Genetic regulators of cytokine responses in children from West Africa: the impact of BCG vaccination

Collins K. Boahen*
Department of Internal Medicine and Radboud Institute of Molecular Life Sciences (RIMLS), Radboud University Medical Center, Nijmegen, 6525 HP, the Netherlands, Department of Internal Medicine and Radboud Center for Infectious Diseases (RCI), Radboud University Medical Center, Nijmegen, 6525 HP, the Netherlands

S.J.C.F.M. Moorlag*
Department of Internal Medicine and Radboud Institute of Molecular Life Sciences (RIMLS), Radboud University Medical Center, Nijmegen, 6525 HP, the Netherlands, Department of Internal Medicine and Radboud Center for Infectious Diseases (RCI), Radboud University Medical Center, Nijmegen, 6525 HP, the Netherlands

Kristoffer Jarlov Jensen
Odense Patient Data Explorative Network, University of Southern Denmark/Odense University Hospital, Odense, DK-5000, Denmark, Center for Clinical Research and Prevention, Frederiksberg and Bispebjerg Hospital, DK-2000 Frederiksberg, Denmark

Vasiliki Matzaraki
Department of Internal Medicine and Radboud Institute of Molecular Life Sciences (RIMLS), Radboud University Medical Center, Nijmegen, 6525 HP, the Netherlands, Department of Internal Medicine and Radboud Center for Infectious Diseases (RCI), Radboud University Medical Center, Nijmegen, 6525 HP, the Netherlands

Stephanie Fanucchi
Lemba Therapeutics, Nijmegen, 6525 GA, the Netherlands

Ivan Monteiro
Bandim Health Project, Indepth Network, Bissau, codex 1004, Guinea-Bissau

Charlotte de Bree
Department of Internal Medicine and Radboud Institute of Molecular Life Sciences (RIMLS), Radboud University Medical Center, Nijmegen, 6525 HP, the Netherlands, Department of Internal Medicine and Radboud Center for Infectious Diseases (RCI), Radboud University Medical Center, Nijmegen, 6525 HP, the Netherlands

Ezio T. Fok
Epigenomics & Single Cell Biophysics Group, Department of Cell Biology FNWI, Radboud Institute for Molecular Life Sciences (RIMLS), Nijmegen, 6525 HP, the Netherlands, Department of Human Genetics, Radboud University Medical Center, 6525 HP, the Netherlands

Musa Mhlanga
Epigenomics & Single Cell Biophysics Group, Department of Cell Biology FNWI, Radboud Institute for Molecular Life Sciences (RIMLS), Nijmegen, 6525 HP, the Netherlands, Department of Human Genetics, Radboud University Medical Center, 6525 HP, the Netherlands

Leo A.B Joosten
Department of Internal Medicine and Radboud Institute of Molecular Life Sciences (RIMLS), Radboud University Medical Center, Nijmegen, 6525 HP, the Netherlands, Department of Internal Medicine and Radboud Center for Infectious Diseases (RCI), Radboud University Medical Center, Nijmegen, 6525 HP, the Netherlands

Peter Aaby
Odense Patient Data Explorative Network, University of Southern Denmark/Odense University Hospital, Odense, DK-5000, Denmark Bandim Health Project, Statens Serum Institut, Copenhagen S, DK-2300, Denmark

Christine Stabell Benn
Odense Patient Data Explorative Network, University of Southern Denmark/Odense University Hospital, Odense, DK-5000, Denmark Danish Institute for Advanced Study, University of Southern Denmark, Odense, DK-5000, Denmark

Mihai G. Netea

Department of Internal Medicine and Radboud Institute of Molecular Life Sciences (RIMLS), Radboud University Medical Center, Nijmegen, 6525 HP, the Netherlands, Department of Internal Medicine and Radboud Center for Infectious Diseases (RCI), Radboud University Medical Center, Nijmegen, 6525 HP, the Netherlands Department for Immunology and Metabolism, Life and Medical Sciences Institute (LIMES), University of Bonn, Germany

Vinod Kumar (v.kumar@radboudumc.nl)

Department of Internal Medicine and Radboud Institute of Molecular Life Sciences (RIMLS), Radboud University Medical Center, Nijmegen, 6525 HP, the Netherlands, Department of Internal Medicine and Radboud Center for Infectious Diseases (RCI), Radboud University Medical Center, Nijmegen, 6525 HP, the Netherlands University of Groningen, University Medical Center Groningen, Department of Genetics, Groningen, 9700 RB, the Netherlands Nitte (Deemed to be University), Nitte University Centre for Science Education and Research (NUCSER), Medical Sciences Complex, Deralakatte, Mangalore, 575018, India

---

Research Article

Keywords: cytokines, genetic variation, Africa, immunogenomics, QTL, BCG

Posted Date: April 1st, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1502935/v1

Abstract

Genetic variation is a key factor influencing cytokine production capacity, but which genetic loci regulate cytokine production before and after vaccination, particularly in African population is unknown. Here, we aimed to identify single-nucleotide polymorphisms (SNPs) controlling cytokine responses (cQTLs) after microbial stimulation in infants of West-African ancestry, comprising of low-birth-weight neonates randomized to BCG vaccine-at-birth (intervention) or to the usual delayed BCG (control). Genome-wide cytokine QTL-mapping revealed 12 independent cQTLs, of which the *LINC01082-LINC00917* locus influenced more than half of the cytokine-stimulation pairs assessed. Furthermore, nine distinct cQTLs were found among infants randomized to BCG. Functional validation confirmed that several complement genes affect cytokine response after BCG vaccination. We observed limited overlap of common cQTLs between the West-African infants and cohorts of Western European individuals. These data reveal strong population-specific genetic effects on cytokine production and may indicate new opportunities for therapeutic intervention and vaccine development in African populations.

Introduction

Cytokines are key mediators of the immune response and inflammatory process and their regulation has direct consequences for susceptibility to many inflammatory and infectious diseases. Several functional genomic studies, conducted mainly in Western European populations, have established the significant role of both environmental and host genetic variation in affecting the inter-individual variability in cytokine production capacity to pathogens[1–3]. Identifying such factors not only helps in understanding molecular mechanisms that regulate cytokine production, but also in prioritizing the targets for the personalized treatment of immune-mediated and infectious diseases.

The genetic factors identified in the European populations are not easily translatable to African populations[4]. Africa is considered as the birthplace of modern humans and, from the limited genetic data available from the African populations[5], it is clear that it harbors the much higher genetic diversity than other continents. Lack of comprehensive genomics and genetics studies on Africans could hamper the development of a personalized treatment, which leads to population disparities in precision medicine. Given the high burden of infectious diseases in low-income countries, it is important that we understand the influences of genetic variation on human cytokine responses in African populations. This may provide novel insights into population-specific genetic regulation of cytokine responses, and lead to improvements in clinical care.

Routine vaccination that provides active acquired immunity to infectious diseases is estimated to save over 3 million lives annually[6]. Tuberculosis (TB) is a severe infectious disease: the World Health Organization (WHO) estimates that 1 million children below 15 years develop TB globally, with over 25% of TB deaths occurring in the African Continent. Vaccination with bacillus Calmette-Guérin (BCG) is an integral part of childhood immunization programs in countries with a high TB burden. Besides protection against mycobacterial infections, BCG has been shown to provide non-specific protection against child mortality, mainly due to protection against respiratory tract infections and neonatal sepsis[7–9]. This effect appears to be at least partially linked to the induction of functional and epigenetic reprogramming in myeloid cells and natural killer (NK) cells, a process termed trained immunity or innate immune memory[10]. For example, BCG vaccination of human healthy volunteers strongly enhances the production of pro-inflammatory cytokines, such as IL-1β, TNF and IL-6, by monocytes for up to 3 months after vaccination upon secondary *ex vivo* stimulation with unrelated pathogens[11]. In human experimental challenge models, the BCG-induced increase in cytokine production has been shown to be associated with lower viral load after yellow fever vaccine[12] and with reduced parasitemia in a subgroup of individuals after an experimental malaria infection[13]. Several studies have demonstrated that the host genetic variation has a strong influence on vaccination-induced immune responses, e.g. to influenza, Hepatitis B or measles vaccination[14–16]. However, which genetic variants and pathways affect cytokine responses to secondary infections upon BCG vaccination, particularly in the African population, remains largely unknown.

In this study, we investigated cytokine production at 4 weeks of age in a West-African (Guinea-Bissau) cohort consisting of more than 400 low weight neonates, randomized to receive either a BCG vaccination at birth (intervention group) or the usual delayed BCG vaccine (control group)[17]. Cytokine production capacity was assessed after 24 hr by whole blood stimulations with microbial and non-microbial stimuli[18]. Next, we performed genome-wide SNP genotyping followed by QTL mapping to prioritize important genes and pathways that regulate cytokine responses and assessed genetic variants that may specifically affect BCG-induced immune

responses. In addition, we compared the identified cytokine QTLs (cQTLs) with cQTLs found in two cohorts of individuals of European ancestry.

Results

High inter-individual variation in cytokine production capacity upon stimulation in both control and BCG vaccinated groups

In order to test whether stimulation has any effect on inter-individual variation in cytokine production capacity, we firstly measured the production of seven cytokines upon 24-hour whole-blood stimulations with either control culture medium or five different stimuli: PMA, PPD, LPS, CLO75 and PAM (see **Fig. 1**). Details of these stimulations and concentrations used are described in the Methods section. In general, stimulation resulted in a strong increase in overall average cytokine production to all stimulants compared to cell culture medium controls (**Fig. 2A-BS**). In both control and BCG groups, stimulation with the TLR-8 agonist CLO75, resulted in the strongest increase in cytokine responses, whereas the least increase was observed for the mycobacterial specific stimulus PPD. Stratified analysis on sex did not show any significant differences in cytokine responses between male and female, both in the control and BCG groups (data not shown). There was a strong inter-individual variability in cytokine responses upon stimulation, arranged in descending order of median values per stimuli (**Fig. 2C-D**). These results indicate the role of individual-specific factors in regulating the cytokine production upon stimulation.

Identification of genetic variants affecting cytokine production capacity in the unvaccinated children

In order to better understand the factors that determine variation in immune responses in the unvaccinated pediatric population, we next assessed the impact of genetic variation on cytokine production in response to *in vitro* stimulation. First, we performed genome-wide SNP cQTL mapping to identify genetic loci that regulate cytokine responses in the group of infants that had not received BCG. After adjusting for the effect of age and sex, we identified a total of 13 genome-wide significant ($P < 5x10^{-8}$) lead SNPs in 12 independent loci associated with cytokine production capacity upon stimulation with different stimuli (**Fig. 3A**). **Table 1** summarizes the results of the genome-wide significant loci identified. Of the 12 loci, two cQTLs were associated with IL-6, one with IL-1β, two with TNF-α, one with IL-5, three with IL-17, and three with IFN-γ production. The SNP rs11918168 on chromosome 3 showed strongest association (2.4×10^{-9}) with interleukin-6 (IL-6) concentrations after LPS stimulation. We further observed that these genome-wide significant cQTLs exhibited strong a pleiotropic effect (**Fig. 3B**), indicating shared genetic regulatory pathways. For example, two independent loci, rs62051095 and rs11643992 both located on chromosome 16, showed multiple associations (P values varying between 5×10^{-2} to 5×10^{-6}), with approximately half of the total cytokine stimulation pairs used in this study.

To better understand the functional consequences of cQTLs variants, we mapped the identified genome-wide significant signals and their proxies (r^2 ³ 0.8) against known gene structures using Ensemble Variant Effect Predictor[19]. Interestingly, the vast majority of the cQTLs, as high as 63% identified in the unvaccinated group, were predominant in the intronic regions. Functional category such as intergenic prediction was only 2% **(Fig. S1A).** Besides interrogating the functional consequences of cQTLs, we further examined whether the genome-wide significant hits and their proxies disrupt transcription factor binding sites (TFBS). The three top most significant transcription factors (TFs) identified, whose binding sites are located on genes containing the cQTL variants are *Hand1_Tcfe2a, YY1,* and *MZF1_5-13* **(Fig. S1B).** The identification of TF-binding variants is crucial to understand the molecular mechanism by which cQTLs variants affect cytokine responses.

Functional annotation of genome-wide significant cQTLs: the control group

In order to prioritize the potential causal genes at the 13 genome-wide significant cQTL loci, we applied two main approaches. First, we prioritized genes based on their proximity to the lead SNP using HaploReg v4.1[20]. Next, we assessed whether the genome-wide significant SNPs have been reported to correlate with gene expression levels in whole blood (eQTL analysis) by making use of publicly available eQTL datasets[21]. In total, we identified 12 unique potential causal genes at the 12 cQTL loci (**Table 1**). SNP rs11918168 on chromosome 3 showed the strongest association ($p= 2.4 \times 10^{-9}$), influencing IL-6 production capacity upon LPS stimulation. This SNP is located in the *CPNE4* gene, which encodes copine IV, a calcium-dependent membrane-binding protein.

Of the 12 genome-wide significant loci, 2 independent loci located on chromosome 16 significantly affected the production of IL-5, as well as the production of TNF-α upon PMA stimulation. Both loci are found near long noncoding RNAs (lncRNAs) (see **Fig. 3C and D** for regional association plots of these loci). Interestingly, recent studies have indicated lnRNAs as important regulators of cytokine gene expression[22].

Next, we investigated whether the genes located in the vicinity of the cQTLs within a window of 250kb are overrepresented for particular pathways, using the Web-based Gene Set analysis (WebGestalt) tool[23]. Pathway enrichment analysis revealed that the cQTLs were enriched for 'mismatch repair', 'linoleic acid metabolism' , 'autophagy' and 'interleukins'(**Fig. S2A-B** and **Supplementary Table 1**).

Distinct genetic variants affect cytokine production capacity in BCG-vaccinated children

A previous study investigating the differences in cytokine production between the BCG-vaccinated infants and controls in this cohort revealed significantly increased cytokine responses in the BCG group as compared to the control group[18], similar to findings in various other BCG vaccination studies[11,12,24]. However, little is known about the genetic variants and pathways influencing cytokine responses to secondary stimulation after BCG vaccination. Therefore, we next investigated the impact of genetic variation on cytokine production in the BCG-vaccinated group. In total, we identified 9 independent genome-wide significant loci affecting cytokine responses (**Fig. 4A and Table 2**). The strongest association signal was at rs13246275, located on chromosome 7, which was associated with TNF-α after PAM stimulation ($p = 2.66 \times 10^{-9}$). Some of the genome-wide significant loci also showed nominal association with other cytokines (**Fig. 4B**). For example, SNP rs202131186, located on chromosome 20, affected ten other cytokine stimulation pairs. This SNP is located near *LOC284788*, which belongs to the class of lncRNAs. As this locus significantly influences PPD-induced IFN-γ production ($p = 3.96 \times 10^{-8}$), this might suggest that this locus could be important for the regulation of the specific immune response against *M. tuberculosis* upon BCG vaccination. In addition to this locus, rs5887565 showed a significant association with PPD-induced TNF-α levels. This SNP is located in *PLXNA4*, which plays an important role in various processes involving the immune system, including the regulation of T-cell responses[25]. Both T-cells and TNF-α production play a prominent role in host defense against mycobacterial infections[26].

Interestingly, for two loci (rs3759416 and rs6834860) with cQTL effects in the BCG-vaccinated group of infants, eQTL analysis indicated strong eQTL effect from two loci on *C1RL* and *INPP4B* genes, respectively (regional association plots of these loci are shown in **Fig. 4 C-D**). *INPP4B* encodes the inositol polyphosphate 4-phosphatase type II and has been shown to be involved as a negative regulator of PI3K/AKT pathway[27] which is critical for the induction of trained immunity[12,28]. *C1RL* encodes for the Complement C1r subcomponent-like protein. Interestingly, *C1R* and *C1RL* genes are located next to each other, and the rs3759416 SNP shows opposite eQTL effects on these two genes. This association warrants further elucidation on the role of complement in trained immunity. As described previously, introns harbored 48% of the identified cQTLs in the BCG group and their proxies, contrary to just 2% of the variants located in the intergenic regions (**Fig. S1C**). The cQTLs identified in the BCG group also alter TFBS with the most significantly enriched ones being *RORA_2, Foxa2, NFYB, Tcf12,* and *FOXA1*(**Fig. S1D**).

To assess whether the genome-wide significant signals identified in the BCG-vaccinated infants overlap with cQTLs found in the infants who did not receive BCG, we intersected the nine independent genome-wide significant hits from the BCG vaccinated group with cQTLs results from the BCG unvaccinated group at a nominal p-value threshold of 0.05. Strikingly, none of these loci showed association with cQTLs in the BCG unvaccinated cohort. This might indicate that these nine variants specifically regulate cytokine responses modulated by BCG vaccination. In line with these findings, pathway analysis on the prioritized genes of the genome-wide significant cQTLs in the BCG-vaccinated group revealed different pathways affecting cytokine responses as compared to the pathways found in the control group (**Fig. S2 C-D Supplementary Table 2**). The T cell receptor signaling pathway, which plays a crucial role in immune function, emerged as the statistically most significant enriched pathway in the BCG-vaccinated group, with genes such as *PPP3R1*, *PTPN6*, *CD4*, *PAK4* and *NFKBI* involved. Among the other pathways, we found glycolysis, inositol phosphate metabolism and the JAK-STAT signaling as top pathways in BCG-vaccinated group. Intriguingly, these pathways have already been shown to be major regulatory pathways of BCG-induced trained immunity in adults[29]. These results might indicate that similar pathways are also important in the induction of BCG-induced trained immunity in infants in an African population.

Combined analyses across BCG-vaccinated and unvaccinated groups identified five independent genome-wide significant cQTLs

It is possible that some of the genetic variants affect cytokine production capacity in children independent of their vaccination status. Therefore, we aimed to identify such cQTLs. Interestingly, five independent genome-wide significant cQTLs were found that showed same direction of effect in both groups (**Table 3**). Among these five genetic variants, SNP rs4674520 influences PPD-induced IL-17 production. This SNP has been linked to changes in mRNA expression levels of *PPIL3, CFLAR and CASP10*. *CASP10* encodes for the apoptotic cysteine protease CASP10, which was shown to be differentially expressed in monocytes upon stimulation with pathogenic mycobacteria[30]. Furthermore, rs62313427, influencing CLO75-induced TNF-α has been linked to changes in expression levels of 3'-Phosphoadenosine 5'-phosphosulfate synthase 1 (PAPSS1). How these *cis-genes* regulate the abundance of cytokines warrants further mechanistic studies.

Enrichment of BCG-vaccination specific cQTLs at complement genes

Among the genome-wide significant cQTLs identified in the BCG group, one of the strongest signals was an intronic SNP rs3759416 residing in the *CLSTN3* locus flanked by *C1R* and *C1RL* complement genes **Fig. 4C**. SNP rs3759416 also showed strong eQTL effects on *C1R* and *C1RL* genes in this locus **Table 2**. Given the fact that genes annotated in the complement system do not function in isolation but interact with each other to activate the complement cascade, we hypothesized that several other complement genes might be involved in regulating cytokine responses upon BCG vaccination. To test this, we focused on 65 complement genes and performed competitive gene set enrichment analysis (see Methods). Interestingly, we observed significant enrichment of cQTLs identified in the BCG-vaccinated group at several complement genes **Fig. 5A**. In summary, we observed a clear enrichment of SNPs near *C1RL, C1R, C1S, C7, ITGAX, ITGAM,* and *C3* to be affecting cytokine production upon BCG vaccination **Fig. 5B**.
Interestingly, mining publicly available ChIP-Seq datasets revealed that SNP rs3759416 overlapped a primed enhancer region with evidence of interaction among the nearest gene and other neighbouring genes (**Fig. 5C**). We next set up functional experiments to elucidate the impact of complement system on cytokines by targeting *C1RL, C1R, CLSTN3* and *RBP5* with antisense oligonucleotides (ASOs) to reduce their expression in THP1 monocytes. We observed significant reduction in *IL1B* expression after LPS stimulation following depletion of CIR and C1RL (**Fig. 5D**).

Minimal overlap between African and European cQTLs

Next, we aimed to investigate whether similar genetic variants influence cytokine responses in the West-African cohort and the European population. To achieve this, we used as comparator a cohort of healthy adults of Western European ancestry (500FG cohort), in which genetic variants influencing pro-inflammatory cytokine production after bacterial, fungal and viral stimulations have been identified[31].

First, we compared the genome-wide significant cQTLs from the 500FG cohort (BCG unvaccinated samples) with the African BCG-unvaccinated group cQTLs to contrast the strength of association between the two cohorts. We performed this analysis for the same cytokine-SNP pairs that were common to both studies since majority of the stimulations used to induce cytokine responses differed between the two cohorts. There were six 500FG genome-wide significant loci comparable (index SNP present in both cohorts used for QTL mapping) with the BCG-unvaccinated African cohort. We tested for association of the 6 loci with all the 5 stimulations utilized in the present study, making a total of 30 comparisons. We found that only 3 of them showed nominal association with cytokine production in the African population, while 27 SNPs did not have a significant effect on cytokine production in the BCG-unvaccinated African cohort (**Supplementary Table 3**). For example, SNP rs28393318 at *TLR10-TLR1-TLR6* locus on chromosome 4 showed the strongest association[31] with multiple cytokine-stimulation pairs in the European cohort, and this locus is also under positive selection[32]. However, the SNP rs28393318 in the African cohort (BCG-unvaccinated group) showed only nominal association ($P = 1.4\times10^{-3}$) with IL-1β production (**Fig. 6A).** Interestingly, we observed much stronger association ($P = 8.2\times10^{-5}$) between IL-1β and a different SNP (rs4832789) in this locus (**Fig. 6B**). These findings suggest that different genetic variants affect the regulation of cytokine production capacity in the African population as compared to the European population.

Next, we sought to investigate the proportion of cQTLs common to both cohorts using the same cytokine-stimulation pairs. Both the African (Guinea-Bissau cohorts) and European (500FG and 300BCG) cohorts have *lipopolysaccharide* (LPS) as a common stimulus in the studies. To study the extent of overlap, we intersected cQTL results using a nominal p-value cut-off of 5×10^{-2} and suggestive association p-value of 10^{-5}. With a p-value < 5×10^{-2} threshold, initial overlap comparison was performed in the African BCG-unvaccinated group and the 500FG unvaccinated group, with IL-1β being the common cytokine. As expected, we observed a minimal percentage overlap of 1.5%. Venn diagram depicting this shared and unique genetic variants is shown in (**Fig. 6C**).

Also, in the 300 BCG and the African BCG-vaccinated cohorts, where IL-6 and TNF-a were the two common cytokines measured, similar observations were made. The percentage overlaps of the investigated SNPs were 1.5% and 1.6% emanating from IL-6 and TNF-a overlap analysis respectively (**Fig. 6D and E**). We did not observe any overlapping variants that show suggestive association of p-value < 10^{-5} (data not shown).

Discussion

As most genomic studies assessing cytokine responses have been performed in populations of European descent, characterization of the architecture of cytokine responses and the impact of genetic variation on these responses in an African population is warranted. However, to the best of our knowledge, this aspect is yet to be thoroughly investigated.

In this study, we assessed the impact of genetic polymorphisms on cytokine production capacity in response to several different stimuli in a cohort of low-birth-weight neonates from Guinea-Bissau, who were randomized to be vaccinated with BCG at birth or to the usual delayed administration of BCG. BCG vaccination has been shown to reduce all-cause mortality in children by providing protection against a wide range of infections in addition to its protective specific effects against tuberculosis. In neonates, for instance, randomized trials have shown that BCG reduced all-cause neonatal mortality by more than a third[17]. However, clearly many neonates die in spite of being BCG-vaccinated. Furthermore, considerable variation has been observed in the immune responses following BCG[33]. Characterization of the genetic variants influencing BCG-induced immune response may increase our understanding of pathways and mechanisms driving these responses. We further aimed to compare the strength of association as well as the proportion of overlapping cQTLs between individuals of European and African ancestries.

Generally, we observed strong inter-individual variability in cytokine production after in vitro challenge with microbial and non-microbial stimuli. Though the BCG group had enhanced responses due to BCG trained immunity compared with the control group[18], both BCG-vaccinated and unvaccinated responded to stimulation with increased cytokine production. These observations are consistent with what was reported about cytokine response architecture using individuals of predominantly European ancestry[31]. Furthermore, we identified 13 genome-wide significant cQTLs in 12 independent loci that affect cytokine levels in the control group. None of the identified SNPs were located within a gene or in the cis-regions of genes encoding for a particular cytokine, suggesting that the identified cQTLs are rather located in regulatory regions of other genes that modulate cytokine responses or, in other words, mainly trans-regulatory pathways are involved in cytokine regulation. This is consistent with what we observed also with studies in the Human Genomics Projects consisting of European cohorts[3,31].

To reveal genetic factors and pathways influencing cytokines responses upon *in vitro* stimulation after BCG vaccination, we performed a separate genome-wide genetic analysis in the BCG-vaccinated infants. We identified 9 independent genome-wide significant loci affecting cytokine responses, including 4 cQTLs affecting cytokine responses upon stimulation with PPD, suggesting that these cQTLs are possible candidates that specifically influence innate memory responses against mycobacteria induced by BCG vaccination. An argument for this assumption is the fact that none of these loci were associated with cQTLs in the BCG-unvaccinated infants. In support of this, various pathways that were overrepresented for the genes in the vicinity of the cQTLs in BCG-vaccinated infants (e.g. T-cell- and B-cell signaling, JAK/STAT signaling, glycolysis), have been identified in previous BCG vaccination studies[29]. Moreover, in the clinical settings, a purified protein derivate (PPD) skin test is used to determine whether a person has been infected with TB or not, and the host immune cells release of interferon gamma (IFN-γ) is an indication of positive result. Interestingly, a cQTL rs202131186 show strong correlation with interferon gamma (IFN-γ) upon stimulation with purified protein derivate (PPD). This finding indicates that the host genetic variation might influence PPD-based TB test results. Furthermore, two cQTLs rs6834860 and rs5887565 resides in *INPP4B* and *PLXNA4* genes, respectively. Previous studies have indicated that

INPP4B expression restrains cancer cell growth and metastasis by regulating the activation of PI3K/Akt signaling[34], while via genome-wide association study, several single nucleotide polymorphisms (SNPs) in *PLXNA4* gene have been reported to play a role in Alzheimer's disease (AD) and tau phosphorylation[35]. Intriguingly, BCG vaccination has been associated with a reduced risk of lung cancer and a reduced risk of Alzheimer in epidemiological studies[36,37], and it remains to be studied if *PLXNA4*-dependent effects may play a role in this.

Comprehensive understanding of the specific and non-specific immune response elicited upon vaccination is crucial to developing efficacious vaccines for infectious diseases, and to understand and optimize their non-specific effects on other diseases. Thus, an important insight emerging from our analysis being the specific enrichment of cQTLs identified upon BCG vaccination within complement genes, while not such effect was seen in BCG unvaccinated infants. It is an important finding given that the complement system forms an integral part of innate immunity. While previous work demonstrated that *M. bovis* BCG activates the complement system through classical and alternative pathways[38], little is known concerning the exact contribution of the compliment cascade and its impact on TB progression[39]. With this observation, we establish a clear link between the polymorphisms within the complement genes and response to BCG vaccine. Further longitudinal studies are needed to understand the implications of these polymorphisms in conferring susceptibility to TB, which will further deepen our knowledge of tuberculosis etiology and identify potential therapeutic targets, and in conferring non-specific protection against severe infections/mortality. Interestingly, we not only observed eQTL effect of SNP rs3759416 residing in the CLSTN3 locus on multiple genes including *C1R*, *C1RL*, *CLSTN3*, and *RBP5* but also validated their effect on IL-1β production. Future studies are needed to disentangle how all these genes participate in regulating cytokine responses upon vaccination.

Several GWAS have revealed variability in susceptibility to infectious diseases among people from different populations. Genetic determinants of disease and their effect sizes vary significantly between the European and the non-European populations, especially individuals of African ancestry[5], and indeed, genetic findings from populations of European ancestry could often not be replicated in non-European populations[40]. Recently, to comprehensively understand the genetic mechanisms underlying these variations in immune response, Quach et al performed eQTL analysis and reported differences in *cis-eQTLs* identified between the Africans and the Europeans, especially genes involved in inflammatory and viral responses[41]. The current analysis comparing Africans and Europeans in the context of genetics of cytokine responses upon stimulation with pathogens revealed a similar pattern, as we observed minimal overlap of genetic variants between the populations. Our findings highlight the importance of inclusion of populations with diverse genetic background in functional genomic studies.

Strengths of the study includes that this was based on a randomized trial, ensuring that BCG-vaccinated and BCG-unvaccinated cohorts were comparable. Limitations include that samples were only collected at 4 weeks after vaccination. As neonatal mortality was reduced by a third among BCG vaccinated children; those that did not survive may have had a different genetic make-up, especially rare variants, and BCG may have saved such children. It would have been ideal to have performed immunological assays prior to randomization, but this was unfortunately not feasible. The current study was performed in newborns, whereas the comparator genomic studies assessed were all performed in European adults. Hence, age and/or other factors like OPV could potentially explain the differences in the degree of overlap and strength of association. It is recommended for future analyses to consider comparable age groups, similar cytokine stimulation pairs as well as similar cellular systems (whole blood or PBMCs). Nevertheless, findings from our data in the context of comparing cQTLs of Europeans and Africans still provides valuable information.

In conclusion, by performing a genome-wide genetic analysis in a population of African infants, we identified important genetic factors that affect cytokine production and revealed gene candidates that may affect BCG-induced immunological responses. These findings increase our understanding of the molecular mechanisms that affect cytokine responses and the non-specific effects of BCG vaccine and open new avenues for the development of therapeutics and design of vaccines in African populations.

Methods

Study design and participants

This study was performed in a cohort of newborns in Guinea-Bissau, West Africa, part of the Bandim Health Project (BHP) (clinicaltrials.gov: NCT00625482). A detailed description of study design, inclusion and follow-up has been described previously[17,18]

Briefly, low-birth-weight (<2.5 kg) newborns were randomly assigned to early BCG (BCG given at birth) or delayed BCG (BCG given around six weeks of age, control group). Blood was drawn at four weeks of age (before individuals in the control group were vaccinated with BCG). All infants received oral polio vaccine (OPV) at birth. Out of the 467 individuals with measurements of *ex vivo* cytokine production from the original study[18], the blood clot had been preserved and was available from 453 individuals and genome-wide SNP genotype data for 429 individuals. After quality control, we obtained combined genetic and cytokine data for 187 individuals (105 females and 82 males) in the BCG group and 163 individuals (92 females and 71 males) in the control group. A general overview of the study design and analysis is shown in **Fig. 1** while overview of the study cohort is presented in **Fig. S3**.

The study was approved by the National Committee on Health ethics of the Ministry of Health in Guinea-Bissau and the Danish National Committee on Biomedical Research Ethics.

In vitro whole blood stimulation assay

In vitro whole blood stimulation assays were performed as previously described[18] Briefly, heparinized blood was diluted 1:9 with RPMI culture medium (Roswell Park Memorial Institute medium, Life Technologies Europe BV) supplemented with 1mM Pyruvate (Lonza, Copenhagen, Denmark) and L-Glutamine-Penicillin-Streptomycin 1x (Gibco Technologies). Whole blood was incubated at 37 °C for 24 hours either with culture medium (RPMI) only as a negative control, phorbol 12-myristate 13-acetate (PMA) (100 ng/mL; Sigma-Aldrich) and ionomycin (1 µg/mL) (Sigma-Aldrich) as a positive control, purified protein derivative (PPD) from *M. tuberculosis* (Statens Serum Institut, Copenhagen, Denmark) (10 µg/mL) to investigate the mycobacterial specific response, Toll-like receptor (TLR) 4 agonist lipopolysaccharide (LPS) (10 ng/mL) (Sigma-Aldrich), TLR1/2 agonist (S)-(2,3-bis(palmitoyloxy)-(2-RS)-propyl)-N-palmitoyl-(R)-Cys-(S)-Ser-(S)-Lys4-OH,trihydrochloride (Pam3CSK4) (1 µg/mL) (InvivoGen) or a TLR7-8 agonist (Thiazoloquinoline Compound (CL075), 1 µg/mL, InvivoGen). Supernatants were collected and stored at -20°C until assayed.

Cytokine quantification

Concentrations of interleukin (IL)-1β, IL-5, IL-6, IL-10, IL-17, interferon (IFN)-γ and tumor necrosis factor (TNF)-α measured from whole-blood stimulation assays were analyzed by an immunobead-based multiplexed assay (Luminex).

European cohorts

In order to compare the data from the Guinea-Bissau cohort to studies performed in the European population, we made use of two independent cohorts of ~ 500 and ~300 healthy individuals of Western European ancestry from the Human Functional Genomics Project (500FG and 300BCG cohorts, wwww.humanfunctionalgenomics.org). The 500FG cohort consists of 237 males and 296 females (age range 18-75 years), the 300BCG cohort consists of 141 males and 184 females (age range 18 – 71 years). In 500FG, whole blood, peripheral blood mononuclear cells (PBMCs), and macrophages were stimulated with microbial and non-microbial stimuli to induce cytokine production and the impact of genetic heritability was assessed. Individuals that took part in the 300BCG study received 0.1ml of BCG (BCG vaccine Bulgaria strain, Intervax). Blood sampling was performed before BCG vaccination, two weeks and three months after vaccination. At all timepoints, PBMCs were stimulated *ex vivo* with different microbial stimuli and the impact of genetic variants on cytokine production before vaccination was assessed. In addition, the effect of genetic variation on the BCG-induced trained immunity response was assessed. Both studies were approved by the Arnhem-Nijmegen Medical Ethical Committee (500FG: NL42561.091.12; 300BCG: NL58553.091.16) and performed in accordance with the Declaration of Helsinki. All individuals gave written informed consent to donate venous blood for research.

Cytokine data quality control

Distributions of raw *in vivo* cytokine concentrations deviated from normality. Transformation of phenotypes to normality aims to deal with skewness of cytokine distributions and to permit parametric testing. Moreover, one characteristic of quantitative traits is that variations in individuals within a natural population follow a normal distribution. Therefore, raw cytokine concentrations were

first log2-transformed and then normalized to follow standard Gaussian distributions using inverse ranked-based normal transformation function in the GenABEL R package[42] . We show an example of such transformation in **Fig. S4A**. Unsupervised hierarchical clustering analysis was also performed using Pearson's correlation as a measure of similarity to identify potential outliers. In total, 3 outliers were removed from downstream analysis based on upon visual inspection of the dendrogram.

Genotyping, Quality control and Imputation of genetic data

DNA was extracted from dried blood spots using the QIAamp 96 DNA Blood Kit (Qiagen). Default settings of Opticall 0.70[43] was used to perform genotype calling. Genotyping was performed using the Global Screening Arrays (GSA). Prior to imputation of the genetic data, SNP quality control (QC) was performed by excluding variants with call rate exceeding 0.1, minor allele frequencies (MAF < 0.01) and SNPs deviating from Hardy-Weinberg Equilibrium (HWE) with a p-value < 1×10^{-4} . Two individuals with high missing genotype data (>0.1) and five individuals with high or low heterozygosity rates were removed. After pre-imputation quality control, a total of 429 individuals and 357,960 variants were available. Strand alignment to a reference panel with the aid of Genotype harmonizer[44] was performed using 1000 Genome reference panel dataset.

To improve genome coverage, genotype imputation was performed for all chromosomes except the sex chromosomes using the Minimac4 software through the publicly available Michigan Imputation Server[45]. The Human Reference Consortium (HCR r.1.1 2016) was used as reference panel and the dataset was phased using Eagle v2.3. We filtered out variants with imputation quality score (R^2) < 0.3. Sample quality control such as extreme heterozygosity, cryptic relatedness and sex discordance were performed. Based on these criteria, a total of 52 samples that failed QC were removed. No genetic outlier was identified using multi-dimensional scaling plots as our samples exhibited a homogenous population structure after merging with 1000 Genome data (**Fig. S4B**). Similar pattern was observed after projecting the control and BCG group samples separately on individuals of African origin (**Fig. S4C**) Upon genome-wide SNP genotyping and imputation, a total of 6,739,950 variants in 377 individuals were generated.

Previously, genotyping, quality control and imputation procedures for the European cohorts: 500FG and 300BCG were described[31,46].

Cytokine QTL mapping

To test for association between cytokine concentrations and genotypes (SNP dosages), we excluded SNPs with minor allele frequencies (MAF) below 5% and SNPs deviating from Hardy-Weinberg Equilibrium (HWE) with a p-value < 1×10^{-6}. We further considered a subset of 187 BCG-vaccinated individuals and 163 controls with complete matched cytokines, covariate and genetic data. For association analyses, inverse ranked normalized cytokine concentrations were adjusted for age and sex and subsequently mapped on 6,735,702 autosomal single nucleotide polymorphisms (SNPs) independently in both cohorts. The genome-wide significance threshold was considered at p-value < 5×10^{-8}. Linkage disequilibrium (LD) analysis around lead SNPs was used to determine independent loci. To perform LD analysis, pairwise LD, as measured by R^2 (>= 0.8) between all SNPs within 5 Mb upstream and downstream of the significant lead SNPs, was calculated using our imputed dataset as reference data.

Thereafter, to identify SNPs associated with cytokine responsiveness irrespective of BCG vaccination, meta-analysis was conducted by integrating summary statistics of cQTL association results from the BCG-vaccinated group and control group and with each cytokine stimulation pairs independently. Heterogeneity statistics based on chi-square test were calculated for all SNPs to estimate heterogeneity of effect sizes in both cohorts. Statistically significant SNPs were considered for a meta-p value $\leq 5\times10^{-8}$, with consistent effect size direction and statistically not significant heterogeneity between the two cohorts.

Downstream analyses of significant cQTLs

SNP2TFBS web-based annotation tool [47], was applied to study functional effects of genetic variants lying in transcription factor binding sites (TFBS) of the human genome. Briefly, a SNP's effect on transcription factor (TF) binding is estimated with position weight matrix (PWM) model for the binding specificity of the corresponding factor. P-value less than 0.05 was declared as statistically significant threshold for TF enrichment.

Complement system annotation and enrichment analysis

We retrieved the set of genes annotated as part of the complement system in the HUGO Gene Nomenclature Committee (HGNC) repository[48]. In total, 65 approved genes were included in the curated list (**Supplementary Table 4**). However, two genes on chromosome X (*CFP* and *VSIG4*) were excluded from further analysis as cQTL mapping was performed only on autosomal SNPs. Next, we performed competitive gene set enrichment analysis which uses gene-level regression model as implemented in the MAGMA (Multi-marker analysis of genomic annotation) [49] tool for gene and gene set analysis. The actual analysis is proceeded with the annotation step, which involves mapping SNPs or aggregating SNP associations in a specific genomic window around genes. Several window sizes ranging from 0kb to 250kb with a step size of 50kb for gene annotation were tested. Subsequently, gene analysis on SNP p-values was performed and used as input for competitive gene-set analysis.

Cell culture

The THP-1 human monocytic cell line (derived from an acute leukaemia patient) was obtained from InvivoGen. THP1 cells were maintained in complete media, which is composed of RPMI 1640, 1% (2mM) GlutaMAX L-glutamine supplement, 25 mM HEPES, 10% FBS, 100 µg/ml Normocin, Pen-Strep (100 U/ml), Blasticidin (10 µg/ml) and Zeocin (100 µg/ml). Prior to transfection, the THP-1 monocyte culture was split by 50% to enable the cells to re-enter an exponential growth phase. Each gapmer compound tested was added to the THP-1 cells at a final concentration of 10 µM and mixed gently. Plates were incubated at 37 ˚C at 5% CO_2 for 24 hours. Then, LPS (10 ng/mL) was added to each well, and plates were incubated at 37 ˚C at 5% CO_2 for another 24 hours.

cDNA synthesis and qPCR (RT-qPCR)

All qPCR measurements were made on complementary DNA (cDNA) produced from whole RNA. RNA was extracted using the Zymo RNA extraction kit and converted to cDNA by reverse transcription using the SuperScript IV cDNA kit (Invitrogen). qPCR was performed in technical duplicates on a CFX-96 Real-Time PCR Detection System (Bio-Rad). Reactions were set up using the SsoAdvanced Universal Sybr Green Supermix (Bio-Rad) according to the manufacturer's protocol. Relative quantification was performed using the delta-delta Ct method, using *RPL37A* as the reference gene.

ChIP-Seq Analysis

Publicly available datasets were examined using the UCSC genome browser. Specifically, the data shown here are traces from the ENCODE consortium and the GeneHancer database[50]

Statistical analysis

All statistical analyses were performed in R: a free software environment for statistical computing and graphics[51]. SNP QC and preprocessing analysis as well as LD estimation to determine independent loci were performed with PLINK 1.90[52].

For cQTL mapping, we utilized an additive linear model as implemented in the Matrix-eQTL R package[53]. Age and sex covariates were adjusted in the linear model. The meta-analysis was carried out using fixed effects sample size weighted analysis method implemented in the package METAL[54], based on cQTL summary statistics (p-values).

Declarations

Data and Code Availability

cQTL summary statistics will be deposited to the EBI GWAS Catalog. Other data supporting this manuscript and code are available from the corresponding authors upon reasonable request.

Author contributions

V.K., C.S.B., and M.G.N. designed the study. V.K., L.A.B.J., and M.G.N. supervised and generated genotype data. C.S.B., K.J.J., I.M., and P.A. recruited the African cohort and generated cytokine data. C.K.B., V.M., C.d.B., and S.J.C.F.M.M. performed the data analysis. C.K.B conducted cytokine QTL mapping and other statistical analyses. V.K., M.G.N, L.A.B.J., C.K.B., and S.J.C.F.M.M. conducted data interpretation. E.T.F., S.F., M.M. conducted functional experiments. C.K.B. prepared the figures and tables. C.K.B., and S.J.C.F.M.M. drafted the manuscript. V.K., C.S.B., M.G.N., K.J.J., and P.A. reviewed and edited the manuscript. All authors contributed to critical revisions and approved the final manuscript.

The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Acknowledgements

The authors thank all volunteers from the BHP cohort and 500FG and 300BCG cohorts for participation in the studies. Financial support. M.G.N. was supported by a Spinoza grant of the Netherlands Organization for Scientific Research and an ERC Advanced Grant (grant number 833247). The original trial was supported by The European Research Council (starting grant ERC-2009-StG-243149); the Novo Nordisk Foundation (research professorship grant to P. A.); the Danish National Research Foundation (grant DNRF108); and DANIDA, European Union FP7, and OPTIMUNISE (grant Health-F3-2011–261375 to the Bandim Health Project). V.K was supported by a Hypathia tenure track grant RadboudUMC

Competing Interests

The authors declare no competing interests.

Supplemental information

Supplementary data include four figures and four tables.

References

1. ter Horst, R. *et al.* Host and Environmental Factors Influencing Individual Human Cytokine Responses. *Cell* (2016). doi:10.1016/j.cell.2016.10.018

2. Oosting, M. *et al.* Functional and Genomic Architecture of Borrelia burgdorferi-Induced Cytokine Responses in Humans. *Cell Host Microbe* (2016). doi:10.1016/j.chom.2016.10.006

3. Li, Y. *et al.* Inter-individual variability and genetic influences on cytokine responses to bacteria and fungi. *Nat. Med.* (2016). doi:10.1038/nm.4139

4. Boahen, C. K. *et al.* A functional genomics approach in Tanzanian population identifies distinct genetic regulators of cytokine production compared to European population. *Am. J. Hum. Genet.* **109**, 471–485 (2022).

5. Bentley, A. R., Callier, S. L. & Rotimi, C. N. Evaluating the promise of inclusion of African ancestry populations in genomics. *npj Genomic Medicine* (2020). doi:10.1038/s41525-019-0111-x

6. Organization, W. H. Immunization coverage. *Fact sheet N 378* (2015).

7. Aaby, P. *et al.* Randomized trial of BCG vaccination at birth to low-birth-weight children: Beneficial nonspecific effects in the neonatal period? *J. Infect. Dis.* (2011). doi:10.1093/infdis/jir240

8. Biering-Sørensen, S. *et al.* Small randomized trial among low-birth-weight children receiving bacillus Calmette-Guéerin vaccination at first health center contact. *Pediatr. Infect. Dis. J.* (2012). doi:10.1097/INF.0b013e3182458289

9. Biering-Sørensen, S. *et al.* Early BCG-Denmark and Neonatal Mortality Among Infants Weighing. *Clin. Infect. Dis. An Off. Publ. Infect. Dis. Soc. Am.* (2017). doi:10.1093/CID/CIX525

10. Netea, M. G. *et al.* Defining trained immunity and its role in health and disease. *Nature Reviews Immunology* (2020). doi:10.1038/s41577-020-0285-6

11. Kleinnijenhuis, J. *et al.* Bacille Calmette-Guérin induces NOD2-dependent nonspecific protection from reinfection via epigenetic reprogramming of monocytes. *Proc. Natl. Acad. Sci. U. S. A.* (2012). doi:10.1073/pnas.1202870109

12. Arts, R. J. W. *et al.* BCG Vaccination Protects against Experimental Viral Infection in Humans through the Induction of Cytokines Associated with Trained Immunity. *Cell Host Microbe* (2018). doi:10.1016/j.chom.2017.12.010

13. Walk, J. *et al.* Outcomes of controlled human malaria infection after BCG vaccination. *Nat. Commun.* (2019). doi:10.1038/s41467-019-08659-3

14. Franco, L. M. *et al.* Integrative genomic analysis of the human immune response to influenza vaccination. *Elife* (2013). doi:10.7554/eLife.00299

15. Tan, P. L., Jacobson, R. M., Poland, G. A., Jacobsen, S. J. & Pankratz, V. S. Twin studies of immunogenicity - Determining the genetic contribution to vaccine failure. in *Vaccine* (2001). doi:10.1016/S0264-410X(00)00468-0

16. Ovsyannikova, I. G. *et al.* The association of CD46, SLAM and CD209 cellular receptor gene SNPs with variations in measles vaccine-induced immune responses: A replication study and examination of novel polymorphisms. *Hum. Hered.* (2011). doi:10.1159/000331585

17. Biering-Sørensen, S. *et al.* Early BCG-Denmark and Neonatal Mortality among Infants Weighing <2500 g: A Randomized Controlled Trial. *Clin. Infect. Dis.* (2017). doi:10.1093/cid/cix525

18. Jensen, K. J. *et al.* Heterologous immunological effects of early BCG vaccination in low-birth-weight infants in guinea-bissau: A randomized-controlled trial. *J. Infect. Dis.* (2015). doi:10.1093/infdis/jiu508

19. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, (2016).

20. Ward, L. D. & Kellis, M. HaploReg v4: Systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. *Nucleic Acids Res.* (2016). doi:10.1093/nar/gkv1340

21. Ghoussaini, M. *et al.* OUP accepted manuscript. *Nucleic Acids Res.* 1–10 (2020). doi:10.1093/nar/gkaa840

22. Carpenter, S. & Fitzgerald, K. A. Cytokines and long Noncoding RNAs. *Cold Spring Harb. Perspect. Biol.* (2018). doi:10.1101/cshperspect.a028589

23. Wang, J., Vasaikar, S., Shi, Z., Greer, M. & Zhang, B. WebGestalt 2017: A more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. *Nucleic Acids Res.* (2017). doi:10.1093/nar/gkx356

24. Cirovic, B. *et al.* BCG Vaccination in Humans Elicits Trained Immunity via the Hematopoietic Progenitor Compartment. *Cell Host Microbe* (2020). doi:10.1016/j.chom.2020.05.014

25. Yamamoto, M. *et al.* Plexin-A4 negatively regulates T lymphocyte responses. *Int. Immunol.* (2008). doi:10.1093/intimm/dxn006

26. Kaufmann, S. H. E. & Parida, S. K. Tuberculosis in Africa: Learning from Pathogenesis for Biomarker Identification. *Cell Host and Microbe* (2008). doi:10.1016/j.chom.2008.08.002

27. Gewinner, C. *et al.* Evidence that Inositol Polyphosphate 4-Phosphatase Type II Is a Tumor Suppressor that Inhibits PI3K Signaling. *Cancer Cell* (2009). doi:10.1016/j.ccr.2009.06.006

28. Cheng, S. C. *et al.* MTOR- and HIF-1α-mediated aerobic glycolysis as metabolic basis for trained immunity. *Science (80-.).* (2014). doi:10.1126/science.1250684

29. Hasso-Agopsowicz, M., Scriba, T. J., Hanekom, W. A., Dockrell, H. M. & Smith, S. G. Differential DNA methylation of potassium channel KCa3.1 and immune signalling pathways is associated with infant immune responses following BCG vaccination. *Sci. Rep.* (2018). doi:10.1038/s41598-018-31537-9

30. Mcgarvey, J. A., Wagner, D. & Bermudez, L. E. Differential gene expression in mononuclear phagocytes infected with pathogenic and non-pathogenic mycobacteria. *Clin. Exp. Immunol.* (2004). doi:10.1111/j.1365-2249.2004.02490.x

31. Li, Y. *et al.* A Functional Genomics Approach to Understand Variation in Cytokine Production in Humans. *Cell* (2016). doi:10.1016/j.cell.2016.10.017

32. Mathieson, I. *et al.* Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* (2015). doi:10.1038/nature16152

33. Finan, C., Ota, M. O. C., Marchant, A. & Newport, M. J. Natural variation in immune responses to neonatal Mycobacterium bovis Bacillus Calmette-Guerin (BCG) vaccination in a cohort of Gambian infants. *PLoS One* (2008). doi:10.1371/journal.pone.0003485

34. Chen, Y. *et al.* INPP4B restrains cell proliferation and metastasis via regulation of the PI3K/AKT/SGK pathway. *J. Cell. Mol. Med.* (2018). doi:10.1111/jcmm.13595

35. Han, Q. *et al.* Common Variants in PLXNA4 and Correlation to CSF-related Phenotypes in Alzheimer's Disease. *Front. Neurosci.* (2018). doi:10.3389/fnins.2018.00946

36. Usher, N. T. *et al.* Association of BCG Vaccination in Childhood with Subsequent Cancer Diagnoses: A 60-Year Follow-up of a Clinical Trial. *JAMA Netw. Open* **2**, (2019).

37. Gofrit, O. N. *et al.* Can immunization with Bacillus Calmette-Guérin (BCG) protect against Alzheimer's disease? *Med. Hypotheses* **123**, (2019).

38. Carroll, M. V., Lack, N., Sim, E., Krarup, A. & Sim, R. B. Multiple routes of complement activation by Mycobacterium bovis BCG. *Mol. Immunol.* **46**, (2009).

39. Jagatia, H. & Tsolaki, A. G. The role of complement system and the immune response to tuberculosis infection. *Medicina (Lithuania)* **57**, (2021).

40. Ioannidis, J. P. A. Population-wide generalizability of genome-wide discovered associations. *Journal of the National Cancer Institute* (2009). doi:10.1093/jnci/djp298

41. Quach, H. *et al.* Genetic Adaptation and Neandertal Admixture Shaped the Immune System of Human Populations. *Cell* (2016). doi:10.1016/j.cell.2016.09.024

42. Aulchenko, Y. S., Ripke, S., Isaacs, A. & van Duijn, C. M. GenABEL: An R library for genome-wide association analysis. *Bioinformatics* (2007). doi:10.1093/bioinformatics/btm108

43. Shah, T. S. *et al.* OptiCall: A robust genotype-calling algorithm for rare, low-frequency and common variants. *Bioinformatics* (2012). doi:10.1093/bioinformatics/bts180

44. Deelen, P. *et al.* Genotype harmonizer: Automatic strand alignment and format conversion for genotype data integration. *BMC Res. Notes* (2014). doi:10.1186/1756-0500-7-901

45. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat. Genet.* (2016). doi:10.1038/ng.3656

46. Moorlag, S. J. C. F. M. *et al.* An integrative genomics approach identifies KDM4 as a modulator of trained immunity. *Eur. J. Immunol.* (2021). doi:10.1002/eji.202149577

47. Kumar, S., Ambrosini, G. & Bucher, P. SNP2TFBS-a database of regulatory SNPs affecting predicted transcription factor binding site affinity. *Nucleic Acids Res.* **45**, (2017).

48. Seal, R., Gordon, S., Lush, M., Wright, M. & Bruford, E. HGNC: The Why and How of Standardised Gene Nomenclature. *Nat. Preced.* (2009). doi:10.1038/npre.2009.3182.1

49. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLoS Comput. Biol.* **11**, (2015).

50. Fishilevich, S. *et al.* GeneHancer: Genome-wide integration of enhancers and target genes in GeneCards. *Database* **2017**, (2017).

51. R Development Core Team, R. *R: A Language and Environment for Statistical Computing*. *R Foundation for Statistical Computing* (2011). doi:10.1007/978-3-540-74686-7

52. Purcell, S. *et al.* PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* (2007). doi:10.1086/519795

53. Shabalin, A. A. Matrix eQTL: Ultra fast eQTL analysis via large matrix operations. *Bioinformatics* (2012). doi:10.1093/bioinformatics/bts163

54. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* (2010). doi:10.1093/bioinformatics/btq340

Tables

Table 1. Summary genome-wide significant cQTL results of control group

Loci	SNPs	Chr	Base pair	Cytokines	Stimuli	P value	Causal genes
1	rs11918168	3	131397888	IL-6	LPS	2.42×10^{-9}	CPNE4[a]
2	rs11129738	3	36901580	IL-6	PPD	2.49×10^{-8}	TRANK1[a,b], LRRFIP2[b]
3	rs17073317	4	183369097	IL-17	LPS	1.14×10^{-8}	ODZ3[a]
4	rs16899560	5	81702810	IFN-γ	LPS	3.14×10^{-8}	ATP6AP1L[a]
5	rs2441704	8	102133238	IL-17	PPD	4.12×10^{-8}	FLJ42969[a]
6	rs34713905	8	69314096	IFN-γ	LPS	2.33×10^{-8}	C8orf34[a]
7	rs75715336	10	120225199	IL-1β	PPD	1.75×10^{-8}	AL356865.2[a]
8	rs62051095	16	86314831	IL-5	PMA	3.0×10^{-8}	LOC146513[a]
9	rs11643992	16	86296221	TNF-α	PMA	4.69×10^{-8}	LOC146513[a]
10	rs74652081	17	4744397	TNF-α	PPD	3.82×10^{-8}	MINK1[a]
11	rs58229361	20	30511738	IFN-γ	LPS	3.36×10^{-10}	TTLL9[a]
	rs144893358	20	30507643	IFN-γ	LPS	3.57×10^{-9}	TTLL9[a]
12	rs9975790	21	22289731	IL-17	CLO75	2.73×10^{-8}	NCAM2[a]

[a] The gene in closest proximity to the cytokine QTL SNPs

[b] Expression of this gene show association with Cytokine QTL SNP in whole blood

le 2. Summary of genome-wide significant cQTL results of BCG group

i SNPs	Chr	Base pair	Cytokines	Stimuli	P value BCG	P value Control	Causal genes
rs541710069	2	68977305	IL-17	PPD	4.81×10^{-8}	9.34×10^{-1}	ARHGAP25[a]
rs6834860	4	143861293	IL-10	CLO75	9.75×10^{-9}	5.50×10^{-1}	INPP4B[a,b], AC139720.1[b], USP38[b], AC104596.1[b]
rs13246275	7	22881568	TNF-α	PAM	2.66×10^{-9}	9.50×10^{-1}	SNORD93[a]
rs5887565	7	131989498	TNF-α	PPD	1.36×10^{-8}	7.20×10^{-2}	PLXNA4[a]
rs16904801	8	134012085	IL-17	PPD	5.80×10^{-9}	8.23×10^{-2}	TG[a]
rs3759416	12	7283646	IL-1β	LPS	3.26×10^{-8}	5.53×10^{-1}	CLSTN3[a,b],GNB3[b], C1R[b],EMG1[b], C1RL[b],RBP5[b]
rs116315540	19	39867913	IFN-γ	PAM	1.95×10^{-8}	9.33×10^{-1}	SAMD4B[a]
rs202131186	20	22359048	IFN-γ	PPD	3.96×10^{-8}	9.68×10^{-1}	LOC284788[a]
rs8138652	22	43955593	IL-5	LPS	1.44×10^{-8}	4.81×10^{-1}	EFCAB6[a]

e gene in closest proximity to the cytokine QTL SNPs
pression of this gene show association with Cytokine QTL SNP in whole blood

Table 3. Genome-wide significant cQTL identified by meta-analysis of control and BCG group

Loci	SNPs	Chr	Base Pair	Cytokines	Stimulation	Zscore	Meta P value	Control P value	BCG P value	Allelic Direction	Causal genes
1	rs4674520	2	201713454	IL-17	PPD	5.492	3.97×10^{-8}	1.61×10^{-5}	4.90×10^{-4}	++	RN5S115[a],PPIL3[b] CASP10[b],CFLAR[b]
2	rs62313427	4	108668193	TNF-a	CLO75	-5.674	1.40×10^{-8}	3.38×10^{-6}	6.21×10^{-4}	--	PAPSS1[a,b]
3	rs7717530	5	31076540	IL-6	PPD	5.479	4.29×10^{-8}	2.42×10^{-5}	3.82×10^{-4}	++	CDH6[a]
4	rs10955956	8	121122623	IL-1β	PAM	5.637	1.73×10^{-8}	4.20×10^{-4}	9.87×10^{-6}	++	COL14A1[a]
5	rs11877080	18	29888196	IL-5	PAM	5.662	1.50×10^{-8}	1.02×10^{-4}	3.82×10^{-5}	++	FAM59A[a]

[a] The gene in closest proximity to the cytokine QTL SNPs
[b] Expression of this gene show association with Cytokine QTL SNP in whole blood

Supplementary Table 1. Comparing genome-wide significant cQTL identified in European cohort with control group cQTL

European 500FG					Guinea-Bissau					
						PAM	*PPD*	*LPS*	*PMA*	*CLO75*
SNPs	**Chr**	**Cytokines**	**Stimulation**	**P value**	**Cytokines**	**P value**	**P value**	**P value**	**P value**	**P value**
rs4496335	2	[p]TNF-a	*Cryptococcus*	4.2×10^{-9}	TNF-a	7.31×10^{-1}	5.97×10^{-1}	9.18×10^{-1}	9.02×10^{-1}	5.96×10^{-1}
rs9941692	2	[p]IL-6	*C. burnetii*	3.7×10^{-10}	IL-6	5.25×10^{-1}	9.19×10^{-1}	1.74×10^{-1}	8.90×10^{-1}	1.21×10^{-1}
rs28393318	4	[p]IL-1b	Poly(I:C)	2.8×10^{-11}	IL-1b	6.82×10^{-1}	5.05×10^{-1}	3.36×10^{-2}	7.21×10^{-1}	1.35×10^{-3}
rs10790723	11	[p]IL-1b	*C. burnetii*	1.0×10^{-8}	IL-1b	2.73×10^{-1}	2.29×10^{-1}	4.87×10^{-1}	5.26×10^{-1}	5.72×10^{-1}
rs7310164	12	[p]IL-6	*C. burnetii*	4.7×10^{-9}	IL-6	7.48×10^{-1}	9.42×10^{-1}	9.54×10^{-1}	9.71×10^{-1}	8.37×10^{-2}
rs7256586	19	[w]IL-6	*C. albicans*	8.5×10^{-9}	IL-6	9.61×10^{-1}	2.01×10^{-2}	6.88×10^{-1}	2.01×10^{-1}	5.67×10^{-1}

[p]Cytokines profiled from PBMCs [w]Cytokines profiled from whole blood

Figures

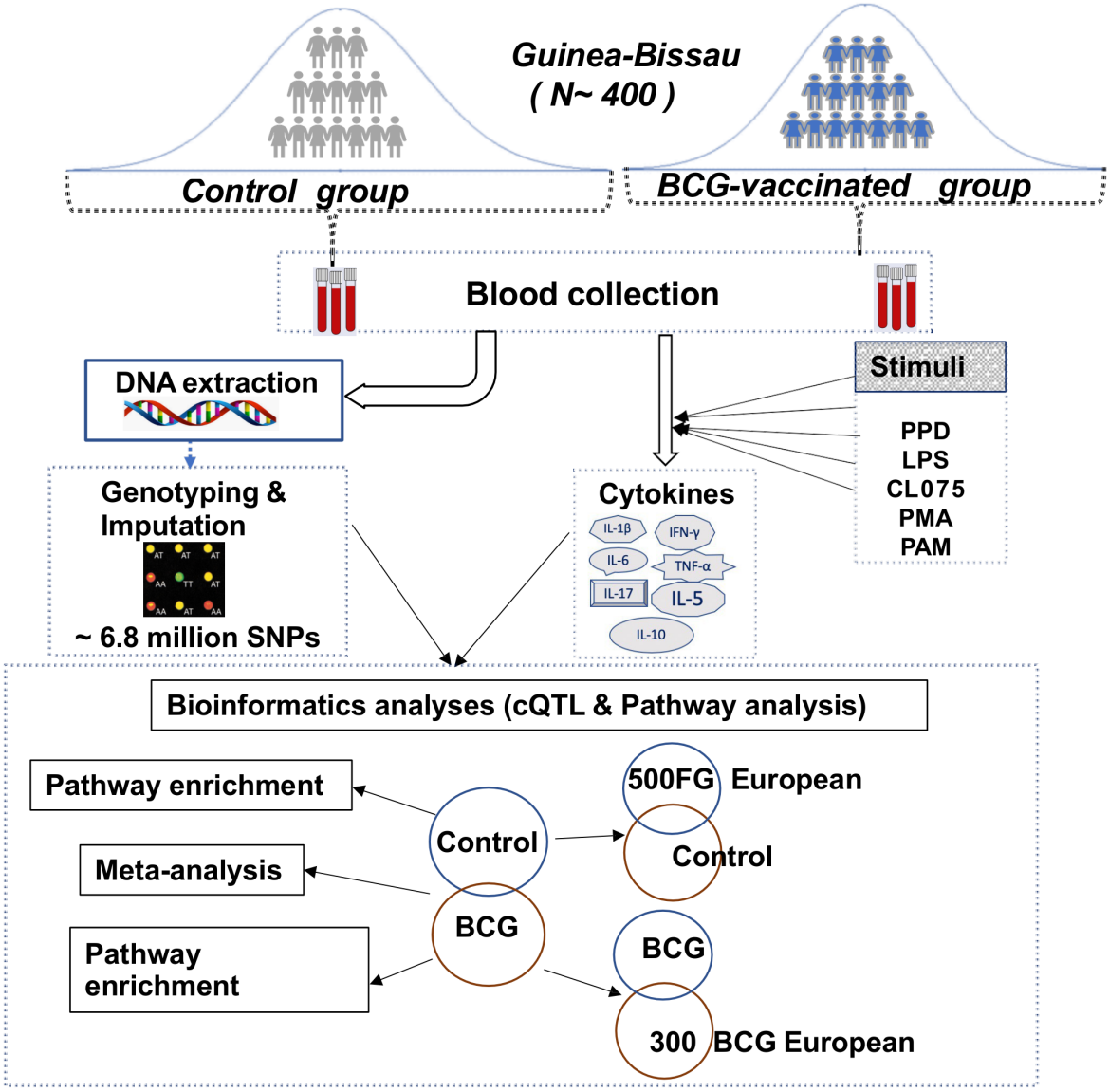

Figure 1

Schematic representation of study design and bioinformatics analysis

We collected more than 400 blood samples from infants randomized to BCG vaccination or no BCG vaccination from West-Africa. Genotyping was performed on isolated DNA using Global Screening Arrays (GSA). To profile cytokines expressions, whole blood was stimulated with 5 different stimuli. Cytokine QTLs were mapped by integrating the imputed genotypes data with cytokine measurements for the control and BCG groups independently and combined analysis was performed to further explore genetics of both groups irrespective of BCG vaccinatsion. Pathway enrichment analysis was performed on the identified genome-wide significant loci to gain biological insights. Cytokine QTLs identified in control and BCG groups cohort were compared with cQTLs identified previously in a European population.

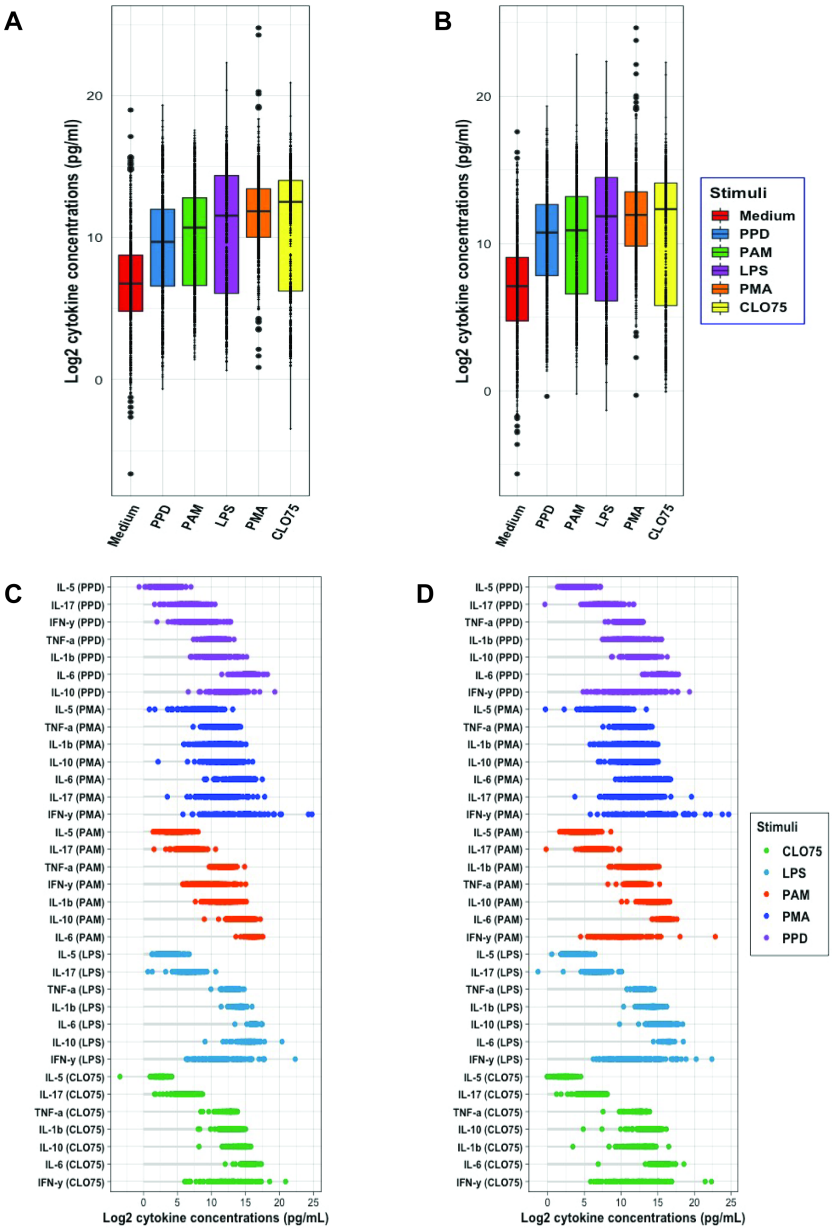

Figure 2

Overview of cytokine responses architecture

(A) Boxplot comparing cell culture medium and all stimuli cytokine production capacity (Control group) **(B).** Boxplot comparing cell culture medium and all stimuli cytokine production capacity (BCG group). For all boxplots, y-axes represent all cytokines levels in their respective Medium and stimuli (PPD, PAM, LPS, PMA, and CLO75) were merged together. **(C)** Dot plot depicting inter-personal differences in cytokine response upon stimulation (Control group). **(D)** Dot plot showing inter-personal differences in cytokine response upon stimulation (BCG-vaccinated group). The y-axes in the dot plots represent all cytokine-stimulation pairs used in this study. Labels on the y-axes are arranged in descending order of median values per stimuli.

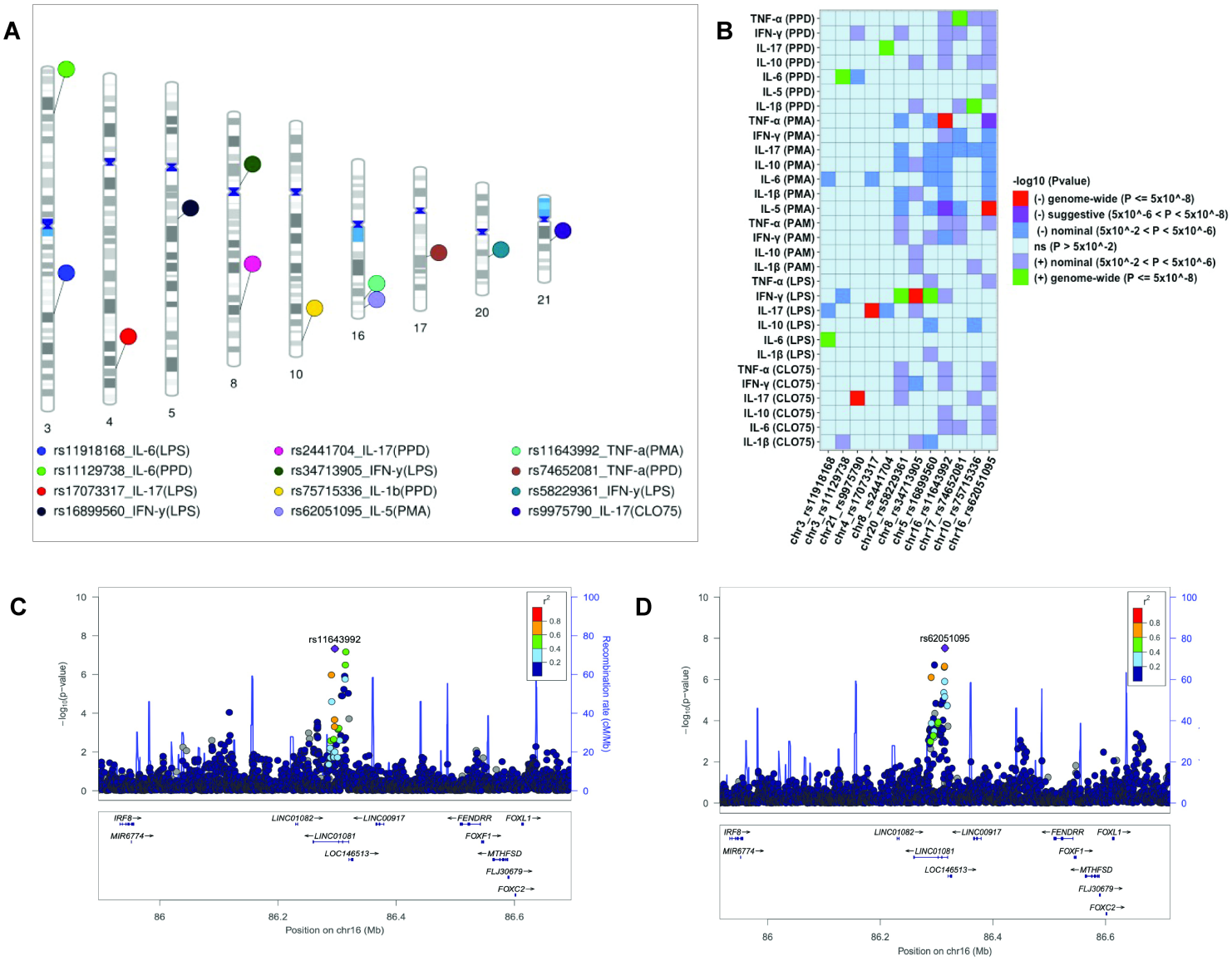

Figure 3

Summary of control group cytokine quantitative trait loci (cQTLs) results

(A) Phenogram depicting the 12 genome-wide significant loci associated with different cytokines. The legend shows the lead SNP rs ID, underscore cytokine and stimulations in brackets. **(B)** Heatmap showing the 12 genome-wide significant hits and their shared associations. The x-axis shows all the 12 genome-wide significant loci and the y-axis shows cytokine-stimulation pairs with at least one association with any of the 12 loci. SNP-cytokine associations with p-values greater than 5×10^{-2} considered statistically non-significant(ns). **(C)** Regional association plot at the *LOC145613* locus associated with Tumor necrosis factor alpha (TNF-α) upon PMA stimulation. **(D)** Regional association plot at the *LOC145613* locus associated with PMA induced interleukin-5(IL-5) concentrations. The regional plots for the SNPs in the region flanking a window of 400kb of two lead SNPs (rs11643992 and rs62051095) which shows multiple association with different cytokines. P-values in -log10 scale are shown on the left vertical axis, the recombination rates are shown on the right vertical axis, and the position (NCBI Build GRCh37) of the chromosomes are shown on the horizontal axis. Pairwise linkage disequilibrium estimates are colored with respect to the cQTL index SNP and displayed as a heatmap from dark blue ($0 \geq r^2 > 0.2$) to red ($0.8 \geq r^2 > 1.0$). The bottom panel of the plot shows the name and location of the genes which are shown as arrows.

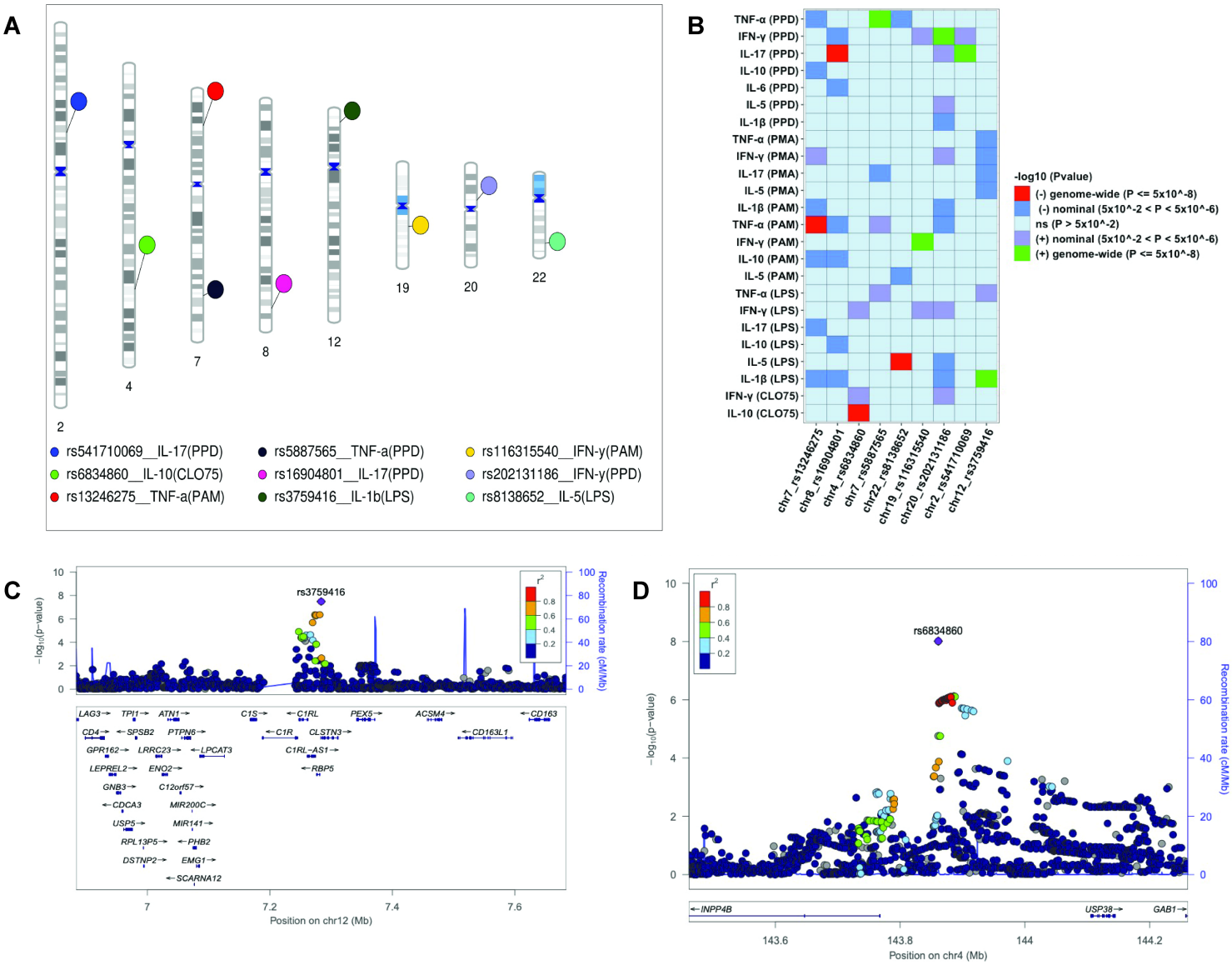

Figure 4

Summary of BCG group cytokine quantitative trait loci (cQTLs) results

(A) Phenogram displaying the 9 genome-wide significant loci associated with different cytokines. **(B)** Heatmap showing the 9 genome-wide significant hits and their shared associations. The x-axis shows all the 9 genome-wide significant loci and the y-axis shows cytokine-stimulation pairs with at least one association with any of the 9 loci. SNP-cytokine associations with p-values greater than 5×10^{-2} are considered statistically non-significant(ns). **(C)** Regional association plot of the *CLSTN3* locus associated with IL-1β concentrations upon LPS stimulation. **(D)** Regional plot of the INPP4B locus associated with interleukin-10(IL-10) concentrations upon CLO75 stimulation. P-values in -log10 scale are shown on the left vertical axis, the recombination rates are shown on the right vertical axis, and the position (NCBI Build GRCh37) of the chromosomes are shown on the horizontal axis. Pairwise linkage disequilibrium estimates are colored with respect to the cQTL index SNP and displayed as a heatmap from dark blue ($0 \geq r^2 > 0.2$) to red ($0.8 \geq r^2 > 1.0$). The bottom panel of the plot shows the name and location of the genes which are shown as arrows.

Figure 5

Enrichment of cQTLs at complement genes in BCG group

(**A**) Bar chart showing evidence of cQTLs enrichment within the complement system in the BCG group after MAGMA analysis. **(B)** Bar chart representing enrichment of complement genes (horizontal axis) from cQTL variants. Horizontal dashes lines depict the level of significance (MAGMA p-value less than 0.05). Genes are arranged in ascending order of P-values. **(C)** rs3759416 falls in an enhancer region within the first intron of CLSTN3. This region is primed with H3K4me1 in CD14+ monocytes, and is bound by several transcription factors, including STAT5 and RELA. **(D)** Antisense inhibition of RBP5, C1R, CLSTN3 and C1RL revealed that depletion of C1R and C1RL resulted in a significant reduction in IL-1b transcription.

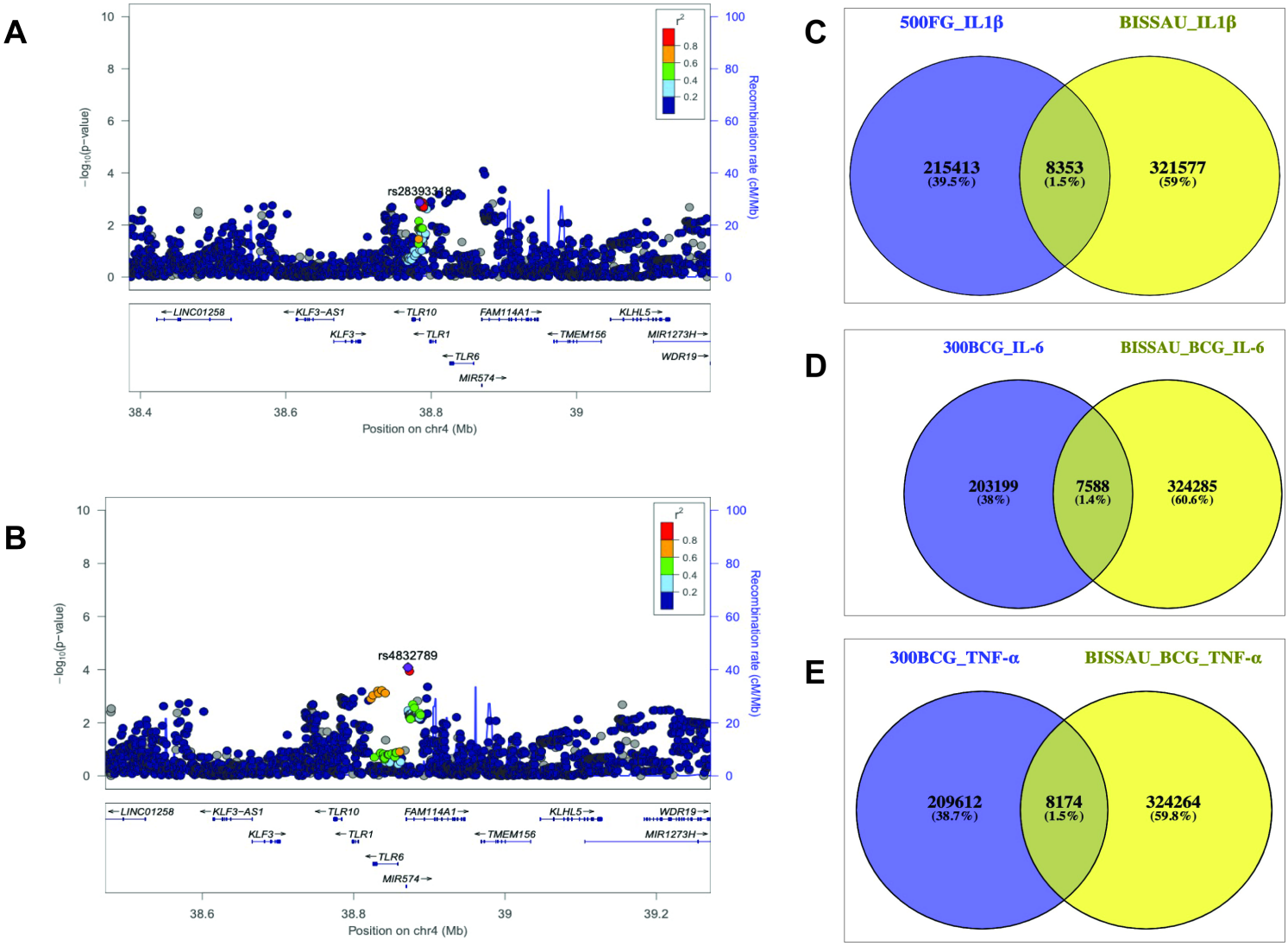

Figure 6

Comparison of cytokine QTLs between Africans and Europeans

(A) Regional association plot of SNPs flanking 400kb on either side of SNP rs28393318 associated with CLO75 induced IL-1β concentration at TLR10-TLR1-TLR6 locus. SNP rs28393318 showed nominal association but the strongest association among the genome-wide significant loci in the European cohort was at this locus. **(B)** Regional association plot of the lead SNP rs4832789 at TLR10-TLR1-TLR6 locus in the African cohort. **(C)** Veen diagram showing the overlap between cQTLs of LPS induced IL-1β concentration in the 500FG cohort and the BCG-unvaccinated group. **(D)** Venn diagram depicting the overlap between cQTLs of IL-6 after stimulation with LPS for the 300BCG cohort and the BCG group from the African cohort. **(e)** Veen diagram showing the overlap between cQTLs of TNF-a after stimulation with LPS in the 300BCG cohort and the BCG group from the African cohort. SNPs with nominal p-value threshold of 5×10^{-2} were considered for the overlapping analysis.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TableS1.xlsx
- TableS2.xlsx
- TableS4.csv
- SupplementalInformation.pdf