Risk of nontyphoidal Salmonella bacteraemia in African children is modified by STAT4

Gilchrist, James J.

2018-03-09


http://hdl.handle.net/10138/233913
https://doi.org/10.1038/s41467-017-02398-z

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.
Risk of nontyphoidal *Salmonella* bacteraemia in African children is modified by *STAT4*

James J. Gilchrist1,2, Anna Rautanen1, Benjamin P. Fairfax1, Tara C. Mills1, Vivek Naranbhai1, Holly Trochet1, Matti Pirinen1,3, Esther Muthumbi4, Salim Mwarumba4, Patricia Njuguna4, Neema Mturi4, Chisomo L. Msefula5,6, Esther N. Gondwe5, Jenny M. MacLennan5,7, Stephen J. Chapman1,8, Malcolm E. Molyneux5, Julian C. Knight1, Chris C.A. Spencer1, Thomas N. Williams4,9, Calman A. MacLennan5,10, J.Anthony G. Scott4,11 & Adrian V.S. Hill1,10

Nontyphoidal *Salmonella* (NTS) is a major cause of bacteraemia in Africa. The disease typically affects HIV-infected individuals and young children, causing substantial morbidity and mortality. Here we present a genome-wide association study (180 cases, 2677 controls) and replication analysis of NTS bacteraemia in Kenyan and Malawian children. We identify a locus in *STAT4*, rs13390936, associated with NTS bacteraemia. rs13390936 is a context-specific expression quantitative trait locus for *STAT4* RNA expression, and individuals carrying the NTS-risk genotype demonstrate decreased interferon-γ (IFNγ) production in stimulated natural killer cells, and decreased circulating IFNγ concentrations during acute NTS bacteraemia. The NTS-risk allele at rs13390936 is associated with protection against a range of autoimmune diseases. These data implicate interleukin-12-dependent IFNγ-mediated immunity as a determinant of invasive NTS disease in African children, and highlight the shared genetic architecture of infectious and autoimmune disease.
ontophysoidal *Salmonella* (NTS) is a common, and frequently fatal, cause of bacteraemia in children and HIV-infected adults in sub-Saharan Africa. NTS is estimated to cause 1.9 million episodes of invasive disease resulting in 390,000 deaths annually in Africa. This burden of disease reflects inadequate control strategies for NTS in sub-Saharan Africa, with expanding antibiotic resistance, and no anti-NTS vaccine available for use in humans. There is marked inter-individual variation in susceptibility to invasive NTS (NTS) disease. Both heritable and acquired risk factors for disease are well documented, with sickle cell disease, HIV infection, malnutrition and malaria all contributing to susceptibility to iNTS disease in African children.

Genome-wide association studies (GWAS) have been highly successful in identifying common genetic variants associated with hundreds of disease traits in diverse populations, including infectious diseases in African populations. We have previously conducted a GWAS of all-cause bacteraemia in Kenyan children, in which we were unable to identify genetic determinants of invasive bacterial disease beyond sickle cell disease, an association we observed across a broad range of bacterial pathogens, including NTS. Despite a smaller sample size, an analysis of pneumococcal bacteraemia within the same collection, however, did demonstrate evidence of novel genetic associations at genome-wide significance, suggesting that in this context the loss of study power resulting from a substantial reduction in case numbers is more than offset by the provision of a more precisely defined phenotype. This observation and the lack of unbiased, population-based data assessing host genetic susceptibility to iNTS disease in African populations led us to conduct a GWAS of NTS bacteraemia in the Kenyan collection with additional replication in a Malawian case–control sample collection.

In this study we report a GWAS of susceptibility to NTS bacteraemia in African children. We identify a susceptibility locus for NTS bacteraemia at signal transducer and activator of transcription 4 (STAT4), rs13390936. This NTS-associated genetic variation is a context-specific, expression quantitative trait locus (eQTL) for STAT4 RNA expression in stimulated leucocytes, where the NTS-associated genotype results in reduced STAT4 RNA expression. We demonstrate that carriage of the NTS-associated genotype at STAT4 is associated with reduced IFNγ protein production in stimulated natural killer (NK) cells, and with reduced serum IFNγ levels in African children with iNTS infection. Delivery of an effective anti-NTS vaccine is urgently required. By establishing the importance of IFNγ-mediated immunity in the control of NTS infection in African children, our data highlight the need to consider the induction of cell-mediated immunity in the design and evaluation of such a vaccine.

**Results**

**NTS GWAS and replication analysis.** Following quality control (QC; Supplementary Tables 1 & 2) and genome-wide imputation, 2847 Kenyan samples (180 cases, 2677 controls) were included in the GWAS discovery analysis of 5,585,198 (additive model) and 4,669,480 (genotypic model) autosomal single-nucleotide polymorphisms (SNPs). Inspection of quantile–quantile plots (Supplementary Fig. 1) and the genomic control inflation factors (λ\textsubscript{additive} = 1.01, λ\textsubscript{genotypic} = 1.01) indicate that inclusion of the four major principal components (Supplementary Fig. 2) as covariates in the analysis adequately controls for population substructure. In the discovery analysis, 67 SNPs at 16 loci were suggestively associated (P < 1 × 10\textsuperscript{-6}) with NTS bacteraemia (Supplementary Table 3 and Supplementary Fig. 3). SNPs in each of these associated loci were selected for genotyping in Kenyan and Malawian replication samples. Imputation accuracy of SNPs taken forward for replication genotyping was confirmed in 930 of the Kenyan discovery samples (33% of the discovery samples; 180 cases, 750 controls) by Sequenom MASSArray (mean r\textsuperscript{2} = 0.97, Supplementary Table 4).

Following the discovery analysis, an additional 1374 Kenyan samples (38 cases, 1336 controls) and 489 Malawian samples (150 cases, 339 controls) were included in the replication genotyping (Supplementary Table 4). In the replication analysis, an intronic SNP (rs13390936, chr2:191954816) in the STAT4 gene demonstrated evidence of association with NTS bacteraemia in both the Kenyan and Malawian replication samples (Fig. 1 and Table 1). Comparison of the NTS association at rs13390936 under alternative genetic models demonstrated that the observed association is best explained by a recessive model (Supplementary Table 5), with carriage of the minor, TT genotype being associated with increased risk of NTS bacteraemia. Re-analysis of the STAT4 locus in the Kenyan discovery samples (181 cases, 2759 controls; this includes samples excluded from the main analysis for relatedness) using a linear mixed model, to account for population structure and relatedness, confirms the association of rs13390936 genotype with NTS bacteraemia (recessive model; P = 7.14 × 10\textsuperscript{-7}).

Re-imputation of the STAT4 region, with 1000G Phase 3 as a reference panel, confirmed rs13390936 as the variant most strongly associated with NTS in both the Kenyan discovery and replication samples (Fig. 1). A Bayesian analysis of the re-imputed data demonstrated that a set of seven SNPs, including rs13390936, has a >95% probability of containing the causal variant, under the assumption that we have accurately imputed the single causal SNP in our data (Fig. 1 and Supplementary Table 6). To more fully confirm imputation accuracy at rs13390936 in the Kenyan discovery samples, we re-genotyped rs13390936 in the complete sample collection with high-resolution melting curve analysis (HRMA). HRMA and imputation concordance at rs13390936 confirmed imputation accuracy (r\textsuperscript{2} = 0.97), and confirmed the association with NTS bacteraemia (Table 1). In the combined analysis of all samples, under a recessive model, the association of rs13390936 with NTS bacteraemia exceeds genome-wide significance (fixed-effects meta-analysis, P = 8.62 × 10\textsuperscript{-10}, odds ratio (OR) 7.61, 95% confidence interval (CI), 3.98–14.55).

---

**Fig. 1** Association plot of NTS bacteraemia susceptibility at the STAT4 region in Kenyan children. NTS association at the STAT4 region under the recessive model in Kenyan discovery samples (n = 180 cases, 2677 controls). rs13390936 is highlighted in blue. SNPs are coloured according to the strength of linkage disequilibrium (r\textsuperscript{2}) to rs13390936. SNPs in a credible SNP set, which includes the causal variant with >95% probability, are ringed with black circles.
Established NTS-risk factors and rs13390936. Malaria, HIV infection and malnutrition are important, acquired risk factors for NTS bacteraemia in African children\(^5\). To address whether the observed association of rs13390936 with NTS bacteraemia is independent of these acquired risk factors, we conducted a Bayesian analysis comparing models of association at rs13390936 with NTS bacteraemia with and without acquired risk factors for NTS in the Kenyan discovery samples (Supplementary Fig. 4). In that analysis, the most probable models are those in which rs13390936 is associated with susceptibility to NTS bacteraemia with the same effect size in children with and without each of malaria and malnutrition. The numbers of HIV-infected children in the Kenyan discovery samples (\(n = 24\), none of which carry the rs13390936:TT genotype) are too small to permit a stratified analysis. To ensure HIV infection did not confound the observed association between rs13390936 genotype and NTS bacteraemia, we repeated the association analysis at rs13390936 including only HIV-uninfected NTS cases (\(n = 97\)). In that analysis, the observed effect size (recessive model; OR 8.49, 95% CI, 2.37–24.28) is in keeping with that observed in the association analysis including all samples (Table 1), further supporting the association between rs13390936 genotype and NTS bacteraemia.

Control samples used in the Kenyan discovery analysis are taken from a birth cohort, and acquired risk factor data are not available to perform a regression analysis of NTS bacteraemia risk adjusted for malaria, malnutrition and HIV. To address this limitation, we fitted a regression model of rs13390936 association with NTS bacteraemia in the Malawian replication samples, including HIV infection, malaria parasitaemia and severe malnutrition as covariates. In that model, including 396 Malawian children (109 cases, 287 controls) for which complete covariate data are available, rs13390936 is associated with NTS bacteraemia under a recessive model (logistic regression, \(P = 9.96 \times 10^{-5}\); OR 7.74, 95% CI, 6.18–9.30) independent of HIV, malaria and malnutrition. This lack of confounding by acquired risk factors for NTS is in keeping with the absence of observed association between genetic variation at the \(\text{STAT4}\) locus and HIV and malaria-related phenotypes in GWAS published to date\(^1,11,12\). To further address the possibility of confounding secondary to malaria in Kenyan and Malawian children, we used publicly available GWAS summary statistics from the MalariaGEN consortium GWAS of severe malaria in Gambian, Kenyan and Malawian children\(^12\) (total 5126 cases, 5287 controls; Gambia 2428 cases, 2489 controls; Malawi 1193 cases, 1321 controls; Kenya 1505 cases, 1477 controls) to assess for any evidence of association between rs13390936 and severe malaria. In that dataset, rs13390936 is not associated with severe malaria in African children (logistic regression; additive model, \(P = 0.31\), OR 0.96, 95% CI, 0.88–1.04; dominant model \(P = 0.47\), OR 0.97, 95% CI, 0.88–1.06; recessive model \(P = 0.15\), OR 0.76, 95% CI, 0.52–1.11; heterozygous advantage model \(P = 0.98\), OR 0.98, 95% CI, 0.89–1.08).

Rates of HIV and malaria co-infection, but not malnutrition, among Kenyan and Malawian children with NTS included in the study are significantly different (Supplementary Table 7). The observed effect size of the rs13390936 association with NTS bacteraemia is lower in Malawian children than that observed in Kenyan children, albeit with no evidence of inter-study heterogeneity of effect (Table 1). While there is no evidence that the observed association between rs13390936 genotype and NTS bacteraemia is the result of confounding secondary to acquired risk factors for NTS, we further investigated whether there is evidence that the effect of rs13390936 genotype on NTS disease risk is modified by HIV, malaria or malnutrition. We fitted regression models of NTS bacteraemia and rs13390936 association, with and without interaction terms between rs13390936 and each acquired risk factor, in the same set of Malawian children. In that analysis, there is no evidence for interaction between carriage of the rs13390936:TT genotype and HIV status (likelihood ratio test, \(P = 0.81\)), malaria co-infection (likelihood ratio test, \(P = 0.07\)) or malnutrition (likelihood ratio test, \(P = 0.82\)) in NTS bacteraemia risk.

The sickle cell locus (rs334) has been previously demonstrated to be associated with increased risk of NTS bacteraemia in Kenyan children\(^5\) (sickle cell disease—HbSS), and with protection against malaria, and as a consequence, bacteraemia\(^7\) (sickle cell trait—HbAS). In the Kenyan discovery samples (164 cases, 2342 controls), sickle cell disease is associated with increased risk of NTS bacteraemia (logistic regression, \(P = 8.30 \times 10^{-5}\); OR 4.89, 95% CI, 2.10–10.40), and there is no statistically significant evidence (logistic regression, \(P = 0.65\)) for an effect of sickle cell trait (HbAS) on risk of NTS bacteraemia (Supplementary Fig. 5). As carriage of HbSS is associated with risk of NTS bacteraemia, and carriage of HbAS is associated with a key risk factor for NTS bacteraemia (malaria), we sought to assess whether the observed association at the \(\text{STAT4}\) locus is independent of genotype at rs334. In regression models conditioned on genotype at rs334 (Supplementary Fig. 5), the observed association at rs13390936 with NTS bacteraemia persists when conditioned on HbAS (OR 7.16, 95% CI, 2.68–17.41) and HbSS (OR 8.08, 95% CI, 2.71–21.89).

Table 1 Association of rs13390936 with NTS bacteraemia in African children

<table>
<thead>
<tr>
<th>Population</th>
<th>Case Freq.</th>
<th>TT/AT/AA</th>
<th>Control Freq.</th>
<th>TT/AT/AA</th>
<th>Genotypic model</th>
<th>Recessive model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kenya (discovery)</td>
<td>0.151</td>
<td>8/40/138</td>
<td>0.107</td>
<td>15/529/2068</td>
<td>1.76 × 10^{-5}</td>
<td>3.33 × 10^{-6}</td>
</tr>
<tr>
<td>Kenya (replication)</td>
<td>0.146</td>
<td>3/4/30</td>
<td>0.106</td>
<td>12/254/1053</td>
<td>1.83 × 10^{-3}</td>
<td>6.54 × 10^{-4}</td>
</tr>
<tr>
<td>Malawi</td>
<td>0.15</td>
<td>6/31/106</td>
<td>0.136</td>
<td>3/86/249</td>
<td>0.039</td>
<td>0.026</td>
</tr>
<tr>
<td>Combined</td>
<td>0.15</td>
<td>8/40/138</td>
<td>0.107</td>
<td>15/529/2068</td>
<td>1.76 × 10^{-5}</td>
<td>3.33 × 10^{-6}</td>
</tr>
</tbody>
</table>

rs13390936 is directly genotyped in all samples by HRMA. Freq, minor allele frequency. OR odds ratio, CI confidence interval. HRMA high-resolution melting curve analysis
evidence of differential LD at the NTS-associated locus between the Kenyan discovery GWAS samples and the European samples used in the eQTL analyses (Supplementary Fig. 6). In naïve immune cells, rs13390936 genotype was not significantly associated with STAT4 RNA expression, which was most abundant in NK cells (Fig. 2a). By contrast, rs13390936 was a significant expression quantitative trait locus (eQTL) for STAT4 RNA expression in monocytes following stimulation with lipopolysaccharide for 2 (linear regression, \( P = 4.52 \times 10^{-5} \)) or 24 h (linear regression, \( P = 2.65 \times 10^{-5} \)), or IFN\(\gamma\) for 24 h (linear regression, \( P = 2.63 \times 10^{-5} \)), with carriers of the NTS-risk genotype expressing the least STAT4 in each stimulation condition (Fig. 2b).

rs13390936 is associated with IFN\(\gamma\) production in NK cells. There is no established role for STAT4 in monocytes during

- Figure a shows normalized STAT4 expression in different cell types with different genotypes at rs13390936.
- Figure b illustrates the normalized STAT4 expression following different stimulations with lipopolysaccharide or IFN\(\gamma\), showing significant associations with gene expression.
- Figure c displays the distribution of NK and CD3+ cells, highlighting the protective AA homozygote and susceptible TT homozygote genotypes.
- Figure d presents the percentage of IFN\(\gamma\)-produced NK cells, indicating significant differences between genotypes.
- Figure e shows serum IFN\(\gamma\) levels with a box plot illustrating the distribution and significance of the differences among genotypes.
Salmonella infection. By contrast, a major consequence of STAT4 activity is IFNγ production in NK cells and CD4+ T cells. To better understand the NTS-association phenotype of rs13390936 at the protein level in these cells, we recruited 54 healthy European adults of known genotype at rs13390936 (36 donors carrying the NTS-protective AA genotype; 18 carrying the NTS-susceptible TT genotype) via a genotype-selectable bioresource (Oxford Biobank). In these individuals, we investigated genotype-dependent IFNγ production in naïve and stimulated NK cells and CD4+ T cells by intracellular cytokine staining and flow cytometry. IFNγ production was quantified at 6 h in naïve cells and in cells following stimulation with IL-12 and NTS (Fig. 2c). Donors with the NTS-risk rs13390936 genotype produced a reduced fraction of IFNγ+ NK cells following stimulation with IL-12 (linear regression, \( P = 1.91 \times 10^{-5}, \text{mean}_{AA} = 1.54\%\), \( \text{mean}_{TT} = 0.91\%\)) or NTS (linear regression, \( P = 0.049, \text{mean}_{AA} = 17.09\%, \text{mean}_{TT} = 10.76\%\)) (Fig. 2d). We observed this genotype-dependent effect in NK cells but not in CD4+ T cells (Supplementary Fig. 7).

rs13390936 is associated with IFNγ in acute NTS bacteraemia. To extend our understanding of the IFNγ protein phenotype of rs13390936 to African children and to determine the effect of rs13390936 on total IFNγ production during episodes of NTS bacteraemia, we measured serum IFNγ levels in Malawian children admitted to hospital with NTS bacteraemia (n = 106) during the acute phase of disease and correlated this with genotype at rs13390936 (Fig. 2c). There is no evidence for association between genotype at rs13390936 and co-morbidity, clinical features or outcome in these children (Supplementary Table 8). In keeping with our results in stimulated NK cells in European adults, serum IFNγ levels in Malawian children with acute NTS bacteraemia are dependent on rs13390936 genotype (linear regression, \( P = 0.016\), independent of age, sex, malnutrition, HIV status and malaria co-infection. Malawian children carrying the NTS-risk genotype have reduced circulating IFNγ concentrations (mean\(_{AA}\) = 613.9 pg/ml, mean\(_{TT}\) = 281.4 pg/ml). Given the importance of CD4+ T cells as a source of IFNγ, to ensure HIV infection did not confound this association, we also performed this analysis including only HIV-uninfected children (n = 61). This confirms the association between serum IFNγ levels during acute NTS bacteraemia and rs13390936 genotype (linear regression, \( P = 0.038\), Supplementary Fig. 8). In HIV-infected children (n = 43), there is no evidence for association between serum IFNγ levels during acute NTS bacteraemia and rs13390936 genotype (linear regression, \( P = 0.23\), Supplementary Fig. 8). The lack of association between rs13390936 genotype and serum IFNγ may simply reflect a lack of power in HIV-infected children, or may reflect other sources of variation in IFNγ response in HIV-infected children (e.g. CD4+ T cell count).

Bacteraemia risk conferred by rs13390936 is specific to NTS. To investigate whether the susceptibility to bacteraemia associated with rs13390936 is specific to NTS or results in susceptibility to other causes of bacteraemia, we conducted a Bayesian analysis comparing models of association at rs13390936 with the major causes of bacteraemia in the Kenyan discovery samples (Fig. 3a, b). In that analysis, the most probable model is one in which rs13390936 is associated with susceptibility to NTS bacteraemia but not bacteraemia caused by other pathogens. This suggests that bacteraemia susceptibility conferred by genetic variation at STAT4 is specific to NTS in this context.

rs13390936 is associated with risk of autoimmune disease. Previously published genetic association studies have implicated variation at other loci within STAT4 in the pathogenesis of several autoimmune diseases18–20. To assess the evidence for any shared genetic aetiology between NTS bacteraemia and autoimmune disease at rs13390936, we estimated effect sizes and 95% confidence intervals for association at the NTS-associated locus with eight autoimmune diseases from summary statistics in the ImmunoBase database21. We observed a protective association between the NTS-risk allele and a range of autoimmune diseases, with the strongest evidence of association seen with type 1 diabetes, coeliac disease, juvenile idiopathic arthritis, primary biliary cirrhosis and rheumatoid arthritis (Fig. 3c). The direction of this effect suggests that, in contrast to its role in NTS-risk, increased STAT4 expression and IFNγ production enhances the risk of a range of autoimmune diseases.

Discussion
In this study, we demonstrate that genetic variation in STAT4 is a determinant of NTS bacteraemia in African children. The NTS-associated SNP in STAT4 (rs13390936) is intronic, and is associated with disease under a recessive model with a large effect size. In common with many examples of trait-associated genetic variation identified by GWAS13, rs13390936 acts as, or is in LD with, a regulatory determinant of STAT4 expression. Regulatory genetic variation is commonly specific to cell type and context14, and this is the case for rs13390936 and STAT4. The role of rs13390936 as an eQTL for STAT4 RNA in immune cells is only evident following innate immune stimulation. Moreover, the downstream
effects of that regulatory variation at the protein level exhibit both cell type and context specificity, with rs13390936 genotype modulating IFNγ production capacity in NK cells, but not CD4+ T cells. We further demonstrate that carriage of the NTS-risk allele at rs13390936 associates with reduced IFNγ levels in the serum of African children during acute iNTS disease.

STAT4 is a member of the STAT family of transcription factors. In NK cells and CD4+ T cells, STAT4 phosphorylation in response to IL-12 signalling results in type 1 T helper (TH1) differentiation and IFNγ production (Fig. 4). The importance of this pathway in the control of NTS infection in humans has been established by studies of a group of genetically heterogeneous, rare primary immunodeficiencies, characterized by extreme susceptibility to poorly pathogenic mycobacteria and NTS, collectively designated Mendelian susceptibility to mycobacterial disease (MSMD). While mutations in STAT4 have not been reported as a cause of MSMD, the described MSMD loci are all in genes with roles in IL-12-dependent IFNγ production and IFNγ-mediated immunity (Fig. 4).

The identification of genetic variation modifying IFNγ production in NK cells, but not CD4+ T cells, as a risk factor for NTS bacteremia is noteworthy. These observations are in keeping with findings in the mouse model of NTS infection, in which NK cells have been shown to have a protective role in anti-Salmonella immunity, in an IFNγ-dependent manner. The observed association between NTS-associated genetic variation, NK cell IFNγ production and serum IFNγ production in acute NTS bacteremia suggests that NK cells are an important source of IFNγ in NTS infections in African children. The lack of a similar association between NTS-associated genetic variation and CD4+ T cell IFNγ production does not preclude an important role for CD4+ T
cell-derived IFNγ in anti-Salmonella immunity in African children. It is, however, consistent with the observation that acquisition of Salmonella-specific CD4+ T cells in Malawian children does not result in effective anti-Salmonella immunity.\(^{32}\)

A striking feature of MSMD is the narrow and specific range of pathogens to which affected individuals are susceptible. In keeping with this, we demonstrate that bacteraemia susceptibility conferred by genetic variation at STAT4 is specific to NTS among the most frequent causes of bacteraemia in Kenyan children. By contrast, we observe a protective effect of the NTS-risk allele at rs13390936 across a range of autoimmune diseases. The observation that NTS shares a genetic susceptibility locus with a variety of autoimmune diseases is consistent with previous examples of a shared genetic architecture between autoimmune and infectious diseases\(^{33}\), and supports the hypothesis that selection pressure imposed by infectious agents has impacted the evolution of autoimmune disease.

A consequence of our use of a birth cohort as the control population for the Kenyan samples in this study is that we were unable to correct for transient risk factors (e.g. malaria and malnutrition) in the discovery analysis. We demonstrate that risk of NTS bacteraemia conferred by rs13390936 genotype is independent of malaria, malnutrition and HIV status in Malawian children. It is important to recognize, that while our analysis provides no evidence for acquired NTS-risk factors modifying the effect of rs13390936 genotype on risk of NTS disease, this study has limited power to detect such an interaction. It will therefore be important to establish in future studies, whether the association between rs13390936 and risk of NTS bacteraemia is observed consistently across populations of African children with diverse rates of HIV, malaria and malnutrition. The limited sample size, in particular with respect to case numbers, restricted this analysis to the contribution of common genetic variation (minor allele frequency >10%) to risk of NTS bacteraemia. Larger sample sizes will be required to assess the contribution of less common genetic variants, and variants with smaller effect sizes, on the risk of NTS bacteraemia in African children.

Taken together, these results identify regulatory genetic variation in STAT4 as a key determinant of susceptibility to NTS bacteraemia in African children. We identify IL-12-dependent IFNγ production in NK cells as a mechanism by which genetic variation in STAT4 modifies risk of NTS disease. In future studies it will be important to establish whether NTS-associated genetic variation also perturbs Th17 responses and IFNγ production in other cell types. Our data have important implications for the design of an anti-NTS vaccine for African children, supporting the development of vaccine candidates that induce robust cell-mediated immunity to NTS.

**Methods**

**NTS GWAS study cohorts.** For the Kenyan study cohort, children (under 13 years of age) presenting to Kilifi District Hospital, Kenya between 1 August 1998 and 30 October 2010 with all-cause bacteraemia were recruited.\(^{10}\) Blood samples for bacterial culture (BACTEC 9050, Becton Dickinson) were taken from every child – one from each child with a new bacteraemia episode and a second from children admitted to hospital with fever. Cases were defined as children under the age of 12 months with a new bacteraemia episode and an age-matched control. For the Malawian study, children (under 13 years of age) admitted to the Queen Elizabeth Central Hospital in Blantyre, Malawi between 1 January 2007 and 30 April 2008 were recruited. Cases were defined as children under the age of 12 months with a new bacteraemia episode and an age-matched control. For the Malawian study, children (under 13 years of age) admitted to the Queen Elizabeth Central Hospital in Blantyre, Malawi between 1 January 2007 and 30 April 2008 were recruited. Cases were defined as children under the age of 12 months with a new bacteraemia episode and an age-matched control.

From this collection, we extracted children with NTS bloodstream infection (n = 267) from the bacteraemia cases, and all of the control children (n = 4444) to perform a genome-wide association analysis of NTS bacteraemia in Kenyan children. NTS serotyping was performed according to the Kauffman–White scheme using commercial antisera. As the discovery and replication sample sets had been genotyped using different platforms (see below), we analysed the discovery (cases = 218, controls = 3000) and replication sets separately (n = 49, controls = 1444). Demographics of study cases and controls in both the discovery and replication sample sets are presented in Supplementary Table 9.
frequencies of NTS serotypes and co-morbidities in discovery and replication case samples is presented in Supplementary Table 7. Following explanation of the study, written informed consent was obtained from the parent or guardian of each child included in the study. Ethical approval for study was obtained from the Kenya Medical Research Institute (KEMRI) National Scientific Steering and Research Committees and the Oxford Tropical Research Ethics Committee.

For the Malawian replication study, case samples (n = 150) were recruited from children following admission to Queen Elizabeth Central Hospital, Blantyre, Malawi in 2006. Bacterial culture of blood samples was performed using a BacT/Alert 3D system (bioMérieux) in all children admitted with suspected sepsis. Blantyre, Malawi in 2006. Bacterial culture of blood samples was performed using a

recruited from children following admission to Queen Elizabeth Central Hospital, Blantyre, Malawi between October and December 2005, and September and November 200634. Unrelated, healthy children under 16 years of age were eligible for inclusion in the study. Following explanation of the study, written informed consent was obtained from the parent or guardian of each child included in the study, and a buccal swab was taken for DNA extraction. Ethical approval for the study was granted by the College of Medicine Research and Ethics Committee, College of Medicine, University of Malawi.

NTS discovery GWAS. Genomic DNA was extracted using QIAGEN DNA blood mini kits (Qiagen). Extracted genomic DNA was whole-genome amplified with GenomiPhi (GE Healthcare) according to the manufacturer’s instructions. Before whole-genome genotyping, DNA quality was assessed; DNA concentration was assessed using PicoGreen (Invitrogen), agarose electrophoresis was used to detect DNA degradation and QC genotypes were assessed using the Sequenom iPLEX assay35. Genome-wide genotyping of Kenyan discovery cases and controls was performed with an Affymetrix SNP 6.0 chip. SNP QC filters were applied as follows: minor allele frequency (MAF) <1%, info <975, Hardy–Weinberg equilibrium (HWE) P < 1×10−5, plot efficiency <10% and SNP missingness <2%. Following SNP and sample QC (Supplementary Tables 1 & 2) 787,861 autosomal SNPs from 1536 bacteria (including 180 NTS) case and 2677 control individuals were taken forward for genome-wide imputation with SHAPEIT236 and IMPUTE237 using all available populations in 1000G Phase3 as a reference panel. The first four principal components of the genome-wide genotyping data differentiate self-reported ethnicity (Supplementary Fig. 2) and were included in the NTS association analysis to control for underlying population structure. Association analysis was performed using the frequentist score test (accounting for imputed genotype uncertainty) in SNPTEST238 under additive and genotypic models of association. Following association analysis, genotyped SNPs nominally associated (P < 1×10−5) with NTS were identified with SNPs and with poor cluster separation excluded from further analysis. Further SNP QC exclusions were applied as follows: model-specific imputation info score <0.8 and Hardy–Weinberg equilibrium P < 1×10−5. Given the study sample size and limited power to replicate associations at less common SNPs (Supplementary Fig. 10), we further excluded SNPs with MAF <10% from the analysis.

NTS GWAS replication. SNPs in 16 genomic regions putatively associated with NTS bacteriaemia (P < 1×10−8) in the Kenyan discovery samples were genotyped in the Malawian replication samples using the Sequenom MassARRAY platform. Inspection of cluster plots (TYPER4.0) revealed good cluster separation at 25 SNPs, all of which having a call rate in excess of 95%. Samples were excluded with call rates <80%. Following QC, 416 Malawian samples (135 cases, 281 controls) and 318 Kenyan samples (36 cases, 282 controls) were included in the replication analysis. The Sequenom MASSArray platform was also used to confirm imputation accuracy in 930 Kenyan discovery samples (180 cases, 750 controls) at loci included in the replication analysis (Supplementary Table 4).

The multiplexed Sequenom MASSArray genotyping reaction chosen for replication did not include the STAT4 region SNP rs13390936 most significantly associated with NTS in the Kenyan discovery analysis. This SNP was therefore genotyped by HRMA, using the following PCR primers and an unlabelled probe with a 3’-amino-C7 modification: TAGTAGGGCCTAAGTAAATTTGGGAC (F), GCCCTACAGTGTCTCCATCATCT (R), GTGATGACTCTTGACAATATTTATATTTACAA (probe). Ten microlitres of PCR reactions (1 μl 10X PCR buffer, 1 μl 25 mM MgCl2, 1 μl dNTPs (2 mM each), 0.05 μM forward primer (10 μM), 0.025 μM reverse primer (100 μM), 0.025 μM unlabelled probe (100 μM), 1 μl LC Green dye, 0.06 μl HotStarTaq DNA polymerase (Qiagen), 5 μl water and 0.5 μl template DNA (25 ng/μl) were cycled (95 °C for 10 min, 55 cycles of 30 s at 95 °C, 56 °C and 72 °C), under mineral oil. Following the PCR reactions, the samples were denatured (95 °C) for 10 min, cooled and cooled to room temperature for 30 min. High-resolution melting was then performed on a LightScanner (Idaho Technology) at 0.1 °C/s increments between 45 °C and 98 °C. Melt curves were analysed with LightScanner Call-IT 2.0 (Idaho Technology) using derivative normalized melting plots between 62 °C and 72 °C. Supplementary Table 4. Genotyping of rs13390936 was performed in total RNA from naïve cells, and protein phenotypes for rs13390936 in immune cell subsets without pre-phasing with IMPUTE2 using 1000G Phase 3 as a reference panel. Imputation was performed with 1 Mb buffer regions and template haplotypes for phasing increased to 200. This region was also imputed in the Kenyan replication samples from Affymetrix genotyping data using the same IMPUTE2 settings. Following QC of the Kenyan discovery case and control samples, association analysis of the re-imputed genotypes was performed using the frequentist score methods in SNPTEST2 under a recessive model of association. Re-imputed genotypes were further used to test for association at STAT4, under a recessive model of association, using a linear mixed model to account for population structure and sample relatedness in GEMMA39. In this analysis, samples were not excluded according to pairwise relatedness, but QC thresholds were otherwise as described for the Kenyan discovery samples.

We used a Bayesian approach to identify a set of SNPs with 95% probability of containing the causal locus at the NTS-associated region41,42. Approximate Bayes’ factors35,36 were calculated under a recessive model for each SNP in the NTS-associated region with a prior distribution of N(0,1). SNPs with Bayes’ factors were considered equally likely to be the causal variant a priori. A set of SNPs with 95% probability of containing the causal SNP was defined as the smallest number of SNPs for which the summed posterior probabilities exceed 0.95.

RNA expression quantitative trait analysis of rs13390936. Using RNA expression and genome-wide genotyping data from published43,44 and unpublished eQTL studies of naive and stimulated primary immune cell subsets in healthy European adults, we correlated rs13390936 genotype with gene expression. CD19+ B cells, CD14+ monocytes and CD63+CD36+ NK cells were separated from peripheral blood mononuclear cells by magnetic activating cell sorting (Miltenyi). CD16+ neutrophils were isolated from granulocytes with CD16+ microbeads. Gene expression was quantified in total RNA from naive cells, and CD107a was quantified with lipopolysaccharide (for 2 or 24 h) or IFNγ (for 24 h), with the Illumina HumanHT-12 v4 BeadChip gene expression array platform. Genome-wide genotyping was performed with the Illumina HumanOmniExpress-12v1.0 Beadchip and imputation undertaken using 1000G phased1 as the reference panel.

Following QC of gene expression and genotyping data as previously described14,24, rs13390936 genotype was correlated with STAT4 RNA expression in each cell type: B cells (n = 279), NK cells (n = 245), neutrophils (n = 101), naive monocytes (n = 414) and stimulated monocytes (LPS 2 h, n = 261; LPS 24 h, n = 322; IFNγ 24 h, n = 367). Random-spline normalized and variance-stabilization transformed STAT4 expression was correlated with genotyped genotypes using linear regression and analysis of variance (ANOVA), including the first principal components of gene expression data in each cell type/condition to account for confounding variation. P values are calculated with F-tests (1 df). Statistical analysis was performed in R.

LD at STAT4 in study populations. LD plots of the STAT4 region in the Kenyan GWAS discovery samples (n = 2857) and the European samples used in the RNA eQTL analyses (n = 421) were constructed in Haploview v4.24. To quantify differential LD between the Kenyan and European samples we used varLD45, quantifying LD variation between the two populations in 50-SNP windows. Raw varLD scores across all of chromosome 2 were standardized, and regions with standard deviation (SD) <10 and |varLD| scores >95th centile (standardized |varLD| >1.86) considered to have evidence of differential LD between the two populations.

Protein phenotypes for rs13390936 in immune cell subsets. In archived genomic DNA from 5911 individuals registered to the Oxford Biobank (NHIR, Oxford Biomedical Research Centre), rs13390936 was genotyped by the Oxford
Biobank laboratory using a TaqMan assay (Applied Biosystems). Fifty-four volunteers (32 females; median age 44 years, range 30–51 years) were recruited to the study according to the model under analysis and within the NTS-protective AA genotype, 18 with the NTS-susceptible TT genotype. The study was approved by the Oxfordshire Research Ethics Committee (COREC reference 06/Q1605/55), and written informed consent was obtained from each volunteer.

Whole blood from each volunteer was collected into Lithium Heparin containing collection tubes (Vacutainer System, Becton Dickinson). One millilitre of aliquots of whole blood from each volunteer were left unstimulated, or stimulated with 10 ng/ml recombinant human IL-12 (catalogue #219-IL, R&D Systems) or 10 μg/ml NTS (Malawian clinical S. Typhimurium isolate D23580) grown to mid-log phase, within 2 h of collection. In addition, 1 ml whole blood aliquots were stimulated with phorbol 12-myristate 13-acetate (100 ng/ml)/ionomycin (1 μg/ml, Sigma) to act as a positive control. Following stimulation, samples were incubated at 37 °C. Following ex vivo stimulation of each sample, intracellular IFN-γ staining was performed. Brefeldin A (Becton Dickinson) was added at 2 h post-stimulation to a final concentration of 2 μg/ml. At 6 h post-stimulation, samples were surface immunostained (20 min, 4 °C) following erythrocye lysis (Versalyse, Beckman Coulter) with allophycocyanin-conjugated anti-CD3 (clone UCHT1, BD Biosciences), fluorescein isothiocyanate-conjugated anti-CD4 (clone RPA-T4, BD Biosciences) and phycoerythrin (PE)-conjugated anti-CD56 (clone CMSSB, eBioscience) monoclonals. Samples were fixed with Fixation Medium A (Invitrogen), before permeabilization with Permeabilization Medium B (Invitrogen) and intracellular staining with PE-Cy7-conjugated anti-IFN-γ monoclonal (clone B27, BD Biosciences) for 20 min at 4 °C.

Immunostained samples were acquired with a BD FACSCanto II and BD FACSDiva® (BD Biosciences), and the data were analysed using Flowjo v10. For each batch, validation across batches was observed prior to stimulus and recruitment to become the same manufacturing lot number. Prior to analysis, CS&T beads of the same lot were used to calibrate the flow cytometer, to ensure equivalent photomultiplier tube voltages between experiments. Proportions of IFN-γ+ NK cells were log transformed prior to analysis and genotype-phenotype associations tested with linear regression and ANOVA, adjusted for age and sex. P values were calculated with F-tests (1 d.f.). Transformation did not normalise IFN-γ+ CD4+ T cell proportions, and genotype-phenotype associations were tested with nonparametric Wilcoxon’s rank-sum tests. There was no evidence of significant batch effect (P > 0.05) on IFN-γ+ T cell proportion for either NK cells (linear regression model, adjusted for rs13390936 genotype and stimulation conditions) or CD4+ T cells (Kruskal–Wallis test). There was no evidence of significant inter-genotype inequality of variances (Bartlett test; P > 0.05) of IFN-γ+ NK cell proportions. Statistical analysis was performed in R.

**Protein phenotypes of rs13390936 in acute NTS bacteraemia.** We collected serum samples from a subset of the Malawian replication cases (n = 106). Methods used for bacterial culture and NTS identification and serotyping are as described above. Following identification of NTS in blood culture, the parent or guardian of the child was approached for recruitment to the study. Having obtained written informed consent, a venous blood sample was taken from the child during their acute admission with NTS bacteraemia (median 1 day following collection of the NTS-positive blood culture, range 0–5 days). *Plasmodium falciparum* parasitaemia was tested with thick and thin Giemsa-stained blood films. HIV status was determined using Determine (Abbott Laboratories) and UniGold (Trinity Biotech) rapid tests, and HIV infection was confirmed in children <18 months by PCR. Children with weight-for-age Z-scores >3 standard deviations below WHO median values were classified as being malnourished. ***rs13390936 genotypes were assayed as for the Malawian GWAS replication samples (above). Ethical approval for the study was granted by the College of Medicine Research and Ethics Committee, College of Medicine, University of Malawi. The study was granted by the College of Medicine Research and Ethics Committee, College of Medicine, University of Malawi. The study was approved by the College of Medicine Research and Ethics Committee, College of Medicine, University of Malawi.

**Model comparisons of rs13390936 in autoimmune diseases.** To compare evidence for association at rs13390936 across a range of autoimmune diseases, we downloaded from ImmunoBase summary statistics of disease associations at rs16832339 (in perfect LD with rs13390936 in CEU 1000 Genomes Pilot 1 population; r2 = 1, D = 1) for nine association studies of eight autoimmune diseases (coeliac disease, rheumatoid arthritis, primary biliary cirrhosis, multiple sclerosis, narcolepsy, juvenile idiopathic arthritis, psoriasis and type 1 diabetes—one case-only retrospective study and one meta-analysis of case–control-based and family-based studies) conducted using the ImmunoChip in individuals of European ancestry to a set of 17 selected autoimmune diseases and were included in the analysis. To compare evidence for association at rs13390936 across a range of autoimmune diseases, we downloaded from ImmunoBase summary statistics of disease associations at rs16832339 (in perfect LD with rs13390936 in CEU 1000 Genomes Pilot 1 population; r2 = 1, D = 1) for nine association studies of eight autoimmune diseases (coeliac disease, rheumatoid arthritis, primary biliary cirrhosis, multiple sclerosis, narcolepsy, juvenile idiopathic arthritis, psoriasis and type 1 diabetes—one case-only retrospective study and one meta-analysis of case–control-based and family-based studies) conducted using the ImmunoChip in individuals of European ancestry to a set of 17 selected autoimmune diseases and were included in the analysis. To compare evidence for association at rs13390936 across a range of autoimmune diseases, we downloaded from ImmunoBase summary statistics of disease associations at rs16832339 (in perfect LD with rs13390936 in CEU 1000 Genomes Pilot 1 population; r2 = 1, D = 1) for nine association studies of eight autoimmune diseases (coeliac disease, rheumatoid arthritis, primary biliary cirrhosis, multiple sclerosis, narcolepsy, juvenile idiopathic arthritis, psoriasis and type 1 diabetes—one case-only retrospective study and one meta-analysis of case–control-based and family-based studies) conducted using the ImmunoChip in individuals of European ancestry to a set of 17 selected autoimmune diseases and were included in the analysis. To compare evidence for association at rs13390936 across a range of autoimmune diseases, we downloaded from ImmunoBase summary statistics of disease associations at rs16832339 (in perfect LD with rs13390936 in CEU 1000 Genomes Pilot 1 population; r2 = 1, D = 1) for nine association studies of eight autoimmune diseases (coeliac disease, rheumatoid arthritis, primary biliary cirrhosis, multiple sclerosis, narcolepsy, juvenile idiopathic arthritis, psoriasis and type 1 diabetes—...


ARTICLE NATURE COMMUNICATIONS | DOI: 10.1038/s41467-017-02398-z

NATURE COMMUNICATIONS | www.nature.com/naturecommunications

10

Supplementary Information

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-017-02398-z.

Competing interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

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