Genetic variation in *VAC14* is associated with bacteremia secondary to diverse pathogens in African children

Authors: James J. Gilchrist[1,2], Alexander J. Mentzer[1], Anna Rautanen[1], Matti Pirinen[1,3], Salim Mwarumba[4], Patricia Njuguna[4], Neema Mturi[4], Wellcome Trust Case-Control Consortium 2[+], The Kenyan Bacteraemia Study Group[+], Christopher C. A. Spencer[1], Thomas N. Williams[4,5], J. Anthony G. Scott[4,6], Adrian V. S. Hill[1,7].

Affiliations:

[1]Wellcome Trust Centre for Human Genetics, Roosevelt Drive, University of Oxford, Oxford OX3 7BN, UK.

[2]Department of Paediatrics, University of Oxford, Oxford OX3 9DU, UK.

[3]Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki 00014, Finland.

[4]KEMRI-Wellcome Trust Clinical Research Programme, Kilifi 80108, Kenya.

[5]Department of Medicine, Imperial College, Norfolk Place, London, W2 1PG, UK.

[6]Department of Infectious Disease Epidemiology, London School of Hygiene & Tropical Medicine, Keppel Street, London, WC1E 7HT, UK.

[7]The Jenner Institute, University of Oxford, Old Road Campus Research Building, Oxford, OX3 7DQ, UK.

[+]Memberships of consortia are provided in the Supplementary Information

Performing a genome-wide association study (GWAS) of *S.* Typhi invasion, Alvarez et al. (1) identify a trait-associated single nucleotide polymorphism (SNP), rs8060947, in *VAC14*. rs8060947 is an expression quantitative trait locus (eQTL) for *VAC14* RNA expression, and carriage of the A allele is associated with reduced *VAC14* RNA and protein expression, and increased invasion of *S.* Typhi. VAC14-associated inhibition of *S.* Typhi invasion is mediated by a reduction in host cell membrane cholesterol. Carriage of the A allele at rs8060947 is associated with typhoid fever in Vietnamese individuals (cases = 496, controls = 500; $P = 0.01$, additive odds ratio [OR] = 1.38). The authors further identify a SNP in high linkage disequilibrium with rs8060947, rs8044133, located in a transcription factor binding site, which merits further investigation as the causative SNP.

Alvarez et al. note that cholesterol has been implicated in the invasion and pathogenesis of a wide range of pathogens (2-5). We therefore used GWAS data from Kenyan children with bacteraemia (6) secondary to diverse pathogens (n = 1,536) and healthy control children from the same population (n = 2,677), to explore whether genetic variation at the *VAC14* locus is associated with susceptibility to invasive bacterial infections other than typhoid fever (SI Methods). In Kenyan children, rs8060947 is significantly associated with all-cause bacteremia: $P_{\text{additive}} = 0.02$, OR = 1.11 (95% confidence interval [CI] 1.02-1.22). In keeping with the effect observed in typhoid, carriage of the A allele at rs8060947 increases risk of bacteremia.

To understand whether risk of bacteremia conferred by rs8060947 is shared across pathogens causing bacteremia in Kenyan children, we conducted a Bayesian analysis (SI Methods) comparing models of association at rs8060947 with the major causes of

bacteremia in this population (Fig. 1*A*). The most probable model is one in which rs8060947 is associated with susceptibility to bacteraemia caused by nontyphoidal *Salmonella* (NTS), *S. pneumoniae, E. coli* and *Acinetobacter* species, but not bacteremia caused by other pathogens (Fig. 1*B*). We performed imputation-based mapping of the association at the *VAC14* locus with bacteraemia secondary to NTS, *S. pneumoniae, E. coli* and *Acinetobacter* species (Fig. 1*C*). In that analysis, there is evidence for association between bacteremia secondary to these four pathogens and both rs8060947 ($P_{\text{additive}} = 4.76\text{x}10^{-3}$, OR = 1.17 [95% CI = 1.05-1.30]) and rs8044133 ($P_{\text{additive}} = 4.58\text{x}10^{-3}$, OR = 1.17 [95% CI = 1.05-1.30]).

Alvarez et al. demonstrate that genetic variation in *VAC14* is a determinant of clinical typhoid fever in Vietnamese individuals. Our data expand on this observation, demonstrating that the same risk allele at rs8060947, with a comparable effect size, increases risk of bacteraemia secondary to diverse pathogens in Kenyan children. This observation may reflect a previously unrecognised role for cholesterol in the pathogenesis of diverse bacterial pathogens, or a role for cholesterol in a shared risk factor for these pathogens in this population, e.g. malaria or HIV. Future studies will be required to further define the range of clinical diseases associated with *VAC14*, and to fine-map the genetic signal at *VAC14*.

References

1. Alvarez MI, et al. (2017) Human genetic variation in VAC14 regulates Salmonella invasion and typhoid fever through modulation of cholesterol. *Proceedings of the National Academy of Sciences* 114(37):E7746–E7755.

2. Hacke M, et al. (2015) Inhibition of Ebola virus glycoprotein-mediated cytotoxicity by targeting its transmembrane domain and cholesterol. *Nat Commun* 6:7688.

3. Jutras I, Abrami L, Dautry-Varsat A (2003) Entry of the lymphogranuloma venereum strain of Chlamydia trachomatis into host cells involves cholesterol-rich membrane domains. *Infection and Immunity* 71(1):260–266.

4. Voisset C, et al. (2005) High density lipoproteins facilitate hepatitis C virus entry through the scavenger receptor class B type I. *J Biol Chem* 280(9):7793–7799.

5. Samuel BU, et al. (2001) The role of cholesterol and glycosylphosphatidylinositol-anchored proteins of erythrocyte rafts in regulating raft protein content and malarial infection. *J Biol Chem* 276(31):29319–29329.

6. Kenyan Bacteraemia Study Group, et al. (2016) Polymorphism in a lincRNA Associates with a Doubled Risk of Pneumococcal Bacteremia in Kenyan Children. *Am J Hum Genet* 98(6):1092–1100.

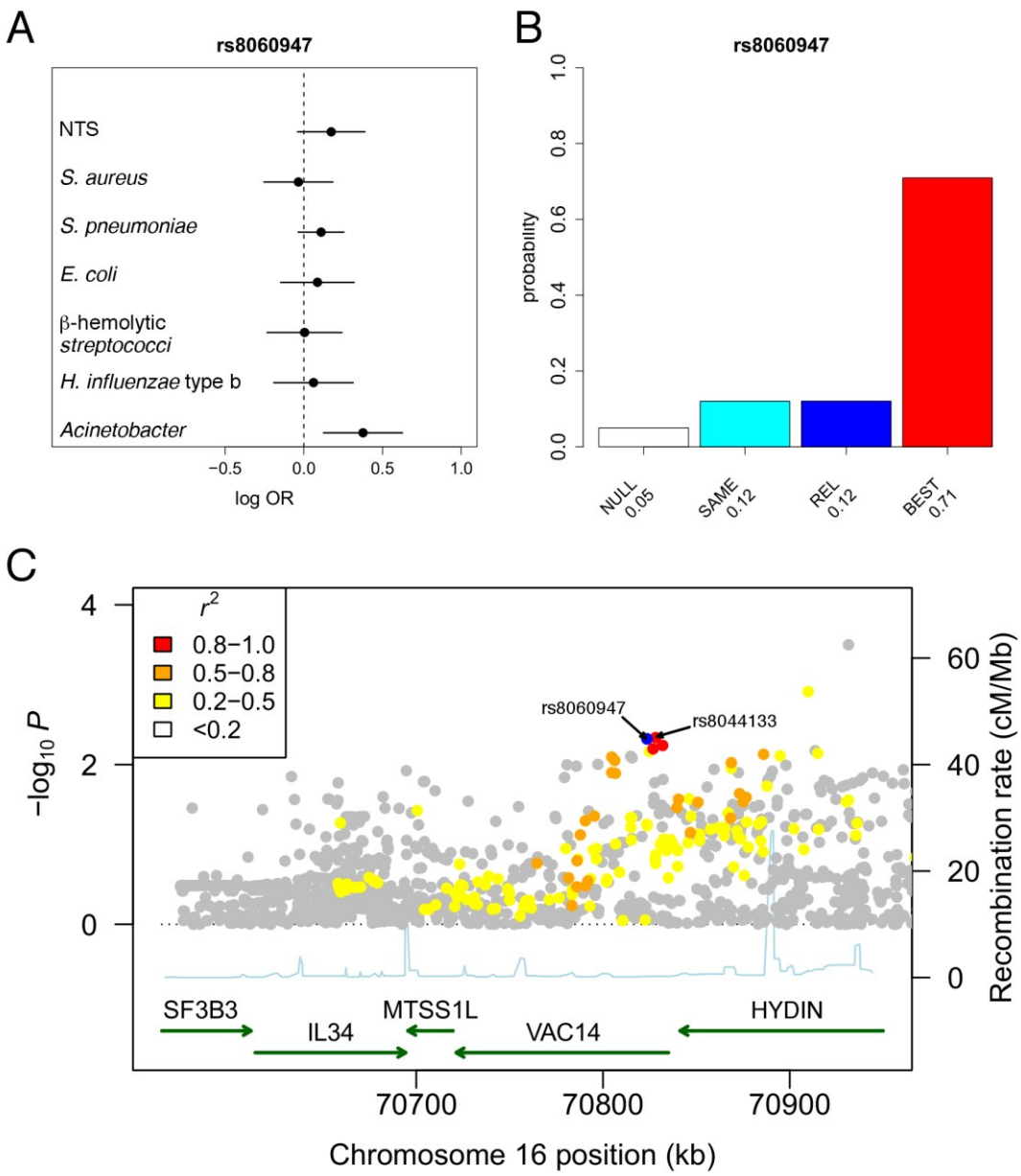


Fig. 1. Association between the *VAC14* locus and major causes of bacteremia in Kenyan children. (*A*) rs8060947 association with the major causes of bacteremia in Kenyan children; nontyphoidal *Salmonella*, n = 180, *S. aureus*, n = 175, *S. pneumoniae*, n = 426, *E. coli*, n = 151, β-hemolytic *Streptococci*, n = 146, *H. influenza* type b, n = 128, *Acinetobacter* species, n = 130. Log-transformed odds ratios and 95% confidence intervals of rs8060947 association (additive model) with each pathogen. (*B*) Posterior probabilities of models of association at rs8060947: NULL, no association with any bacterial pathogen; SAME, the same effect across all

bacterial pathogens; REL, related effects across all bacterial pathogens; BEST, a non-zero effect in bacteremia secondary to nontyphoidal *Salmonella*, *S. aureus*, *S. pneumoniae* and Acinetobacter species alone; this model (highlighted in red) is the most probable (Bayes factor c.f. NULL = 14). (*C*) Association plot of bacteremia susceptibility at the *VAC14* locus. SNPs are colored according to strength of linkage disequilibrium (r^2) to rs8060947. Association statistics are calculated using an additive model including bacteraemia cases secondary to nontyphoidal *Salmonella*, *S. aureus*, *S. pneumoniae* and Acinetobacter species (n = 887) and shared, healthy controls (n = 2,677).

Supporting Information

SI Methods

Genetic association analysis of bacteraemia in Kenyan children at VAC14

The GWAS of all-cause bacteraemia in Kenyan children has been described in detail elsewhere (1). In brief, Kenyan children under the age of 13 years, presenting to Kilifi District Hospital, Kenya, between 1st August 1998 and 30th October 2010, with community-acquired bacteraemia were recruited to the GWAS. Healthy community controls were recruited as a birth cohort from the same population between 1st May 2006 and 30th April 2008. Following DNA extraction and quality control (1), genome-wide genotyping was performed using an Affymetrix SNP 6.0 chip. Genome-wide imputation was performed with SHAPEIT (2) and IMPUTE2 (3) using 1000G Phase 1 as a reference panel. Association analysis was performed under an additive model of association, in SNPTEST2 (4), including the first four principal components of genome-wide genotyping data to correct for population structure.

In this analysis of the *VAC14* locus, we included common (minor allele frequency >0.05), well-imputed SNPs (imputation info score >0.4), with no evidence for departure from Hardy Weinberg equilibrium ($P>1x10^{-10}$). rs8060947 was directly genotyped on the Affymetrix SNP 6.0 chip (Fig. S1). In this population, rs8060947 has a minor allele frequency of 0.44, with no evidence of departure from Hardy Weinberg equilibrium ($P = 0.11$). rs8044133 is imputed (imputation info score = 0.996), and has a minor allele frequency of 0.47, with no evidence of departure from Hardy Weinberg equilibrium ($P = 0.73$). Following quality control, 1,536 cases and 2,677 control samples were included in the analysis.

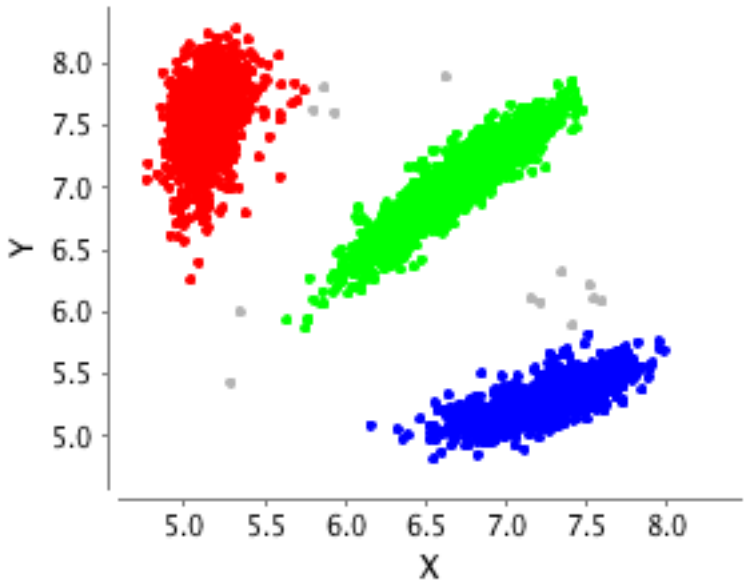
Bayesian comparison of models of association

We compared models of association at candidate loci across the seven most frequently isolated bacterial pathogens (NTS, n = 180; *S. pneumoniae*, n = 426; *S. aureus*, n = 175; *H. influenza* type b, n = 128; β-haemolytic *Streptococci*, n = 146; *E. coli*, n = 151; *Acinetobacter*, n = 130) among cases of bacteraemia in the Kenyan discovery samples in the all-cause bacteraemia GWAS (1). Effect size estimates and 95% confidence intervals were calculated by multinomial logistic regression under a recessive model, using control status and each of the bacterial pathogen subgroups as strata, and the first two principal components as covariates. We considered three models of effect across the bacterial pathogens, defined by the prior distributions on the effect size:

NULL: effect size = 0, i.e. no association with any pathogen.
SAME: effect size ~ N(0,0.2) and fixed between pathogens (ρ=1).
REL: effect size ~ N(0,0.2) and correlated (ρ=0.96), but not fixed, between pathogens.

These three models were considered alongside a fourth model defined as a combination of pathogens with the same non-zero effect (with no effect for other pathogens), for which the data provides support. For each model we calculated approximate Bayes factors (5) and posterior probabilities, assuming each model to be equally likely *a priori*.

SI Figures



SI Fig. 1. Cluster plot of rs8060947 in Kenyan children (n = 4,924).

SI References

1. Kenyan Bacteraemia Study Group, et al. (2016) Polymorphism in a lincRNA Associates with a Doubled Risk of Pneumococcal Bacteremia in Kenyan Children. *Am J Hum Genet* 98(6):1092–1100.

2. Delaneau O, Marchini J, Zagury J-F (2012) A linear complexity phasing method for thousands of genomes. *Nat Methods* 9(2):179–181.

3. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR (2012) Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet* 44(8):955–959.

4. Marchini J, Howie B, Myers S, McVean G, Donnelly P (2007) A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* 39(7):906–913.

5. Wakefield J (2009) Bayes factors for genome-wide association studies: comparison with P-values. *Genet Epidemiol* 33(1):79–86.

Consortia Membership

The Kenyan Bacteraemia study group

Principal Investigators: Adrian V S Hill (Chair), Thomas N Williams, J Anthony G Scott, Stephen J Chapman

Key Personnel: Anna Rautanen, Tara C Mills, Kirk Rockett, Anne W Ndungu, Vivek Naranbhai, Alex W Macharia, Sophie Uyoga, Carolyne Ndila, Neema Mturi, Patricia Njuguna, Shebe Mohammed, James A Berkley, Isaiah Mwangi, Salim Mwarumba, Barnes S Kitsao, Brett S Lowe, Susan C Morpeth, Iqbal Khandwalla

The Kilifi DNA extraction Group: Alex W Macharia, Sophie Uyoga, Herbert Opi, Carolyne Ndila, Emily Nyatichi, Prophet Ingosi, Barnes Kitsao, Clement Lewa, Johnstone Makale, Adan Mohamed, Kenneth Magua, Mary Njoroge, Gideon Nyutu, Ruth Mwarabu, Metrine Tendwa and Thomas N Williams.

The Kilifi Bacteraemia Surveillance Group: Ismail Ahmed, Samuel Akech, Alexander Balo Makazi, Mohammed Bakari Hajj, Andrew Brent, Charles Chesaro, Hiza Dayo, Richard Idro, Patrick Kosgei, Kathryn Maitland, Kevin Marsh, Laura Mwalekwa, Shalton Mwaringa, Charles Newton, Mwanajuma Ngama, Allan Pamba, Norbert Peshu, Anna Seale, Alison Talbert and Thomas N Williams.

Wellcome Trust Case Control Consortium 2

Management Committee: Peter Donnelly (Chair), Ines Barroso (Deputy Chair), Jenefer M Blackwell, Elvira Bramon, Matthew A Brown, Juan P Casas, Aiden Corvin, Panos Deloukas, Audrey Duncanson, Janusz Jankowski, Hugh S Markus, Christopher G Mathew, Colin NA Palmer, Robert Plomin, Anna Rautanen, Stephen J Sawcer, Richard C Trembath, Ananth C Viswanathan, Nicholas W Wood

Data and Analysis Group: Chris C A Spencer, Gavin Band, Céline Bellenguez, Colin Freeman, Garrett Hellenthal, Eleni Giannoulatou, Matti Pirinen, Richard Pearson, Amy Strange, Zhan Su, Damjan Vukcevic, Peter Donnelly

DNA, Genotyping, Data QC and Informatics Group: Cordelia Langford, Sarah E Hunt, Sarah Edkins, Rhian Gwilliam, Hannah Blackburn, Suzannah J Bumpstead, Serge Dronov, Matthew Gillman, Emma Gray, Naomi Hammond, Alagurevathi Jayakumar, Owen T McCann, Jennifer Liddle, Simon C Potter, Radhi Ravindrarajah, Michelle Ricketts, Matthew Waller, Paul Weston, Sara Widaa, Pamela Whittaker, Ines Barroso, Panos Deloukas.

Publications Committee: Christopher G Mathew (Chair), Jenefer M Blackwell, Matthew A Brown, Aiden Corvin, Chris C A Spencer