production in Humans. *Cell* 167, 1099–1110.e14. <https://doi.org/10.1016/j.cell.2016.10.017>.
  7. Wurfel, M.M., Gordon, A.C., Holden, T.D., Radella, F., Strout, J., Kajikawa, O., Ruzinski, J.T., Rona, G., Black, R.A., Stratton, S., et al. (2008). Toll-like receptor 1 polymorphisms affect innate immune responses and outcomes in sepsis. *Am. J. Respir. Crit. Care Med.* 178, 710–720.
  8. Mikacenic, C., Reiner, A.P., Holden, T.D., Nickerson, D.A., and Wurfel, M.M. (2013). Variation in the TLR10/TLR1/TLR6 locus is the major genetic determinant of interindividual difference in TLR1/2-mediated responses. *Genes Immun.* 14, 52–57. <https://doi.org/10.1038/gene.2012.53>.
  9. Nédélec, Y., Sanz, J., Baharian, G., Szpiech, Z.A., Pacis, A., Dumaine, A., Grenier, J.C., Freiman, A., Sams, A.J., Hebert, S., et al. (2016). Genetic Ancestry and Natural Selection Drive Population Differences in Immune Responses to Pathogens. *Cell* 167, 657–669.e21. <https://doi.org/10.1016/j.cell.2016.09.025>.
  10. Retshabile, G., Mlotshwa, B.C., Williams, L., Mwesigwa, S., Mboowa, G., Huang, Z., Rustagi, N., Swaminathan, S., Katarigira, E., Kyobe, S., et al. (2018). Whole-Exome Sequencing Reveals Uncaptured Variation and Distinct Ancestry in the Southern African Population of Botswana. *Am. J. Hum. Genet.* 102, 731–743. <https://doi.org/10.1016/j.ajhg.2018.03.010>.
  11. Magalhães, T.R., Casey, J.P., Conroy, J., Regan, R., Fitzpatrick, D.J., Shah, N., Sobral, J., and Ennis, S. (2012). HGDP and HapMap analysis by Ancestry Mapper reveals local and global population relationships. *PLoS ONE* 7, e49438. <https://doi.org/10.1371/journal.pone.0049438>.
  12. Sirugo, G., Williams, S.M., and Tishkoff, S.A. (2019). The Missing Diversity in Human Genetic Studies. *Cell* 177, 26–31. <https://doi.org/10.1016/j.cell.2019.02.048>.
  13. Bentley, A.R., Callier, S.L., and Rotimi, C.N. (2020). Evaluating the promise of inclusion of African ancestry populations in genomics. *NPJ Genom. Med.* 5, 5. <https://doi.org/10.1038/s41525-019-0111-x>.
  14. Wang, L., Pittman, K.J., Barker, J.R., Salinas, R.E., Stanaway, I.B., Williams, G.D., Carroll, R.J., Balmat, T., Ingham, A., Gopalakrishnan, A.M., et al. (2018). An Atlas of Genetic Variation Linking Pathogen-Induced Cellular Traits to Human Disease. *Cell Host Microbe* 24, 308–323.e6.
  15. Karimuribo, E.D., Mboera, L.E., Mbugi, E., Simba, A., Kivaria, F.M., Mmbuji, P., and Rweyemamu, M.M. (2011). Are we prepared for emerging and re-emerging diseases? Experience and lessons from epidemics that occurred in Tanzania during the last five decades. *Tanzan. J. Health Res.* 13 (5, Suppl 1), 387–398.
  16. Kumburu, H.H., Sonda, T., Mmbaga, B.T., Alifrangis, M., Lund, O., Kibiki, G., and Aarestrup, F.M. (2017). Patterns of infections, aetiological agents and antimicrobial resistance at a tertiary care hospital in northern Tanzania. *Trop. Med. Int. Heal.* 22, 454–464.
  17. Netea, M.G., Wijmenga, C., and O'Neill, L.A.J. (2012). Genetic variation in Toll-like receptors and disease susceptibility. *Nat. Immunol.* 13, 535–542.
  18. Karlsson, E.K., Kwiatkowski, D.P., and Sabeti, P.C. (2014). Natural selection and infectious disease in human populations. *Nat. Rev. Genet.* 15, 379–393.
  19. Temba, G.S., Kullaya, V., Pecht, T., Mmbaga, B.T., Aschenbrenner, A.C., Ulas, T., Kibiki, G., Lyamuya, F., Boahen, C.K., Kumar, V., et al. (2021). Urban living in healthy Tanzanians is associated with an inflammatory status driven by dietary and metabolic changes. *Nat. Immunol.* 22, 287–300. <https://doi.org/10.1038/s41590-021-00867-8>.
  20. Aulchenko, Y.S., Ripke, S., Isaacs, A., and van Duijn, C.M. (2007). GenABEL: an R library for genome-wide association analysis. *Bioinformatics* 23, 1294–1296. <https://doi.org/10.1093/bioinformatics/btm108>.
  21. Shah, T.S., Liu, J.Z., Floyd, J.A., Morris, J.A., Wirth, N., Barrett, J.C., and Anderson, C.A. (2012). optiCall: a robust genotype-calling algorithm for rare, low-frequency and common variants. *Bioinformatics* 28, 1598–1603. <https://doi.org/10.1093/bioinformatics/bts180>.
  22. Deelen, P., Bonder, M.J., van der Velde, K.J., Westra, H.J., Winder, E., Hendriksen, D., Franke, L., and Swertz, M.A. (2014). Genotype harmonizer: automatic strand alignment and format conversion for genotype data integration. *BMC Res. Notes* 7, 901. <https://doi.org/10.1186/1756-0500-7-901>.
  23. Das, S., Forer, L., Schönherr, S., Sidore, C., Locke, A.E., Kwong, A., Vrieze, S.I., Chew, E.Y., Levy, S., McGue, M., et al. (2016). Next-generation genotype imputation service and methods. *Nat. Genet.* 48, 1284–1287. <https://doi.org/10.1038/ng.3656>.
  24. Shabalina, A.A. (2012). Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* 28, 1353–1358. <https://doi.org/10.1093/bioinformatics/bts163>.
  25. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. <https://doi.org/10.1086/519795>.
  26. Willer, C.J., Li, Y., and Abecasis, G.R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 26, 2190–2191. <https://doi.org/10.1093/bioinformatics/btq340>.
  27. Shin, J.H., Blay, S., McNeney, B., and Graham, J. (2006). LDheatmap: An R function for graphical display of pairwise linkage disequilibria between single nucleotide polymorphisms. *J. Stat. Softw.* 16, 1–9.

28. Team, R.C.R. (2016). *A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing).
29. Cerillo, G., Rees, A., Manchanda, N., Reilly, C., Brogan, I., White, A., and Needham, M. (1998). The oestrogen receptor regulates NFkappaB and AP-1 activity in a cell-specific manner. *J. Steroid Biochem. Mol. Biol.* *67*, 79–88. [https://doi.org/10.1016/S0960-0760\(98\)00078-8](https://doi.org/10.1016/S0960-0760(98)00078-8).
30. Hug, H., Mohajeri, M.H., and La Fata, G. (2018). Toll-like receptors: Regulators of the immune response in the human gut. *Nutrients* *10*, E203. <https://doi.org/10.3390/nu10020203>.
31. Areal, H., Abrantes, J., and Esteves, P.J. (2011). Signatures of positive selection in Toll-like receptor (TLR) genes in mammals. *BMC Evol. Biol.* *11*, 368. <https://doi.org/10.1186/1471-2148-11-368>.
32. Shang, L., Smith, J.A., Zhao, W., Kho, M., Turner, S.T., Mosley, T.H., Kardia, S.L.R., and Zhou, X. (2020). Genetic Architecture of Gene Expression in European and African Americans: An eQTL Mapping Study in GENOA. *Am. J. Hum. Genet.* *106*, 496–512. <https://doi.org/10.1016/j.ajhg.2020.03.002>.
33. Quach, H., Rotival, M., Pothlichet, J., Loh, Y.E., Dannemann, M., Zidane, N., Laval, G., Patin, E., Harmant, C., Lopez, M., et al. (2016). Genetic Adaptation and Neandertal Admixture Shaped the Immune System of Human Populations. *Cell* *167*, 643–656.e17. <https://doi.org/10.1016/j.cell.2016.09.024>.
34. Kwok, A.J., Mentzer, A., and Knight, J.C. (2020). Host genetics and infectious disease: new tools, insights and translational opportunities. *Nat. Rev. Genet.* *22*, 137–153. <https://doi.org/10.1038/s41576-020-00297-6>.
35. Willing, E.M., Dreyer, C., and van Oosterhout, C. (2012). Estimates of genetic differentiation measured by F(ST) do not necessarily require large sample sizes when using many SNP markers. *PLoS ONE* *7*, e42649.
36. Schirmer, M., Smeekens, S.P., Vlamakis, H., Jaeger, M., Oosting, M., Franzosa, E.A., Ter Horst, R., Jansen, T., Jacobs, L., Bonder, M.J., et al. (2016). Linking the Human Gut Microbiome to Inflammatory Cytokine Production Capacity. *Cell* *167*, 1125–1136.e8. <https://doi.org/10.1016/j.cell.2016.10.020>.
37. Patin, E., Hasan, M., Bergstedt, J., Rouilly, V., Libri, V., Urrutia, A., Alanio, C., Scepanovic, P., Hammer, C., Jönsson, F., et al. (2018). Natural variation in the parameters of innate immune cells is preferentially driven by genetic factors. *Nat. Immunol.* *19*, 302–314. <https://doi.org/10.1038/s41590-018-0049-7>.
38. Bakker, O.B., Aguirre-Gamboa, R., Sanna, S., Oosting, M., Smeekens, S.P., Jaeger, M., Zorro, M., Vösa, U., Withoff, S., Netea-Maier, R.T., et al. (2018). Integration of multi-omics data and deep phenotyping enables prediction of cytokine responses. *Nat. Immunol.* *19*, 776–786. <https://doi.org/10.1038/s41590-018-0121-3>.
39. Barreiro, L.B., Ben-Ali, M., Quach, H., Laval, G., Patin, E., Pickrell, J.K., Bouchier, C., Tichit, M., Neyrolles, O., Gicquel, B., et al. (2009). Evolutionary dynamics of human Toll-like receptors and their different contributions to host defense. *PLoS Genet.* *5*, e1000562. <https://doi.org/10.1371/journal.pgen.1000562>.
40. Dannemann, M., Andrés, A.M., and Kelso, J. (2016). Introgression of Neandertal- and Denisovan-like Haplotypes Contributes to Adaptive Variation in Human Toll-like Receptors. *Am. J. Hum. Genet.* *98*, 22–33. <https://doi.org/10.1016/j.ajhg.2015.11.015>.
41. Laayouni, H., Oosting, M., Luisi, P., Ioana, M., Alonso, S., Ricaño-Ponce, I., Trynka, G., Zhernakova, A., Plantinga, T.S., Cheng, S.C., et al. (2014). Convergent evolution in European and Roma populations reveals pressure exerted by plague on Toll-like receptors. *Proc. Natl. Acad. Sci. USA* *111*, 2668–2673.
42. Lucek, K., and Willi, Y. (2021). Drivers of linkage disequilibrium across a species' geographic range. *PLoS Genet.* *17*, e1009477.
43. Pickrell, J.K., Coop, G., Novembre, J., Kudaravalli, S., Li, J.Z., Absher, D., Srinivasan, B.S., Barsh, G.S., Myers, R.M., Feldman, M.W., and Pritchard, J.K. (2009). Signals of recent positive selection in a worldwide sample of human populations. *Genome Res.* *19*, 826–837. <https://doi.org/10.1101/gr.087577.108>.
44. Ness, R.B., Haggerty, C.L., Harger, G., and Ferrell, R. (2004). Differential distribution of allelic variants in cytokine genes among African Americans and White Americans. *Am. J. Epidemiol.* *160*, 1033–1038. <https://doi.org/10.1093/aje/kwh325>.
45. Manry, J., Laval, G., Patin, E., Fornarino, S., Itan, Y., Fumagalli, M., Sironi, M., Tichit, M., Bouchier, C., Casanova, J.L., et al. (2011). Evolutionary genetic dissection of human interferons. *J. Exp. Med.* *208*, 2747–2759.
46. Sheriff, F.G., Manji, K.P., Manji, M.P., Chagani, M.M., Mpenbeni, R.M., Jusabani, A.M., Alwani, Z.R., and Karimjee, T.S. (2010). Latent tuberculosis among pregnant mothers in a resource poor setting in Northern Tanzania: a cross-sectional study. *BMC Infect. Dis.* *10*, 52.
47. Lankheet, I., Vicente, M., Barbieri, C., and Schlebusch, C. (2021). The performance of common SNP arrays in assigning African mitochondrial haplogroups. *BMC Genom. Data* *22*, 43.
48. Johnston, H.R., Hu, Y.J., Gao, J., O'Connor, T.D., Abecasis, G.R., Wojcik, G.L., Gignoux, C.R., Gourraud, P.A., Lizee, A., Hansen, M., et al.; CAAPA Consortium (2017). Identifying tagging SNPs for African specific genetic variation from the African Diaspora Genome. *Sci. Rep.* *7*, 46398.



**The American Journal of Human Genetics, Volume 109**

**Supplemental information**

**A functional genomics approach in Tanzanian  
population identifies distinct genetic regulators of  
cytokine production compared to European population**

**Collins K. Boahen, Godfrey S. Temba, Vesla I. Kullaya, Vasiliki Matzaraki, Leo A.B. Joosten, Gibson Kibiki, Blandina T. Mmbaga, Andre van der Ven, Quirijn de Mast, Mihai G. Netea, and Vinod Kumar**

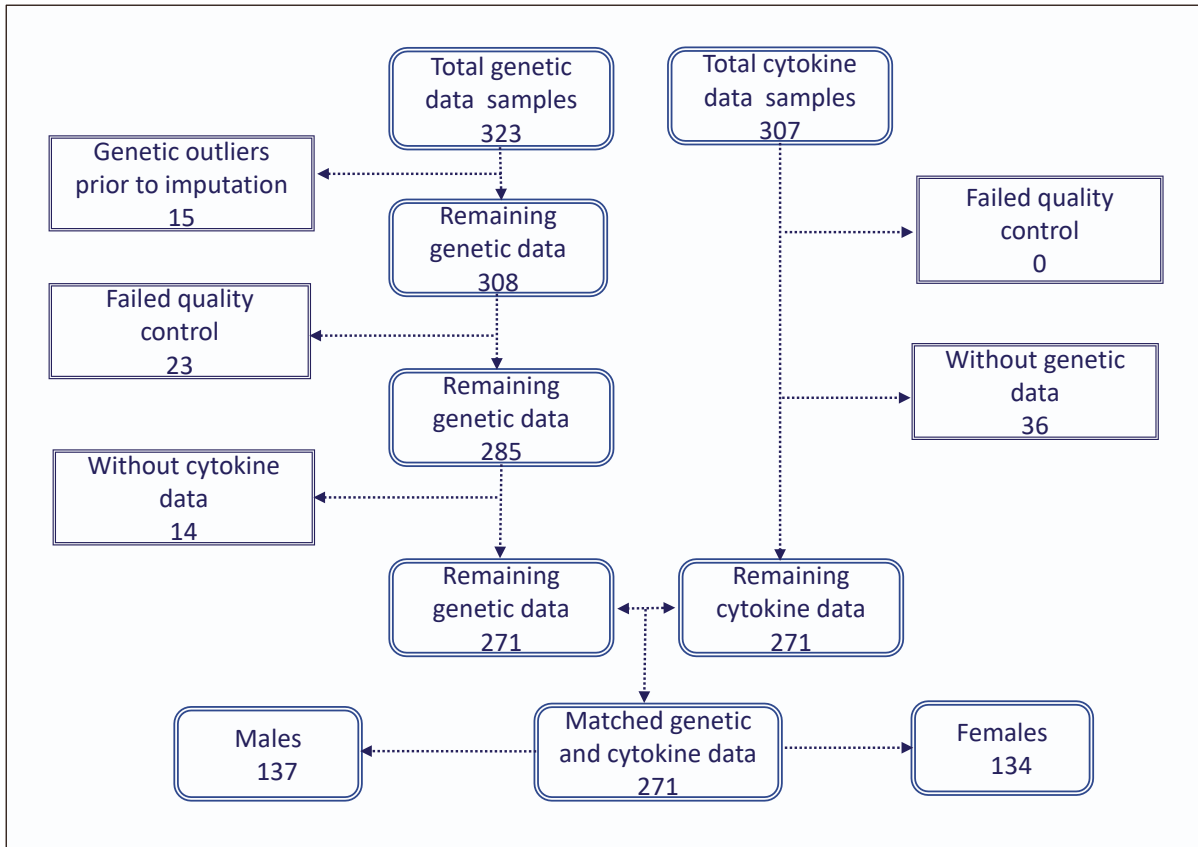


Figure S1. Schematic overview of Tanzania study cohort quality control.

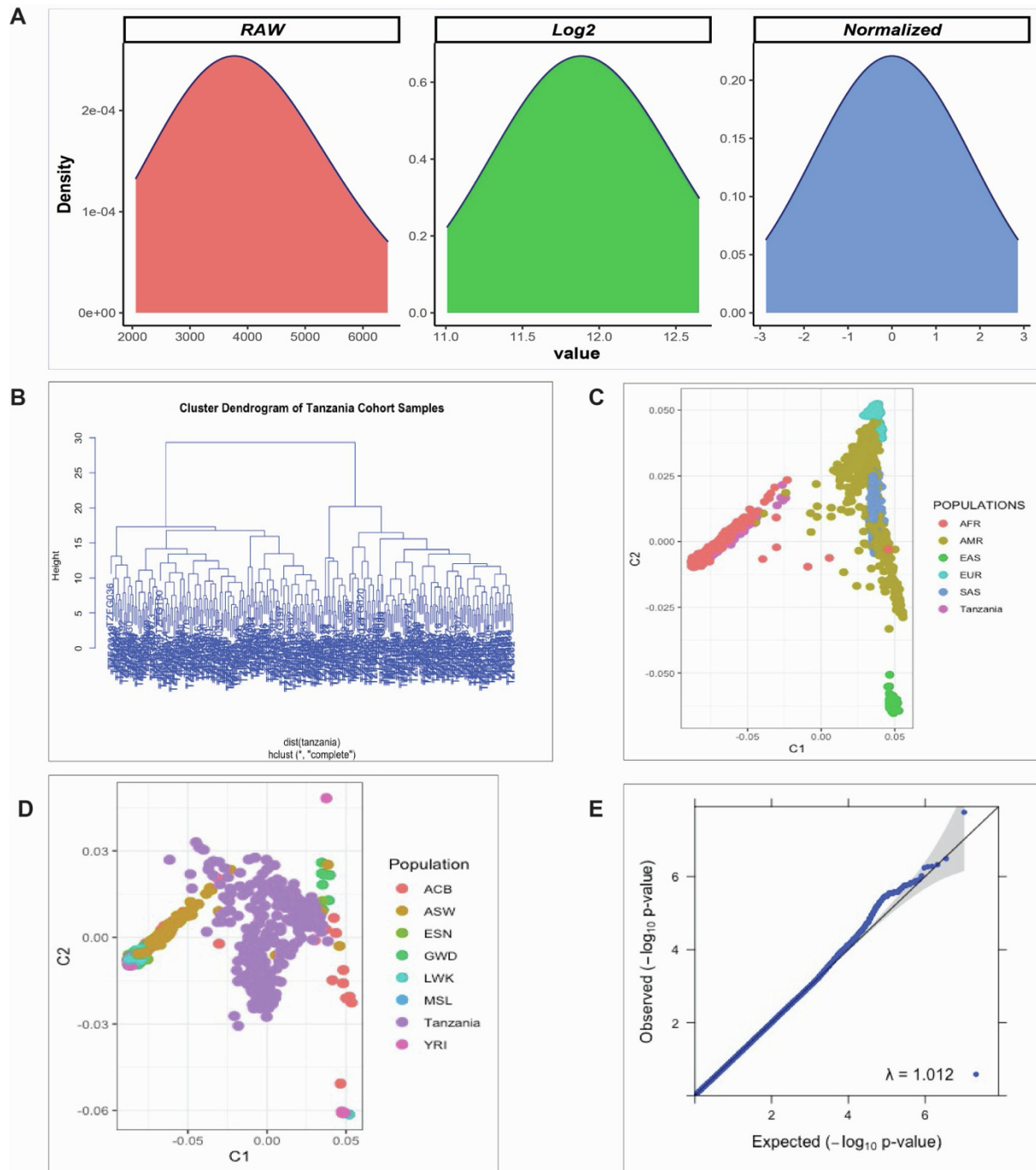


Figure S2. Quality control of genetic and cytokine data.

(A) Density plots depicting the distributions of *S. enteritidis* induced-IL-1 $\beta$  concentrations. The density plot is faceted as follows: raw (RAW), Log2 transformed (Log2), and inverse-ranked normalization of IL-1 $\beta$  distributions respectively. (B) Dendrogram visualization of unsupervised hierarchical clustering of individuals used for cQTL mapping. The “Euclidean” method in the “hclust” function was used for

computing distances. (C) Multi-dimensional scaling (MDS) plot of populations in the 1000 Genome project against individuals in Tanzania cohort. The first two MDS components (C1 and C2) are plotted against each other. Individuals representing various continental groups are color coded as shown in the color legend. The color legend populations in the 1000 Genome project are: AFR (Africans), AMR (American), EAS (East Asians), EUR (Europeans), and SAS (South Asians). Clustering of populations revealed genetically homogenous cluster of Tanzania cohort individuals with African ancestry individuals in the 1000 Genome project data. (D) MDS plot with Tanzanian data projected on only Africans (dominated by west Africans) in the 1000G data. (E) Quantile-Quantile (QQ) plot depicting association results of IL-1 $\beta$  after *S. enteritidis* stimulation. Lambda ( $\lambda$ ) value indicates lack of inflation of the test statistics.

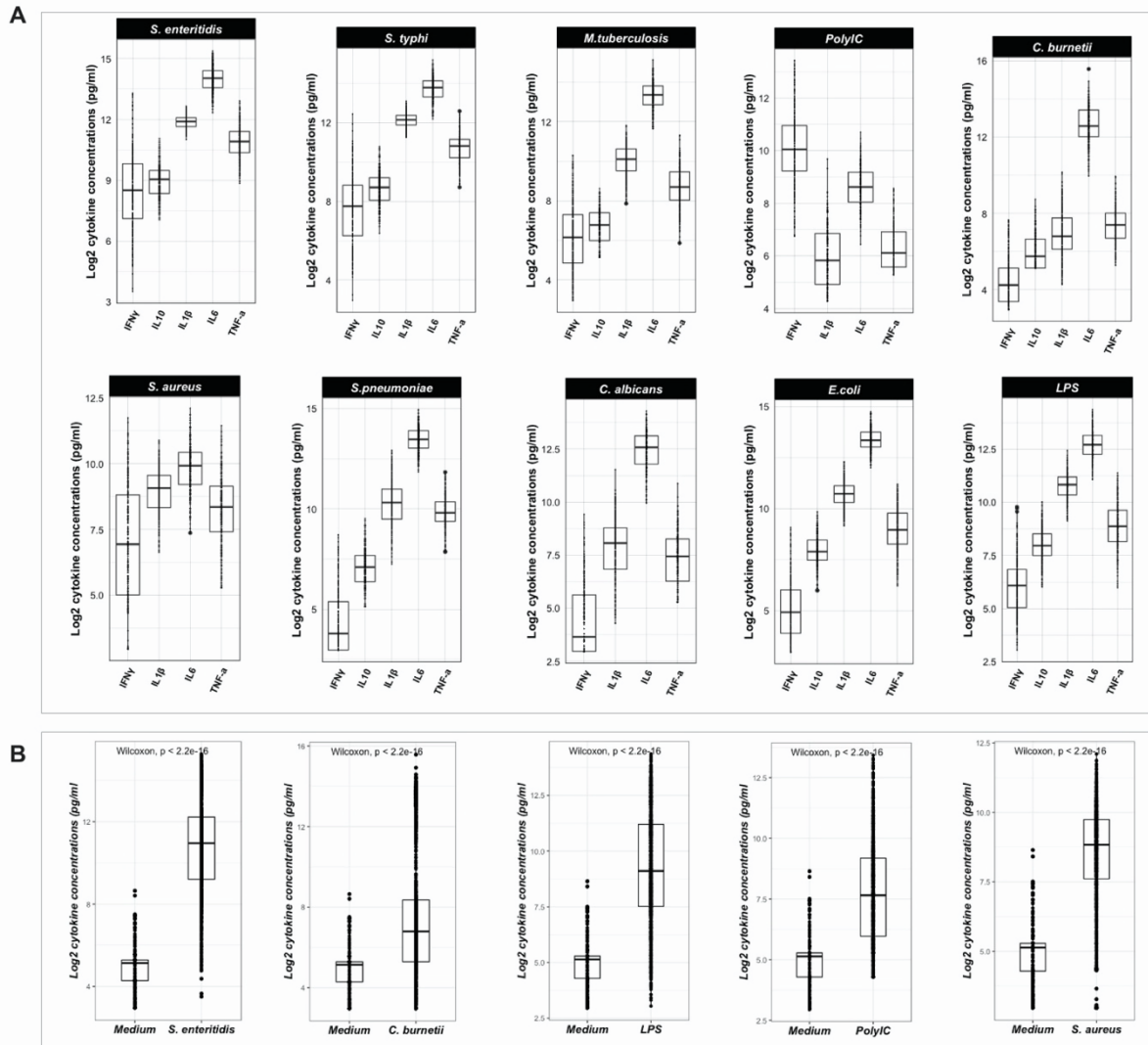


Figure S3: Interindividual variability of cytokine profiles upon stimulation. (A)

Boxplots indicating strong variation of IFN $\gamma$  responses compared to other cytokines in most stimulations. (B) Boxplot of Wilcoxon test depicting the level of significant differences of cytokines response upon stimulation compared to unstimulated conditions (Medium).

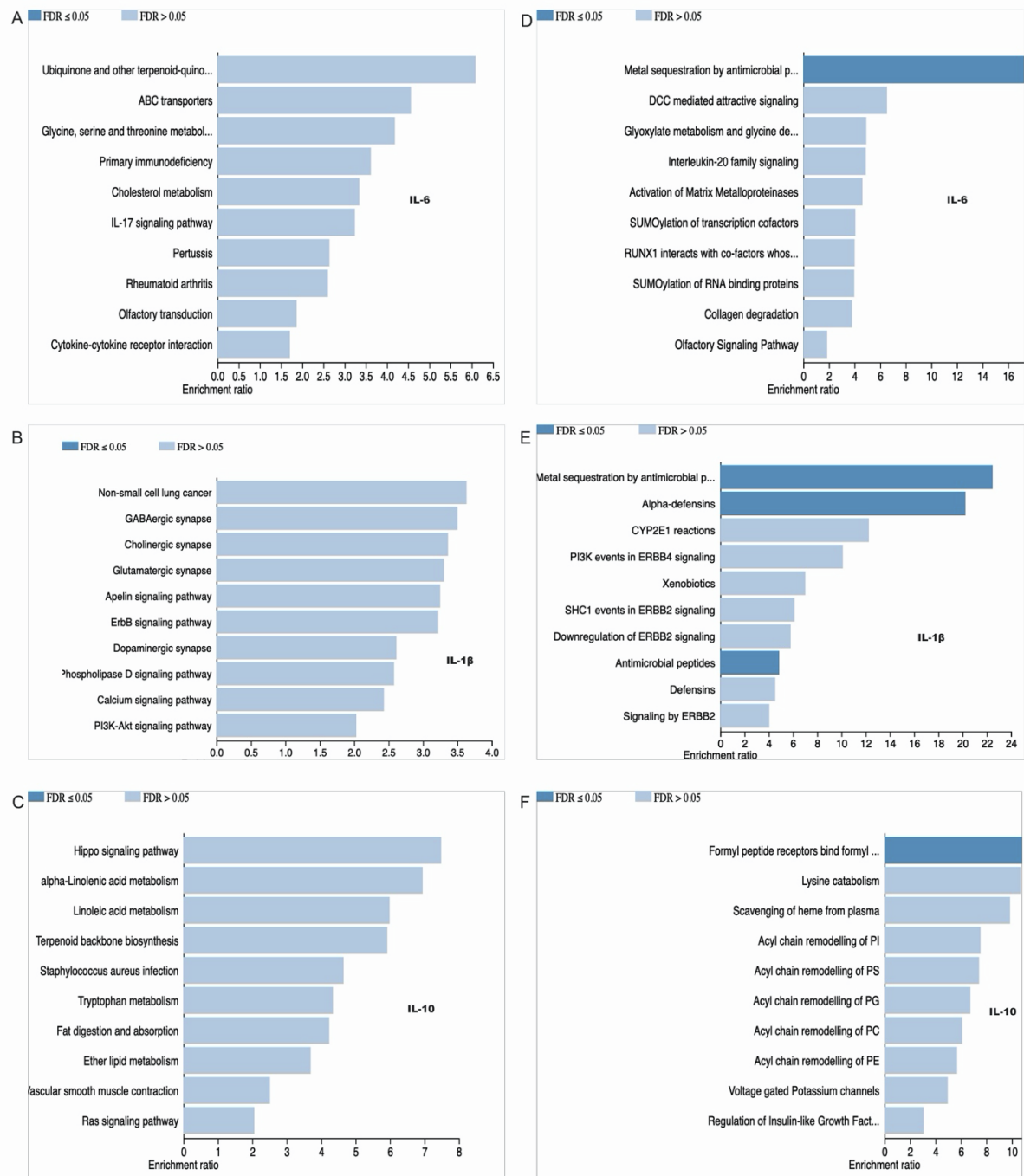


Figure S4. Bar chart representation of IL-6, IL-1β- and IL-10-based pathways.

(A , B and C) Top 10 enriched KEGG pathways for IL-6, IL-1β and IL-10 respectively.

(D , E and F) Top 10 enriched Reactome pathways for IL-6, IL-1β and IL-10 respectively.

Darker blue bars represent statistically significant pathways and lighter blue bars represent non-significant pathways after multiple testing. We set the false discovery rate (FDR)-adjusted P-value to 0.05 as significance level.

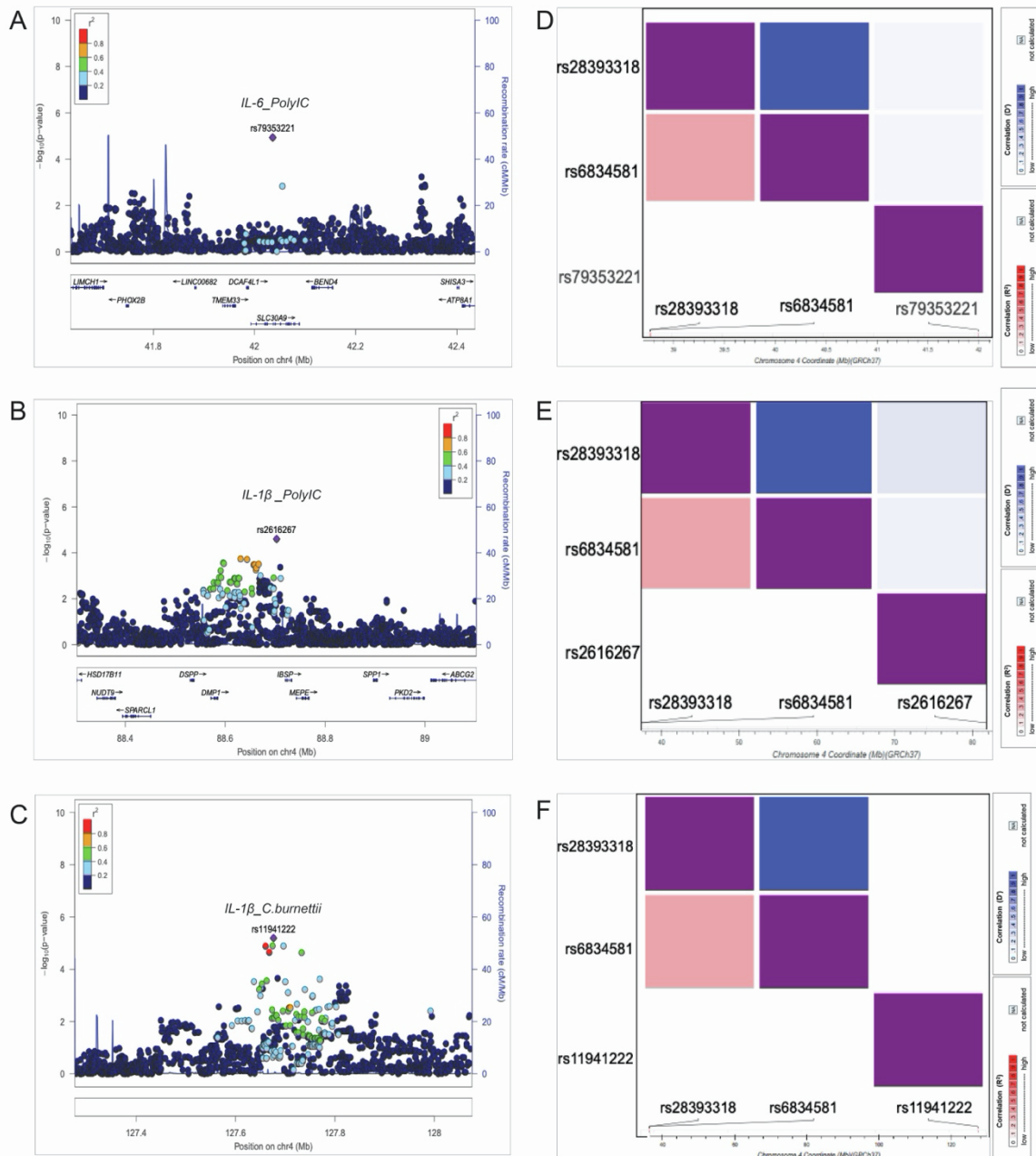


Figure S5. Different haplotypes regulate cytokine production between Africans and Europeans.

(A, B and C) Regional association plots of SNPs flanking genomic window of 400kb around top SNPs: rs79353221, rs2616267 and rs11941222 associated with IL-6 and IL-1 $\beta$  concentrations at *TLR10-TLR1-TLR6* locus. None of these variants are in linkage disequilibrium with the top variants identified in this locus in the European

population. (D, E and F) Heatmaps showing lack of correlation between lead SNPs in the African and the European dataset at the *TLR10-TLR1-TLR6* locus.