Genetic Susceptibility to Invasive Nontyphoidal *Salmonella* Disease in African Children

James John Gilchrist

Keble College

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas Term 2016

In memory of my mother

Judy Gilchrist

Abstract

Nontyphoidal *Salmonella* (NTS) causes invasive, and frequently fatal, disease in African children. The burden of disease secondary to NTS reflects inadequacy of *Salmonella*-control strategies in Africa, with expanding antibiotic resistance, and no licensed anti-NTS vaccine. The delivery of improved interventions to prevent, diagnose, and treat invasive NTS (iNTS) infection, will be facilitated by an improved understanding of the biological determinants of susceptibility to iNTS, including host genetic factors.

To identify host genetic determinants of iNTS disease, we performed a GWAS and replication analysis of NTS bacteraemia in African children. This analysis identified and validated a common genetic variant in *STAT4* associated with increased iNTS risk.

To characterise the function of the NTS-associated *STAT4* variant, we utilised a genotype-selectable bioresource of healthy European adults and samples from African children with iNTS disease. In these experiments, the risk genotype at *STAT4* is associated with reduced *STAT4* RNA expression in stimulated leukocytes, and reduced IFNγ production in both *ex vivo* stimulated natural killer cells and in the serum of African children with acute NTS bacteraemia.

To validate genetic variation suggestively associated with NTS bacteraemia in the GWAS, NTS-associated loci with evidence of regulatory function were prioritised for functional characterisation. Using *in vitro* models of intracellular *Salmonella* infection and RNA interference, I characterise the role of a candidate NTS-susceptibility determinant, EVI5L, in *Salmonella* infections.

Finally, applying a pathway enrichment analysis to the NTS bacteraemia GWAS demonstrated that NTS-associated genetic variation in African children is enriched for methionine salvage enzymes. I further investigate the potential for host-pathogen interaction in this pathway, generating and characterising *Salmonella* mutants deficient in methionine metabolism.

Taken together, these data represent the first unbiased assessment of genetic susceptibility to iNTS disease in unselected populations. These results have important implications for the design of *Salmonella*-control strategies for use in Africa.

Acknowledgements

Firstly, I would like to thank my supervisor Adrian Hill. Being mentored by Adrian has been an absolute privilege, and I couldn't have asked for a more intellectually stimulating and supportive environment over the last three years. Over and above contributing to the supervision of my thesis, Cal MacLennan has mentored my early clinical academic career for more than a decade now. Over that time he has been a constant source of advice and friendship, for which I am enormously grateful. He is also responsible for inspiring my initial interest in *Salmonella* in Africa as a medical student. Despite the focus of his lab being in directions other than *Salmonella*, Chris Tang has provided a ready and valuable source of advice throughout my fellowship for which I am very grateful.

A great deal (more than I'd initially intended) of my DPhil fellowship was spent at Imperial College, in David Holden's *Salmonella* pathogenesis group. David could not have been more welcoming and encouraging during my time in London. I am very grateful for the way he integrated me into his group, his patient and thoughtful advice, and his enthusiasm for my work.

Two post-doctoral researchers supervised my day-to-day work more than anyone else. In Oxford, I am very grateful to Anna Rautanen for teaching me the practicalities of modern genetics. In London, Teresa Thurston tolerated

my attempts to turn my hand to molecular biology with patience and good humour.

I have been lucky enough to work with a huge number of fantastic people both in Oxford and in London over the last three years. In Adrian's group I would especially like to thank, Tara Mills, Vivek Naranbhai, Alex Mentzer, Tom Parks, Daniel O'Connor, Anne Ndungu, Kate Elliot, and Kate Auckland. In David's group, I owe particular thanks to Sophie Matthews, Regina Gunster and Megan Winterbotham. My fellowship has also been greatly enhanced by collaborations with Ben Fairfax and Julian Knight in Oxford, and Dennis Ko at Duke University.

As ever, I have been supported over the last three years by my wonderful friends and family, in particular my dad and my brother, Alex. Finally, I want to thank my wife, Ella, for continuing to be the kindest, most supportive person I have ever met. Thank you for letting me work too hard, and more importantly, for telling me when I needed to stop.

Attributions

This thesis is a report of research work carried out at the Wellcome Trust Centre for Human Genetics, University of Oxford, between September 2013 and August 2016. The work was funded by the Wellcome Trust. All work presented in this thesis is original, and was conducted by the author, except where stated in the text.

The research presented in this thesis was made possible thanks to clinical sample collections undertaken in Kenya and Malawi.

In Kenya, sample collections were directed by Anthony Scott and Tom Williams, with contributions from; Alex Macharia, Sophie Uyoga, Carolyne Ndila, Neema Mturi, Patricia Njuguna, Shebe Mohammed, James Berkley, Isaiah Mwangi, Salim Mwarumba, Barnes Kitsao, Brett Lowe, Susan Morpeth, Iqbal Khandwalla, Herbert Opi, Emily Nyatichi, Prophet Ingosi, Barnes Kitsao, Clement Lewa, Johnstone Makale, Adan Mohamed, Kenneth Magua, Mary Njoroge, Gideon Nyutu, Ruth Mwarabu, Metrine Tendwa, 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, and Alison Talbert.

In Malawi, sample collections were directed by Cal MacLennan, with contributions from; Chisomo Msefula, Esther Gondwe, Jenny MacLennan, and Malcolm Molyneux.

The genome-wide genotyping presented in this thesis was directed by the Wellcome Trust Case Control Consortium 2. The membership of which is as follows; Peter Donnelly, Ines Barroso, Jenefer Blackwell, Elvira Bramon, Matthew Brown, Juan Casas, Aiden Corvin, Panos Deloukas, Audrey Duncanson, Janusz Jankowski, Hugh Markus, Christopher Mathew, Colin Palmer, Robert Plomin, Anna Rautanen, Stephen J Sawcer, Richard Trembath, Ananth Viswanathan, Nicholas Wood, Chris Spencer, Gavin Band, Céline Bellenguez, Colin Freeman, Garrett Hellenthal, Eleni Giannoulatou, Matti Pirinen, Richard Pearson, Amy Strange, Zhan Su, Damjan Vukcevic, 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 McCann, Jennifer Liddle, Simon Potter, Radhi Ravindrarajah, Michelle Ricketts, Matthew Waller, Paul Weston, Sara Widaa, Pamela Whittaker, and Panos Deloukas. Affymetrix genotyping was performed by the Affymetrix Service Laboratory, and ImmunoChip genotyping by the Wellcome Trust Sanger Institute. Sequenom iPlex genotyping was performed by the core genomics group at the Wellcome Trust Centre for Human Genetics.

Associated publications

Gilchrist JJ, MacLennan CA, Hill AV. Genetic susceptibility to invasive Salmonella disease. Nat Rev Immunol. 2015 Jul;15(7):452-63.

Gilchrist JJ, Mills TC, Naranbhai V, Chapman SJ, Fairfax BP, Knight JC, Williams TN, Scott JA, MacLennan CA, Rautanen A, Hill AV; Wellcome Trust Case Control Consortium 2. Genetic variants associated with non-typhoidal Salmonella bacteraemia in African children. Lancet. 2015 Feb 26;385 Suppl 1:S13.

Gilchrist JJ, Heath JN, Msefula CL, Gondwe EN, Naranbhai V, Mandala W, MacLennan JM, Molyneux EM, Graham SM, Drayson MT, Molyneux ME, MacLennan CA. Cytokine Profiles during Invasive Nontyphoidal *Salmonella* Disease Predict Outcome in African Children. Clin Vaccine Immunol. 2016 Jul 5;23(7):601-9.

Contents

List of Figures

Chapter 1 – Introduction & Literature Review

Chapter 3 – Genome-wide association study of nontyphoidal *Salmonella* bacteraemia

Chapter 4 – Functional basis for the nontyphoidal *Salmonella* association at *STAT4*

Chapter 5 – Characterisation of *EVI5L* as a susceptibility locus for nontyphoidal *Salmonella* bacteraemia

Chapter 6 – Pathway analysis of nontyphoidal *Salmonella* bacteraemia susceptibility

List of Tables

Chapter 1 – Introduction & Literature Review

Chapter 2 – Methods: genetic & statistical concepts

Chapter 3 – Genome-wide association study of nontyphoidal *Salmonella* bacteraemia

Chapter 4 – Functional basis for the nontyphoidal *Salmonella* association at *STAT4*

Chapter 5 – Characterisation of *EVI5L* as a susceptibility locus for nontyphoidal *Salmonella* bacteraemia

Chapter 6 – Pathway analysis of nontyphoidal *Salmonella* bacteraemia susceptibility

Abbreviations

Chr……..….chromosome

CI……confidence interval

DNA……..……deoxyribonucleic acid

dNTP…….……deoxyribonucleotide triphosphate

eQTL…….…expression quantitative trait locus

eSNP……..expression-associated SNP

G6PD……..……..glucose-6-phosphate dehydrogenase

GWAS……..…genome-wide association study

HIV……..…..human immunodeficiency virus

HRMA……..……..high resolution melt-curve analysis

HWE……..Hardy-Weinberg equilibrium

IBD……..…identity by descent

IBS……identity by state

IFN……..interferon

IL-12……interleukin-12

iNOS……..…….inducible nitric oxide synthase

iNTS…….……invasive nontyphoidal *Salmonella*

LD……..linkage disequilibrium

MAF……..……minor allele frequency

MSMD……...……..Mendelian susceptibility to mycobacterial disease

NK……..……natural killer

NLR……..NOD-like receptor

NTS……nontyphoidal *Salmonella*

OR……odds ratio

PAMP……pathogen-associated molecular pattern

PCA……principal components analysis

PCR……polymerase chain reaction

QC……quality control

QQ plot……quantile-quantile plot

RNA……ribonucleic acid

RES……reticulo-endothelial system

RNAi……RNA interference

ROS……reactive oxygen species

SCV……Salmonella-containing vacuole

shRNA……small hairpin RNA

siRNA……small interfering RNA

SNP……single nucleotide polymorphism

SPI……*Salmonella* pathogenicity island

TLR……toll-like receptor

TNF……tumour necrosis factor

TTSS……type-three secretion system

Chapter 1 – Introduction & Literature Review

The overarching aims of this thesis are to identify, validate and describe the functional consequences of common genetic variation associated with invasive nontyphoidal *Salmonella* (NTS) disease in African children. In this chapter, I review the evidence published to date outlining the epidemiology, clinical features, microbiology, and transmission of NTS in Africa. I further outline our current understanding of the microbial pathogenesis of, and immunity and genetic susceptibility to, invasive *Salmonella* infection in humans.

Epidemiology of nontyphoidal *Salmonella* disease in Africa

Burden of nontyphoidal Salmonella gastroenteritis

NTS are an important cause of infectious diarrhoeal disease, with the global burden of NTS gastroenteritis estimated to be 93.5 million in 2006 (Majowicz et al., 2010). Rates of NTS gastroenteritis are particularly high in East and Southeast Asia and Central Europe (Figure 1.1A), with the estimated

incidence in sub-Saharan Africa reported to be modest (median 190 cases per 100,000 person-years; range 70-470). Diarrhoeal disease secondary to NTS is estimated to have resulted in 155,000 deaths in 2006 globally, with 4,100 of those deaths estimated to have occurred in sub-Saharan Africa.

Burden of invasive nontyphoidal Salmonella disease

In contrast to the modest contribution of Africa to the global burden of NTS gastroenteritis, sub-Saharan Africa is the principal region affected by invasive NTS (iNTS) infection (Figure 1.1B). In a meta-analysis of the causes of bacteraemia identified in prospective, hospital-based studies in Africa, NTS is the leading cause of community-acquired bacteraemia in African adults, and the second most commonly identified pathogen in African children (Reddy, Shaw, & Crump, 2010). In 2010, of an estimated 3.4 million global cases of iNTS disease, 1.9 million were estimated to have occurred in sub-Saharan Africa (Ao et al., 2015).

Invasive NTS disease in African populations primarily affects HIV-infected adults and young children, as is reflected by the observed bimodal distribution of iNTS cases seen in Malawi (Figure 1.2). In the limited number of African studies with defined denominator populations (Enwere et al., 2006; Mandomando et al., 2015; Mtove et al., 2011; Muthumbi et al., 2015; Oneko et al., 2015; Tabu et al., 2012; Verani et al., 2015), the median estimated incidence of iNTS disease in children is 195 cases (range 26-1,870) per 100,000 person-years of observation (Table 1.1). The importance of NTS as a

cause of invasive bacterial infection in African children seems likely to increase as access to available anti-bacterial vaccines (in particular pneumococcal and *Haemophilus influenza* type b vaccines) improves in Africa, and anti-*Salmonella* vaccines remain only available for typhoid fever (MacLennan, Martin, & Micoli, 2014).

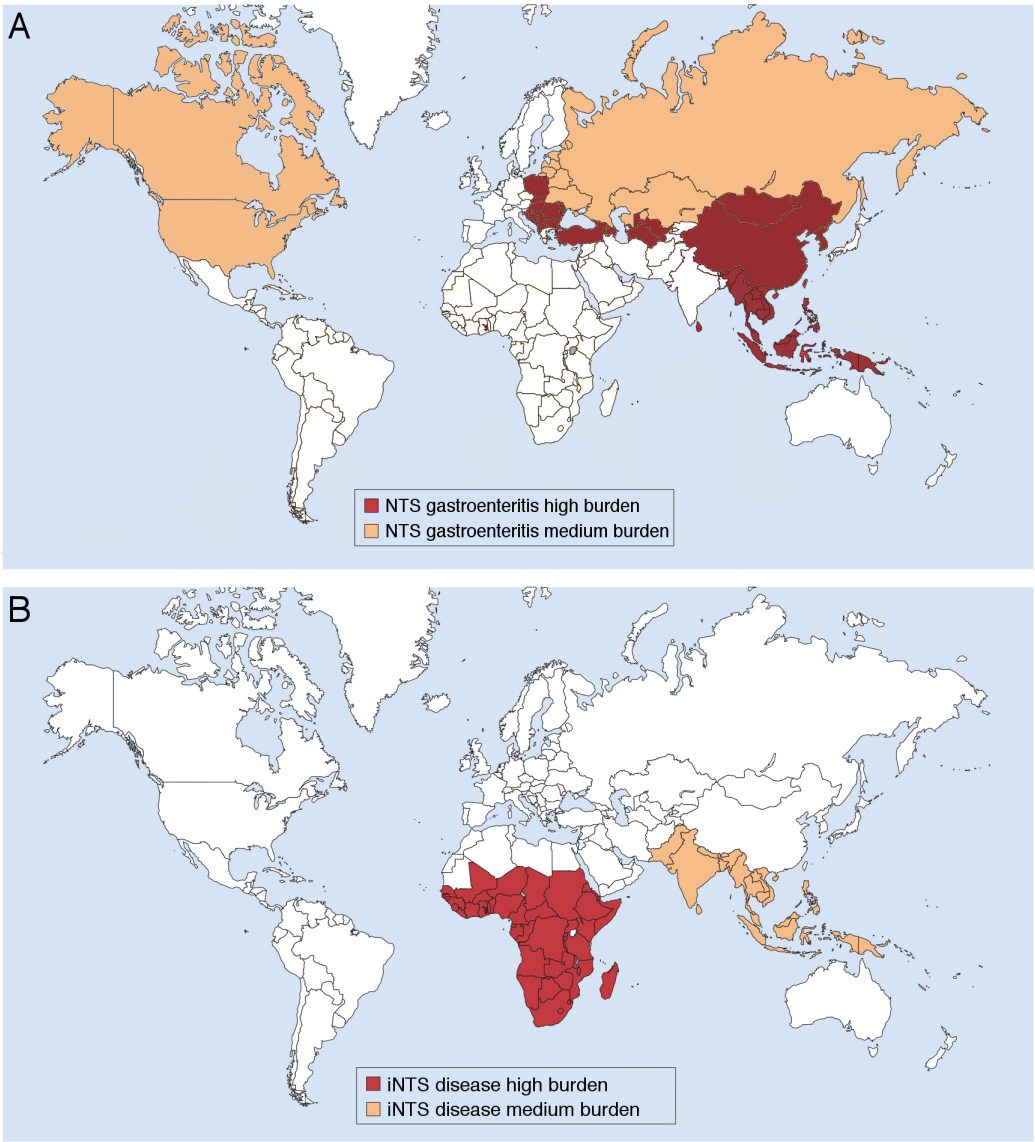

Figure 1.1 Global distribution of diarrhoeal and invasive nontyphoidal *Salmonella* disease.

For diarrhoeal disease (**A**), high burden of disease is defined as >1,000 episodes per 100,000 person-years, and medium burden as 500-1,000 per 100,000 persons years. For iNTS disease (**B**), high burden of disease is defined as >100 episodes per 100,000 person-years, and medium burden as 10-100 per 100,000 person-years. Estimates of diarrhoeal disease burden as taken from (Majowicz et al., 2010).

Figure modified from Reference (Gilchrist, MacLennan, & Hill, 2015). First published in Nature Reviews Immunology, volume 15, pages 452-63, by Nature Publishing Group.

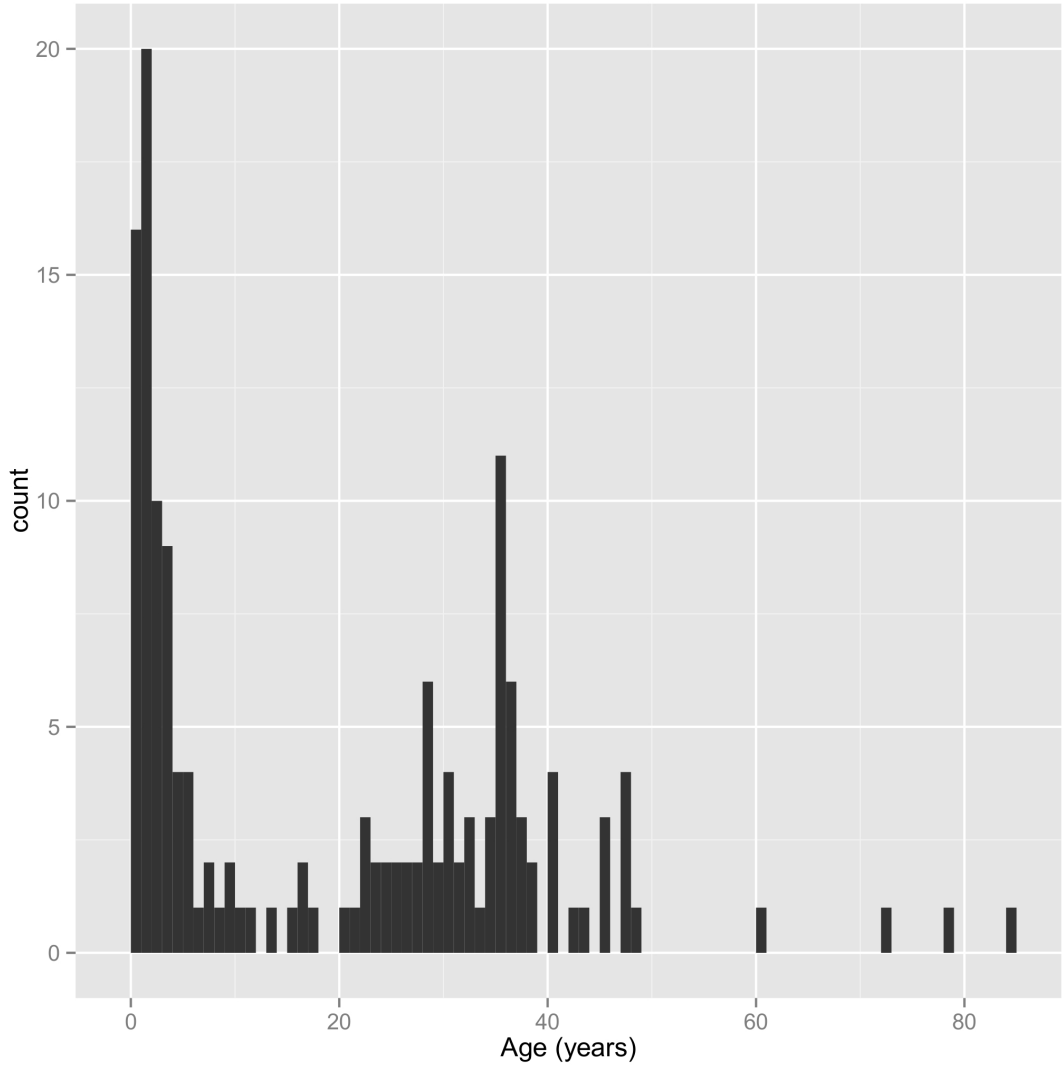

Figure 1.2 Invasive nontyphoidal Salmonella disease age distribution.

Age distribution of invasive nontyphoidal *Salmonella* disease at Queen Elizabeth Central Hospital, Blantyre, Malawi; 2011-2012.

Table 1.1 Incidence estimates for invasive nontyphoidal Salmonella infection in African children.

Country	Region	Dates	Crude Incidence per 100,000 person-years	Adjusted incidence estimates	Changing incidence over study period, IRR	Reference
Kenya	Kilifi	1998-2014	25.6 (22.0–29.5)	36.4 (35.6–37.1)	0.84 (0.81–0.86)	(Muthumbi et al., 2015)
Kenya	Lwak	2006-2009	206 (138–298)	2085 (1181–2990)	2.69 (1.11-8.50)*	(Tabu et al., 2012)
Kenya	Lwak	2009-2014	501.8 (411.6–611.6)	3914.3 (3646.5–4201.9)	0.73 (0.56-0.93)*	(Verani et al., 2015)
Kenya	Kibera	2006-2009	52 (17–122)	260 (102–419)	ND	(Tabu et al., 2012)
Kenya	Kibera	2009-2014	254.9 (192.1–338.2)	997.9 (864.9–1151.3)	0.81 (0.68-0.97)*	(Verani et al., 2015)
Kenya	Western Kenya	2009-2013	1870 (1540-2280)	ND	ND	(Oneko et al., 2015)
Mozambique	Manhica	2001-2014	184.5 (164.5-207)	ND	0.82 (0.79-0.85)*	(Mandomando et al., 2015)
Tanzania	Muheza	2006-2010	29.7 (25.3-34.9)	ND	0.41 (0.22-0.67)*	(Mtove et al., 2011)
Gambia	Upper & Central River Divisions	2000-2004	300 (222-397)	ND	ND	(Enwere et al., 2006)
Malawi	Blantyre	2008-2010	158 (145.2-172.1)	ND	0.58 (0.41-0.80)*	**

*Where not reported, annualised IRRs are calculated from incidence rates by Poisson regression in R.

** Calculated from annual paediatric cases in Feasey, *et al.*(Feasey, Everett, et al., 2015a) and Blantyre urban population estimates from 2008 census (www.nsomalwi.mw/2008-population-and-housing-census).

ND, no data; IRR, incidence rate ratio.

It should be noted that there are significant limitations to estimates of both NTS gastroenteritis and invasive disease in Africa. Burden of disease estimates for NTS gastroenteritis in Africa, included in the most complete global burden of disease analysis to date (Majowicz et al., 2010), are derived

solely from returning traveller data, which are unlikely to be representative of much of sub-Saharan Africa, in particular rural and remote settings. Available estimates of invasive disease are substantially more reliable, making use of multiple, prospective studies in diverse African settings. That notwithstanding, current estimates are limited by a bias towards studies with hospital-based ascertainment of disease, and to the few African settings with ready availability of bacterial culture services.

Mortality secondary to invasive nontyphoidal Salmonella disease

Mortality secondary to iNTS disease in both African adults and children is high. Case-fatality rates for iNTS disease in African adults have been estimated to be higher than those seen in children, with a 37% case-fatality in Kenyan adults (Muthumbi et al., 2015) and 47% reported in Malawi (Gordon et al., 2002). These high rates of mortality have improved, in parallel with the falling incidence of iNTS disease in ARV-treated African adults. Provision of ARV treatment programmes in Malawi resulted in the NTS-associated case-fatality rate in Malawian adults declining to 11% (Feasey, Houston, et al., 2014b). In studies reporting case-fatality rates in African children (Table 1.2), the median case-fatality is 12% (range 3%-24%) (Enwere et al., 2006; Graham, Walsh, Molyneux, Phiri, & Molyneux, 2000; Mandomando et al., 2015; Muthumbi et al., 2015; Oneko et al., 2015). Adopting a consensus case fatality rate of 20% for iNTS disease, the estimated global iNTS disease-related mortality is 680,000 deaths annually, of which 390,000 occur in Africa (Ao et al., 2015).

Table 1.2 Clinical features and outcome of African children with invasive nontyphoidal *Salmonella* disease.

		(Muthumbi et al., 2015) Kenya Kilifi 1998-2014	(Oneko et al., 2015) Kenya Western Kenya 2009-2013	(Mandomando et al., 2015) Mozambique Manhica 2001-2014	(Enwere et al., 2006) Gambia Upper & Central River Divisions 2000-2004	(Graham et al., 2000) Malawi Blantyre 1996-1998	(Falay et al., 2016) DRC Oriental 2009-2014
Age, months (median)		15	17.5	21.5	20	14	24
Male		55.70%	44.7%	54.6%	52.2%	59.1%	58.3%
Clinical features at presentation	Fever	74.4%	99.0%	97.0%	100.0%	96.6%	NR
	Tachycardia	NR	51.0%	NR	NR	NR	NR
	Tachypnoea	NR	68.8%	NR	77.2%	72.1%	NR
	Diarrhoea	29.9%	45.8%	33.0%	37.8%	47.5%	28.0%
	Vomiting	32%	NR	30.0%	62.2%	NR	45.3%
	Clinical ARI	17.1%	NR	35.0%	NR	NR	15.7%
	Hepatomegaly	17.6%	NR	NR	NR	37.1%	18.8%
	Splenomegaly	19%	NR	36.0%	NR	44.7%	26.1%
Mortality		22.1%	3.1%	10.3%	5.8%	23.8%	13.3%

ARI, acute respiratory infection; NR, not recorded, DRC, Democratic Republic of Congo.

While these rates of iNTS-related mortality represent the most robust estimates generated to date, there remains considerable uncertainty regarding their accuracy. It is noteworthy that the two lowest estimates of case fatality (3.1% and 5.8%) were reported in studies performing blood cultures for fever surveillance as part of vaccine trials (Enwere et al., 2006; Oneko et al., 2015) including children managed in an outpatient setting, whereas the two highest case fatality estimates (22.1% and 23.8%) are reported in studies undertaken exclusively in hospital settings (Graham et al., 2000; Muthumbi et al., 2015). Some of these differences will be secondary to differential availability of ARV drugs between study populations, and

differential rates of mortality-associated co-morbidities, but some of these differences seem likely to be secondary to ascertainment bias in hospital-based studies resulting in estimated mortality rates not representative of those seen in the general population.

Nontyphoidal Salmonella meningitis

The majority of invasive NTS infection encountered in African children is bacteraemia without apparent focal infection. However, a small, but significant, proportion of iNTS disease in African children is NTS meningitis (Keddy et al., 2015), commonly complicated by bacteraemia. Among Kenyan children with iNTS disease, 22 of 321 cases (6.9%) were cases of NTS meningitis (Muthumbi et al., 2015). African children with NTS meningitis are typically younger than children with non-focal bacteraemic disease, with a median age of 7 months in one Malawian series (E. M. Molyneux et al., 2009), and 8 of 22 Kenyan children with NTS meningitis were aged less than 7 days (Muthumbi et al., 2015). Mortality among children with NTS meningitis is particularly high, with case fatality rates of circa 50% (E. M. Molyneux et al., 2009; Muthumbi et al., 2015), with neurological sequelae reported in 56% of survivors. These rates of mortality and neurological sequelae are unaffected by adjunctive steroid treatment (E. M. Molyneux et al., 2002).

Acquired risk factors for invasive nontyphoidal Salmonella disease

Invasive NTS infection in African adults is overwhelmingly HIV-associated, with reported rates of HIV co-infection being reported in excess of 95% of African adults with NTS bacteraemia (Gordon et al., 2002; 2008). The role of HIV as a key factor in the emergence of NTS as an important cause of invasive bacterial infection in Africa has been further clarified by whole-genome sequencing-based phylogenetic analysis of *S*. Typhimurium isolates causing invasive infection in the continent over a period of several decades (Okoro, Kingsley, Connor, et al., 2012a). In that analysis, *S*. Typhimurium strains causing invasive infection are composed of two highly related lineages that are genetically distinct from gastroenteritis-causing strains. The emergence of the first of these lineages in each African country sampled is temporally associated with the emergence of HIV in each region. In keeping with this, the minimum incidence of iNTS disease among anti-retroviral drug-naïve, HIV-infected Malawian adults is estimated at 93 per 100,000 person-years, which falls to 38 per 100,000 following 9 months of ARV treatment (Feasey, Houston, et al., 2014b), albeit in parallel to a similar fall in incidence of all-cause bacteraemia.

In African children, risk factors for iNTS disease are more diverse, with HIV infection, malnutrition (Berkley et al., 2005) and malaria (Scott et al., 2011) all predisposing to disease. The relationship between iNTS disease and malaria is particularly complex, with concurrent parasitaemia (Berkley et al., 2009), recent malaria (Brent et al., 2006) and severe malarial anaemia (Bronzan et al., 2007) all being associated with disease. The prevalence of these key risk factors among cases of paediatric iNTS disease in several African settings

(Enwere et al., 2006; Falay et al., 2016; Graham et al., 2000; Mandomando et al., 2015; Muthumbi et al., 2015; Oneko et al., 2015; Tabu et al., 2012; Verani et al., 2015) is summarised in Table 1.3.

Table 1.3 Risk factors for invasive nontyphoidal *Salmonella* disease in African children

Study	Country	Region	Dates	HIV	Malnutrition	Malaria	Severe anaemia
(Muthumbi et al., 2015)	Kenya	Kilifi	1998-2014	24.1%	31.4%	28.1%	NR
(Tabu et al., 2012)	Kenya	Lwak	2006-2009	20.0%	NR	NR	NR
(Verani et al., 2015)	Kenya	Lwak	2009-2014	28.6%	NR	28.6%	NR
(Verani et al., 2015)	Kenya	Kibera	2009-2014	NR	NR	8.3%	NR
(Oneko et al., 2015)	Kenya	Western Kenya	2009-2013	16.7%	8.3%	58%**	32.2%***
(Mandomando et al., 2015)	Mozambique	Manhica	2001-2014	NR	31.0%	43.0%	19%
(Enwere et al., 2006)	Gambia	Upper & Central River Divisions	2000-2004	NR	32.6%	24.7%	54.4%***
(Graham et al., 2000)	Malawi	Blantyre	1996-1998	19.7%*	14.5%	31.4%	26.8%
(Falay et al., 2016)	DRC	Oriental	2009-2014	NR	15.9%	69.3%	29.7%

*Clinical diagnosis **Includes malaria up to 2 weeks prior to iNTS disease ***Hb<8g/dl

NR, not recorded.

Declining incidence of paediatric invasive nontyphoidal Salmonella disease

African centres reporting longitudinal surveillance of iNTS disease rates in children over the last decade, have reported declines in paediatric iNTS incidence (Figure 1.3). The median incidence rate ratio (IRR) during

surveillance is 0.81 (range 0.41 to 2.69), suggesting an annual reduction of 19% in paediatric iNTS incidence. In coastal Kenyan and Gambian settings this has been shown to be secondary to waning malaria transmission (Mackenzie et al., 2010; Scott et al., 2011). The role of falling malaria transmission in declining iNTS disease rates has been recently confirmed by a study mathematically modelling the observed changes in incidence in Malawian children (Feasey, Everett, et al., 2015a). In addition however, the same study found evidence for changing rates of HIV and malnutrition as having significant effects on iNTS incidence.

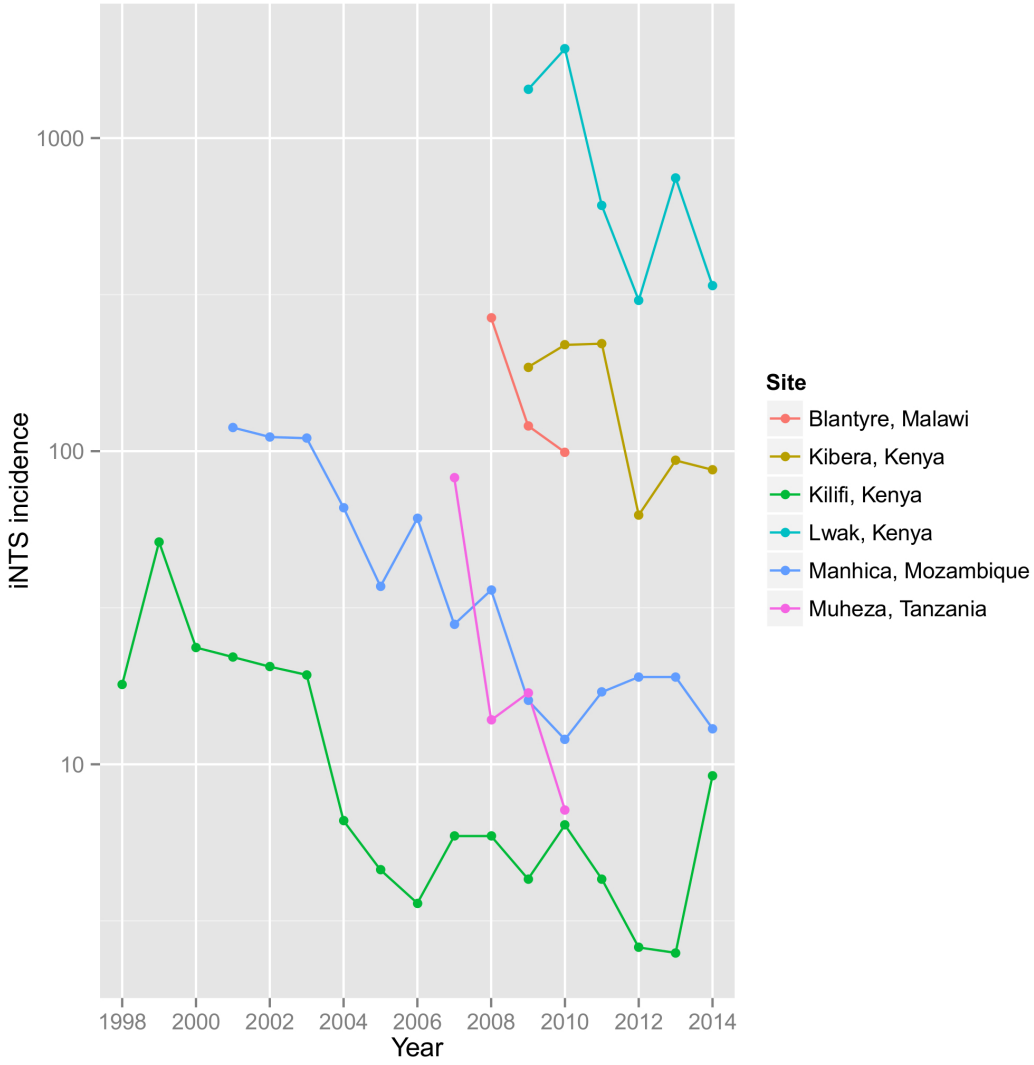

Figure 1.3 Declining invasive nontyphoidal *Salmonella* incidence in Africa.

Declining invasive NTS disease incidence in African children at six sites. Incidence per 100,000 person-years observation in children under 5 years.

Clinical aspects of invasive nontyphoidal *Salmonella* disease

Clinical presentation

The reported clinical presentation of African children with iNTS disease across several African study sites is strikingly consistent (Table 1.3; (Enwere et al., 2006; Falay et al., 2016; Graham et al., 2000; Mandomando et al., 2015; Muthumbi et al., 2015; Oneko et al., 2015)). Disease is predominantly seen in young infants (study median age, 19 months; range, 14-24 months). Fever at presentation is common (study median, 97%; range 74-100%), but other clinical features are highly variable. Despite causing diarrhoeal disease in other settings, a history of diarrhoea in children with invasive NTS is uncommon (study median, 35%; range, 28-48%). The lack of reported diarrhoea seems likely to represent attenuated host immune responses in a susceptible and immunosuppressed population, and pathogen-directed attenuation of intestinal inflammatory responses in invasive African strains of NTS (for a discussion of both of these features, see below). It seems likely that blunted enteric inflammatory responses facilitate dissemination, enhancing the likelihood of systemic infection.

That the clinical picture represents one of undifferentiated sepsis is supported by the observation that a clinical diagnosis of pneumonia at admission is made in a median of 19% of cases (range, 16-35%). On examination, children are commonly tachypnoeic (study median, 72%; range, 69-77%), and, where reported, tachycardic (prevalence in a single study, 52%). The observed

tachypnoea is likely to be secondary to acidosis in the context of sepsis, but may well contribute to the clinical impression of community-acquired pneumonia a in subset of children. Splenomegaly is also commonly reported (study median, 31%; range, 19-45%), which in this context is likely to represent recent malaria infection.

Diagnosis

The lack of a pathognomonic clinical presentation makes clinical diagnosis of invasive NTS infection in African children impossible. There are no validated point-of-care biomarkers of NTS infection, and no available rapid diagnostic serological or PCR-based tool. This means that bacterial culture of blood or cerebro-spinal fluid is the only available method for diagnosis of invasive NTS disease. In Africa, where blood culture facilities are not commonly available, the lack of available diagnostic tests compromises both clinical care, and the ability to generate robust epidemiological estimates of the NTS-associated disease burden that are representative of Africa as a whole.

Management

The effective clinical management of invasive NTS disease is focussed on supportive care, and delivery of antimicrobial agents with adequate anti-*Salmonella* activity and adequate penetration into tissues harbouring *Salmonellae*. Even in the few African settings in which bacterial culture services are available, the diagnostic challenge associated with iNTS in

children will mean that definitive microbiological diagnosis is delayed following admission. As a consequence, African children treated for invasive NTS disease will be initially managed with empirical antibiotics, and so empirical antibiotic prescribing guidelines need to reflect the local burden and antibiotic resistance profile of NTS. Expanding antibiotic resistance among NTS isolates causing invasive disease in African populations (see below), has meant that empirical antibiotic guidelines for African children commonly fail to provide adequate anti-*Salmonella* cover. For instance, chloramphenicol and gentamicin have been until recently used as empirical antibiotic treatment for sepsis in Malawian children, with the majority of invasive NTS isolates having chloramphenicol resistance, and gentamicin lacking intracellular penetration, making it ineffective for the treatment of NTS. Similarly, the widespread use of penicillins for the treatment of pneumonia in African children will result in inadequate anti-*Salmonella* treatment for those children with NTS initially diagnosed with community-acquired pneumonia.

Microbiology of invasive nontyphoidal *Salmonella* disease in Africa

Salmonellae are motile, gram-negative, facultative anaerobes, belonging to the family *Enterobactericeae*. The genus *Salmonella* is composed of two species, *Salmonella bongori* and *Salmonella enterica*, and is estimated to have diverged from *Escherichia coli* approximately 100 million years ago (Doolittle, Feng, Tsang, Cho, & Little, 1996). *S. bongori* principally causes disease in reptiles, and has only rarely been reported to be associated with disease in humans. In cases where *S. bongori* has acted as a human pathogen, it has resulted in uncomplicated diarrhoeal disease in infants and young children (Giammanco et al., 2002). By contrast, the species *Salmonella enterica* is composed of thousands of human, disease-causing serovars.

Salmonella enterica serovars can be broadly divided into two groups; typhoidal and nontyphoidal strains. The typhoidal serovars, *Salmonella enterica* serovar Typhi (*S*. Typhi) and *Salmonella enterica* serovars Paratyphi A, B and C (*S.* Paratyphi A, B and C), are host-adapted, human-restricted organisms, causing a single clinical syndrome – enteric fever – in immunocompetent individuals. Enteric fever, also known as typhoid fever, is characterised by a systemic, febrile, influenza-like illness, in which gastrointestinal symptoms (while common), are often not a cardinal presenting feature (Parry, Hien, Dougan, White, & Farrar, 2002).

Nontyphoidal *Salmonella* serovars have a broad vertebrate host range and cause a self-limiting colitis in the majority of immunocompetent individuals in developed settings (Hohmann, 2001). In the UK, where national surveillance of *Salmonella* infections has been in place since 1945, epidemics of diarrhoeal disease secondary to NTS have been caused most commonly by serovars Typhimurium and Enteritidis, but significant epidemics have also been caused by serovars Hadar and Agona (Lane et al., 2014). NTS serovars causing invasive disease in Africa are diverse, but in common with diarrhoeal disease, the large majority are either *Salmonella enterica* serovars Typhimurium or Enteritidis (Table 1.4).

Multi-drug resistant nontyphoidal Salmonella

Multi-drug resistant (MDR) strains of NTS causing invasive disease, with non-susceptibility to ampicillin, co-trimoxazole and chloramphenicol, are widespread in Africa (Table 1.4).

Table 1.4 Nontyphoidal *Salmonella* isolates causing invasive disease in Africa

Study	(Muthumbi et al., 2015)	(Tabu et al., 2012)	(Tabu et al., 2012)	(Oneko et al., 2015)	(Mandomando et al., 2015)	(Kalonji et al., 2015)	(Feasey, Masesa, et al., 2015b)	(Feasey et al., 2010)
Country	Kenya	Kenya	Kenya	Kenya	Mozambique	DRC	Malawi	South Africa
Region	Kilifi	Lwak	Kibera	Western Kenya	Manhica	4 sites	Blantyre	NRU
Dates	1998-2014	2006-2009	2006-2009	2009-2013	2001-2014	2011-2014	1998-2014	2003-2004
MDR	23.9%	80.0%*	NR	52.1%	NR	84.9%	65.1%	NR
3rd generation cefalosporin resistance**	6.0%	0.0%	NR	16.7%	0.0%	0.0%	0.2%	NR
Decreased ciprofloxacin sensitivity	NR	NR	NR	1.1%	0.0%	1.9%	0.2%	NR
S. Typhimurium	43.3%	88.3%	85.7%	70.6%	68.9%	49.2%	79.0%	67.0%
S. Enteritidis	45.6%	10.0%	14.3%	29.4%	18.6%	49.9%	16.0%	10.0%
Nontypable/ other	11.1%	1.7%	0.0%	0.0%	12.4%	0.9%	5.1%	23.0%

MDR, multi-drug resistance (resistance to co-trimoxazole, chloramphenicol, ampicillin); NRU, national reference unit; NR, not recorded.

The emergence of these strains represents a major public health concern for a region in which these cheap, and widely available, antibiotics are commonly used to empirically treat children with clinical syndromes indistinguishable from iNTS disease. The prevalence of MDR strains among invasive NTS isolates in Africa has varied considerably in centres undertaking active surveillance over the last 20 years (Feasey, Masesa, et al., 2015b; Muthumbi et al., 2015). In Malawi between 1998 and 2014, high rates of MDR among *S*. Enteritidis isolates have progressively fallen (2.6% decrease annually, beta regression $P = 0.05$) while the proportion of *S*. Typhimurium isolates with

MDR rose significantly (4.3% increase annually, beta regression $P = 6.6x10^{-4}$; Figure 1.4).

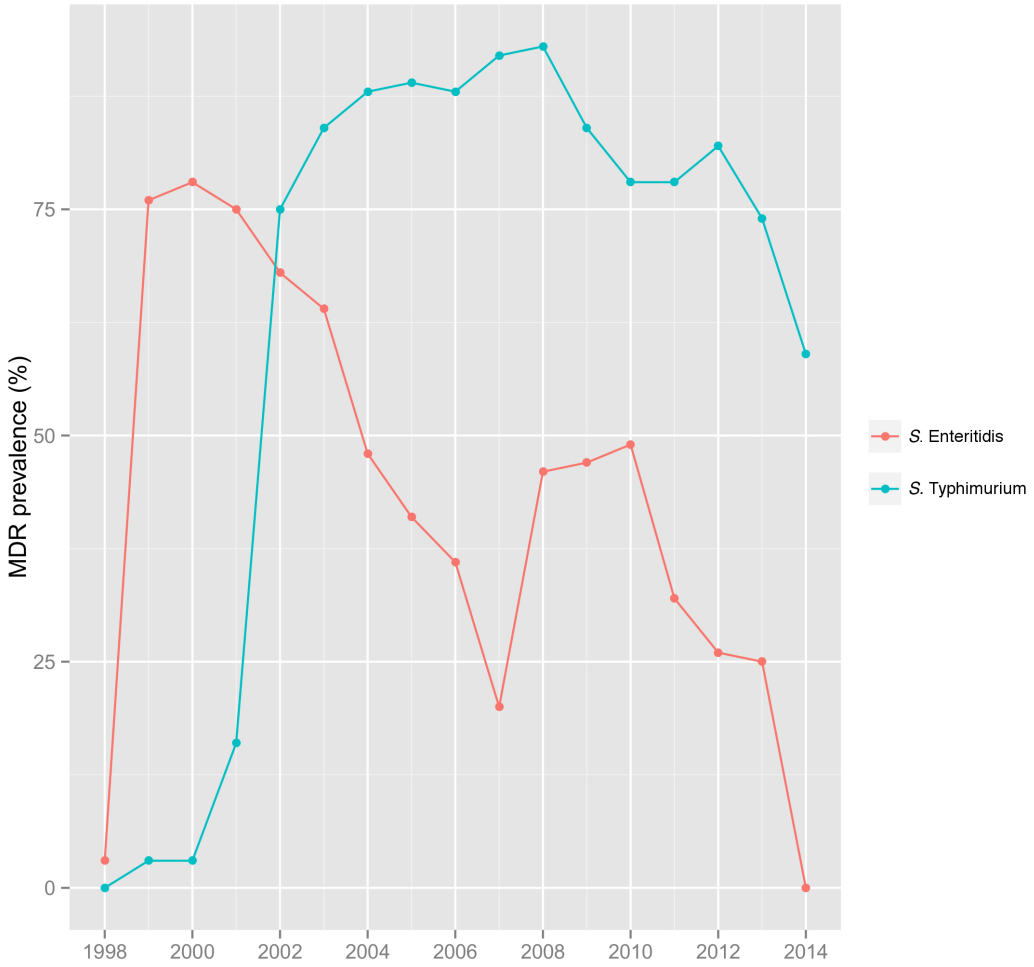

Figure 1.4 Multidrug resistant invasive nontyphoidal *Salmonella*.

Prevalence of multidrug resistance (MDR) among invasive nontyphoidal *Salmonella* isolates from paediatric and adult patients at Queen Elizabeth Central Hospital, Blantyre, Malawi (1998-2014). Data extracted from Reference (Feasey, Masesa, et al., 2015b).

In both Kenya and Malawi, there is a temporal relationship between the emergence of high rates of MDR among *S.* Typhimurium isolates, and the subsequent replacement of *S.* Enteritidis by *S.* Typhimurium as the dominant causative agent of iNTS disease (Gordon et al., 2008; Muthumbi et al., 2015). In keeping with this, declining MDR rates among *S.* Enteritidis isolates are

significantly correlated with falling incidence of *S*. Enteritidis disease in a Poisson regression model ($P = 1.9 \times 10^{-4}$, Figure 1.5).

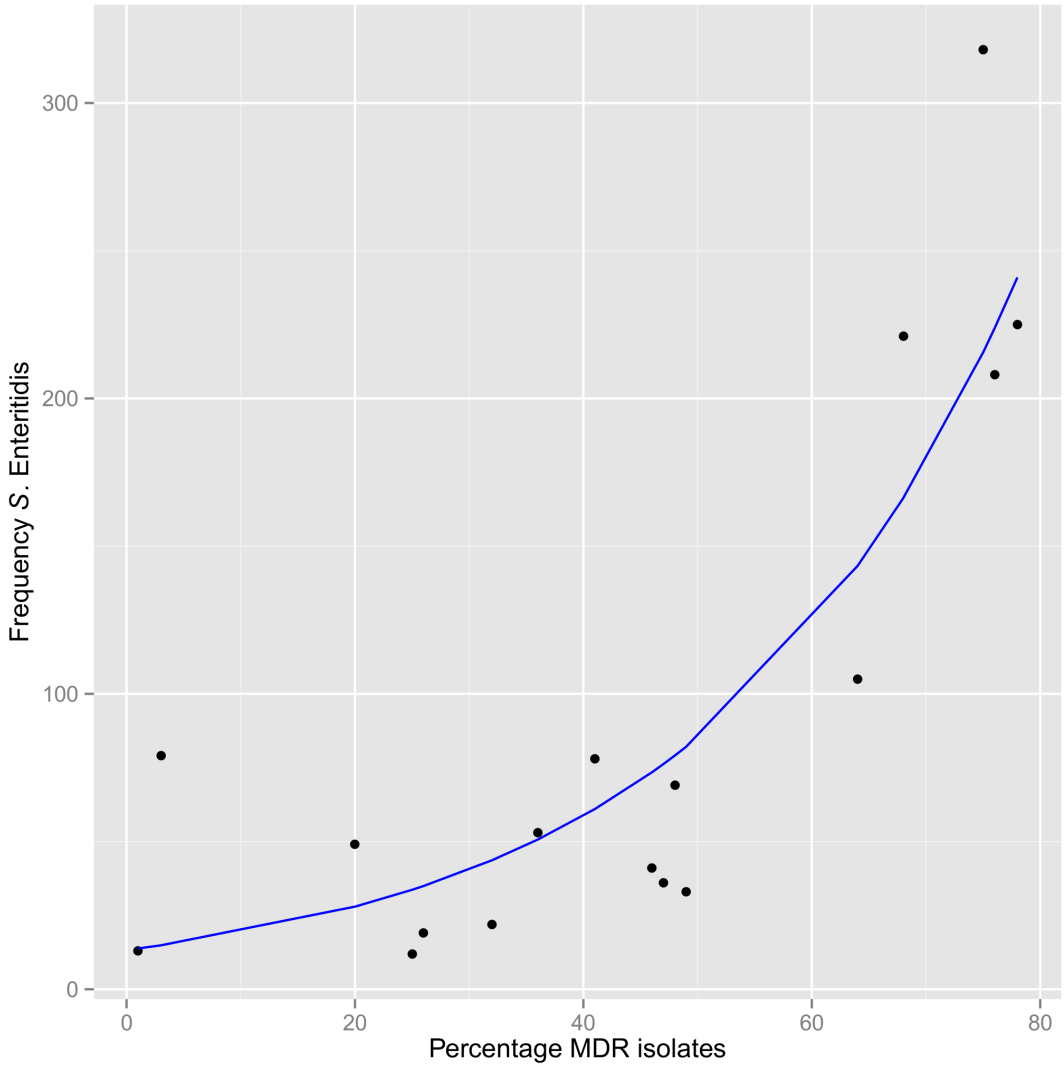

Figure 1.5 Multidrug resistance and incidence of invasive nontyphoidal *Salmonella*.

The relationship between the annual frequency of invasive disease caused by *S*. Enteritidis and the rate of *S*. Enteritidis multi-drug resistance. Poisson regression curve is plotted in blue. Data extracted from Reference (Feasey, Masesa, et al., 2015b).

Whole-genome sequencing and phylogenetic analysis of African MDR *S*. Typhimurium strains has been highly informative, identifying a dominant *S*. Typhimurium multi-locus sequence type pathovar (ST313) (Kingsley et al., 2009). Invasive ST313 in Africa appears to have emerged as two distinct

lineages, with the second lineage clonally replacing the first following the acquisition of chloramphenicol resistance, in the context of considerable chloramphenicol use for iNTS disease/undifferentiated sepsis (Okoro, Kingsley, Connor, et al., 2012a).

The prevalence of MDR iNTS isolates has meant that ceftriaxone has been increasingly relied upon to empirically treat undifferentiated sepsis in Africa (Feasey, Cain, et al., 2014a). Alongside the use of ciprofloxacin following a microbiologically confirmed case of iNTS disease, these two antimicrobials now form the mainstay of anti-*Salmonella* antimicrobial therapy in Africa. It is therefore of particular concern that isolates resistant to third generation cephalosporins and with decreased ciprofloxacin susceptibility have been reported in several African settings (Table 1.4) (Kalonji et al., 2015; Muthumbi et al., 2015; Oneko et al., 2015). In contrast to the emergence of MDR ST313 isolates, which appears not to have had an appreciable impact on case-fatality rate, mortality secondary to 3rd generation cephalosporin-resistant ST313 isolates appears to be increased (Muthumbi et al., 2015). Of particular concern is the recent report of an ST313 isolate in a Malawian HIV-infected adult patient, resistant to both ciprofloxacin and ceftriaxone (Feasey, Cain, et al., 2014a), a combination of antibiotic resistance which would make this isolate untreatable in many African settings.

Salmonella Typhimurium ST313

Whole-genome sequence-based, phylogenetic analysis of a large collection of *S*. Typhimurium isolates causing invasive disease, collected in Africa over 70 years, demonstrated a high degree of relatedness among invasive strains (Okoro, Kingsley, Connor, et al., 2012a). The invasive *S*. Typhimurium isolates, as well as having a high degree of relatedness to one another, are distinct from strains causing diarrhoeal disease globally, and form two distinct lineages. The second Typhimurium ST313 lineage is characterised by acquisition of chloramphenicol resistance, which is hypothesized to have driven its clonal replacement of lineage 1 (see discussion of MDR NTS above). Estimates of the timescale of emergence of the first lineage in several African settings demonstrates coincident emergence with the African HIV pandemic. Typhimurium ST313 causes invasive disease in both African children and adults (Msefula et al., 2012), and results in both bacteraemia and meningitis (Keddy et al., 2015). Importantly, ST313 Typhimurium strains are also a cause of uncomplicated diarrhoeal disease in Africa (Paglietti et al., 2013).

Despite a high degree of synteny to sequenced non-invasive *S*. Typhimurium strains, ST313 Typhimurium has several distinct genomic features, including a collection of five prophage-like elements (BTP1-5), and a distinct pattern of genomic degradation (Kingsley et al., 2009). That pattern of genomic degradation comprises the loss of 77 open reading frames through pseudogene formation (23 of which are specific to the ST313 isolate D23580,

as compared to sequencing non-invasive Typhimurium strains), and loss of a further 20 genes through four genomic deletions. Analysis of the major chromosomal virulence-determining regions of *Salmonella*, the *Salmonella* pathogenicity islands (SPI), demonstrates no clear pattern of variation in these regions in the ST313 lineages (Okoro et al., 2015). Several pseudogenes unique to ST313 isolates inactivate genes encoding enzymes implicated in aerobic metabolism. In keeping with this observation, ST313 isolates grown in low-salt LB broth have a distinct metabolic profile as compared to non-ST313 strains grown in the same conditions (Okoro et al., 2015). Whether this metabolic profile represents a correlate of *Salmonella* virulence is currently unclear.

In addition to the bacterial chromosome, the ST313 isolate D23580 has four plasmids, three of which have no assigned phenotype (pBT1-3) and a fourth, pSLT-BT. pSLT-BT exhibits high levels of homology to pSLT, a virulence-associated plasmid encoding the *spv* (*Salmonella* plasmid virulence) operon (Guiney et al., 1995). NTS serovars cured of pSLT have attenuated ability to cause systemic infection in poultry (Barrow, Simpson, Lovell, & Binns, 1987). pSLT carriage in *S*. Typhimurium strains is variable, and data from developed settings suggest an over-representation of pSLT-positive isolates causing bacteraemia as compared to isolates causing diarrhoeal disease (Fierer, Krause, Tauxe, & Guiney, 1992), in keeping with the hypothesis that pSLT carriage is a determinant of invasive potential in human infection. ST313-associated pSLT-BT contains a Tn*21*-like element, containing genes responsible for the ST313 MDR phenotype. Both of the documented ST313

lineages carry pSLT-BT, with nearly all isolates within each lineage carrying a *Tn*21-like element (Okoro, Kingsley, Connor, et al., 2012a). Within each lineage, the insertion sites of the transposon are conserved, but differentiate the two lineages suggesting independent acquisition of the *Tn*21-like element. A further key difference is the content of the transposable element, with the lineage II *Tn*21-like element almost universally containing the chloramphenicol resistance gene. In addition to carriage of pSLT-BT, comparative genomics of the bacterial chromosome in ST313 and non-ST313 lineages, identified a conserved genomic island between ST313 and invasive S. Dublin strains, containing a putative virulence fragment, ST313-TD (Leekitcharoenphon et al., 2013). The candidacy of ST313-TD as a virulence determinant for ST313 Typhimurium needs to be confirmed in further studies, with larger numbers of isolates with complete genome sequences.

Pathogenesis of invasive nontyphoidal *Salmonella* disease

Central to the pathophysiology of an invasive *Salmonella* infection, is the ability of *Salmonella* to invade mammalian epithelial cells and to survive and replicate within macrophages. Both of these behaviours are dependent on genes encoded within *Salmonella* pathogenicity islands. Pathogenicity islands are genetic regions containing multiple virulence-related genes, the presence of which together confer a virulence-related phenotype (Blum et al., 1994). Several genetic features associated with pathogenicity islands (including, flaking insertion sequences, island GC content not representative of the flanking region, and interruption of normally contiguous genes in non-pathogenic strains of the same species) suggest that their acquisition in pathogenic bacteria is by horizontal gene transfer, allowing the gain of a virulence mechanism in a single evolutionary step (Groisman & Ochman, 1996).

Studies utilising the mouse model of invasive *Salmonella* infection in particular, have helped to inform the current model of how the pathogenic mechanisms of *Salmonella* interact with the host to produce a systemic infection (Mastroeni, Grant, Restif, & Maskell, 2009) (Figure 1.6). Following oral ingestion, a proportion of luminal bacteria invade the apical surface of intestinal epithelial cells. Infecting *Salmonellae*, then can transmigrate to the basolateral surface of the intestinal mucosa, where they are engulfed by submucosal macrophages. Infected macrophages migrate to mesenteric lymph nodes, and into the bloodstream, both as extracellular bacteria and

CD18+ monocytic cell-associated bacteria. Blood-borne bacteria can then seed new infectious foci at distant sites, principally within the reticuloendothelial system.

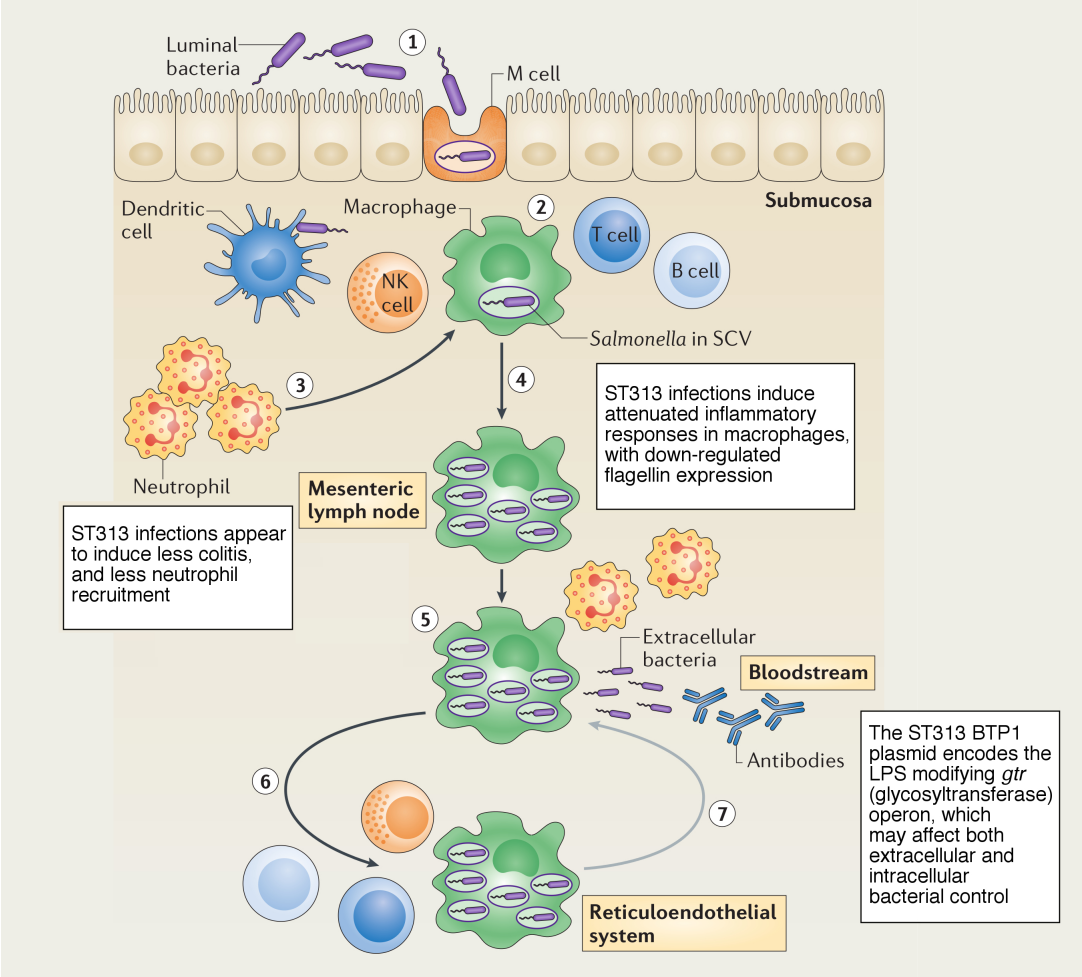

Figure 1.6 The pathogenesis of invasive nontyphoidal *Salmonella* infections.

Numbered steps in the progression of disseminated infection are as follows: 1, epithelial invasion of luminal bacteria; 2, infection of submucosal tissue phagocytes; 3, pro-inflammatory cytokine mediated recruitment of neutrophils to the infected get (this is limited in invasive infection); 4, migration of infected phagocytes to the mesenteric lymph nodes; 5, systemic dissemination takes place with intracellular and extracellular bacteraemia; 6, establishment of new foci of infection systemically, in particular in the reticuloendothelial system; 7, periodic re-circulation, establishing new infectious foci. Text boxes highlight phenotypes thought to be associated with the ST313 pathovar that may enhance invasiveness. Figure modified from (Gilchrist et al., 2015). First published in Nature Reviews Immunology, volume 15, pages 452-63, by Nature Publishing Group.

Invasion of intestinal epithelial cells

Following oral ingestion, *Salmonella* passes through the acidic environment of the stomach, before invading gut mucosal epithelial cells and microfold cells (M-cells). *Salmonella* invades epithelial cells in a SPI-1-dependent manner, inducing bacterial-mediated endocytosis. Following adherence to the apical surface of enterocytes, *Salmonella* induces cytoskeletal changes in the enterocyte, mediating membrane ruffling, and internalisation of the bacterium (Francis, Starnbach, & Falkow, 1992). This process is SPI-1 dependent (Galán & Curtiss, 1989), requiring the delivery of bacterial effector proteins into the host cell cytosol (Hardt, Chen, Schuebel, Bustelo, & Galán, 1998), directing cytoskeletal rearrangements, and *Salmonella* uptake. Delivery of bacterial effectors takes place via a SPI-1-encoded type-3 secretion system (TTSS), which forms a needle-like complex across the bacterial inner and outer membranes (Kubori et al., 1998), allowing translocation of effectors into the mammalian cell. SPI-1-deficient NTS strains are attenuated in the mouse model of invasive *Salmonella* infection, following oral delivery, but have normal infection kinetics when delivered via the intraperitoneal route (i.e. bypassing the need for epithelial invasion)(Galán & Curtiss, 1989). SPI-1-delivered effectors direct host cell cytoskeletal re-modelling by interacting directly with the actin cytoskeleton and its regulators. SopE is a GTP/GDP exchange factor for key regulators of the mammalian actin cytoskeleton (CDC42 and RAC1)(Hardt et al., 1998) and SipA and SipC directly bind actin (Hayward & Koronakis, 1999; Zhou, Mooseker, & Galán, 1999).

SPI-1-mediated enterocyte/M-cell trans-epithelial migration represents the most common route for *Salmonella* invasion across the gut epithelial barrier. A SPI-1-independent mechanism, in which CD18+ monocytic cells facilitate systemic dissemination directly from the intestinal lumen, has also been described (Vazquez-Torres et al., 1999).

Survival and persistence within macrophages

Following SPI-1 mediated transmigration of the intestinal mucosa, *Salmonella* encounters tissue-resident macrophages in the intestinal submucosa. Following phagocytosis, invasive strains of *Salmonellae* are able to survive and replicate within macrophages, and strains lacking this ability are unable to establish systemic infection. This ability to survive within macrophages is encoded by multiple virulence factors under the control of a two component response system (the PhoP/PhoQ regulon) and SPI-2.

The PhoP/PhoQ regulator consists of a transmembrane sensor element, PhoQ, which activates a cytosolic transcriptional activator, PhoP, which in turn coordinates the up- and down-regulation of a cohort of over 40 virulence-associated genes (Miller, Kukral, & Mekalanos, 1989). PhoQ senses several environmental correlates of uptake into the macrophage phagosome, including low pH(Prost et al., 2007), Mg^{2+} concentrations (García Véscovi, Soncini, & Groisman, 1996) and the presence of antimicrobial peptides (Bader et al., 2005). Several PhoP/PhoQ regulated effectors function to extensively remodel the *Salmonella* outer membrane, with distinct effectors modifying the

acetylation (Guo et al., 1998), deacylation (Trent, Pabich, Raetz, & Miller, 2001) and hydroxylation(Gibbons, Lin, Cotter, & Raetz, 2000) of lipid A, alongside a further group of effectors directing alterations in core lipopolysaccharide (LPS) glycosylation (Gunn, Ryan, Van Velkinburgh, Ernst, & Miller, 2000). The extensive remodelling of the *Salmonella* outer membrane induced by transition into an intra-macrophage environment has at least two complementary roles. Firstly, LPS modifications increase bacterial resistance to cationic antimicrobial peptides in the *Salmonella*-containing vacuole (SCV) (Gunn et al., 2000; Guo et al., 1998). Secondly, changes in LPS induced by transition to an intra-macrophage environment, attenuate host pro-inflammatory responses to intracellular *Salmonella* (Guo et al., 1997).

SPI-2 (Hensel et al., 1995), in common with SPI-1, encodes a TTSS and is induced by environmental sensing, two-component regulatory systems, responding to phagosomal acidification and inorganic phosphate restriction (Löber, Jäckel, Kaiser, & Hensel, 2006). While PhoP/PhoQ regulated effector proteins function to increase the resistance of *Salmonella* to the intra-phagosomal environment, SPI-2 effectors function to modify the phagosomal environment, mitigating its bactericidal properties. There are over 30 recognised SPI-2-delivered effectors described, with diverse effects that include the manipulation of SCV trafficking resulting in the inhibition of; phagosome-lysosome fusion (Uchiya et al., 1999), and trafficking of inducible nitric oxide synthase (iNOS) and NADPH oxidase to the SCV (Chakravortty, Hansen-Wester, & Hensel, 2002; Vazquez-Torres et al., 2000). SPI-2 effectors, in common with the PhoP/PhoQ effectors, also function to down-

regulate host pro-inflammatory responses (Haneda et al., 2012; Mazurkiewicz et al., 2008). In fact, a subset of these anti-inflammatory effectors, e.g. SpvC, are translocated into the host cell cytosol of epithelial cells as well as macrophages in SPI-1 and SPI-2 dependent processes, attenuating mucosal and systemic inflammatory responses (Haneda et al., 2012).

Extracellular survival of Salmonellae

Cell death of infected macrophages facilitates the spread of systemic infection, and exposes *Salmonella* to extracellular components of the host immune system, most notably antibody and complement. *Salmonella* virulence determinants interfering with serum killing, block assembly and function of the complement membrane attack complex. A major virulence determinant in this regard is LPS O-antigen, with bacteria with increased O-antigen polysaccharide length fixing complement (including the membrane attack complex - MAC), but remaining resistant to complement mediated lysis, with long O-antigen polysaccharide chain holding MAC distant to the outer membrane (Joiner, Hammer, Brown, & Frank, 1982a; Joiner, Hammer, Brown, Cole, & Frank, 1982b). *Salmonella* genes that determine O antigen length, also determine virulence in the mouse model by increasing complement resistance (Murray, Attridge, & Morona, 2003), resistance that is inducible following serum exposure (Murray, Attridge, & Morona, 2005). In addition, the *rck* locus contributes to complement resistance in *Salmonella*. Rck is an outer membrane protein, which appears to prevent normal C9

polymerisation, which is required to produce a fully functional MAC (Heffernan et al., 1992).

Pathogenic features of ST313 isolates

Following the identification of a clearly defined and homogenous genetic architecture of African invasive NTS isolates in ST313, there has been considerable interest in defining phenotypic correlates of the ST313 invasive lineage. Several features of ST313 contribute to its invasiveness that are not confined to ST313 (or even Typhimurium) isolates. ST313's pSLT-BT plasmid encodes the *spv* operon, which has been linked to invasive disease in diverse *Salmonella* serovars. In keeping with this, and with the observed phenotype in humans, experimental infection in the murine model of invasive *Salmonella* infection, demonstrates that ST313 isolates cause invasive infection, resulting in comparable systemic bacterial burdens to non-ST313 murine invasive strains (ST19 strains) (Okoro, Kingsley, Connor, et al., 2012a).

Efforts to define correlates of ST313-specific genomic features have demonstrated that the BTP1 plasmid encodes a glycosyltransferase (*gtr*) operon, which modifies LPS polysaccharide chain length (Kintz et al., 2015). While it has not been established that this is a correlate of virulence, this seems plausible given the extensive LPS modification directed by multiple *Salmonella* virulence factors, which determines extracellular and intra-macrophage survival. Comparisons of single ST313 (D23580) and ST19

(SL1344) isolates, have suggested phenotypic differences in invasive potential (Parsons et al., 2013; J. Yang et al., 2015) and acid-tolerance (J. Yang et al., 2015) (with the ST313 strain exhibiting greater invasiveness and acid-tolerance). Comparison of ST313 and non-ST313 (ST19) infection in the mouse (streptomycin pre-treated) and the calf ligated ileal loop models of *Salmonella*-induced colitis, demonstrates reduced inflammation in ST313 infections, with reduced neutrophil recruitment (Okoro et al., 2015). *In vitro* models of epithelial and phagocytic infection demonstrate reduced SPI-1 mediated epithelial invasion in parallel with reduced expression of the SPI-1 effector SopE2 (Carden, Okoro, Dougan, & Monack, 2015), and reduced inflammatory responses in phagocytic infection in parallel with reduced flagellin expression (Carden et al., 2015). It is noteworthy however, that replication of these findings in the murine and rhesus macaque model of *Salmonella* infection has not been consistent, with attenuated intestinal inflammation not being found in ST313 strains, and epithelial invasiveness being increased rather than decreased (Singletary et al., 2016). Many of these discrepancies seem likely to be secondary to the use of different comparator strains, and these findings will need to be replicated using a range of ST313 and non-ST313 isolates. At present it is unclear whether any of the observed differences are ST313-defining phenotypic characteristics.

Transmission of invasive nontyphoidal *Salmonella* in Africa

There are limited data describing the sources and routes of transmission of invasive NTS isolates in an African context. ST313 Typhimurium isolates exhibit marked genomic degradation (Kingsley et al., 2009). Genomic degradation has been observed to be associated with host adaption in host-restricted *Salmonella* serovars (e.g. *S*. Typhi, *S*. Paratyphi A (Holt et al., 2009) and *S*. Gallinarum (Thomson et al., 2008)), and in other human-restricted pathogens that evolved from an ancestral pathogen with greater host promiscuity (e.g. *Mycobacterium leprae* (S. T. Cole et al., 2001) and *Yersinia pestis* (Parkhill et al., 2001)). Moreover, 11 of the 77 observed pseudogenes in the ST313 isolate D23580, have also been lost by pseudogene formation in *S*. Typhi. This observation has led to the hypothesis that ST313 isolates causing invasive disease in Africa, may be restricted, or be in the process of becoming restricted, to humans.

The suggestion that invasive *Salmonella* disease-causing isolates may, in large part, be human-restricted is supported by studies comparing the epidemiology of NTS isolates in human and animal populations in African settings. In those studies, isolated *Salmonella* strains in humans and animals were distinct (Dione et al., 2011; Kariuki et al., 2002; 2006). This is in contrast to the genetic epidemiology of NTS isolates causing invasive disease in Malawian children and adults, which has demonstrated that, despite divergent risk factors for NTS disease in the adult and paediatric population, they appear to share a common infectious source (Msefula et al., 2012).

The limited phenotypic characterisation of ST313 isolates undertaken to date has demonstrated that the studied ST313 strains lack catalase activity (Singletary et al., 2016) and have reduced ability to form biofilms (Ramachandran, Aheto, Shirtliff, & Tennant, 2016; Singletary et al., 2016). Both characteristics suggest that ST313 isolates may have reduced capacity for environmental persistence, as compared to gastroenteritis-causing ST19 strains, a phenotype which may be in keeping with increased anthroponotic transmission. In common, however, with the putative ST313 changes in pathogenesis, these phenotypic differences with need to be verified using larger panels of ST313 and non-ST313 strains. It is also important to note that, while the epidemiological, genetic and phenotypic data do appear to suggest increased host-restriction and anthroponotic transmission among ST313 isolates, ST313 isolates, as opposed to truly host-restricted *Salmonella* serovars, retain the ability to cause systemic infection in mammals other than humans (Okoro et al., 2015; Parsons et al., 2013).

In keeping with humans acting as the infectious source for invasive NTS disease in African populations, it has been demonstrated that HIV-infected African adults can carry invasive NTS isolates for many months, resulting in multiple recurrent infections with the same organism (Gordon et al., 2002; Okoro, Kingsley, Quail, et al., 2012b). This observation, considered alongside the emergence of ST313 lineages alongside the HIV pandemic in Africa, suggests a hypothesis whereby introduction of a permissive host environment may have resulted in the emergence of an increasingly host-adapted

organism causing epidemics of invasive disease in a susceptible population. This hypothesis has clear parallels to the recently described in-host evolution of an NTS isolate from an individual with a primary immune defect, resulting in extreme susceptibility to invasive NTS disease (Klemm et al., 2016).

Immunity to invasive nontyphoidal *Salmonella* disease

Much of our understanding of immunity to invasive *Salmonella* infections has been derived from studies in the mouse model of invasive *Salmonella* disease (Santos et al., 2001). Both innate and adaptive responses facilitate the control of *Salmonella* infection, with components of each contributing to control of *Salmonella* during the distinct phases of a murine *Salmonella* infection (Hormaeche, 1979).

Innate immune responses to nontyphoidal Salmonella infection

The ability of the host, in the experimental mouse model of a primary, invasive *Salmonella* infection, to control early bacterial growth within the reticuloendothelial system (RES) is dependent on innate pattern-recognition receptors, activation of which results in the production of key pro-inflammatory cytokines. This phase of a primary infection, in a mouse, is independent of T or B cell immunity (Hess, Ladel, Miko, & Kaufmann, 1996; Hormaeche, Mastroeni, Arena, Uddin, & Joysey, 1990).

Toll-like receptor signalling

Toll-like receptors (TLRs) are membrane-expressed, innate pattern-recognition receptors, which ligate pathogen-associated molecular patterns (PAMPs). Four human-expressed TLR molecules ligate *Salmonella*-expressed PAMPs; TLR2, TLR4, TLR5 and TLR9 bind lipoprotein, LPS, flagellin, and CpG DNA respectively (Figure 1.7) (Kawai & Akira, 2011).

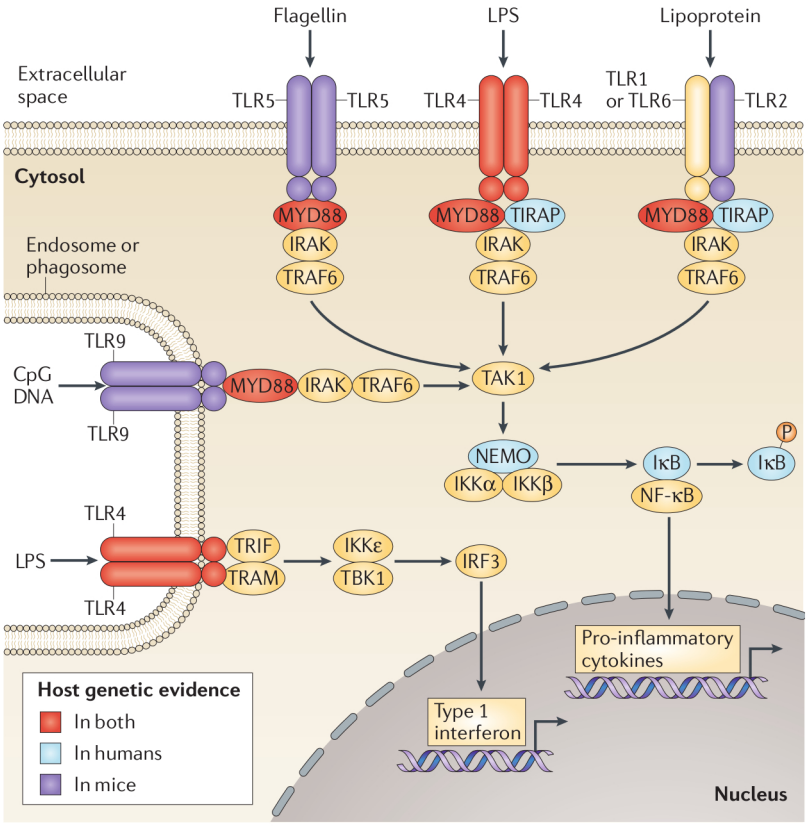

Figure 1.7 Toll-like receptor signalling in *Salmonella* infections.

Toll-like receptor (TLR) signalling in response to *Salmonella* pathogen-associated molecular patterns (PAMPs). Ligation of Salmonella PAMPS by their cognate TLR results in recruitment of cytosolic adaptor proteins and initiation of anti-bacterial transcriptional programmes. TLR2 and TLR4 stimulation facilitates activation of myeloid differentiation primary response 88 (MYD88) and Toll-interleukin-1 receptor domain-containing adaptor protein (TIRAP), thus activating interleukin-1 receptor associated kinase (IRAK). IRAK recruits tumour necrosis factor receptor-associated factor 6 (TRAF6), activating transforming growth factor-β-activated kinase 1 (TAK1). TLR5 and TLR9 PAMP ligation activate TAK1 in an analogous, but TIRAP-independent manner. Activated TAK1 acts on the

inhibitor of nuclear factor-κB (NF- κB) kinase (IKK) complex (IKKα, IKKβ, and NF-κB essential modulator - NEMO). The IKK phosphorylates inhibitor of NF-κB (IκB), which releases NF-κB, allowing nuclear translocation and initiation of NF-κB-mediated transcription. Genes for which there is host genetic evidence as a determinant of NTS susceptibility are coloured according to whether that evidence is derived from the mouse model, human studies or both. Figure reproduced from (Gilchrist et al., 2015). First published in Nature Reviews Immunology, volume 15, pages 452-63, by Nature Publishing Group.

Mice deficient in Tlr4 have increased susceptibility to early, primary NTS infection (Poltorak et al., 1998; Robson & Vas, 1972). The importance of LPS-recognition by Tlr4 in the mouse is further demonstrated by the observation that functional Tlr4 signalling compensates for loss of Tlr2 (D. S. Weiss, Raupach, Takeda, Akira, & Zychlinsky, 2004), Tlr5 (Feuillet et al., 2006) and Tlr9 (Arpaia et al., 2011) signalling. The effects of each of these TLR molecules on the outcome of a *Salmonella* infection only become apparent in mice additionally deficient in Tlr4. As would be expected, loss of TLR-signalling diminishes the ability of the host to control infection with *Salmonella*. Interestingly however, loss of signalling via multiple TLR molecules (in mice deficient in all of Tlr2, Tlr4 and Tlr9) results in less *Salmonella* susceptibility as compared to mice deficient in only two TLR molecules (either lacking both Tlr2 and Tlr4, or Tlr4 and Tlr9) (Arpaia et al., 2011). This effect appears to be secondary to a requirement for phagosomal acidification by *Salmonella* in order to induce expression of key virulence factors (e.g. via the PhoP/PhoQ regulon or SPI-2, see above). In keeping with this, while mice deficient in all TLR signalling molecules are highly susceptible to *Salmonella* infection, this susceptibility is secondary to uncontrolled extracellular replication, with pan-TLR deficient macrophages not supporting intracellular *Salmonella* growth (Sivick et al., 2014).

Inflammasome signalling

By analogy to TLRs, NOD-like receptors (NLRs) are cytosolic pattern-recognition receptors, detecting intra-cytoplasmic microbial PAMPs and signals of cellular stress, resulting in inflammasome activation. Inflammasomes are multimeric, cytoplasmic, protein scaffolds, composed of PAMP detector molecules (NLRs), adaptor molecules, and effector caspases. Inflammasome activation results in a pro-inflammatory form of programmed cell death, termed pyroptosis, characterised by loss of membrane integrity and release of pro-inflammatory cytokines (IL-18 and IL-1β) (Franchi, Warner, Viani, & Núñez, 2009).

Three distinct inflammasomes detect cytosolic *Salmonella* PAMPs. The canonical NLRP3 (NOD-, LRR-, and pyrin containing 3) and NLRC4 (NOD-, LRR-, and CARD containing 4) inflammasomes respond to *Salmonella*, with NLRC4 detecting flagellin and TTSS proteins (the specificity of the NLRP3 inflammasome with respect to *Salmonella* remains unclear) (Broz et al., 2010). The non-canonical (caspase-11 rather than caspase-1 dependent) *Salmonella*-sensing inflammasome detect *S*. Typhimurium LPS (Kayagaki et al., 2013).

Experimentation in the mouse model of invasive *Salmonella* infection has demonstrated the importance of canonical inflammasome signalling, and production of its canonical cytokines, in host control of NTS. Mice deficient in caspase-1, NLRP3-NLRC4, IL18, and IL-1β, all demonstrate diminished

ability to successfully control challenge with NTS, with increased early bacterial dissemination to the RES (Broz et al., 2010; 2013; Raupach, Peuschel, Monack, & Zychlinsky, 2006). In keeping with the redundancy observed in mice deficient in TLR signalling molecules, *Salmonella* PAMP ligation by one of NLRP3 or NLRC4 appear to be sufficient for effective host *Salmonella* control, with Salmonella-susceptibility being evident only in mice lacking both molecules (Broz et al., 2013). IL-18 and IL-1β production have complementary but distinct roles in controlling a *Salmonella* challenge. IL-18 is a key determinant of anti-*Salmonella* immunity, regardless of the route of administration, and is a key determinant of immunity to disseminated *Salmonella* infection. By contrast, IL-1β-deficient mice only have a clear *Salmonella*-susceptibility phenotype following oral challenge (Raupach et al., 2006).

Despite sensing cytosolic *S*. Typhimurium LPS, a non-redundant role for the non-canonical, caspase-11-dependent inflammasome, has been less forthcoming. Caspase-11 deficient mice appear to have no increase in susceptibility to *Salmonella* infectious challenge, and mice deficient in both caspase-1 and caspase-11 have improved *Salmonella* control as compared to mice deficient in caspase-1 alone (Broz et al., 2013). This apparently paradoxical effect appears to be secondary to a loss of neutrophil-mediated control of extracellular *Salmonella* in caspase-1-deficient mice, with delayed pyroptosis in macrophages of mice deficient in both caspase-1 and caspase-11 reducing the extracellular burden of bacteria.

Interferon-γ and tumour necrosis factor production

A key function of TLR-mediated signalling in the context of a *Salmonella* infection, is the production of the inflammatory cytokines, IL-12 and IFNγ (Kawai, Adachi, Ogawa, Takeda, & Akira, 1999). *Salmonella* infection of mammalian phagocytes results in IL-12 release, which signals via the IL-12 receptor complex on lymphoid lineage cells, resulting in the IFNγ production and, in the case of CD4+ T cells, T_H1 polarisation. Released IFNγ then serves to up-regulate anti-bacterial effector mechanisms of infected phagocytes, facilitating control and resolution of the intracellular *Salmonella* infection (Figure 1.8). IL-12-dependent IFNγ production plays a key role early control of a *Salmonella* infectious challenge in the mouse (Mastroeni et al., 1998). Innate and adaptive immune responses to *Salmonella* contribute distinct sources of IFNγ, with the principle innate source being natural killer (NK) cells (Kupz et al., 2013).

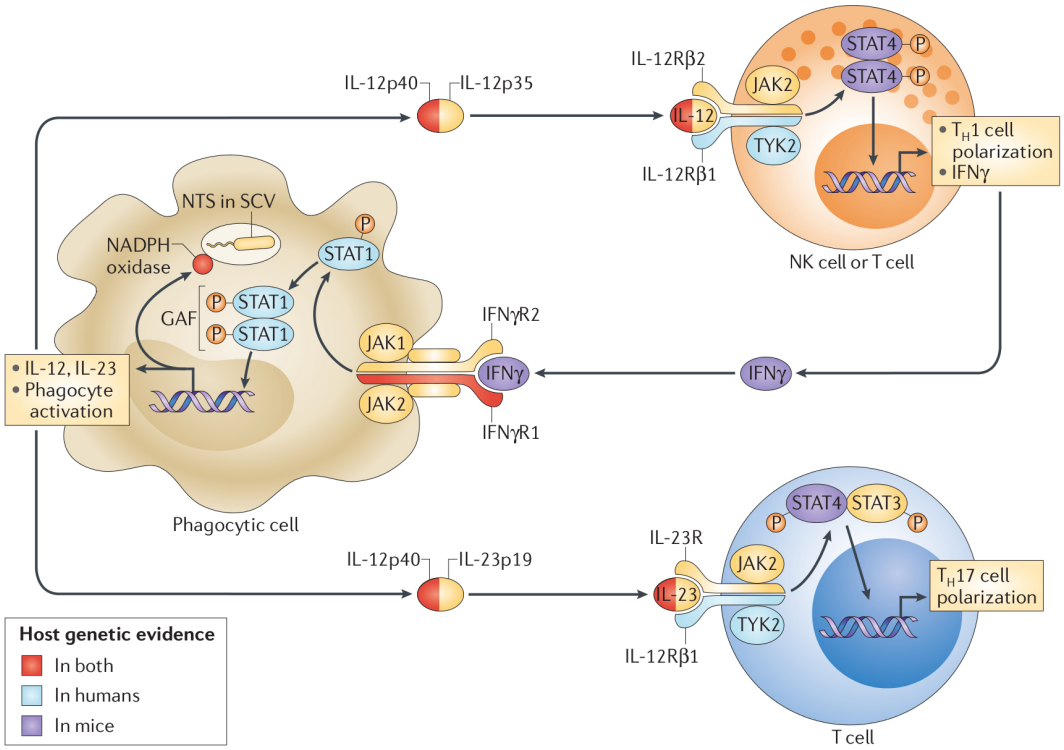

Figure 1.8 Cell-mediated immunity to intracellular nontyphoidal *Salmonella* infections.

Following uptake of nontyphoidal *Salmonella* (NTS) by a host phagocyte, IL-12 and IL-23 (hetero-dimeric cytokines with one shared and one distinct subunit) are released. In the case of IL-12, signalling occurs via the IL-12 receptor complex at natural killer (NK) cells and $CD4^+$ T cells, resulting in T_H1 polarisation and IFNγ release. IFNγ signalling at the IFNγ receptor complex on infected phagocytes results in STAT1 homo-dimerization and nuclear translocation, leading to phagocyte activation and control of the intracellular NTS infection. In the case of IL-23 signalling, signalling via the IL-23 receptor complex on CD4+ T cells results in T_H17 polarisation, facilitating gut mucosal immunity. Genes for which there is host genetic evidence of for acting as a determinant of NTS susceptibility are coloured according to whether that evidence is derived from the mouse model, human studies or both. Figure reproduced from Reference (Gilchrist et al., 2015). First published in Nature Reviews Immunology, volume 15, pages 452-63, by Nature Publishing Group.

In addition to IL-12 production, TLR signalling within *Salmonella*-infected macrophages results in tumour necrosis factor (TNF) release (Kawai et al., 1999). TNF is a poly-functional, pro-inflammatory cytokine with effects that are highly synergistic to IFNγ (Ohmori, Schreiber, & Hamilton, 1997). Interruption

of TNF signalling by either deficiency of the cytokine or its receptor results in failure to control *Salmonella* infection, with increased, early, systemic dissemination (Everest, Roberts, & Dougan, 1998; Mastroeni et al., 1991).

Intracellular anti-bacterial effector mechanisms

IFNγ signalling facilitates the cell-autonomous immunity of *Salmonella*-infected phagocytes in part through the up-regulation of anti-bacterial effector mechanisms. Key effector mechanisms in the context of an invasive *Salmonella* infection are phagolysosomal maturation, oxidative and nitrosative burst, divalent cation restriction, and autophagy (MacMicking, 2012). It is noteworthy that many key *Salmonella* virulence factors are released in response to these changes (e.g. phagolysosomal acidification, and divalent cation restriction), and function to mitigate the effects of these anti-bacterial effector mechanisms (see above).

Oxidative burst within infected phagocytes allows the production of reactive oxygen species (ROS), which have direct bactericidal properties, and their delivery to the SCV. ROS is produced by the NADPH oxidase system, and mice with deficiencies in NADPH oxidase subunits, and thus oxidative burst, are highly susceptible to *Salmonella* challenge, failing to control early *Salmonella* growth in the RES (R. T. Khan, Yuki, & Malo, 2014; Mastroeni, Vazquez-Torres, Fang, Xu, Khan, Hormaeche, et al., 2000b).

Solute carrier family 11 member 1 (*Slc11a1*, previously termed *Nramp1*) is the major host genetic determinant of early *Salmonella*-susceptibility in inbred mice (Vidal et al., 1995). *Slc11a1* encodes an IFNγ-regulated divalent cation transporter, which mediates host-directed efflux of key divalent cations (e.g. Mn^{2+} and Fe^{2+}) from the SCV (Govoni, Gauthier, Billia, Iscove, & Gros, 1997; Jabado et al., 2000), and further restricts intracellular iron availability through its up-regulation of lipocalin-2, which binds bacterial siderophores, preventing bacterial iron sequestration (Fritsche, Nairz, Libby, Fang, & Weiss, 2012). The role of iron availability in particular, in host control of *Salmonella* infections is further illustrated by the reduced *Salmonella*-susceptibility seen in mice with haemochromatosis (Hfe-deficient mice), in which macrophages become paradoxically iron-poor, despite widespread iron-overload in other tissues (Nairz et al., 2009).

In addition to the induction of bacterial effector mechanisms functioning to restrict *Salmonella* growth within phagosomes, IFNγ signalling facilitates control of cytosolic bacteria through its up-regulation of autophagy (Birmingham, Smith, Bakowski, Yoshimori, & Brumell, 2006). The role of autophagy in host control of an invasive *Salmonella* infection has been demonstrated in mice with autophagy-deficient intestinal epithelial cells. Loss of autophagy in the intestinal epithelium leads to increased bacterial dissemination to the RES (Conway et al., 2013).

Extracellular antibacterial effector mechanisms

While many host determinants of *Salmonella* infection, and many *Salmonella* virulence factors, determine the ability of *Salmonella* to survive and replicate intracellularly, the periodic escape of *Salmonellae* from within its intracellular niche, exposes it to extracellular effectors, which restrict bacterial growth and survival. LPS O-antigen length and composition are key determinants of extracellular *Salmonella* persistence (see above). This is in large part mediated by differential susceptibility to complement-mediated lysis, with differences in virulence among isogenic *Salmonella* strains differing only in O-antigen composition, being abolished in mice deficient in alternative complement pathway components (Liang-Takasaki, Saxén, Mäkelä, & Leive, 1983).

The role of neutrophils in controlling extracellular bacteria during *Salmonella* infection is being increasingly appreciated. Neutropenic mice are highly susceptible to *Salmonella* infections (Conlan, 1997). As noted above, the susceptibility of caspase-1 deficient mice to disseminated *Salmonella* infection is, in part, secondary to a decrease in neutrophil-mediated clearance of extracellular bacteria (Broz et al., 2013). Similarly, the *Salmonella*-susceptibility observed in oxidative burst deficient mice appears likely to be secondary to a loss of ROS production in neutrophils as well as *Salmonella*-infected macrophages. In keeping with this, haemolysis regardless of its aetiology, results in a neutrophil-specific defect in oxidative burst, with considerable *Salmonella* susceptibility (Ampel, Van Wyck, Aguirre, Willis, & Popp, 1989; Cunnington, de Souza, Walther, & Riley, 2012a; Roy et al., 2007; Yuki et al., 2013).

Acquired immune responses to nontyphoidal Salmonella infection

In contrast to innate immunity to invasive *Salmonella* infections, acquired immune responses are the key determinants of the late stages of a primary *Salmonella* infection, and susceptibility to a secondary infection.

T cells in nontyphoidal Salmonella infections

Athymic and αβ-T cell-deficient mice control the early phases of a *Salmonella* infection in the normal manner, but are unable to clear Salmonella during late primary infection, and fail to generate protective responses following vaccination (Hess et al., 1996; Sinha, Mastroeni, Harrison, de Hormaeche, & Hormaeche, 1997). $CD4^+$ T cells are required for late phase primary immunity, as are $CD8^+$ T cells, but to a lesser extent (Hess et al., 1996; S.-J. Lee, Dunmire, & McSorley, 2012b). Both $CD4^+$ and $CD8^+$ T cells are required for effective immunity to secondary infection (Mastroeni, Villarreal-Ramos, & Hormaeche, 1992). In keeping with the importance of IFNγ production in the innate stages of a primary *Salmonella* infection, $CD4^+$ T cell immunity to secondary *Salmonella* infection appears to be in large part mediated by acquisition of a T_H1 phenotype, with IFNγ-deficient mice exhibiting decreased anti-*Salmonella* immunity following vaccination (Mastroeni et al., 1992). The role of $CD4^+$ T cell immunity in acquired immunity to *Salmonella*, is further demonstrated by linkage studies in inbred mice, identifying MHC (H2) haplotypes as key genetic determinants of late primary infection in inbred mice (Hormaeche, Harrington, & Joysey, 1985).

B cells and antibody in nontyphoidal Salmonella infections

In the mouse model of invasive *Salmonella* infection, B cells and antibody are dispensable in the control of a primary infectious challenge (Mastroeni, Simmons, Fowler, Hormaeche, & Dougan, 2000a). In secondary infections, following prior infectious challenge or vaccination, however, B cells are required in order to generate effective immunity (Mastroeni, Simmons, Fowler, Hormaeche, & Dougan, 2000a). Interestingly, this requirement for B cells extends beyond antibody production, with passive transfer of immune serum not fully restoring immunity of B cell-deficient mice to secondary infectious challenge (Mastroeni, Villarreal-Ramos, & Hormaeche, 1993). The antibody-independent functions of B cells in anti-*Salmonella* immunity appear to, in large part, be secondary to a role for B cells in generation of robust $CD4^+$ T_H1 responses (Nanton, Way, Shlomchik, & McSorley, 2012). The importance of antibody in African children has been demonstrated by the observation that anti-*Salmonella* antibody acquisition is a correlate of protection against invasive NTS disease (MacLennan et al., 2008).

Acquired impairment of anti-Salmonella immunity

In African children, three well-defined, acquired risk factors modify risk of invasive NTS disease; HIV infection, malaria, and malnutrition. The predicted effects of these risk factors on immunity to invasive NTS disease are summarised in Figure 1.9.

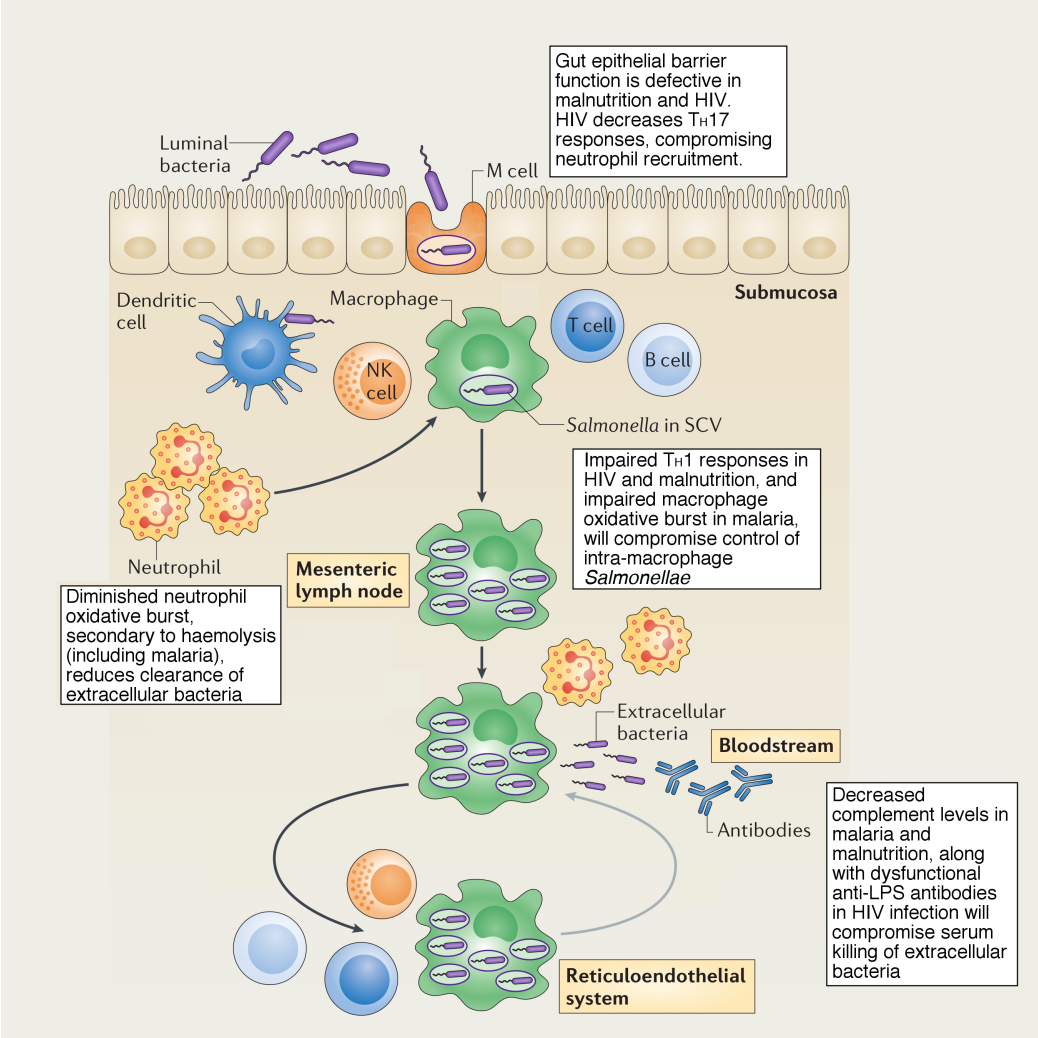

Figure 1.9 Acquired impairment of anti-*Salmonella* immunity in African children.

Predicted consequences of HIV, malaria and malnutrition on immunity to nontyphoidal *Salmonella* (NTS) infection in African children. Text boxes highlight steps in the immunopathogenesis of a disseminated NTS infection predicted to be affected by NTS-associated co-morbidities. Figure modified from (Gilchrist et al., 2015). First published in Nature Reviews Immunology, volume 15, pages 452-63, by Nature Publishing Group.

Effect of HIV co-infection on anti-NTS immunity

Loss of $CD4^+$ T cell numbers is a key clinical indicator of progression of HIV infection and increased immunodeficiency, with increased risk of opportunistic infection. In African adults with NTS bacteraemia, HIV infection is the dominant risk factor for disease, with the majority of these individuals demonstrating significant immunosuppression (median $CD4^+$ T cell count at presentation; 99 cells/μl, range 6-313) (Gordon et al., 2002). While HIV co-infection in African children with NTS bacteraemia is less frequent than in adults, it remains an important risk factor, with a median study prevalence of 20% (Table 1.3).

The immune defects seen in HIV infection contribute to invasive NTS susceptibility in several ways. Firstly, HIV-induced depletion of the $CD4^+$ T cell compartment leads to the loss of T_H1 cells (Geldmacher et al., 2008), and loss of *Salmonella*-specific IFNγ responses. Diminished IFNγ production capacity will result in inadequate induction of anti-bacterial effector mechanisms in *Salmonella*-infected phagocytes, resulting in a failure to control intracellular infections. This appears to have two consequences, firstly a greatly increased risk of developing an episode of invasive NTS disease, but also a tendency towards recurrent infections, with recrudescence hypothesized to be occurring from intracellular sanctuary sites (Okoro, Kingsley, Quail, et al., 2012b). In addition to the loss of T_H1 $CD4^+$ T cells, gut mucosal T_H17 cells are preferentially depleted in early HIV infection (Dandekar, George, & Bäumler,

2010). In the primate model of *Salmonella* infection, simian immunodeficiency virus (SIV)-infected macaques demonstrated attenuated gut mucosal T_H17 responses (Raffatellu et al., 2008). Attenuated T_H17 responses impair intestinal recruitment of neutrophils, which facilitates bacterial dissemination and establishment of systemic disease. Attenuated T_H17 responses, with diminished inflammatory cell recruitment to the infected gut, may in part explain why diarrhoea is uncommon among Africans with invasive NTS disease (Table 1.2).

In addition to impaired cell-mediated immunity, antibody dysfunction in the context of HIV infection is well-described. HIV-infected African adults generate high titres of anti-LPS IgG, which is able to fix complement, but is not bactericidal, with assembled MACs unable to insert into the bacterial outer membrane (MacLennan et al., 2010). The resultant impairment of extracellular killing has clear parallels to the modification of LPS O-antigen length and composition as a virulence determinant in NTS strains (see above). This raises the hypothesis that these susceptibility and virulence determinants may have interacted, and the emergence of HIV, with dysfunctional antibody immunity, may have driven selection of invasive NTS isolates of a given O-antigen chain length in Africa.

Effect of malaria on anti-NTS immunity

In common with HIV co-infection, malaria infection impairs the host's ability to control both intracellular and extracellular *Salmonellae*. In macrophages,

malaria pigment ingestion impairs oxidative burst capacity (Schwarzer et al., 1992), which will impair intracellular control of co-existing NTS infection. Complement consumption has been documented in the context of malaria infection (Nyakoe, Taylor, Makumi, & Waitumbi, 2009), which in addition to impairing bactericidal killing of extracellular NTS, may also interfere with induction of neutrophil oxidative burst (Gondwe et al., 2010). In keeping with this, haemolysis regardless of its aetiology (but including secondary to malaria infection), via induction of haem oxygenase-1 drives the egress of an immature population of oxidative burst-deficient neutrophils from the bone marrow, impairing neutrophil-mediated control of extracellular NTS (Cunnington, de Souza, Walther, & Riley, 2012a; Cunnington, Njie, Correa, Takem, Riley, et al., 2012b). This association with haemolysis, rather than malaria *per se*, may in part explain the increased association of NTS with severe malarial anaemia (Bronzan et al., 2007), and the greater association between NTS and recent malaria than with concurrent malaria (Brent et al., 2006).

Effect of malnutrition on anti-NTS immunity

Under-nutrition has pleiotropic effects on the immune system, which will impair the ability of the host to control a *Salmonella* exposure, both within the gut and during systemic disease, and for intracellular and extracellular bacteria. Epithelial barrier function is impaired in malnourished children, facilitating translocation of enteric bacteria into blood, facilitating systemic bacterial dissemination (Brewster, Manary, Menzies, O'Loughlin, & Henry,

1997). Children with malnutrition have reduced levels of complement components (particularly C3) (Rytter, Kolte, Briend, Friis, & Christensen, 2014), and neutrophils from under-nourished children have diminished microbicidal capacity (Schopfer & Douglas, 1976), both of which will impair the clearance of extracellular NTS.

Intracellular control of invasive NTS infection is also very likely to be deficient in malnourished children. T_H1 cytokines are reduced (in favour of a T_H2 response), which will impair the ability of *Salmonella*-infected macrophages to respond to intracellular bacteria (Rytter et al., 2014). The reduced circulating B cell numbers seen in malnourished children may also impair the host's ability to resolve an NTS infection (Rikimaru, Taniguchi, Yartey, Kennedy, & Nkrumah, 1998). As reduced B cell numbers appear to not be accompanied by either reduced antibody levels or altered vaccine responses (Prendergast, 2015), any effect of reduced B cell numbers appears likely to be independent of antibody. By analogy to the mouse model, this mechanism may be to further impair generation of T_H1 responses to Salmonella antigens.

Adult-onset, acquired immunodeficiency with invasive NTS infection

A group of patients with an adult-onset, acquired immunodeficiency, in the absence of HIV infection, has been recently described, presenting with invasive NTS disease and poorly-pathogenic mycobacteria (Browne et al., 2012). The immune defect in these individuals has been demonstrated to be secondary to the presence of auto-antibodies directed against IFNγ. While

anti-cytokine antibodies are unlikely to represent a common cause of invasive NTS infection in African children, the observation is informative, as it underscores the importance of IFNγ-mediated immunity in host control of NTS.

A further patient group, with considerable potential to provide insights into immunity to invasive infection, comprise individuals with autoimmune disease receiving targeted immunosuppressive agents. By analogy to observations from the mouse model of invasive NTS infection, patients treated with anti-TNF agents would be predicted to have increased risk of invasive NTS disease. In a Spanish cohort of anti-TNF treated individuals with rheumatoid arthritis, NTS infections were not found to be more common than in patients not treated with anti-TNF agents. That said, the rates of invasive infection among anti-TNF treated patients was substantial, with 7 of 17 developing bacteraemia (Peña-Sagredo et al., 2009).

Genetic susceptibility to invasive nontyphoidal *Salmonella* disease in humans

Primary immunodeficiencies predisposing to invasive nontyphoidal Salmonella disease

Studies of rare primary immunodeficiencies in humans have allowed our understanding of anti-*Salmonella* immunity provided by the mouse model to be extended into the sphere of human disease. The complementary nature of insights from these studies, epidemiological data defining acquired risk factors, and studies in the mouse model is striking, and together have produced a highly detailed model of NTS immunopathogenesis. In addition to confirming the relevance of mouse susceptibility loci in humans, the relative frequencies of NTS infections within these immunodeficiency states are informative. For instance, individuals with Mendelian susceptibility to mycobacterial disease (MSMD) and sickle cell disease are highly enriched for NTS as a causative agent, whereas individuals with TLR pathway defects are susceptible to invasive *Salmonella* infection as part of a much broader immunodeficiency.

Mendelian Susceptibility to Mycobacterial Disease

MSMD is a genetically heterogeneous group of primary, monogenic immunodeficiencies, characterised by extreme susceptibility to poorly pathogenic mycobacteria and NTS infections (Table 1.5).

To date defects causing MSMD have been identified in ten genes; eight of which are autosomal (*IFNGR1, IFNGR2, STAT1, IRF8, ISG15, IL12B, IL12RB1,* and *TYK2*) and two X-linked (*CYBB* and *IKBKG*). The immunological roles of these genes are highly related, all encoding products which function to mediate either IL-12-dependent IFNγ production or IFNγ-stimulated effector functions (Figure 1.8). Invasive NTS disease has been described in patients with MSMD caused by mutations in *IFNRG1* (Dorman et al., 2004; Newport et al., 1996), *IL12B* (Prando et al., 2013), *IL12RB1* (de Beaucoudrey et al., 2010), *STAT1* (Averbuch, Chapgier, Boisson-Dupuis, Casanova, & Engelhard, 2011), *TYK2* (Minegishi et al., 2006), and *IKBKG* (Filipe-Santos, 2006).

Table 1.5 Susceptibility to NTS infection at Mendelian Susceptibility to Mycobacterial Disease loci

Mendelian susceptibility to mycobacterial disease			Complementary mouse model evidence			References
Gene	Phenotype & inheritance	Reported NTS infections	Variation	Challenge (strain, route)	Phenotype	
IL12B	MSMD - complete, autosomal recessive	11 of 44 (25%)	*Il12p40*$^{-/-}$	Attenuated *S.* Enteritidis (*ade*$^{-}$/*his*$^{--}$) i.p.	Increased mortality & increased RES bacterial load D3, 4, 20	(Lehmann et al., 2001; Prando et al., 2013)
IL12RB1	MSMD – complete and partial autosomal recessive	57 of 132 (43%)	*Il12rb1*$^{-/-}$	Murine IL-12 i.p. (daily for 5 days)	Reduced IFNγ production D5	(de Beaucoudrey et al., 2010; Jong, 1998; Wu, Ferrante, Gately, & Magram, 1997)
TYK2	Hyper-IgE syndrome with MSMD – autosomal recessive	1 of 2 cases	No studies			(Minegishi et al., 2006)
IFNGR1	MSMD – complete autosomal recessive, and partial, dominant & recessive	5 of 60 (8%)	*Ifngr*$^{-/-}$	Attenuated STm SL7207 (*aroA*$^{-}$) i.v.	Increased mortality from D16	(Dorman et al., 2004; Hess et al., 1996; Jouanguy et al., 2000)
STAT1	MSMD – autosomal recessive partial	2 of 9 (22%)	No studies			(Averbuch et al., 2011)
IKBKG	MSMD - X-linked recessive	1 of 6 cases	No studies			(Filipe-Santos, 2006)
IFNGR2	MSMD – complete autosomal recessive, and partial, dominant & recessive	None	No studies			
IRF8	MSMD - autosomal dominant negative	None	No studies			
ISG15	MSMD with intracranial calcification - complete, autosomal recessive	None	No studies			
CYBB	MSMD - X-linked recessive	None	No studies			

MSMD, Mendelian susceptibility to mycobacterial disease; NTS, nontyphoidal *Salmonella*

Alongside complementary evidence from the mouse model, these studies have established the importance of IFNγ-mediated immunity in host control of NTS infections. In addition to highlighting the importance of IFNγ production in *Salmonella* infections, the distribution of NTS infections observed in patients with MSMD also highlights the importance of T_H17-mediated immunity in NTS host defence. Individuals with MSMD caused by mutations in the IL-12p40 subunit or the IL-12Rβ1 receptor chain, are at far greater risk of NTS disease than are individuals with mutations in genes affecting IFNγ signalling (MacLennan et al., 2004). This may be secondary to the IL-12p40 and IL-12Rβ1 cytokine and receptor subunits also contributing to IL-23 signalling and thus T_H17 immunity (Figure 1.8).

Primary immunodeficiencies affecting TLR signalling

In keeping with increased susceptibility to NTS infection seen in mice deficient in TLR signalling molecules, invasive NTS infections have been described in patients with immunodeficiencies caused by mutations in TLR signalling molecules. Loss of function mutations in *IKBKG* (encodes nuclear factor-κB essential modulator - NEMO) cause anhydrotic ectodermal dysplasia with immunodeficiency (EDA-ID). The immunodeficiency associated with EDA-ID is characterised by susceptibility to invasive bacterial infection, including NTS in 4% of cases (Hanson et al., 2008). Recurrent, invasive NTS infections have also been described in a patient with complex mosaicism for a dominant, hypermorphic *NFKBIA* (encodes inhibitor of NF-κBα) mutation, who lack IL-12 prouction in stimulated monocytes, but with no other apparent phenotype

(Janssen et al., 2004). Patients with immunodeficiencies secondary to MYD88 deficiency are susceptible to pyogenic bacterial infections, with NTS infections being reported in 7% of cases (Picard, Casanova, & Puel, 2011).

Chronic granulomatous disease

In keeping with data from studies demonstrating NTS susceptibility in mice lacking oxidative burst (see above), humans with chronic granulomatous disease (CGD), a primary immunodeficiency characterised by defective oxidative burst (Segal, Leto, Gallin, Malech, & Holland, 2000), are at increased risk of invasive NTS infections. Patients with CGD are susceptible to invasive infections, commonly complicated by abscess formation, with catalase-positive organisms (*Staphylococcal* and *Aspergillus* species being most common) (Winkelstein et al., 2000). However, NTS are the commonest cause of bacteraemia in patients with CGD (Winkelstein et al., 2000).

Sickle cell disease

It has been recognised for many years that NTS are a common cause of invasive infection, especially osteomyelitis, in the context of sickle cell disease (Adeyokunnu & Hendrickse, 1980). In addition to osteomyelitis, NTS is an important cause of bacteraemia in African children with sickle cell disease, and sickle cell disease is a highly significant risk factor for invasive NTS disease (Williams et al., 2009). The immune dysfunction associated with sickle cell disease is multi-faceted, contributed to by functional asplenia

(Pearson, Spencer, & Cornelius, 1969), defective opsono-phagocytosis and serum bactericidal activity secondary to alternative pathway complement deficiency (Johnston, Newman, & Struth, 1973), and impaired granulocyte oxidative burst secondary to haemolysis (Cunnington, de Souza, Walther, & Riley, 2012a). Impaired complement-mediated lysis and neutrophil oxidative burst are likely to be responsible for a significant portion of the NTS-susceptibility seen in sickle cell disease, both being predicted to compromise the host's ability to clear extracellular NTS.

Immunodeficiencies characterised by T cell, antibody and complement deficiency

In contrast to the mouse model and the extreme NTS susceptibility seen in HIV positive patients, T cells immunodeficiency states are not typically characterised by invasive NTS disease. There are no reports of NTS infection in severe-combined immunodeficiency or complete DiGeorge syndrome, although it should be noted that patients with these conditions present in early infancy with invasive infections requiring early haematopoietic stem cell transplant, and so their window for potential *Salmonella* exposure may well be restricted. That said milder T cell immunodeficiency states are also not commonly associated with invasive NTS disease, with NTS reported in 3 of 259 cases of idiopathic $CD4^+$ T cell lymphocytopoenia (Ahmad, Esmadi, & Steinmann, 2013). NTS infections may be more common in MHC class II deficiency, with NTS infection reported in 7 of 30 cases (although it is unclear

whether these were cases of invasive disease) (Klein, Lisowska-Grospierre, LeDeist, Fischer, & Griscelli, 1993).

Despite the clear role of protective antibody acquisition in protecting African children against invasive NTS disease, NTS infections are unusual complications of antibody deficiency states. NTS infection is described rarely in common variable immunodeficiency (Hermaszewski & Webster, 1993), X-linked agammaglobulinaemia (Winkelstein et al., 2006), and X-linked hyper-IgM syndrome (Levy, Espanol-Boren, Thomas, & Fischer, 1997), and when NTS does cause disease it manifests as chronic or recurrent diarrhoea without systemic disease.

While defective complement-mediated lysis has been identified as a contributory factor underlying several well-established risk factors for invasive NTS disease (e.g. sickle cell disease, malaria, HIV infection), isolated complement deficiency states are rarely associated with invasive NTS disease. In a large series of patients with complement deficiency, invasive NTS infection was described in only 2 of 242 patients(Ross & Densen, 1984), suggesting that while complement dysfunction can contribute to NTS susceptibility it appears not to represent a dominant determinant of disease in isolation.

Population-based studies of genetic susceptibility to invasive nontyphoidal Salmonella disease

With the exception of studies investigating the role of sickle cell disease as a determinant of risk of NTS bacteraemia in African children (Williams et al., 2009), studies evaluating genetic risk factors for invasive NTS disease in human populations are limited. Two candidate gene studies of bacteraemia in Kenyan children, in which NTS was the second most common cause of disease, have evaluated the role of genetic variation in *CISH* (Khor et al., 2010) and *TIRAP* (Khor et al., 2007) as risk factors for disease. A missense variant in *TIRAP* is associated with bacteraemia in Kenyan children, supporting the hypothesis that variability in TLR signalling is an important determinant of NTS susceptibility in human populations. Similarly genetic variation in *CISH* was found to associate with risk of all-cause bacteraemia. CISH is a negative regulator of IL-2 signalling, a key determinant of T cell development and differentiation, implicating variation in anti-bacterial T cell responses in disease susceptibility at the population level. That both the *TIRAP* and *CISH* variants are associated with all-cause bacteraemia susceptibility, is also in keeping with human primary immunodeficiency data, and experimentation in mouse models, demonstrating that defects in TLR signalling and T cell-mediated immunity results in broad immune defects, with susceptibility to a wide range of pathogens. To date, however, there have been no published unbiased assessments of genetic risk of NTS bacteraemia in unselected populations.

In contrast to the lack of studies assessing genetic risk of NTS disease in human populations, several candidate gene studies of typhoid fever have been reported, and more recently the first genome-wide association study of typhoid fever (Table 1.6).

Table 1.6 Population-based, genetic association studies of enteric fever susceptibility

Gene	Disease	Study design	Population	Variant	*P*	OR	Reference
TLR4	Typhoid	Candidate gene	Malay (304 cases, 250 controls)	rs4986790	0.014	2.54	(Bhuvanendran et al., 2011)
	Typhoid	Candidate gene	Vietnamese (372 cases & controls)	10 variants	NS	NS	(Hue et al., 2009)
TLR5	Typhoid	Candidate gene	Vietnamese (565 cases, 281 controls)	rs5744168	NS	NS	(Dunstan et al., 2005)
CASP1 & *CASP11*	Typhoid & paratyphoid	Candidate gene	Indonesian (116 cases, 659 controls)	rs580253	NS	NS	(Ali et al., 2007)
IL1B	Typhoid & paratyphoid	Candidate gene		rs1143634 rs16944	NS	NS	
IL12B	Typhoid & paratyphoid	Candidate gene		rs17860508	NS	NS	
IFNG	Typhoid & paratyphoid	Candidate gene		rs2430561	NS	NS	
IFNGR1	Typhoid & paratyphoid	Candidate gene		*IFNGR1* $(CA)_n$ repeat	NS	NS	
TNF & *TNFRSF1A*	Typhoid	Candidate gene	Vietnamese (380 cases & controls)	MHC class III haplotype	Bayesian analysis; prior probability = 0.5, posterior probability = 0.82		(Dunstan et al., 2007)
	Typhoid & paratyphoid	Candidate gene	Indonesian (116 cases, 659 controls)	rs361525 rs1800587 rs767455	NS	NS	(Ali et al., 2007)
SLC11A1	Typhoid	Candidate gene	Vietnamese (217 cases, 288 controls)	274C/T, 469+14G/C, 1465-85G/C, D543N, Promoter (GT)n repeat, D2S1471	NS	NS	(Dunstan et al., 2001)
PARK2	Typhoid & paratyphoid	Candidate gene	Indonesian (116 cases, 659 controls)	rs935658	0.03	1.51	(Ali et al., 2006)
MHC	Typhoid with complications	Candidate gene	Javanese (19 cases, 44 uncomplicated)	*HLA-DRB1*12*	0.05	0.3	(Dharmana et al., 2002)
	Typhoid & paratyphoid	GWAS	Vietnamese (583 cases, 2679 controls) Nepalese (595 cases, 386 controls)	*HLA-DRB1*04:05*	$2.6x10^{-11}$	0.14	(Dunstan et al., 2014)

NS, not significant; GWAS, genome-wide association study; OR, odds ratio

In other contexts, attempts to replicate genetic associations identified using candidate gene methodologies have been disappointing. In some instances this will be secondary to heterogeneity of genetic effects between studied populations, but a large percentage of these associations seem likely to be false-positives, in the context of small, underpowered studies, reporting results with insufficiently robust correction for multiple association testing. In light of these concerns, apparent associations with typhoid, identified in candidate gene studies, with marginal evidence of significant association, should be interpreted with caution. By contrast, the recently published genome-wide association study (GWAS) of enteric fever in Vietnamese and Nepalese individuals (Dunstan et al., 2014) provides convincing evidence for *HLA-DRB1* alleles acting as susceptibility determinants for enteric fever. The relevance of these data to invasive NTS infections in African children is however questionable. Despite the causative agents representing serovars of the same bacterial species, the lack of a shared set of acquired risk factors for enteric fever and invasive NTS disease, suggests that a shared genetic architecture between the two diseases is perhaps unlikely.

Conclusions and overview of thesis

NTS is an important cause of invasive infection in African children. Despite limited availability of robust epidemiological data describing the incidence of invasive bacterial disease in African children, in African settings for which data are available, NTS consistently imposes substantial paediatric morbidity and mortality. The limited number of centres reporting longitudinal data have consistently reported declining incidence of NTS bacteraemia in African children over the last two decades, in parallel with declining malaria incidence at these sites. Despite this, NTS remains a leading cause of invasive bacterial infection in these settings, and the increased provision of available anti-bacterial vaccines is likely to increase the relative importance of NTS as a pathogen.

Studies of the clinical features of African children with NTS bacteraemia consistently describe a clinical presentation characterised by fever with no pathognomonic signs or symptoms. Diagnosis is limited to blood culture, which, in the rare African setting in which bacterial culture is available, delays diagnosis. The lack of diagnostic features or tests available at presentation, combined with high rates of MDR among African NTS isolates, all contribute to the high rates of NTS-associated mortality borne by African children. Improving our understanding of the biological determinants of invasive NTS infections in African children will facilitate the delivery of much-needed, novel control strategies for NTS in Africa.

Epidemiological data have identified both heritable and acquired risk factors for disease in African children including sickle cell disease, HIV infection, current or recent malaria infection, and malnutrition. In addition to these host susceptibility factors, NTS serovars causing invasive disease in Africa are distinct, with a dominant sequence type of *S*. Typhimurium (ST313), responsible for much of the epidemic of invasive disease, rarely encountered outside Africa. A distinct pattern of genomic degradation seen in these isolates, suggests that they may be in the process of becoming more host-restricted, although they currently retain the ability to cause invasive disease in species other humans. It remains unclear whether there are phenotypic correlates of the ST313 pathovar, which contribute to its invasive potential in African populations.

A detailed picture of the host and pathogen determinants of invasive NTS disease has emerged through studies conducted in the mouse model of invasive *Salmonella* infection. These studies have been highly informative, describing striking overlap between *Salmonella* virulence factors and host immune factors in determining the ability of *Salmonella* to invade, persist intracellularly, and survive extracellularly. These findings have been complemented by studies of human primary immunodeficiencies causing susceptibility to *Salmonella*, revealing considerable overlap between susceptibility factors identified in the mouse model and determinants of disease in human patients. It is, however, unclear the extent to which these findings are generalizable to invasive infection in unselected humans populations.

In this thesis, I address this key question by defining the genetic basis of susceptibility to invasive NTS infections in African children. Starting by performing the first genome-wide association study of NTS bacteraemia, the overarching aim of this thesis is to identify, validate and functionally characterise, novel genetic correlates of susceptibility to NTS in African children. It is hoped that the improved understanding of host determinants of invasive *Salmonella* infection in African children that this provides, will facilitate the development and delivery of much-needed, novel control strategies for invasive NTS infection in Africa.

Chapter two presents a general discussion of methods and concepts employed in modern genetic epidemiology studies, with particular reference to techniques used throughout this thesis. Detailed experimental and analytical methods are provided in the relevant results chapters (chapters three to six).

Chapter three describes the first GWAS of NTS bacteraemia; including a discovery GWAS in Kenyan children and a replication analysis in Kenyan and Malawian children. Within this analysis, I further present a candidate gene analysis of previously well-defined, infection-associated loci.

In chapter four, I functionally characterise and contextualise the single validated, novel genetic correlate of invasive NTS disease identified in the GWAS and replication analysis, an intronic SNP in *STAT4*. I present data identifying the regulatory consequences of the NTS-associated SNP on

STAT4 RNA expression and downstream IFNγ protein production. Beyond its association with NTS, we assess the evidence for *STAT4* association with bacteraemia secondary to other pathogens and autoimmune diseases. In addition, I present an analysis of whether SNPs associated with NTS are globally enriched for genes perturbing IFNγ immunity.

Chapter five describes the validation and functional characterisation of *EVI5L* as a susceptibility locus for NTS bacteraemia. Here I use expression quantitative trait locus (eQTL) datasets to prioritise a discovery phase NTS-associated locus (*EVI5L*) without evidence of replication. I use RNA interference to define EVI5L-associated phenotypes in the context of *in vitro* models of epithelial and phagocytic NTS infection, and a genotype-selectable bioresource to investigate its role in IFNγ production and immunity.

Chapter six presents a pathway-based enrichment analysis of the NTS bacteraemia GWAS, describing methionine metabolism as a determinant of invasive NTS disease. I further characterise the role of methionine metabolism in NTS infection, using λ-red recombinase-driven homologous recombination to generate *Salmonella* mutants deficient in key methionine metabolism enzymes and regulators, characterising these mutants using *in vitro* models of *Salmonella* infection.

In chapter seven, I provide a discussion of the key findings presented in this thesis. I further discuss the limitations of the work presented, and the opportunities it suggests for future work.

Chapter 2 – Methods: genetic & statistical concepts

In this chapter, I outline key genetic and statistical concepts and methods used throughout this thesis. It is intended to outline a framework within which the presented genetic association analyses were conducted. As such, chapters three to six each contain methods sections describing the details of the statistical and experimental methods particular to those chapters.

Genetic epidemiology

Genetic epidemiology aims to define genetic determinants of diseases or traits within a population. Studies identifying the genetic determinants of a trait in a population have historically proceeded with efforts to establish that a trait clusters within families, that the observed familial clustering is heritable, i.e. that is has a genetic component, through to efforts to model its inheritance. More recently, molecular genetics has facilitated efforts to map the genetic determinants of a trait, be that at the level of the chromosome, the gene or a genetic variant. While the focus of this thesis in defining the genetic epidemiology of invasive NTS disease in African children makes use of modern genetic association studies (specifically GWAS), it is informative to

consider these approaches in the wider context of genetic epidemiological studies. The discussion in this chapter focuses on defining genetic determinants for complex traits; that is traits whose occurrence is determined by multiple interacting genetic and non-genetic factors; and traits that are binary, although the described techniques are in large part readily applicable to continuous traits.

Familial clustering, heritability and segregation analysis

The simplest approach to the estimation of whether a trait has a genetic component is to assess the evidence for clustering of that trait within families. This is most commonly estimated, for a binary trait, as the recurrence risk ratio for a given relative type (λ_R), most commonly the sibling risk ratio (λ_S), which is calculated as the ratio of disease prevalence in relatives of affected family members to the disease prevalence in the general population (Burton, Tobin, & Hopper, 2005).

Familial clustering of a trait can be secondary to genetic and non-genetic effects. The genetic contribution to liability of a trait within such studies can be estimated as the proportion of total variance explained by additive genetic effects: the trait's heritability. The additive genetic effects for these estimates can be modelled within family studies according to the expected sharing of alleles according to the degree of relative. Twin studies, comparing disease liability between monozygotic and dizygotic twin pairs, allow estimation of

heritability in the broader sense, that is the proportion of trait variance accounted for by total genetic effects.

The observed familial clustering of a trait can be further used to estimate the likely existence of a major trait-associated gene(s) and its likely mode of inheritance. Segregation analysis of the distribution of a trait within multi-case extended pedigrees allows estimation of the number, frequency and inheritance of major genetic determinants of the trait of interest within studied families (Elston, 1981).

Approaches to mapping trait-associated genetic determinants

Linkage studies

In a manner analogous to segregation analysis, linkage analysis assesses co-segregation of genetic markers of known genomic position with the trait of interest within families to estimate the chromosomal region containing a disease associated genetic determinant (Risch, 1990). This approach is based on the concept of linkage. Two genetic loci are linked within a family if their co-transmission occurs more frequently than would be expected if they were independently inherited. Linkage is related to linkage disequilibrium (LD), with the latter describing the tendency of alleles at two loci to be co-inherited at the population level. Markers in LD will be in linkage within families, but the converse is not necessarily true, with recombination frequencies required to maintain LD at the population level being much lower

than those required to maintain linkage within a family over a small number of generations. This means that while markers can be linked within families over large genetic distances, the distances that LD can operate over are much smaller. A consequence of this is that linkage-based mapping of trait-associated genetic loci implicate genomic regions far larger than LD-based mapping approaches.

Genetic association studies

Linkage studies have been highly successful in identifying genetic determinants of large effect causing essentially monogenic disorders (Peltonen & McKusick, 2001). They are, however, inefficient tools for the detection of trait-associated genetic determinants of modest effect. Rather than using family-based sample collections, genetic association studies test for association of a genetic marker with a trait in unrelated individuals. A key difference between the methodologies is that while in a linkage study the allele linked to the underlying genetic determinant of disease can vary between families, in a genetic association the linked allele is shared across the population. In common with linkage studies, evidence for location of the trait-associated genetic determinant is based, in the majority of circumstances, on indirect observation of the causative genetic variant. However, as the genetic markers in a genetic association study are linked to trait-associated loci by LD, rather than by within-family linkage, genetic association studies permit far more accurate localisation of trait-associated genetic variation than is possible with linkage studies. In addition, the

feasibility of recruiting large numbers of unrelated case and control samples, as compared to family-based studies, allow genetic association studies to efficiently identify genetic loci of modest effect. Genetic association studies can test for association between genetic determinants and a trait of interest at the level of a candidate genetic variant, a candidate gene, or across the whole genome. Replication of findings from candidate variant and candidate gene studies, in which the study's genetic marker is specified based on prior knowledge of a putative biological role for that gene in the disease of interest, has been largely disappointing (Malaria Genomic Epidemiology Network, 2014). By contrast, large-scale genetic association studies testing for trait-associated genetic variation across the entire genome, GWAS, have been highly successful in identifying robust genetic correlates of hundreds of traits.

Genome-wide association studies

A GWAS proceeds by testing for association between alleles at typed genetic markers and a trait of interest in unrelated individuals (Burton et al., 2007). In common with linkage studies, GWAS are unbiased in that they assume no prior knowledge of the genomic location of trait-associated genetic determinants. For GWAS, as markers are linked to underlying, unobserved genetic determinants by LD, rather than linkage, the required marker density is greatly increased as compared to linkage studies. This requirement has been met by single nucleotide polymorphism (SNP) genotyping. SNPs are single base-pair genetic loci at which at least two possible alleles are present in a population. Their high frequency across coding and non-coding regions of

the genome, combined with the ability to genotype large numbers of SNPs in parallel, have made SNPs the dominant genetic marker used in GWAS. Modern design of SNP genotyping arrays allows massively parallel genotyping of hundreds of thousands of SNPs. Genotyping arrays are designed to capture the largest amount of unobserved genetic variation by LD. It is important to note, however, that until recently, these chips were designed in for use in European populations, using patters of LD in populations of European ancestry to inform their design. Consequently, their capture of unobserved African variation has been suboptimal (Teo, Small, & Kwiatkowski, 2010).

Quality control considerations in genome-wide association studies

The massively parallel number of association tests performed within a GWAS makes them vulnerable to spurious association results. To mitigate this, quality control (QC) filters are applied at both the sample and SNP levels. Specific thresholds for QC filters are detailed in experimental methods in chapters three to six. The following describes their rationale and calculation.

Missingness

SNPs with high rates of missingness are removed prior to imputation and association analysis. High rates of missingness at a SNP suggest a poorly performing assay with a high likelihood of erroneous genotype calls. Similarly, samples with high rates of missing data across all genotyped SNPs are likely

to indicate a poor quality or contaminated DNA sample resulting in unreliable genotypes, and are also removed prior to imputation and association testing. Equally, however, it is important to appreciate that while stringent thresholds for genotype calling will improve the accuracy of genotypes included in the study, missingness can also be a function of genotype (if one genotype at a locus is particularly difficult to assign), which also has the potential to introduce bias (McCarthy et al., 2008). For these reasons the missingness/call-rate threshold adopted aims to balance both of these competing considerations.

Hardy-Weinberg Equilibrium

Hardy-Weinberg equilibrium (HWE) describes the expected distribution of genotypes at a locus given the observed allele frequencies in the population (Hardy, 1908). It assumes that carriage of an allele at a locus is independent of the identity of the allele at the same locus on the homologous chromosome. Where *p* and *q* represent allele frequencies at a bi-allelic locus of alleles A and a respectively, the expected genotype frequencies are as follows:

$$p^2 + 2pq + q^2 = 1$$

Where p^2 and q^2 give the AA and aa homozygous genotype frequencies, and $2pq$ the Aa heterozygous genotype frequency.

Departure of the observed genotype frequencies from those expected given the allele frequencies under HWE is calculated using the Pearson's χ^2 test with one degree of freedom. The χ^2 tests statistic is calculated as follows, where O_i are observed genotype counts, and E_i expected genotype counts:

$$\chi^2 = \sum_i \frac{(O_i - E_i)}{E_i}$$

Departure of observed genotypes from those expected under HWE can be caused by many factors including migration, non-random mating, and selection. Extreme deviation from HWE in the context of GWAS, however, is likely to be indicative of genotyping error, and SNPs exhibiting extreme HWE deviation among control samples are excluded from imputation and analysis. It is important to note however that both missingness and departure from HWE is poorly powered to detect genotyping error, and genotyping cluster plots of putatively trait-associated SNPs require careful inspection post-analysis.

Heterozygosity

Mean heterozygosity of an individual describes the proportion of polymorphic loci at which an individual is heterozygous. Heterozygosity is thus calculated as:

$$H_O = \frac{(N - O)}{N}$$

Where H_O represents observed heterozygosity, N represents the number of called genotypes and O the number of homozygous loci. Extreme values of heterozygosity can indicate DNA sample contamination (high values) or in-breeding (low values). Observed levels of heterozygosity will vary between populations, and absolute thresholds of heterozygosity used to exclude samples are set according to the sample population.

Relatedness

A premise of genetic association studies is that they are conducted using samples from unrelated individuals. Inclusion of related individuals in an association analysis will confound allele frequency estimates as a consequence of marker linkage between related individuals, with the potential to drive false positive associations. The identification of cryptic relatedness in study samples is performed by calculation of pairwise identity by state. A given genetic region shared between two individuals is identical by state (IBS). A subset of IBS loci will be identical by descent (IBD); that is alleles are shared by virtue of inheritance from a common ancestor (Weir, Anderson, & Hepler, 2006). In sample collections of apparently unrelated individuals, genome wide estimates of the proportion of alleles shared IBD ($\hat{\pi}$) between a pair of individuals is calculated as:

$$\hat{\pi} = \pi_2 + \frac{\pi_1}{2}$$

Where π_2 represents the probability of a pair of individuals sharing two alleles IBD, and π_1 represents the probability of sharing one allele IBD. This allows estimation of pairwise relationships, e.g. identical twins $\hat{\pi} = 1$, siblings $\hat{\pi} = 0.5$, and first cousins $\hat{\pi} = 0.125$.

To prevent confounding, in the analyses presented here, we have excluded one of related pairs of individuals (by preference case samples, and samples with the higher call rates were retained), with $\hat{\pi}$ > 0.4. An alternative approach, not employed here, would be to retain all individuals regardless of relatedness in the analysis, and account for that allele sharing by IBD by including a model of relatedness as a covariate in the analysis, e.g. inclusion of a genetic relatedness matrix using a linear mixed model.

Population structure

Genetic association analysis in structured populations has the potential for confounding by underlying population structure (Jallow et al., 2009). Population stratification, whereby differences in ancestry between cases and controls result in differential observed genotype frequencies, leads to apparently significant trait-associations not related to genetic determinants of disease. Population stratification, as a source of confounding genetic variation, has the potential to be particularly problematic for GWAS performed in African populations (Teo et al., 2010), with African populations having

increased levels of genetic diversity as compared to non-African populations, and being highly structured (Gurdasani et al., 2015).

Population structure in study samples can be effectively modelled using principal components analysis (PCA) of genome-wide genotyping data(Price et al., 2006). PCA is a multivariate statistical technique, which constructs independent linear variables (principal components - PCs) based on, in this case, study genotypes. These PCs are ordered according to their ability to account for the most variance in the dataset. By virtue of accounting for the most observed variance among genotypes, the top PCs typically capture genetic ancestry with the study samples. Inclusion of these top PCs as covariates in a genetic association analysis will therefore account for confounding secondary to population structure. It is important to note, however, it is not necessarily the case that the top PCs will capture genetic ancestry, and may capture other important sources of confounding variation if present, e.g. genotyping batch effects. For this reason it is important to confirm that the major PCs included in the association analysis are tagging population structure, by comparison with, for instance, self-reported ethnicity.

Principal components of genome-wide genotyping data are calculated using data from founders (identified as individuals without IBD>0.05 to other study participants) alone to minimise the potential for bias introduced by including related samples, and a pruned set of independent markers (markers in high LD are removed).

Imputation

Imputation statistically infers genotypes at SNPs not directly assayed on study SNP arrays. It facilitates fine-mapping of trait-associated variation, and by increasing the density of observable markers and thus minimising LD between the peak associated marker and the unobserved genetic determinant, increases study power. Imputed genotypes presented in this thesis have been generated using the software packages SHAPEIT (segmented haplotype estimation and imputation tool (Delaneau, Marchini, & Zagury, 2012)) and IMPUTE2 (Marchini, Howie, Myers, McVean, & Donnelly, 2007). This section thus focuses on methods employed for imputation by these packages, although the principles are analogous to other implementations of genotype imputation.

We used the pre-phasing approach to genome-wide genotype imputation (Howie, Fuchsberger, Stephens, Marchini, & Abecasis, 2012), which proceeds in two stages. Firstly, phasing is undertaken, in which study sample haplotypes are estimated from genotype data, using reference samples with a known set of genotypes and haplotypes. Taking a common set of markers from un-phased study samples and reference samples with known haplotypes, IMPUTE2 uses hidden Markov models (HMM) to model the haplotypes underlying marker genotypes in the study samples by re-sampling from reference haplotypes. In IMPUTE2 the complexity of this model scales quadratically with the number of reference haplotypes, imposing a

considerable computational burden as reference panel sizes increase. SHAPEIT improves on the IMPUTE2 phasing framework by summarising reference haplotype information and the possible study sample haplotypes given the genotypes, within graph structures. This simplification allows the complexity of the haplotype estimation model to scale linearly with reference haplotype numbers, facilitating the phasing of study samples using large reference panels.

Following phasing, genotypes not typed in study samples, but observed in reference samples, are imputed into the estimated study haplotypes, in our case using IMPUTE2. Probabilities of alleles at untyped markers in each study haplotype are derived using an HMM with reference panel haplotypes as the template. At each imputed marker, genotype probabilities can then be estimated from the allele probability, assuming HWE. This last assumption illustrates why imputation has the potential to perform poorly at loci under strong selection pressure, e.g. the sickle cell locus.

Quality control for imputed genotypes

The probabilistic nature of genotype calls derived by imputation allows measures of confidence in the genotype call to be incorporated into both the subsequent association test at that SNP, and into quality control parameters to exclude poorly-imputed SNPs. To evaluate imputation quality we used the IMPUTE2 post-imputation information score (info score). This metric represents, for a given SNP, the relative statistical information for the imputed

allele frequency as compared to a SNP directly typed in all study samples. It is bounded at 0 (no confidence in the SNP genotypes) and 1 (maximal confidence in the SNP genotypes), with its value proportional to the number of study samples having imputed genotypes with perfect statistical information at that SNP (Marchini & Howie, 2010).

Association testing

Association tests between genetic markers and a trait of interest have been universally performed by regression throughout this thesis. The principal advantage of regression in this context, as compared to other common single locus association tests, is the ability to easily include covariates in the analysis, facilitating, for instance, the inclusion of genotyping data PCs to account for population stratification.

For a dichotomous outcome, as is the case for a case-control study, we use logistic regression, fitting a generalised linear model, with the *logit* link function, of the form:

$$logit\{E(Y_i \mid X_i)\} = \log\left[\frac{E(Y_i \mid X_i)}{1 - E(Y_i \mid X_i)}\right] = \beta_0 + \beta_1 X_i$$

Where Y_i represents the trait of interest, X_i the genetic marker, β_0 and β_1 unknowns to be estimated by the model. The significance of inclusion of a parameter of interest can then be assessed comparing models with and without that parameter, using the likelihood ratio test under a χ^2 distribution.

The model can be readily extended to include parameters to account for confounding variation, e.g.:

$$logit\{E(Y_i \mid X_i, PC_i)\} = \beta_0 + \beta_1 X_i + \zeta PC_i$$

Where PC_i represents a PC to account for confounding variation secondary to population stratification, and ζ the associated coefficient to be estimated. Altering genotype coding further allows this model to be applied to association testing under different genetic models (Table 2.1). Genotype-phenotype associations in models fitted with additive, dominant, recessive and heterozygous advantage genotype coding are tested using likelihood ratio tests under a χ^2 distribution with one degree of freedom. Genotype-phenotype associations in models fitted with genotypic genotype coding, in which two genotype parameters are fitted, are instead tested with likelihood ratio tests under a χ^2 distribution with two degrees of freedom.

Table 2.1 Genotype coding for logistic regression models

	Model					
					Genotypic	
Genotype	Additive	Dominant	Recessive	Heterozygous advantage	Parameter 1	Parameter 2
AA	2	1	1	0	0	1
Aa	1	1	0	1	1	0
aa	0	0	0	0	0	0

Statistical genetic software packages, e.g. SNPTEST2 (Marchini & Howie, 2010), further extend these models to account for uncertainty in imputed genotypes. In doing so they provide model-specific post-imputation

information scores, analogous to those provided by IMPUTE2. These model-specific scores are weighted according to the imputation information of informative individuals within the fitted model. Under additive models of association, these model-specific scores largely approximate the IMPUTE2 info scores. They can, however, diverge significantly in cases of association tests performed with less common SNPs under a recessive model of association, in which cases a comparatively small number of individuals can be highly informative for the model.

Quality control post-association analysis

Post-association analysis inspection of quantile-quantile (QQ) plots, and calculation of the genomic inflation factor (λ), allow assessment of whether adequate control of confounding variation has been achieved in the analysis. QQ plots depict the ordered distribution of *p-values* observed in the analysis compared to the distribution expected under H_0 from a χ^2 distribution with one degree of freedom (additive model) and a χ^2 distribution with two degrees of freedom (genotypic). λ is calculated as the ratio of the median χ^2 test statistic observed in the study over the median of that expected under H_0. Values of between 0.95 and 1.05 are considered to be consistent with those expected if confounding variation is adequately controlled.

Meta-analysis

Throughout this thesis, evidence for trait-association across discovery and replication sample collections has been combined using fixed-effects meta-

analysis (Pfeiffer, Gail, & Pee, 2009). To derive pooled estimates of effect size and variance, we used the inverse variance weighting to account for differences in study size. The effect sizes (θ) from i studies with variance σ^2 are pooled as follows:

$$\theta_{meta} = \frac{\sum {\theta_i}/{\sigma_i^2}}{\sum {1}/{\sigma_i^2}}$$

With the pooled variance calculated:

$$\sigma^2_{meta} = \frac{1}{\sum {1}/{\sigma_i^2}}$$

This approach assumes that the underlying genetic effect is fixed in the included study populations. While this approach is more powerful than those that allow for heterogeneity in effect sizes between study populations, the underlying assumption may not be valid in certain circumstances. To take invasive NTS disease in African children as an example, it seems highly plausible that heterogeneity in the burden of acquired risk factors between populations of African children in different study settings will affect the true effect size of a genetic determinant in each population.

Multiple association testing

Family-wise error rate

The family-wise error rate (FWER) is defined as the probability of at least one type-1 (false positive) error occurring among a group of hypothesis tests assuming that the null hypothesis (H_0) is true for all tests.

Bonferroni correction

The Bonferroni correction (Dunn, 1961) controls the FWER by reducing the significance threshold (α) required to reject H_0 by a factor of the number of independent tests being performed (N), such that:

$$\alpha_{FWER} = \frac{\alpha}{N}$$

In the case of GWAS, the number of independent tests being performed is not simply the number of SNPs tested for trait-association, as LD between SNPs mean that a substantial proportion of typed and imputed SNPs are not independent of one another. Rather, the commonly accepted threshold of genome-wide significance, $\alpha = 5 \times 10^{-8}$, was derived from empirical simulations of the testing burden performing a GWAS encompassing all common variation.

False-discovery rate

FWER seeks to control the rate of type-1 error by constraining the probability of at least one incorrect rejection of H_0 being made, assuming H_0 is universally true. The false-discovery rate (FDR) instead seeks to control type-1 error by controlling the rate of false rejections of H_0 within the total number of tests declared significant, without assumptions regarding the number of H_0 that are true. As a consequence, FDR provides less stringent control of the type-1 error, with the result that the procedure is more powerful. The FDR is calculated as follows:

$$E\left(\frac{FP}{FP+TP}\right)$$

Where FP is the number of false positives in a group of tests, and TP the number of true positives.

Benjamini-Hochberg corrections

Benjamini-Hochberg corrections(Benjamini & Hochberg, 1995) control the FDR by ordering p-values in descending significance: 1, …, *i*. Then for the defined FDR (α_{FDR}), H_0 is rejected for $p_{(1)}$ to $p_{(k)}$, for the largest value of k, such that:

$$p_{(k)} \leq \frac{k}{N}\alpha_{FDR}$$

Permutation-based procedures

An alternative approach to control study error is to sample test statistics where H_0 is true using re-sampling/permutation procedures. For a given genetic association study, the distribution of test statistics at a given marker under H_0 can be observed by permutation (repeated random reassignment) of phenotypes among study samples, and repeated recalculation of the test statistics for association with the marker. It is theoretically possible to apply this genome-wide, recording for each permutation of phenotypes, the minimum *p-value* observed under H_0 at all markers included in the study. By repeated permutation, a minimum *p-value* distribution is generated, which can be compared with the observed *p-values* obtained in the original un-permuted analysis (North, Curtis, & Sham, 2002; 2003). This comparison is used to derive an empirical *p-value* as follows:

$$p_{adj} = \frac{(M+1)}{(N+1)}$$

Where N represents the number of permutations, and M the number of permutation *p-values* more extreme than the observed *p-value.* Permutation procedures are theoretically highly advantageous, and appear to be more powerful than the FWER/FDR approaches outlines above, they are however extremely computationally demanding to perform genome-wide. For this

reason, permutation-based tests performed in this study are restricted to enrichment analyses, where the burden of multiple testing is more limited.

Power

Statistical power is the probability of rejecting the null hypothesis (H_0) when the alternative hypothesis (H_1) is true, or 1-β where β represents the type-2 error rate (the probability of the test returning a false negative result). Defined in this way, it is apparent that a study lacking adequate statistical power risks failing to identify true genetic correlates of a trait. Equally however, study power influences the false positive report probability; that is the likelihood that a finding represents a false positive despite achieving a given statistical threshold (Wacholder, Chanock, Garcia-Closas, Ghormli, & Rothman, 2004).

Power in a genetic association study is dependent on the allele frequency and effect size of the putative genetic determinant and the strength of LD of an observed genetic marker to the (likely) unobserved genetic determinant (Sham & Purcell, 2014). Power calculations presented are calculated by estimation of the non-centrality parameter (λ):

$$\lambda = \frac{2r^2 N_{case} N_{ctrl} (F_{case} - F_{ctrl})^2}{NF(1-F)}$$

where N represents sample numbers, F represents allele frequencies and r^2 is the LD between marker and the putative genetic determinant. λ is then used to specify the statistical power of the study for a given level of α. It is noteworthy that this estimation of study power does not account for several important features of GWAS that contribute to the study's power. Most notably

in terms of the GWAS presented in this thesis, these calculations do not account for the loss in study power imposed by fitting covariates in models of association, e.g. PCs to account for population structure.

Bayesian methods of statistical inference

In the frequentist approach to hypothesis testing, H_0 is rejected in favour of H_1 if the *p-value* associated with H_0 is below a pre-defined threshold. A key limitation of this approach to statistical inference is that a positive result does not imply a probability that the identified association is a true positive finding. This is a consequence of the way in which a frequentist test is constructed, in which the *p-value* only refers to the likelihood of H_0 rather than H_1. Importantly, as noted above in the discussion of study power, the likelihood that a given *p-value* in a given study is associated with a true positive finding is dependent on the power of that study.

A Bayesian approach to statistical inference allows estimation of the probability that H_1 is true, or in terms of a genetic association study, that a genetic marker is truly associated with the trait of interest. A Bayesian approach to hypothesis testing combines evidence from the study that the collected data is associated with the model under evaluation, with prior beliefs about that model (the prior probability) to give a posterior probability of the tested model (H_1) being true (Stephens & Balding, 2009).

The prior probability of H_1 (π) reflects the prior beliefs about H_1, which will be a function of several factors, including, the prior evidence relating to that model, and the number tests being performed within the study. For instance, in a single test of a hypothesis in which the prior beliefs about H_0 and H_1 make them equally likely, π would be set at 0.5. By contrast in the case of a

GWAS, in which the prior expectation could be that association at each SNP is equally likely and a true association would be observed in 1 in every 1,000,000 SNPs, π could be set at 10^{-6}.

Having set π, a Bayes factor (BF) for the association test is calculated by comparing the probability of observing the data given H_1 with the probability of observing the data given H_0:

$$\mathrm{BF} = \frac{P(\mathrm{data} \mid \mathrm{H}_1)}{P(\mathrm{data} \mid \mathrm{H}_0)}$$

In the case of a case-control genetic association test at a given marker, assuming additive effects of genotype on the trait, H_0 and H_1 will be defined according to the expected effect size (the log of the odds ratio) of carriage of the minor allele (θ). For H_0, $\theta = 0$, and for H_1, θ equals the true effect size. As the true effect size under H_1 is unknown, the probability of the data under H_1 is calculated across a distribution of plausible values of θ, termed the prior distribution. The prior distribution of θ is most commonly set as $N(0,\sigma)$. By defining σ, one can define the plausible size of θ. For example, within SNPTEST, the default prior distribution for an additive genotype phenotype association is $N(0,0.2)$, whereas for a recessive association the prior distribution is $N(0,0.5)$, reflecting the expectation of larger genetic effects under a recessive model of association.

Approximate Bayes factors (ABF) can also be derived from effect size estimates calculated using standard frequentist approaches (Wakefield,

2009). Given a maximum likelihood estimate for θ of $N(0,V)$ and the Wald statistic (Z) from the logistic regression model of a genotype-phenotype association, where the prior distribution of θ is defined as $N(0,W)$:

$$\mathrm{ABF} = \sqrt{\frac{V+W}{V}}\exp\left(-\frac{Z^2}{2}\frac{W}{(V+W)}\right)$$

The calculated BF or ABF can be then used to derive the posterior odds (PO) of H_1 being true given the observe Bayes factor and the prior probability, which can then be used to derive the posterior probability (PP) that H_1 is true:

$$\mathrm{PO} = \mathrm{BF}\frac{\pi}{1-\pi}$$

$$\mathrm{PP} = \frac{\mathrm{PO}}{\mathrm{PO}+1}$$

In the work presented in this thesis, we use ABFs to compare models of association among related diseases at disease-associated SNPs identified using the conventional frequentist approaches described above.

Chapter 3 – Genome-wide association study of nontyphoidal *Salmonella* bacteraemia

Background

The outcome of an infectious exposure, be that clearance, asymptomatic carriage, clinical disease or death, is determined by host and pathogen factors, and the manner in which they interact. The inter-individual variation in the outcome of a given infectious exposure is considerable. In infants given oral BCG vaccination contaminated with virulent *Mycobacterium tuberculosis* in pre-antibiotic era Germany – the Lübeck disaster – 228 of 251 children (91%) developed clinical or radiographic evidence of disease, whereas only 72 infants died (29%), and 62 (25%) survived but with severe disease (Fox, Orlova, & Schurr, 2016). Some of this variability in the Lübeck case seems likely to have been secondary to differences in the infective inoculum, however several cases

of severe and fatal disease occurred in infants given the lowest inoculae, demonstrating that host factors are likely to have determined outcome for at least a portion of the Lübeck children. Host determinants of infectious mortality have been demonstrated to be highly heritable in large-scale epidemiological studies (Sørensen, Nielsen, Andersen, & Teasdale, 1988), and modern population-based genetic studies have more recently demonstrated a considerable heritable component for disease susceptibility in many infectious diseases (Chapman & Hill, 2012).

Despite a clear role for host genetic factors as determinants of invasive NTS disease in individuals with rare primary immunodeficiencies and in the mouse model of invasive *Salmonella* infection, no published study to date has adopted an unbiased, population-based approach in order to identify genetic correlates of invasive NTS disease in unselected populations. GWAS have been highly successful in identifying common genetic variants associated with hundreds of disease traits in diverse populations (Altshuler, Daly, & Lander, 2008); including infectious diseases in African populations (Timmann et al., 2013). The lack of unbiased, population-based data assessing host genetic susceptibility to iNTS disease in African populations, led us to conduct a GWAS and replication analysis of NTS bacteraemia in Kenyan and Malawian children.

In addition to the unbiased assessment of the genetic architecture of NTS bacteraemia provided by this analysis, we further consider evidence for genetic

correlates of invasive NTS disease at loci previously suggested to predispose to invasive *Salmonella* disease, with a high degree of biological plausibility. Carriage of both class I and class II HLA alleles have been implicated in host susceptibility to a variety of infectious diseases (Fellay et al., 2007; International HIV Controllers Study et al., 2010; Zhang et al., 2009), and recently, carriage of a DRB1 allele has been found to protect against the other form of invasive *Salmonella* disease in humans, enteric fever (Dunstan et al., 2014). Using genotyping data generated as part of the GWAS, we estimate classical HLA alleles to investigate evidence of association with common class I and II HLA alleles and NTS bacteraemia, and other common causes of bacteraemia in Kenyan children.

Sickle cell disease is a risk factor for NTS bacteraemia in Kenyan children, and sickle cell trait protects against development of bacteraemia caused by diverse pathogens (Williams et al., 2009). The sickle cell locus therefore provides an informative positive control for the GWAS of NTS bacteraemia, and we assess evidence for association with sickle cell homozygosity and heterozygosity with bacteraemia secondary to NTS and other common pathogens. The immune defect associated with sickle cell disease is complex, but a proportion of the resultant NTS susceptibility may be secondary to defective oxidative burst secondary to haemolysis (Cunnington, de Souza, Walther, & Riley, 2012a). To investigate whether other genetic causes of haemolysis in African children could

predispose to NTS bacteraemia, we further perform a candidate gene analysis of glucose-6-phosphate dehydrogenase (G6PD) deficiency and NTS bacteraemia.

Methods

Nontyphoidal Salmonella genome-wide association study

Study design and participants

We performed a case-control, GWAS of susceptibility to NTS bacteraemia in African children. Kenyan (n=267) and Malawian (n=150) children with culture-confirmed NTS bacteraemia were recruited to the study. Kenyan cases were recruited among children admitted to Kilifi District Hospital, Kilifi. Malawian cases were recruited among children admitted to Queen Elizabeth Central Hospital, Blantyre. Kenyan control participants (n=4,444) were healthy infants resident in Kilifi district, recruited from consecutive births between 1st May 2006 and 30th April 2008. Malawian control participants (n=339) were healthy children resident in Blantyre, recruited in 2005 and 2006.

Kenyan study cohort

Children (under 13 years of age) presenting to Kilifi District Hospital, Kenya between 1st August 1998 and 30th October 2010 with all-cause bacteraemia were recruited as described previously (Kenyan Bacteraemia Study Group et al., 2016). In brief, blood samples for bacterial culture (BACTEC 9050, Becton

Dickinson) were taken from every child admitted to the hospital during the study period, with the exception of elective surgical admissions and children admitted following minor accidents. Control samples were collected as part of a birth cohort study from consecutive births between 1st May 2006 and 30th April 2008, among the same population as the case samples in Kilifi district. All control children were recruited under the age of 12 months. Longitudinal follow-up of the control children suggests that the risk of case-control misclassification is negligible (cases of bacteraemia among controls during follow-up = 12; mortality among controls during follow-up = 49). These case and control samples were divided into discovery (cases = 1,885, controls = 3,000) and replication (cases = 532, controls = 1,444) sets for the Wellcome Trust Case Control Consortium 2 (WTCCC2) all-cause bacteraemia genome-wide association study (GWAS) (Kenyan Bacteraemia Study Group et al., 2016).

From this collection, we extracted children with nontyphoidal *Salmonella* (NTS) bloodstream infection (n=267) from the bacteraemia cases, and all of the control children (n=4,444) 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=3,000) and replication sets (n=49, controls=1,444) separately. Demographics of study cases and controls in both the discovery and replication sample sets are presented in Table

3.1. The frequencies of NTS serotypes and co-morbidities in discovery and replication case samples are presented in Table 3.2.

Table 3.1 Demographic data for cases and controls in Kenyan discovery, Kenyan replication and Malawian replication samples sets

			Kenya discovery		Kenya replication		Malawi replication	
			Cases	**Controls**	**Cases**	**Controls**	**Cases**	**Controls**
		Numbers*	180	2,677	38	1,336	150	339
		Median age in months (range)	15 (0-143)	BC	14 (0-52)	BC	16 (0-180)	18 (0-182)
		Females	47.2%	49.6%	50.0%	49.5%	45.3%	49.6%
Reported ethnicity	Kenya	Giriama	50.0%	45.8%	33.3%	40.0%		
		Chonyi	30.4%	37.1%	41.6%	44.1%		
		Kauma	7.6%	11.9%	8.3%	10.3%		
	Malawi	Chewa					20.7%	11.8%
		Yao					14.0%	19.2%
		Lomwe					28.7%	28.6%
		Ngoni					16.7%	24.8%

*Numbers included in final analysis. BC, birth cohort.

Table 3.2 Distribution of NTS serovars and common NTS-associated co-morbidities in the Kenyan discovery, Kenyan replication and Malawian replication NTS cases

		Kenya discovery (n=180)	Kenya replication (n=38)	Malawi replication (n=150)
	Age (months)	15 (0-143)	14 (0-52)	16 (0-180)
	Female	85/180 (47.2%)	19/38 (50.0%)	68/150 (45.3%)
NTS serovar	Typhimurium	66/166 (39.8%)	12/33 (36.4%)	135/150 (90.0%)
	Enteritidis	84/166 (50.6%)	20/33 (60.6%)	12/150 (8.0%)
	Not typable*	16/166 (9.6%)	1/33 (3.0%)	3/150 (2.0%)
Co-morbidities	HIV-infection	24/117 (20.5%)	6/30 (20.0%)	49/126 (38.9%)
	Malaria	55/179 (30.7%)	15/38 (39.5%)	17/150 (11.3%)
	Severe wasting**	62/160 (38.8%)	10/32 (31.3%)	48/133 (36.1%)
	Mortality	38/178 (21.3%)	10/36 (27.8%)	17/150 (11.3%)

*with available antisera.

**Severe wasting defined as weight-for-age z-scores <-3.

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 (OxTREC).

Malawian study cohort

Malawian replication 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. Identification of NTS was performed using API 20E kits (bioMérieux) and NTS serotyping performed as for the Kenyan isolates. Every child under 16 years of age, admitted with NTS in blood culture was approached for recruitment to the study.

Malawian replication control samples (n = 339) were collected from healthy children at the Ndirande Health Centre, Blantyre, Malawi between October and December 2005, and September and November 2006 (Mandala et al., 2010). 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.

DNA sample preparation

Blood samples from the Kenyan children were collected into ethylene-diamine tetra-acetic acid (EDTA)-containing tubes, and stored at -80C prior to extraction. Buccal samples were collected and stored at -20C prior to extraction. Genomic DNA was extracted using QIAamp DNA blood mini kits (Qiagen). Extracted genomic DNA was quantified using a PicoGreen (Invitrogen) assay. The PicoGreen assay is a fluorescence-based quantification assay, in which the PicoGreen reagent fluoresces when bound to double-stranded DNA (emitting negligible fluorescence in the presence of other nucleic acids). Following extraction and quantification, all DNA samples were whole-genome amplified using Genomiphi (GE Healthcare). Genomiphi uses a pool of random DNA hexamers to initiate PCR DNA amplification at multiple, random intervals giving even coverage across the genome, and uniform whole-genome amplification. Whole-genome amplification was undertaken according to the manufacturer's instructions, except in cases where the input DNA concentration was <10ng/µl, in which case the input DNA volume was increased threefold.

Kenyan discovery sample genotyping

Genome-wide genotyping of Kenyan discovery cases and controls was performed with an Affymetrix SNP 6.0 chip at the Affymetrix Service Laboratory, and genotypes called with a modified version of Chiamo software (Burton et al., 2007). Case and control samples were randomized across batches for genotyping. SNP quality control (QC) filters were applied as follows: minor allele frequency (MAF)<1%, info<0.975, Hardy-Weinberg equilibrium $P<1x10^{-20}$, plate effect $P<1x10^{-6}$ and SNP missingness>2%. Following SNP and sample QC (Table 3.3) 787,861 autosomal SNPs from 1,536 bacteraemia (including 180 NTS) case and 2,677 control individuals were taken forward for genome-wide imputation.

Table 3.3 Sample exclusions in the Affymetrix SNP 6.0 chip-genotyped Kenyan discovery and ImmunoChip-genotyped Kenyan replication sets

		Kenyan discovery samples		Kenyan replication samples	
		Cases	**Controls**	**Cases**	**Controls**
	Total before QC	218	3,000	49	1,444
Exclusion criteria	Call rate & extreme heterozygosity	7	99	6	40
	HapMap PCA outliers	15	68	2	18
	Outlying channel intensity	0	12	1	1
	Sequenom discordance	0	13	0	8
	Discrepant or undetermined gender	4	14	1	2
	Duplicate samples	4	16	2	3
	Relatedness (IBD>0.4)	1	83	0	28
	Duplication of discovery samples in replication	NA	NA	1	12
	Genotyping failures	8	37	0	0
	Total after QC	180	2,677	38	1,336

QC, quality control; IBD, Identity by descent.

Kenyan discovery sample phasing & imputation

To estimate genotypes at SNPs not assayed by the Affymetrix SNP 6.0 chip, we performed whole-genome imputation. Genotypes passing QC were pre-phased, estimating sample haplotypes, using SHAPEIT (Delaneau et al., 2012). Following pre-phasing, estimated sample haplotypes were used to infer unobserved genotypes with IMPUTE2 (Howie et al., 2012) using the 1000 Genomes Project Phase 1 release as a reference panel. Imputation was performed in 5Mb chromosomal chunks, with the software-recommended input parameters (buffer regions of 250kb, effective population size 20,000, 30 Markov chain Monte Carlo iterations).

Kenyan discovery association analysis

To model population structure within the Kenyan discovery samples, we calculated principal components of genome-wide genotyping data using EIGENSTRAT (Price et al., 2006). Principal components were calculated from a linkage-disequilibrium (LD) pruned set of genotypes in PLINK (Purcell et al., 2007), pruning within 50kb windows in 5 variant steps, with an r^2 threshold of 0.2. The first four principal components of the genome-wide genotyping data differentiate self-reported ethnicity (Figure 3.1) 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 SNPTEST2 (Marchini et al., 2007) under additive and genotypic models of association. Following association analysis, genotyped SNPs nominally associated (P<1x10^{-3}) with NTS were inspected with Evoker and SNPs with poor cluster separation excluded from further analysis.

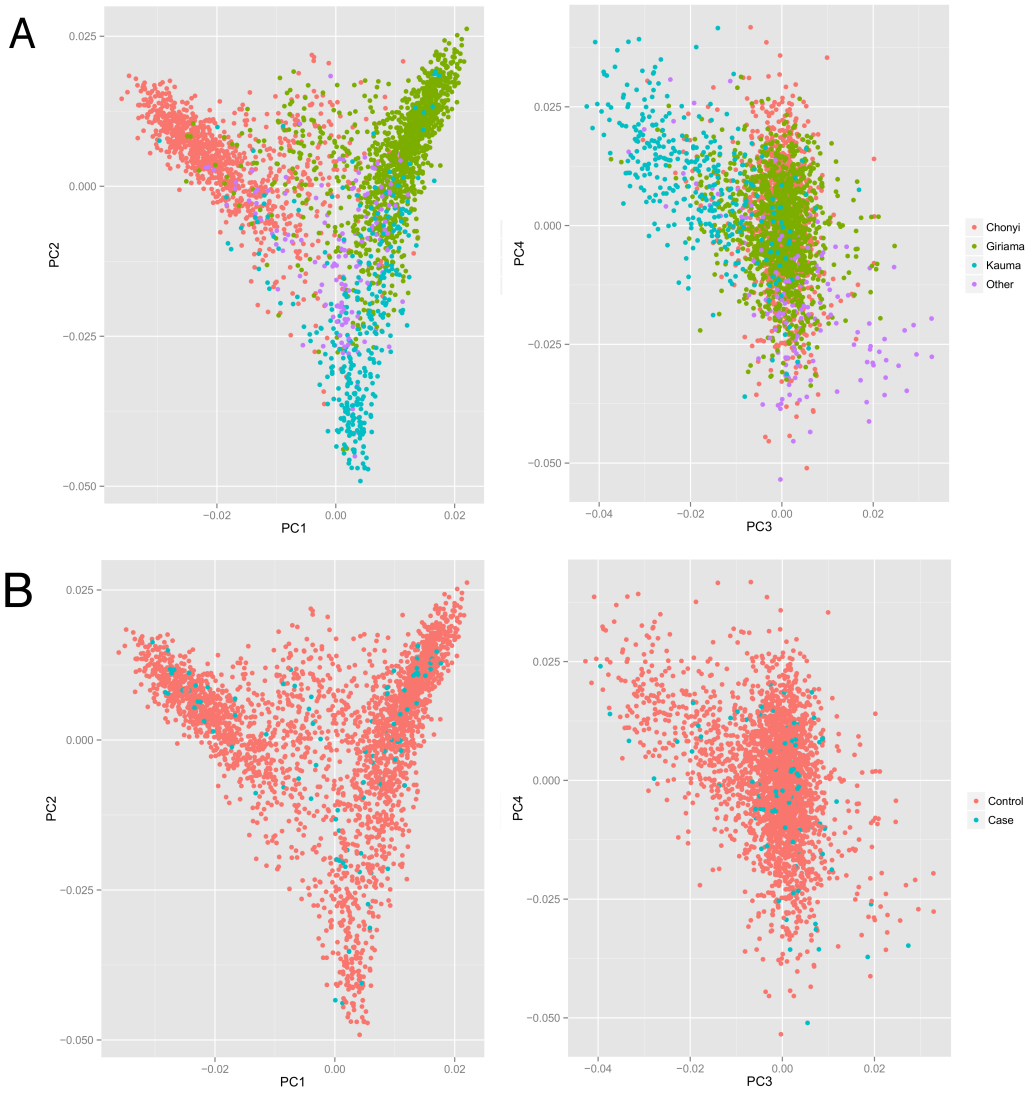

Figure 3.1 The first four principal components of Kenyan discovery genome-wide genotyping data.

Individuals are color-coded according to (A) self-reported ethnicity and (B) case-control status.

Further SNP QC exclusions were applied as follows: model-specific imputation info score<0.8, Hardy-Weinberg equilibrium $P<1x10^{-10}$. Given the study sample size and limited power to replicate associations at less common SNPs (Figure 3.2), we further excluded SNPs with MAF<10% from the analysis.

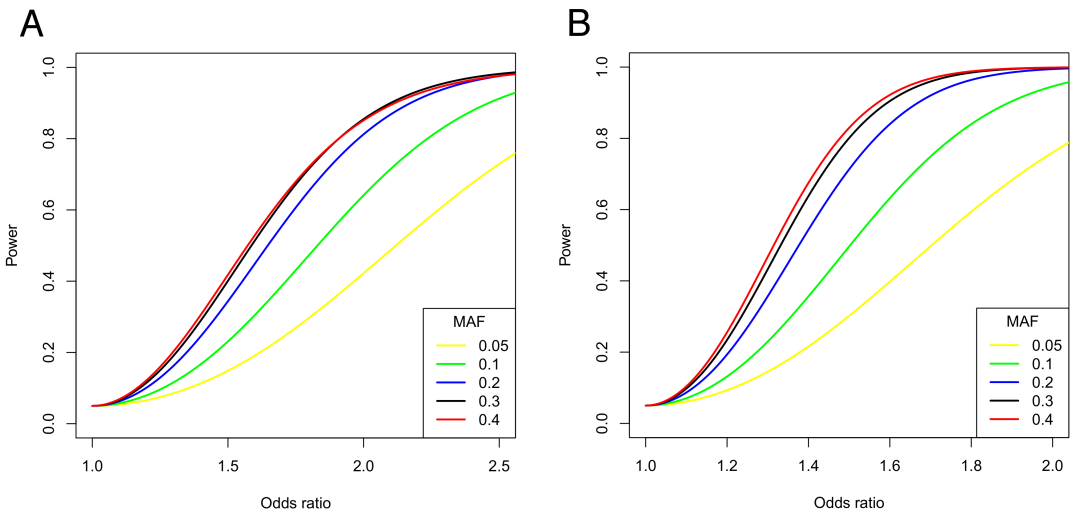

Figure 3.2 Statistical power in replication sample collections.

Statistical power to detect an association (p<0.05) of varying effect size in (A) Kenyan replication samples (38 cases, 1,336 controls) and (B) Malawian replication samples (150 cases, 339 controls) given different minor allele frequencies (MAF).

Re-imputation of genomic regions of interest

1Mb regions centred in regions of interest were intensively re-imputed in the Kenyan discovery samples without pre-phasing with IMPUTE2 using 1000G Phase3 as a reference panel. Imputation was performed with 1Mb buffer regions

and template haplotypes for phasing increased to 200. This region was also imputed in the Kenyan replication samples from ImmunoChip genotyping data (see below) using the same IMPUTE2 settings.

Replication genotyping & analysis

Kenyan replication cases and controls were genotyped using the ImmunoChip (Cortes & Brown, 2011). Sample QC filters (Table 3.1) were applied as for the discovery samples, resulting in 38 NTS case and 1,336 control samples being included in the analysis. As for the discovery phase samples, an LD-pruned set of SNPs (generated in PLINK) with MAF>1%, call rate>95% (99% if MAF<5%), and Hardy-Weinberg equilibrium $P<1x10^{-10}$ were used to generate principal components of replication sample genotypes in EIGENSOFT (Figure 3.3).

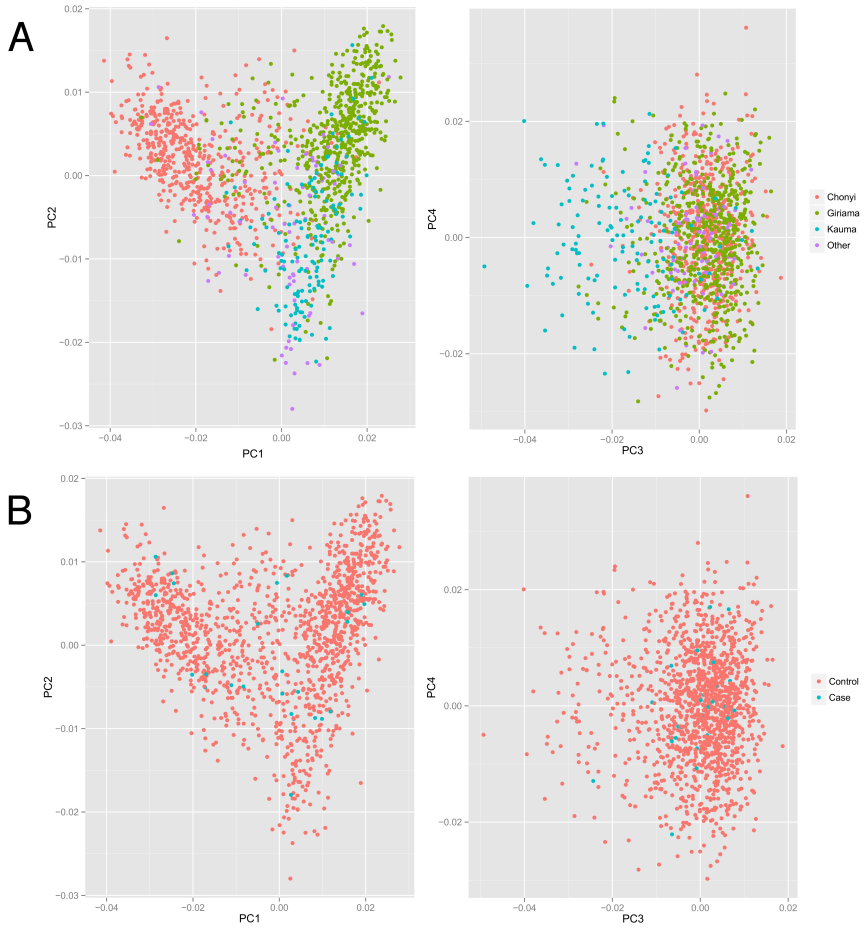

Figure 3.3 Principal components of Kenyan replication genome-wide genotyping data.

Individuals are color-coded according to (A) self-reported ethnicity and (B) case-control status.

SNPs in genomic regions putatively associated with NTS bacteraemia (P<1x10^{-6}) in the Kenyan discovery samples were genotyped in the Kenyan and Malawian replication samples using the Sequenom MASSArray platform. Samples were excluded with call rates <80%. 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.

The most significant *STAT4* region SNP (rs13390936) in both the Kenyan discovery and replication samples following imputation, failed design for inclusion in a multiplexed Sequenom MASSArray genotyping reaction. This SNP was therefore genotyped by High-Resolution Melt-curve Analysis (HRMA), using the following PCR primers and an unlabelled probe with a 3'-amino-C7 modification: TAGTGAGCCCTAATGTAAATTATGGGAC (F), CCCTCACCAGTTTCTCCTATATCT (R), GTGATGTACTTGTTACAAATTTATATTATTACAATA (probe). 10µl PCR reactions (1µl 10x PCR buffer, 1µl 25mM MgCl2, 1µl dNTPs (2mM each), 0.05µl forward primer (10µM), 0.025µl reverse primer (100µM), 0.025µl unlabelled probe (100µM), 1µl LC Green dye, 0.06µl HotStarTaq DNA polymerase (Qiagen), 5µl water, and 0.5µl template DNA(25ng/µl)) were cycled (95°C for 10 minutes, 55 cycles of 30 seconds at 95°C, 56°C, and 72°C), under mineral oil. Following the PCR reaction, the DNA was again denatured (95°C for 1 minute) and cooled to room temperature for 30 minutes. 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 3.0 (Idaho Technology) using derivative normalized melting plots between 62°C and 71°C.

Association analysis of Malawian and Kenyan replication samples was performed in PLINK (Purcell et al., 2007). For the Kenyan replication samples, the first four principal components of ImmunoChip genotyping data were included in the model to account for population structure. We considered a 2-tailed p-

value<0.05, with the same risk allele in the same model of association as observed in the Kenyan discovery set, as evidence of replication. For the combined analysis, a fixed-effects meta-analysis of the three sample collections was performed in PLINK.

Candidate gene analysis

Imputation and association analysis of classical HLA alleles

We used HLA*IMP:02 (Dilthey et al., 2013) to estimate classical HLA alleles to 4-digit resolution at HLA-A, HLA-B, HLA-C, HLA-DQA1, HLA-DQB1, and HLA-DRB1. For HLA imputation we included Kenyan discovery and replication individuals passing QC thresholds for inclusion for genotype imputation (see above). We generated classical HLA allele estimations in 4,203 discovery (Affymetrix SNP Chip 6.0) individuals (1,526 cases and 2,677 controls), and 1,770 replication (Immunochip) individuals (434 cases, 1,336 controls). Directly genotyped SNPs passing QC thresholds for genotype imputation (see above) in the HLA region (chr6: 28,000,000-35,000,000) were used to generate HLA allele estimates, with 2,245 Affymetrix-genotyped SNPs and 8,296 Immunochip-genotyped SNPs being included in the analysis. Generation of the HLA*IMP HLA alleles was performed by Dr Alex Dilthey, Wellcome Trust Centre for Human Genetics, University of Oxford.

Individuals were included in the analysis at each HLA locus if both HLA alleles were imputed with reasonable confidence, for which we set a threshold absolute posterior probability of the imputed allele at 0.7. Carriage of common 4-digit HLA alleles (minor allele frequency > 0.05) were tested for association with NTS and all-cause bacteraemia in the Kenyan discovery samples, using logistic regression models, including the first two principal components of genome-wide genotyping data to account for population structure. Multinomial regression models of common HLA alleles were also fitted at each locus to test for HLA associations among common bacterial pathogens. In models suggestive of HLA allele association with bacteraemia, the model of disease association best supported by the data was defined using a Bayesian framework (see above), and a replication analysis undertaken using the Kenyan replication samples. Analysis was performed in R.

Sickle cell locus

Imputation quality control metrics suggest that imputation was of acceptable quality (info > 0.8) at the sickle cell locus (rs334). We converted genotype probabilities to absolute genotype calls using GTOOL (http://www.well.ox.ac.uk/~cfreeman/software/gwas/gtool.html), with an absolute posterior probability threshold of 0.8 required to call a genotype at rs334. We tested for association of genotype at rs334 with bacteraemia secondary to NTS and to all pathogens under additive, recessive and heterozyogote advantage

models. The recessive and heterozygote advantage models representing disease risk associated with sickle cell disease and sickle cell trait respectively. We further construct multinomial regression models of rs334 association under each model of inheritance (additive, recessive and heterozygote advantage) with each common cause of bacteraemia. We compare models of association in a Bayesian framework (see below), defining the models of pathogen-specific association at rs334 best supported by the data.

Glucose-6-phosphate dehydrogenase deficiency

A single genetic locus (rs1050828) determines G6PD deficiency in the coastal Kenyan population in which the GWAS was performed. rs1050828 is directly genotyped in the Kenyan discovery collection. To investigate any association between G6PD deficiency and disease we tested for association at rs1050828 in an additive linear model, with rs1050828 genotypes coded to reflect a monotonic change in G6PD biochemical activity: girls homozygous, and boys hemizygous, for the wildtype rs1050828:C allele, with normal G6PD activity; heterozygous girls with intermediate G6PD activity; girls homozygous, and boys hemizygous, for the alternate rs1050828:T allele, with G6PD deficiency. We further tested for an altered risk of bacteraemia in heterozygous girls and G6PD deficient children separately, fitting regression models of each using wildtype rs1050828:C allele homozygous girls and hemizygous boys as the baseline. Each of these models was fitted to assess risk of NTS and all-cause bacteraemia (logistic regression),

and risk of bacteraemia secondary to the common bacterial pathogens observed in the Kenyan cohort (multinomial logistic regression). The first two principal components of genome-wide genotyping data were included in each model to account for population sub-structure. We compare models of association in a Bayesian framework (see below), defining the models of pathogen-specific association at rs1050828 best supported by the data.

Bayesian comparison of models of association in major bacterial pathogens in Kenyan children

We compared models of association at candidate loci across the seven most frequently isolated bacterial pathogens (NTS, *S. pneumoniae, S. aureus*, *H. influenza* type b, β-haemolytic *Streptococci*, *E. coli*, *Acinetobacter*) among cases of bacteraemia in the Kenyan discovery samples in the all-cause bacteraemia GWAS(Kenyan Bacteraemia Study Group et al., 2016). 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(0,1) and fixed between pathogens (ρ=1).

REL: effect size ~ N(0,1) 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 instance, if the data best supports a model in which a genetic locus is associated with NTS bacteraemia alone, the fourth model would be specified as follows:

NTS: effect size ~ N(0,1) for NTS and is zero for all other pathogens.

For each model we calculated approximate Bayes factors (Wakefield, 2009) and posterior probabilities, assuming each model to be equally likely *a priori*. Statistical analysis was performed in R, with code supplied by Dr Matti Pirinen, University of Helsinki.

Results

Genome-wide association study of nontyphoidal Salmonella bacteraemia

Following quality control, genotypes at 5,585,198 SNPs (additive model) and 4,669,480 SNPs (genotypic model) from 2,847 individuals (180 cases, 2,677 controls) were included in the discovery analysis. Inspection of quantile-quantile plots (Figure 3.4) and the genomic control inflation factors ($\lambda_{additive}$=1.01, $\lambda_{genotypic}$=1.01) indicate that inclusion of the four major principal components as covariates in the analysis adequately controls for population substructure. In this analysis, 67 SNPs at 16 loci were suggestively associated ($P<1x10^{-6}$) with NTS bacteraemia (Figure 3.5, Table 3.4).

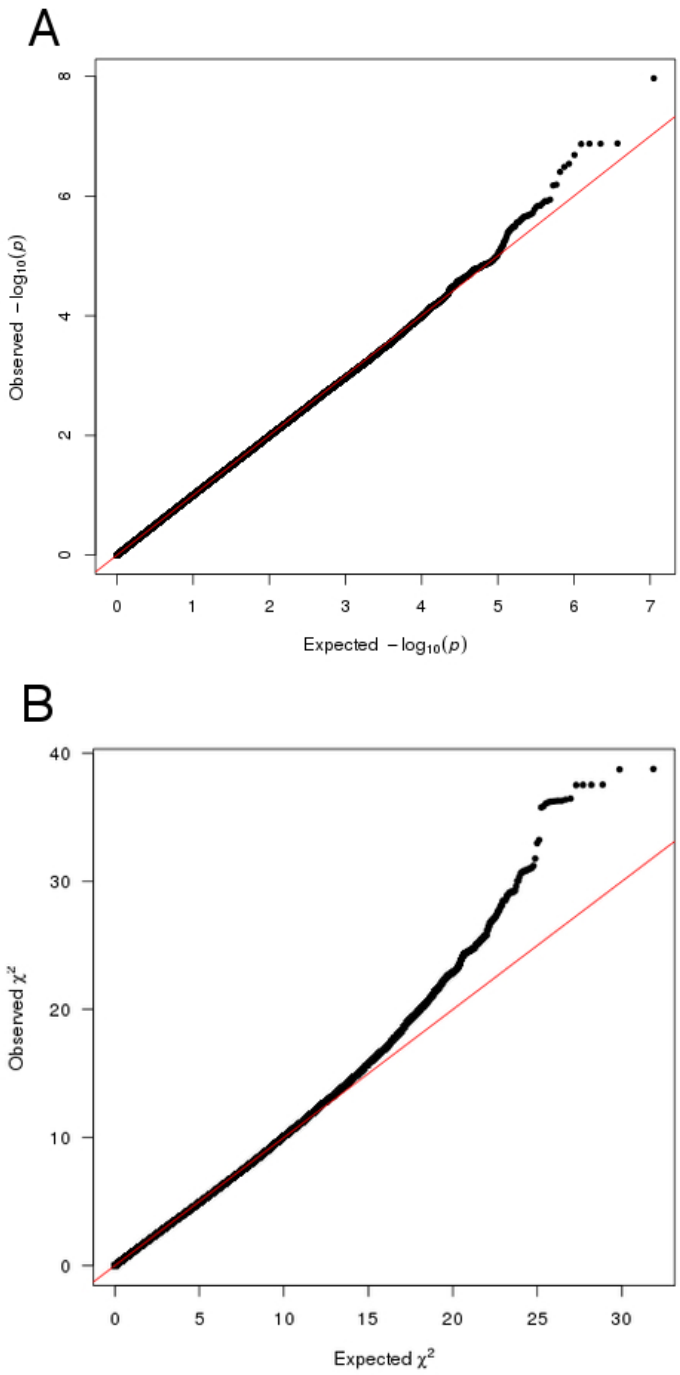

Figure 3.4 Quantile-quantile plots of nontyphoidal *Salmonella* genome-wide association study.

Quantile-quantile plots of NTS association in Kenyan discovery samples under (A) additive (5,585,198 SNPs, $\lambda = 1.01$) and (B) genotypic (4,669,480 SNPs, $\lambda = 1.01$) models. Observed $-\log_{10}$(p-values)/test statistics are plotted against those expected under a χ^2 distribution with 1 degree of freedom (additive model) and two degrees of freedom (genotypic model).

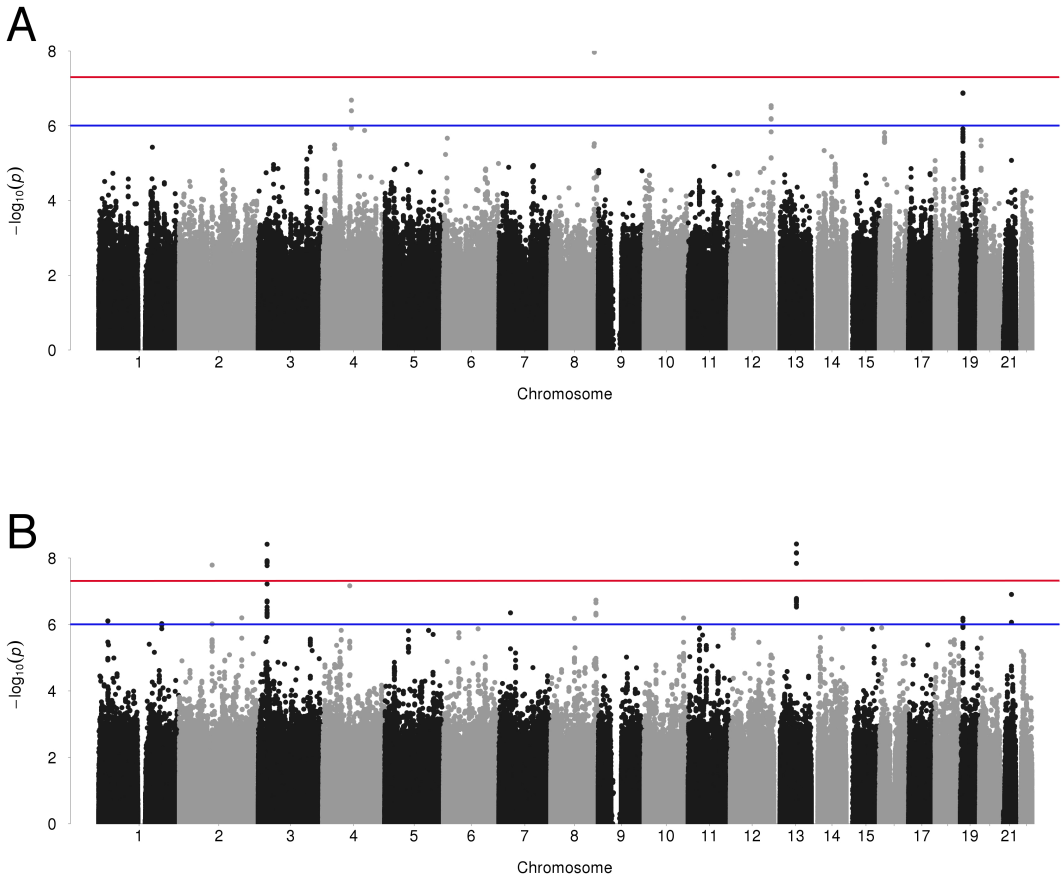

Figure 3.5 Manhattan plots of nontyphoidal *Salmonella* genome-wide association study

Manhattan plots of NTS association in Kenyan discovery samples under (A) additive (5,585,198 SNPs) and (B) genotypic (4,669,480 SNPs) models. The upper (red) horizontal line denotes genome-wide significance (P=5x10^{-8}), the lower (blue) horizontal denotes suggestive association (P=1x10^{-6}) used as the threshold for replication.

Table 3.4 Nontyphoidal *Salmonella*-associated loci (p<1x10^{-6}) in discovery GWAS

SNP				NTS association			
CHR	BP (b37)	RSID	Freq.	Imputation Info Score	*P*	Model	r^2 to peak SNP in locus
1	29053720	rs4654369	0.264	0.969	7.88E-07	Genotypic	1.000
1	29054752	rs10799120	0.264	0.971	7.85E-07	Genotypic	Peak
1	194946012	rs866059	0.115	0.976	9.46E-07	Genotypic	1
2	100773069	rs184422605	0.190	0.961	9.64E-07	Genotypic	0.770
2	100789798	rs75776491	0.232	0.988	9.56E-07	Genotypic	0.978
2	100795043	rs62150289	0.238	0.923	1.63E-08	Genotypic	Peak
2	191954816	rs13390936	0.124	0.878	6.33E-07	Genotypic	Peak

3	26761622	rs1488241	0.147	0.995	3.68E-07	Genotypic	0.780
3	26768800	rs2171512	0.135	1.000	4.75E-07	Genotypic	0.730
3	26768818	rs2129855	0.135	1.000	5.83E-07	Genotypic	0.727
3	26769354	rs11713454	0.135	0.999	4.63E-07	Genotypic	0.730
3	26769392	rs7430878	0.135	0.999	4.63E-07	Genotypic	0.730
3	26769429	rs6790025	0.135	0.999	4.63E-07	Genotypic	0.730
3	26771574	rs139655208	0.135	0.995	4.35E-07	Genotypic	0.730
3	26792607	rs11914329	0.145	0.951	1.31E-08	Genotypic	0.803
3	26792972	rs961399	0.145	0.951	1.31E-08	Genotypic	0.803
3	26793518	rs1488248	0.145	0.953	1.39E-08	Genotypic	0.803
3	26794821	rs1488250	0.136	1.000	5.28E-07	Genotypic	0.736
3	26795026	rs1505577	0.145	1.000	6.04E-08	Genotypic	0.817
3	26795765	rs1488251	0.135	0.999	3.00E-07	Genotypic	0.751
3	26796445	rs6768401	0.134	1.000	2.00E-07	Genotypic	0.752
3	26796598	rs1586506	0.134	0.999	2.00E-07	Genotypic	0.753
3	26796765	rs1586507	0.134	0.999	1.96E-07	Genotypic	0.752
3	26798069	rs1488254	0.143	0.985	1.70E-08	Genotypic	0.819
3	26798614	rs1032885	0.142	0.979	1.25E-08	Genotypic	0.814
3	26798992	rs4973809	0.142	0.978	1.20E-08	Genotypic	0.814
3	26799072	rs6778280	0.143	0.978	1.35E-08	Genotypic	0.820
3	26799395	rs1038636	0.143	0.977	1.33E-08	Genotypic	0.820
3	26800829	rs6551141	0.157	0.848	2.12E-07	Genotypic	0.801
3	26804954	rs10865806	0.158	0.932	3.85E-09	Genotypic	Peak
4	83324569	rs11099533	0.111	0.858	6.86E-08	Genotypic	Peak
4	88413965	rs17012693	0.214	0.920	2.06E-07	Additive	Peak
4	88420287	rs2169502	0.222	0.923	3.95E-07	Additive	0.939
7	35394619	rs138302119	0.115	0.848	4.47E-07	Genotypic	Peak
8	73087935	rs1809447	0.113	0.825	6.61E-07	Genotypic	1.000
8	73090061	rs13251057	0.113	0.822	6.49E-07	Genotypic	Peak
8	134298504	rs16904882	0.112	0.954	1.08E-08	Additive	Peak
8	139215025	rs11995943	0.184	0.836	5.20E-07	Genotypic	0.980
8	139215155	rs11985359	0.184	0.831	5.05E-07	Genotypic	0.980
8	139215247	rs13265099	0.184	0.826	4.60E-07	Genotypic	0.983
8	139215574	rs4404965	0.189	0.806	1.85E-07	Genotypic	Peak
8	139215649	rs7387858	0.188	0.806	2.23E-07	Genotypic	0.993
10	121497290	rs73355807	0.154	0.915	6.40E-07	Genotypic	Peak
12	125613150	rs78857905	0.123	0.832	2.89E-07	Additive	Peak
12	125614614	rs76336213	0.118	0.806	6.65E-07	Additive	0.998
12	125615610	rs75239828	0.118	0.805	6.44E-07	Additive	0.998
12	125615627	rs75627386	0.122	0.804	3.26E-07	Additive	0.964
13	69573095	rs73211763	0.112	0.830	2.99E-07	Genotypic	0.901
13	69576658	rs73211791	0.112	0.836	2.56E-07	Genotypic	0.901
13	69578250	rs73213609	0.112	0.832	2.10E-07	Genotypic	0.901
13	69579168	rs55763812	0.126	0.999	1.45E-08	Genotypic	0.998
13	69579594	rs73213613	0.124	1.000	7.10E-09	Genotypic	1.000
13	69580147	rs1340723	0.112	0.827	1.80E-07	Genotypic	0.901
13	69580255	rs9541701	0.124	1.000	7.05E-09	Genotypic	1.000
13	69580286	rs9541702	0.124	1.000	7.04E-09	Genotypic	1.000
13	69581350	rs7321066	0.112	0.824	1.65E-07	Genotypic	0.901
13	69581690	rs9541704	0.124	1.000	6.99E-09	Genotypic	1.000
13	69582697	rs12429197	0.111	0.811	1.87E-07	Genotypic	0.899
13	69582740	rs9541709	0.124	0.955	3.78E-09	Genotypic	Peak
19	7914568	rs550134	0.349	0.908	1.32E-07	Additive	Peak
19	7915451	rs520802	0.440	0.847	6.52E-07	Genotypic	0.720
19	7916475	rs2452012	0.349	0.908	1.34E-07	Additive	1.000
19	7916620	rs580790	0.349	0.908	1.33E-07	Additive	1.000
19	7922672	rs2059820	0.348	0.907	1.35E-07	Additive	1.000
19	7932459	rs523648	0.432	0.834	7.70E-07	Genotypic	0.740
21	35289246	rs2834301	0.158	1.000	8.59E-07	Genotypic	0.422
21	35303773	rs8129054	0.275	0.805	1.25E-07	Genotypic	Peak

Genetic association between NTS bacteraemia and suggestively associated SNPs ($P<1x10^{-6}$, n = 67) at 16 loci in Kenyan discovery samples. CHR, chromosome; BP, base-pair position (Human Genome Reference GRCh37); RSID, SNP identifier; Freq, minor allele frequency.

Table 3. 5 Replication results and imputation accuracy of top SNPs (P<1x10^{-6}) in NTS Kenyan discovery analysis

				Kenyan discovery samples					Kenyan replication samples				Malawian replication samples				Combined analysis		
CHR	BP (b37)	SNP	Gene	MAF cases imputed	MAF controls imputed	Imputed accuracy	p-value imputed	OR (95% CI) imputed	MAF cases	MAF controls	p-value	OR (95% CI)	MAF cases	MAF controls	p-value	OR (95% CI)	p-value	OR (95% CI)	Model
1	29053720	rs4654369		0.32	0.26	0.97	7.88x10^{-7}		0.22	0.27	0.436		0.29	0.28	0.981				Genotypic
2	100795043	rs62150289	*AFF3***	0.32	0.23	0.95	1.63x10^{-8}		0.18	0.21	0.588		0.18	0.22	0.471				Genotypic
2	191940260	rs16833239	*STAT4***	0.17	0.13	0.95	1.80x10^{-5}		0.13	0.11	2.51x10^{-3}		0.15	0.13	0.080				Genotypic
2	**191954816**	**rs13390936**	***STAT4***	**0.17**	**0.12**	**0.97**	**6.33x10^{-7}**		**0.15**	**0.11**	**1.83x10^{-3}**		**0.15**	**0.14**	**0.039**				**Genotypic**
2	191968420	rs13407419	*STAT4***	0.17	0.13	0.97	1.18x10^{-5}		0.16	0.12	0.014		NA	NA	NA				Genotypic
2	191970330	rs13401064	*STAT4***	0.17	0.13	0.97	1.16x10^{-5}		0.13	0.11	4.03x10^{-3}		0.15	0.13	0.072				Genotypic
3	26768818	rs2129855	*LRRC3B**	0.17	0.13	NA	5.82x10^{-7}		0.14	0.14	0.940		0.13	0.14	0.218				Genotypic
3	26769354	rs11713454	*LRRC3B*	0.17	0.13	1.00	4.63x10^{-7}		0.14	0.14	0.926		0.13	0.13	0.206				Genotypic
3	26792607	rs11914329	*LRRC3B*	0.19	0.14	0.99	1.31x10^{-8}		0.15	0.15	0.779		0.13	0.14	0.451				Genotypic
3	26794821	rs1488250	*LRRC3B*	0.17	0.13	1.00	5.28x10^{-7}		0.14	0.14	0.940		0.14	0.14	0.461				Genotypic
3	26796445	rs6768401	*LRRC3B**	0.17	0.13	NA	2.00x10^{-7}		0.14	0.14	0.922		0.13	0.14	0.490				Genotypic
3	26798614	rs1032885	*LRRC3B*	0.18	0.14	1.00	1.25x10^{-8}		0.15	0.15	0.848		0.13	0.14	0.210				Genotypic
4	88413965	rs17012693	*SPARCL1*	0.33	0.21	0.95	2.06x10^{-7}	2.04 (1.56-2.66)	0.21	0.24	0.787	0.88 (0.35-2.23)	0.21	0.21	0.936	0.98 (0.6-1.6)	1.14x10^{-5}	1.66 (1.32-2.08)	Additive
8	73087935	rs1809447		0.16	0.11	0.99	6.61x10^{-7}		0.15	0.10	0.328		0.06	0.07	0.733				Genotypic
8	134298504	rs16904882	*NDRG1*	0.20	0.11	0.95	1.08x10^{-8}	2.78 (1.95-3.94)	0.00	0.11	0.988	NA	0.14	0.11	0.295	1.28 (0.81-2.01)	1.96x10^{-7}	2.08 (1.58-2.74)	Additive
8	139215155	rs11985359	*FAM135B*	0.16	0.19	0.99	5.05x10^{-7}		0.19	0.22	0.580		0.18	0.18	0.527				Genotypic
10	121497290	rs73355807	*INPP5F*	0.21	0.15	0.99	6.40x10^{-7}		0.17	0.14	0.630		0.08	0.11	0.246				Genotypic
12	125614614	rs76336213	*AACS*	0.19	0.11	0.93	6.65x10^{-7}	2.56 (1.77-3.70)	0.15	0.12	0.395	1.34 (0.68-2.63)	0.09	0.08	0.632	1.14 (0.67-1.93)	5.44x10^{-5}	1.80 (1.35-2.39)	Additive
13	69576658	rs73211791		0.15	0.11	0.99	2.56x10^{-7}		0.11	0.12	0.972		0.11	0.15	0.137				Genotypic
13	69582740	rs9541709		0.16	0.12	0.99	3.78x10^{-9}		0.11	0.12	0.993		0.18	0.17	0.956				Genotypic
19	7900650	rs652710	*EVI5L*	0.24	0.37	0.94	1.21x10^{-6}	0.56 (0.45-0.71)	0.40	0.36	0.550	1.17 (0.7-1.95)	0.33	0.36	0.474	0.89 (0.66-1.21)	1.60x10^{-4}	0.71 (0.60-0.85)	Additive
19	7903484	rs604938	*EVI5L***	0.34	0.46	0.99	2.17x10^{-5}	0.62 (0.50-0.78)	0.46	0.45	0.416	1.19 (0.75-1.9)	0.44	0.42	0.589	1.08 (0.81-1.44)	0.016	0.81 (0.69-0.96)	Additive
19	7910854	rs794442	*EVI5L***	0.33	0.45	0.99	1.27x10^{-5}	0.62 (0.50-0.77)	0.67	0.45	0.629	1.11 (0.69-1.76)	0.42	0.41	0.754	1.05 (0.79-1.39)	4.27x10^{-3}	0.79 (0.67-0.93)	Additive
19	7916620	rs580790	*EVI5L*	0.22	0.36	0.87	1.33x10^{-7}	0.53 (0.42-0.67)	0.40	0.31	0.110	1.51 (0.91-2.49)	0.30	0.34	0.382	0.88 (0.65-1.18)	1.23x10^{-4}	0.71 (0.6-0.85)	Additive
19	7922672	rs2059820	*EVI5L*	0.22	0.36	0.87	1.35x10^{-7}	0.53 (0.42-0.67)	0.42	0.32	0.356	1.39 (0.69-2.81)	0.29	0.31	0.529	0.9 (0.64-1.26)	1.72x10^{-5}	0.67 (0.55-0.80)	Additive
21	35289246	rs2834301	*ATP5O**	0.24	0.15	NA	8.59x10^{-7}		0.10	0.17	0.202		0.17	0.13	0.366				Genotypic

Table 3.5 (preceding page) Notes. The replication genotyping presented was performed with the Sequenom MASSArray platform (Kenyan replication, 38 cases, 297 controls; Malawian replication, 150 cases, 339 controls), with the exception of Kenyan replication samples for which ImmunoChip genotyping data was available (**) and rs13390936 (highlighted in bold) which was HRMA genotyped (Kenyan replication, 38 cases, 1,336 controls; Malawian replication, 150 cases, 339 controls). Imputation accuracy (proportion of correctly imputed genotypes) is calculated for a subset of Sequenom genotyped Kenyan discovery samples (180 cases, 750 controls), SNPs directly genotyped in the discovery analysis are highlighted (*).CHR, chromosome; BP (b37), base-pair position Human Genome Reference GRCh37; MAF, minor allele frequency; OR, odds ratio; CI, confidence interval; NA, not applicable.

Replication genotyping & analysis

We included 1,374 Kenyan samples (38 cases, 1,336 controls) and 489 Malawian samples (150 cases, 339 controls) in the replication genotyping. Immunochip genotyping data was available for the Kenyan replication sample at 6 SNPs included in the replication genotyping, in which case the full complement of Kenyan replication samples was included in the analysis. In cases where Immunochip genotypes were not available, 284 Kenyan replication control samples were included in the genotyping assays along with all of the NTS case samples.

We designed multiplexed Sequenom MASSArray genotyping assays, with the aim of including at least one NTS-associated SNP from the discovery analysis in each genomic region (n = 16). Inspection of cluster plots (TYPER4.0) revealed good cluster separation at 25 SNPs in 13 regions, all of which having call-rates in excess of 95%. Following QC, 416 Malawian samples (135 cases, 281 controls) and 318 Kenyan samples (36 cases, 282 controls) were

included in the Sequenom MASSArray replication analysis (Table 3.5). Imputation accuracy of SNPs taken forward for replication genotyping was also confirmed in 930 Kenyan discovery samples (33%) by Sequenom MASSArray (mean r^2=0.97). The lead *STAT4* region variant, rs13390936, was genotyped by HRMA, which generated clearly differentiated genotypes at the locus in 1,356 Kenyan replication (37 cases, 1,319 controls) and 481 Malawian replication (143 cases, 338 controls) samples (Figure 3.6).

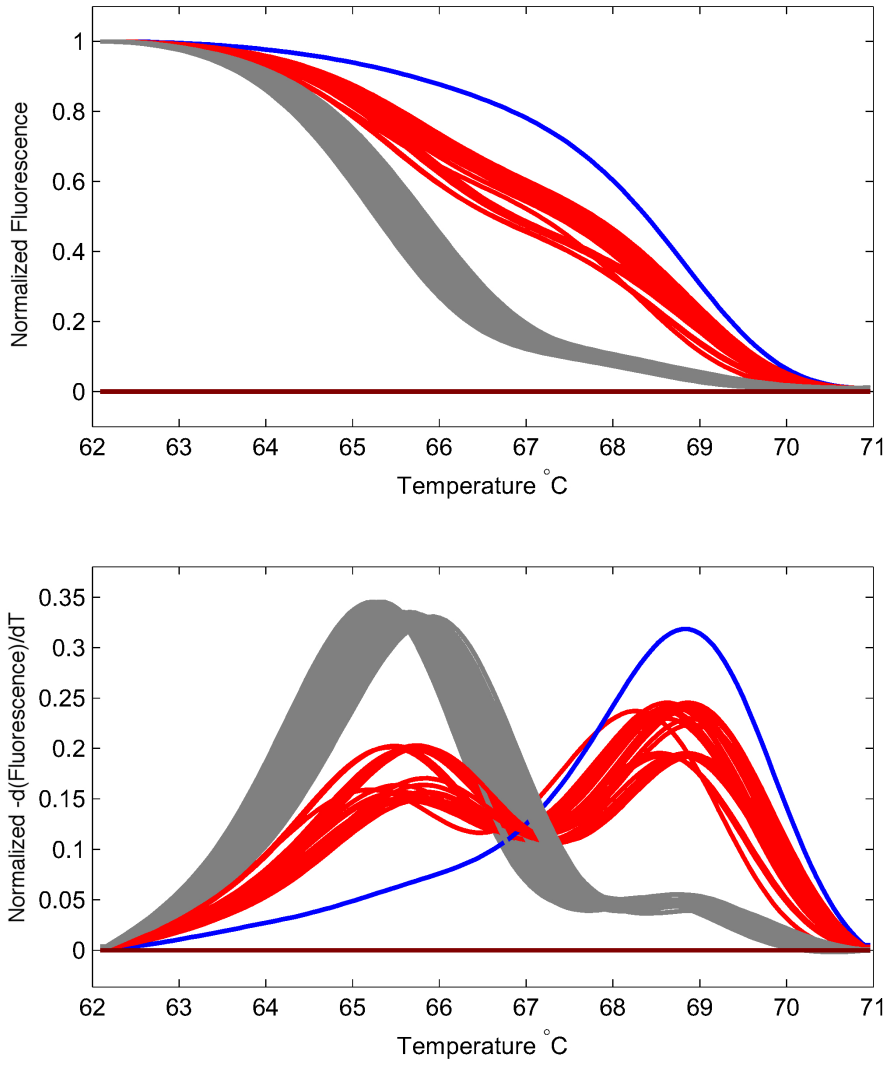

Figure 3.6 High-resolution melt-curve analysis genotyping of rs13390936

Representative normalized melting curves (top) and peaks (bottom) from rs13390936 HRMA genotyping. Data is taken from a single assay (96 individuals). Genotypes are color-coded as follows; AA, grey, AT red, TT blue.

In the replication analysis, one SNP (rs13390936, chr2:191954816) located in the 3rd intron of the signal transducer and activator of transcription 4 (*STAT4*) gene demonstrated evidence of association with NTS bacteraemia in both the Kenyan and Malawian replication samples (Tables 3.5 & 3.6).

Table 3. 6 Association of rs13390936 with NTS bacteraemia in Kenyan discovery, Kenyan replication and Malawian replication samples

	Population	Case		Control		Genotypic model	Recessive model	
		Freq.	TT/AT/AA	Freq.	TT/AT/AA	*P*-value	*P*-value	OR (95% CI)
rs13390936	Kenya (discovery)	0·151	8/40/138	0·107	15/529/2068	$1{\cdot}76x10^{-5}$	$3{\cdot}33x10^{-6}$	8·03 (3·34-19·31)
STAT4	Kenya (replication)	0·146	3/4/30	0·106	12/254/1053	$1{\cdot}83x10^{-3}$	$6{\cdot}54x10^{-4}$	9·96 (2·66-37·36)
	Malawi	0·150	6/31/106	0·136	3/86/249	0·039	0·026	4·89 (1·21-19·83)
	Combined						$8{\cdot}62x10^{-10}$	7·61 (3·98-14·55)

rs13390936 is directly genotyped in all samples by HRMA. Freq., minor allele frequency; OR, odds ratio; CI, confidence interval.

Comparison of the NTS association at rs13390936 under alternative genetic models demonstrated that the observed association is best explained by a recessive model (Table 3.7), with carriage of the minor, TT genotype being associated with increased risk of NTS bacteraemia. Intensive 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 (Figure 3.7). 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 rs13390396 confirmed imputation accuracy (r^2=0·97), and confirmed the association with NTS bacteraemia (Table 3.6). In the combined analysis of all samples, under a recessive model, the association of rs13390936 with NTS bacteraemia exceeds genome-wide significance (P=$8{\cdot}62x10^{-10}$, OR 7·61).

Table 3.7 Comparison of the observed rs13390936 association with NTS bacteraemia under alternative genetic models

	Genetic model									
	Genotypic		Additive		**Recessive**		Dominant		Heterozygous advantage	
Population	*P*-value	AIC	*P*-value	AIC	***P*-value**	**AIC**	*P*-value	AIC	*P*-value	AIC
Kenya (discovery)	$1.76x10^{-5}$	1351	0.012	1359	**$3.33x10^{-6}$**	**1349**	0.141	1363	0.782	1365
Kenya (replication)	$1.83x10^{-3}$	340	0.398	346	**$6.54x10^{-4}$**	**339**	0.869	347	0.209	345
Malawi	0.039	586	0.559	589	**0.026**	**584**	0.917	589	0.379	589

AIC, Akaike information criterion.

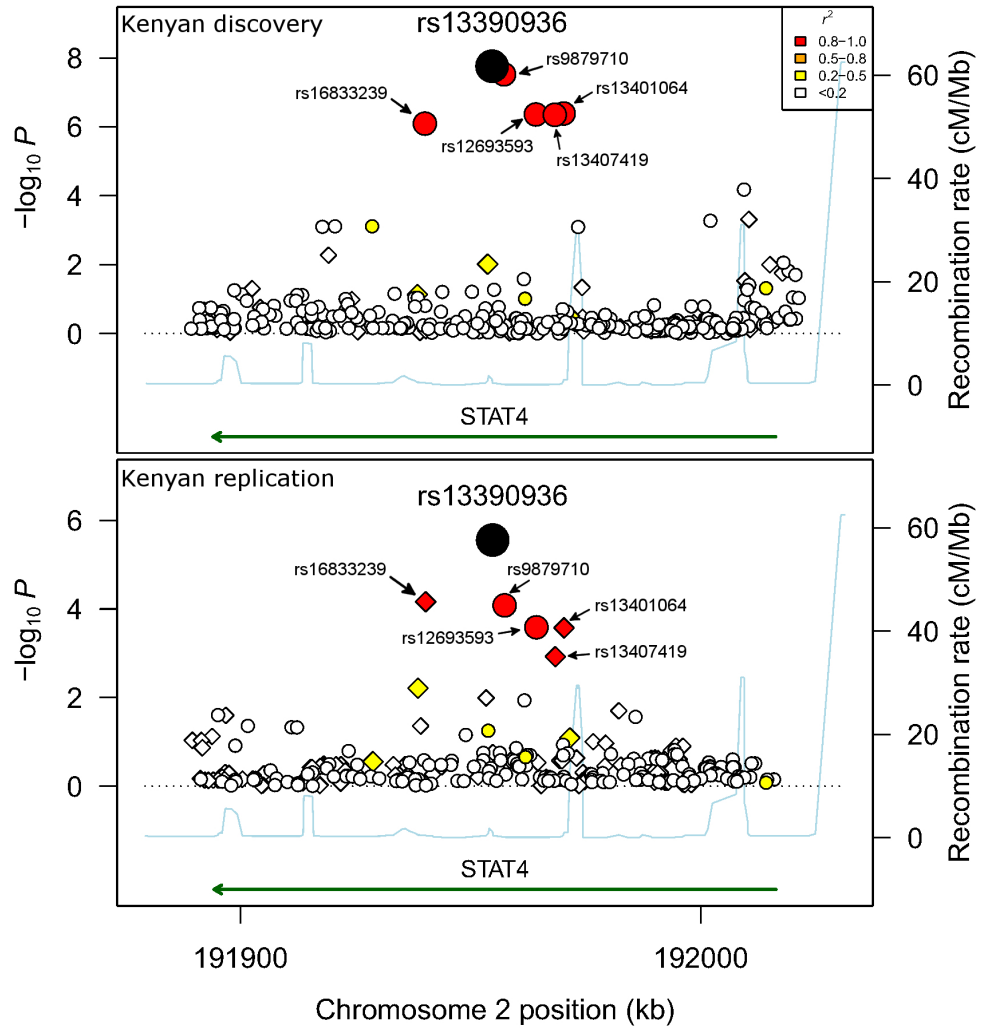

Figure 3.7 Association plot of NTS bacteraemia susceptibility at the *STAT4* region in Kenyan children

NTS-association at the *STAT4* region under a recessive model in Kenyan discovery samples (n = 180 cases, 2,677 controls) and Kenyan replication samples (n = 38 cases, 1,336 controls). rs13390936 is represented as a black circle in each plot. Imputed SNPs (following intensive re-imputation) are plotted as circles, directly genotyped SNPs as diamonds. SNPs are colored according to strength of linkage disequilibrium (r^2) to rs13390936.

Classical HLA alleles

HLA-A

Both HLA-A alleles were well imputed (posterior probability >0.7) in 4,074 individuals of 4,203 Kenyan children included in the Kenyan all-cause bacteraemia discovery analysis. Of those individuals, 171 were cases of NTS bacteraemia and 2,594 were control samples. Twenty-seven, four-digit, classical HLA-A alleles were identified, seven of which have a minor allele frequency of at least 5%, and were included in the association analysis (Figure 3.8A). There is no evidence of association between any common HLA-A allele and NTS bacteraemia (Figure 3.8B).

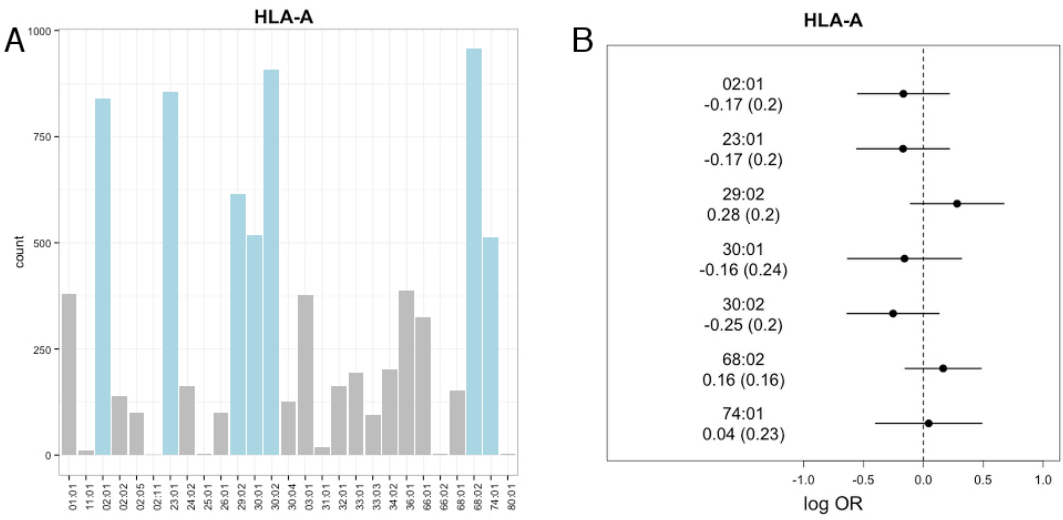

Figure 3.8 HLA-A allelic association with nontyphoidal *Salmonella* bacteraemia in Kenyan children.

Distribution of observed HLA-A alleles in Kenyan children (n = 4,074) included in the analysis (A). Log-transformed odds ratios and 95% confidence intervals of NTS bacteraemia association with common (minor allele frequency > 5%) HLA-A alleles (B).

There is no evidence for significant association of common HLA-A alleles with bacteraemia caused by any pathogen. This is the case both for regression models of all-cause bacteraemia (1,480 cases), and for multinomial logistic regression models, in which common causes of bacteraemia (NTS, n = 171; *S. aureus*, n = 158; *S. pneumonia*, n = 404; *H influenza*e, n = 130; *E. coli*, n = 149; β-haemolytic *Streptococci*, n = 153; *Acinetobacter*, n = 125) are included as strata in the analysis (Figure 3.9).

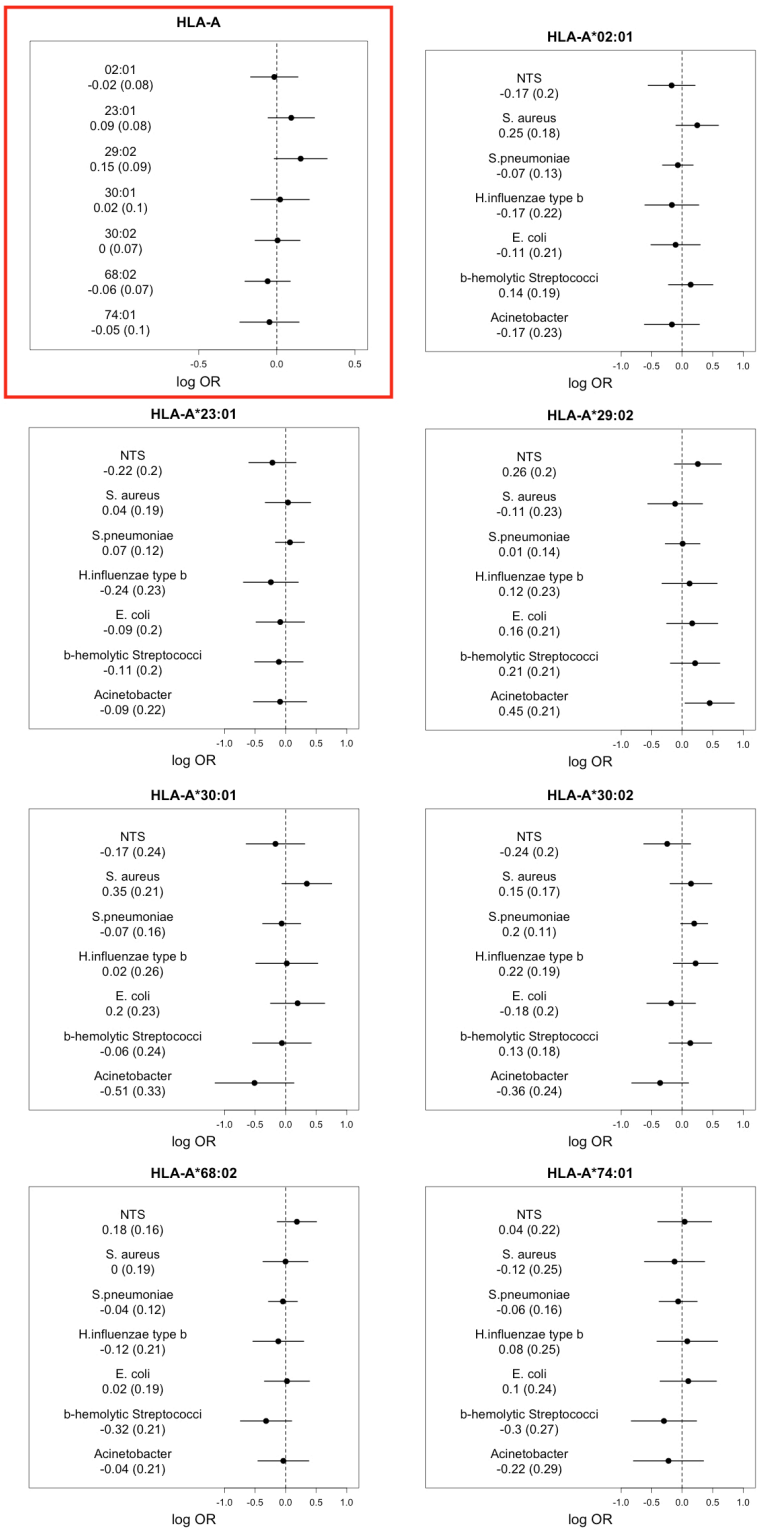

Figure 3.9 HLA-A allelic association with all-cause bacteraemia in Kenyan children.

Log-transformed odds ratios and 95% confidence intervals of all-cause bacteraemia association with common HLA-A alleles (minor allele frequency > 5%). HLA-A association analysis is displayed for all-cause bacteraemia (red box), and for common bacterial pathogens at each HLA-A allele.

HLA-B

Both HLA-B alleles were well imputed (posterior probability >0.7) in 2,596 individuals of 4,213 Kenyan children included in the Kenyan all-cause bacteraemia discovery analysis. Of those individuals, 104 were cases of NTS bacteraemia and 1,674 were control samples. Forty-one, four-digit, classical HLA-B alleles were identified, seven of which have a minor allele frequency of at least 5%, and were included in the association analysis (Figure 3.10A). There is no evidence of association between any common HLA-B allele and NTS bacteraemia (Figure 3.10B).

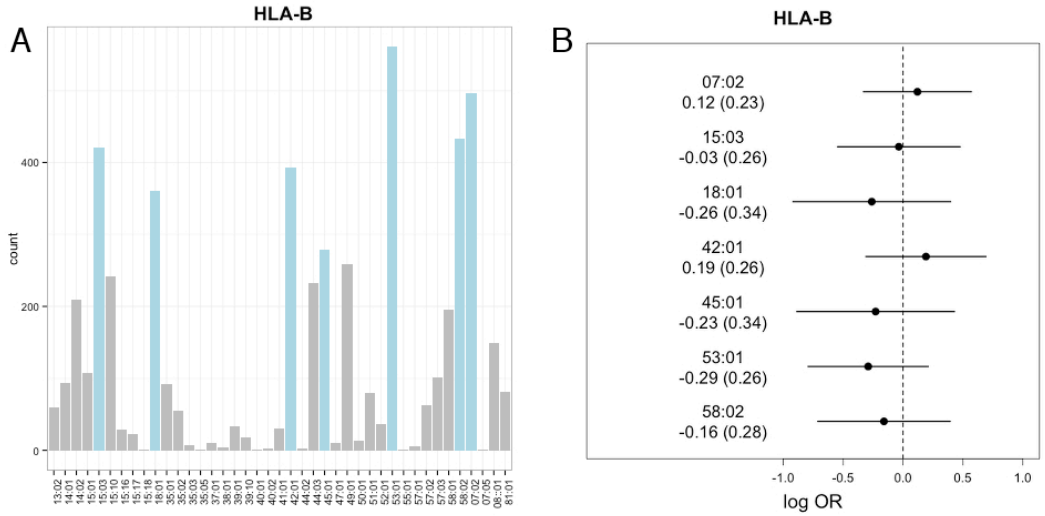

Figure 3.10 HLA-C allelic association with nontyphoidal *Salmonella* bacteraemia in Kenyan children.

Distribution of observed HLA-B alleles in Kenyan children (n = 2,596) included in the analysis (A). Log-transformed odds ratios and 95% confidence intervals of NTS bacteraemia association with common (minor allele frequency > 5%) HLA-B alleles (B).

There is no evidence for significant association of common HLA-B alleles with bacteraemia caused by any pathogen. This is the case both for regression models of all-cause bacteraemia (922 cases), and for multinomial logistic regression models, in which common causes of bacteraemia (NTS, n = 104; *S. aureus*, n = 94; *S. pneumonia*, n = 252; *H influenza*e, n = 78; *E. coli*, n = 89; β-haemolytic *Streptococci*, n = 107; *Acinetobacter*, n = 69) are included as strata in the analysis (Figure 3.11).

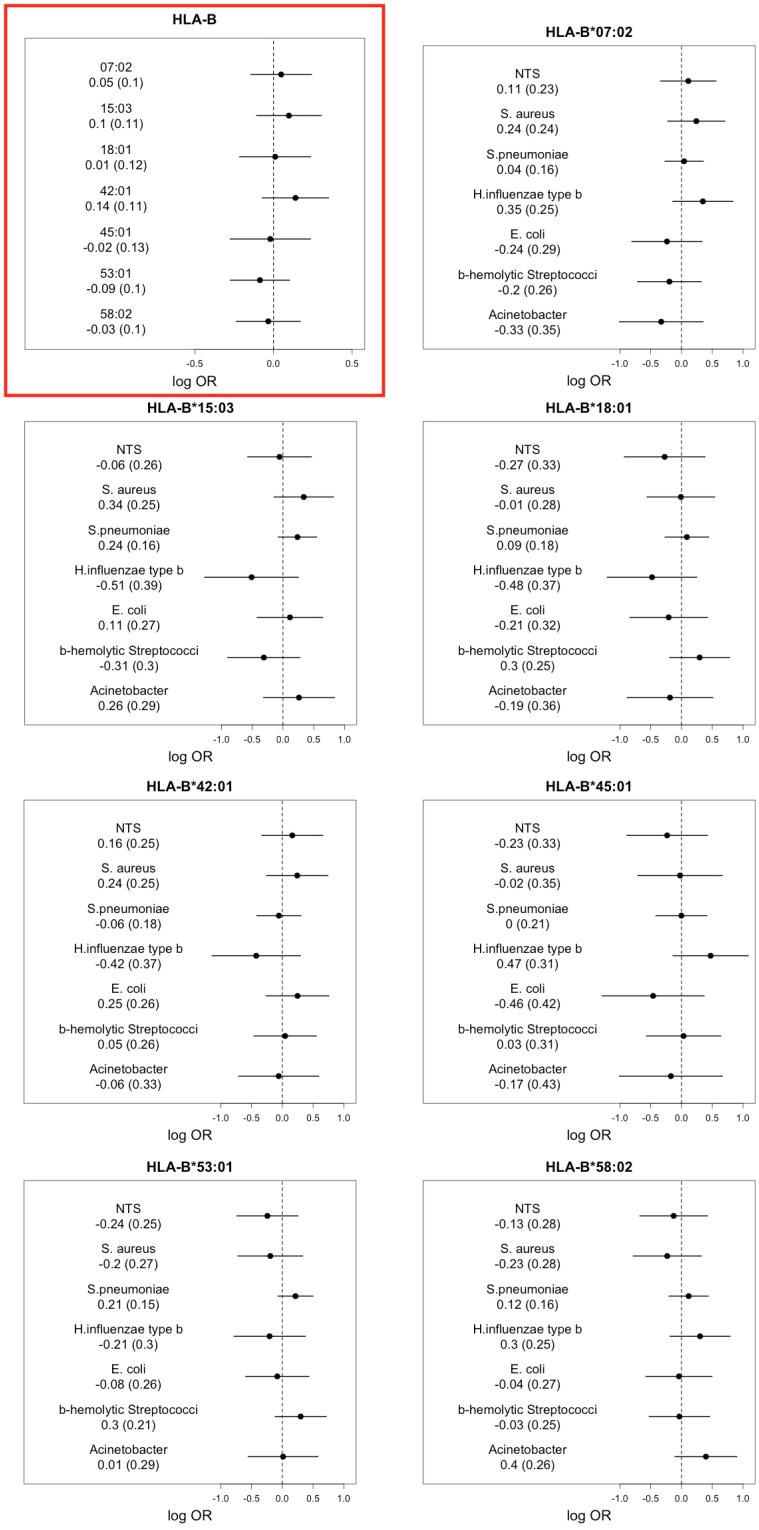

Figure 3.11 HLA-B allelic association with all-cause bacteraemia in Kenyan children.

Log-transformed odds ratios and 95% confidence intervals of all-cause bacteraemia association with common HLA-B alleles (minor allele frequency > 5%). HLA-B association analysis is displayed for all-cause bacteraemia (red box), and for common bacterial pathogens at each HLA-B allele.

HLA-C

Both HLA-C alleles were well imputed (posterior probability >0.7) in 3,502 individuals of 4,213 Kenyan children included in the Kenyan all-cause bacteraemia discovery analysis. Of those individuals, 144 were cases of NTS bacteraemia and 2,236 were control samples. Twenty-three, four-digit, classical HLA-C alleles were identified, ten of which have a minor allele frequency of at least 5%, and were included in the association analysis (Figure 3.12A). In an additive linear model, there is evidence of association between NTS bacteraemia and carriage of HLA-C*17:01 (P = 6.8x10^{-3}, OR = 1.71, 95% CI 1.14-3.48) and HLA-C*18.01 (P = 0.018, OR = 0.37, 95% CI 0.14-0.77) (Figure 3.12B).

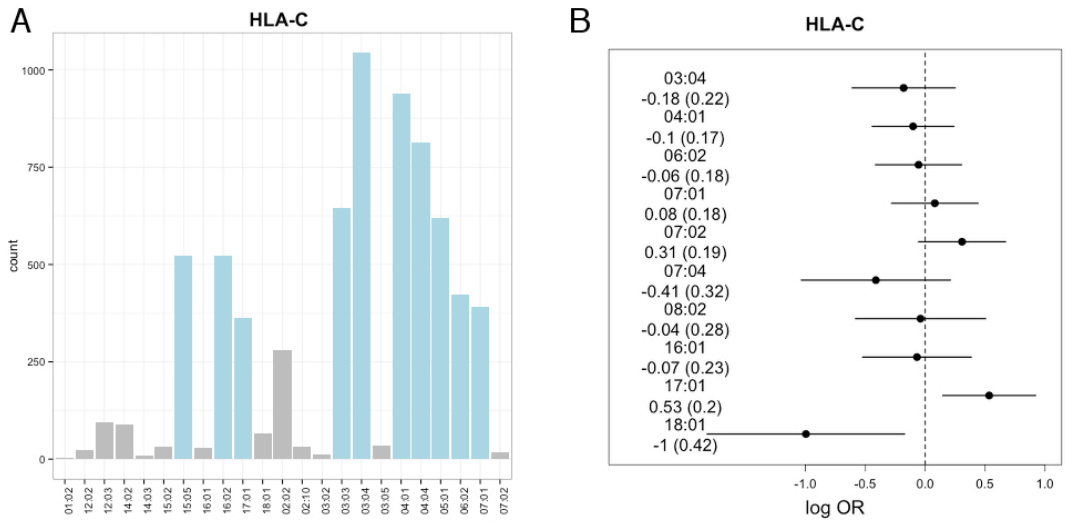

Figure 3.12 HLA-C allelic association with nontyphoidal *Salmonella* (NTS) bacteraemia in Kenyan children.

Distribution of observed HLA-C alleles in Kenyan children (n = 3,502) included in the analysis (A). Log-transformed odds ratios and 95% confidence intervals of NTS bacteraemia association with common (minor allele frequency > 5%) HLA-C alleles (B).

In an association analysis of all-cause bacteraemia and HLA-C alleles (Figure 3.13), in keeping with the NTS data, carriage of HLA-C*17:01 is associated with risk of all cause bacteraemia (P = 3.3x10^{-3}, OR = 1.31, 95% CI 1.09-1.58). In the multinomial regression model of HLA-C*17:01 carriage and risk of bacteraemia secondary to common pathogens (NTS, n = 144; *S. aureus*, n = 121; *S. pneumonia*, n = 354; *H influenza*e, n = 110; *E. coli*, n = 123; β-haemolytic *Streptococci*, n = 138; *Acinetobacter*, n = 104), the strength of the observed association is strongest for NTS, with the data best supporting a model in which HLA-C*17:01 carriage increases risk of bacteraemia secondary to NTS, *S. aureus, E. coli, Acinetobacter* (Bayes factor c.f. null = 4). In the association analysis of all-cause bacteraemia there is no evidence of association with HLA-C*18:01.

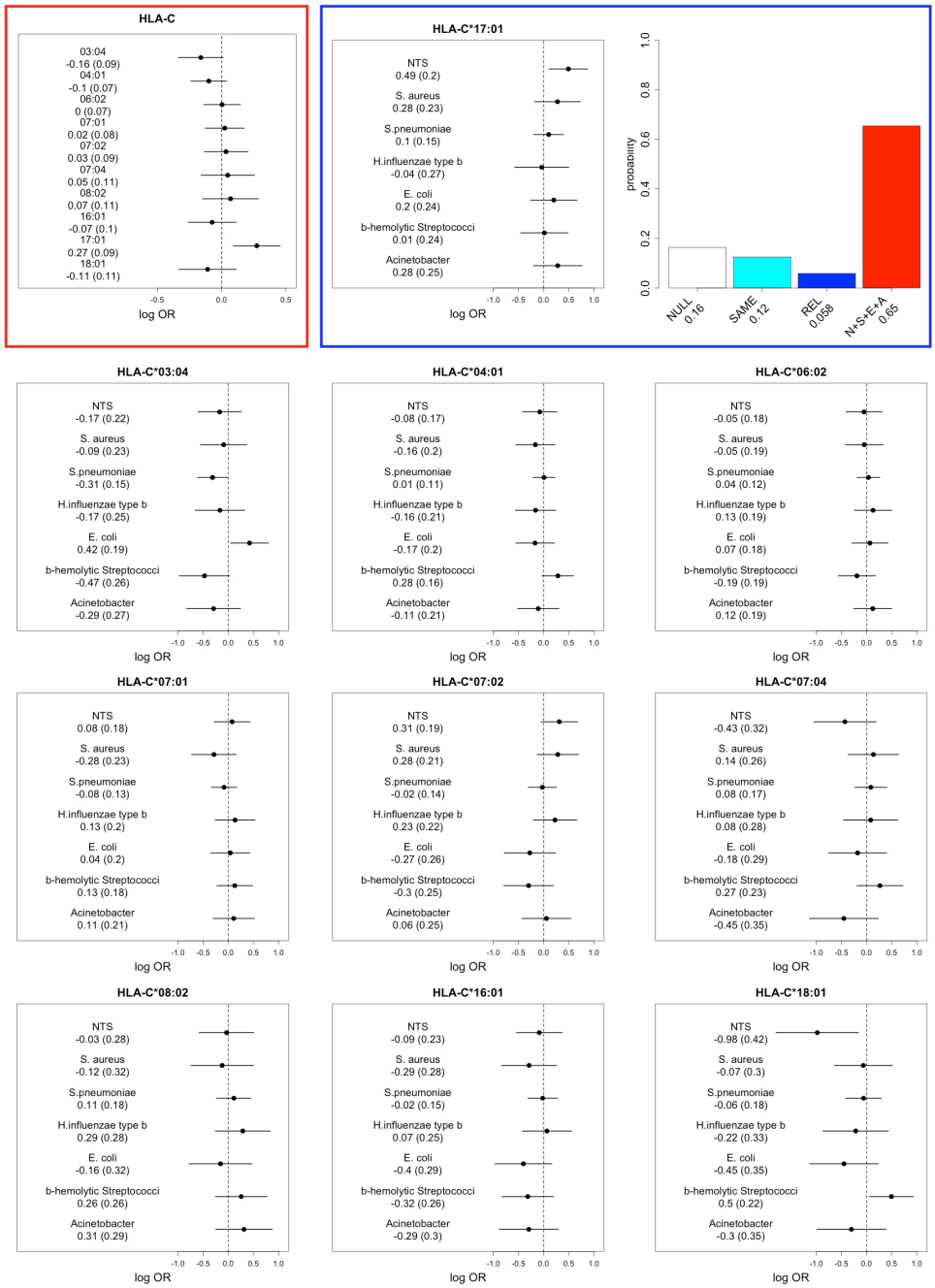

Figure 3.13 HLA-C allelic association with all-cause bacteraemia in Kenyan children.

Log-transformed odds ratios and 95% confidence intervals of all-cause bacteraemia association with common HLA-C alleles (minor allele frequency > 5%). HLA-C association analysis is displayed for all-cause bacteraemia (red box), and for common bacterial pathogens at each HLA-C allele. Multinomial regression model and posterior probabilities of models of association at HLA-C*17:01 (blue box): NULL, no association with any bacterial pathogen; SAME, the same effect across all bacterial pathogens; REL, related effects across all bacterial pathogens; N+S+E+A, a non-zero effect in the NTS, *S. aureus, E. coli* and *Acinetobacter* alone.

In the Kenyan replication samples, both HLA-C alleles were well imputed (posterior probability >0.7) in 1,625 individuals of 1,770 included in the all-cause bacteraemia replication analysis. These individuals were included in a replication analysis of HLA-C*17:01 association with bacteraemia (NTS cases, n = 33; *S. aureus*, n = 42; *S. pneumonia*, n = 105; *H influenza*e, n = 29; *E. coli*, n = 40; β-haemolytic *Streptococci*, n = 41; *Acinetobacter*, n = 41; controls, n = 1,229). In that analysis, there is no evidence of replication at HLA-C*17:01 with NTS (*P* = 0.45, OR = 1.36, 95% CI 0.56-2.82) or all-cause (*P* = 0.24, OR = 0.87, 95% CI 0.60-1.13) bacteraemia. A Bayesian analysis of the evidence for the most likely model observed in the discovery analysis (Figure 3.14), demonstrates no evidence for that model in the replication samples (Bayes factor c.f. null = 0.22).

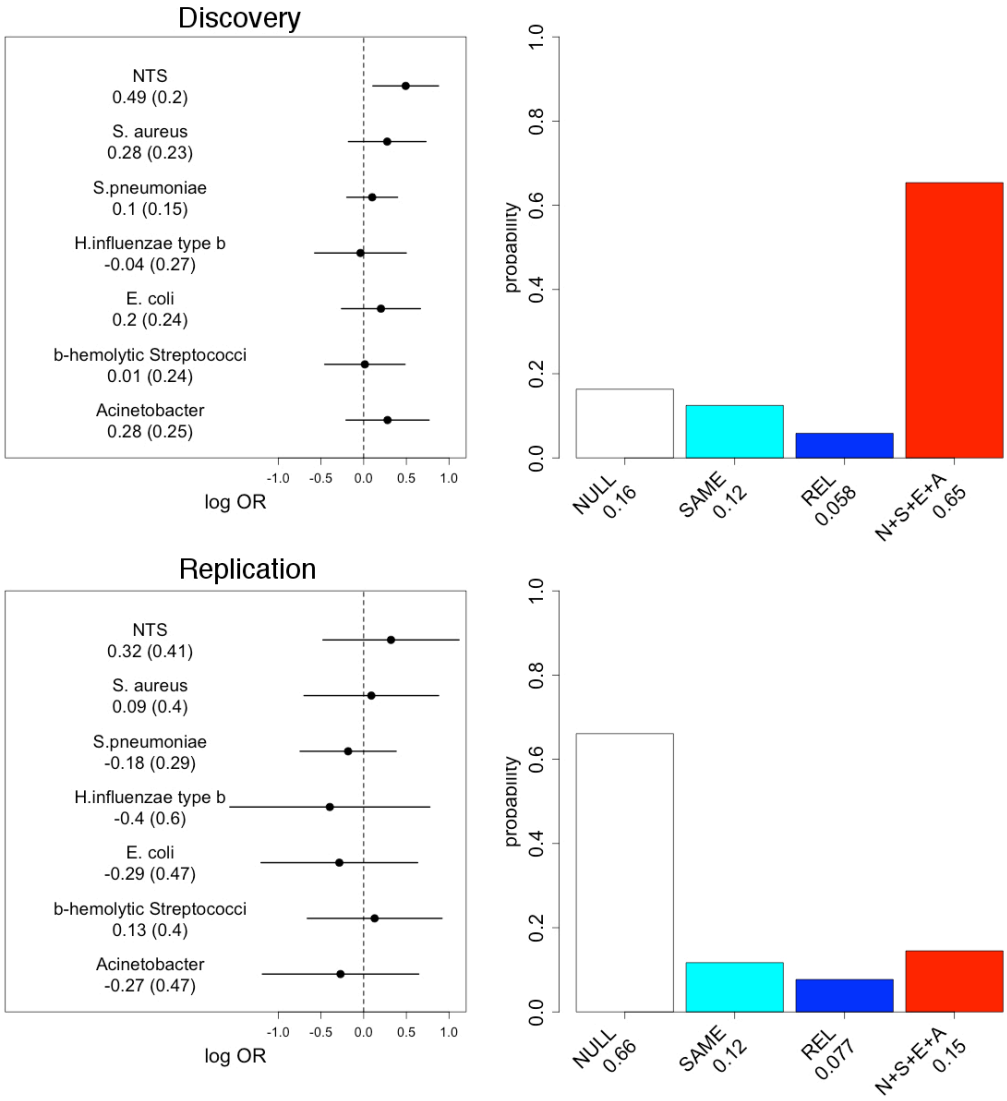

Figure 3.14 HLA-C*17:01 allelic association with all-cause bacteraemia in Kenyan discovery and replication samples.

Log-transformed odds ratios and 95% confidence intervals of HLA-C*17:01 association, in Kenyan discovery (top), and replication (bottom) samples; left panels. Posterior probabilities of models of association at HLA-C*17:01 (right panels): NULL, no association with any bacterial pathogen; SAME, the same effect across all bacterial pathogens; REL, related effects across all bacterial pathogens; N+S+E+A, a non-zero effect in bacteraemia secondary to NTS, *S. aureus, E. coli* and *Acinetobacter*.

HLA-DQA1

Both HLA-DQA1 alleles were well imputed (posterior probability >0.7) in 3,397 individuals of 4,213 Kenyan children included in the Kenyan all-cause bacteraemia discovery analysis. Of those individuals, 148 were cases of NTS bacteraemia and 2,174 were control samples. Eight, four-digit, classical HLA-DQA1 alleles were identified, six of which have a minor allele frequency of at least 5%, and were included in the association analysis (Figure 3.15A). There is no evidence of association between any common HLA-DQA1 allele and NTS bacteraemia (Figure 3.15B).

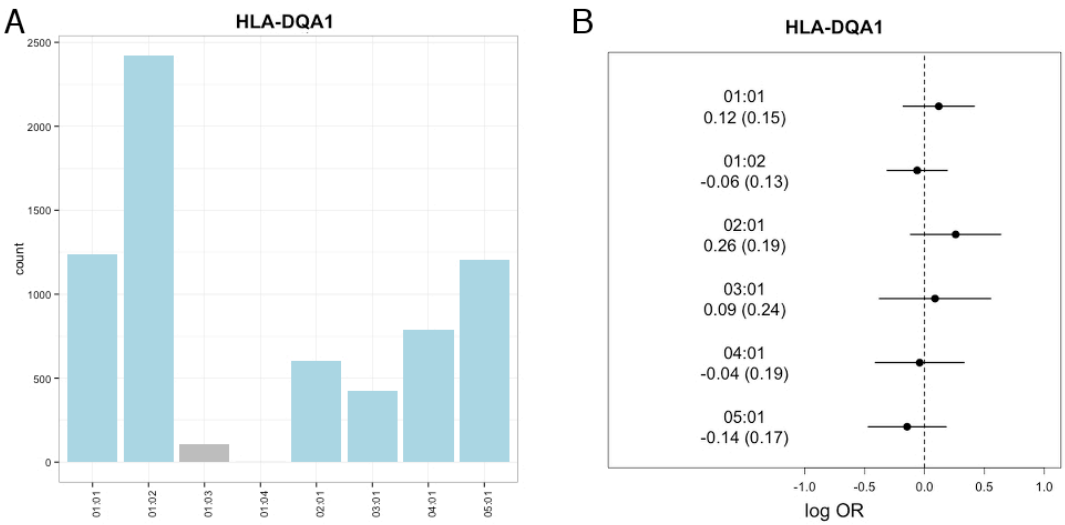

Figure 3.15 HLA-DQA1 allelic association with nontyphoidal *Salmonella* (NTS) bacteraemia in Kenyan children.

Distribution of observed HLA-DQA1 alleles in Kenyan children (n = 3,397) included in the analysis (A). Log-transformed odds ratios and 95% confidence intervals of NTS bacteraemia association with common (minor allele frequency > 5%) HLA-DQA1 alleles (B).

There is no evidence for significant association of common HLA-DQA1 alleles with bacteraemia caused by any pathogen. This is the case both for regression models of all-cause bacteraemia (1,223 cases), and for multinomial logistic regression models, in which common causes of bacteraemia (NTS, n = 148; *S. aureus*, n = 124; *S. pneumonia*, n = 335; *H influenzae*, n = 110; *E. coli*, n = 120; β-haemolytic *Streptococci*, n = 116; *Acinetobacter*, n = 100) are included as strata in the analysis (Figure 3.16).

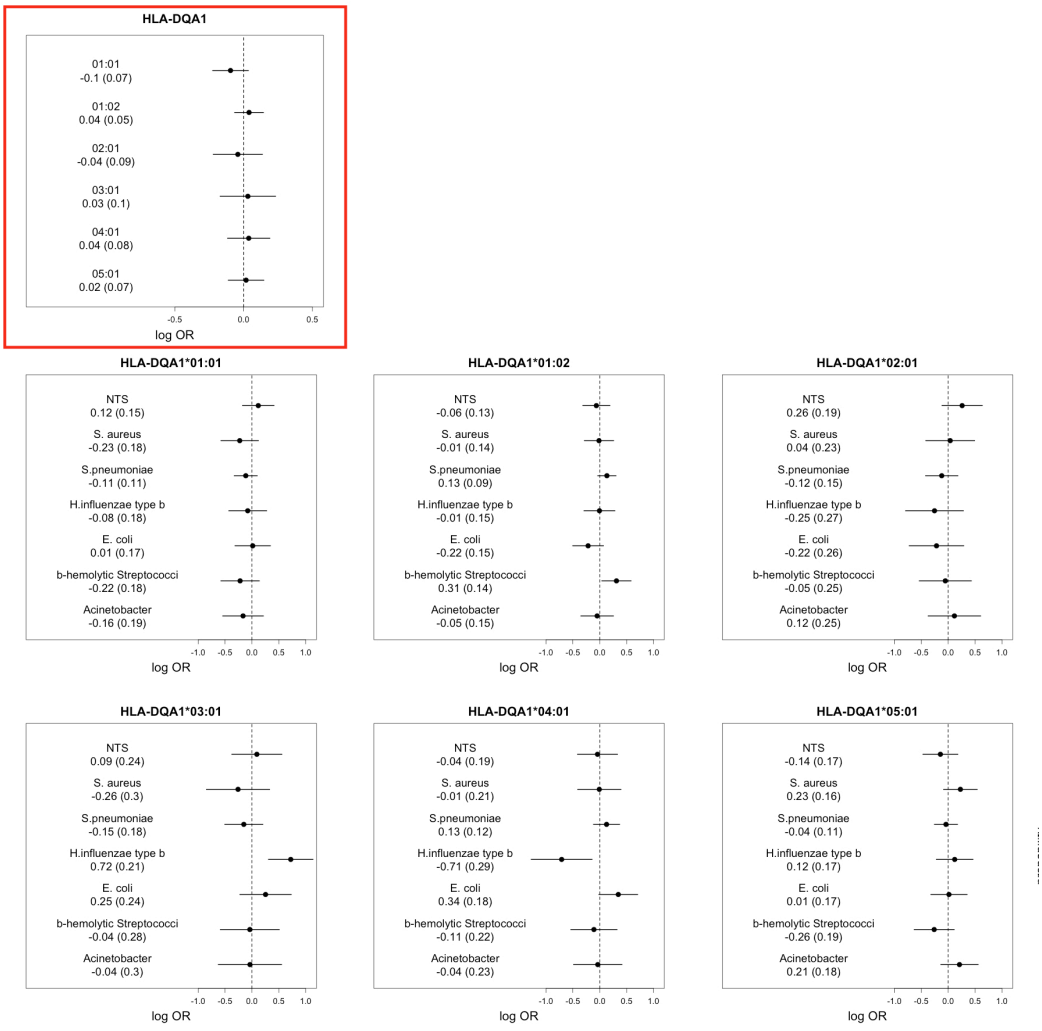

Figure 3.16 HLA-DQA1 allelic association with all-cause bacteraemia in Kenyan children.

Log-transformed odds ratios and 95% confidence intervals of all-cause bacteraemia association with common HLA-DQA1 alleles (minor allele frequency > 5%). HLA-DQA1 association analysis is displayed for all-cause bacteraemia (red box), and for common bacterial pathogens at each HLA-DQA1 allele.

HLA-DQB1

Both HLA-DQB1 alleles were well imputed (posterior probability >0.7) in 3,275 individuals of 4,213 Kenyan children included in the Kenyan all-cause bacteraemia discovery analysis. Of those individuals, 135 were cases of NTS bacteraemia and 2,100 were control samples. Fourteen, four-digit, classical HLA-DQB1 alleles were identified, six of which have a minor allele frequency of at least 5%, and were included in the association analysis (Figure 3.17A). There is no evidence of association between any common HLA-DQB1 allele and NTS bacteraemia (Figure 3.17B).

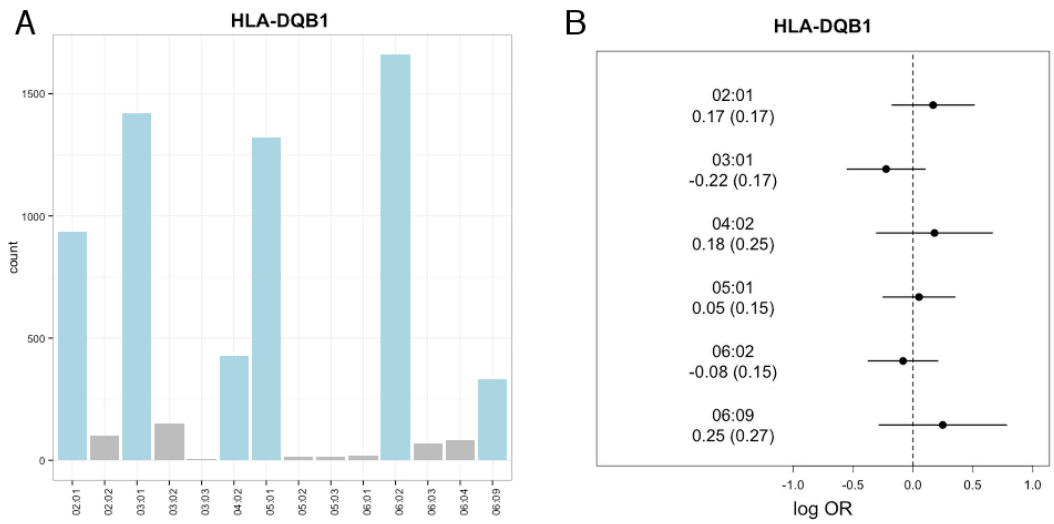

Figure 3. 17 HLA-DQB1 allelic association with nontyphoidal *Salmonella* (NTS) bacteraemia in Kenyan children.

Distribution of observed HLA-DQB1 alleles in Kenyan children (n = 3,275) included in the analysis (A). Log-transformed odds ratios and 95% confidence intervals of NTS bacteraemia association with common (minor allele frequency > 5%) HLA-DQB1 alleles (B).

There is no evidence for significant association of common HLA-DQB1 alleles with bacteraemia caused by any pathogen. This is the case both for

regression models of all-cause bacteraemia (1,175 cases), and for multinomial logistic regression models, in which common causes of bacteraemia (NTS, n = 135; *S. aureus*, n = 127; *S. pneumonia*, n = 337; *H influenzae*, n = 95; *E. coli*, n = 108; β-haemolytic *Streptococci*, n = 120; *Acinetobacter*, n = 97) are included as strata in the analysis (Figure 3.18).

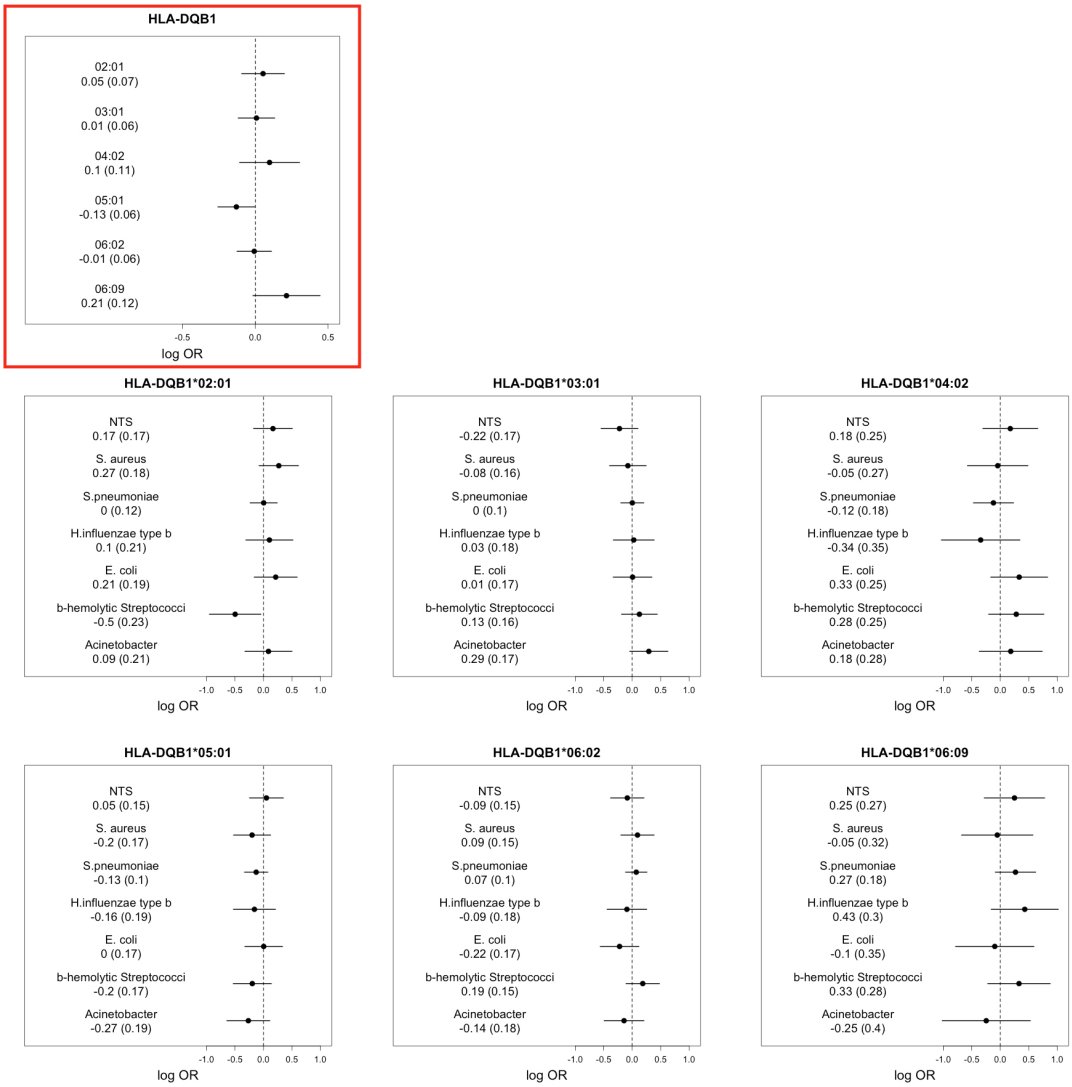

Figure 3.18 HLA-DQB1 allelic association with all-cause bacteraemia in Kenyan children.

Log-transformed odds ratios and 95% confidence intervals of all-cause bacteraemia association with common HLA-DQB1 alleles (minor allele frequency > 5%). HLA-DQB1 association analysis is displayed for all-cause bacteraemia (red box), and for common bacterial pathogens at each HLA-DQB1 allele.

HLA-DRB1

Both HLA-DRB1 alleles were well imputed (posterior probability >0.7) in 3,521 individuals of 4,213 Kenyan children included in the Kenyan all-cause bacteraemia discovery analysis. Of those individuals, 153 were cases of NTS bacteraemia and 2,245 were control samples. Twenty-eight, four-digit, classical HLA-DRB1 alleles were identified, eight of which have a minor allele frequency of at least 5%, and were included in the association analysis (Figure 3.19A). In an additive linear model, there is evidence of association between NTS bacteraemia and carriage of HLA-DRB1*11:01 (*P* = 0.014, OR = 0.68, 95% CI 0.50-0.92) (Figure 3.19B).

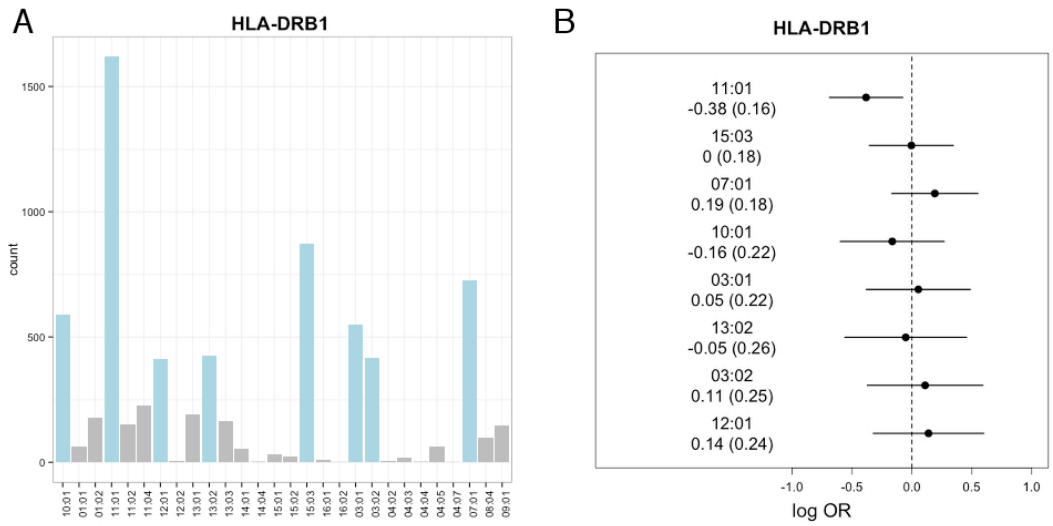

Figure 3.19 HLA-DRB1 allelic association with nontyphoidal *Salmonella* (NTS) bacteraemia in Kenyan children.

Distribution of observed HLA-DRB1 alleles in Kenyan children (n = 3,275) included in the analysis (A). Log-transformed odds ratios and 95% confidence intervals of NTS bacteraemia association with common (minor allele frequency > 5%) HLA-DRB1 alleles (B).

In a multinomial logistic regression model of bacteraemia risk, HLA-DRB1*11:01 carriage is most strongly associated with NTS bacteraemia (Figure 3.20). The data provides modest evidence to support a model in which DRB1*11:01 carriage is associated with decreased risk of NTS bacteraemia alone (Bayes factor c.f. null = 2.7).

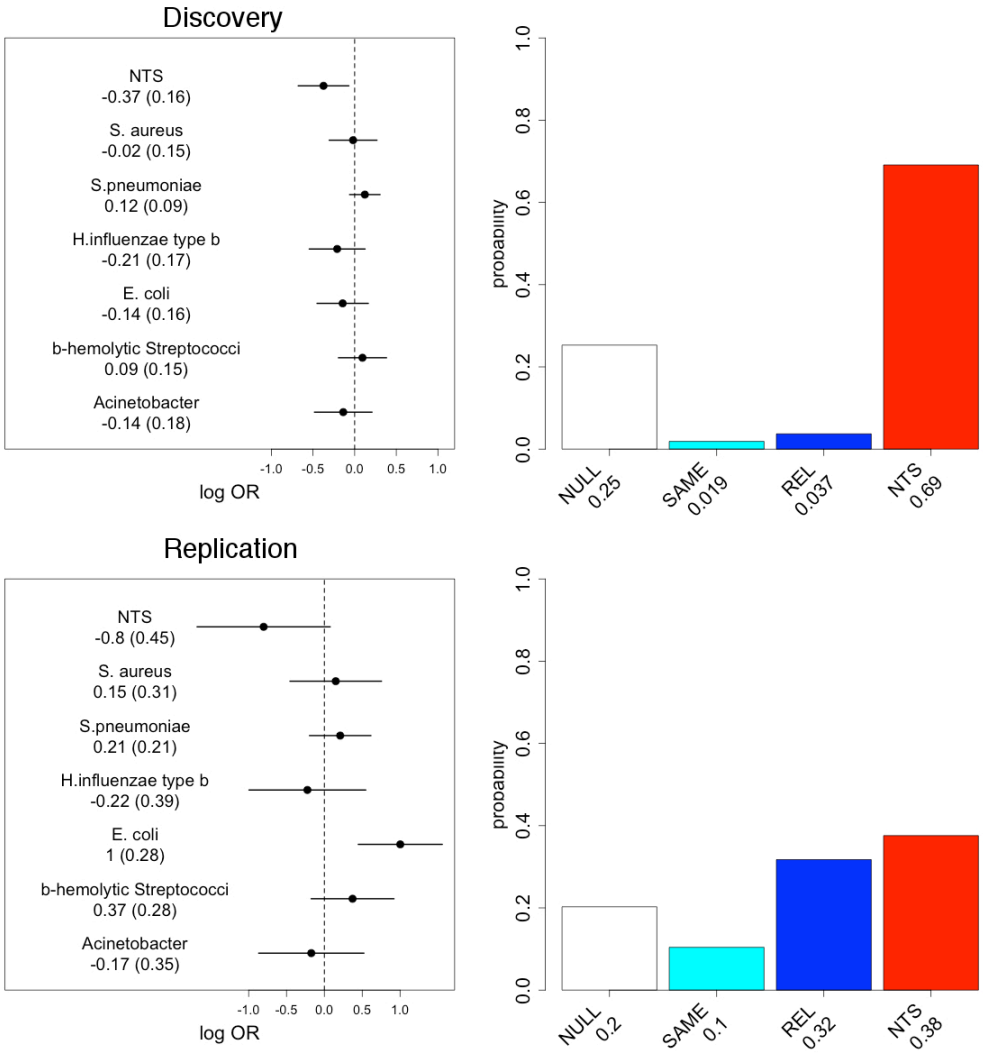

Figure 3.20 HLA-DRB1*11:01 allelic association with all-cause bacteraemia in Kenyan discovery and replication samples.

Log-transformed odds ratios and 95% confidence intervals of HLA-DRB1*11:01 association, in Kenyan discovery (top), and replication (bottom) samples; left panels. Posterior probabilities of models of association at HLA-DRB1*11:01 (right panels): NULL, no association with any bacterial pathogen; SAME, the same effect across all bacterial pathogens; REL, related effects across all bacterial pathogens; NTS, a non-zero effect for association with NTS bacteraemia alone.

In the Kenyan replication samples, both HLA-DRB1 alleles were well imputed (posterior probability >0.7) in 1,327 individuals of 1,770 included in the all-cause bacteraemia replication analysis. These individuals were included in a replication analysis of HLA-DRB1*11:01 association with bacteraemia (NTS cases, n = 31; *S. aureus*, n = 36; *S. pneumonia*, n = 84; *H influenza*e, n = 25; *E. coli*, n = 35; β-haemolytic *Streptococci*, n = 35; *Acinetobacter*, n = 32; controls, n = 993). In that analysis, there is a trend towards replication of the protective association of DRB1*11:01 observed for NTS bacteraemia in the discovery analysis (*P* = 0.073, OR = 0.45, 95% CI 0.17-0.99). In a multinomial regression model of bacteraemia risk in the replication samples (Figure 3.20), the data provides further, modest evidence for a model in which DRB1*11:01 carriage is associated with protection against NTS bacteraemia alone (Bayes factor c.f. null = 1.9).

In an association analysis of all-cause bacteraemia and HLA-DRB1 alleles (Figure 3.21), carriage of HLA-DRB1*10:01 is associated with decreased risk of all cause bacteraemia (*P* = 0.024, OR = 0.81, 95% CI 0.68-0.97). In the multinomial regression model of HLA-DRB1*10:01 carriage and risk of bacteraemia secondary to common pathogens (NTS, n = 153; *S. aureus*, n = 139; *S. pneumonia*, n = 345; *H influenza*e, n = 111; *E. coli*, n = 129; β-haemolytic *Streptococci*, n = 129; *Acinetobacter*, n = 100), there is a marginal protective effect observed across all pathogens, however the data best supports a model in which HLA-DRB1*10:01 carriage is not associated with bacteraemia (Figure 3.21). In keeping with this, there is no evidence of

replication of the DRB1*10:01 association in the replication analysis of all-cause bacteraemia (*P* = 0.49, OR = 0.90, 95% CI 0.66-1.21).

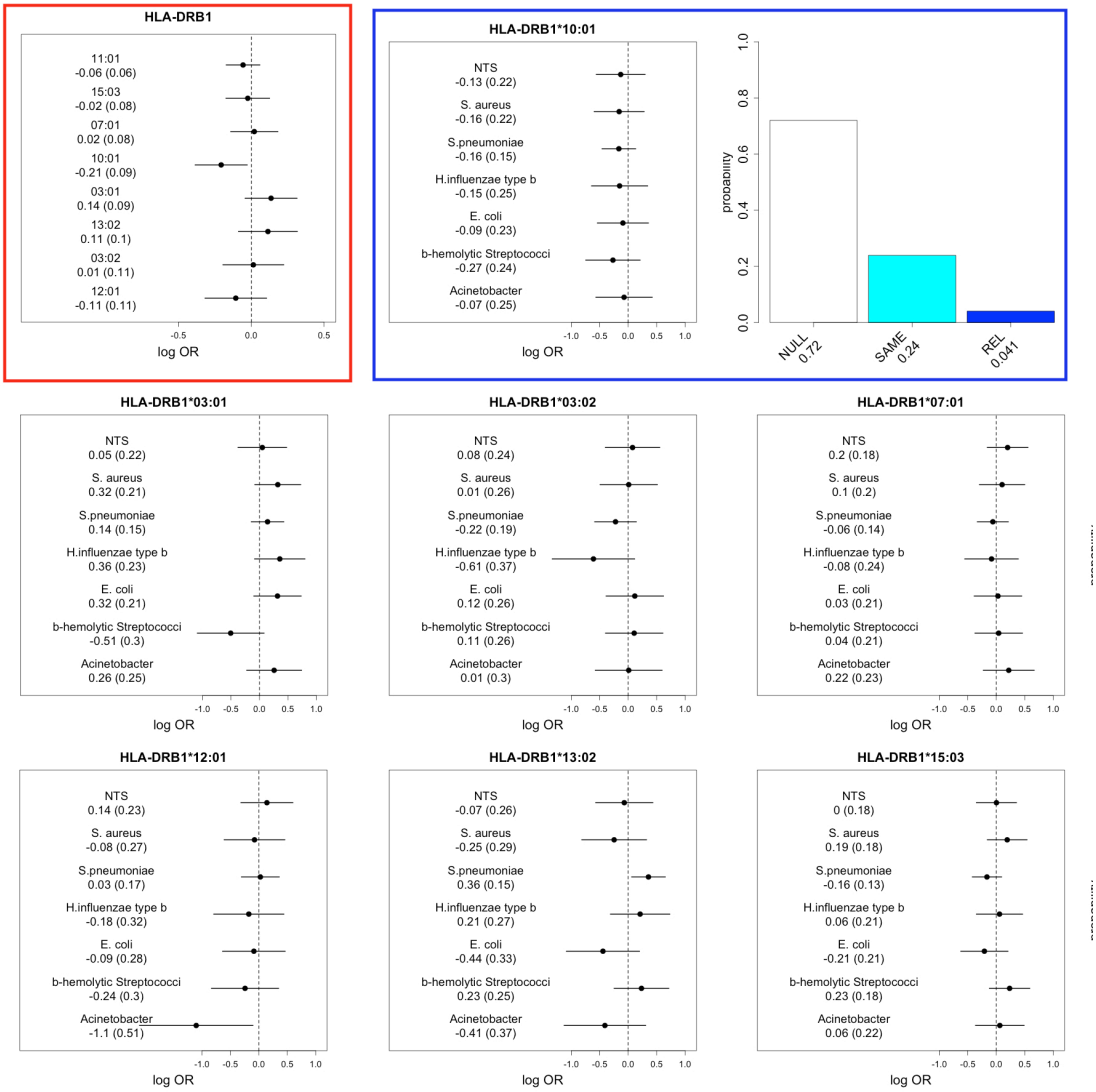

Figure 3.21 HLA-DRB1 allelic association with all-cause bacteraemia in Kenyan children.

Log-transformed odds ratios and 95% confidence intervals of all-cause bacteraemia association with common HLA-DRB1 alleles (minor allele frequency > 5%). HLA-DRB1 association analysis is displayed for all-cause bacteraemia (red box), and for common bacterial pathogens at each HLA-DRB1 allele (with the exception of DRB1*11:01 – see Figure 3.20).

HLA allele homozygosity

Of the 1,436 individuals with good quality imputation at all six class I and class II alleles, 552 individuals are homozygous for at least one four-digit HLA allele. In an association analysis of HLA homozygosity and NTS bacteraemia (NTS cases, n = 53; controls, n = 953), homozygosity at any of the six major class I and class II loci was not found to be associated with risk of NTS bacteraemia (Figure 3.22).

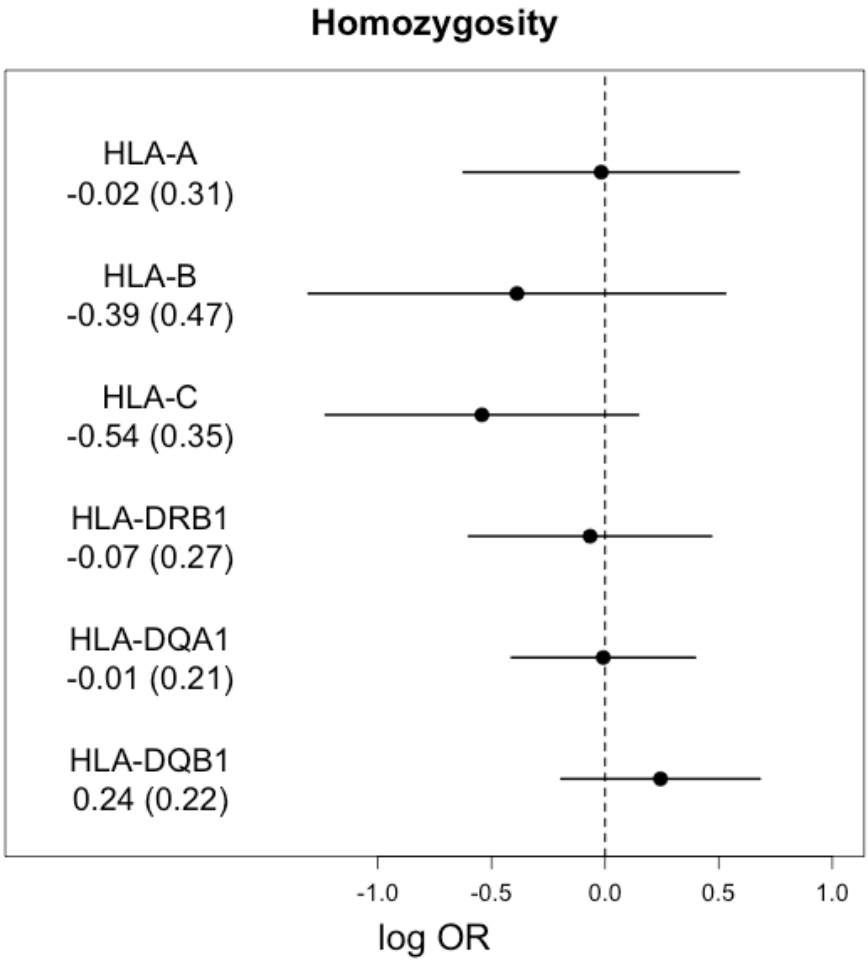

Figure 3.22 HLA homozyogosity and nontyphoidal *Salmonella* (NTS) bacteraemia risk in Kenyan children.

Log-transformed odds ratios and 95% confidence intervals of NTS bacteraemia association with homozygosity at each HLA locus.

Similarly, there is no evidence for significant association between homozygosity at any HLA allele with bacteraemia caused by any pathogen. This is the case both for regression models of all-cause bacteraemia (483 cases), and for multinomial logistic regression models, in which common causes of bacteraemia (NTS, n = 53; *S. aureus*, n = 45; *S. pneumonia*, n = 148; *H influenza*e, n = 40; *E. coli*, n = 39; β-haemolytic *Streptococci*, n = 56; *Acinetobacter*, n = 35) are included as strata in the model (Figure 3.23).

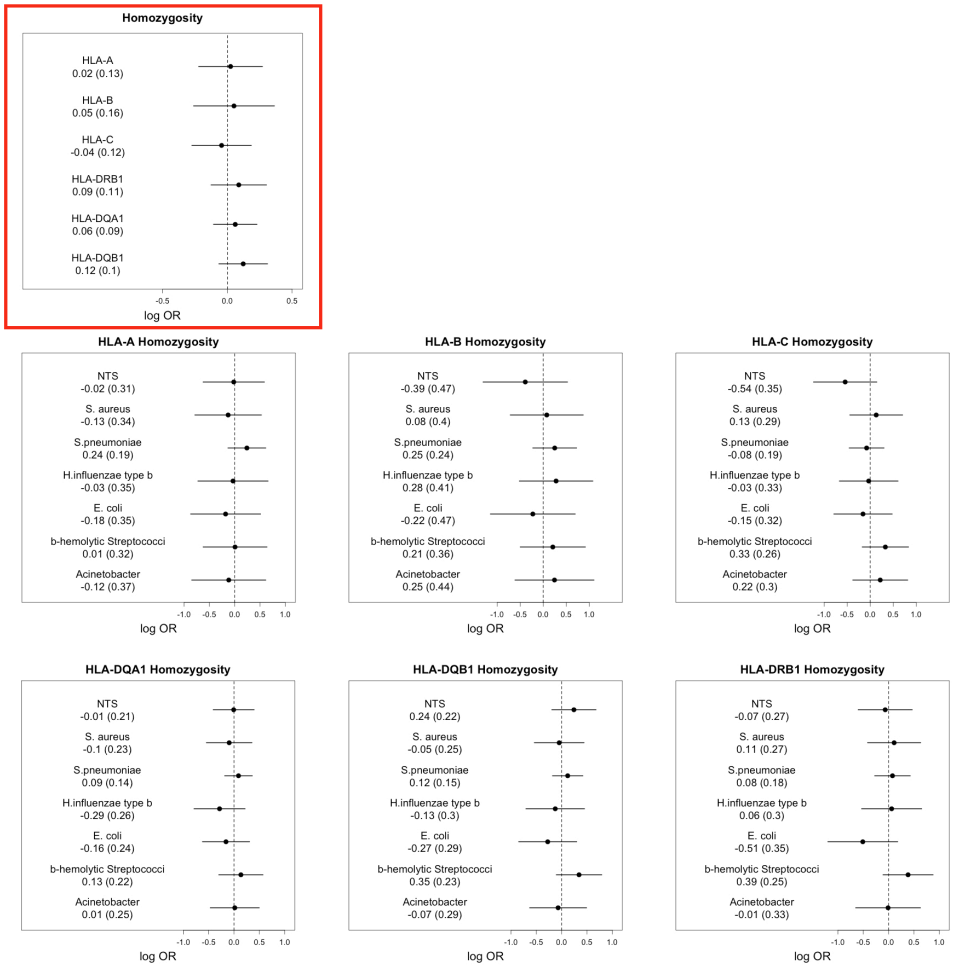

Figure 3.23 The association of homozygosity at HLA loci with all-cause bacteraemia in Kenyan children.

Log-transformed odds ratios and 95% confidence intervals of all-cause bacteraemia association with homozygosity at class I and class II loci. Association of HLA homozygosity with all-cause bacteraemia is displayed (red box), alongside multinomial association analysis for common bacterial pathogens with homozygosity at each HLA locus.

Sickle cell homozygosity & trait

Sickle cell disease (HbSS) and sickle cell trait (HbAS) are common among the population of Kenyan children included in the study. Of children included in the study, 2.7% (84 of 3,152) have sickle cell disease, and 19.0% (599 of 3,152) sickle cell trait.

Under an additive model of inheritance, including the first two principal components of genome-wide genotyping data, there is no evidence of an association between NTS bacteraemia and the sickle cell disease locus (rs334) in Kenyan children (cases, 180; controls, 2,677): P = 0.072, OR = 1.35 (95% CI 0.96-1.85). As previously reported (Williams et al., 2009), homozygosity at the sickle cell locus, is associated with increased risk of NTS disease: P (recessive model) = 4.41×10^{-5}, OR = 5.03 (95% CI 2.19-10.53). In contrast to previous observations however (Williams et al., 2009), under a heterozygote advantage model, we found no evidence for protection against NTS bacteraemia in HbAS individuals: P = 0.46, OR = 0.85 (95% CI 0.54-1.25).

In a multinomial regression model of bacteraemia risk at the sickle cell locus, in keeping with the data for NTS bacteraemia alone, the data best supports an increased risk of NTS, *S. pneumoniae*, and *H. influenza* type b bacteraemia with HbS carriage, under an additive model of inheritance (Bayes factor c.f. null = 542, Figure 3.24A), albeit with marginal effects in each associated stratum. Under a recessive model of inheritance, the data best support a non-zero effect in bacteraemia secondary to NTS, *S. pneumoniae*, *H. influenza*

type b, and *Acinetobacter* species; with increased risk of bacteraemia in children with HbSS (Bayes factor c.f. null = 7.07×10^{12}, Figure 3.24B). Under the heterozygous advantage model, the data provide limited support for a model in which children with HbAS are protected against bacteraemia secondary to *S. pneumoniae*, *E. coli*, and *Acinetobacter* species (Bayes factor c.f. null = 15, Figure 3.24C).

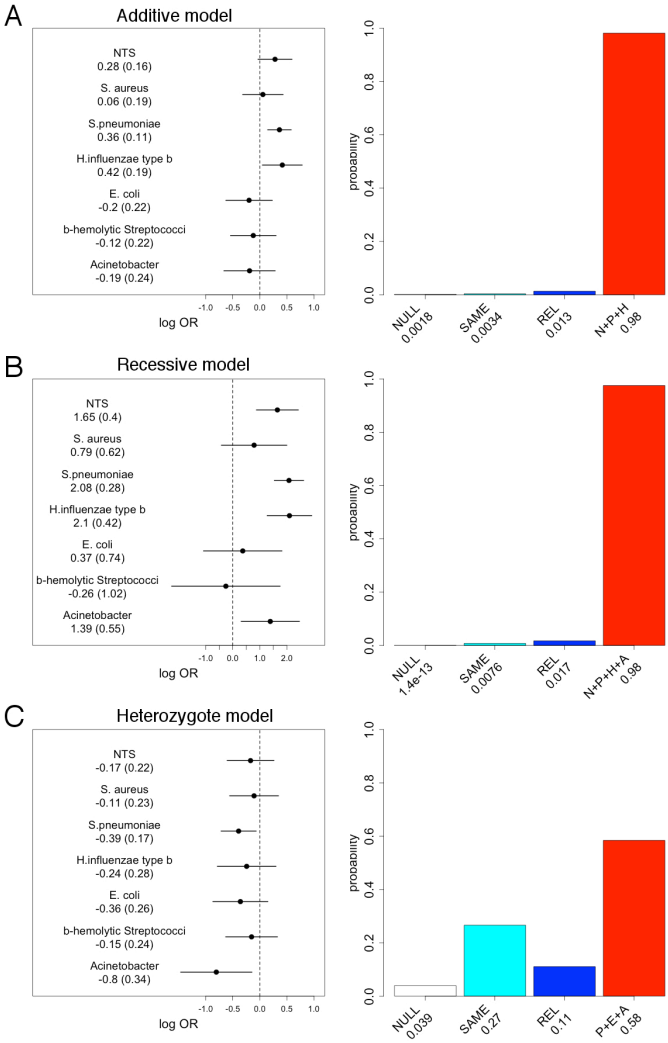

Figure 3.24 Sickle cell locus association with major bacterial pathogens in Kenyan children.

Log-transformed odds ratios and 95% confidence intervals of rs334 association, under additive (A), recessive (B) and heterozygous advantage models (C) in Kenyan discovery samples; left panels. Posterior probabilities of models of association at rs334: NULL, no association with any bacterial pathogen; SAME, the same effect across all bacterial pathogens; REL, related effects across all bacterial pathogens; N+P+H, N+P+H+A, and P+E+A, a non-zero effect in the specified pathogens alone, where N = NTS, P = *S. pneumoniae*, H = *H. influenza*, E = *E. coli*, A = *Acinetobacter*.

Glucose-6-phosphate dehydrogenase deficiency

Genotyping of rs1050828 in the Kenyan discovery samples passes SNP QC thresholds (with the exception of extreme deviation from HWE [$P = 5.8 \times 10^{-224}$], as is expected for a non-autosomal locus). The genotyping clusters are not perfectly distinct between major homozygote and heterozygote calls, but are sufficiently differentiated to permit an exploratory analysis (Figure 3.25). G6PD deficiency is common in this population, with 11.9% having homozygous or hemizygous deficiency, and 32.9% of girls being heterozygous for rs1050828:T carriage.

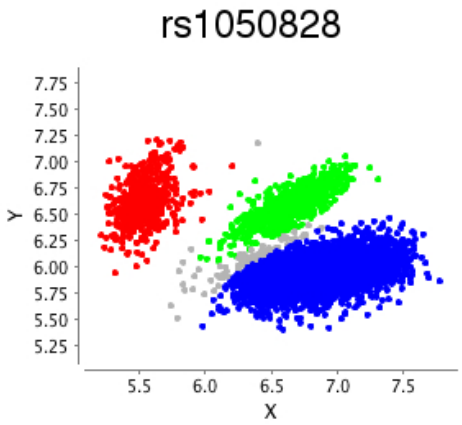

Figure 3.25 Cluster plot of the G6PD deficiency locus in Kenyan discovery samples.

Individuals called as minor homozygotes, heterozygotes, and major homozygotes are coloured red, green, and blue respectively.

Under an additive model, there is no evidence of association between NTS bacteraemia and the G6PD deficiency locus (rs1050828) in Kenyan children (cases, 180; controls, 2,677): $P = 0.846$, OR = 0.98 (95% CI 0.57-1.22). Nor is there evidence for modified bacteraemia risk in girls heterozygous at

rs1050828 (*P* = 0.722; OR = 1.08, 95% CI 0.71-1.59), or in G6PD deficient boys and girls (*P* = 0.712; OR = 0.91, 95% CI 0.53-1.47). Similarly, in an additive model, rs1050282 is not associated risk of all-cause bacteraemia in the same population of Kenyan children (cases, 1,526; controls, 2,677): *P* = 0.757, OR = 1.01 (95% CI 0.93-1.11). However, while G6PD deficiency does not modify all-cause bacteraemia risk (*P* = 0.171; OR = 1.15, 95% CI 0.94-1.39), there is modest evidence that heterozygous carriage of the rs1050828:T allele protects against bacteraemia (*P* = 0.014; OR = 0.80, 95% CI 0.67-0.95).

In a multinomial regression model of bacteraemia risk at rs1050828 (NTS cases, 180; *S. aureus*, 177; *S. pneumonia*, 429; *H. influenzae*, 133; *E. coli*, 159; β-haemolytic *Streptococcus*, 158; *Acinetobacter*, 130; controls, 2,677), in keeping with the all-cause bacteraemia analysis, rs1050828 is not associated with bacteraemia secondary to any pathogen under an additive model of inheritance (Figure 3.26A). By contrast, in a model of bacteraemia risk of females heterozygous at rs1050828, the data provides support for a model in which heterozygous rs1050828:T carriage protects against pneumococcal bacteraemia alone (Figure 3.26B, Bayes factor = 3.7). In addition, in a model of bacteraemia risk of G6PD deficiency risk (i.e. rs1050828:T hemizygous boys and homozygous girls), the data provides modest support for a model in which G6PD deficiency increases risk of pneumococcal bacteraemia alone (Figure 3.26C, Bayes factor = 2.3).

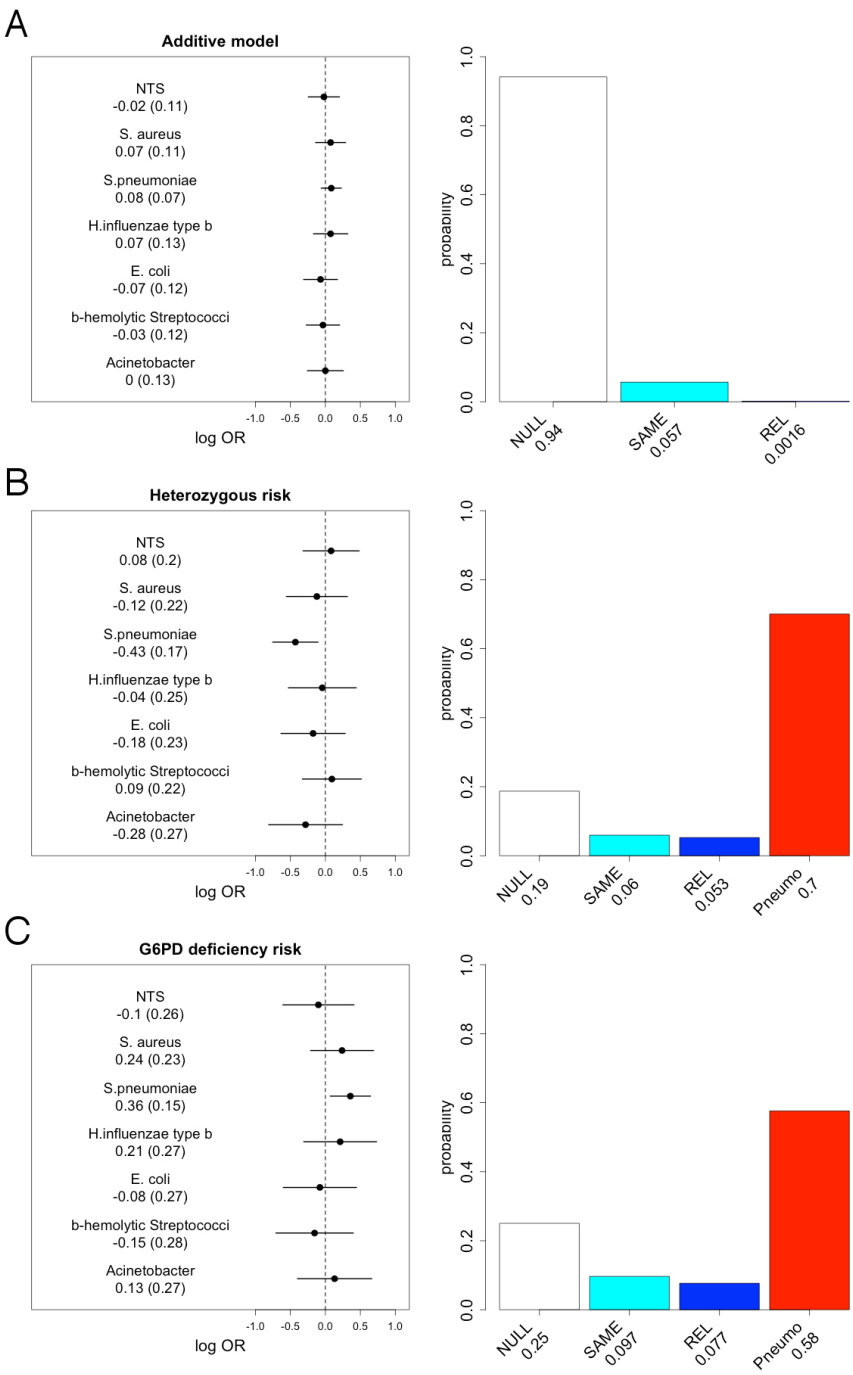

Figure 3.26 Glucose-6-phosphate dehydrogenase deficiency association with major bacterial pathogens in Kenyan children.

Log-transformed odds ratios and 95% confidence intervals of rs1050828 association, under models of additive risk (A), heterozygous risk (B) and G6PD deficiency risk (C) in Kenyan discovery samples; left panels. Posterior probabilities of models of association at rs1050828: NULL, no association with any bacterial pathogen; SAME, the same effect across all bacterial pathogens; REL, related effects across all bacterial pathogens; Pneumo, a non-zero effect for pneumococcal bacteraemia alone.

Conclusions

In performing, to our knowledge, the first genome-wide association study of invasive NTS disease, we demonstrate that genetic variation in *STAT4* is a determinant of NTS bacteremia in African children. The NTS-associated SNP (rs13390936) is located in the third intron of *STAT4*, and is associated with disease under a recessive model. The effect size of homozygous carriage of rs13390936:T is substantial, with the combined odds ratio for NTS bacteraemia in Kenyan and Malawian children being 7.6. In the Kenyan discovery population, this odds ratio is of the order of the effect observed for sickle cell disease on NTS bacteraemia risk (OR = 5.0).

STAT4 is a member of the STAT family of transcription factors (Watford et al., 2004). In NK cells and $CD4^+$ T cells, STAT4 phosphorylation in response to IL-12 signaling results in T_H1 differentiation and IFNγ production. 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 (Gilchrist et al., 2015; Jong, 1998). 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 immunity (Figure 3.27).

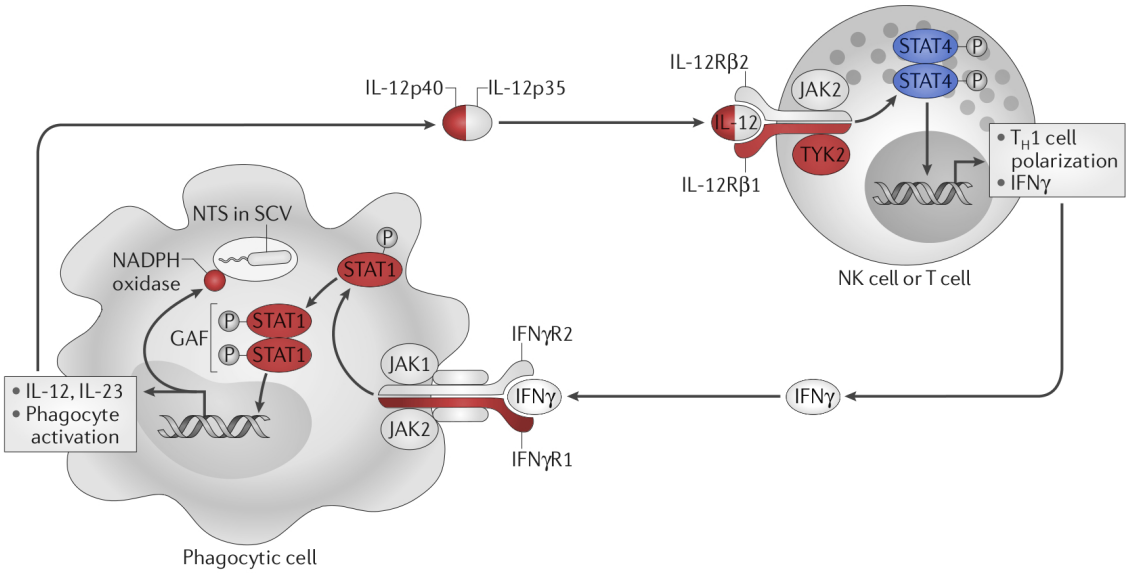

Figure 3.27 The role of STAT4 in the control of intracellular *Salmonella* infection.

Internalization of NTS by a phagocyte, in a *Salmonella*-containing vacuole (SCV), results in interleukin (IL)-12 release. IL-12 signals via the IL-12 receptor complex on NK cells and T cells leading to the phosphorylation and activation of STAT4 (highlighted in blue). Activated STAT4 upregulates IFNγ production from both T cells and NK cells, and results in T_H1 polarization. Released IFNγ activates infected phagocytes via the IFNγ receptor, resulting in the phosphorylation and homodimerization of STAT1 to form γ-activating factor (GAF), which upregulates anti-bacterial effector mechanisms. Genes implicated in Mendelian Susceptibility to Mycobacterial Disease and resulting in susceptibility to disseminated NTS infection are highlighted in red. Figure modified from(Gilchrist et al., 2015). First published in Nature Reviews Immunology, volume 15, pages 452-63, by Nature Publishing Group.

The role of STAT4 in host immunity to NTS infection is further supported by the enhanced susceptibility of Stat4-deficient mice to systemic NTS infection (Eva et al., 2013). The identification of genetic variation affecting IL-12-dependent IFNγ-mediated immunity as a risk factor for NTS bacteremia highlights the common biological determinants of susceptibility to infectious diseases at the population level and in individuals with rare primary immunodeficiencies.

Despite high quality imputation of class I and II HLA alleles, we did not identify robust evidence for carriage of any HLA allele modifying risk of NTS bacteraemia, or bacteraemia secondary to other common pathogens in Kenyan children. Nor did we identify any evidence that HLA allele 4-digit homozygosity is associated with risk of disease. In the analysis of the Kenyan discovery samples, we do identify modest evidence of association between carriage of both HLA-C*17:01 and HLA-DRB1*11:01 and risk of NTS disease, but without robust evidence for replication in the Kenyan replication samples. It is however noteworthy that there is a trend towards replication of the DRB1*11:01 association in the Kenyan replication samples. While this could not be considered to provide convincing evidence of replication, it should be noted that the replication analysis included only 31 case samples, and this association clearly warrants more rigorous assessment in larger sample collections.

Within the NTS bacteraemia GWAS we analyse two further well-characterised loci associated with infectious diseases in African children, the sickle cell disease locus, and G6PD deficiency. We confirm sickle cell homozygosity as a significant risk factor for bacteraemia secondary to NTS, *S. pneumoniae* and *H. influenzae* type b, and heterozygosity as being protective against bacteraemia secondary to *S. pneumoniae, E. coli,* and *Acinetobacter.* These observations provide no additional insights into the interaction of HbS with common causes of bacteraemia in this population beyond those already defined in previously-published studies (Williams et al., 2009); studies which contain a significant degree of sample overlap with the GWAS presented

here. They do however provide an informative positive control, demonstrating the feasibility of performing a genome-wide association analysis using imputed genotypes in the manner described in this population. Analysis of the G6PD deficiency locus in this collection provides no evidence for an association of G6PD deficiency with NTS bacteraemia. We do, however, identify evidence of increased risk of pneumococcal bacteraemia in G6PD deficient patients, and a protective effect of heterozygous rs1050828:T carriage against pneumococcal bacteraemia. The identification of G6PD heterozygous carriage decreasing risk of pneumococcal bacteraemia is in keeping with recent data in the same Kenyan population describing a reduction in malaria risk in G6PD heterozygous girls (Uyoga et al., 2015), malaria being a well-established risk factor for community-acquired bacteraemia in this context (Scott et al., 2011). The identification of G6PD deficiency as a risk factor for pneumococcal bacteraemia is interesting, and may reflect the deleterious effects of haemolysis on oxidative burst in neutrophils (Cunnington, de Souza, Walther, & Riley, 2012a).

Taken together, data presented in this chapter describe, to our knowledge, the first GWAS of NTS bacteraemia. This analysis identifies robust evidence for an established, sickle cell disease, and a novel genetic correlate of invasive NTS disease in *STAT4*.

Chapter 4 – Functional basis for the nontyphoidal *Salmonella* association at *STAT4*

Background

In identifying genetic variation in *STAT4* as risk factor for invasive NTS disease in African children, our data highlight the key role of IFNγ-mediated immunity in host control of *Salmonella* infection in unselected populations of African children. To better understand the biological consequences of the NTS-associated genetic variation in STAT4, we designed a series of experiments to assess the consequences of carriage of the rs13390936:TT genotype at the RNA and protein levels.

GWAS-identified trait-associated loci are enriched for regulatory DNA elements (Maurano et al., 2012). Given rs13390936 is non-coding, we explored whether it is associated with gene expression. Mapping regulatory genetic variants associated with gene expression, using expression quantitative trait locus (eQTL) studies (Emilsson et al., 2008), is often

informative but has demonstrated that regulatory variation can be specific to cell-type or context (Fairfax et al., 2014). This highlights the need for experimentation in biologically relevant cell-types and conditions. To address this we analysed eQTL data from previously published and unpublished datasets (Fairfax et al., 2014; 2012; Naranbhai et al., 2015) of naïve and stimulated primary immune cells, assessing the effect of rs13390936 genotype on *STAT4* RNA expression.

A major consequence of STAT4 activity, in particular in the context of invasive *Salmonella* infection, is IFNγ production in NK cells and $CD4^+$ T cells. To better understand the NTS-associated phenotype of rs13390936 at the protein level in these cells, we recruited healthy European adults of known genotype at rs13390936 via a genotype-selectable bioresource (Oxford Biobank), and assayed STAT4 phosphorylation and IFNγ production in NK and CD4 T^+ cell subsets following *ex vivo* stimulation. As regulatory genetic variation can demonstrate specificity to human populations (Raj et al., 2014), we further characterised the effect of rs13390936:T carriage on total serum IFNγ protein production in African children with acute invasive NTS infection.

The identification of genetic variation in *STAT4* as a risk factor for invasive NTS infection provides an informative example of the shared biology underlying susceptibility to a single pathogen in patients with rare primary immunodeficiencies and individuals in unselected populations. Rare individuals with Mendelian Susceptibility to Mycobacterial Disease (MSMD),

with deleterious mutations in genes affecting IL-12-dependent IFNγ–mediated immunity other than *STAT4*, are highly susceptible to invasive NTS infections. Given the striking specificity of the pathogen susceptibility seen in individuals with MSMD, we investigated whether the NTS-associated genetic variant results in similar pathogen specificity, comparing the evidence for rs13390936 disease association across commonly identified pathogens in the all-cause bacteraemia GWAS in the same population of Kenyan children (Kenyan Bacteraemia Study Group et al., 2016).

Previously published genetic association studies have implicated variation at other loci within *STAT4* in the pathogenesis of several autoimmune diseases (Eyre et al., 2012; Liu et al., 2012; Trynka et al., 2011). To assess the evidence for any shared genetic aetiology between NTS bacteraemia and autoimmune disease at rs13390936, we used publically-available summary statistics in the ImmunoBase database (Onengut-Gumuscu et al., 2015), comparing evidence for association at rs13390936 with eight autoimmune diseases.

In highlighting the role of genetic determinants of IFNγ-mediated immunity in invasive NTS disease in African children, our data identifies overlap between *Salmonella* susceptibility in individuals with rare primary immunodeficiency (MSMD) and unselected populations. In order to assess whether common genetic variation at MSMD loci could also act as a determinant of NTS immunity in unselected populations, we performed a candidate gene analysis

of NTS bacteraemia at the eight autosomal loci, and two sex-linked loci, identified as causes of MSMD.

Methods

RNA expression quantitative trait analysis of rs13390936

Study samples, expression and genotyping data

Using RNA expression and genome-wide genotyping data from published and unpublished expression quantitative trait locus (eQTL) studies of naïve and stimulated primary immune cell subsets in healthy European adults, we correlated rs13390936 genotype with STAT4 RNA expression. As previously described, (Fairfax et al., 2012; 2014; Naranbhai et al., 2015) $CD19^+$ B cells, $CD14^+$ monocytes and $CD56^+CD3^-$ NK cells were separated from peripheral blood mononuclear cells (PBMCs) by magnetic activating cell sorting (MACS - Miltenyi). $CD16^+$ neutrophils were isolated from granulocytes with $CD16^+$ microbeads. Gene expression was quantified in total RNA from naïve cells, and monocytes stimulated with lipopolysaccharide (for 2 or 24 hours) or IFNγ (for 24 hours), 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 phase1 as the reference panel.

Statistical analysis

rs13390936 genotype was correlated with STAT4 RNA expression in each cell type: B cells (n=279), NK cells (n=245), neutrophils (n=101), naïve monocytes (n=414), and stimulated monocytes (LPS 2 hours, n=261; LPS 24 hours, n=322; IFNγ 24 hours, n=367). Normalized *STAT4* RNA expression was correlated with genotype under by linear regression and analysis of variance (ANOVA), including the first 25 principal components of gene expression data in each cell-type/condition to account for confounding variation. P-values are calculated with *F*-tests (1 d.f.). Statistical analysis was performed in R.

Protein phenotypes of rs13390936 in immune cell subsets

Study subjects

In archived genomic DNA from 5,911 individuals registered to the Oxford Biobank (NIHR Oxford Biomedical Research Centre), rs13390936 was genotyped by the Oxford Biobank laboratory using a TaqMan assay (Applied Biosystems). 54 volunteers (32 female; median age 44 years, range 30-51 years) were recruited to the study according to rs13390936 genotype (36 with 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.

Ex vivo whole blood stimulations

Whole blood from each volunteer was collected into Lithium Heparin-containing collection tubes (Vacutainer System, Becton Dickinson). 1ml aliquots of whole blood from each volunteer were left unstimulated, or stimulated with 10ng/ml recombinant human IL-12 (catalog# 219-IL, R&D systems) or 10^6cfu/ml NTS (Malawian clinical *S*. Typhimurium isolate D23580) grown to mid-log phase, within 2 hours of collection. In addition, 1ml whole blood aliquots were stimulated with phorbol 12-myristate 13-acetate (PMA - 100ng/ml)/ionomycin (1µg/ml - Sigma) to act as a positive control. Following stimulation samples were incubated at 37°C.

Whole blood intracellular IFNγ staining

Brefeldin A (Becton Dickinson) was added at 2 hours post-stimulation to a final concentration of 2µg/ml. At 6 hours post-stimulation, samples were surface immunostained (20 minutes, 4°C) following erythrocyte lysis (Versalyse, Beckman Coulter) with allophycocyanin (APC)-conjugated anti-CD3 (clone UCHT1, BD Biosciences), fluorescein isothiocyanate (FITC)-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 permeabilisation with Permeabilization Medium B (Invitrogen) and intracellular staining with PE-cyanine 7 (PE-Cy7)-conjugated anti-IFNγ monoclonal (clone B27, BD Biosciences) for 20 minutes at 4°C.

Whole blood intracellular pSTAT4 staining

For intracellular pSTAT4 staining, at 15 minutes post-stimulation, erythrocytes were lysed and the samples fixed with BD Phosflow Lyse/Fix buffer (BD Biosciences). Samples were surface immunostained (20 minutes, 4°C) with APC-conjugated anti-CD3 (clone UCHT1, BD Biosciences), FITC-conjugated anti-CD4 (clone RPA-T4, BD Biosciences) and PE-Cy7-conjugated anti-CD7 (clone M-T701, BD Biosciences) monoclonals. Samples were methanol permeabilized on ice with BD Phosflow Perm Buffer III (BD Biosciences) for 30 minutes, followed by intracellular immunostaining (60 minutes, room temperature) with PE-conjugated anti-STAT4 pY693 (clone p38/pStat4, BD Biosciences).

Data acquisition and statistical analysis

Immunostained samples were acquired with a BD FACSCanto II and BD FACSDiVa (BD Biosciences), and the data analyzed using FlowJo v10 (TreeStar). 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. To minimise variation across batches, all

antibodies were obtained prior to volunteer recruitment and belonged to the same manufacturing lot number. Gating strategies to identify NK and $CD4^+$ T cells in IFNγ and pSTAT-stained cells are described in Figures 4.1 and 4.2 respectively.

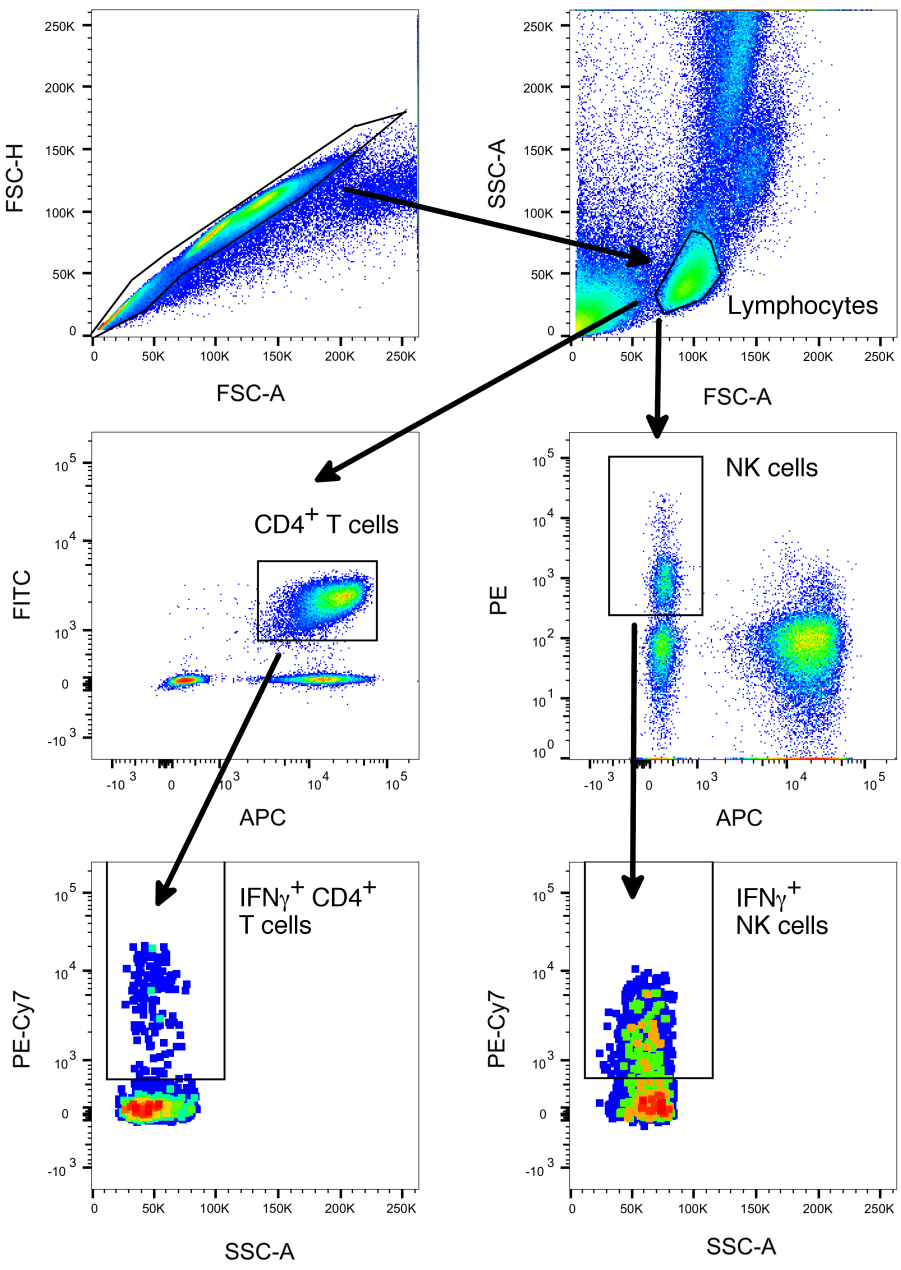

Figure 4.1 Gating strategy for IFNγ stained cells.

Anti-CD3, CD4, CD56, and IFNγ monoclonals are conjugated to APC, FITC, PE, and PE-Cy7 respectively. $CD4^+$ T cells are defined as $CD3^+CD4^+$ lymphocytes; NK cells are defined as $CD3^-CD56^+$ lymphocytes.

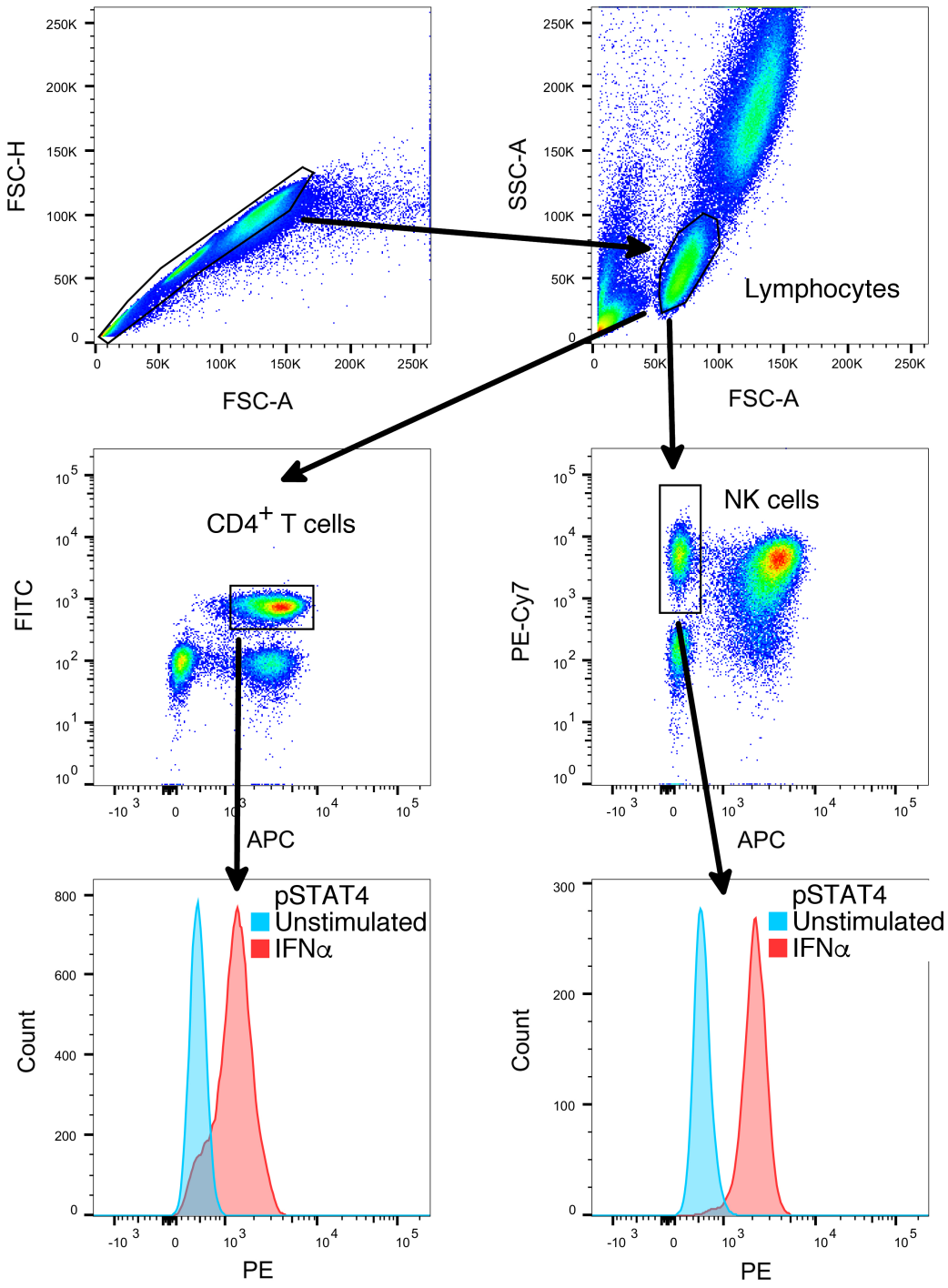

Figure 4.2 Gating strategy for phospho-STAT4 stained cells.

Anti-CD3, CD4, CD7, and pSTAT4 monoclonals are conjugated to APC, FITC, PE-Cy7, and PE respectively. $CD4^+$ T cells are defined as $CD3^+CD4^+$ lymphocytes; NK cells are defined as $CD3^-CD7^+$ lymphocytes.

Proportions of IFNγ$^+$ NK cells were logit transformed prior to analysis, and genotype-phenotype associations tested with linear regression and ANOVA, adjusted for age and sex. P-values are calculated with *F*-tests (1 d.f.). Transformation did not normalise IFNγ$^+$ CD4$^+$ T cell proportions, and genotype-phenotype associations were tested with nonparametric Wilcoxon rank-sum tests. There was no evidence of significant batch effect ($p<0.05$) on IFNγ$^+$ cell proportion for either NK cells (linear regression model, adjusted for rs13390936 genotype and stimulation conditions) or CD4$^+$ T cells (Kruskal-Wallis test).

In the case of pSTAT4 immunostaining results, geometric mean fluorescence intensities (gMFI) were log-transformed prior to analysis, and genotype-phenotype associations tested with a linear regression model. There was evidence of significant batch effect, with two (of 12) batches significantly ($P<0.05$) associated with pSTAT4 gMFI in NK cells, and four in CD4$^+$ T cells. All batches were included in the analysis and regression models adjusted for batch membership, along with age and sex. Statistical analysis was performed in R.

IFNγ protein phenotype of rs13390936 in acute NTS bacteraemia

Study subjects

Serum samples were collected from a subset of the Malawian replication cases (n = 106). Methods used for bacterial culture and NTS identification and

serotyping are as described for the GWAS Malawian replication samples. 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 (Abbot Laboratories) and UniGold (Trinity Biotech) rapid tests, and HIV infection was confirmed in children less than 18 months by PCR. Children with weight-for-age z-scores greater than 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.

Serum IFNγ quantification

Sera were separated from clotted whole blood within 2 hours of venesection, and stored at -80°C prior to analysis. Serum IFNγ concentrations were assayed using a Bio-Plex Pro 27-plex human cytokine fluorescent bead–based assay (Bio-Rad Laboratories) according to the manufacturer's instructions on a Luminex-100 instrument (BioRad Laboratories) using Bio-Plex Manager 4.1.1 software (BioRad). Assays were performed by Dr Jenny

Heath at the University of Birmingham, and were run in three batches with rs13390936 genotypes randomized across batches. IFNγ measurements below the detection limit of the assay were assigned values of the half of the lower detection limit, and were included in the analysis.

Statistical analysis

Correlation of rs13390936 genotype with serum IFNγ concentration during acute NTS bacteraemia was performed by linear regression and ANOVA, adjusting for age, sex, and NTS-associated co-morbidities (HIV, malnutrition and malaria). Serum IFNγ concentrations were cube-root transformed prior to analysis. P-values are calculated with *F*-tests (1 d.f.). Statistical analysis was performed in R.

Models of association at rs13390936 in infectious and autoimmune diseases

Major bacterial pathogens in Kenyan children

We compared models of association at rs13390936 across the six most frequently isolated bacterial pathogens among cases of bacteraemia in the Kenyan discovery samples in the all-cause bacteraemia GWAS (Kenyan Bacteraemia Study Group et al., 2016). As described previously, multinomial logistic regression models were calculated for association at rs13390936

under a recessive model, with control status and each of the six most common bacterial pathogens as strata. We considered four 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(0,1) and fixed between pathogens (ρ=1).
REL: effect size ~ N(0,1) and correlated (ρ=0.96), but not fixed, between pathogens.
NTS: effect size ~ N(0,1) for NTS and is zero for all other pathogens.

For each model we calculated approximate Bayes factors (Wakefield, 2009) and posterior probabilities, assuming each model to be equally likely *a priori*. Statistical analysis was performed in R.

Autoimmune diseases in populations of European ancestry

To compare models of association at rs13390936 across a range of autoimmune diseases, we downloaded from ImmunoBase (Onengut-Gumuscu et al., 2015) summary statistics of disease associations at rs16833239 (in perfect linkage disequilibrium with rs13390936 in CEU 1000 Genomes Pilot 1 population; r^2 = 1, D' = 1) for nine association studies of eight autoimmune diseases (celiac disease, rheumatoid arthritis, primary biliary cirrhosis, multiple sclerosis, narcolepsy, juvenile idiopathic arthritis, psoriasis and type 1 diabetes – one case-control study and one meta-analysis

of case-control and family-based studies) conducted using the ImmunoChip (Cortes & Brown, 2011) in individuals of European ancestry (Eyre et al., 2012; Faraco et al., 2013; Hinks et al., 2013; International Multiple Sclerosis Genetics Consortium (IMSGC) et al., 2013; Liu et al., 2012; Onengut-Gumuscu et al., 2015; Trynka et al., 2011; Tsoi et al., 2012). Standard errors were estimated from reported odds ratios and p-values calculated under an additive model (Fig. 3B). For each combination of disease associations at rs16833239 we calculated an approximate Bayes factor (Wakefield, 2009) against the null model of no association in any disease, assuming a prior with a mean log odds ratio of 0, and a standard error of 0.2. We assumed effects between autoimmune diseases to be correlated but not fixed (ρ=0.1). Marginal probabilities of association at rs16833239 for each autoimmune disease were calculated as the sum of posterior probabilities for each model in which the disease is associated with rs16833239. Statistical analysis was performed in R using code developed by Holly Trochet, Wellcome Trust Centre for Human Genetics.

Candidate gene analysis of Mendelian susceptibility to mycobacterial disease loci

Genetic case-control study of NTS bacteraemia risk at Mendelian susceptibility to mycobacterial disease loci

To evaluate the evidence for modification of risk of NTS bacteraemia at MSMD loci, we extracted association statistics for NTS bacteraemia from the

Kenyan GWAS at the ten described MSMD loci. We included association tests at any SNP within 10kb of a known MSMD-associated gene, passing QC thresholds employed in the GWAS, and with a minor allele frequency of at least 5%. At the eight autosomal loci (*IFNGR1, IFNGR2, STAT1, IRF8, ISG15, IL12B, IL12RB1,* and *TYK2*) we included association statistics calculated under additive, dominant, recessive and heterozygous advantage models. At the two X chromosome loci (*CYBB* and *IKBKG*) we included association statistics calculated under an additive model, and statistics calculated under an additive model stratified by sex. The whole X chromosome was imputed in IMPUTE2 (Howie et al., 2012), following pre-phasing in SHAPEIT (Delaneau et al., 2012), using 1000 Genomes Project Phase 1 release as a reference panel. Association analysis, accounting for imputation uncertainty, was performed in SNPTEST (Marchini et al., 2007) using the frequentist score test.

We considered association at $P < 1x10^{-4}$ as suggestive evidence of association in the discovery analysis. Replication at associated loci was performed in the Kenyan replication sample collection. Genome-wide imputation of Immunochip-genotyped samples was undertaken in IMPUTE2 (Howie et al., 2012), following pre-phasing in SHAPEIT (Delaneau et al., 2012), using 1000 Genomes Project Phase 3 release as a reference panel.

Enrichment analysis at Mendelian susceptibility to mycobacterial disease loci

To evaluate the evidence for enrichment of NTS-associated genetic variation at MSMD loci beyond single variant association tests, we used a permutation-based approach to assess for evidence of global enrichment of NTS association at MSMD loci. We included 1,124 SNPs in the analysis, with SNPs being included if they were within 10kb of an MSMD-associated gene, well-imputed (info>0.8), common (minor allele frequency >5%), and passing HWE thresholds ($P<1x10^{-10}$, except X chromosomal loci). Samples were included according to GWAS criteria detailed previously. Case-control status was permuted 10,000 times and association testing performed for each permutation under an additive model. An empirical enrichment *P*-value is then derived as the proportion of randomly permuted associated test sets in which the summed Z-scores are more extreme than that observed for the NTS case-control status. Statistical analysis was performed in R.

Results

RNA expression quantitative trait analysis of rs13390936

In naïve immune cells, rs13390936 genotype was not significantly associated with *STAT4* RNA expression, which was most abundant in NK cells (Figure 4.3). By contrast, rs13390936 was a significantly associated with *STAT4* RNA expression in monocytes following stimulation with lipopolysaccharide for 2 (P=4·52x10^{-5}) or 24 hours (P=2·65x10^{-5}), or IFNγ for 24 hours (P=2·63x10^{-5}), with carriers of the NTS-risk genotype expressing the least *STAT4* in each stimulation condition (Figure 4.4).

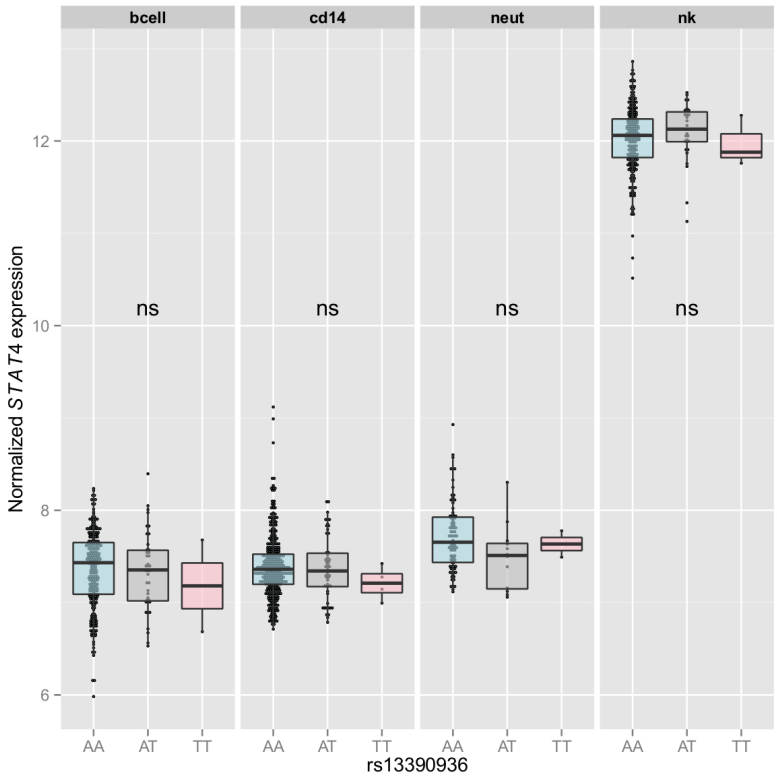

Figure 4.3 The effect of rs13390936 genotype on *STAT4* RNA expression in unstimulated leukocytes.

Genotype-phenotype correlations are analysed by regression and analysis of variance. *P*-values are calculated by *F*-tests with one degree of freedom. NS, not significant.

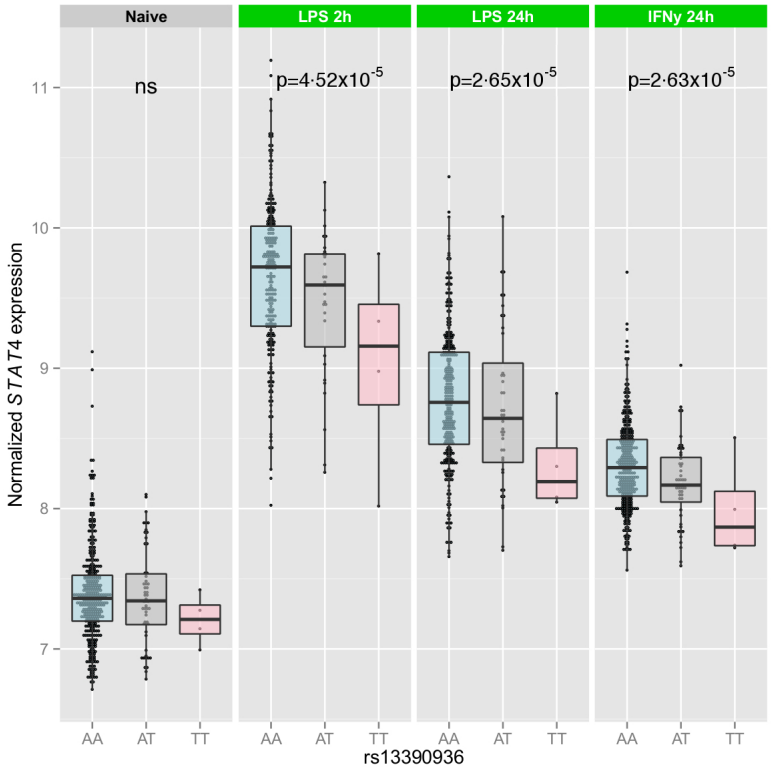

Figure 4.4 The effect of rs13390936 genotype on *STAT4* RNA expression in stimulated monocytes.

Genotype-phenotype correlations are analysed by regression and analysis of variance. *P*-values are calculated by *F*-tests with one degree of freedom. NS, not significant.

IFNγ protein phenotype of rs13390936 in immune cell subsets

In 54 healthy European adults of known genotype at rs13390936 (36 donors carrying the NTS-protective AA genotype; 18 carrying the NTS-susceptible TT genotype), we quantified by flow cytometry, STAT4 phosphorylation and IFNγ production in NK cells and $CD4^+$ T cells, in response to stimulation with IL-12 and NTS. STAT4 phosphorylation, following *ex vivo*, whole blood stimulation with IL-12 and NTS, is not associated with genotype at rs13390936 in either NK cells or $CD4^+$ T cells (Figure 4.5).

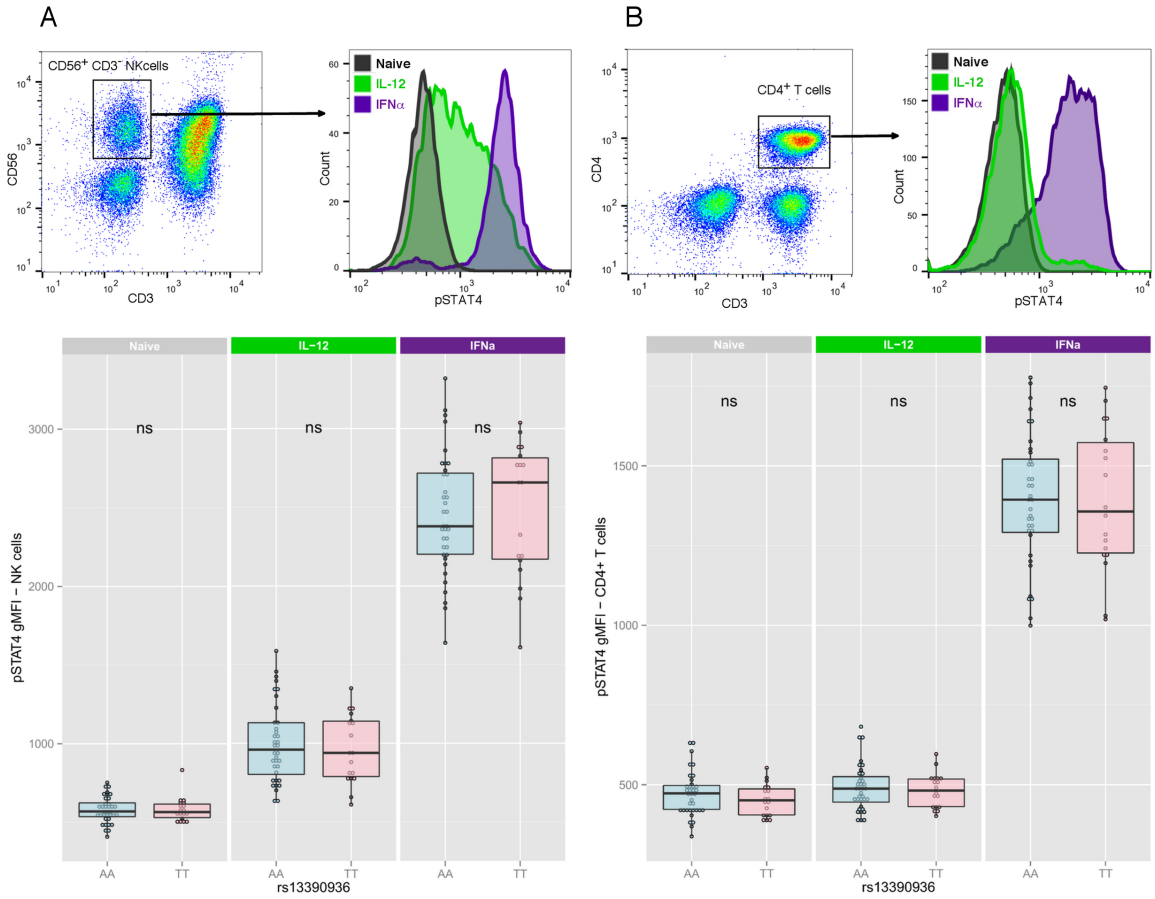

Figure 4.5 Effect of rs13390936 genotype on STAT4 phosphorylation in NK cells and CD4$^+$ T cells.

Gating strategy for intracellular immunostaining of pSTAT4 in NK cells (A - top), and genotype-specific pSTAT4 levels in naïve, IL-12 stimulated and IFNα-stimulated NK cells (A - bottom). Gating strategy for intracellular immunostaining of pSTAT4 in CD4$^+$ T cells (B - top), and genotype-specific pSTAT4 levels in naïve, IL-12 stimulated and IFNα-stimulated NK cells (B - bottom). ns, not significant; pSTAT4, phosphorylated STAT4; gMFI, geometric mean fluorescence intensity.

pSTAT4 geometric mean fluorescence intensity (gMFI) in NK cells is not modified by rs13390936 genotype in naïve ($mean_{AA}$ = 580, $mean_{TT}$ = 583, P = 0.278), IL-12 stimulated ($mean_{AA}$ = 1,000, $mean_{TT}$ = 967, P = 0.974) or IFNα-stimulated ($mean_{AA}$ = 2,450, $mean_{TT}$ = 2,484, P = 0.436) NK cells from

healthy European adults (Figure 4.5A). Similarly, pSTAT4 gMFI in $CD4^+$ T cells is not modified by rs13390936 genotype in naïve ($mean_{AA}$ = 471, $mean_{TT}$ = 452, *P* = 0.859), IL-12 stimulated ($mean_{AA}$ = 490, $mean_{TT}$ = 482, *P* = 0.064) or IFNα-stimulated ($mean_{AA}$ = 1,393, $mean_{TT}$ = 1,392, *P* = 0.375) cells from the same individuals (Figure 4.5B).

Despite the apparent lack of effect of rs13390936 genotype on levels of pSTAT4 following stimulation, donors with the NTS-risk rs13390936 genotype produced a reduced fraction of $IFN\gamma^+$ NK cells following stimulation with IL-12 (*P*=$1{\cdot}91x10^{-3}$, $mean_{AA}$ 1·54%, $mean_{TT}$ 0·91%) or NTS (*P*=0·049, $mean_{AA}$ 17·09%, $mean_{TT}$ 10·76%) (Figure 4.6). This effect appears to be context specific, with rs13390936 genotype having no effect on the numbers of $IFN\gamma^+$ NK cells in the absence of stimulation ($mean_{AA}$ 0·25%, $mean_{TT}$ 0·19%).

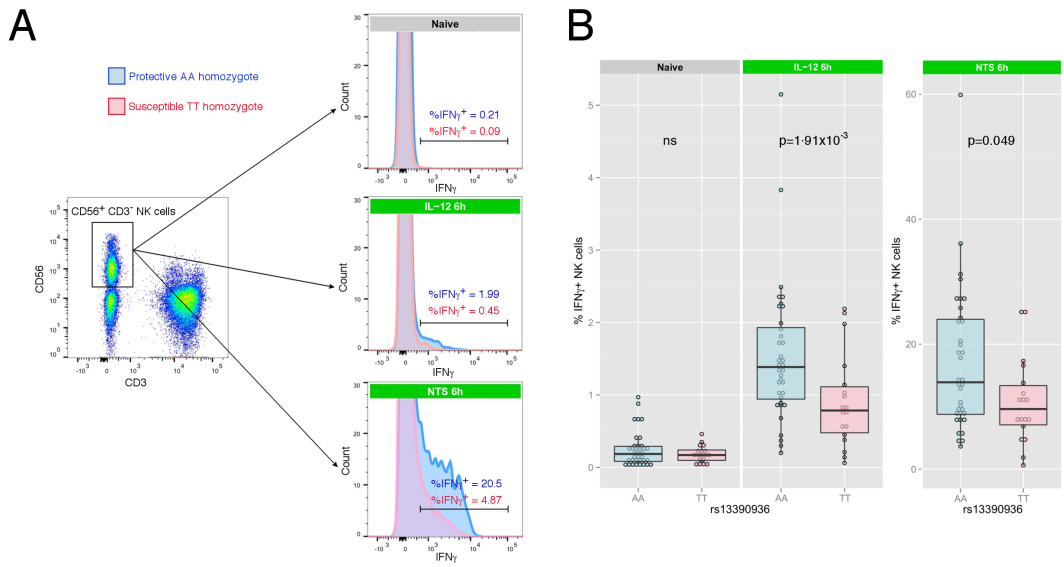

Figure 4.6 Effect of rs13390936 genotype on IFNγ production in NK cells and CD4$^+$ T cells.

Gating strategy for intracellular immunostaining of IFNγ in NK cells (A). Effect of rs13390936 genotype on numbers of IFNγ$^+$ NK cells in unstimulated cells, and cells following stimulation with IL-12 and NTS (B). Genotype-phenotype correlations are analysed by regression and analysis of variance. *P*-values are calculated by *F*-tests with one degree of freedom. NS, not significant.

We observed this genotype-dependent effect of rs13390936 on IFNγ production in NK cells but not in CD4$^+$ T cells. The proportion of IFNγ$^+$ CD4$^+$ T cells was not modified by rs13390936 genotype in naïve (mean$_{AA}$ 0·017%, mean$_{TT}$ 0·016%, p=0.941), IL-12 stimulated (mean$_{AA}$ 0.017%, mean$_{TT}$ 0.015%, p=0.727) or NTS-stimulated (mean$_{AA}$ 0.062%, mean$_{TT}$ 0.055%, p=0.132) CD4$^+$ T cells (Figure 4.7).

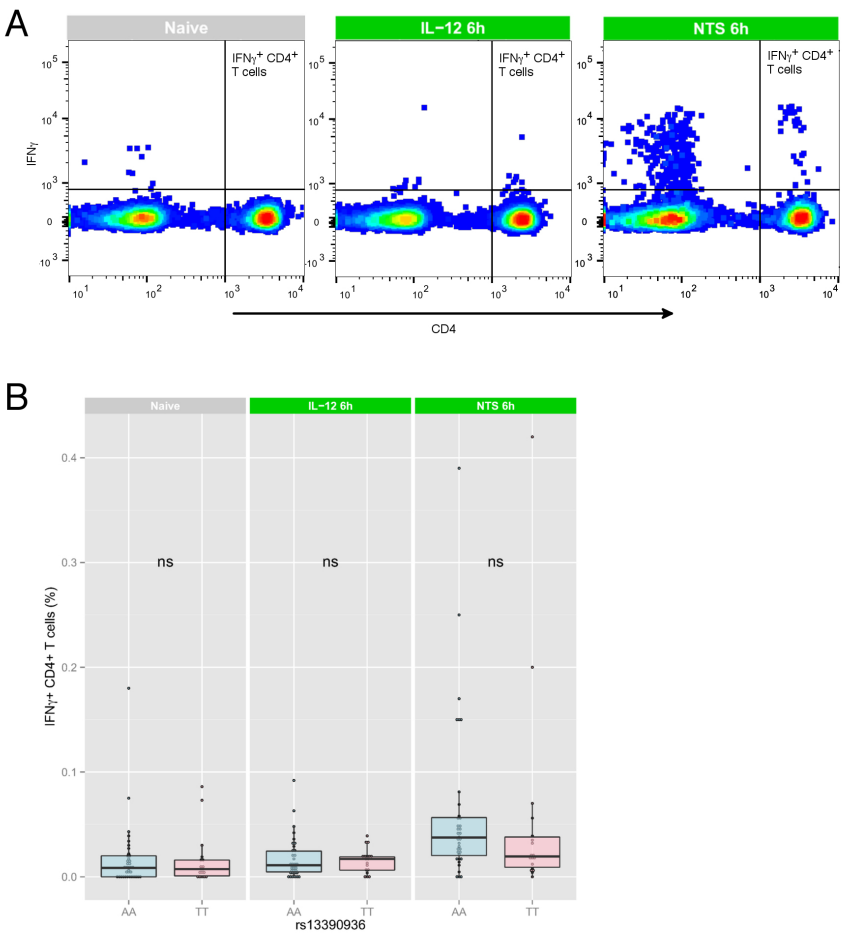

Figure 4.7 Effect of rs13390936 genotype on IFNγ production in CD4$^+$ T cells.

Gating strategy for IFNγ$^+$ CD4$^+$ T cells (A). Representative IFNγ production in naïve, IL-12- and NTS-stimulated CD4$^+$ T cells. Effect of rs13390936 genotype on numbers of IFNγ$^+$ CD4$^+$ T cells in unstimulated cells, and cells following stimulation with IL-12 and NTS (B). Genotype-phenotype correlations are analysed by regression and analysis of variance. *P*-values are calculated by *F*-tests with one degree of freedom. NS, not significant.

IFNγ protein phenotype of rs13390936 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 (Figure 4.8). 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 (p=0·016), and Malawian children carrying the NTS-risk genotype have reduced circulating IFNγ concentrations ($mean_{AA}$ 613·9pg/ml, $mean_{TT}$ 281·4pg/ml).

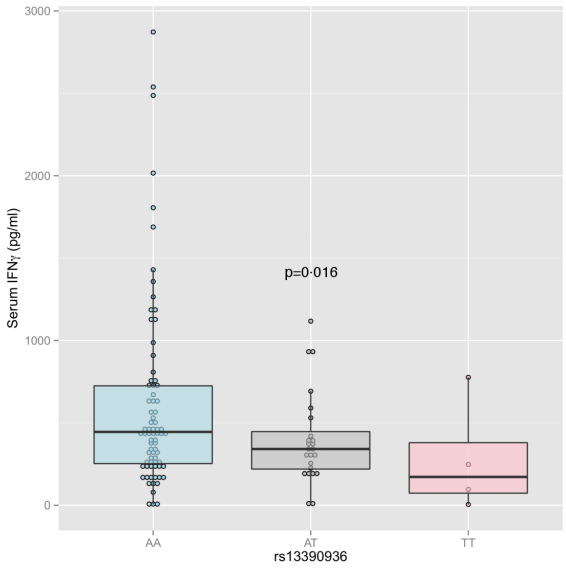

Figure 4.8 Effect of rs13390936 genotype on serum IFNγ levels in Malawian children with acute NTS bacteraemia.

Genotype-phenotype correlations are analysed by regression and analysis of variance. *P*-values are calculated by *F*-tests with one degree of freedom. NS, not significant.

Models of association at rs13390936 in infectious and autoimmune diseases

Multinomial regression model of rs13390936 effect on bacteraemia risk

We fitted a multinomial regression model of bacteraemia risk in the Kenyan discovery GWAS samples, with 2,651 control samples, and 1,222 cases (Streptococcus pneumoniae, 427 cases; NTS, 177 cases; Staphylococcus aureus, 172 cases; Escherichia coli, 157 cases; β-hemolytic Streptococci, 156 cases; Haemophilus influenzae type b, 133 cases) included in the analysis. Comparing possible models of rs13390936 association with bacterial pathogens with a Bayesian analysis, the data supports a model in which rs13390936 is associated with susceptibility to NTS bacteraemia but not bacteraemia caused by other pathogens (Figure 4.9, Bayes factors = 1,327).

In addition, we fitted a multinomial regression model of bacteraemia risk in the Kenyan replication samples, with 1,336 control samples and 161 cases (NTS, 38 cases; *S. aureus*, 45 cases; Haemophilus influenzae type b, 31 cases; Acinetobacter, 47 cases) at rs13390936. In that analysis, the replication data provides further support for a model in which NTS bacteraemia, but not bacteraemia secondary to other pathogens, is associated with rs13390936 genotype (Figure 4.9, Bayes factor = 12).

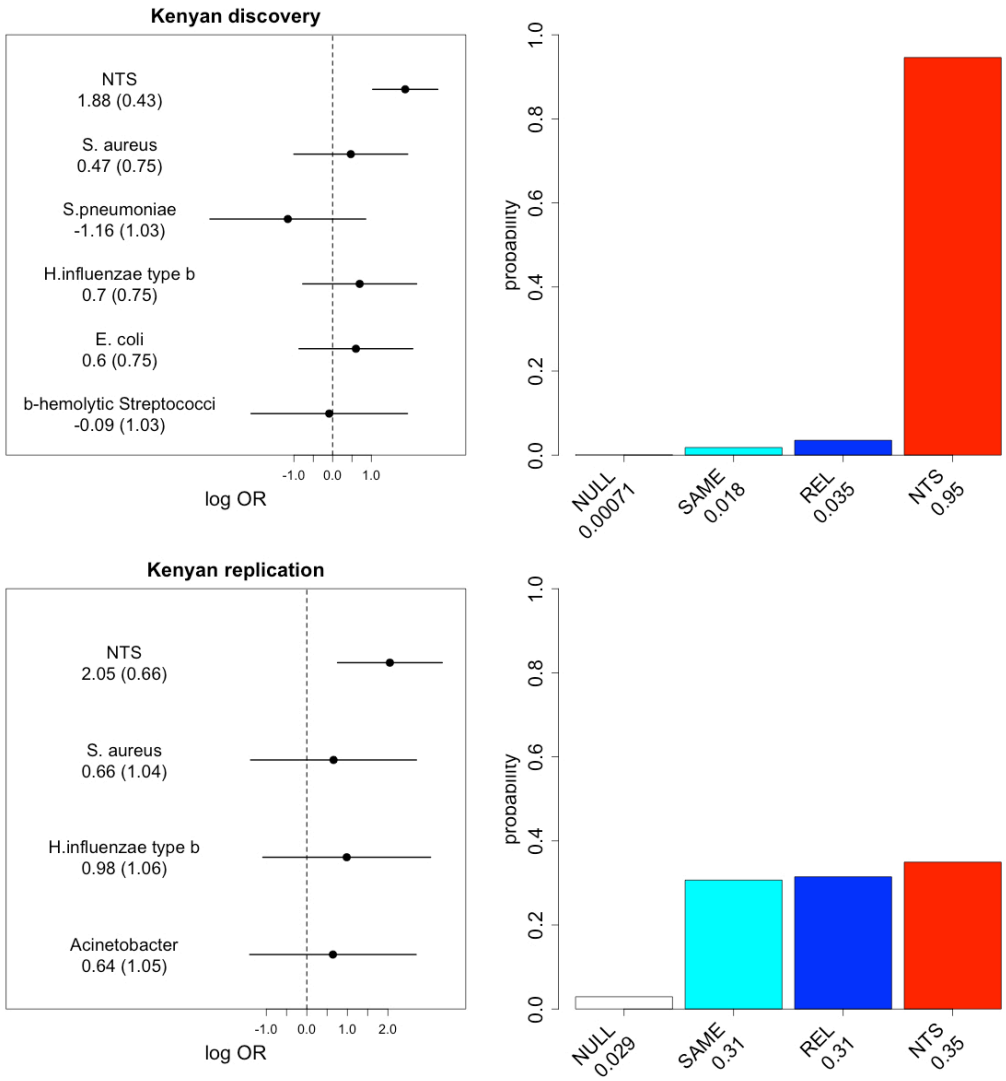

Figure 4.9 rs13390936 association with major bacterial pathogens in Kenyan children.

Log-transformed odds ratios and 95% confidence intervals of rs13390936 association under a recessive model in Kenyan discovery (top) and replication (bottom) samples; left panels. Posterior probabilities of models of association at rs13390936: NULL, no association with any bacterial pathogen; SAME, the same effect across all bacterial pathogens; REL, related effects across all bacterial pathogens; NTS, a non-zero effect for rs13390936 genotype in NTS alone.

Multinomial regression model of rs13390936 effect on autoimmune disease risk

Previously published genetic association studies have implicated variation at other loci within *STAT4* in the pathogenesis of several autoimmune diseases (Eyre et al., 2012; Liu et al., 2012; Trynka et al., 2011). 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 rs16833239 (in perfect linkage disequilibrium with rs13390936 in European populations; $r^2 = 1$, $D' = 1$) with eight autoimmune diseases from summary statistics in the ImmunoBase database (Onengut-Gumuscu et al., 2015). We observed a protective association between the NTS-risk allele and a range of autoimmune diseases, with the data best supporting a model in which type 1 diabetes (both the case-control and the meta-analysis of case-control and family-based studies), celiac disease, juvenile idiopathic arthritis, narcolepsy, primary biliary cirrhosis and rheumatoid arthritis are associated with rs16833239 (Figure 4.10 Bayes factor = 547).

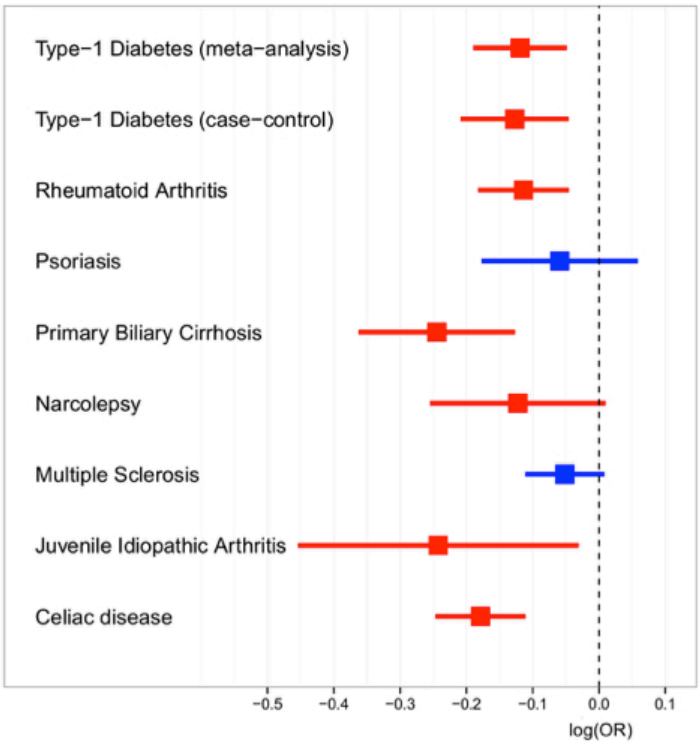

Figure 4.10 *STAT4* association with autoimmune disease in populations of European ancestry.

Log-transformed odds ratios and 95% confidence intervals of the rs16833239 (in perfect linkage disequilibrium with rs13390936 in European populations; $r^2 = 1$, $D' = 1$) association under an additive model with autoimmune disease in European populations, were calculated from summary statistics downloaded from the ImmunoBase Database(Onengut-Gumuscu et al., 2015). In a Bayesian model comparison of all possible combinations of rs16833239 association with autoimmune diseases, the model in which type 1 diabetes (both the case-control and the meta-analysis of case-control and family-based studies), celiac disease, juvenile idiopathic arthritis, narcolepsy, primary biliary cirrhosis and rheumatoid arthritis (all highlighted in red) are associated with rs16833239 is the most probable (Bayes factor = 547).

Candidate gene analysis of Mendelian susceptibility to mycobacterial disease loci

Genetic case-control study of NTS bacteraemia risk at Mendelian susceptibility to mycobacterial disease loci

In the GWAS analysis of Kenyan bacteraemia in the Kenyan discovery samples, no SNPs in established MSMD loci were taken forward for replication genotyping. To evaluate sub-genome-wide significant evidence for NTS association at MSMD loci, we considered evidence for NTS association at SNPs within 10kb of established MSMD loci. At the eight autosomal loci, we considered evidence under additive, dominant, recessive, and heterozygous advantage models of inheritance. At the two X chromosome loci, we considered evidence under an additive model, and an additive model stratified by sex. Evidence for NTS association at the MSMD loci in the Kenyan discovery samples (180 NTS cases; 2,677 controls) is summarised in Figure 4.11.

In the discovery GWAS of Kenyan bacteraemia, five SNPs, at three MSMD loci, are nominally associated with disease, with $P<1x10^{-4}$ (Table 4.1). There is no evidence of association at these SNPs in the Kenyan replication samples (38 NTS cases; 1,336 controls).

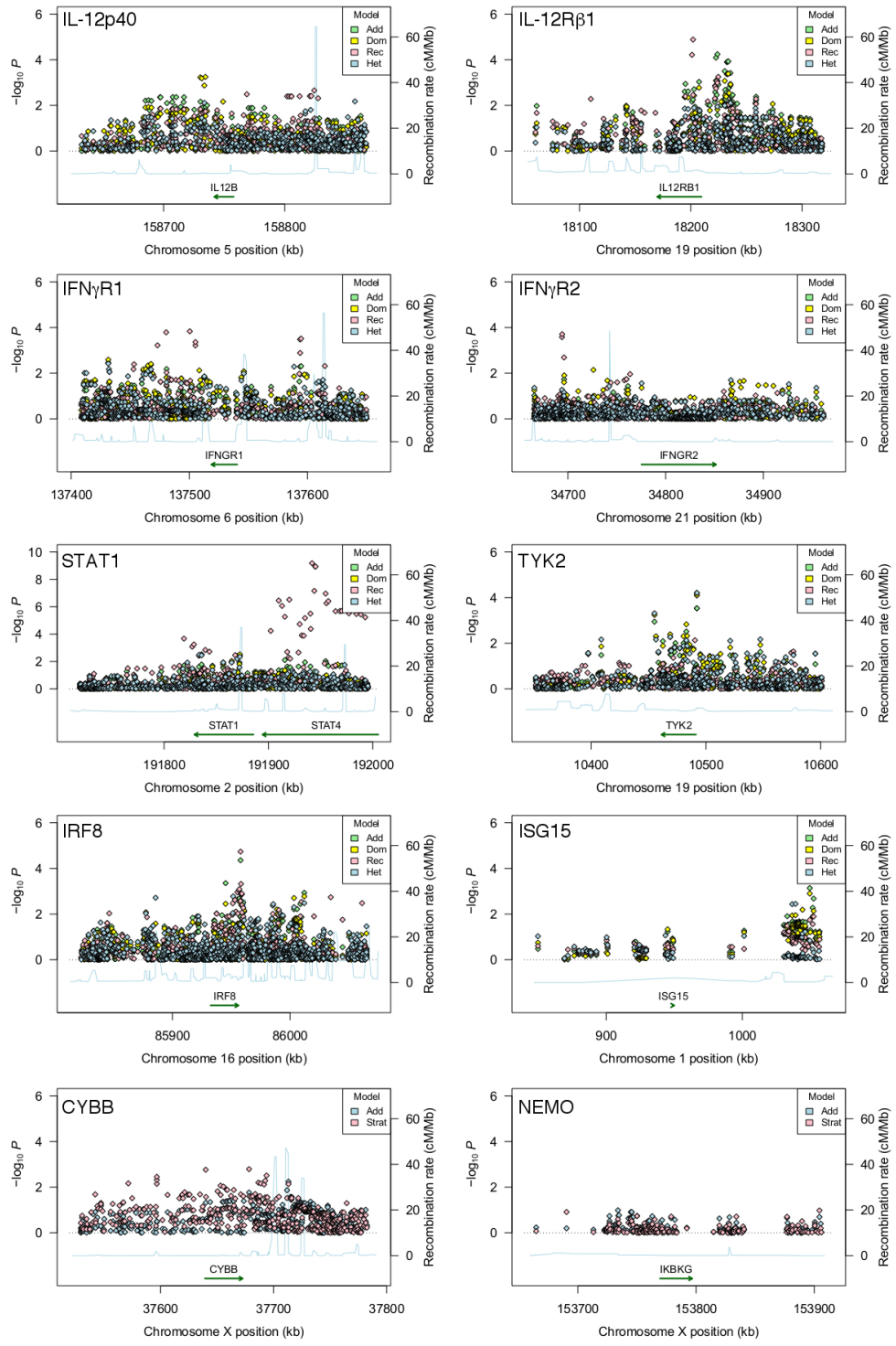

Figure 4.11 Association analysis of NTS bacteraemia at Mendelian susceptibility to mycobacterial disease loci in Kenyan children

Regional association plots within 100kb of each MSMD associated locus are displayed. SNPs are coloured according to the model of association employed. Add, additive; dom, dominant; rec, recessive; and het, heterozygous advantage; in the case of autosomal loci. Add, additive; strat, additive model stratified by sex; in the case of X chromosome loci.

Table 4.1 Association of NTS bacteraemia with SNPs in Mendelian susceptibility to mycobacterial disease loci in Kenyan discovery, and Kenyan replication samples

		Kenyan discovery samples					Kenyan replication samples			
SNP	Gene	MAF cases	MAF controls	Model	*P*-value	OR (95% CI)	MAF cases	MAF controls	*P*-value	OR (95% CI)
rs405165	IL12RB1	0.14	0.10	Rec	1.31E-05	3.2 (1.2-8.5)	0.17	0.13	0.54	2.1 (0.3-15.9)
rs114832087	IL12RB1	0.13	0.09	Rec	6.09E-05	3.2 (1.2-8.6)	0.16	0.12	0.41	2.5 (0.3-19.4)
rs273511	IL12RB1	0.39	0.28	Add	5.74E-05	1.6 (1.3-2.0)	0.31	0.30	0.92	1.0 (0.6-1.6)
rs273509	IL12RB1	0.38	0.28	Add	8.06E-05	1.6 (1.3-2.0)	0.30	0.29	0.91	1.0 (0.6-1.6)
rs76965124	TYK2	0.17	0.11	Het	6.21E-05	2.4 (2.0-2.9)	0.10	0.11	0.79	1.1 (0.3-2.0)
rs12720211	TYK2	0.17	0.11	Het	6.24E-05	2.4 (2.0-2.9)	0.10	0.11	0.77	1.1 (0.3-2.0)
rs9932835	IRF8	0.39	0.30	Rec	1.84E-05	2.5 (1.6-3.8)	0.25	0.29	0.48	0.6 (0.2-2.4)

Add, additive; rec recessive; het, heterozygous advantage; MAF, minor allele frequency.

Enrichment analysis at Mendelian susceptibility to mycobacterial disease loci

In contrast to the lack of validated association between NTS bacteraemia and MSMD loci in Kenyan children at the single variant level, there is evidence of global enrichment of NTS association at MSMD loci. In a permutation analysis of NTS association at MSMD loci, SNPs within 10kb of genes implicated in MSMD are significantly enriched for NTS association (Figure 4.12A, permutation *P*-value<$1x10^{-4}$).

In keeping with the pathogen specificity observed at MSMD loci in primary immunodeficiencies, this enrichment appears to be specific to NTS, with no evidence of enrichment observed in permutation analysis of association at MSMD loci with bacteraemia other than NTS (1,356 case samples; 2,677

controls) in the Kenyan discovery samples (Figure 4.12B, permutation *P*-value = 0.883).

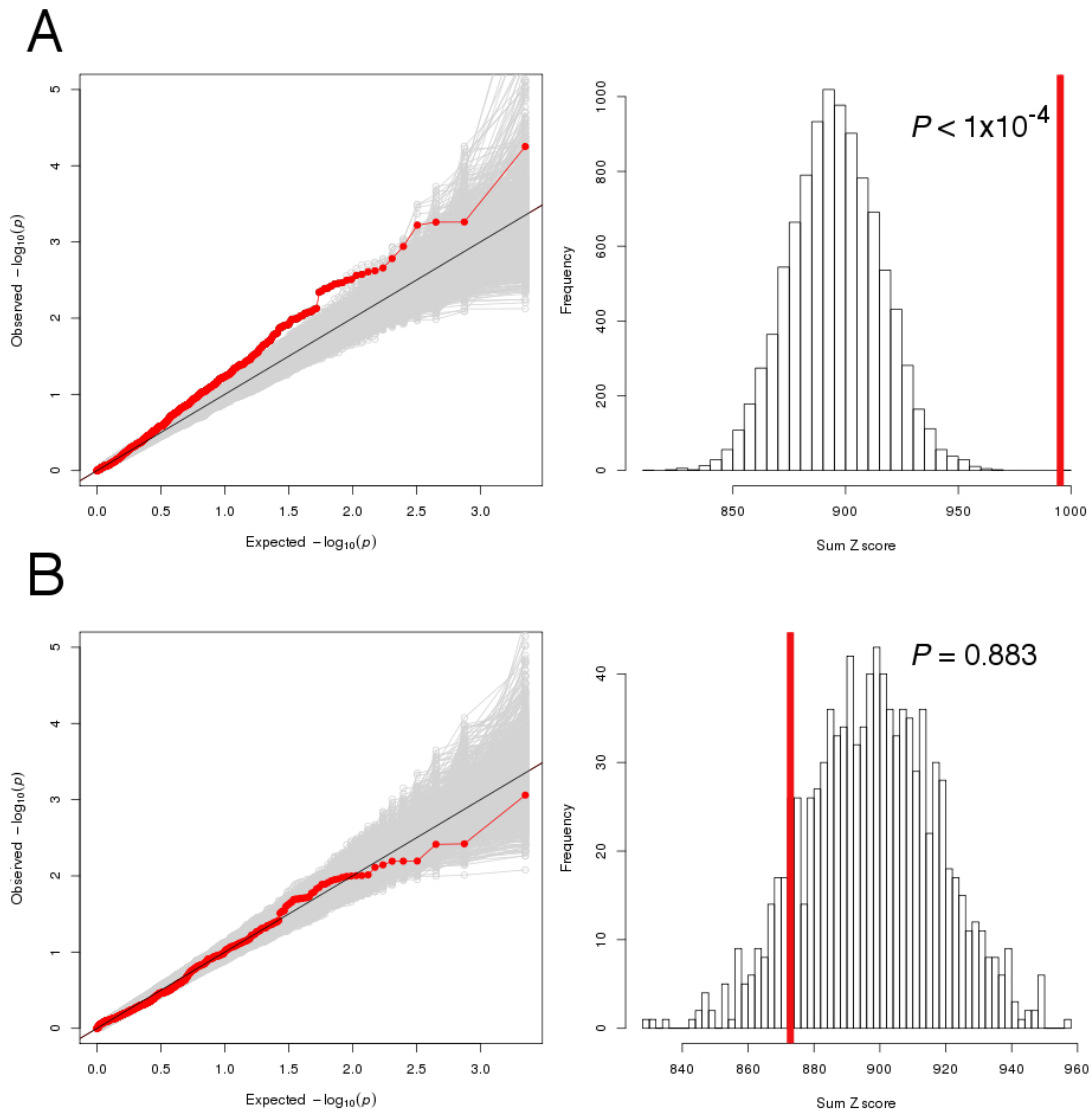

Figure 4.12 Enrichment analysis of NTS bacteraemia at Mendelian susceptibility to mycobacterial disease loci in Kenyan children

1,124 SNPs at 10 MSMD loci were included in a permutation analysis of NTS bacteraemia (180 cases, 2,677 controls; Panel A), and bacteraemia due to pathogens other than NTS (1,356 cases, 2,677 controls; Panel B). Observed *P*-value distributions are plotted in red, with P-value distribution of 1,000 case-control random permutations plotted in grey (left). The empirical distribution of summed Z-scores for each case-control permutation is plotted in grey, with the observed Z-score sum highlighted in red (right).

Conclusions

In this chapter, to characterize the biological functions of NTS-associated genetic variation at *STAT4,* we utilized a genotype-selectable bioresource of healthy European adults and samples from African children with invasive NTS disease. In common with many examples of trait-associated genetic variation identified by GWAS (Maurano et al., 2012), rs13390936 acts as, or is in linkage disequilibrium with, a regulatory determinant of *STAT4* RNA expression. Regulatory genetic variation is commonly specific to cell type and context (Fairfax et al., 2014), 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 results in reduced IFNγ levels in the serum of African children during acute invasive NTS disease.

These data establish that NTS-associated genetic variation in *STAT4* functions to impair IL-12-dependent IFNγ-mediated immunity, in a manner analogous to the immune defects seen in patients with MSMD. 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 bacteremia susceptibility conferred by genetic variation at *STAT4* is specific to NTS among the most frequent causes of bacteremia 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 (Jostins et al., 2012), and supports the hypothesis that selection pressure imposed by infectious agents has, in part, directed the evolution of autoimmune disease.

Finally motivated by the similarities between the immune defect caused by genetic variation in *STAT4* and those seen in MSMD, we investigated whether common genetic variation at MSMD loci are associated with invasive NTS disease risk in Kenyan children. We find no robust evidence of association of NTS bacteraemia with MSMD loci at the single SNP level in Kenyan children, but in a permutation-based enrichment analysis, we identify highly significant enrichment of NTS-association among SNPs located in MSMD loci. In keeping with the specificity of rs13390936 for risk of NTS disease, this enrichment for NTS-association at MSMD loci is specific to NTS bacteraemia.

Chapter 5 – Characterisation of *EVI5L* as a susceptibility locus for nontyphoidal *Salmonella* bacteraemia

Background

The identification of the *STAT4* intronic variant, rs13390936, as a risk factor for NTS bacteraemia in African children represents the only instance of an NTS-associated locus in the discovery GWAS that we have been able to validate in the replication sample collections. A significant proportion of loci suggestively associated with NTS in the discovery GWAS will represent false positive associations. However, given our limited power to confirm associations in our replication sample collections (Figure 3.2), a portion of NTS-associated loci in the discovery GWAS may represent true associations, with our failure to replicate these findings being secondary to inadequate replication sample sizes.

We therefore sought to prioritise loci associated with NTS in the discovery GWAS, for biological validation in *in vitro* models of NTS infection. Loci were prioritised according to: strength of evidence for NTS association in discovery samples; modest effect sizes, with limited power to replicate discovery findings in the replication samples; and evidence that NTS-associated variants act as regulatory determinants of gene expression. An example of an association signal meeting these criteria was located in the gene, ecotropic viral integration site 5 like (*EVI5L*). The lead SNP in the NTS GWAS was rs2059820 (ch19:7922672), highly associated with disease (P = 1.35x10^{-7}; OR 0.53) under an additive model, but with limited power to replicate an equivalent association in either the Kenyan or Malawian replication collections.

In this chapter, we use eQTL datasets to assess for evidence of enrichment of *cis* regulatory variation among NTS-associated SNPs in the *EVI5L* region. Results of these analyses, coupled with our current understanding of the biological functions of EVI5L, led us to prioritise EVI5L for further investigation as a susceptibility locus for *Salmonella* susceptibility. EVI5L is a Rab GTPase activating protein (RABGAP), functioning to inactivate Rab proteins, and thus regulate intracellular trafficking of endosomes (Frasa, Koessmeier, Ahmadian, & Braga, 2012). A key virulence determinant of *Salmonella* is its ability to tightly regulate Rab recruitment to the *Salmonella*-containing vacuole, and thus to determine its trafficking and intracellular fate (Smith et al., 2007). Indeed, at least one *Salmonella* effector protein, sifA, functions in part by acting on mammalian Rab proteins (McGourty et al., 2012). EVI5L has been

suggested to have relatively broad Rab-specificity, with evidence of interaction with Rabs 4, 7,10 and 23 (Frasa et al., 2012). The evidence for its interaction with Rab23, however, is the most convincing, hydrolysing Rab23-bound GTP in the context of cilium formation in retinal epithelial cells (Yoshimura, Egerer, Fuchs, Haas, & Barr, 2007). Rab23 is actively excluded from the *Salmonella*-containing vacuole during the early phases of intracellular *Salmonella* infection, and has been implicated in phagosome-lysosome fusion (Smith et al., 2007).

We therefore hypothesized that common genetic variation in *EVI5L* may function to modulate the risk of invasive NTS disease in African children, by modifying *EVI5L* expression levels, Rab protein trafficking during invasion and/or intracellular infection, and thus productivity of intracellular NTS infection. We test this hypothesis using RNA interference (RNAi) to define the function of EVI5L in *Salmonella* infection using in vitro models of epithelial and phagocytic infection.

In parallel to these experiments, using an unbiased enrichment analysis of genetic correlates of infection-associated cellular traits, we derive a second hypothesis for the function of EVI5L in the context of *Salmonella* infection. These enrichment analyses suggest that EVI5L may function to determine IFNγ production by infected cells. To expand on these observations, in parallel to our efforts to understand the regulatory consequences of NTS-associated variation in *STAT4*, we recruited healthy European adults via the Oxford Biobank, and assayed STAT4 phosphorylation and IFNγ production in

NK and CD4 T^+ cell subsets following *ex vivo* stimulation, correlating this with genotypes at the *EVI5L* locus.

Methods

Re-imputation of EVI5L region in Kenyan GWAS samples

To maximise imputation quality at *EVI5L*, a 1Mb region centred at *EVI5L* was intensively re-imputed in the Kenyan discovery and replication samples without pre-phasing with IMPUTE2 using 1000G Phase3 as a reference panel. Imputation was performed with 1Mb buffer regions and template haplotypes for phasing increased to 200.

RNA expression quantitative trait analysis of EVI5L

Study samples, expression and genotyping data

Using RNA expression and genome-wide genotyping data from published and unpublished eQTL studies of naïve and stimulated primary immune cell subsets in healthy European adults, we correlated genotypes in cis to *EVI5L* (within 1Mb) with *EVI5L* RNA expression. As previously described (Fairfax et al., 2012; 2014; Naranbhai et al., 2015), $CD19^+$ B cells, $CD14^+$ monocytes and $CD56^+CD3^-$ NK cells were separated from peripheral blood mononuclear cells (PBMCs) by magnetic activating cell sorting (MACS - Miltenyi). $CD16^+$ neutrophils were isolated from granulocytes with $CD16^+$ microbeads. Gene expression was quantified in total RNA from naïve cells, and monocytes stimulated with lipopolysaccharide (for 2 or 24 hours) or IFNγ (for 24 hours), 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 phase1 as the reference panel.

Statistical analysis

Genotypes were correlated with EVI5L RNA expression in each cell type: B cells (n=279), NK cells (n=245), neutrophils (n=101), naïve monocytes (n=414), and stimulated monocytes (LPS 2 hours, n=261; LPS 24 hours, n=322; IFNγ 24 hours, n=367). Normalized *EVI5L* RNA expression was correlated with genotype by linear regression and ANOVA, including the first 25 principal components of gene expression data in each cell-type/condition to account for confounding variation. *P*-values are calculated with *F*-tests (1 d.f.). *P*-values were adjusted for multiple comparisons using the Benjamini-Hochberg method. Statistical analysis was performed in R.

Cell culture

HeLa cells

The human epithelial cell line – HeLa – was obtained from the European Collection of Cell Culture. Cells were maintained in Dulbecco's Modified Eagle's medium (DMEM, Life Technologies) supplemented with 10% fetal bovine serum (Sigma-Aldrich), at 37°C and 5% CO_2.

Immortalized bone marrow-derived macrophages

Primary immortalized bone marrow-derived macrophages (iBMDM) were obtained from C57BL/6 WT mice (Charles River, U.S.A.) by Dr Sophie Matthews, MRC Centre for Bacteriology and Infection, Imperial College, London. Cells were maintained in Dulbecco's Modified Eagle's medium (DMEM; Life Technologies) supplemented with 10% fetal bovine serum (Sigma Aldrich), 20% L929 cell–conditioned medium (National Institute for Medical Research), and 1mM sodium pyruvate (Sigma Aldrich, U.K.) at 37°C and 5% CO_2.

Bacterial strains and infections

Bacterial strains

Infections were performed with the *S*. Typhimurium strain 12023, and its mutant derivative Δ*invA*/pRI203 (Δ*invA*/Inv). For experiments requiring fluorescent bacteria, isogenic strains expressing the plasmid pFPV25.1, resulting in constitutive green-fluorescent protein (GFP) expression were used. Bacteria were grown in Luria Bertani (LB) broth, with shaking (200 r.p.m.) at 37°C, supplemented with carbenicillin (50 μg/ml) where appropriate.

Epithelial cell infection

For epithelial (HeLa) infection, SPI1 expression was induced with sub-culture of an overnight bacterial culture (1:33), followed by incubation with shaking (200 r.p.m.) at 37°C for 3.5 hours. Before infection, HeLa cells were sub-cultured into 24-well tissue culture plates (containing coverslips in the case of microscopy experiments), with $5x10^4$ cells per well, and grown overnight. On the day of infection, each well was infected with 10µl of mid-log phase bacteria to give a multiplicity of infection (m.o.i.) of 100x (40 µl with an m.o.i. of 400x in the case of *invA::inv* strains). At 15 minutes post-infection, media containing extracellular bacteria was aspirated and the cells washed (twice with phosphate-buffered saline, PBS) before replacing the media with complete media supplemented with 100µg/ml gentamicin. To prevent eventual intracellular penetration of the antibiotic (and intracellular bacterial killing), media was replaced at 2 hours post-infection with complete media containing 20µg/ml gentamicin. This antibiotic concentration was maintained for the remainder of the infection, and cells incubated at 37°C with 5% CO_2.

Phagocytic infection

For phagocytic infections, using iBMDMs, overnight stationary phase (non-SPI1-inducing conditions) bacterial culture was used. iBMDMs were sub-cultured into 24-well tissue culture plates, with $1.5x10^5$ cells per well, and grown overnight. Before infection, stationary phase bacteria were opsonised in mouse serum (1:1 by volume) for 20 minutes. Bacteria were diluted 40x in complete media, before addition of 40µl (m.o.i. 100x) to wells containing iBMDMs, and the infection synchronised by centrifugation (200g, 5 minutes).

At 25 minutes post-infection, media containing extracellular bacteria was aspirated and the cells washed (twice with PBS) before replacing the media with complete media supplemented with 100μg/ml gentamicin. Media was replaced at 2 hours post-infection with complete media containing 20 μg/ml gentamicin, and cells incubated at 37°C with 5% CO_2.

Assays of intracellular replication

In the case of colony-forming unit (CFU) assays, to determine net intracellular bacterial replication, at appropriate time points, infected cells were washed (2x PBS), and lysed in 0.1% Triton (Sigma-Aldrich) in PBS at 4°C. Lysed cells were re-suspended, and following serial dilution (fold dilution dependent on strain, infected cell, and time-point) plated in duplicate onto LB agar. Plates were incubated overnight and bacterial colonies enumerated with an automated colony counter (ACOlyte, Synbiosis). Bacterial recovery for each infection condition, and each time-point, was measured in triplicate. Net intracellular replication within each infection condition was calculated as fold-change from bacteria recovered at 2 hours post-infection. Summary statistics (mean and standard errors) for a given infection condition were calculated with averages of at least three independent experiments, each conducted in triplicate. Statistical analysis was performed in R.

Assays of net intracellular replication, e.g. the CFU assay described above, can result in apparent differences in intracellular replication between experimental groups, secondary to differences other than the rate of

intracellular replication, e.g. differential rates infected cell cytotoxicity during the course of an experiment. To more accurately determine whether an observed difference in net replication is secondary to altered intracellular growth, we also assayed intracellular replication of GFP-expressing bacteria by flow cytometry. For these experiments, adherent cells were collected into Optimem media (Life Technologies), following re-suspension by incubation in trypsin-EDTA (TE, Sigma-Aldrich) buffer at 37°C for 1 minute (HeLa cells), or by scraping (iBMDMs). Samples were immediately run on a FACS Calibur flow cytometer (Becton Dickinson). GFP-fluorescent, infected cells were identified by fluorescence intensity in the FL-1 channel, with 10,000 infected-cell events recorded per sample. Flow cytometric analysis was performed in FlowJo 10.0.8. Intracellular bacterial numbers were estimated as the median GFP fluorescence per infected cell. Intracellular bacterial burden in each infection condition, at each time-point, was measured in triplicate. Intracellular replication within each infection condition was calculated as fold-change from median GFP fluorescence per infected cell at 2 hours post-infection. Summary statistics (mean and standard errors) for a given infection condition were calculated with averages of at least three independent experiments, each conducted in triplicate. Statistical analysis was performed in R.

Microscopy

At appropriate time points, infected cells were washed (2x PBS), and fixed in 4% paraformaldehyde (20 minutes, room temperature). Fixed cells were again washed (2x PBS) and excess PFA quenched with 50mM ammonium chloride.

Coverslips were stored at 4°C prior to labelling and analysis. For anti-Flag labelling, cell monolayers were washed (2x 0.1% Triton in PBS) before primary antibody incubation (1:200 rabbit monoclonal anti-Flag, Sigma Aldrich) in 0.1% Triton in PBS with 10% horse serum at room temperature for 2 hours. Again cell monolayers were washed (2x 0.1% Triton in PBS, 1x PBS) before secondary antibody incubation (1:500 Alexa Fluor 488-conjugated anti-rabbit, Life Technologies) and DNA staining (1:10,000 4′,6-Diamidine-2′-phenylindole dihydrochloride, DAPI, Sigma-Aldrich) in 0.1% Triton in PBS with 10% horse serum at room temperature for 2 hours. Coverslips were then washed, blotted and mounted onto slides with polyvinyl alcohol. Following labelling, micrographs were analysed with a fluorescent microscope (BX50, Olympus) or a confocal, scanning microscope (LSM710, Zeiss). For quantitative estimates, e.g. co-localisation, sample infection conditions for each slide were obscured and shuffled prior to analysis.

Cytotoxicity

Cytotoxicity during phagocytic infection was estimated by lactate dehydrogenase (LDH) release into culture supernatant during infection. Phagocytic infections were performed as described above, with the exception that infected cells were maintained in a single concentration of gentamicin (20μg/ml) throughout the experiment, to ensure that the measured LDH release included the first two hours of infection. At appropriate time-points, culture supernatants were harvested from wells containing infected cells, briefly centrifuged (13,000 rpm, 1 minute) and the cell-free supernatants

stored at -80°C prior to analysis. Supernatants from wells of uninfected cells of the same density were harvested in the same manner, to act as negative controls. To estimate maximal LDH release (100% cytotoxicity), wells of uninfected cells at the same density as those used in the experiment, were frozen in the tissue culture plate at -80°C. LDH concentrations in each sample was assayed with the CytoTox 96 Non-Radioactive Cytotoxicity Assay (Promega) according to manufacturer's instructions. In a 96-well plate, samples were incubated in a 1:1 ratio with CytoTox 96 reagent, for 30 minutes at room temperature, before addition of CytoTox 96 Stop Solution. Absorbance at 490nm was measured using a Tecan Infinite PRO 200 plate reader (Tecan). Cytotoxicity, as the proportion of observed cell death, was calculated as the LDH release in uninfected cells, subtracted from LDH release in infected cells, divided by maximal LDH release. Cytotoxicity in each infection condition, at each time-point, was measured in triplicate. Summary statistics (mean and standard errors) for a given infection condition were calculated with averages of at least three independent experiments, each conducted in triplicate. Statistical analysis was performed in R.

Knockdown of EVI5L expression by RNA interference

EVI5L RNA knockdown in HeLa cells

Two short interfering RNA (siRNA) molecules targeting human *EVI5L* mRNA were used (Ambion Silencer Select siRNA Cat. Nos. s41854 & s41856, ThermoFisher Scientific), alongside scrambled negative siRNA control 21mer

duplex molecules (Ambion Silencer Select siRNA Cat. No. 4390843) designed to not anneal to any human mRNA sequence were used. HeLa cells were seeded into tissue culture plates one day prior to siRNA transfection. Complete media was substituted for Optimem 4 hours prior to transfection. siRNA oligos and Lipofectamine 2000 (1µl per 24-well sample, Life Technologies) were incubated in Optimem for 10 minutes at room temperature, before mixing and incubation together for 15 minutes. Lipofectamine/siRNA complexes were then added to the cells at a final siRNA concentration of 80nM, and incubated for 4 hours, before changing media to complete DMEM. Experiments were performed at 72 hours post-transfection.

EVI5L knockdown in iBMDMs

To knockdown EVI5L mRNA expression in murine iBMDMs, we used pooled, puromycin-selectable short hairpin RNA (shRNA) molecules targeting murine EVI5L mRNA (Santa Cruz Biotechnology, Cat. No. sc-144965-V), alongside a scrambled negative, puromycin-selectable shRNA (Santa Cruz Biotechnology, Cat. No. sc-108080) delivering an oligo targeting no known murine RNA transcript. shRNA-containing lentiviral particles were used to transduce shRNA molecules into iBMDMs. Prior to transduction, $3x10^5$ iBMDMs were seeded into 12-well tissue culture plates, and rested overnight. 10^5 lentiviral particles were added to each well in complete medium containing polybrene (final concentration of 8µg/ml), and plates then centrifuged (1,800rpm, room temperature) for 2 hours. At 48 hours, shRNA expressing cells were expanded, before selection with puromycin (final concentration =

2.5µg/ml). Experiments were performed on shRNA-expressing cells, following three passages in puromycin-containing media.

Confirmation of RNAi knockdown

To test EVI5L RNAi knockdown, RNA was extracted from EVI5L-targeting RNAi-treated cells, and cells treated with scrambled negative control RNAi oligos with RNeasy kits (Qiagen) according to manufacturer's instructions. Complementary DNA (cDNA) was synthesized from 400ng of extracted RNA using the QuantiTect Reverse Transcription Kit (Qiagen). Gene-specific cDNA concentrations were assayed by quantitative-PCR (qPCR), using the relative expression method on a Rotor Gene RG-3000 (Corbett Life Science). Each 25µl PCR reaction included 0.5µl of sample cDNA, gene-specific qPCR primers at a final concentration of 0.2µM, SYBR® Select Master Mix (containing SYBR® GreenER™ dye, AmpliTaq® DNA Polymerase, and dNTPs: ThermoFisher Scientific), and PCR-grade water (Sigma-Aldrich). Reactions were incubated at 95°C for 10 minutes, before 45 cycles of melting (95°C, 15 seconds), annealing (55°C, 20 seconds), and extension (72°C, 20 seconds), with fluorescence acquisition during the extension phase. Exon-spanning qPCR primers were used for human *EVI5L* (F, 5'-CAAGCTGGTCCTCAAAGCCT -3'; R, 5'- CTTGCCCCTGGATTAACCTATCA -3') and for murine EVI5L (F, 5'- GCTGGTCCTCAAGGCTTATC -3'; R, 5'-GATCTGCTCCTCCATCTCTTTAC-3'). In HeLa cells, *EVI5L* expression was quantified relative to human β-actin, and in iBMDMs, relative to murine m18s, by the ΔCt (cycle threshold) method.

In HeLa cells, to confirm EVI5L RNAi knockdown at the protein level, protein was extracted, and quantified by immunoblot, from EVI5L and control siRNA-treated cells expressing Flag-EVI5L (see below). Cells were solubilised in 2x sodium dodecyl sulphate (SDS) buffer, before sonication. Samples were denatured (95°C for 10 minutes) prior to loading into 10% polyacrylamide gels and separation by electrophoresis (100V). Proteins were transferred to a PVDF membrane (15V, 25 minutes) using a Trans-Blot® Turbo™ Transfer System (BioRad). Membranes were then blocked (5% skimmed milk in TBST - Tris-buffered saline, 0.1% Tween 20) for 2 hours, before washing (5x TBST), and incubated with the primary antibody (anti-Flag M2 mouse monoclonal, Sigma-Aldrich; and anti-α-tubulin mouse monoclonal as a loading control, Abcam), diluted 1:2,000 in 5% skimmed milk/TBST, overnight at 4°C. After washing (5x TBST), membranes were then incubated with secondary antibody (horseradish peroxidase-conjugated anti-mouse IgG polyclonal), diluted 1:5,000 in 5% skimmed milk/TBST for 2 hours at room temperature. Blots were then developed using ECL chemiluminescent reagent (Amersham, GE Life Sciences) and visualised on X-ray film by autoradiography.

Generation of cell-lines stably expressing Flag-tagged EVI5L

Lentiviral particles containing a puromycin-selectable 3rd generation lentiviral transfer vector, expressing human Flag-*EVI5L* (custom-designed, produced by VectorBuilder, Cyagen), were produced by transfection of 293ET cells. Plasmid DNA was prepared from plasmid-expressing *E. coli* strains using

QIAprep Spin Miniprep Kits (Qiagen) according to manufacturer's instructions. Sub-confluent 293-ET monolayers were transfected (using Lipofectamine 2000, as above) with *EVI5L*-containing pro-viral plasmid and packaging plasmids; *gag* and *pol* expression provided by the pRRE plasmid, *rev* expression by pRSV-Rev, and *env* expression by pMD.VSVG. Virus-containing supernatants were harvested at 48 and 72 hours post-transfection, and stored at -20°C before use.

Lentiviral particles were then used to transduce HeLa cells as described above. At 48 hours, Flag-*EVI5L*-expressing cells were expanded, before selection with puromycin (final concentration = 1.0µg/ml). Expression of FLAG-tagged EVI5L in transduced cells was confirmed by anti-FLAG immunoblotting (as described above).

Immunoprecipitation

293-ET cells were co-transfected with vectors expressing Flag-EVI5L and GFP-tagged Rab proteins (Rabs 5, 7, 9 and 23). At 24 hour post-transfection, cells were washed in PBS, and lysed in Lumier lysis buffer (containing protease inhibitors; 1mM dithiothreitol, 1mM benzamidine, 2µg/ml aprotinin, 5µg/ml leupeptin, 1µM phenylmethylsulphonyl fluoride). Lysates were incubated on ice for 15 minutes before centrifugation (13,000rpm for 10 minutes). Cleared post-nuclear supernatants were incubated with anti-GFP beads (on rotor at 4°C) for 2 hours. Beads were then washed, and protein eluted with 2x SDS loading buffer. All steps of immunoprecipitation were

performed at 4°C. Lysates and immunoprecipitated protein were then run by SDS-PAGE and immunoblotted for Flag and GFP (see above).

Enrichment for genetic correlates of cellular traits in the NTS bacteraemia GWAS

Enrichment analysis

GWAS summary statistics of *in vitro*, infection-related cellular traits, were kindly provided by Prof. Dennis Ko (Duke University, North Carolina, U.S.A.). Cellular traits for GWAS were generated using a high-throughput screening platform (Hi-HOST), in which cellular traits in primary cell-lines derived from hundreds of individuals with available genome-wide genotyping data, are assayed and genetic correlates of a given trait sought using standard GWAS methods(Ko et al., 2012). To assess for overlap of genetic correlates between NTS bacteraemia and infection-related cellular traits, we used Fisher's Exact tests to test for evidence of enrichment of SNPs nominally associated ($P<1x10^{-3}$) with NTS and a given cellular trait. Well-imputed SNPs (info>0.8), with a minor allele frequency in excess of 0.05, and no evidence of extreme deviation from Hardy-Weinberg equilibrium ($P>1x10^{-10}$) were included in the analysis. Fold-enrichment of SNPs associated with both traits was calculated by comparison of the observed number of shared SNPs compared with that expected by chance under a χ^2 distribution with one degree of freedom.

IFNγ-related protein phenotypes of EVI5L eQTLs

We re-analysed our data describing phospho-STAT4 and IFNγ production following *ex vivo* IL-12 and NTS stimulation in NK cells and CD4+ T cells, which we had generated to investigate the functional role of NTS-associated genetic variation in *STAT4* (see chapter 4).

Data from 55 healthy European donors were included in the analysis. In these individuals, we tested for association between genotype at EVI5L *cis* eSNPs genotyped with the HumanOmniExpress-12v1.0 BeadChip (data provided by Dr Benjamin Fairfax, Wellcome Trust Centre for Human Genetics) and the proportion of $pSTAT4^+$ and $IFN\gamma^+$ cells in NK and CD4+ T cell subsets following stimulation with NTS (for IFNγ staining) and IL-12 (for pSTAT4 and IFNγ staining).

Proportions of $pSTAT4^+$ and $IFN\gamma^+$ cells were logit transformed prior to analysis, and genotype-phenotype associations tested with linear regression and ANOVA, adjusted for age and sex. P-values are calculated with *F*-tests (1 d.f.). In the case of the pSTAT data, there was evidence of significant batch effect, with two (of 12) batches significantly ($P<0.05$) associated with pSTAT4 in NK cells, and one in $CD4^+$ T cells. All batches were included in the analysis and regression models adjusted for batch membership, along with age and sex. Statistical analysis was performed in R.

Results

Common genetic variation at EVI5L as a risk factor for invasive nontyphoidal Salmonella disease in African children

In the GWAS of NTS bacteraemia in the Kenyan discovery samples (NTS cases, n = 180; controls, n = 2,677), common genetic variation in *EVI5L* is associated with risk of NTS disease (Figure 5.1). There is no evidence of replication in the Kenyan (Figure 5.1) or Malawian (Table 3.5) replication sample collections. This lack of replication is in keeping with limited power to detect the estimated effect size at the peak-associated SNP, assuming an additive model of association, given the observed minor allele frequency, at $\alpha<0.05$; 12% in the Kenyan replication samples, and 24% Malawian samples.

We fitted a multinomial regression model of bacteraemia risk at the peak *EVI5L* SNP (rs2059820) in the Kenyan discovery samples, including 2,677 control samples, and 1,336 case samples (NTS, 180; *S. aureus*,175; *S. pneumoniae*, 426; *H. influenzae* type b, 128; *E. coli*, 151; β-haemolytic *Streptococci*, 146; *Acinetobacter*, 130). In that analysis, the data best supports a model in which rs2059820 is associated with NTS disease alone, and not with risk of bacteraemia secondary to other pathogens (Figure 5.2, Bayes factor c.f. null = 14,012). In keeping with the lack of evidence of replication, in a multinomial regression model of bacteraemia risk at rs2059820 in the Kenyan replication samples (controls, 1,336; NTS cases, 38; *S. aureus*, 45; *S. pneumoniae*, 113; *H. influenzae* type b, 31; *E. coli*, 44; β-

haemolytic *Streptococci*, 45; *Acinetobacter*, 47), the data do not support an association with NTS bacteraemia, with the most likely model being one in which *E. coli* bacteraemia alone is associated with rs2059820 (Figure 5.2, Bayes factor c.f. null = 6).

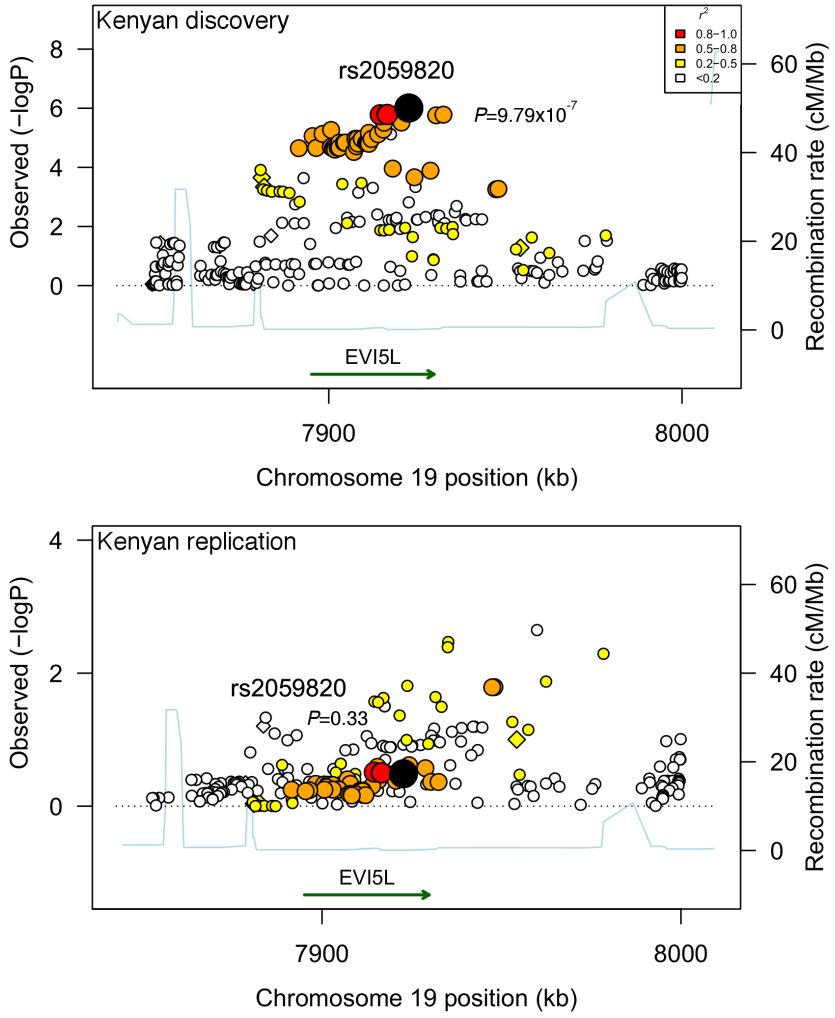

Figure 5.1 Association plot of NTS bacteraemia susceptibility at the *EVI5L* region in Kenyan children.

NTS-association at the *EVI5L* region under an additive model in Kenyan discovery samples (n = 180 cases, 2,677 controls) and Kenyan replication samples (n = 38 cases, 1,336 controls). rs2059820 is represented as a black circle in each plot. Imputed SNPs (following intensive re-imputation) are plotted as circles, directly genotyped SNPs as diamonds. SNPs are coloured according to strength of linkage disequilibrium (r^2) to rs2059820.

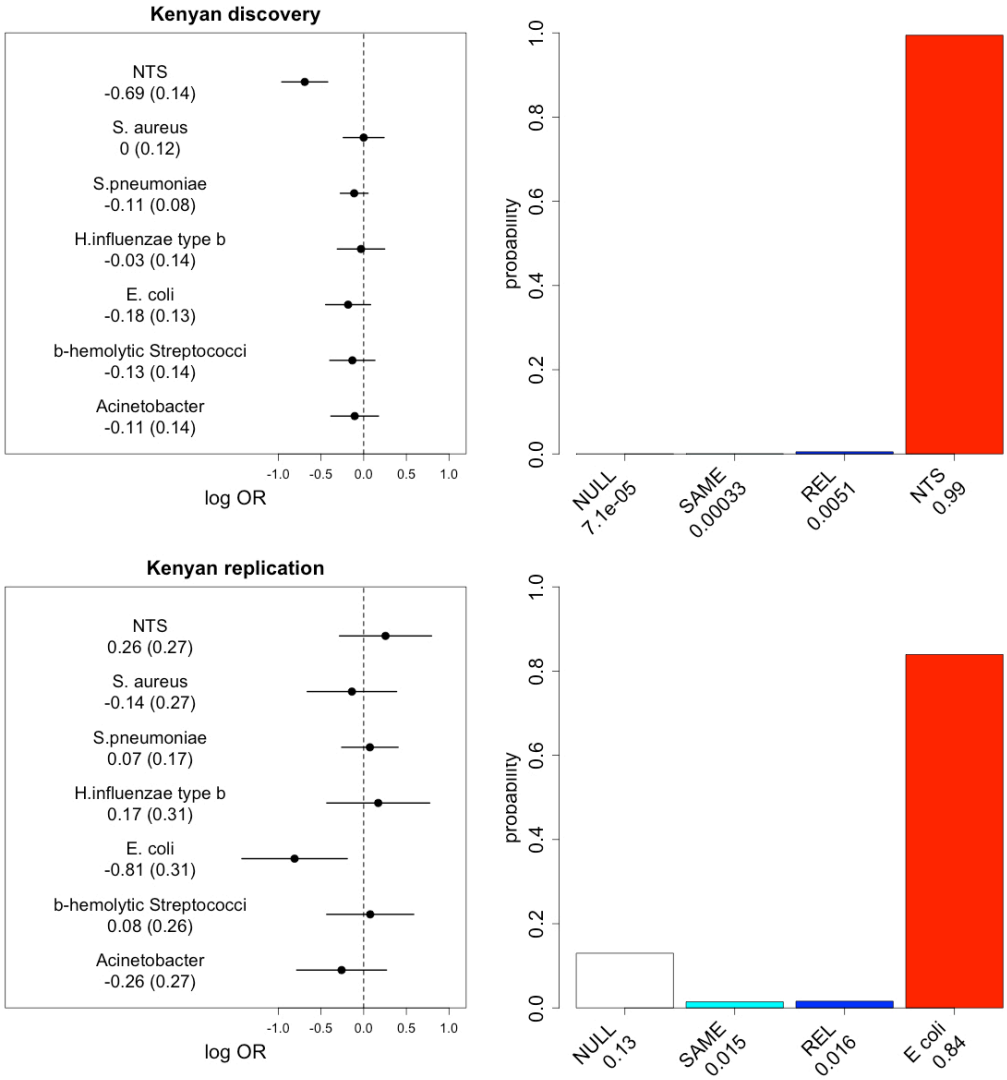

Figure 5.2 *EVI5L* association with major bacterial pathogens in Kenyan children

Log-transformed odds ratios and 95% confidence intervals of rs2059820 association under a recessive model in Kenyan discovery (top) and replication (bottom) samples; left panels. Posterior probabilities of models of association at rs2059820: NULL, no association with any bacterial pathogen; SAME, the same effect across all bacterial pathogens; REL, related effects across all bacterial pathogens; NTS, a non-zero effect for rs2059820 genotype in NTS alone; E coli, a non-zero effect for rs2059820 genotype in E. coli alone.

RNA expression quantitative trait analysis of EVI5L

EVI5L cis expression quantitative trait loci in immune cell subsets

In healthy adults of European ancestry, following correction for multiple association testing, there are no significant *cis* eQTLs for *EVI5L* expression in naïve NK cells (n = 245 individuals), B cells (n = 279) and neutrophils (n = 101). However, in naïve monocytes (n = 414), and monocytes stimulated with LPS for 2hours (n = 261), LPS for 24 hours (n = 322), and IFNγ for 24 hours (n = 367), there are significant (FDR-adjusted *P*-value<0.05) cis eQTLs determining *EVI5L* RNA expression (Figure 5.3). Many of the identified monocyte eSNPs are shared between stimulation conditions (Figure 5.4). The peak cis eSNP for *EVI5L* in naïve monocytes, rs111777172, is shared by monocytes stimulated for 24 hours with either IFNγ or LPS, but not by monocytes stimulated for 2 hours with LPS (Figure 5.5). A conditional *cis* eQTL analysis of *EVI5L* expression, with rs111777172 included as a covariate in the regression model, demonstrates that a single regulatory determinant appears to be shared between naïve monocytes and monocytes stimulated for 24 hours with LPS, in that no significant eSNPs remain in the conditional analysis. This is not the case for monocytes stimulated with LPS for 2 hours or IFNγ for 24 hours, in which significant eSNPs for *EVI5L* expression persist in the conditional analysis. In the context of LPS stimulation for 2 hours, and IFNγ stimulation for 24 hours, monocytes have significant eSNPs independent of rs111777172 (Figure 5.6).

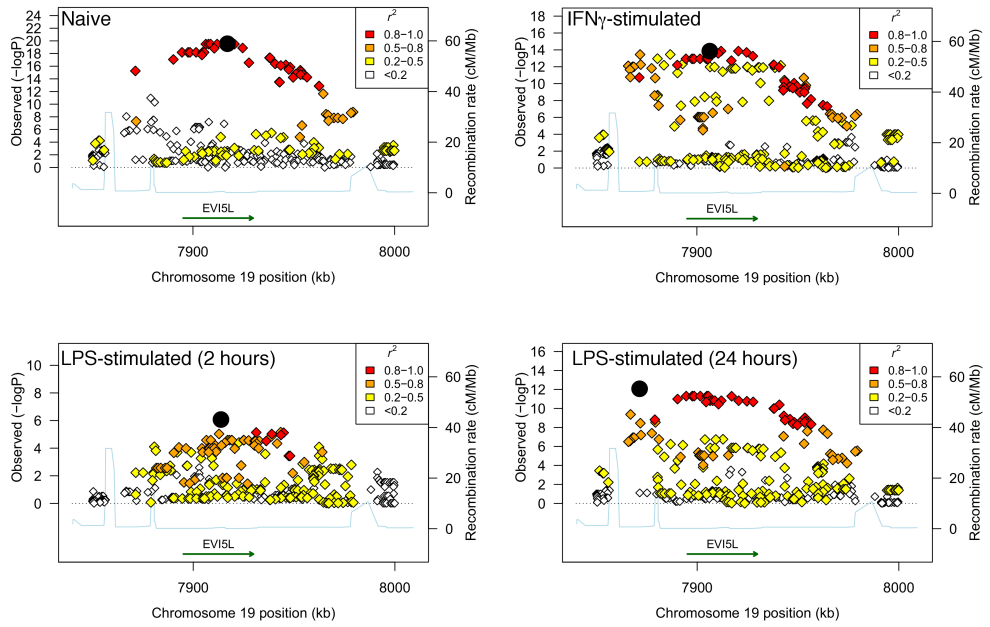

Figure 5.3 Regional association plot of cis genetic loci with *EVI5L* RNA expression in monocytes.

Association plots are shown for cis eQTL analysis of *EVI5L* in naïve and stimulated (LPS 2 hours, LPS 24 hours and IFNγ 24 hours) monocytes. Peak eSNPs in each condition are represented as a black circle. SNPs are coloured according to strength of linkage disequilibrium (r^2) to the peak eSNP.

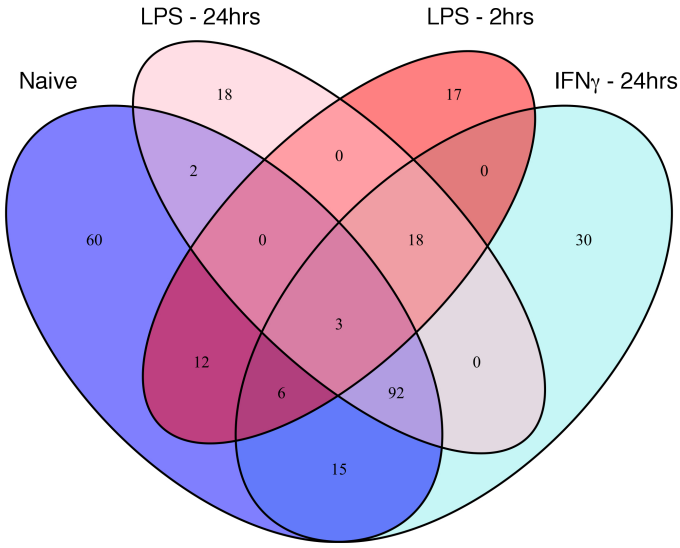

Figure 5.4 Sharing of *EVI5L* eSNPs between naïve and stimulated monocytes.

Euler plot illustrating sharing of significant *EVI5L* eSNPs (FDR-adjusted *P*-value<0.05) between naïve and stimulated monocytes.

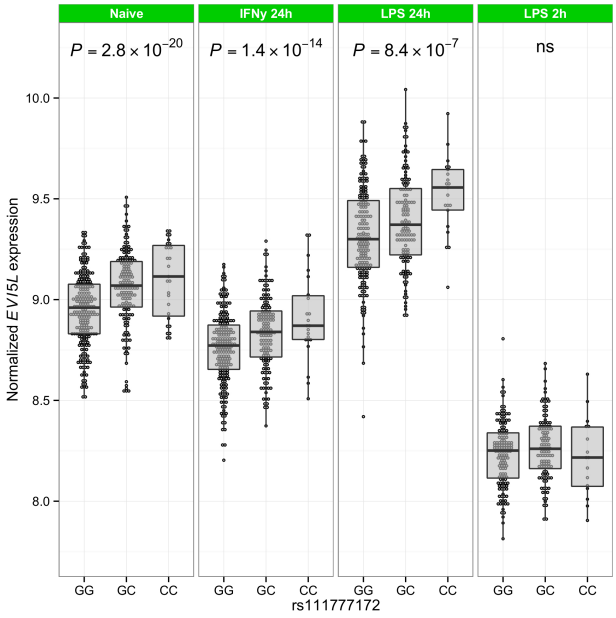

Figure 5.5 Effect of rs111777172 on *EVI5L* RNA expression in naïve and stimulated monocytes

Expression of *EVI5L* RNA is dependent on rs111777172 genotype in naïve monocytes, and monocytes stimulated with IFNγ and LPS for 24 hours, but not LPS for 2 hours.

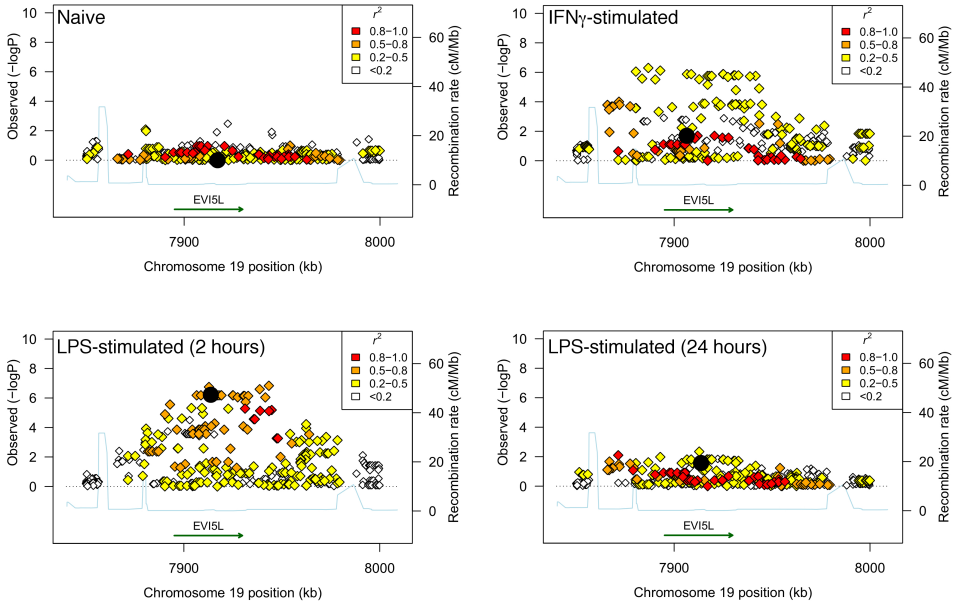

Figure 5.6 Regional eQTL plot of *EVI5L* expression in monocytes conditioned on rs111777172.

Association plots are shown for cis eQTL analysis of *EVI5L*, conditioned on the peak eSNP in naïve monocytes (rs111777172). Analysis is shown for naïve and stimulated (LPS 2 hours, LPS 24 hours and IFNγ 24 hours) monocytes. Peak eSNPs (in the unconditioned analysis) in each condition are represented as a black circle. SNPs are coloured according to strength of linkage disequilibrium (r^2) to the peak eSNP.

Taken together, the analysis of cis eQTLs for EVI5L in monocytes, demonstrates that naïve monocytes and monocytes stimulated with LPS for 24 hours appear to share a single regulatory determinant of *EVI5L* RNA expression levels. By contrast, monocytes stimulated with LPS for 2 hours or with IFNγ, have evidence of cis regulatory determinants of EVI5L RNA expression that are independent of those operating in naïve monocytes. In the case of monocytes stimulated with LPS for 2 hours, conditioning on the peak naïve monocyte eSNP has very little appreciable effect on the distribution of EVI5L eSNPs, suggesting an independent regulatory determinant in this context. In the case of monocytes stimulated with IFNγ for 24 hours however, conditioning on the peak naïve monocyte eSNP abolishes the unconditioned peak association, leaving a second, independent eQTL peak. There is considerable overlap between this second group of eSNPs in IFNγ-stimulated monocytes and those seen in monocytes stimulated with LPS for 2 hours (30 shared eSNPs of 54 in IFNγ), suggesting that these peaks may represent shared regulatory determinants between these two stimulation contexts. In keeping with this there are no residual significant *EVI5L* eSNPs in IFNγ-stimulated monocytes when the analysis is conditioned on both the peak naïve monocyte eSNP and the peak eSNP in monocytes stimulated with LPS for 2 hours (rs9710465).

Enrichment of NTS-associated SNPs for EVI5L eSNPs in monocytes

There is highly significant enrichment of EVI5L *cis* eSNPs (FDR-adjusted *P*-value<0.05) among NTS-associated SNPs ($P<1x10^{-4}$) within 1Mb of *EVI5L* in

the Kenyan discovery GWAS samples. This enrichment is observed in naïve and stimulated monocytes (Table 5.1, Figure 5.7), and is particularly striking for monocytes stimulated with LPS for 2 hours.

To understand whether increased or decreased expression of *EVI5L* in monocytes is associated with NTS disease, we compared the observed effect size (and direction of effect) at SNPs found to be associated with NTS disease ($P<1x10^{-4}$), and also significant EVI5L eSNPs (FDR-adjusted $P<0.05$). This analysis reveals a complex relationship between NTS disease and EVI5L expression level (Figure 5.8), with alleles increasing EVI5L expression in naïve and LPS (2 hours)-stimulated monocytes increasing NTS risk, but alleles decreasing *EVI5L* expression in monocytes stimulated with LPS and IFNγ for 24 hours increasing disease risk.

Table 5.1 Enrichment of EVI5L eSNPs in naïve and stimulated monocytes among NTS-associated SNPs

Stimulation condition	*P*-value	Fold enrichment
Naïve	$1.38x10^{-8}$	6.6
LPS - 2 hours	$1.12x10^{-19}$	23.2
LPS - 24 hours	$2.26x10^{-6}$	7.1
IFNγ - 24 hours	$1.86x10^{-5}$	5.5

P-values are calculated with Fisher's exact test. Fold enrichment is the ratio of observed shared SNPs observed between EVI5L eSNPs and NTS bacteraemia, and that expected under a χ^2 distribution.

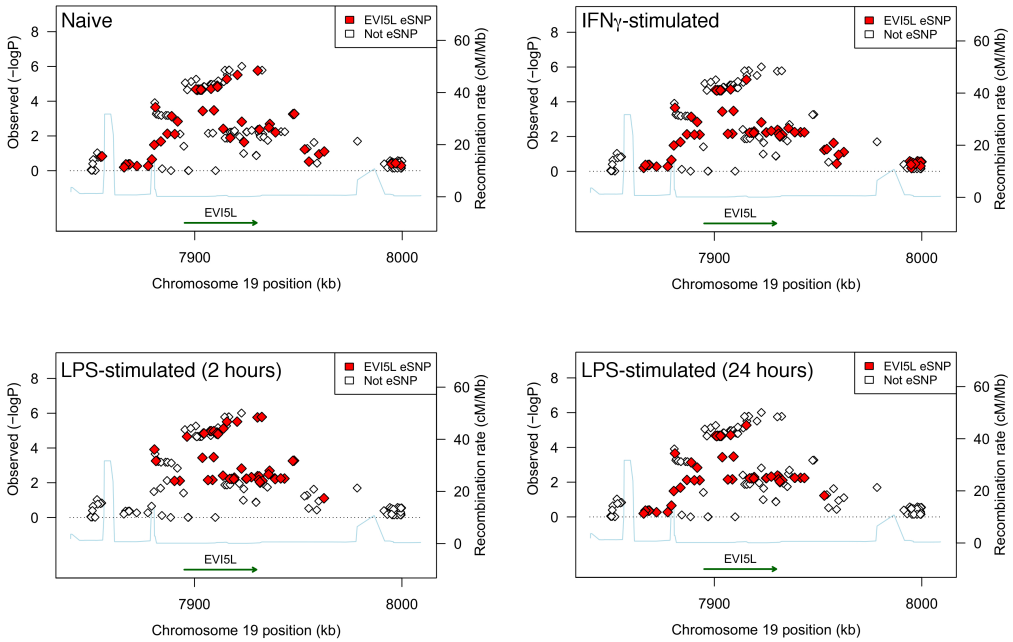

Figure 5.7 Enrichment of EVI5L eSNPs in naïve and stimulated monocytes in NTS-associated SNPs

Each panel depicts NTS-association at the *EVI5l* region under an additive model in Kenyan discovery samples (n = 180 cases, 2,677 controls), annotated for significant eSNPs (FDR-adjusted *P*-value<0.05) in monocytes in four stimulation conditions. Significant eSNPs are highlighted in red.

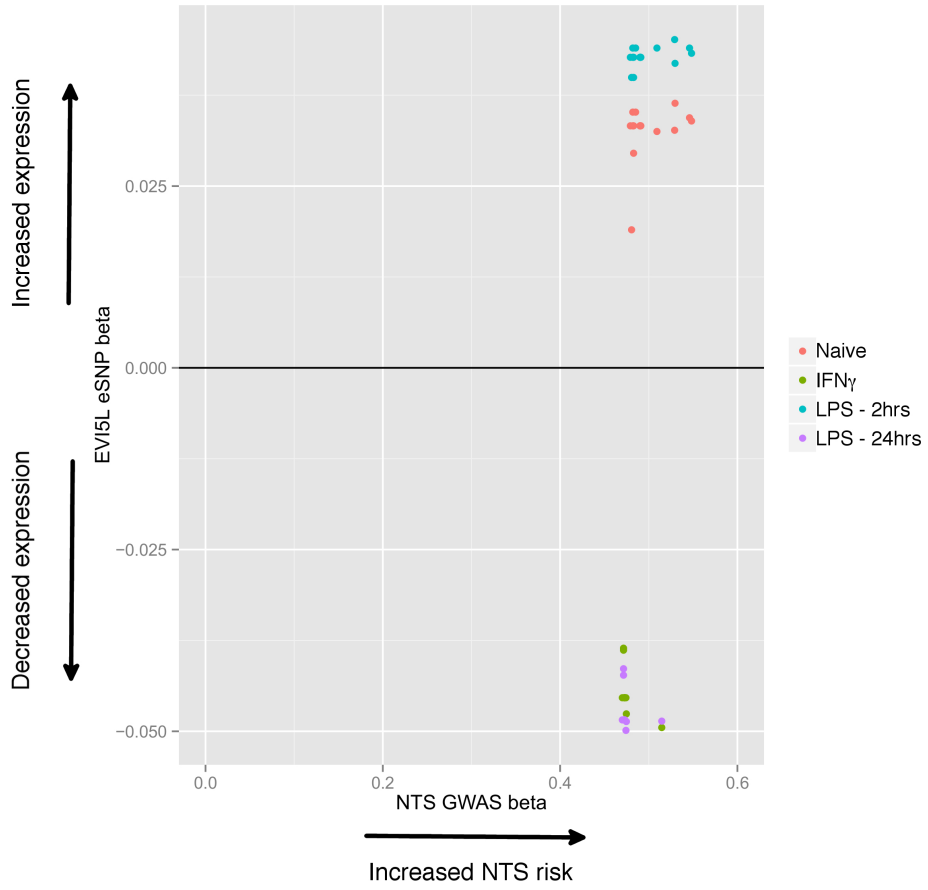

Figure 5.8 Effect size and direction of shared NTS risk loci and *EVI5L* eSNPs in monocytes.

All plotted SNPs are associated with NTS disease (P=1x10^{-4}), and are significant *EVI5L* eSNPs (FDR-adjusted P<0.05) for monocytes in at least one stimulation condition (see colour-coding).

EVI5L cellular distribution and interaction partners

Subcellular localisation of EVI5L

In uninfected human epithelial (HeLa) cells, EVI5L is diffusely cytosolic (Figure 5.9). Infection with either wild-type or Δ*invA*/Inv mutant 12023 results in co-localisation of EVI5L with intracellular *Salmonellae* (Figure 5.9). EVI5L/*Salmonella* co-localisation is transient, peaking at 15 minutes post-infection, and appearing to rapidly dissociate, with no persistent co-localisation observed beyond 30 minutes post-infection (Figure 5.10). EVI5L:*Salmonella* co-localisation is not SPI-1 dependent, with co-localisation seen during infection with both wild-type and Δ*invA*/Inv *Salmonella*, with no significant differences in frequency or kinetics of co-localisation (Figure 5.10).

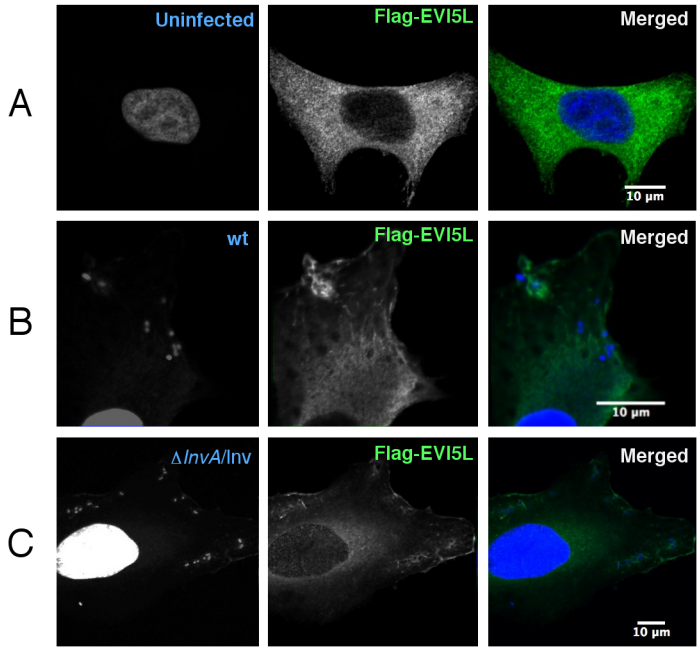

Figure 5.9 Co-localisation of EVI5L with *Salmonella* during epithelial cell infection.

HeLa cells stably expressing Flag-EVI5L were left uninfected (A), or infected with wild-type *Salmonella* (B) and Δ*InvA*/Inv *Salmonella*. At 15 minutes post-infection, cells were fixed, immunolabelled for Flag, and DNA-stained with DAPI.

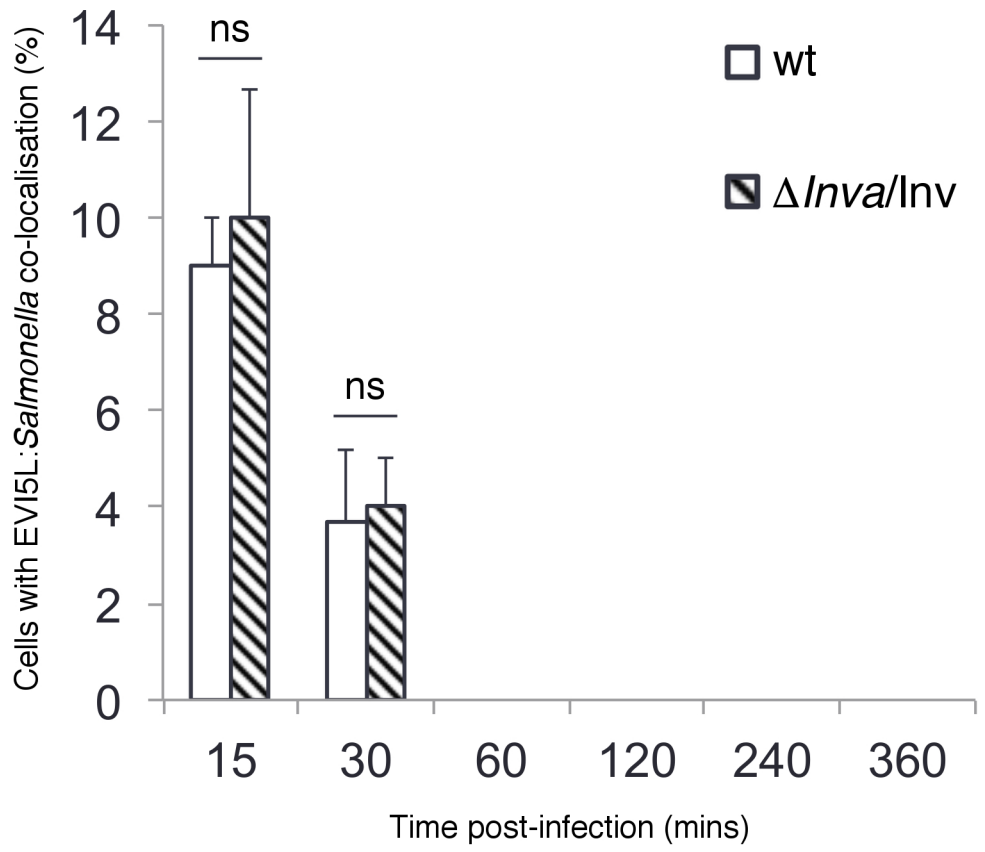

Figure 5.10 Co-localisation of EVI5L with *Salmonella* during epithelial cell infection.

HeLa cells stably expressing Flag-EVI5L were infected with wild-type *Salmonella* and Δ*InvA*/Inv *Salmonella*. The number of cells with EVI5L/*Salmonella* co-localisation was quantified by microscopy, counted per 100 infected cells. Summary statistics are calculated from three independent experiments. Error bars indicate SEM. ns, not significant.

EVI5L interacts with Rabs 23 and 9

In 293ET cells co-transfected with Flag-EVI5L and GFP-tagged Rabs (Figure 5.11), EVI5L interacts with Rab23, and to a lesser extent Rab9. The interaction of EVI5L with Rab23 is in keeping with previously published data in retinal epithelial cells, but we were unable to recapitulate previously published

interactions between Rabs 7 and 10 (Frasa et al., 2012; Yoshimura et al., 2007).

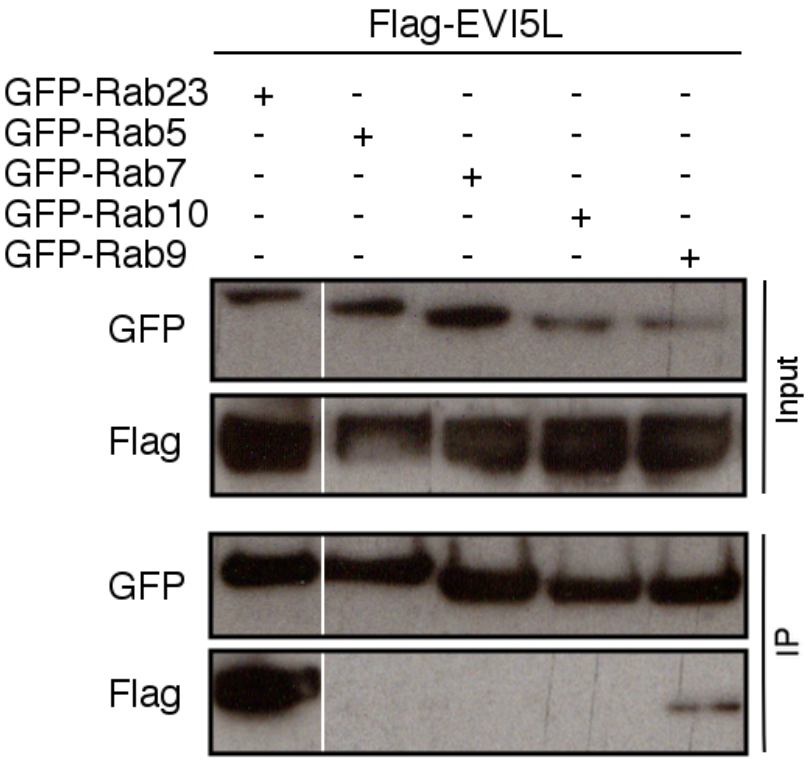

Figure 5.11 Co-immunoprecipitation of EVI5L and candidate cognate Rab proteins.

Lysates of 293ET cells co-transfected with Flag-EVI5L and GFP-tagged Rab proteins and proteins immunoprecipitated with anti-GFP antibody-conjugated beads. GFP-Rabs and Flag-EVI5L were detected in cell lysates before (input) and after (IP) immunoprecipitation by SDS-PAGE and immunoblotting.

EVI5L and Rab23 in Salmonella infection of epithelial cells

RNAi knockdown of EVI5L and Rab23 in epithelial cells

We could achieve good inhibition of EVI5L RNA expression in HeLa cells with two siRNA oligo duplexes compared to negative control delivery of scrambled negative oligo duplexes, with at least 90% knockdown for either oligo (Figure 5.12A). EVI5L RNA knockdown was maximal using 80nM concentrations of siRNA (5.12B), a concentration we used for a subsequent experiments. EVI5L knockdown was also demonstrable at the protein level, with good evidence of knockdown by immunoblot following siRNA treatment of Flag-EVI5L expressing HeLa cells (Figure 5.12C).

Similarly, we achieved reproducible knockdown (80%) of Rab23 expression with one siRNA oligo duplex compared to scrambled negative control (Figure 5.13A). We further confirmed knockdown at the protein level by co-transfecting 293ET cells with GFP-tagged Rab23 and siRNA duplexes (Figure 5.13B).

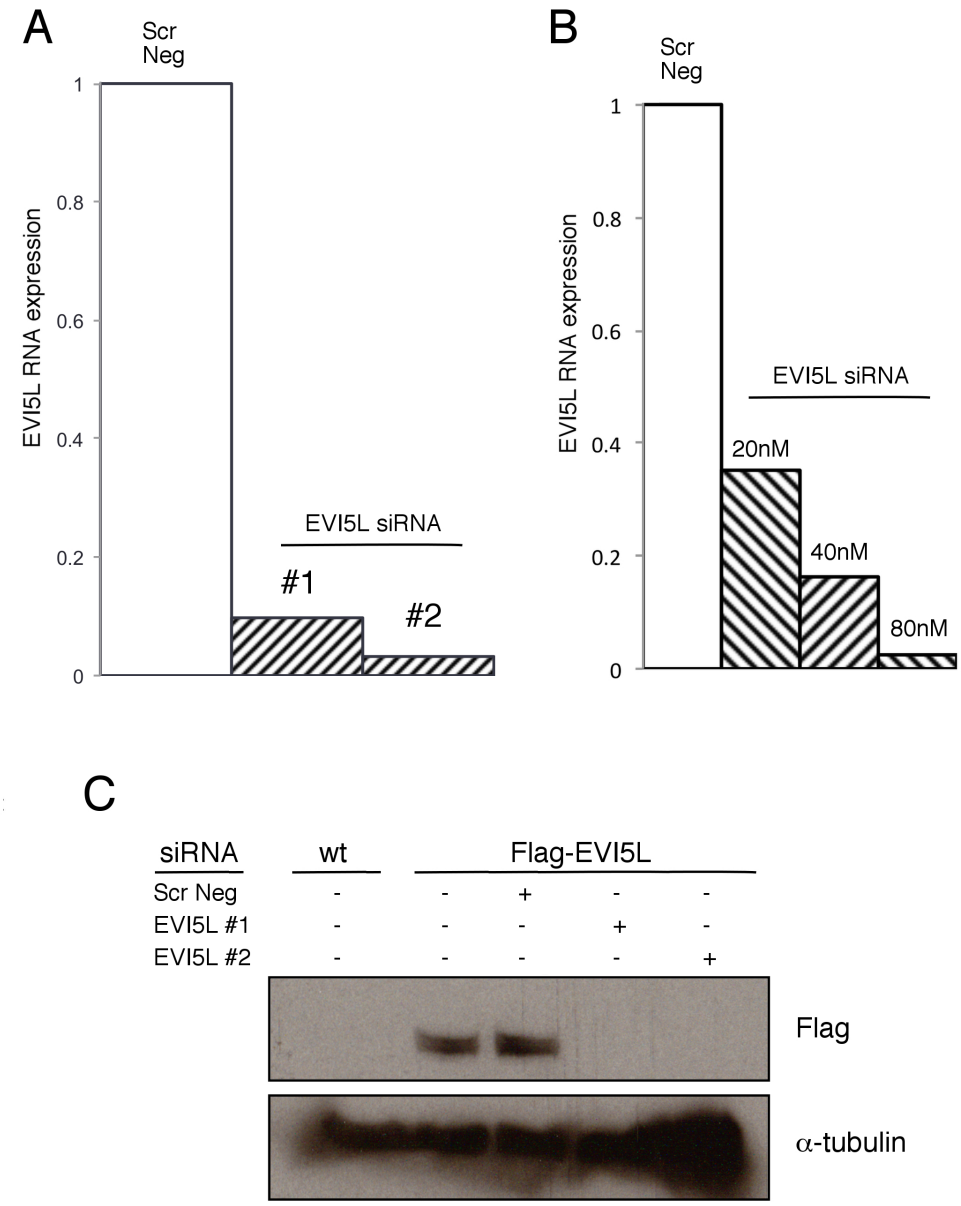

Figure 5.12 RNAi knockdown of EVI5L in epithelial (HeLa) cells.

RNA expression of *EVI5L*, assayed by qPCR relative to β-actin, following transfection of scrambled negative (Scr Neg), and two *EVI5L* siRNA oligos (A), and following transfection of scrambled negative and EVI5L siRNA oligos of increasing concentration (B). Protein levels of Flag-EVI5L, assayed by anti-Flag immunoblot, following transfection of scrambled negative, and two *EVI5L* siRNA oligos (C).

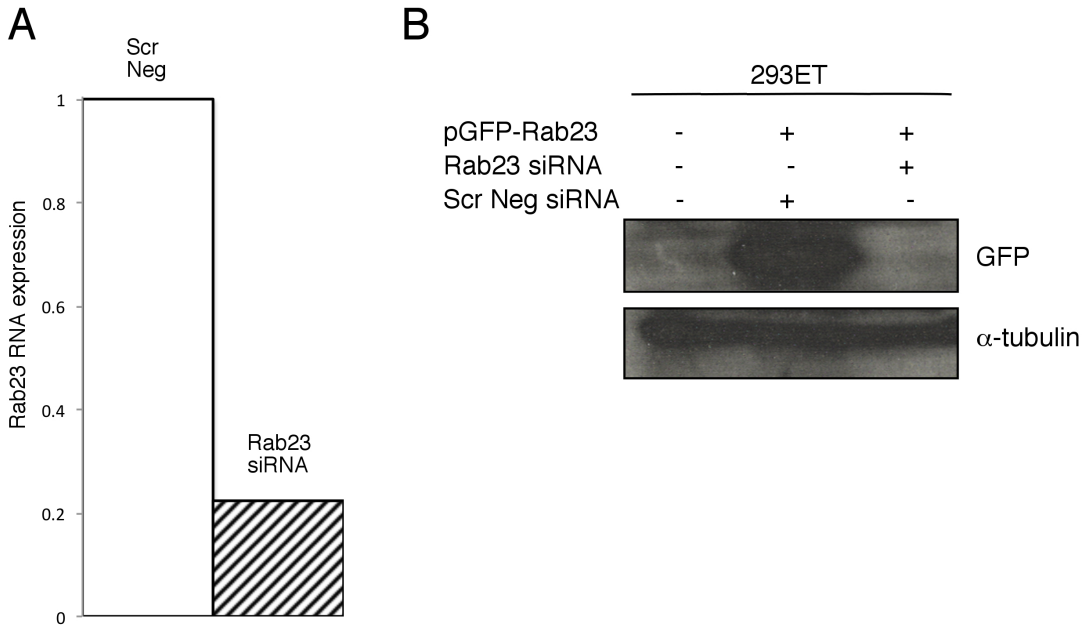

Figure 5.13 RNAi knockdown of Rab23

RNA expression of *EVI5L*, assayed by qPCR relative to β-actin, following transfection of scrambled negative (Scr Neg), and a *RAB23* siRNA oligo (A). Protein levels of GFP-EVI5L (B), assayed by anti-GFP immunoblot, following transfection co-transfection of GFP-tagged Rab23 and siRNA oligos (scrambled negative and a *RAB23* oligo).

Effect of EVI5L knockdown on Salmonella infection in epithelial cells

Following siRNA knockdown of EVI5L in HeLa cells, cells were infected with *S*. Typhimurium (wild-type 12023) and intracellular replication assayed by bacterial recovery following infected cell lysis at 2, 4, 6 and 8 hours post-infection. In these assays, EVI5L knockdown with either siRNA oligo did not have a significant effect on intracellular replication as measured by net bacterial recovery (Figure 5.14). To determine whether this lack of significant defect in EVI5L-deficient cells could be secondary to the limitations of measuring intracellular replication by net bacterial recovery, we recapitulated these experiments using flow cytometry to monitor intracellular replication

following infection with GFP-expressing *Salmonella*. Again, we observed no significant change in intracellular replication in EVI5L-deficient epithelial cells (Figure 5.15). Given the early and transient nature of the observed EVI5L co-localisation with *Salmonella* during epithelial infection (Figure 5.9), we used microscopy (data not shown) and the flow cytometric assay of *Salmonella* infection to investigate whether EVI5L-deficiency could affect *Salmonella* epithelial invasion efficiency. In these assays, EVI5L again had no significant effect on *Salmonella* invasion in epithelial cells (Figure 5.16).

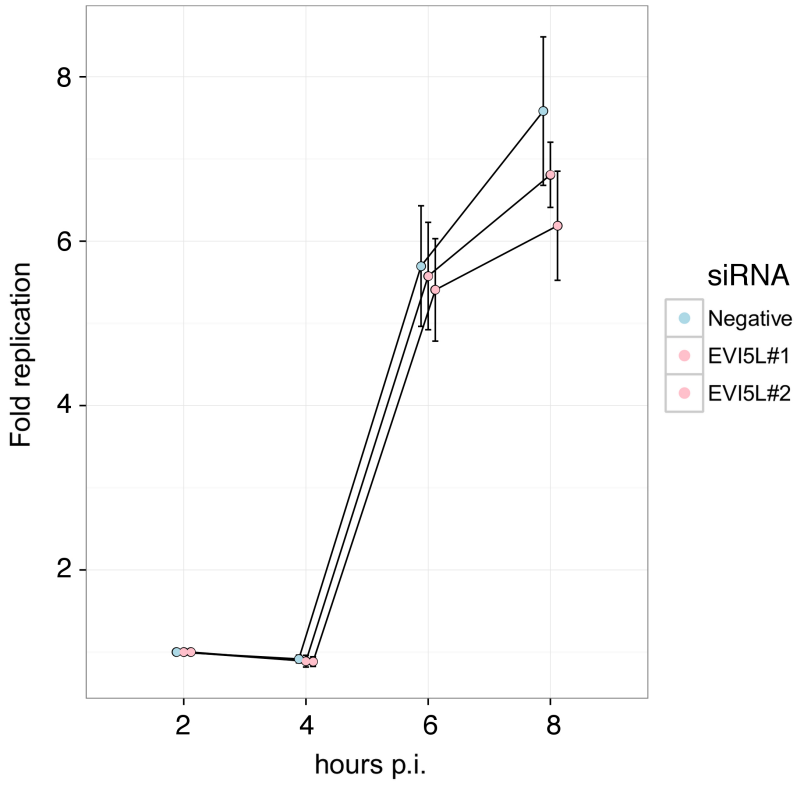

Figure 5.14 Net bacterial replication in *Salmonella*-infected epithelial cells following EVI5L RNAi

HeLa cells were infected with wild-type *S.* Typhimurium 12023, following transfection of scrambled negative (Scr Neg), and two *EVI5L* siRNA oligos. Fold replication at each time point represents net bacterial recovery relative to recovery at 2 hours post-infection. Summary statistics are calculated from three independent experiments. Error bars represent SEM. p.i., post infection.

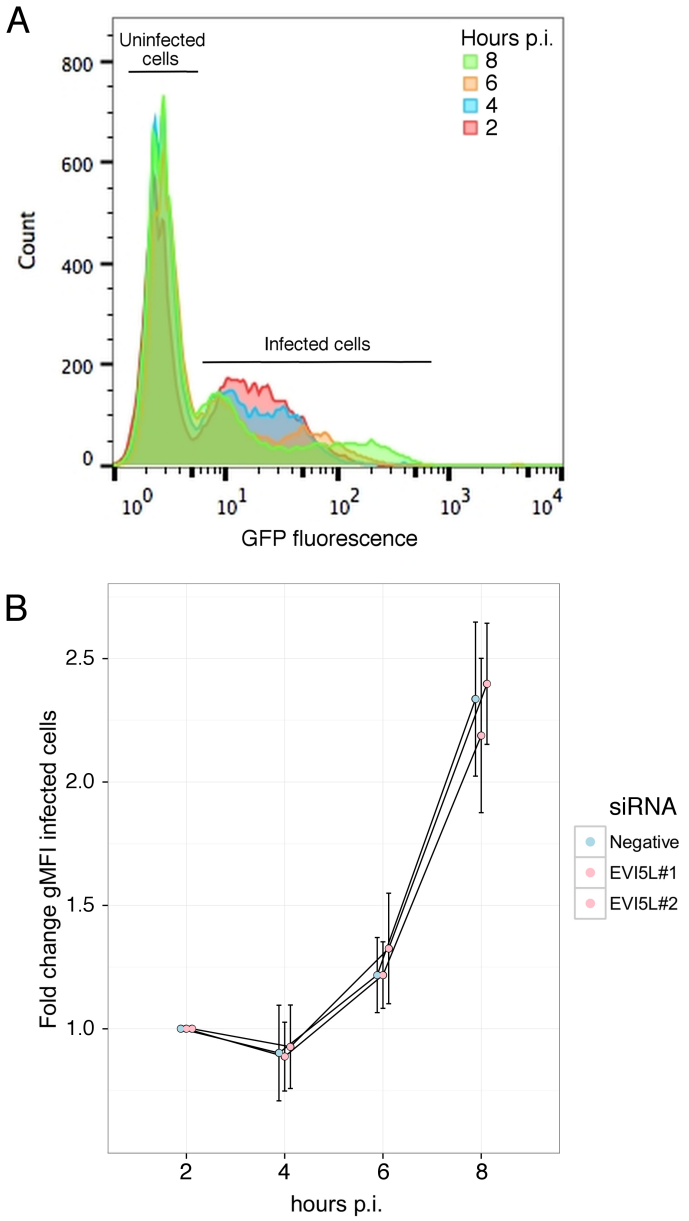

Figure 5.15 Flow cytometric determination of intracellular bacterial replication in *Salmonella*-infected epithelial cells following EVI5L RNAi

HeLa cells were infected with GFP-expressing *S.* Typhimurium 12023, following transfection of scrambled negative (Scr Neg), and two *EVI5L* siRNA oligos. Panel A depicts changes in GFP fluorescence over time in *Salmonella*-infected HeLa cells in a representative infection.
Fold intracellular replication at each time point is estimated as fold change in geometric mean fluorescence intensity (gMFI) relative to gMFI at 2 hours post-infection. Summary statistics are calculated from three independent experiments. Error bars represent SEM. p.i., post infection.

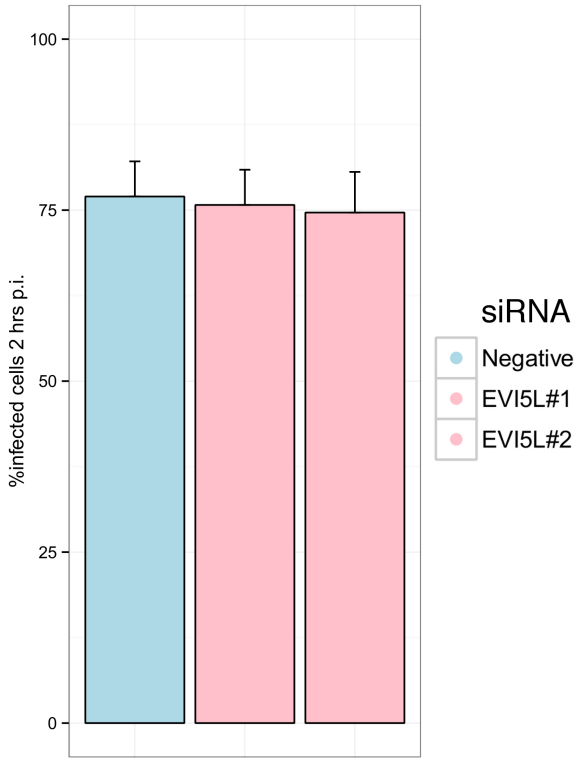

Figure 5.16 Flow cytometric determination of *Salmonella* invasion into epithelial cells following EVI5L RNAi

HeLa cells were infected with GFP-expressing *S.* Typhimurium 12023, following transfection of scrambled negative (Scr Neg), and two *EVI5L* siRNA oligos. Invasion efficiency was estimated at the proportion of infected cells (identified by GFP fluorescence) at 2 hours post-infection. Summary statistics are calculated from three independent experiments. Error bars represent SEM. p.i., post infection.

Effect of Rab23 knockdown on Salmonella infection in epithelial cells

Given the observed interaction between EVI5L and Rab23 we investigated whether Rab23 knockdown would have phenotypic effects during *Salmonella*-infection of epithelial cells. Following siRNA knockdown of Rab23 in HeLa cells, cells were infected with *S*. Typhimurium and intracellular replication

assayed by net bacterial recovery and flow cytometry. In these assays, Rab23 knockdown did not have a significant effect on intracellular replication as measured by net bacterial recovery (Figure 5.17), or by flow cytometry (Figure 5.18). Similarly, there was no evidence of a phenotype for Rab23 in *Salmonella* invasion of epithelial cells (Figure 5.18).

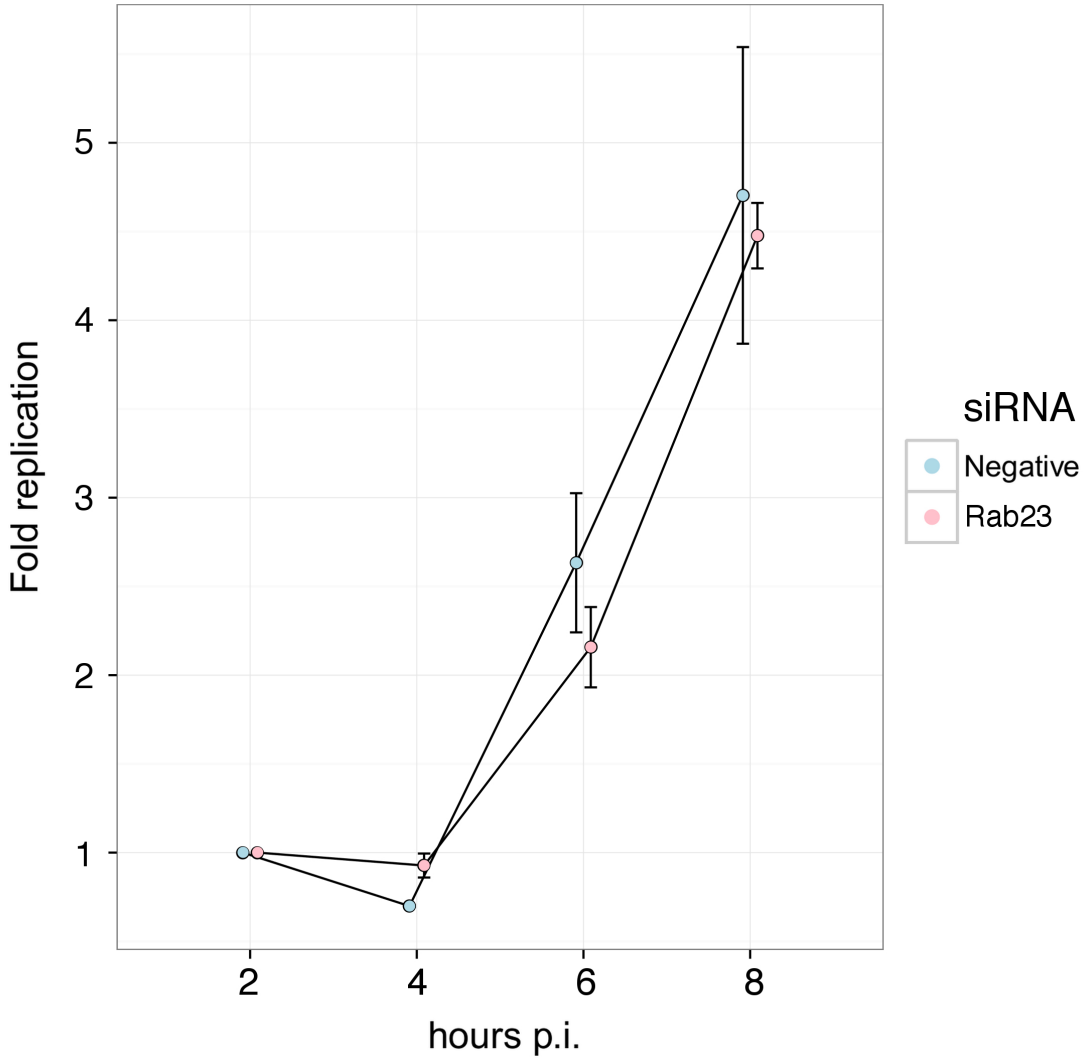

Figure 5.17 Net bacterial replication in *Salmonella*-infected epithelial cells following Rab23 RNAi

HeLa cells were infected with wild-type *S.* Typhimurium 12023, following transfection of scrambled negative (Scr Neg), and *RAB23* siRNA oligos. Fold replication at each time point represents net bacterial recovery relative to recovery at 2 hours post-infection. Summary statistics are calculated from three independent experiments. Error bars represent SEM. p.i., post infection.

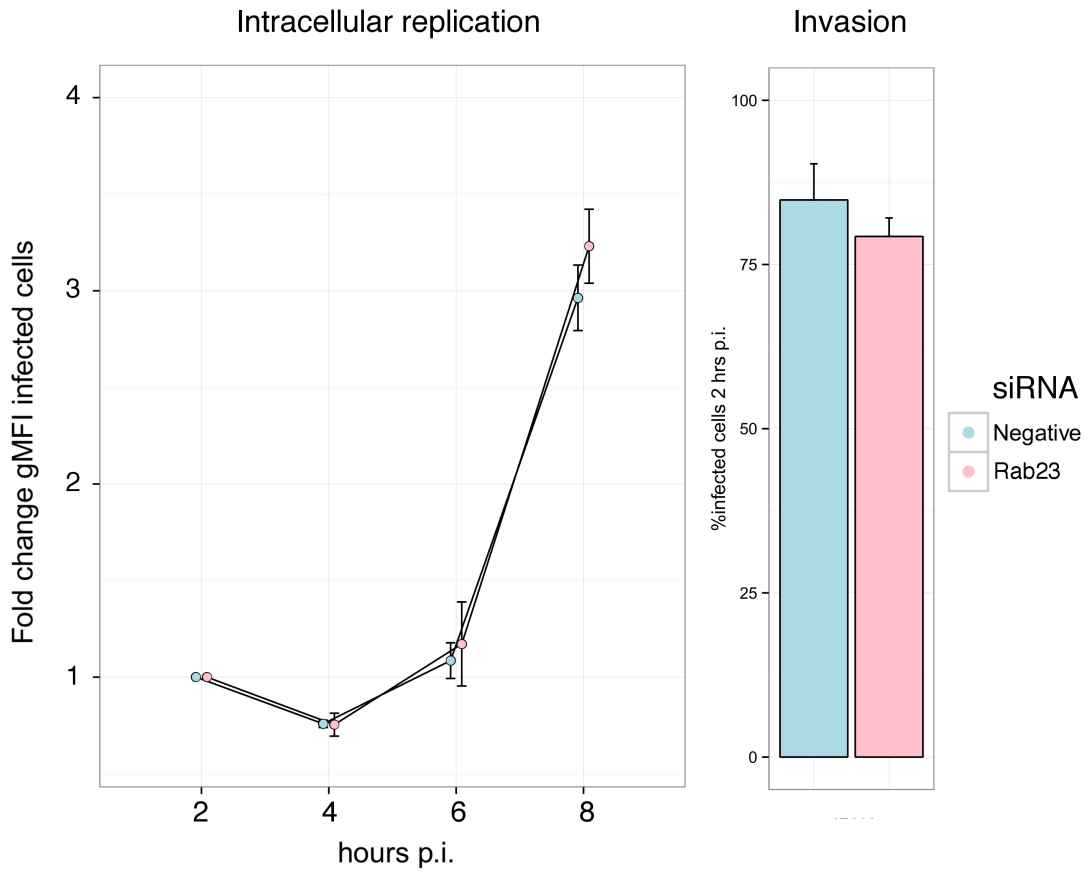

Figure 5.18 Flow cytometric determination of bacterial invasion and replication in *Salmonella*-infected epithelial cells following Rab23 RNAi

HeLa cells were infected with GFP-expressing *S.* Typhimurium 12023, following transfection of scrambled negative (Scr Neg), and *Rab23* siRNA oligos. Fold intracellular replication at each time point is estimated as fold change in geometric mean fluorescence intensity (gMFI) relative to gMFI at 2 hours post-infection (left). Invasion efficiency was estimated at the proportion of infected cells (identified by GFP fluorescence) at 2 hours post-infection (right). Summary statistics are calculated from three independent experiments. Error bars represent SEM. p.i., post infection.

Effect of EVI5L knockdown on epithelial cell infection with ΔinvA/Inv Salmonella

Rab23 has been shown to associate with mature phagolysosomes (Smith et al., 2007) during epithelial infection SPI-1 deficient *Salmonellae*, which fail to

replicate intracellularly. We used the SPI-1 deficient mutant *ΔinvA*/Inv, to assess whether EVI5L could modify *Salmonella* survival and replication in mature phagosomes. As has been previously described, the *ΔinvA*/Inv mutant derivative of *S*. Typhimurium 12023 is able to efficiently invade cultured epithelial cells (HeLas), but fails to replicate intracellularly (Figure 5.19).

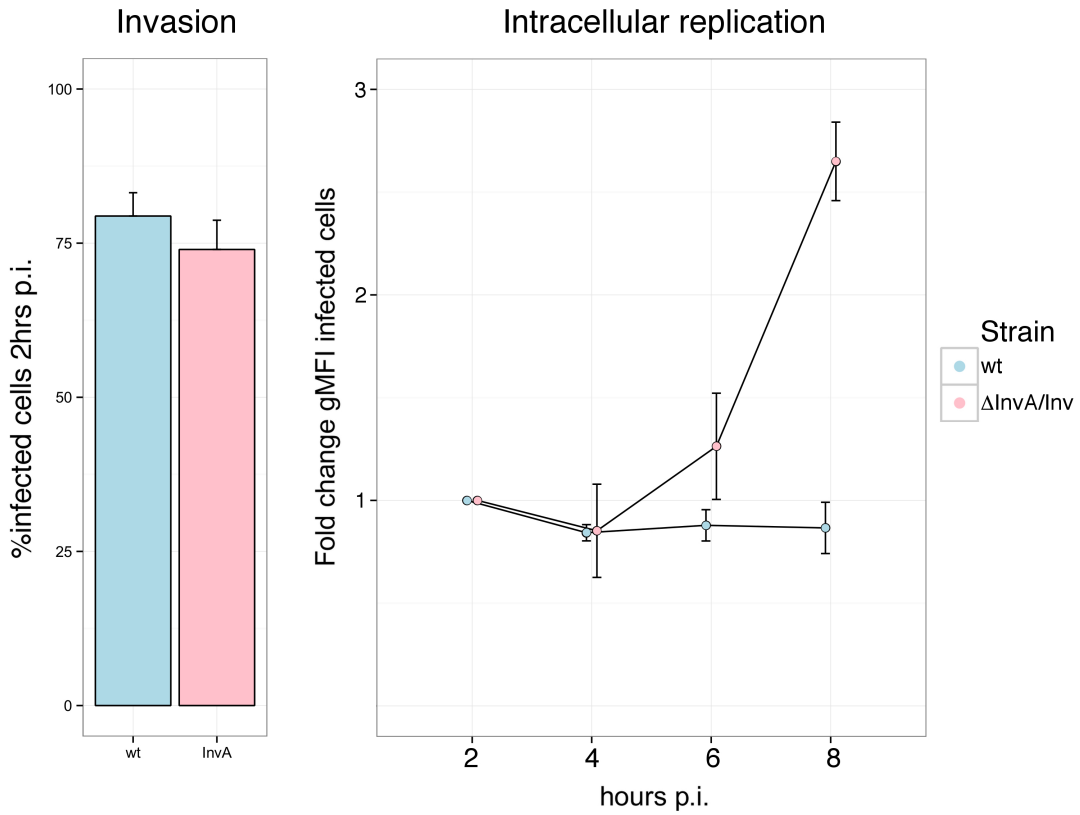

Figure 5.19 Invasion and replication of SPI-1-deficient *Salmonella* in epithelial cells.

HeLa cells were infected with GFP-expressing wild-type *S.* Typhimurium 12023 and its mutant derivative *ΔinvA*/Inv. Fold intracellular replication at each time point is estimated as fold change in geometric mean fluorescence intensity (gMFI) relative to gMFI at 2 hours post-infection (right). Invasion efficiency was estimated at the proportion of infected cells (identified by GFP fluorescence) at 2 hours post-infection (left). Summary statistics are calculated from three independent experiments. Error bars represent SEM. p.i., post infection.

Following siRNA knockdown of EVI5L in HeLa cells, cells were infected with wild-type and *ΔinvA*/Inv *S*. Typhimurium and invasion and intracellular

replication assayed by flow cytometry. In these assays, EVI5L knockdown did not have a significant effect on invasion or intracellular replication as measured by flow cytometry (Figure 5.20).

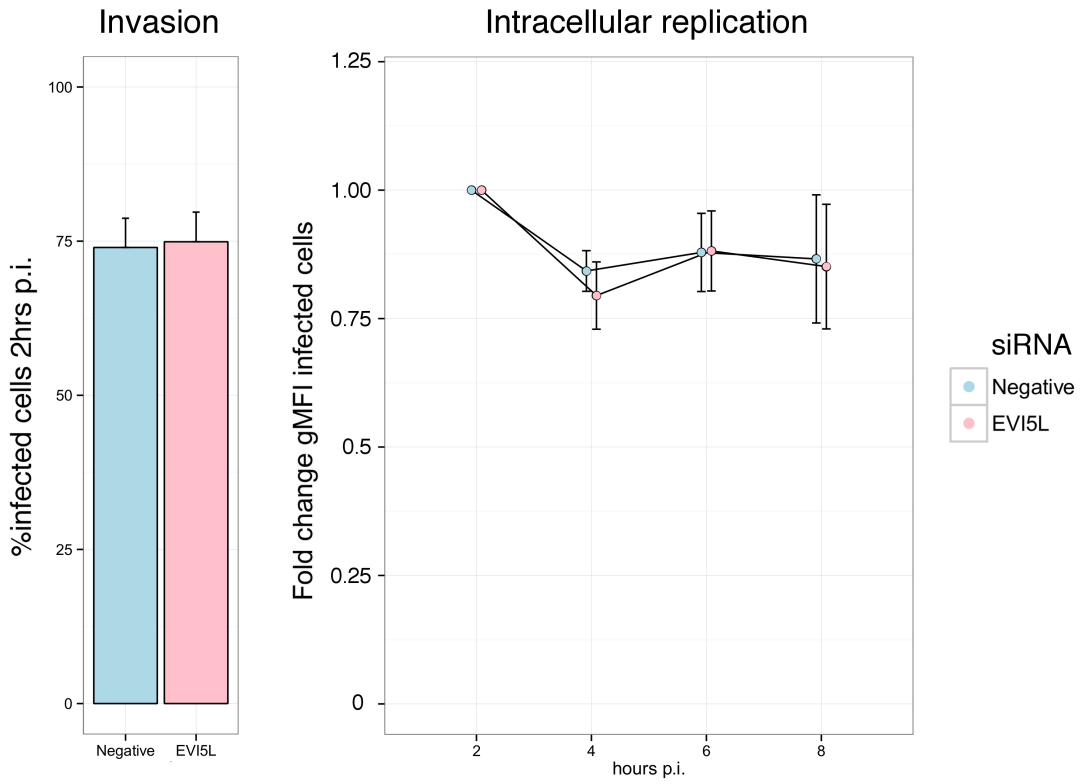

Figure 5.20 Invasion and replication of SPI-1-deficient *Salmonella* in epithelial cells

HeLa cells were infected with GFP-expressing *ΔinvA*/Inv *S.* Typhimurium 12023, following transfection of scrambled negative (Scr Neg), and *EVI5L* siRNA oligos. Fold intracellular replication at each time point is estimated as fold change in geometric mean fluorescence intensity (gMFI) relative to gMFI at 2 hours post-infection (right). Invasion efficiency was estimated at the proportion of infected cells (identified by GFP fluorescence) at 2 hours post-infection (left). Summary statistics are calculated from three independent experiments. Error bars represent SEM. p.i., post infection.

EVI5L in Salmonella infection of phagocytic cells

The lack of an identifiable phenotype for EVI5L in *Salmonella* infection of epithelial cells led us to investigate the role of EVI5L in the context of phagocytic infection. A role for EVI5L in this context appears plausible given the enrichment of we observe between SNPs increasing NTS disease risk and EVI5L eSNPs in monocytes.

RNAi knockdown of EVI5L in phagocytic cells

We achieved stable inhibition of EVI5L RNA expression in murine iBMDMs using a puromycin-selectable EVI5L shRNA, transduced using a lentiviral vector. In comparison with iBMDMs transduced with, and stably expressing, a scrambled negative control shRNA, EVI5L RNA expression was reduced by 80% (Figure 5.21).

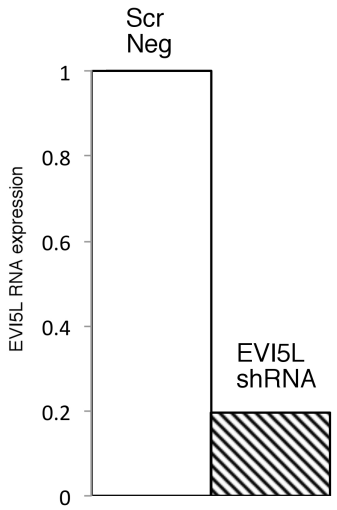

Figure 5.21 RNAi knockdown of EVI5L in phagocytic (iBMDM) cells

RNA expression of *EVI5L*, assayed by qPCR relative to m18s, following transduction and selection of cells expressing scrambled negative (Scr Neg) *EVI5L* shRNAs.

Effect of EVI5L knockdown on Salmonella infection in phagocytic cells

Following selection of iBMDMs stably expressing shRNAs, cells were infected with GFP-expressing *S*. Typhimurium (wild-type 12023) and invasion and intracellular replication assayed by flow cytometry at 2, 4, 6, 8, 10 and 16 hours post-infection. In these experiments, EVI5L knockdown did not have a significant effect on intracellular replication in phagocytes (Figure 5.22). Similarly, there was no evidence of a phenotype for EVI5L in *Salmonella* invasion of phagocytic cells (Figure 5.22).

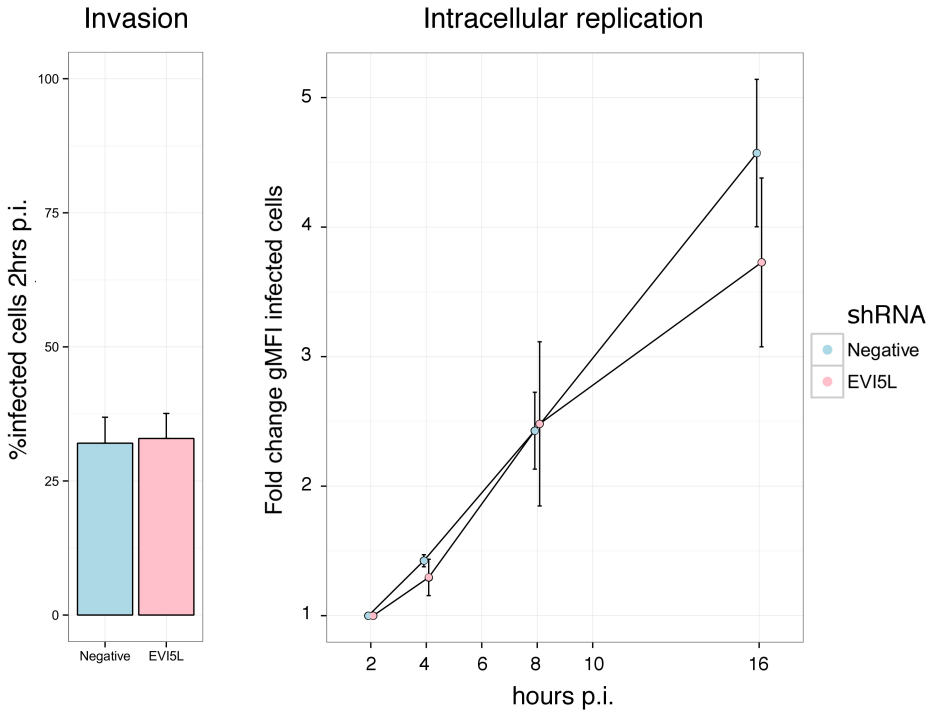

Figure 5.22 Invasion and replication of *Salmonella* in EVI5L RNAi-treated phagocytic cells.

iBMDMs stably expressing scrambled negative and EVI5L shRNAs, were infected with GFP-expressing *S*. Typhimurium 12023. Fold intracellular replication at each time point is estimated as fold change in geometric mean fluorescence intensity (gMFI) relative to gMFI at 2 hours post-infection (right). Invasion efficiency was estimated at the proportion of infected cells (identified by GFP fluorescence) at 2 hours post-infection (left). Summary statistics are calculated from three independent experiments. Error bars represent SEM. p.i., post infection.

To assess whether EVI5L could modify *Salmonella*-induced cytotoxicity in iBMDMs, we assayed LDH release at regular time-points to 16 hours post-infection. In these experiments, LDH release in infected iBMDMs was not significantly affected by EVI5L (Figure 5.23).

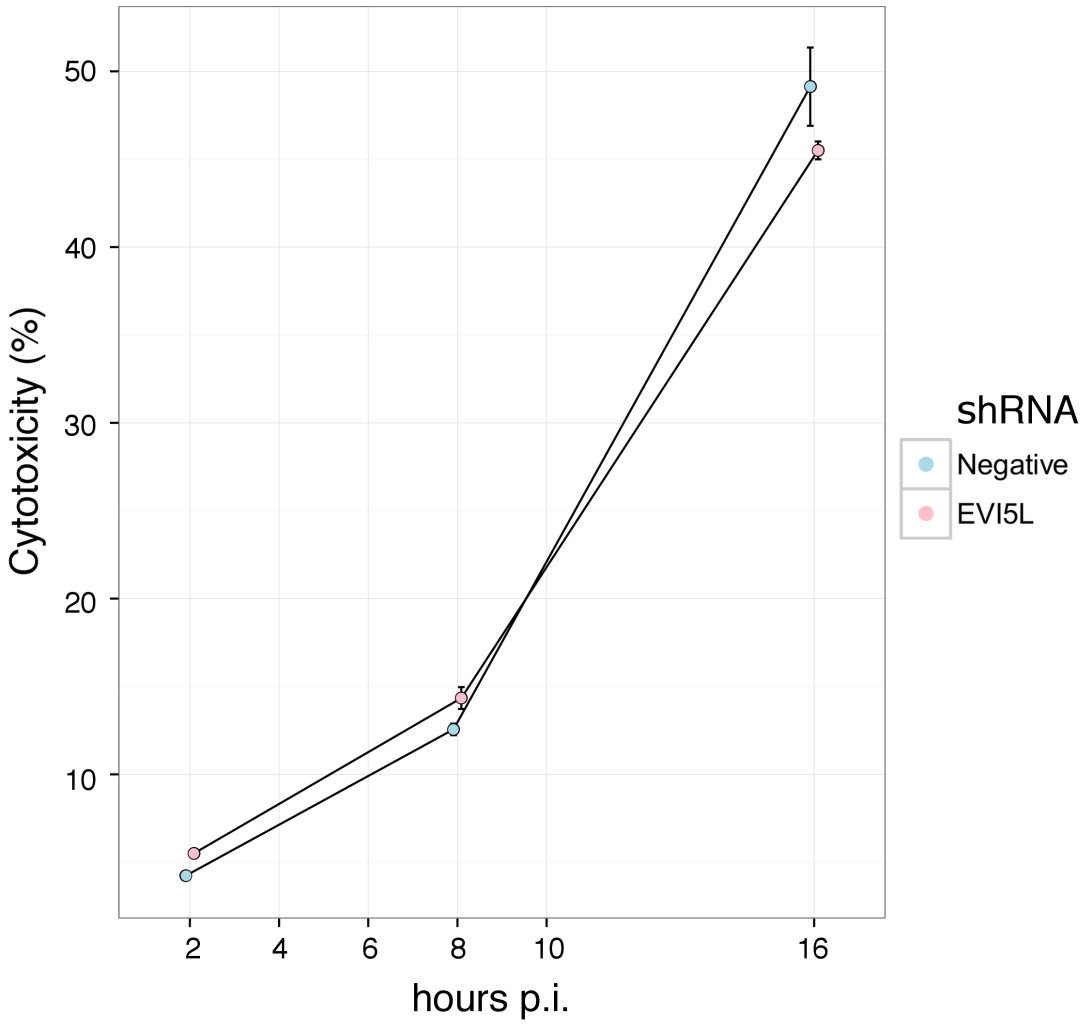

Figure 5. 23 Cytotoxicity of *Salmonella*-infected macrophages following EVI5L RNAi.

iBMDMs expressing scrambled negative and EVI5L shRNAs, were infected with wild-type *S.* Typhimurium 12023. Percentage cytotoxicity was calculated from assayed LDH-release into culture supernatants. Summary statistics are calculated from three independent experiments. Error bars represent SEM. p.i., post infection.

Enrichment for genetic correlates of cellular traits in the NTS bacteraemia GWAS

Enrichment analysis

Enrichment analysis of NTS GWAS associated SNPs and SNPs associated with IFNγ production in *Chlamydia*-stimulated lymphoblastoid cell lines demonstrated highly significant levels of enrichment between the two traits (Table 5.2, Fisher's Exact $P=2.74 \times 10^{-6}$; fold-enrichment c.f. null = 2.73x). Well-imputed SNPs shared between both studies were included in the analysis (n = 7,592,495). Interestingly, that enrichment is entirely driven by SNPs in the *EVI5L* locus, with 19 of 21 SNPs associated with both traits being located in *cis* to *EVI5L*.

Table 5.2 Enrichment of NTS- and *Chlamydia*-induced IFNγ-associated SNPs

		NTS GWAS	
		No association	*P*<0.001
Chamydia-induced IFNγ	No association	7,592,495	7,690
	P<0.001	7,564	21

There is highly significant enrichment of EVI5L *cis* eSNPs (FDR-adjusted *P*-value<0.05) among *Chlamydia*-stimulated IFNγ production associated-SNPs (P<0.001) within 1Mb of *EVI5L* in the Kenyan discovery GWAS samples. This enrichment is observed in naïve and stimulated monocytes (Table 5.3).

Table 5. 3 Enrichment of EVI5L eSNPs in naïve and stimulated monocytes among *Chlamydia*-stimulated IFNγ production associated-SNPs

Stimulation condition	*P*-value	Fold enrichment
Naïve	2.33×10^{-9}	9,3
LPS - 2 hours	1.73×10^{-6}	14.6
LPS - 24 hours	5.82×10^{-10}	13.1
IFNγ - 24 hours	7.62×10^{-9}	10.1

P-values are calculated with Fisher's exact test. Fold enrichment is the ratio of observed shared SNPs observed between EVI5L eSNPs and IFNγ production associated-SNPs and that expected under a χ^2 distribution with one degree of freedom.

To better understand the relationship between EVI5L expression levels and IFNγ production, we compared the observed effect size (and direction of effect) at SNPs found to be associated with *Chlamydia*-stimulated IFNγ production ($P<1 \times 10^{-3}$), and also to be significant EVI5L eSNPs (FDR-adjusted $P<0.05$). In keeping with the relationship between EVI5L expression and NTS disease risk, the relationship between EVI5L expression and IFNγ production is complex (Figure 5.24), with the relationship between EVI5L expression and IFNγ production in monocytes being dependent on stimulation context.

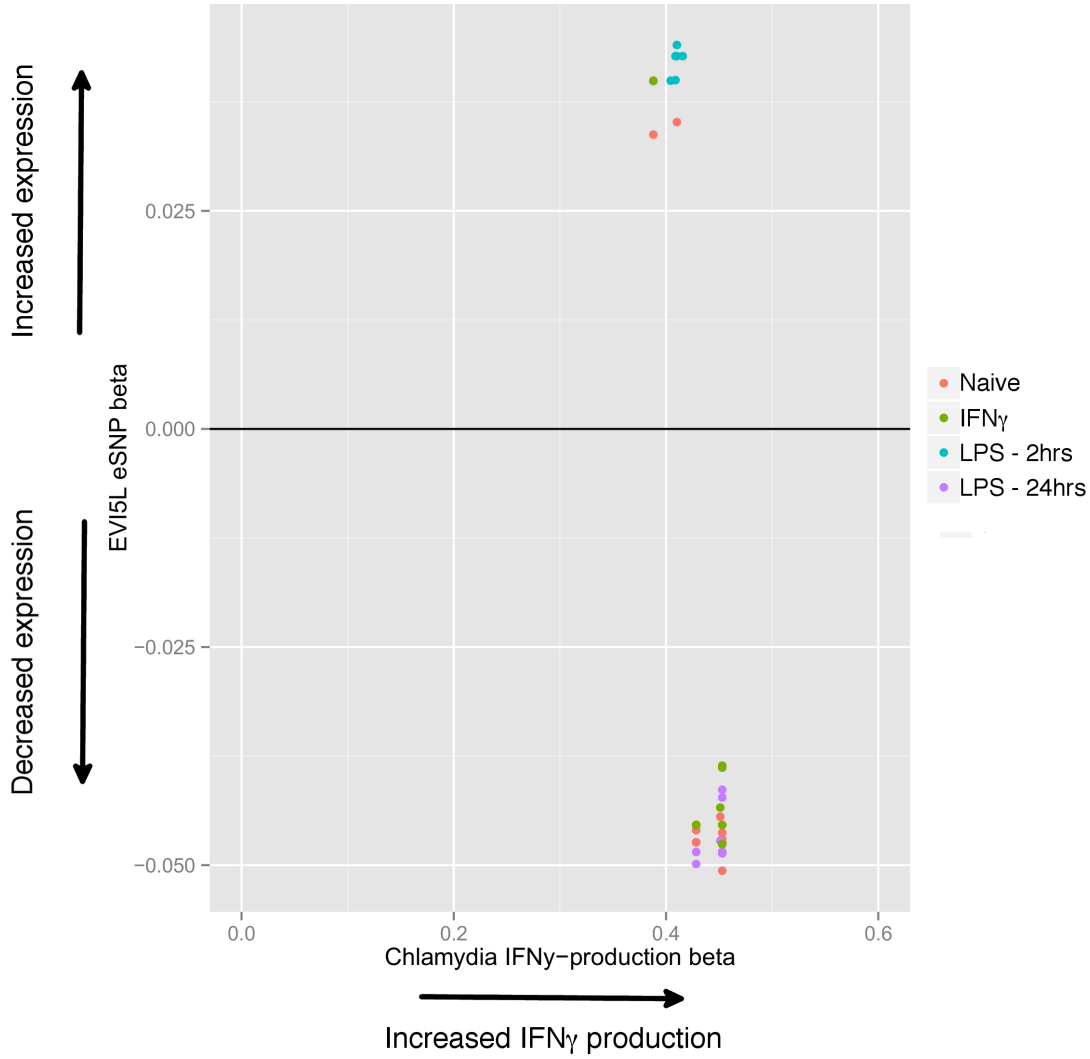

Figure 5. 24 Effect size and direction of shared *Chlamydia*-stimulated IFNγ production loci and *EVI5L* eSNPs in monocytes

All plotted SNPs are associated with IFNγ production (P=1x10^{-3}), and are significant *EVI5L* eSNPs (FDR-adjusted P<0.05) for monocytes in at least one stimulation condition (see colour-coding).

IFNγ-related protein phenotypes of EVI5L eQTLs

Data from 55 healthy European adults were included in the analysis. In those individuals, genotypes at 22 EVI5L *cis* eSNPs (FDR-adjusted *P*-value<0.05) were correlated with STAT4 phosphorylation and IFNγ production following IL-

12 and NTS stimulation (Table 5.3). In that analysis, one SNP (rs13306446, chr19: 7129064), an eSNP for EVI5L expression in naïve monocytes, is associated with the proportion of pSTAT4$^+$ CD4$^+$ T cells (P = 1.37x10^{-3}) following IL-12 stimulation and with the proportion of IFNγ$^+$ CD4$^+$ T cells (P = 0.016) following NTS stimulation (Figure 5.25).

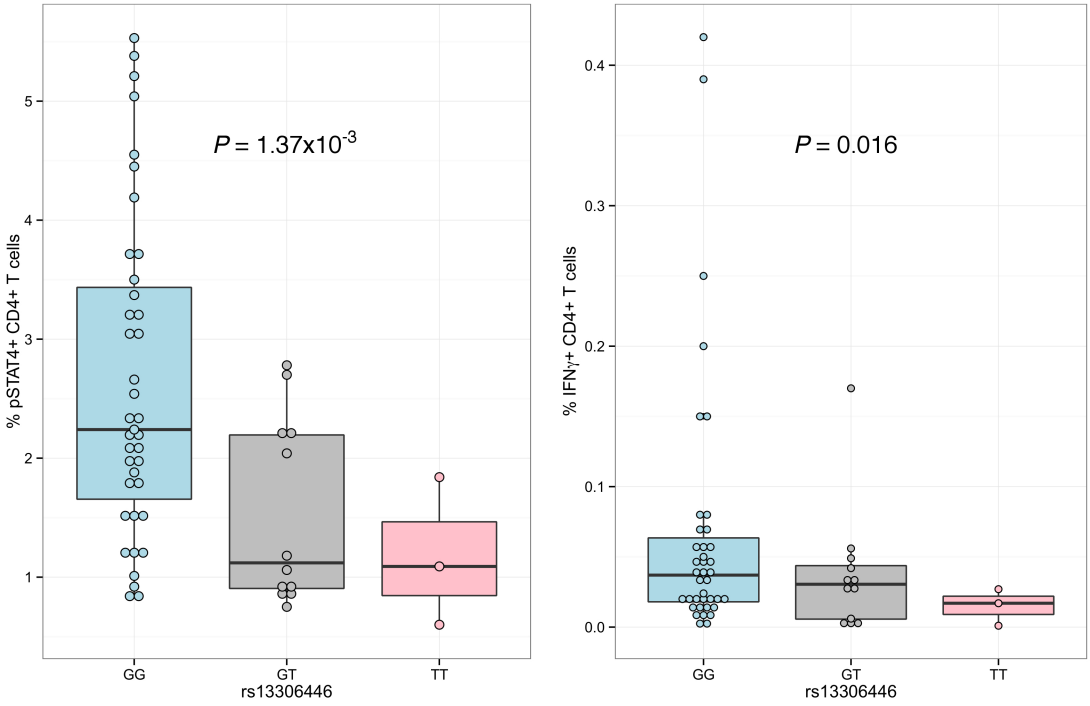

Figure 5. 25 Effect of rs13306446 genotype on STAT4 phosphorylation and IFNγ production in CD4$^+$ T cells.

Genotype-specific pSTAT4$^+$ cell proportions in IL-12 stimulated in CD4$^+$ T cells (A). Genotype-specific IFNγ$^+$ cell proportions in NTS-stimulated in CD4$^+$ T cells (B). pSTAT4, phosphorylated STAT4.

The same locus, rs13306446, is associated with the proportion of IFNγ$^+$ NK cells (P = 0.028) following NTS stimulation, but not with the proportion of pSTAT4$^+$ NK cells (P = 0.714) following IL-12 stimulation (Figure 5.26). The rs13306446:T allele is associated with reduced STAT4 phosphorylation, and

with reduced IFNγ production, and is associated with increased EVI5L RNA expression in naïve monocytes (Figure 5.27).

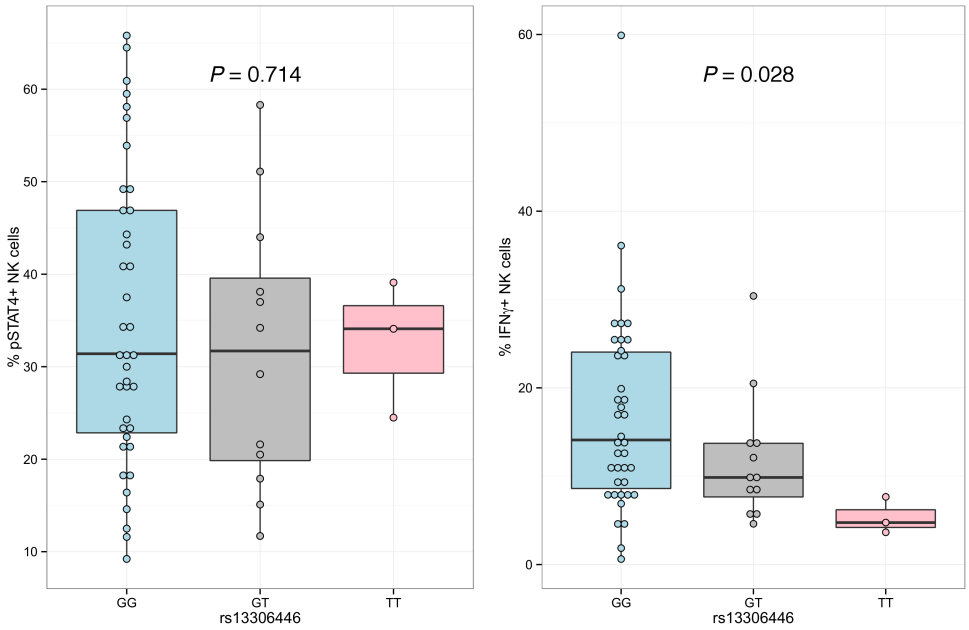

Figure 5.26 Effect of rs13306446 genotype on STAT4 phosphorylation and IFNγ production in NK cells.

Genotype-specific pSTAT4^{+} cell proportions in IL-12 stimulated NK cells (A). Genotype-specific IFNγ^{+} cell proportions in NTS-stimulated NK cells (B). pSTAT4, phosphorylated STAT4.

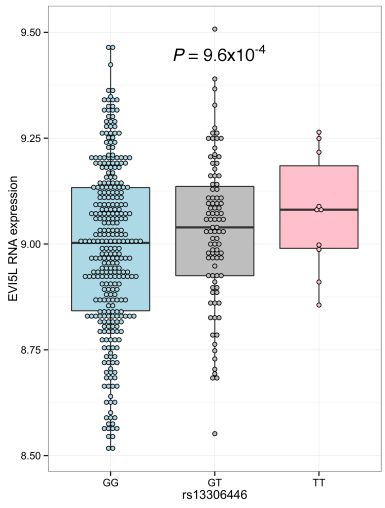

Figure 5.27 Effect of rs13306446 genotype on EVI5L RNA expression in naïve monocytes.

Genotype-phenotype correlations are analysed by regression and analysis of variance. *P*-values are calculated by *F*-tests with one degree of freedom.

Conclusions

In this chapter, we selected EVI5L from the NTS bacteraemia GWAS discovery analysis for further detailed characterisation as a candidate NTS-susceptibility locus. Our unsuccessful attempts to replicate a genetic association at this locus, in either the Kenyan or Malawian replication samples, were substantially underpowered, and we thus sought alternative means to characterise and validate the locus.

Mapping of *cis* eQTLs for *EVI5L* in immune cell subsets was informative, lending further support to the plausibility of the locus as a determinant of NTS susceptibility. This analysis revealed significant cis eQTLs in naïve and stimulated monocytes, with evidence for both shared and context-specific eQTLs between stimulation conditions. NTS-associated genetic variation at *EVI5L* is highly enriched for *EVI5L* monocyte eSNPs. This enrichment is seen in naïve and stimulated monocytes, despite eQTLs between certain stimulation conditions being independent, e.g. naïve and monocytes stimulated with LPS for 2 hours. Interestingly, this overlap between cis *EVI5L* eSNPs and NTS-associated SNPs does not facilitate an answer as to whether increased or decreased EVI5L activity is associated with susceptibility to NTS disease. NTS-associated alleles increase *EVI5L* expression in naïve and LPS (2 hours) stimulated monocytes, and decrease *EVI5L* expression in monocytes stimulated for 24 hours (with IFNγ or LPS).

The identification of *EVI5L* eSNPs as determinants of NTS bacteraemia risk increases the likelihood that the NTS association is robust, and also defines EVI5L as the mediator of that association. However, despite promising initial results demonstrating co-localisation of EVI5L with *Salmonella* at the early stages of intracellular infection, we were unable to identify a robust phenotype for EVI5L in the context of *Salmonella* infection.

In parallel to these efforts, an enrichment analysis of genetic correlates of infection-associated cellular traits among NTS-associated SNPs revealed a highly significant enrichment of SNPs correlated with IFNγ production. This enrichment was entirely driven by SNPs in the *EVI5L* region, and in common with NTS-associated SNPs, genetic correlates of IFNγ production are highly enriched for *EVI5L cis* eQTLs. We expand on this observation utilising samples collected from a genotype-selectable bioresource to demonstrate that, in healthy European adults, an eSNP for *EVI5L* in naïve monocytes is a significant determinant of both STAT4 phosphorylation and IFNγ protein production in $CD4^+$ T cells following stimulation with IL-12 and NTS respectively.

Taken together, these data suggest that our failure to define a phenotype for EVI5L using *in vitro* models of epithelial and phagocytic *Salmonella* infection was secondary to a mistaken focus on the infected cell as the likely site of EVI5L function. In identifying EVI5L as a correlate of IFNγ production, we highlight a novel function of this RABGAP, and lend further weight to the notion that IFNγ production capacity, as mediated by genetic variation in

STAT4 and in *EVI5L*, is a key determinant of susceptibility to invasive NTS disease in African children.

Chapter 6 – Pathway analysis of nontyphoidal *Salmonella* bacteraemia susceptibility

Background

Validated, common genetic variation associated with most complex traits accounts for a small proportion of the estimated heritability of a given phenotype. Consortia-based efforts to increase sample sizes well beyond those possible to recruit at single centres, alongside techniques to adequately assay rare genetic variation within those studies, will continue to provide incremental insights into genetic underpinnings of many diseases. Much of the missing heritability will be accounted for by common loci of small effect. Identification of loci with increasingly small effect sizes will continue to be challenging within the constraints of feasible sample sizes and analytical approaches operating at the single locus level, performed under the assumption that each tested variant has an equal prior probability of disease association.

The clustering of sub-genome wide significant trait associations within a common biological pathway, is one clear example of a situation in which statistically up-weighting a group of highly biologically plausible genetic variants would be desirable. Indeed, pathway-based enrichment analyses have provided important, additive biological insights when applied to genome-wide association studies for several complex diseases (Simonson, McQueen, & Keller, 2014; L. Yang et al., 2013). Motivated by this, we performed a pathway-based enrichment analysis within the NTS bacteraemia GWAS using INRICH (P. H. Lee, O'Dushlaine, Thomas, & Purcell, 2012a). INRICH tests for overlap between trait-associated genic intervals and biologically meaningful gene sets (e.g. sets of genes encoding proteins acting within a common biological pathway). In combining an interval-based approach (as opposed to enrichment at the SNP level) and a permutation-based method to calculate enrichment statistics, INRICH has several desirable properties. Most notably, preservation of the genic interval size for random re-sampling makes it robust to confounding secondary to gene size (enrichment otherwise favouring large genes) and clustering of biologically-related genes at a single chromosomal locus. It should be noted that this hypothesis-free analysis is distinct from the hypothesis-driven enrichment analyses I present in chapter 4.

In this chapter, I present the results of the pathway enrichment analysis for the NTS bacteraemia GWAS. This analysis identifies significant enrichment of a single gene set, implicating genetic variation in methionine metabolism as a key determinant of NTS susceptibility in African children. This finding fits well with our emerging understanding of *Salmonella* host-pathogen interaction,

with the NTS-associated metabolic pathway having been previously implicated in *Salmonella*-induced pyroptosis, and in mortality in patients with sepsis. We further characterise the role of genetic variation at these loci in host susceptibility to bacteraemia in African children secondary to pathogens other than NTS, and their role in determining risk of mortality in these children.

Salmonella metabolises methionine and its derivatives in a manner analogous to mammalian cells. In light of the identification of this shared metabolic pathway as a correlate of susceptibility to invasive NTS disease, alongside previous data implicating flux through this pathway as a determinant for pyroptosis in *Salmonella*-infected cells, we hypothesized that *Salmonella* metabolism during intracellular infection could act as a determinant of infected cell death. We test this, using the λ-red recombinase system to generate *Salmonella* mutants deficient in key genes controlling bacterial methionine metabolism, and characterising the phenotype of these mutants in the context of intracellular infection.

Methods

Pathway enrichment analysis of NTS bacteraemia GWAS

To test for the enrichment of biologically related genes at NTS-associated loci we used INRICH, an interval enrichment analysis tool (P. H. Lee et al., 2012a). INRICH tests for overlap of trait-associated, LD-defined, genomic intervals with genomic locations in a gene set. Firstly, INRICH tests for overlap with each gene set and a user-defined number of randomly generated genetic intervals and the trait-associated set of intervals to generate an empirical enrichment *P* value. Secondly, INRICH corrects for multiple testing across the number of gene sets tested within the selected database with a second user-defined number of permutations.

We defined NTS-associated regions in PLINK (Purcell et al., 2007), as non-overlapping genomic regions containing an NTS-associated SNP ($P<1x10^{-5}$ in either an additive or genotypic model), bounded in the 3' and 5' directions by the extent of linkage disequilibrium ($r^2>0.2$) to the most-associated SNP. Sample inclusion criteria were implemented as for the NTS bacteraemia GWAS, resulting in 180 NTS case and 2,677 control samples being included in the analysis. We included common (minor allele frequency > 5%), well-imputed (info > 0.8) autosomal SNPs without evidence of extreme deviation from Hardy-Weinberg equilibrium ($P > 1x10^{-10}$) in the analysis.
We identified 73 independent NTS-associated genomic regions, which we tested for enrichment in 1,320 gene sets in the curated canonical pathway

gene set in GSEA (Subramanian et al., 2005). Empirical *P*-values for enrichment were calculated with 100,000 replicates, and multiple-testing correction of *P*-values performed with 10,000 bootstraps.

Bayesian comparison of models of association in major bacterial pathogens in Kenyan children

We compared models of association at NTS-associated loci implicated in the pathway enrichment analysis, across the seven most frequently isolated bacterial pathogens (NTS, *S. pneumoniae, S. aureus*, *H. influenza* type b, β-haemolytic *Streptococci*, *E. coli*, *Acinetobacter*) among cases of bacteraemia in the Kenyan discovery samples in the all-cause bacteraemia GWAS. Effect size estimates and 95% confidence intervals were calculated by multinomial logistic regression under the most probable model of association observed in the NTS bacteraemia GWAS, using control status and each of the bacterial pathogen subgroups as strata, and the first two principal components as covariates. We considered four 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(0,1) and fixed between pathogens (ρ=1).
REL: effect size ~ N(0,1) and correlated (ρ=0.96), but not fixed, between pathogens.
NTS: effect size ~ N(0,1) for NTS and is zero for all other pathogens.

For each model we calculated approximate Bayes factors (Wakefield, 2009) and posterior probabilities, assuming each model to be equally likely *a priori*. Statistical analysis was performed in R, with code supplied by Dr Matti Pirinen, University of Helsinki.

Genome-wide association study of mortality in Kenyan children with all-cause bacteraemia

We performed a case-control, GWAS of inpatient mortality among Kenyan children with bacteraemia secondary to any pathogen within the all-cause Kenyan bacteraemia GWAS sample collection. Recruitment to the study is as described in chapter 3. Samples with culture-proven bacteraemia were included in the analysis, with case samples defined as children with documented inpatient death during the hospital admission including the episode of bacteraemia, and control samples defined as children documented as having survived to discharge form hospital. Children documented as having left hospital before being medically discharged, or without documented status at discharge, were excluded from the analysis. Among 1,885 children with bacteraemia recruited to the study, 1,785 children had adequately documented survival status, with 446 cases of mortality and 1,339 survivors, a mortality rate of 25%. Demographics of children included in the analysis are presented in Table 6.1.

Table 6.1 Demographic data for Kenyan children included in the analysis of mortality secondary to bacteraemia

		Kenya discovery	
		Died	**Survived**
	Numbers	446	1,339
	Median age in months (IQR)	11 (1-30)	15 (6-39)
	Females	49.33%	41.75%
Ethnicity	Giriama	56.34%	53.54%
	Chonyi	23.73%	25.25%
	Kauma	7.37%	6.06%

Sample genotyping and genome-wide imputation were performed as described for the NTS bacteraemia GWAS (detailed in chapter 3). Sample quality control filters were applied as for the NTS bacteraemia GWAS, and resulted in 386 case (inpatient case fatality) and 1,110 control (survivors to discharge) samples being included in the analysis. SNPs were included in the analysis if they were common (minor allele frequency > 0.05), well-imputed (imputation info score > 0.4) and without evidence of extreme deviation from Hardy-Weinberg equilibrium ($P > 1x10^{-10}$). Following quality control, 14,760,825 SNPs were included in an association analysis of mortality under an additive, dominant, recessive and heterozygous advantage models of genotype/phenotype association, including the first two principal components of genome-wide genotyping data to account for population substructure.

Enrichment analysis of bacteraemia mortality risk at methionine metabolism loci

To evaluate the evidence for enrichment of mortality-associated genetic variation at methionine metabolism loci associated with susceptibility NTS disease (*AMD1, APIP, ENOPH1* and *MTAP*), we used a permutation-based approach to assess for evidence of global enrichment. We included 1,013 SNPs in the analysis, with SNPs being included if they were within 10kb of *AMD1, APIP, ENOPH1* or *MTAP*. SNP and sample quality control threshold were applied as for the mortality genome-wide association analysis (above). Case-control status was permuted 1,000 times and association testing performed for each permutation under an additive model. An empirical enrichment *P*-value is then derived as the proportion of randomly permuted associated test sets in which the summed Z-scores are more extreme than that observed for the mortality case-control status. Statistical analysis was performed in R.

RNA expression quantitative trait analysis at methionine metabolic loci

As for the *EVI5L* and *STAT4* eQTL analyses, we identified *cis* eQTLs of *APIP, AMD1, MTAP,* and *ENOPH1* in primary immune cell subsets from healthy European adults (see chapters 4 and 5). Genotypes in *cis* to each gene (within 1Mb) were correlated with RNA expression in each cell type: B cells (n=279), NK cells (n=245), neutrophils (n=101), naïve monocytes (n=414), and stimulated monocytes (LPS 2 hours, n=261; LPS 24 hours, n=322; IFNγ

24 hours, n=367). Normalized RNA expression was correlated with genotype by linear regression and ANOVA, including the first 25 principal components of gene expression data in each cell-type/condition to account for confounding variation. *P*-values are calculated with *F*-tests (1 d.f.). *P*-values were adjusted for multiple comparisons using the Benjamini-Hochberg method. Statistical analysis was performed in R.

Generation and characterisation of Salmonella mutants

Chromosomal deletion mutants of genes of interest were generated using one-step, λ-red phage recombinase-driven gene disruption by homologous recombination (Datsenko & Wanner, 2000).

Preparation of DNA for transformation

Flippase recognition target (FRT) sequence-flanked antibiotic resistance genes were amplified from the plasmids pKD3 (chloramphenicol resistance) and pKD4 (kanamycin resistance) using PCR primers flanked by 50 bp regions homologous to flanking regions of the gene to be deleted. Plasmid pKD3 and pKD4 DNA was prepared from plasmid-expressing *E. coli* strains using QIAprep Spin Miniprep Kits (Qiagen) according to manufacturer's instructions. Primers used to delete *pfs*, *speE* and *metJ* are listed in Table 6.2. Each primer is composed of a 50bp flanking region homologous to the gene to be deleted and a common primer region to amplify the antibiotic resistance gene.

Table 6.2 Primers for chromosomal gene deletion by homologous recombination

	Sequence	
Gene	Flanking homology region	Common primer region
pfs	F 5'-AAGCGCATGATAAACTATGCCCCCAAAACTGACTTATCGCGAGTAAATCT	GTGTAGGCTGGAGCTGCTTC-3'
pfs	R 5'-TTTGCTTAGCCATGCGCCAGTTTCTGTACCAGCGTTTCGACCATCAGAGT	CATATGAATATCCTCCTTAG-3'
speE	F 5'-ATGATGTTGCGCCCTTTTTTTACGGGTGTTAACAATGGAGGTATCAGCCA	GTGTAGGCTGGAGCTGCTTC-3'
speE	R 5'-CTTTTTCAATTTTTATCACCTCCTTAACGTCACTGTGCGGACAGTGCGTC	CATATGAATATCCTCCTTAG-3'
metJ	F 5'-GGGCTCAGGTTCAGACCTCAATATTAATGACGAAGAGGATTAAGTATCTC	GTGTAGGCTGGAGCTGCTTC-3'
metJ	R 5'-GAATCGTTAAAAAAGCGCGGCCAGAGGCGTTCTGACCGCATGCTTTGCTA	CATATGAATATCCTCCTTAG-3'

Primers were used at a final concentration of 0.2μM in 50μl PCR reactions including 1μl of plasmid DNA, dNTPs at 100μM (Sigma-Aldrich), 1μl Expand™ High Fidelity Taq Polymerase (Sigma-Aldrich), Expand™ High Fidelity sample buffer (giving a final concentration of $MgCl_2$ of 1.5mM), and PCR-grade water (Sigma-Aldrich). Reactions were incubated at 95°C for 5 minutes, before 25 cycles of melting (95°C, 20 seconds), annealing (55°C, 20 seconds), and extension (72°C, 2 minutes), followed by a final prolonged extension (72°C, 7 minutes). PCR products were purified with the QIAquick PCR Purification Kit (Qiagen) according to manufacturer's instructions.

Transformation of Salmonella

We transformed strains of *S.* Typhimurium 12023 harbouring the ampicillin-resistant, temperature-sensitive, λ-red recombinase expression plasmid pKD46. pKD46 *Salmonella* was cultured overnight in LB broth supplemented

with carbenicillin (50 µg/ml), at 30°C with shaking (200rpm). Overnight cultures were sub-cultured 1:100 and grown to an OD_{600} of 0.6-0.8 at 30°C with shaking (200rpm), in 20ml volumes of LB supplemented with carbenicillin (50 µg/ml) and 0.2% L-arabinose to induce λ-red recombinase component expression. Bacteria were made electrocompetent with cooling on ice and washing with ice-cold water (twice) and with 1mM MOPS/20% glycerol (once). Following washing, the bacterial pellet was re-suspended in 200µl 1mM MOPS/20% glycerol.

Electroporation of bacteria in 2mm electroporation cuvettes (ThermoFisher Scientific) was performed with 100µl bacteria and 100ng of PCR product for transformation, using a GenePulser II electroporator (Bio-Rad) at 200Ω, 25µF and 2.2kV. Electroporated bacteria were then grown in 1ml SOC, at 37°C with shaking (200rpm) for 1 hour, before plating onto antibiotic-containing LB agar plates and overnight incubation at 37°C.

Selection of transformants

Chloramphenicol-resistant (pKD3) and kanamycin-resistant (pDK4) transformants were identified on chloramphenicol (10µg/ml) and kanamycin (50 µg/ml) containing plates. Following re-streaking onto agar plates, and curing overnight at 42°C, loss of the pKD46 λ-red recombinase expression plasmid was confirmed by loss of ampicillin resistance (as determined by absence of growth on plates supplemented with carbenicillin at 50 µg/ml).

Verification of gene deletions

Target gene deletion in selected transformants was confirmed by colony PCR. Three PCR reactions were performed for each transformant. Both chromosome/resistance gene junctional regions were identified using an antibiotic-resistance gene-specific primer paired with a locus-specific primer from the region flanking the deleted gene. We also confirmed gene deletion, using locus-specific primer pairs flanking the deleted region to confirm change in PCR-product size having replaced the chromosomal gene with the antibiotic resistance gene. Primers used to verify *pfs*, *speE* and *metJ* deletions are listed in Table 6.3.

Table 6.3 Primers for verification of chromosomal gene deletion

Gene	Reaction	Forward	Reverse
pfs	Junction #1	5'-CGATCGATGCCATATCCTCTC	Common primer #1
pfs	Junction #1	Common primer #2	5'-AAAGCGAGCTCCGTATTGG
pfs	Spanning	5'-CGATCGATGCCATATCCTCTC	5'-AAAGCGAGCTCCGTATTGG
speE	Junction #1	5'-CTCCGTGATCCTATTGCTGAAA	Common primer #1
speE	Junction #1	Common primer #2	5'-GATATAGCCATCGCGCTCTT
speE	Spanning	5'-CTCCGTGATCCTATTGCTGAAA	5'-GATATAGCCATCGCGCTCTT
metJ	Junction #1	5' CATCTGCGACCGCTAACTT	Common primer #1
metJ	Junction #1	Common primer #2	5'-TTTATCCACCGAGGGTTATTCG
metJ	Spanning	5' CATCTGCGACCGCTAACTT	5'-TTTATCCACCGAGGGTTATTCG

Common primers used are specific to the antibiotic resistance cassette introduced. For kanamycin-resistant transformants; common primer #1, 5'-CAGTCATAGCCGAATAGCCT, common primer #2, 5'-CGGCCACAGTCGATGAATCC. For chloramphenicol-resistant transformants; common primer #1, 5'-TTATACGCAAGGCGACAAGG, common primer #2, 5'-GATCTTCCGTCACAGGTAGG.

Primers were used at a final concentration of 0.2μM in 20μl PCR reactions including 5μl of a 20μl suspension of a transformant colony in PCR-grade water (Sigma-Aldrich), dNTPs at 100μM (Sigma-Aldrich), 0.5μl JumpStart™ Taq Polymerase (Sigma-Aldrich), JumpStart™ sample buffer (giving a final concentration of $MgCl_2$ of 1.5mM), and PCR-grade water (Sigma-Aldrich). Reactions were incubated at 95°C for 10 minutes, before 25 cycles of melting (95°C, 20 seconds), annealing (55°C, 20 seconds), and extension (72°C, 30 seconds – 2 minutes for region-spanning reaction), followed by a final prolonged extension (72°C, 10 minutes). PCR products were separated by electrophoresis in a 1% agarose gel containing 1:10,000 SYBR®-safe, and PCR product size visualised by transillumination with UV light.

Removal of antibiotic resistance gene

The antibiotic gene introduced in place of the deleted chromosomal gene is flanked by FRT sequences, allowing its removal by flippase (FLP)-mediated excision, and minimising adventitious effects on genes in *cis* to the deleted gene. To achieve this, the ampicillin-resistant, temperature-sensitive, FLP expression plasmid pCP20 was transformed into *Salmonella* mutants. Plasmid pCP20 DNA was prepared from plasmid-expressing *E. coli* strains using QIAprep Spin Miniprep Kits (Qiagen) according to manufacturer's instructions. Mutant *Salmonella* strains with target gene deletions were cultured overnight in LB broth supplemented with the appropriate antibiotic, at 37°C with shaking (200rpm). Overnight cultures were sub-cultured 1:100 and

grown to an OD_{600} of 0.6-0.8 at 30°C with shaking (200rpm) in 20ml volumes. Bacteria were made electrocompetent as above, and transformed with 100ng of pCP20 DNA by electroporation (settings as above).

Electroporated bacteria were then grown in 1ml SOC, at 30°C with shaking (200rpm) for 1 hour, before plating onto carbenicillin-containing LB agar plates and overnight incubation at 30°C. Ampicillin-resistant clones were re-streaked onto agar plates, and cured overnight at 42°C. Loss of the pCP20 FLP expression plasmid was confirmed by loss of ampicillin resistance. Removal of the antibiotic resistance gene was confirmed by reduction in PCR product size following colony PCR using the same gene-spanning primers used to confirm transformant gene deletion (Table 6.3), and the same PCR reaction conditions.

Generation of green fluorescent protein expressing Salmonella mutants

For characterisation of *Salmonella* mutants requiring bacterial fluorescence, mutant strains of *Salmonella*, following pCP20-mediated removal of antibiotic resistance, were transformed with the ampicillin-resistant green fluorescent protein (GFP) expressing plasmid pFPV25.1. Transformation of electrocompetent bacteria by electroporation was performed as for pCP20 transformation (above). Successful transformants were selected by the presence of ampicillin resistance, and GFP expression confirmed by colony transillumination with UV light.

Complementation of Salmonella mutants

To confirm that the observed phenotype of *Salmonella* mutants was secondary to the targeted gene deletion, we complemented mutant strains with expression of the deleted gene on the ampicillin-resistant, low copy-number plasmid pWSK29. PCR primers (Table 6.4) to clone *metJ* were designed to amplify a 570bp region including 255bp upstream of metJ (to allow cloning of the endogenous promoter) up to, but not including, the *metJ* stop codon (to allow addition of a haemagglutinin - HA - tag). Primers included restriction sites (EcoR1 and BamH1) to allow ligation into pWSK29. This region was amplified from *S*. Typhimurium 12023 genomic DNA, using the Expand™ High Fidelity Taq polymerase system (Sigma-Aldrich) in 50µl reactions, as described above. PCR products were purified with the QIAquick PCR Purification Kit (Qiagen) according to manufacturer's instructions, and the expected PCR product size verified by visualization following separation on a 1% agarose gel (as above).

Table 6.4 *metJ* cloning primers

	GC clamp	Restriction site	*metJ*-specific region
Forward	5'-CGCGGG	GAATTC	TTCCTGGCTGAGTCGGTGAAAGTC
Reverse	5'-CGCGGG	GGATCC	GTATTCCCACGTTTCCGGATCA

Restriction sites are for EcoR1 (forward primer), and BamH1 (reverse primer)

Following purification, the PCR product was digested in 30µl reactions containing EcoR1-HF® and BamH1-HF® enzymes in CutSmart® buffer (all New England Biolabs) for 2 hours at 37°C. The digest was then separated by

electrophoresis in a 1.5% low-melting temperature agarose gel containing 1:10,000 SYBR®-safe, and the band corresponding to the amplified *metJ* gene (570bp) excised following visualisation by UV transillumination. Plasmid pWSK29 DNA was prepared from plasmid-expressing *E. coli* strains using QIAprep Spin Miniprep Kits (Qiagen) according to manufacturer's instructions. Plasmid DNA was digested with EcoR1 and BamH1 in the same manner as for the *metJ* PCR product, before dephosphorylation with 10% calf intestinal alkaline phosphatase (New England Biolabs) for 30 minutes at 37°C to prevent auto-ligation, agarose gel electrophoresis and excision of the digested pWSK29 backbone.

The digested *metJ* PCR product and pWSK29 backbone were in-gel ligated in 50µl ligation reactions including 4µl backbone and *metJ* insert (following melting of gel slices at 70°C), with 0.5µl T4 DNA ligase, 2µl bovine serum (New England Biolabs) albumin and 10% polyethylene glycol 8000 (Sigma-Aldrich), at room temperature for 4 hours. The ligated construct was then heat-shock transformed into chemically competent *E. coli* (DH5α). Aliquots (100µl) of competent cells were incubated with 5µl of ligated construct for 30 minutes on ice, followed by incubation for 5 minutes at 37°C. Cells were then added to 1ml SOC and incubated at 37°C with shaking (200rpm) for 1 hour, before plating onto LB agar plates supplemented with carbenicillin and incubation overnight at 37°C. To confirm successful cloning of metJ, successful transformants were grown overnight (37°C with shaking in carbenicillin-containing LB broth) and plasmid DNA prepared with QIAprep Spin Miniprep Kits (Qiagen) according to manufacturer's instructions. Plasmid

DNA from successful transformants was then Sanger sequenced using the M13 forward universal primer, which binds just upstream of the pWSK29 insertion site. Plasmid pWSK29::*metJ* was then transformed into the *ΔmetJ* derivative of *S*. Typhimurium 12023 using electroporation of electrocompetent bacteria as used for the pCP20 transformations. An empty wPSK29 vector without insert was also transformed into wild-type and *ΔmetJ S*. Typhimurium 12023 in the same manner to acts as experimental controls, accounting for plasmid burden.

Characterisation of Salmonella mutants

The effects of chromosomal deletions of *pfs*, *metJ* and *speE* were characterised by comparison to their parent strain *S*. Typhimurium 12023. We characterised the effect of these gene deletions on bacterial replication, cytotoxicity and pro-inflammatory cytokine release during phagocytic infection.

Phagocytic infections were performed as described in chapter 5 in immortalized bone-marrow derived macrophages (iBMDMs) murine RAW 264.7 macrophages (European Collection of Authenticated Cell Cultures). Primary iBMDMs were obtained from C57BL/6 wildtype, *casp1*$^{-/-}$*/casp11*$^{-/-}$, *ripk3*$^{-\backslash-}$, and *ripk3*$^{-\backslash-}$*/casp8*$^{-\backslash-}$ mice (Charles River, U.S.A.) by Dr Sophie Matthews, MRC Centre for Bacteriology and Infection, Imperial College, London. RAW macrophages were maintained Dulbecco's Modified Eagle's medium (DMEM, Life Technologies) supplemented with 10% fetal bovine

serum (Sigma-Aldrich), at 37°C and 5% CO_2. Murine iBMDMs were maintained as described in chapter 5.

Bacterial replication within phagocytes was assayed by colony-forming unit (CFU) assays, to determine net intracellular replication, and by flow cytometry of cells infected with GFP-expressing bacteria as described previously (chapter 5). Cytotoxicity during phagocytic infection was determined by LDH release into culture supernatants as previously (chapter 5).

Release of interleukin-1β (IL-1β) into culture supernatants was measured using a murine IL-1β sandwich enzyme-linked immunosorbent assay (ELISA) kit (ThermoFisher Scientific), according to manufacturer's instructions. At appropriate time-points, culture supernatants were harvested from wells containing infected cells, briefly centrifuged (13,000 rpm, 1 minute) and the cell-free supernatants stored at -80°C prior to analysis. Supernatants from wells of uninfected cells of the same density were harvested in the same manner, to act as negative controls. Two-fold diluted supernatants, and serial 10-fold dilutions of IL-1β reference standards, were aliquoted into wells of a 96-well, pre-coated with anti-IL-1β antibody and incubated at room temperature for 2 hours. After plate-washing (wash buffer 3x), a second biotinylated anti-IL-1β antibody was added to each well to detect capture-antibody-bound IL-1β, and incubated for 1 hour at room temperature. After plate-washing (wash buffer 3x), horse-radish peroxidase (HRP) conjugated to streptavidin was added to each well (and incubated for 30 minutes at room temperature), binding biotin-labelling detection antibody. After plate washing

(wash buffer 3x) tetramethylbenzidine (TMB) HRP substrate was added to each well, and absorbance read at 550nm.

Bacterial replication, cytotoxicity, and IL-1β release in each infection condition, at each time-point, were measured in triplicate. Summary statistics (mean and standard errors) for a given infection condition were calculated with averages of at least three independent experiments, each conducted in triplicate. Statistical analysis was performed in R.

Results

Pathway enrichment analysis of NTS bacteraemia GWAS

Pathway enrichment analysis with INRICH, revealed a single significantly enriched gene set among genic regions associated with NTS bacteraemia in Kenyan children (empirical *P*-value = 3.0×10^{-5}, adjusted *P*-value = 0.01), "Metabolism of Polyamines". Within that gene set four genic regions overlap with NTS-associated regions: *APIP*, *ENOPH1*, *AMD1*, and *MTAP* (Figure 6.1). *MTAP, APIP* and *ENOPH1* encode enzymes (methylthioadenosine phosphorylase, APAF1-interacting protein, and enolase-phosphatase 1 respectively) catalysing the three of the six enzymatic steps that make up the methionine salvage pathway, converting 5'-*S*-methyl-5'-thioadenosine (MTA) to methionine. AMD1 encodes adenosylmethionine decarboxylase 1, which decarboxylates *S*-adenosyl-*L*-methionine (SAM). The observed associations with NTS bacteraemia, in the most probable model of association, at the four NTS-associated methionine metabolism loci are summarized in Table 6.5.

Table 6.5 Association statistics of peak SNPs in methionine metabolism loci with NTS bacteraemia in Kenyan children

SNP	Gene	Model	MAF	OR	*P*
rs150128889	AMD1	Dom	0.06	2.81	7.69×10^{-8}
rs2433162	APIP	Dom	0.35	0.44	1.73×10^{-6}
rs11099533	ENOPH1	Rec	0.10	6.97	7.72×10^{-6}
rs75965083	MTAP	Add	0.05	2.48	9.66×10^{-7}

MAF, minor allele frequency; dom, dominant; rec, recessive; add, additive; OR, odds ratio.

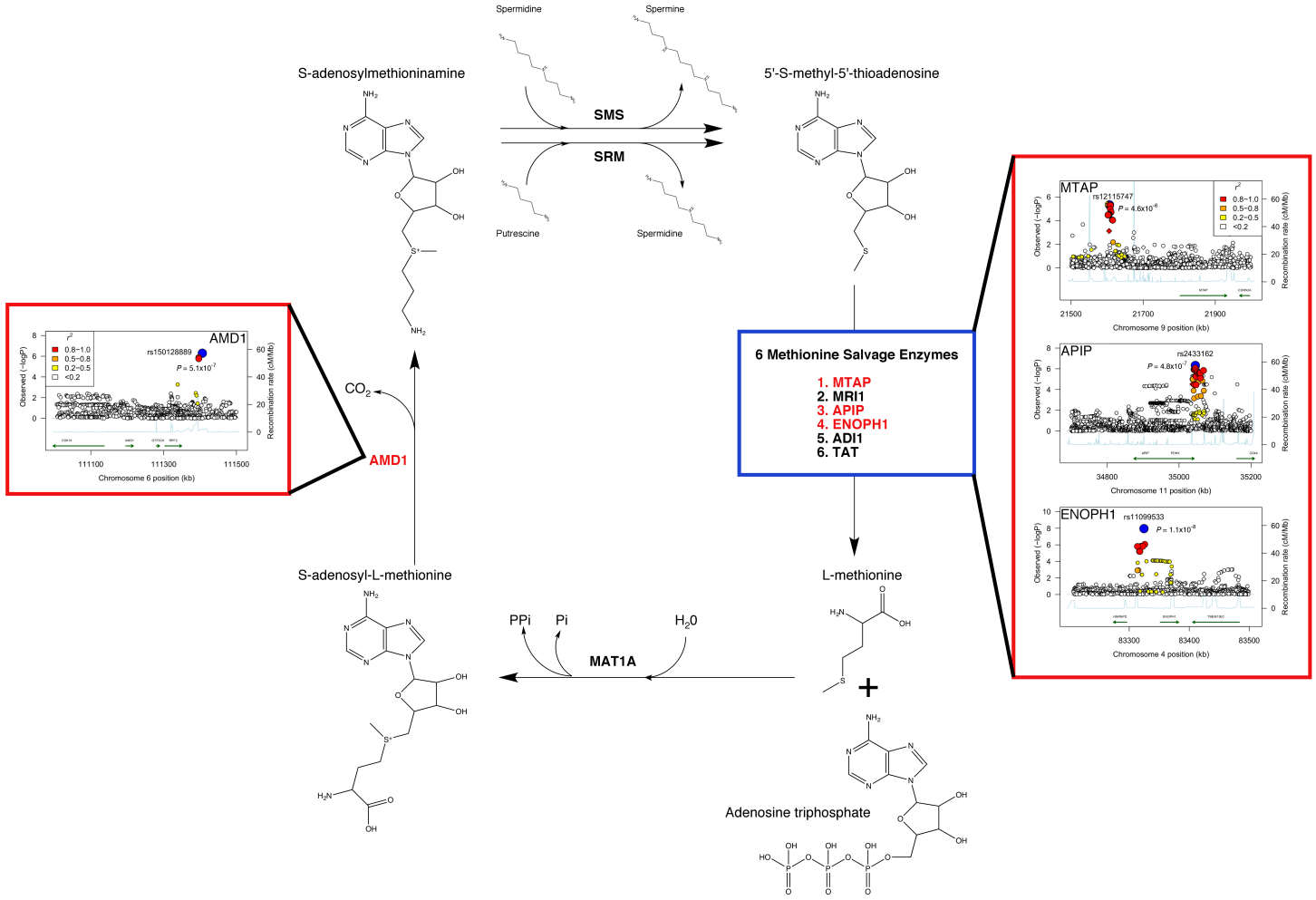

Figure 6.1 Enrichment of genetic loci associated with NTS bacteraemia at genic loci encoding enzymes metabolising methionine and its derivatives.

Figure 6.1 (**preceding page**)
Diagram of methionine metabolism, including methionine salvage (blue box). Enzymes in genic regions with evidence of NTS association are highlighted in red. For these loci, regional association plots of association with NTS bacteraemia in 2,857 children (180 cases; 2,677 controls) under dominant (APIP, AMD1), recessive (ENOPH1) and additive (MTAP) models are presented. SNPs are coloured according to strength of linkage disequilibrium (r^2) to the peak SNP (highlighted in blue). Molecule schematics are drawn in ChemDraw v15.1.

Association of major bacterial pathogens in Kenyan children with methionine metabolism genes

The identification of enrichment of NTS-associated loci in genes coordinating methionine metabolism is striking given previous observations identifying an eQTL for APIP (rs514182) as a determinant of *Salmonella*-induced pyroptosis in an *in vitro* screen (Ko et al., 2012). That study went on to demonstrate that the effect of genetic variation in *APIP* is mediated by altered flux through the methionine salvage pathway, and specifically, intracellular levels of 5'-*S*-methyl-5'-thioadenosine (MTA). The same study went on to demonstrate that rs514182 is associated with mortality (OR = 0.44, P = 0.005) in American adults with sepsis. Given this association of genetic variation in APIP with mortality in patients with sepsis in a setting in which invasive *Salmonella* infection is uncommon, we used a Bayesian approach to assess the evidence for association at NTS-associated SNPs at *APIP*, *MTAP*, *AMD1* and *ENOPH1* with bacteraemia secondary to common bacterial pathogens in Kenyan children.

We included 1,366 case (NTS, 180; *S. aureus*, 177; *S. pneumonia*, 429; *H. influenzae*, 133; *E. coli*, 159; β-haemolytic *Streptococcus*, 158; *Acinetobacter*,

130) and 2,677 control samples in a multinomial regression model of bacteraemia risk at the peak NTS-associated at each of *APIP*, *MTAP*, *AMD1* and *ENOPH1*. In that analysis, the data supports a model in which genotype at the peak *APIP* (rs2433162), *AMD1* (rs150128889) and *ENOPH1* (rs11099533) SNPs is associated with NTS bacteraemia alone (Figure 6.2; Bayes factors c.f. null; 3,357, 4,192, and, 23,656 respectively).

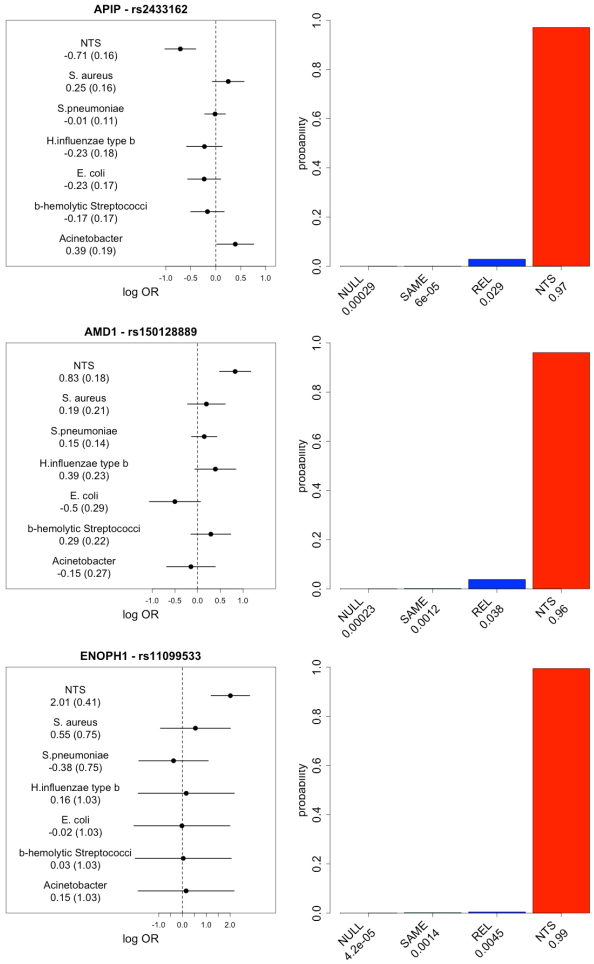

Figure 6.2 Association of methionine metabolic loci with major bacterial pathogens.

Log-transformed odds ratios and 95% confidence intervals of association, under models of dominant (APIP and AMD1) and recessive (ENOPH1) association in Kenyan discovery samples; left panels. Posterior probabilities of models of association: NULL, no association with any bacterial pathogen; SAME, the same effect across all bacterial pathogens; REL, related effects across all bacterial pathogens; NTS, a non-zero effect for NTS bacteraemia alone.

By contrast, data at the MTAP locus (rs75965083) does provide some evidence of association with bacteraemia susceptibility beyond that observed for NTS. In that analysis, the data best supports a model in which genotype at rs75965083 is associated with bacteraemia secondary to both NTS and *H. influenza* type b (Figure 6.3; Bayes factor c.f. null = 87,981).

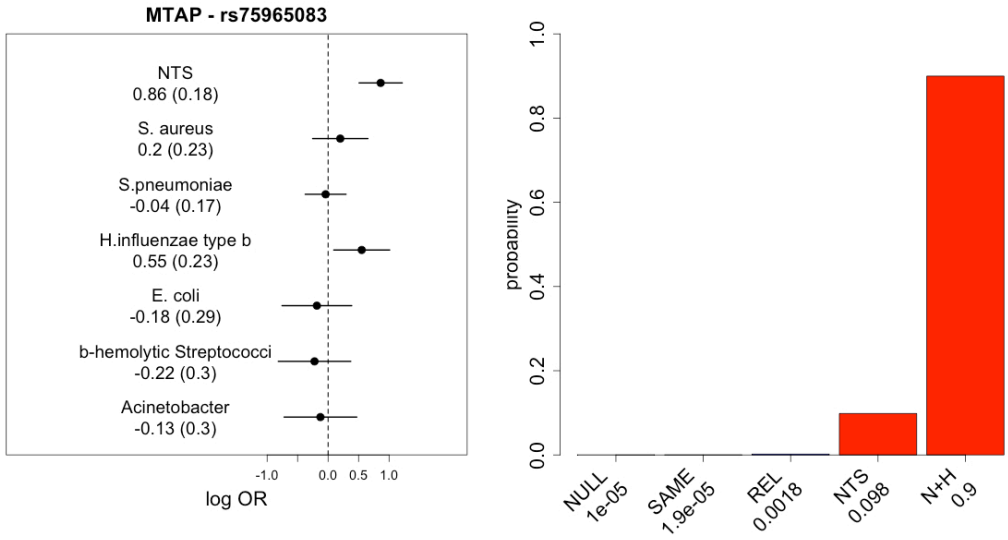

Figure 6.3 Association of rs75965083 with major bacterial pathogens in Kenyan children.

Log-transformed odds ratios and 95% confidence intervals of association, under an additive model of association in Kenyan discovery samples; left panels. Posterior probabilities of models of association: NULL, no association with any bacterial pathogen; SAME, the same effect across all bacterial pathogens; REL, related effects across all bacterial pathogens; NTS, a non-zero effect for NTS bacteraemia alone; N+H, a non-zero effect for bacteraemia secondary to NTS and *H. influenzae*.

Assessment of methionine metabolism loci as genetic correlates of mortality in Kenyan children with bacteraemia

In light of the previous data (Ko et al., 2012) describing an association between an APIP eSNP (rs514182) and risk of mortality in American adults with sepsis, we performed a genome-wide association analysis of in-patient

mortality in Kenyan children with bacteraemia. In that analysis, including genotypes at 14,760,825 SNPs from 1,496 individuals (386 cases, 1,110 controls) were included in the analysis. Inspection of quantile-quantile plots (Figure 6.4) and the genomic control inflation factor (λ=1.01) indicate that inclusion of the two major principal components as covariates in the analysis adequately controls for population substructure.

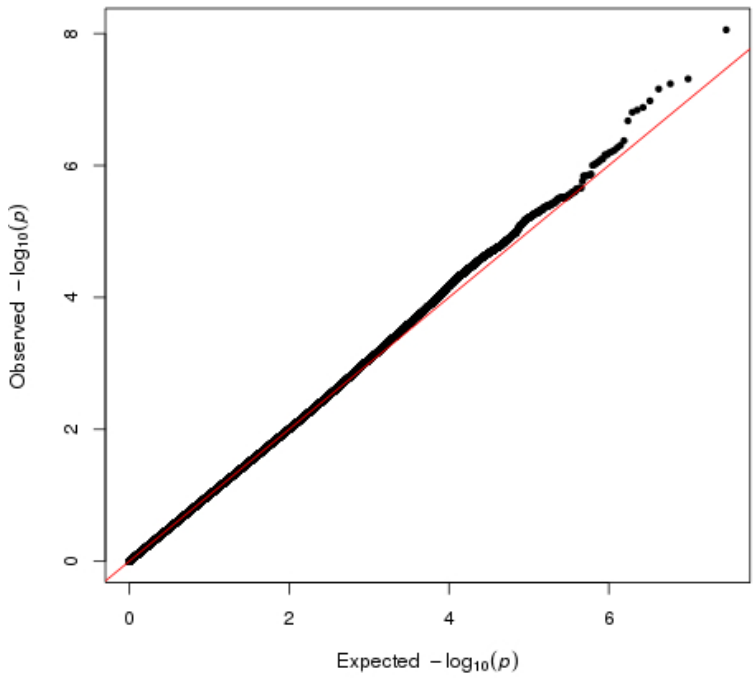

Figure 6.4 Quantile-quantile plot of bacteraemia mortality genome-wide association study.

Quantile-quantile plot of mortality association in Kenyan children under an additive model (14,760,825 SNPs, $\lambda = 1.01$). Observed $-\log_{10}$(p-values)/test statistics are plotted against those expected under a χ^2 distribution with 1 degree of freedom.

We first assessed whether the sepsis mortality-associated variant in *APIP* (rs514182) or NTS-associated variants at *AMD1*, *APIP*, *ENOPH1*, or *MTAP* were associated with mortality in Kenyan children hospitalised with bacteraemia. We find no evidence for association of rs514182 with mortality in

Kenyan children with bacteraemia (OR = 0.92, 95% CI 0.77-1.10; *P* = 0.38; additive model), nor do we find any evidence of association at the peak NTS-associated SNPs at *AMD1* (rs150128889; OR 1.01, 95% CI 0.74-1.38; *P* = 0.94; additive model), *APIP* (rs2433162; OR 0.83, 95% CI 0.69-1.00; *P* = 0.05; additive model), *ENOPH1* (rs11099533; OR 1.05, 95% CI 0.79-1.38; *P* = 0.75; additive model), and *MTAP* (rs75965083; OR 1.05, 95% CI 0.73-1.51; *P* = 0.80; additive model). To evaluate the evidence for association with mortality at *AMD1*, *APIP*, *ENOPH1* and *MTAP* more broadly, we assessed evidence for association under additive, dominant, recessive and heterozygous advantage models at SNPs within 100kb of each locus with mortality. In that analysis (Figure 6.5), no SNP is suggestively associated with mortality ($P < 1x10^{-4}$).

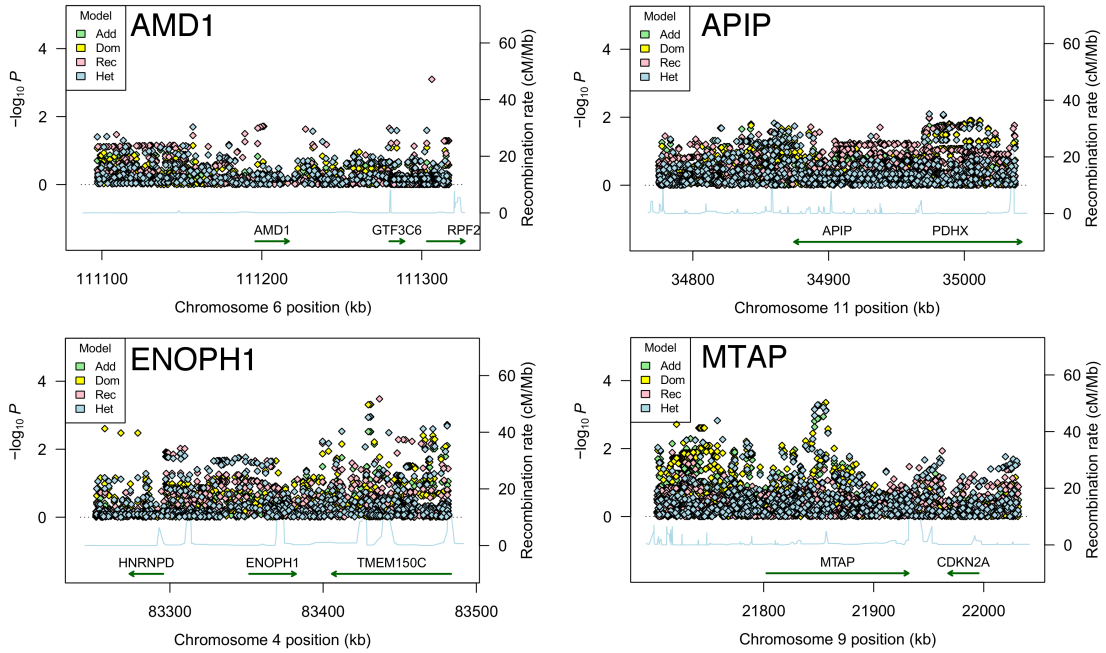

Figure 6.5 Association analysis of mortality in Kenyan children with bacteraemia at NTA-associated methionine metabolism loci

Regional association plots within 100kb of *AMD1, APIP, ENOPH1* and *MTAP* are displayed. SNPs are coloured according to the model of association employed. Add, additive; dom, dominant; rec, recessive; and het, heterozygous advantage.

In keeping with the lack of evidence for association with mortality at *AMD1*, *APIP*, *ENOPH1* and *MTAP* at the single SNP level, there is no evidence of global enrichment of mortality association at these loci (Figure 6.6). In a permutation analysis of mortality association at *AMD1*, *APIP*, *ENOPH1* and *MTAP*, SNPs within 10kb of genes have no evidence of global enrichment of association with mortality (permutation *P*-value=1).

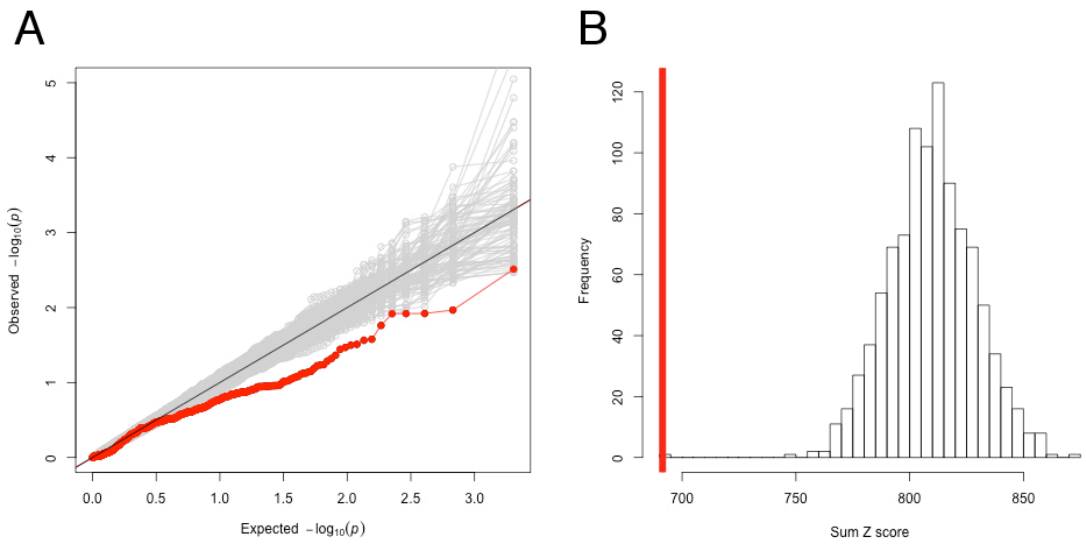

Figure 6.6 Enrichment analysis of mortality at NTS-associated methionine metabolism loci in Kenyan children with bacteraemia

1,013 SNPs at *AMD1*, *APIP*, *ENOPH1* and *MTAP* were included in a permutation analysis of mortality in children with bacteraemia (446 cases, 1,110 controls). Observed *P*-value distributions are plotted in red, with P-value distribution of 1,000 case-control random permutations plotted in grey (A). The empirical distribution of summed Z-scores for each case-control permutation is plotted in grey, with the observed Z-score sum highlighted in red (B).

RNA expression quantitative trait analysis at methionine metabolic loci

No gene expression probe mapping to APIP, in the dataset used for the eQTL analysis, passed probe quality control, and consequently, eQTLs for APIP in primary immune cell subsets were not analysed. As noted previously, an eSNP (rs514182) for APIP RNA expression (Figure 6.7), has been demonstrated to determine levels of *Salmonella*-induced pyroptosis, and modify risk of mortality in patients with sepsis (Ko et al., 2012). In our GWAS of NTS bacteraemia in Kenyan children, we identify limited evidence for an association at rs514182 with susceptibility to NTS bacteraemia in a dominant model of association (OR = 0.71; *P*=0.03), with individuals carrying the rs514182:AA genotype being at increased risk of NTS disease. While the data are limited, they provide some support for a model in which increased APIP expression protects against risk of NTS bacteraemia.

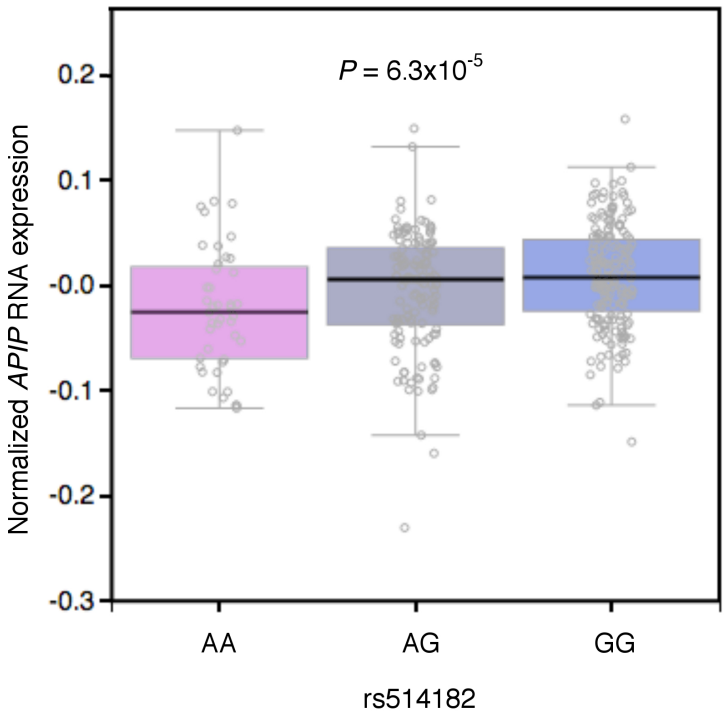

Figure 6.7 The effect of rs514182 genotype on *APIP* RNA expression whole blood.

Data downloaded from GTEx Analysis Release V6 (dbGaP Accession phs000424.v6.p1; http://www.gtexportal.org/).

Gene expression probes passing quality control thresholds do map to *AMD1, MTAP* and *ENOPH1*, and we tested for significant eSNPs at each locus in naïve leukocytes (NK cells, n = 245; B cells, n = 279; neutrophils, n = 101; monocytes, n = 414) and stimulated monocytes (LPS – 2 hours, n = 261; LPS – 24 hours, n = 322; IFNγ - 24 hours, n = 367). We identified no significant eSNPs for *ENOPH1* RNA expression in any immune cell subset, regardless of stimulation condition. We identified a *cis* eQTL for *MTAP* shared between naïve and stimulated monocytes (Figure 6.8), but with no evidence of overlap between NTS-associated SNPs and MTAP eSNPs.

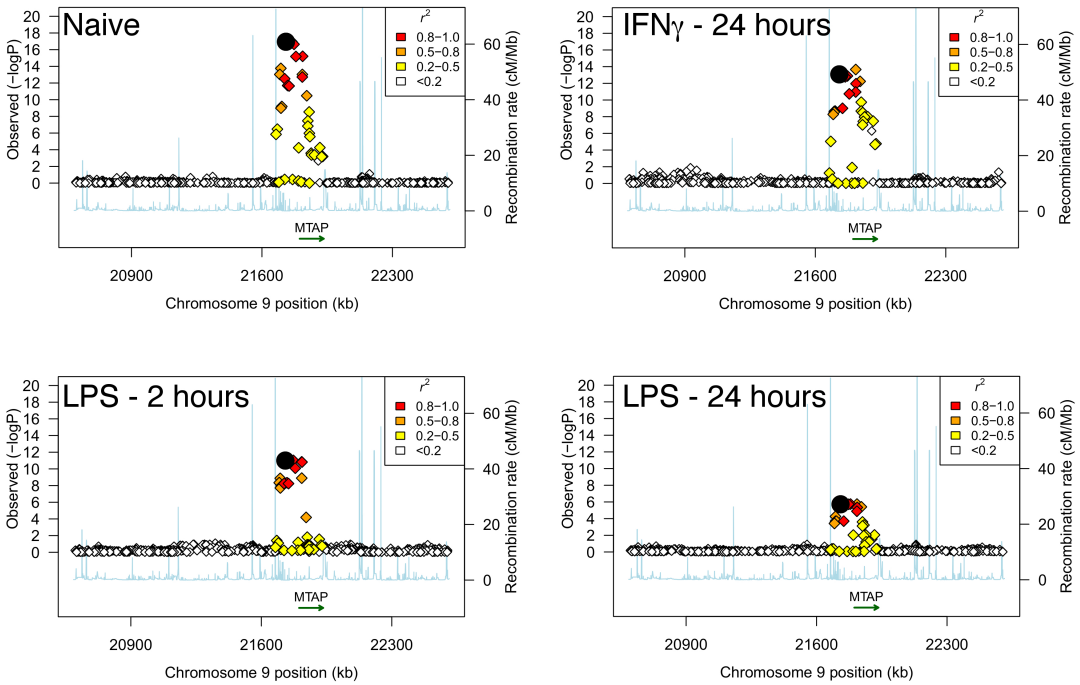

Figure 6.8 Regional association plot of *cis* genetic loci with *MTAP* RNA expression in monocytes.

Association plots are shown for *cis* eQTL analysis of *MTAP* in naïve and stimulated (LPS 2 hours, LPS 24 hours and IFNγ 24 hours) monocytes. The peak eSNP in naïve monocytes (rs6475564) is highlighted as a black in each panel. SNPs are coloured according to strength of linkage disequilibrium (r^2) to the rs6475564.

In the case of *AMD1*, there is a context-specific *cis* eQTL for *AMD1* RNA expression in monocytes, with a significant *cis* eQTL only being identified following stimulation (Figure 6.9). There is limited evidence for overlap between NTS-associated SNPs and AMD1 eSNPs. A highly significant *cis* eSNP for *AMD1* in IFNγ-simulated monocytes, rs802658 ($P = 3.23 \times 10^{-18}$), has some limited evidence of association with NTS bacteraemia in the Kenyan discovery samples ($P = 0.017$).

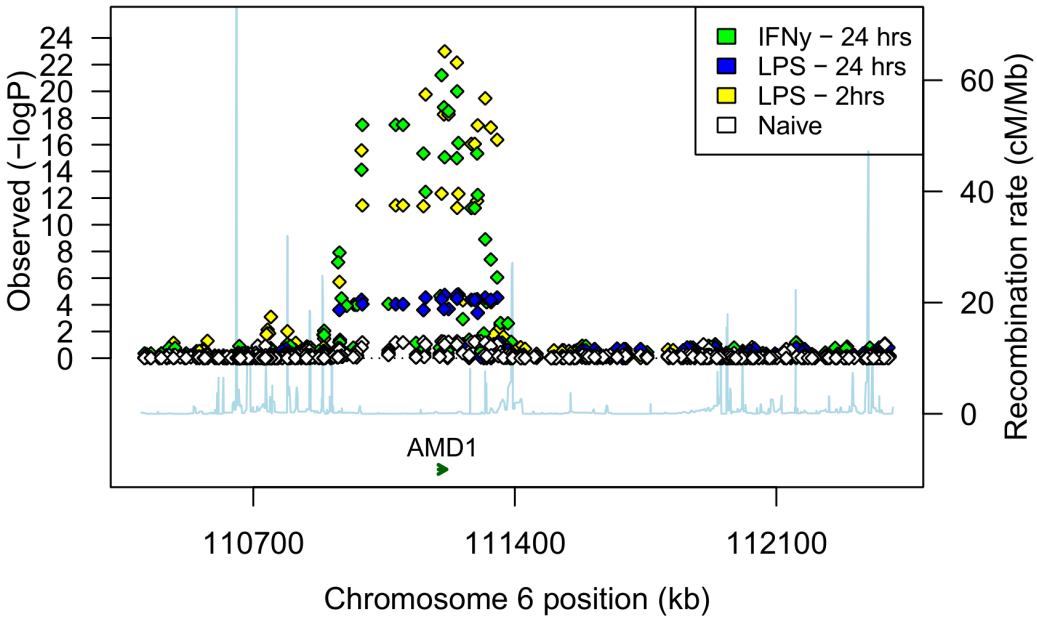

Figure 6.9 Regional association plot of *cis* genetic loci with *AMD1* RNA expression in monocytes.

Association mapping of *cis* eQTLs for *AMD1* in naïve and stimulated (LPS 2 hours, LPS 24 hours and IFNγ 24 hours) monocytes. SNPs are coloured according to stimulation condition.

Generation and characterisation of Salmonella mutants

These data identify host genetic variation in genes encoding methionine-metabolising enzymes as a determinant of susceptibility to *Salmonella* infection. In light of the previously-published observation that levels of a central metabolite in methionine metabolism (MTA) can regulate *Salmonella*-induced pyroptosis, we sought to understand whether bacterial metabolism could influence inflammatory cell death in infected macrophages. This was motivated by the observation that *Salmonella* metabolises MTA (Figure 6.10). We thus sought to perturb bacterial MTA metabolism through the generation of mutant derivatives of *S*. Typhimurium 12023 deficient in pfs, speE and metJ. SpeE catalyses the conversion of *S*-adenosylmethioinamine to MTA, while pfs catalyses the deadenylation of MTA producing *S*-methyl-5'-thio-D-ribose. The *metJ* locus encodes a transcriptional aporepressor, which, along with its co-repressor *S*-adenosyl-L-methionine, represses loci encoding methionine biosynthetic enzymes and transporters (Old, Phillips, Stockley, & Saint Girons, 1991).

We successfully generated mutant derivatives of *S*. Typhimurium strain 12023, with chromosomal deletions of *pfs*, *speE* and *metJ*. All three mutants were viable, although *Δpfs* mutants exhibit a growth defect similar to that observed for *Δpfs N. meningitidis* mutants (Heurlier et al., 2009). Colonies of *Δpfs Salmonella* mutants were consistently and notably smaller following overnight growth on LB agar, and in keeping with this, growth in LB broth was markedly attenuated, with a prolonged lag phase before logarithmic growth (Figure 6.11).

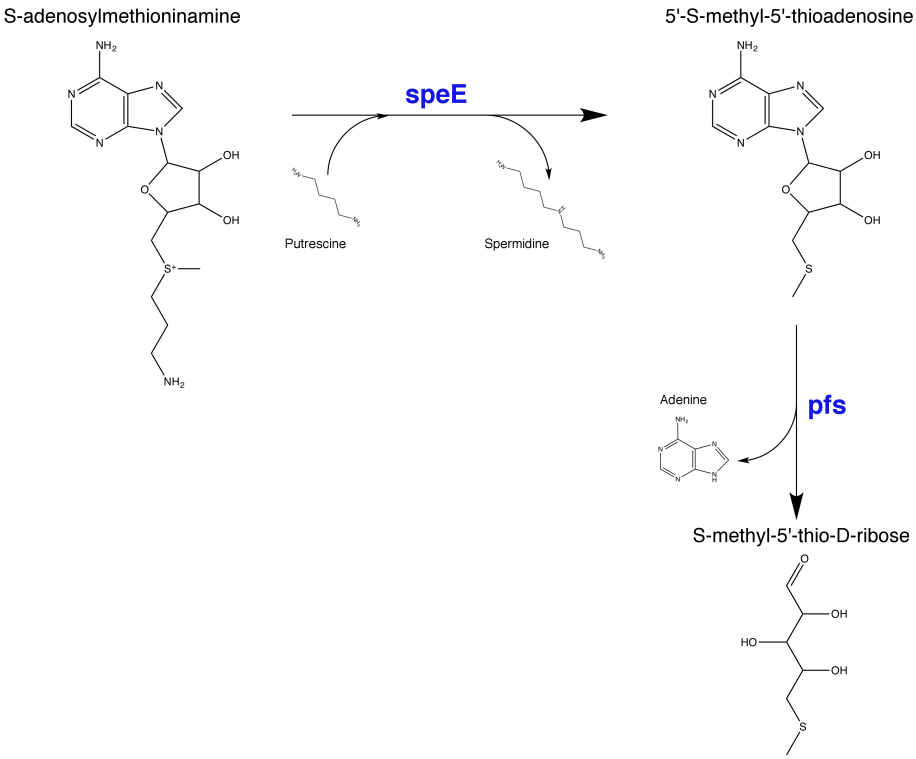

Figure 6.10 Metabolism of 5'S-methyl-5'thioadenosine (MTA) in *Salmonella*

Molecule schematics are drawn in ChemDraw v15.1. Enzymes are highlighted in blue.

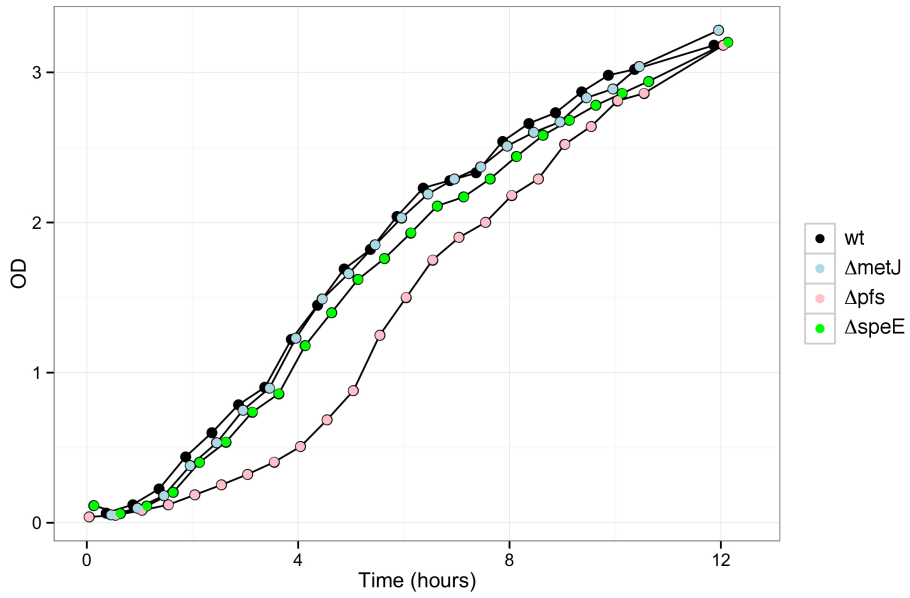

Figure 6.11 Growth curves of mutant derivatives of *Salmonella* deficient in methionine metabolism

Growth curves of *Salmonella* strains in LB broth. Curves are shown for wildtype S. Typhimurium 12023 and its mutant derivatives Δ*metJ,* Δ*pfs* and Δ*speE*.
OD, optical density measured at 600nm; wt, wildtype.

We successfully eliminated the introduced antibiotic resistance gene from the Δ*metJ* and Δ*speE* mutants, ensuring that the deleted genes are replaced by short (circa 80bp) “scars”, with minimal downstream transcriptional effects. All experiments to characterise the Δ*metJ* and Δ*speE* mutants were performed using these mutants. Transduction of pCP20, and removal of the FRT-tagged antibiotic gene, in the Δ*pfs* mutant proved more challenging, and experiments to characterise this mutant were performed with the antibiotic resistance gene used to replace *pfs* still *in situ*.

Intra-macrophage replication of ΔmetJ, Δpfs and ΔspeE mutants

RAW macrophages were infected with *S*. Typhimurium (wild-type 12023 and its mutant derivatives, *ΔmetJ, Δpfs and ΔspeE*) and intracellular replication assayed by bacterial recovery following infected cell lysis at 2, 4, 6, 8, 10 and 16 hours post-infection (Figure 6.12). In these assays, *ΔmetJ Salmonella* had increased rates of intracellular replication at 16 hours post-infection as compared to wild-type (Student's t-test; $P = 0.001$). *ΔspeE Salmonella* had a trend towards increased rates of intracellular replication at 16 hours post-infection as compared to wild-type, but this did not reach statistical significance (Student's t-test; $P = 0.05$). In keeping with its attenuated growth in LB broth, *Δpfs Salmonella* has attenuated replication at 16 hours post-infection in macrophages, as compared to wild-type (Student's t-test; $P = 0.002$).

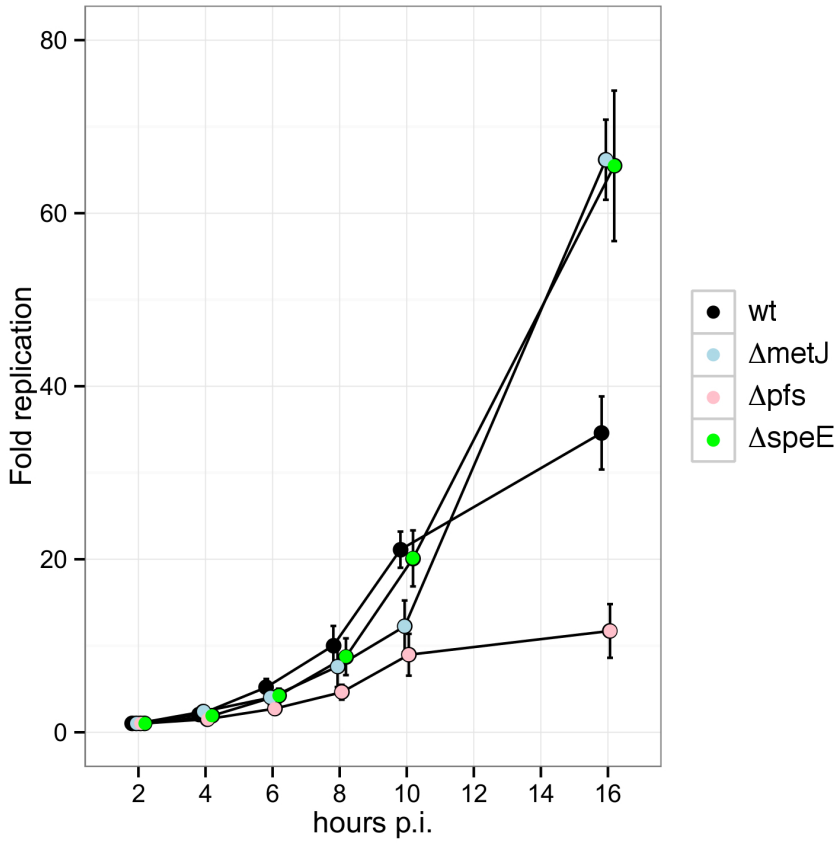

Figure 6.12 Net bacterial replication in RAW macrophages infected with *Salmonella* mutants with perturbed methionine metabolism.

RAW macrophages were infected with wild-type, *ΔmetJ, Δpfs, and ΔspeE S.* Typhimurium 12023. Fold replication at each time point represents net bacterial recovery relative to recovery at 2 hours post-infection. Summary statistics are calculated from five independent experiments. Error bars represent SEM. p.i., post infection.

Salmonella-induced cytotoxicity in macrophages infected with ΔmetJ, Δpfs and ΔspeE Salmonella mutants

RAW macrophages were infected with *S*. Typhimurium (wild-type 12023 and its mutant derivatives, *ΔmetJ, Δpfs and ΔspeE*) and infection-induced cytotoxicity assayed by LDH release at 16 hours post-infection (Figure 6.13). In these experiments, infection with *ΔspeE Salmonella* results in unchanged

levels of cytotoxicity in comparison to wild-type (Student's t-test; $P = 0.91$). In keeping with its attenuated levels of intracellular replication, infection with *Δpfs Salmonella* resulted in reduced cytotoxity as compared to wild-type (Student's t-test; $P = 0.00066$). In contrast however to its apparent increased rates of intracellular replication in RAW macrophages, infection with *ΔmetJ Salmonella* resulted in significantly reduced cellular cytotoxicity (Student's t-test; $P = 0.013$).

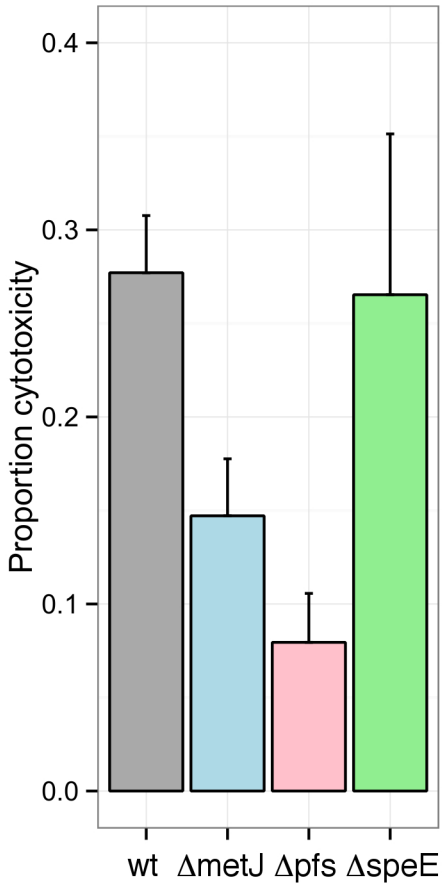

Figure 6.13 Infection-induced cytotoxicity in RAW macrophages infected with *Salmonella* mutants with perturbed methionine metabolism.

RAW macrophages were infected with wild-type, *ΔmetJ, Δpfs, and ΔspeE S.* Typhimurium 12023. The proportion of cellular cytotoxicity was calculated from assayed LDH-release into culture supernatants at 16 hours post-infection. Summary statistics are calculated from six independent experiments. Error bars represent SEM.

Intracellular replication of ΔmetJ mutants in macrophages assayed by flow cytometry

These data suggest that *ΔmetJ Salmonella* exhibits increased intra-macrophage replication while inducing reduced levels of cytotoxicity. We therefore hypothesized that the apparent increase in intracellular replication, as assayed by net bacterial recovery, could represent reduced cytotoxicity during infection with *ΔmetJ Salmonella* rather than an increased intracellular growth rate *per se*. To test this, we directly assayed intracellular replication rate of GFP-expressing wild-type and *ΔmetJ Salmonella* in macrophages by flow cytometry. RAW macrophages were infected with GFP-expressing *S*. Typhimurium (wild-type and *ΔmetJ* 12023) and intracellular replication measured as fold-change in GFP gMFI at 2, 4, 6, 8, 10 and 16 hours post-infection (Figure 6.14). In these experiments, in contrast to the net bacterial recovery experiments, infection with *ΔmetJ Salmonella* resulted in unchanged levels of intracellular replication in comparison to wild-type at 16 hours post-infection (Student's t-test; $P = 0.27$). These data suggest that the apparent effect of metJ gene deletion on intra-macrophage replication is secondary to reduced cytotoxicity of macrophages infected with *ΔmetJ Salmonella*.

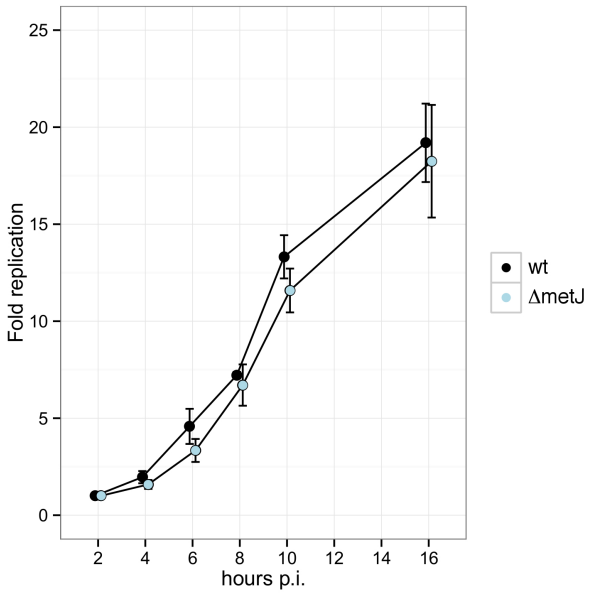

Figure 6.14 Flow cytometric determination of intracellular replication in RAW macrophages infected with *ΔmetJ Salmonella.*

RAW macrophages were infected with GFP-expressing wild-type and *ΔmetJ S.* Typhimurium 12023. Fold intracellular replication at each time point is estimated as fold change in geometric mean fluorescence intensity (gMFI) relative to gMFI at 2 hours post-infection (left). Summary statistics are calculated from four independent experiments. Error bars represent SEM. p.i., post infection.

Complementation of ΔmetJ Salmonella

To confirm that the reduced macrophage cytotoxicity induced by infection with *ΔmetJ Salmonella* is secondary to the deletion of the *metJ* gene, we complemented the *ΔmetJ* mutant by cloning *Salmonella metJ* with its endogenous promoter into a low copy number plasmid (pWSK29), and transforming the metJ expressing plasmid pWSK29::*metJ* into the *ΔmetJ Salmonella* mutant. We then assessed infection-induced cytotoxicity in macrophages infected with *ΔmetJ Salmonella complemented with* pWSK29::*metJ* in comparison to wild-type and *ΔmetJ Salmonella* transduced with the empty pWSK29 vector without insert (Figure 6.15). In these

experiments, as previously, infection of macrophages with *ΔmetJ Salmonella* results in reduced cytotoxicity in comparison to wild-type (Student's t-test; *P* = 0.0009). The phenotype of reduced cytotoxicity is completely abolished by complementation with pWSK29::*metJ* (Student's t-test; *P* = 0.90), demonstrating that the observed *ΔmetJ* mutant phenotype is indeed secondary to loss of *metJ* expression.

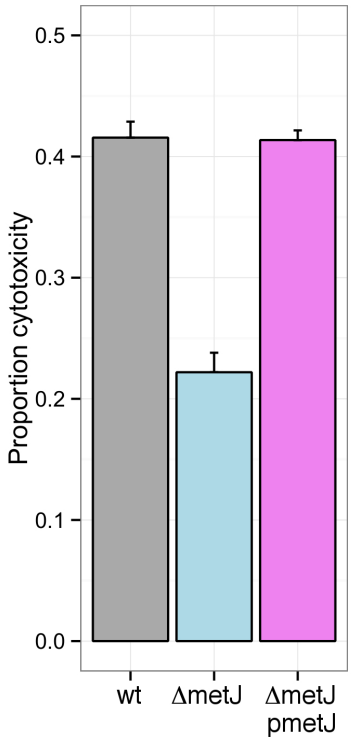

Figure 6.15 Infection-induced cytotoxicity in complemented *ΔmetJ Salmonella* mutants

Immortalised bone-marrow derived macrophages (from wild-type C57BL/6 mice) were infected with wild-type, *ΔmetJ,* and complemented *ΔmetJ S.* Typhimurium 12023. Complementation of *ΔmetJ* mutants was with pWSK29::*metJ*. To control for plasmid burden, wild-type and uncomplemented *ΔmetJ* mutants were transformed with the empty pWSK29 vector. The proportion of cellular cytotoxicity was calculated from assayed LDH-release into culture supernatants at 16 hours post-infection. Summary statistics are calculated from three independent experiments. Error bars represent SEM.

IL-1β production in macrophages infected with ΔmetJ Salmonella

As intracellular MTA levels have been previously implicated in the regulation of *Salmonella*-induced pyroptosis, we hypothesized that the reduced cytotoxicity induced by *ΔmetJ Salmonella* mutants may be due to decreased induction of pyroptosis. To test this we measured the production of IL-1β, the canonical pro-inflammatory cytokine produced by cells undergoing pyroptotic cell death, in macrophages infected with *ΔmetJ* mutant and wild-type *Salmonella* (Figure 6.16). In these experiments, macrophages infected with *ΔmetJ* mutants secrete significantly less IL-1β into culture supernatants during 16 hours of infection than do macrophages infected with wild-type *Salmonella* (Student's t-test; $P = 0.017$). These data are in keeping with the hypothesis that at least a portion of the observed reduction in cytotoxicity in *ΔmetJ Salmonella*-infected macrophages is secondary to reduced levels of pyroptosis.

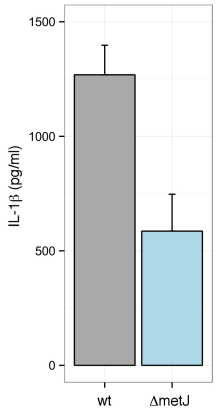

Figure 6.16 IL-1β production in macrophages infected with *ΔmetJ Salmonella*

Immortalised bone-marrow derived macrophages (from wild-type C57BL/6 mice) were infected with wild-type and *ΔmetJ S.* Typhimurium 12023. IL-1β was assayed in culture supernatants harvested at 16 hours post-infection. Summary statistics are calculated from three independent experiments. Error bars represent SEM.

ΔmetJ Salmonella-induced cytotoxicity in macrophages deficient in pyroptosis and necroptosis

To assess whether reduced levels of pyroptosis entirely account for the reduction in infection-induced pyroptosis in *ΔmetJ Salmonella*-infected macrophages, or whether *metJ* deletion could affect other mechanisms of cell death, we assayed infection-induced cytotoxicity in macrophages deficient in both pyroptosis and necroptosis (Figure 6.17). For these experiments we used iBMDMs derived from mice deficient in both caspase1 and caspase11, and thus lacking canonical and non-canonical pyroptosis, and mice deficient in RIPK3, and thus lacking necroptosis. In keeping with our previous data, cytotoxicity in *ΔmetJ Salmonella*-infected wild-type macrophages is reduced by comparison to infection with wild-type *Salmonella* (Student's t-test; P = 0.004). In contrast however to the apparent reduction in IL-1β production in *ΔmetJ Salmonella*-infected macrophages, cytotoxicity in pyroptosis-deficient macrophages remains significantly reduced following infection with the *ΔmetJ* mutant as compared to wild-type (Student's t-test; P = 0.001). By contrast, in macrophages deficient in necroptosis, infection with *ΔmetJ Salmonella* does not result in a significant reduction in cytotoxicity as compared to wild-type *Salmonella* (Student's t-test; P = 0.075). Taken together, these data suggest that rather than limiting cytotoxicity through reduced induction of pyroptosis, *ΔmetJ* mutants may in part reduce cellular cytotoxicity through decreased necroptosis.

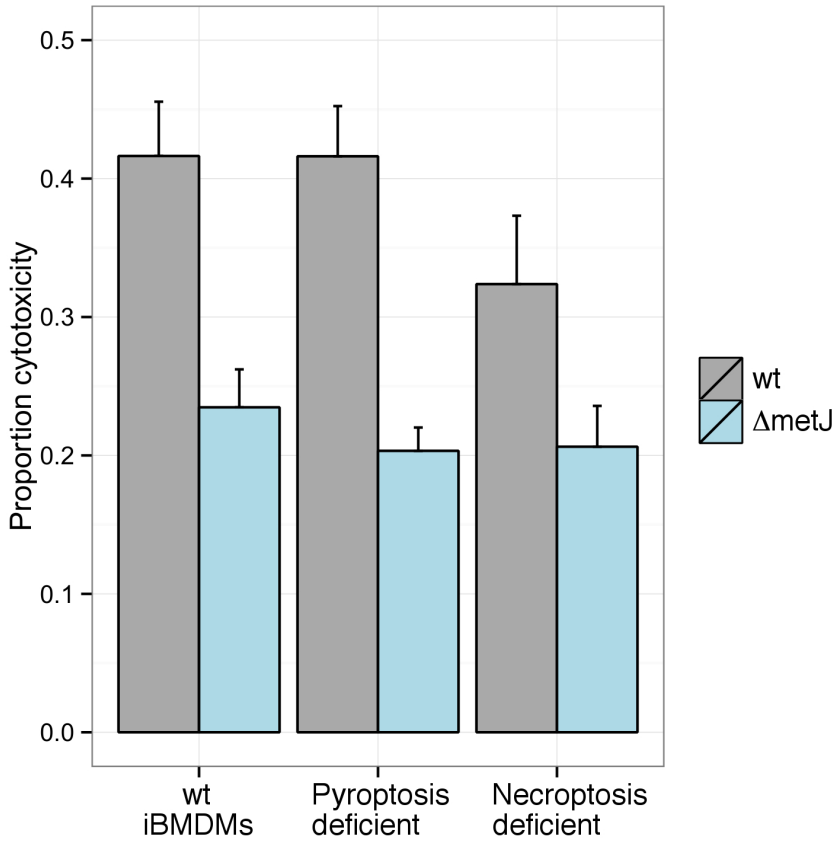

Figure 6.17 Cytotoxicity in pyroptosis- and necroptosis-deficient macrophages infected with *ΔmetJ Salmonella*

Immortalised bone-marrow derived macrophages (from wild-type, *casp1*$^{-/-}$/*casp11*$^{-/-}$, and *ripk3*$^{-/-}$ C57BL/6 mice) were infected with wild-type and *ΔmetJ S.* Typhimurium 12023. The proportion of cellular cytotoxicity was calculated from assayed LDH-release into culture supernatants at 16 hours post-infection. Summary statistics are calculated from three independent experiments. Error bars represent SEM.

Conclusions

By performing a pathway-based enrichment analysis of the NTS bacteraemia GWAS, we identify a single biological pathway significantly enriched for evidence of association with NTS bacteraemia. That signal of enrichment within a gene-set termed “metabolism of polyamines”, is driven by observed NTS-associations in chromosomal regions including *AMD1*, *APIP*, *ENOPH1* and *MTAP*. All of these genes encode enzymes catalysing reactions metabolising methionine and its derivatives. *APIP*, *ENOPH1* and *MTAP* catalyse three of the six steps required to salvage methionine from 5’-*S*-methyl-5’-thioadenosine (MTA), and AMD1 catalyses the decarboxylation of *S*-adenosyl-L-methionine to form *S*-adenosylmethioninamine, the precursor of MTA.

The identification of a set of enzymes central to the metabolism and catabolism of MTA as a determinant of invasive *Salmonella* disease is highly biologically plausible. An *APIP* eQTL has previously been identified as a genetic correlate of *Salmonella*-induced pyroptosis in lymphoid-lineage cells, an effect shown to be secondary to metabolic function of APIP in determining intracellular MTA levels (Ko et al., 2012). That study went on to demonstrate that the APIP eSNP associated with levels of *Salmonella*-induced pyroptosis was a genetic correlate of mortality in American adults with sepsis. Very few, if any, of these individuals will have had sepsis secondary to *Salmonella*, suggesting a role for genetic correlates of MTA/methionine metabolism in immunity to infectious diseases beyond invasive *Salmonella* infections. In

African children, however, we find no robust evidence for association with susceptibility to bacteraemia secondary to any pathogen other than NTS at *AMD1*, *APIP*, *ENOPH1* or *MTAP*. Moreover we find no evidence for association with mortality secondary to bacteraemia, either at the single variant level or performing a permutation-based enrichment analysis, at these loci. The lack of observed replication of the APIP association with sepsis mortality is noteworthy, but not unexpected. The highly disparate study settings, patient characteristics and co-morbidities between a population of American adults admitted to intensive care units with sepsis, and African children admitted to a rural hospital with bacteraemia, make it highly unlikely that determinants of mortality will be shared between the two populations.

Motivated by the observation that trait-associated genetic variation commonly functions to regulate expression of local or distant genes, and more specifically that an eQTL for APIP has been previously associated with *Salmonella*-induced pyroptosis, we assessed the evidence for overlap between significant *cis* eQTLs in immune cell subsets and NTS-associated genetic variation at *AMD1*, *APIP*, *ENOPH1* and *MTAP*. Despite strong evidence for *cis* regulatory determinants of *AMD1* and *MTAP* expression in monocytes, we find no robust evidence for overlap between NTS-associated genetic variation and *cis* eQTLs. This lack of overlap is unexpected, as one would predict that these variants would function as *cis* regulatory determinants. The fact that they were identified by virtue of their proximity to these genes suggest that, if indeed these findings do represent robust evidence of NTS association, that they will function to regulate or modify local

gene function. It seems likely, the possibility that the observed NTS associations represent false positives notwithstanding, that our failure to demonstrate overlap between *cis* eQTLs and NTS-associated variation represents a combination of limited power to adequately map the NTS association, eQTL specificity to a cell-type or a context not represented within the datasets used in the analysis, and the possibility of population specific eQTLs, with the datasets used in the analysis being generated using cells from healthy European adults. The limited power to adequately map NTS associations will be particularly problematic at these loci, given that their identification using an enrichment analysis results in the observed effect sizes being more modest than variants identified using single locus analyses.

The function of intracellular MTA levels in the regulation of pyroptosis is unclear. To test whether bacterial metabolism of MTA during intracellular infection, could also affect the likelihood of host cell pyroptosis, we generated and characterised mutants of *Salmonella* lacking key enzymes and regulators of MTA/methionine metabolism. We generated mutants of *Salmonella* using one-step, λ-red phage recombinase-driven gene disruption by homologous recombination (Datsenko & Wanner, 2000). Of the mutants generated and characterised, *ΔmetJ Salmonella* affects rates of infection-induced cell death in macrophages. Deletion of *metJ* results in reduced levels of host cell death, and phenotype complemented by plasmid expression of *metJ*.

The observed reduction in cell death during infection with *Salmonella* lacking *metJ* is accompanied by reduced production of IL-1β, which suggested that, in

keeping with previous observations that MTA levels determine rates of pyroptosis, *ΔmetJ* mutants may reduce infected-cell death at least in part via an effect on pyroptosis. This suggestion, however, is not in keeping with the observation that the reduction in cell death seen in infections with *ΔmetJ* mutants is persistent in cells deficient in pyroptosis. Interestingly, the reduction in cell death seen in infections with *ΔmetJ* mutants is attenuated in cells deficient in necroptosis. It is also noteworthy that while differential cell death is significantly attenuated in necroptosis-deficient cells, there appears to be a residual effect for *metJ* deletion on cell death (albeit one that in these experiments lacked statistical significance). This apparent discrepancy is consistent with a model in which *metJ* deletion affects more than one cell death pathway, or a model in which metJ deletion leads to a specific reduction in necroptosis, resulting in reduced IL-1β production via non-canonical, caspase-8 mediated processing (Moriwaki, Bertin, Gough, & Chan, 2015).

The *metJ* locus encodes a transcriptional aporepressor, which, along with its co-repressor *S*-adenosyl-L-methionine, represses loci encoding methionine biosynthetic enzymes and transporters (Old et al., 1991). RNASeq data has demonstrated that *metJ* expression is significantly upregulated during infection in macrophages. This observation may simply reflect changes in methionine availability as the bacterium transitions from an extracellular to an intracellular environment. However, considered alongside our data demonstrating a role for metJ in host cell death, this would also be consistent with a model in which *Salmonella* regulates its metabolism, facilitating host cell death, and thus bacterial dissemination.

Chapter 7 – Discussion

Conclusions

The aim of this thesis was to identify, validate and functionally characterise novel genetic correlates of NTS bacteraemia in African children. While many genetic correlates of invasive NTS disease had been previously characterised in the mouse model of invasive *Salmonella* infection, alongside the identification of a growing number of host genetic determinants of invasive *Salmonella* infection in humans with rare primary immunodeficiencies (Gilchrist et al., 2015), the sole, validated genetic correlate of invasive NTS disease in unselected populations remains the sickle cell locus (Williams et al., 2009).

In chapters three and four, performing a GWAS of NTS bacteraemia in Kenyan children, followed by a replication analysis in Kenyan and Malawian children, we identified, and replicated, an association between NTS bacteraemia and a genetic variant in *STAT4*, rs13390936. rs13390936 is a context-specific expression quantitative trait locus for STAT4 RNA expression, and modifies IFNγ protein production in stimulated natural killer cells. We further demonstrate that, genotype at rs13390936 modifies total serum IFNγ

levels in African children with acute invasive NTS disease. Taken together, our data establish IFNγ production capacity as a key determinant of susceptibility to invasive NTS disease in African children.

In chapter five, I present data supporting genetic variation in *EVI5L* as a correlate of NTS bacteraemia. We were unable to replicate this association in either the Kenyan or Malawian replication sample collections, but were considerably underpowered to do so. However, the highly significant enrichment of the NTS-associated SNPs in the *EVI5L* region, both for *EVI5L* cis eSNPs and for correlates of IFNγ production in *Chlamydia*-infected lymphoid lineage cells, lend considerable weight to the candidacy of *EVI5L* as an NTS-susceptibility locus. Preliminary functional experiments exploring the role of EVI5L in IFNγ production identify an eQTL for *EVI5L* in naïve monocytes as determinant of STAT4 phosphorylation and IFNγ production in stimulated $CD4^+$ T cells.

In chapter six, I describe the results of a pathway-based enrichment analysis of the NTS bacteraemia GWAS, identifying genetic variation in the methionine salvage pathway as a determinant of invasive NTS disease in African children. Expanding on previously-published observations that levels of MTA, the methionine salvage precursor, acts as a determinant of *Salmonella*-induced pyroptosis, we generated mutants deficient in key aspects of MTA metabolism and characterised them with respect to induction of cell death in macrophages. We demonstrate that deletion of a transcriptional repressor of methionine metabolism, *metJ*, results in attenuated cell death in macrophages

in manner that appears, at least in part, to be dependent on RIPK3, and thus necroptosis. To explain these observations, we propose a hypothesis in which, sensing of flux through a shared host-pathogen metabolic pathway facilitates detection by the host of an intracellular pathogen, priming protective inflammatory cell death pathways, e.g. pyroptosis or necroptosis. By regulating its own metabolism, *Salmonella* may be able to control the likelihood of its intracellular invasion resulting in host cell death. Our data provide some support for this model but further experiments will be required to establish or refute this hypothesis.

The identification of NTS-associated genetic variation at two independent loci, *STAT4* and *EVI5L*, with apparent roles in IFNγ production provides compelling evidence for the key role of IFNγ-mediated immunity in control of *Salmonella* infection in African children. The identification of the STAT4 locus in particular provides an informative example of the shared biology underlying susceptibility to a single pathogen in patients with rare primary immunodeficiencies and individuals in unselected populations. Rare individuals with MSMD, with deleterious mutations in genes affecting IL-12-dependent IFNγ–mediated immunity other than *STAT4*, are highly susceptible to invasive NTS infections. These data extend our understanding of the importance of IFNγ-mediated immunity in host *Salmonella*-control, from individuals with rare immunodeficiencies to unselected populations of African children. The identification of host genetic correlates of IFNγ-mediated immunity as a risk factor for NTS disease is also very much in keeping with HIV infection as a dominant acquired risk factor for invasive NTS disease in

Africa, with $CD4^+$ T cell depletion, characterised by early loss of the T_H1 compartment (Geldmacher et al., 2008) and *Salmonella*-specific IFNγ responses, being a key predictor of NTS disease susceptibility.

Limitations and future directions

There are several noteworthy limitations to the work presented in this thesis, which are important to highlight and will guide future research efforts. The work's key limitation has been the modest sample sizes, particularly with respect to case samples, of the sample collections available for inclusion in genetic association analyses of NTS bacteraemia. This limitation restricted our power to detect novel associations in the discovery analysis, but most importantly with respect to the work presented here, limited our power to replicate associations. As far as possible, we alleviated that limitation within the analysis as presented, focussing solely on very common genetic variants. Nonetheless, our lack of power means that it remains difficult to interpret our attempts at replication for several loci, most notably for *EVI5L*. It is hoped that on-going sample collection efforts will allow us to better-power future replication efforts, and validate further NTS-associated loci.

A second noteworthy limitation to the work is its reliance on only two study populations. While we provide robust evidence that common genetic variation in *STAT4* is a determinant of NTS bacteraemia in Kenyan and Malawian children, the extent to which this is generalizable to other populations of African children is unclear. A related limitation is our inability to include key

acquired risk factors (most importantly HIV infection) in our genetic association analyses. We were unable to do this discovery analysis, as the control samples represent a birth cohort with no data on HIV infection. That said even were the relevant data available, we would be underpowered to conduct a stratified analysis. It seems plausible that the well-documented differences in the relative importance of acquired risk factors for NTS bacteraemia in African children(Oneko et al., 2015; Verani et al., 2015) in different settings could affect the penetrance of a genetic susceptibility factor. Again, on-going sample collection efforts in settings other than coastal Kenya and Malawi will allow us to begin to address these questions.

The analysis of cell-type and context-specific eQTL datasets alongside the NTS GWAS has been enormously valuable in deriving insights into the likely function of identified NTS-associated genetic variation, particularly with respect to *STAT4*. It is noteworthy, however, that these eQTL datasets were generated with samples from individuals of European ancestry. While the degree of population-specificity of eQTLs may have previously been overstated, there clearly are instances of population-specific regulatory genetic variation(Raj et al., 2014). There is likely to be additive value in annotating African genetic association studies with African-specific eQTL datasets. We hope to be able to address this in the near future, performing an eQTL study in a well-characterised cohort of Ugandan children.

Finally there are several further key questions posed by the work presented that will direct future work. While we define the consequences of the NTS-

associated *STAT4* variant on IFNγ production, we did not design these experiments to assess any potential role for *STAT4* genetic variation in anti-*Salmonella* T_H17 responses. The growing appreciation of the role for T_H17 responses and neutrophils in host control of *Salmonella* infection (Raffatellu et al., 2008), and role of STAT4 in T_H17 polarisation in transducing IL-23 signalling (Figure 1.8), will make experiments designed to assess this particularly interesting to undertake. With respect to the data describing EVI5L as an NTS-susceptibility locus, we have experiments in progress to further explore the role of EVI5L on IFNγ production using RNAi.

Additionally, we have experiments planned to further characterise the *ΔmetJ Salmonella* mutant, initially to characterise the effect of metJ deletion on intracellular concentrations of methionine metabolites. More broadly, the integration of host and pathogen genomic studies within current and future sample collections could help to address the relative contribution of host susceptibility and pathogen virulence to the likelihood of invasive NTS infection in African children.

References

Adeyokunnu, A. A., & Hendrickse, R. G. (1980). Salmonella osteomyelitis in childhood. A report of 63 cases seen in Nigerian children of whom 57 had sickle cell anaemia. *Archives of Disease in Childhood*, *55*(3), 175–184.

Ahmad, D. S., Esmadi, M., & Steinmann, W. C. (2013). Idiopathic CD4 Lymphocytopenia: Spectrum of opportunistic infections, malignancies, and autoimmune diseases. *Avicenna Journal of Medicine*, *3*(2), 37–47. http://doi.org/10.4103/2231-0770.114121

Ali, S., Vollaard, A. M., Kremer, D., de Visser, A. W., Martina, C. A. E., Widjaja, S., et al. (2007). Polymorphisms in Proinflammatory Genes and Susceptibility to Typhoid Fever and Paratyphoid Fever. *Journal of Interferon & Cytokine Research*, *27*(4), 271–280. http://doi.org/10.1089/jir.2006.0129

Ali, S., Vollaard, A. M., Widjaja, S., Surjadi, C., van de Vosse, E., & van Dissel, J. T. (2006). PARK2/PACRG polymorphisms and susceptibility to typhoid and paratyphoid fever. *Clinical and Experimental Immunology*, *144*(3), 425–431. http://doi.org/10.1111/j.1365-2249.2006.03087.x

Altshuler, D., Daly, M. J., & Lander, E. S. (2008). Genetic mapping in human disease. *Science*, *322*(5903), 881–888. http://doi.org/10.1126/science.1156409

Ampel, N. M., Van Wyck, D. B., Aguirre, M. L., Willis, D. G., & Popp, R. A. (1989). Resistance to infection in murine beta-thalassemia. *Infection and Immunity*, *57*(4), 1011–1017.

Ao, T. T., Feasey, N. A., Gordon, M. A., Keddy, K. H., Angulo, F. J., & Crump, J. A. (2015). Global burden of invasive nontyphoidal salmonella disease, 2010(1). *Emerging Infectious Diseases*, *21*(6). http://doi.org/10.3201/eid2106.140999

Arpaia, N., Godec, J., Lau, L., Sivick, K. E., McLaughlin, L. M., Jones, M. B., et al. (2011). TLR signaling is required for Salmonella typhimurium virulence. *Cell*, *144*(5), 675–688. http://doi.org/10.1016/j.cell.2011.01.031

Averbuch, D., Chapgier, A., Boisson-Dupuis, S., Casanova, J.-L., & Engelhard, D. (2011). The clinical spectrum of patients with deficiency of Signal Transducer and Activator of Transcription-1. *The Pediatric Infectious Disease Journal*, *30*(4), 352–355. http://doi.org/10.1097/INF.0b013e3181fdff4a

Bader, M. W., Sanowar, S., Daley, M. E., Schneider, A. R., Cho, U., Xu, W., et al. (2005). Recognition of antimicrobial peptides by a bacterial sensor kinase. *Cell*, *122*(3), 461–472. http://doi.org/10.1016/j.cell.2005.05.030

Barrow, P. A., Simpson, J. M., Lovell, M. A., & Binns, M. M. (1987). Contribution of Salmonella gallinarum large plasmid toward virulence in fowl typhoid. *Infection and Immunity*, *55*(2), 388–392.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B*

Berkley, J. A., Bejon, P., Mwangi, T., Gwer, S., Maitland, K., Williams, T. N., et al. (2009). HIV infection, malnutrition, and invasive bacterial infection among children with severe malaria. *Clinical Infectious Diseases : an Official Publication of the Infectious Diseases Society of America*, *49*(3), 336–343. http://doi.org/10.1086/600299

Berkley, J. A., Lowe, B. S., Mwangi, I., Williams, T., Bauni, E., Mwarumba, S., et al. (2005). Bacteremia among children admitted to a rural hospital in Kenya. *The New England Journal of Medicine*, *352*(1), 39–47. http://doi.org/10.1056/NEJMoa040275

Bhuvanendran, S., Hussin, H. M., Meran, L. P., Anthony, A. A., Zhang, L., Burch, L. H., et al. (2011). Toll-like receptor 4 Asp299Gly and Thr399Ile polymorphisms and typhoid susceptibility in Asian Malay population in Malaysia. *Microbes and Infection*, *13*(10), 844–851. http://doi.org/10.1016/j.micinf.2011.04.007

Birmingham, C. L., Smith, A. C., Bakowski, M. A., Yoshimori, T., & Brumell, J. H. (2006). Autophagy controls Salmonella infection in response to damage to the Salmonella-containing vacuole. *The Journal of Biological Chemistry*, *281*(16), 11374–11383. http://doi.org/10.1074/jbc.M509157200

Blum, G., Ott, M., Lischewski, A., Ritter, A., Imrich, H., Tschäpe, H., & Hacker, J. (1994). Excision of large DNA regions termed pathogenicity islands from tRNA-specific loci in the chromosome of an Escherichia coli wild-type pathogen. *Infection and Immunity*, *62*(2), 606–614.

Brent, A. J., Oundo, J. O., Mwangi, I., Ochola, L., Lowe, B., & Berkley, J. A. (2006). Salmonella Bacteremia in Kenyan Children. *The Pediatric Infectious Disease Journal*, *25*(3), 230–236. http://doi.org/10.1097/01.inf.0000202066.02212.ff

Brewster, D. R., Manary, M. J., Menzies, I. S., O'Loughlin, E. V., & Henry, R. L. (1997). Intestinal permeability in kwashiorkor. *Archives of Disease in Childhood*, *76*(3), 236–241.

Bronzan, R. N., Taylor, T. E., Mwenechanya, J., Tembo, M., Kayira, K., Bwanaisa, L., et al. (2007). Bacteremia in Malawian children with severe malaria: prevalence, etiology, HIV coinfection, and outcome. *The Journal of Infectious Diseases*, *195*(6), 895–904. http://doi.org/10.1086/511437

Browne, S. K., Burbelo, P. D., Chetchotisakd, P., Suputtamongkol, Y., Kiertiburanakul, S., Shaw, P. A., et al. (2012). Adult-onset immunodeficiency in Thailand and Taiwan. *The New England Journal of Medicine*, *367*(8), 725–734. http://doi.org/10.1056/NEJMoa1111160

Broz, P., Newton, K., Lamkanfi, M., Mariathasan, S., Dixit, V. M., & Monack, D. M. (2010). Redundant roles for inflammasome receptors NLRP3 and NLRC4 in host defense against Salmonella. *Journal of Experimental Medicine*, *207*(8), 1745–1755. http://doi.org/10.1084/jem.20100257

Broz, P., Ruby, T., Belhocine, K., Bouley, D. M., Kayagaki, N., Dixit, V. M., & Monack, D. M. (2013). Caspase-11 increases susceptibility to Salmonella infection in the absence of caspase-1. *Nature*, *490*(7419), 288–291. http://doi.org/10.1038/nature11419

Burton, P. R., Cardon, L. R., Duncanson, A., Kwiatkowski, D. P., Barrett, J. C., Davison, D., et al. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, *447*(7145), 661–678. http://doi.org/10.1038/nature05911

Burton, P. R., Tobin, M. D., & Hopper, J. L. (2005). Key concepts in genetic epidemiology. *Lancet*, *366*(9489), 941–951. http://doi.org/10.1016/S0140-6736(05)67322-9

Carden, S., Okoro, C., Dougan, G., & Monack, D. (2015). Non-typhoidal Salmonella Typhimurium ST313 isolates that cause bacteremia in humans stimulate less inflammasome activation than ST19 isolates associated with gastroenteritis. *Pathogens and Disease*, *73*(4), ftu023–ftu023. http://doi.org/10.1093/femspd/ftu023

Chakravortty, D., Hansen-Wester, I., & Hensel, M. (2002). Salmonella pathogenicity island 2 mediates protection of intracellular Salmonella from reactive nitrogen intermediates. *The Journal of Experimental Medicine*, *195*(9), 1155–1166. http://doi.org/10.1084/jem.20011547

Chapman, S. J., & Hill, A. V. S. (2012). Human genetic susceptibility to infectious disease. *Nature Reviews Genetics*, *13*(3), 175–188. http://doi.org/10.1038/nrg3114

Cole, S. T., Eiglmeier, K., Parkhill, J., James, K. D., Thomson, N. R., Wheeler, P. R., et al. (2001). Massive gene decay in the leprosy bacillus. *Nature*, *409*(6823), 1007–1011. http://doi.org/10.1038/35059006

Conlan, J. W. (1997). Critical roles of neutrophils in host defense against experimental systemic infections of mice by Listeria monocytogenes, Salmonella typhimurium, and Yersinia enterocolitica. *Infection and Immunity*, *65*(2), 630–635.

Conway, K. L., Kuballa, P., Song, J.-H., Patel, K. K., Castoreno, A. B., Yilmaz, O. H., et al. (2013). Atg16l1 is required for autophagy in intestinal epithelial cells and protection of mice from Salmonella infection. *Gastroenterology*, *145*(6), 1347–1357. http://doi.org/10.1053/j.gastro.2013.08.035

Cortes, A., & Brown, M. A. (2011). Promise and pitfalls of the Immunochip. *Arthritis Research & Therapy*, *13*(1), 101. http://doi.org/10.1186/ar3204

Cunnington, A. J., de Souza, J. B., Walther, M., & Riley, E. M. (2012a). Malaria impairs resistance to Salmonella through heme- and heme oxygenase-dependent dysfunctional granulocyte mobilization. *Nature Medicine*, *18*(1), 120–127. http://doi.org/10.1038/nm.2601

Cunnington, A. J., Njie, M., Correa, S., Takem, E. N., Riley, E. M., & Walther, M. (2012b). Prolonged neutrophil dysfunction after Plasmodium falciparum malaria is related to hemolysis and heme oxygenase-1 induction. *The Journal of Immunology*, *189*(11), 5336–5346. http://doi.org/10.4049/jimmunol.1201028

Dandekar, S., George, M. D., & Bäumler, A. J. (2010). Th17 cells, HIV and the gut mucosal barrier. *Current Opinion in HIV and AIDS*, *5*(2), 173–178. http://doi.org/10.1097/COH.0b013e328335eda3

Datsenko, K. A., & Wanner, B. L. (2000). One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(12), 6640–6645. http://doi.org/10.1073/pnas.120163297

de Beaucoudrey, L., Samarina, A., Bustamante, J., Cobat, A., Boisson-Dupuis, S., Feinberg, J., et al. (2010). Revisiting Human IL-12Rβ1 Deficiency. *Medicine*, *89*(6), 381–402. http://doi.org/10.1097/MD.0b013e3181fdd832

Delaneau, O., Marchini, J., & Zagury, J.-F. (2012). A linear complexity phasing method for thousands of genomes. *Nature Methods*, *9*(2), 179–181. http://doi.org/10.1038/nmeth.1785

Dharmana, E., Joosten, I., Tijssen, H. J., Gasem, M. H., Indarwidayati, R., Keuter, M., et al. (2002). HLA-DRB1*12 is associated with protection against complicated typhoid fever, independent of tumour necrosis factor alpha. *European Journal of Immunogenetics : Official Journal of the British Society for Histocompatibility and Immunogenetics*, *29*(4), 297–300.

Dilthey, A., Leslie, S., Moutsianas, L., Shen, J., Cox, C., Nelson, M. R., & McVean, G. (2013). Multi-population classical HLA type imputation. *PLoS Computational Biology*, *9*(2), e1002877. http://doi.org/10.1371/journal.pcbi.1002877

Dione, M. M., Ikumapayi, U. N., Saha, D., Mohammed, N. I., Geerts, S., Ieven, M., et al. (2011). Clonal differences between Non-Typhoidal Salmonella (NTS) recovered from children and animals living in close contact in the Gambia. *PLoS Neglected Tropical Diseases*, *5*(5), e1148. http://doi.org/10.1371/journal.pntd.0001148

Doolittle, R. F., Feng, D. F., Tsang, S., Cho, G., & Little, E. (1996). Determining divergence times of the major kingdoms of living organisms with a protein clock. *Science*, *271*(5248), 470–477.

Dorman, S. E., Picard, C., Lammas, D., Heyne, K., van Dissel, J. T., Baretto, R., et al. (2004). Clinical features of dominant and recessive interferon gamma receptor 1 deficiencies. *Lancet*, *364*(9451), 2113–2121. http://doi.org/10.1016/S0140-6736(04)17552-1

Dunn, O. J. (1961). Multiple Comparisons among Means. *Journal of the American Statistical Association*, *56*(293), 52–64.

Dunstan, S. J., Hawn, T. R., Hue, N. T., Parry, C. P., Ho, V. A., Vinh, H., et al. (2005). Host susceptibility and clinical outcomes in toll-like receptor 5-deficient patients with typhoid fever in Vietnam. *The Journal of Infectious Diseases*, *191*(7), 1068–1071. http://doi.org/10.1086/428593

Dunstan, S. J., Ho, V. A., Duc, C. M., Lanh, M. N., Phuong, C. X., Luxemburger, C., et al. (2001). Typhoid fever and genetic polymorphisms at the natural resistance-associated macrophage protein 1. *The Journal of Infectious Diseases*, *183*(7), 1156–1160. http://doi.org/10.1086/319289

Dunstan, S. J., Hue, N. T., Han, B., Li, Z., Tram, T. T. B., Sim, K. S., et al. (2014). Variation at HLA-DRB1 is associated with resistance to enteric fever. *Nature Genetics*, *46*(12), 1333–1336. http://doi.org/10.1038/ng.3143

Dunstan, S. J., Hue, N. T., Rockett, K., Forton, J., Morris, A. P., Diakite, M., et al. (2007). A TNF region haplotype offers protection from typhoid fever in Vietnamese patients. *Human Genetics*, *122*(1), 51–61. http://doi.org/10.1007/s00439-007-0372-9

Elston, R. C. (1981). Segregation analysis. *Advances in Human Genetics*, *11*, 63–120– 372–3.

Emilsson, V., Thorleifsson, G., Zhang, B., Leonardson, A. S., Zink, F., Zhu, J., et al. (2008). Genetics of gene expression and its effect on disease. *Nature*, *452*(7186), 423–428. http://doi.org/10.1038/nature06758

Enwere, G., Biney, E., Cheung, Y. B., Zaman, S. M. A., Okoko, B., Oluwalana, C., et al. (2006). Epidemiologic and clinical characteristics of community-acquired invasive bacterial infections in children aged 2-29 months in The Gambia. *The Pediatric Infectious Disease Journal*, *25*(8), 700–705. http://doi.org/10.1097/01.inf.0000226839.30925.a5

Eva, M. M., Yuki, K. E., Dauphinee, S. M., Schwartzentruber, J. A., Pyzik, M., Paquet, M., et al. (2013). Altered IFN- -Mediated Immunity and Transcriptional Expression Patterns in N-Ethyl-N-Nitrosourea-Induced STAT4 Mutants Confer Susceptibility to Acute Typhoid-like Disease. *The Journal of Immunology*. http://doi.org/10.4049/jimmunol.1301370

Everest, P., Roberts, M., & Dougan, G. (1998). Susceptibility to Salmonella typhimurium infection and effectiveness of vaccination in mice deficient in the tumor necrosis factor alpha p55 receptor. *Infection and Immunity*, *66*(7), 3355–3364.

Eyre, S., Bowes, J., Diogo, D., Lee, A., Barton, A., Martin, P., et al. (2012). High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. *Nature Genetics*, *44*(12), 1336–1340. http://doi.org/10.1038/ng.2462

Fairfax, B. P., Humburg, P., Makino, S., Naranbhai, V., Wong, D., Lau, E., et al. (2014). Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science*, *343*(6175), 1246949–1246949. http://doi.org/10.1126/science.1246949

Fairfax, B. P., Makino, S., Radhakrishnan, J., Plant, K., Leslie, S., Dilthey, A., et al. (2012). Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. *Nature Genetics*, *44*(5), 502–510. http://doi.org/10.1038/ng.2205

Falay, D., Kuijpers, L. M. F., Phoba, M.-F., De Boeck, H., Lunguya, O., Vakaniaki, E., et al. (2016). Microbiological, clinical and molecular findings of non-typhoidal Salmonella bloodstream infections associated with malaria, Oriental Province, Democratic Republic of the Congo. *BMC Infectious Diseases*, *16*(1), 271. http://doi.org/10.1186/s12879-016-1604-1

Faraco, J., Lin, L., Kornum, B. R., Kenny, E. E., Trynka, G., Einen, M., et al. (2013). ImmunoChip study implicates antigen presentation to T cells in narcolepsy. *PLoS Genetics*, *9*(2), e1003270. http://doi.org/10.1371/journal.pgen.1003270

Feasey, N. A., Archer, B. N., Heyderman, R. S., Sooka, A., Dennis, B., Gordon, M. A., & Keddy, K. H. (2010). Typhoid fever and invasive nontyphoid salmonellosis, Malawi and South Africa. *Emerging Infectious Diseases*, *16*(9), 1448–1451. http://doi.org/10.3201/eid1609.100125

Feasey, N. A., Cain, A. K., Msefula, C. L., Pickard, D., Alaerts, M., Aslett, M., et al. (2014a). Drug resistance in Salmonella enterica ser. Typhimurium bloodstream infection, Malawi. *Emerging Infectious Diseases*, *20*(11), 1957–1959. http://doi.org/10.3201/eid2011.141175

Feasey, N. A., Everett, D., Faragher, E. B., Roca-Feltrer, A., Kang'ombe, A., Denis, B., et al. (2015a). Modelling the Contributions of Malaria, HIV, Malnutrition and Rainfall to the Decline in Paediatric Invasive Non-typhoidal Salmonella Disease in Malawi. *PLoS Neglected Tropical Diseases*, *9*(7), e0003979. http://doi.org/10.1371/journal.pntd.0003979

Feasey, N. A., Houston, A., Mukaka, M., Komrower, D., Mwalukomo, T., Tenthani, L., et al. (2014b). A reduction in adult blood stream infection and case fatality at a large African hospital following antiretroviral therapy roll-out. *PLoS ONE*, *9*(3), e92226. http://doi.org/10.1371/journal.pone.0092226

Feasey, N. A., Masesa, C., Jassi, C., Faragher, E. B., Mallewa, J., Mallewa, M., et al. (2015b). Three Epidemics of Invasive Multidrug-Resistant Salmonella Bloodstream Infection in Blantyre, Malawi, 1998-2014. *Clinical Infectious Diseases : an Official Publication of the Infectious Diseases Society of America*, *61 Suppl 4*(suppl 4), S363–71. http://doi.org/10.1093/cid/civ691

Fellay, J., Shianna, K. V., Ge, D., Colombo, S., Ledergerber, B., Weale, M., et al. (2007). A whole-genome association study of major determinants for host control of HIV-1. *Science*, *317*(5840), 944–947. http://doi.org/10.1126/science.1143767

Feuillet, V., Medjane, S., Mondor, I., Demaria, O., Pagni, P. P., Galán, J. E., et al. (2006). Involvement of Toll-like receptor 5 in the recognition of flagellated bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(33), 12487–12492. http://doi.org/10.1073/pnas.0605200103

Fierer, J., Krause, M., Tauxe, R., & Guiney, D. (1992). Salmonella typhimurium bacteremia: association with the virulence plasmid. *The Journal of Infectious Diseases*, *166*(3), 639–642.

Filipe-Santos, O. (2006). X-linked susceptibility to mycobacteria is caused by mutations in NEMO impairing CD40-dependent IL-12 production. *Journal of Experimental Medicine*, *203*(7), 1745–1759. http://doi.org/10.1084/jem.20060085

Fox, G. J., Orlova, M., & Schurr, E. (2016). Tuberculosis in Newborns: The Lessons of the “Lübeck Disaster” (1929-1933). *PLoS Pathogens*, *12*(1), e1005271. http://doi.org/10.1371/journal.ppat.1005271

Franchi, L., Warner, N., Viani, K., & Núñez, G. (2009). Function of Nod-like receptors in microbial recognition and host defense. *Immunological Reviews*, *227*(1), 106–128. http://doi.org/10.1111/j.1600-065X.2008.00734.x

Francis, C. L., Starnbach, M. N., & Falkow, S. (1992). Morphological and cytoskeletal changes in epithelial cells occur immediately upon interaction with Salmonella typhimurium grown under low-oxygen conditions. *Molecular Microbiology*, *6*(21), 3077–3087.

Frasa, M. A. M., Koessmeier, K. T., Ahmadian, M. R., & Braga, V. M. M. (2012). Illuminating the functional and structural repertoire of human TBC/RABGAPs. *Nature Reviews Molecular Cell Biology*. http://doi.org/10.1038/nrm3267

Fritsche, G., Nairz, M., Libby, S. J., Fang, F. C., & Weiss, G. (2012). Slc11a1 (Nramp1) impairs growth of Salmonella enterica serovar typhimurium in macrophages via stimulation of lipocalin-2 expression. *Journal of Leukocyte Biology*, *92*(2), 353–359. http://doi.org/10.1189/jlb.1111554

Galán, J. E., & Curtiss, R. (1989). Cloning and molecular characterization of genes whose products allow Salmonella typhimurium to penetrate tissue culture cells. *Proceedings of the National Academy of Sciences of the United States of America*, *86*(16), 6383–6387.

García Véscovi, E., Soncini, F. C., & Groisman, E. A. (1996). Mg2+ as an extracellular signal: environmental regulation of Salmonella virulence. *Cell*, *84*(1), 165–174.

Geldmacher, C., Schuetz, A., Ngwenyama, N., Casazza, J. P., Sanga, E., Saathoff, E., et al. (2008). Early depletion of Mycobacterium tuberculosis-specific T helper 1 cell responses after HIV-1 infection. *The Journal of Infectious Diseases*, *198*(11), 1590–1598. http://doi.org/10.1086/593017

Giammanco, G. M., Pignato, S., Mammina, C., Grimont, F., Grimont, P. A. D., Nastasi, A., & Giammanco, G. (2002). Persistent endemicity of Salmonella bongori 48:z(35):--in Southern Italy: molecular characterization of human, animal, and environmental isolates. *Journal of Clinical Microbiology*, *40*(9), 3502–3505. http://doi.org/10.1128/JCM.40.9.3502-3505.2002

Gibbons, H. S., Lin, S., Cotter, R. J., & Raetz, C. R. (2000). Oxygen requirement for the biosynthesis of the S-2-hydroxymyristate moiety in Salmonella typhimurium lipid A. Function of LpxO, A new Fe2+/alpha-ketoglutarate-dependent dioxygenase homologue. *The Journal of Biological Chemistry*, *275*(42), 32940–32949. http://doi.org/10.1074/jbc.M005779200

Gilchrist, J. J., MacLennan, C. A., & Hill, A. V. S. (2015). Genetic susceptibility to invasive Salmonella disease. *Nature Reviews Immunology*, *15*(7), 452–463. http://doi.org/10.1038/nri3858

Gondwe, E. N., Molyneux, M. E., Goodall, M., Graham, S. M., Mastroeni, P., Drayson, M. T., & MacLennan, C. A. (2010). Importance of antibody and complement for oxidative burst and killing of invasive nontyphoidal Salmonella by blood cells in Africans. *Proceedings of the National Academy of Sciences*, *107*(7), 3070–3075. http://doi.org/10.1073/pnas.0910497107

Gordon, M. A., Banda, H. T., Gondwe, M., Gordon, S. B., Boeree, M. J., Walsh, A. L., et al. (2002). Non-typhoidal salmonella bacteraemia among HIV-infected Malawian adults: high mortality and frequent recrudescence. *AIDS (London, England)*, *16*(12), 1633–1641.

Gordon, M. A., Graham, S. M., Walsh, A. L., Wilson, L., Phiri, A., Molyneux, E., et al. (2008). Epidemics of invasive Salmonella enterica serovar enteritidis and S. enterica Serovar typhimurium infection associated with multidrug resistance among adults and children in Malawi. *Clinical Infectious Diseases : an Official Publication of the Infectious Diseases Society of America*, *46*(7), 963–969. http://doi.org/10.1086/529146

Govoni, G., Gauthier, S., Billia, F., Iscove, N. N., & Gros, P. (1997). Cell-specific and inducible Nramp1 gene expression in mouse macrophages in vitro and in vivo. *Journal of Leukocyte Biology*, *62*(2), 277–286.

Graham, S. M., Walsh, A. L., Molyneux, E. M., Phiri, A. J., & Molyneux, M. E. (2000). Clinical presentation of non-typhoidal Salmonella bacteraemia in Malawian children. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *94*(3), 310–314.

Groisman, E. A., & Ochman, H. (1996). Pathogenicity islands: bacterial evolution in quantum leaps. *Cell*, *87*(5), 791–794.

Guiney, D. G., Fang, F. C., Krause, M., Libby, S., Buchmeier, N. A., & Fierer, J. (1995). Biology and clinical significance of virulence plasmids in Salmonella serovars. *Clinical Infectious Diseases : an Official Publication of the Infectious Diseases Society of America*, *21 Suppl 2*, S146–51.

Gunn, J. S., Ryan, S. S., Van Velkinburgh, J. C., Ernst, R. K., & Miller, S. I. (2000). Genetic and functional analysis of a PmrA-PmrB-regulated locus necessary for lipopolysaccharide modification, antimicrobial peptide resistance, and oral virulence of Salmonella enterica serovar typhimurium. *Infection and Immunity*, *68*(11), 6139–6146.

Guo, L., Lim, K. B., Gunn, J. S., Bainbridge, B., Darveau, R. P., Hackett, M., & Miller, S. I. (1997). Regulation of lipid A modifications by Salmonella typhimurium virulence genes phoP-phoQ. *Science*, *276*(5310), 250–253.

Guo, L., Lim, K. B., Poduje, C. M., Daniel, M., Gunn, J. S., Hackett, M., & Miller, S. I. (1998). Lipid A acylation and bacterial resistance against vertebrate antimicrobial peptides. *Cell*, *95*(2), 189–198.

Gurdasani, D., Carstensen, T., Tekola-Ayele, F., Pagani, L., Tachmazidou, I., Hatzikotoulas, K., et al. (2015). The African Genome Variation Project shapes medical genetics in Africa. *Nature*, *517*(7534), 327–332. http://doi.org/10.1038/nature13997

Haneda, T., Ishii, Y., Shimizu, H., Ohshima, K., Iida, N., Danbara, H., & Okada, N. (2012). Salmonella type III effector SpvC, a phosphothreonine lyase, contributes to reduction in inflammatory response during intestinal phase of infection. *Cellular Microbiology*, *14*(4), 485–499. http://doi.org/10.1111/j.1462-5822.2011.01733.x

Hanson, E. P., Monaco-Shawver, L., Solt, L. A., Madge, L. A., Banerjee, P. P., May, M. J., & Orange, J. S. (2008). Hypomorphic nuclear factor-kappaB essential modulator mutation database and reconstitution system identifies phenotypic and immunologic diversity. *The Journal of Allergy and Clinical Immunology*, *122*(6), 1169–1177.e16. http://doi.org/10.1016/j.jaci.2008.08.018

Hardt, W. D., Chen, L. M., Schuebel, K. E., Bustelo, X. R., & Galán, J. E. (1998). S. typhimurium encodes an activator of Rho GTPases that induces membrane ruffling and nuclear responses in host cells. *Cell*, *93*(5), 815–826.

Hardy, G. H. (1908). MENDELIAN PROPORTIONS IN A MIXED POPULATION. *Science*, *28*(706), 49–50. http://doi.org/10.1126/science.28.706.49

Hayward, R. D., & Koronakis, V. (1999). Direct nucleation and bundling of actin by the SipC protein of invasive Salmonella. *The EMBO Journal*, *18*(18), 4926–4934. http://doi.org/10.1093/emboj/18.18.4926

Heffernan, E. J., Reed, S., Hackett, J., Fierer, J., Roudier, C., & Guiney, D. (1992). Mechanism of resistance to complement-mediated killing of bacteria encoded by the Salmonella typhimurium virulence plasmid gene rck. *Journal of Clinical Investigation*, *90*(3), 953–964. http://doi.org/10.1172/JCI115972

Hensel, M., Shea, J. E., Gleeson, C., Jones, M. D., Dalton, E., & Holden, D. W. (1995). Simultaneous identification of bacterial virulence genes by negative selection. *Science*, *269*(5222), 400–403.

Hermaszewski, R. A., & Webster, A. D. (1993). Primary hypogammaglobulinaemia: a survey of clinical manifestations and complications. *The Quarterly Journal of Medicine*, *86*(1), 31–42.

Hess, J., Ladel, C., Miko, D., & Kaufmann, S. H. (1996). Salmonella typhimurium aroA- infection in gene-targeted immunodeficient mice: major role of CD4+ TCR-alpha beta cells and IFN-gamma in bacterial clearance independent of intracellular location. *Journal of Immunology (Baltimore, Md. : 1950)*, *156*(9), 3321–3326.

Heurlier, K., Vendeville, A., Halliday, N., Green, A., Winzer, K., Tang, C. M., & Hardie, K. R. (2009). Growth deficiencies of Neisseria meningitidis pfs and luxS mutants are not due to inactivation of quorum sensing. *Journal of Bacteriology*, *191*(4), 1293–1302. http://doi.org/10.1128/JB.01170-08

Hinks, A., Cobb, J., Marion, M. C., Prahalad, S., Sudman, M., Bowes, J., et al. (2013). Dense genotyping of immune-related disease regions identifies 14 new susceptibility loci for juvenile idiopathic arthritis. *Nature Genetics*, *45*(6), 664–669. http://doi.org/10.1038/ng.2614

Hohmann, E. L. (2001). Nontyphoidal salmonellosis. *Clinical Infectious Diseases : an Official Publication of the Infectious Diseases Society of America*, *32*(2), 263–269. http://doi.org/10.1086/318457

Holt, K. E., Thomson, N. R., Wain, J., Langridge, G. C., Hasan, R., Bhutta, Z. A., et al. (2009). Pseudogene accumulation in the evolutionary histories of Salmonella enterica serovars Paratyphi A and Typhi. *BMC Genomics*, *10*(1), 36. http://doi.org/10.1186/1471-2164-10-36

Hormaeche, C. E. (1979). The natural resistance of radiation chimeras to S. typhimurium C5. *Immunology*, *37*(2), 329–332.

Hormaeche, C. E., Harrington, K. A., & Joysey, H. S. (1985). Natural resistance to salmonellae in mice: control by genes within the major histocompatibility complex. *The Journal of Infectious Diseases*, *152*(5), 1050–1056.

Hormaeche, C. E., Mastroeni, P., Arena, A., Uddin, J., & Joysey, H. S. (1990). T cells do not mediate the initial suppression of a Salmonella infection in the RES. *Immunology*, *70*(2), 247–250.

Howie, B., Fuchsberger, C., Stephens, M., Marchini, J., & Abecasis, G. R. (2012). Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nature Genetics*, *44*(8), 955–959. http://doi.org/10.1038/ng.2354

Hue, N. T., Lanh, M. N., Phuong, L. T., Vinh, H., Chinh, N. T., Hien, T. T., et al. (2009). Toll-Like Receptor 4 (TLR4) and Typhoid Fever in Vietnam. *PLoS ONE*, *4*(3), e4800. http://doi.org/10.1371/journal.pone.0004800

International HIV Controllers Study, Pereyra, F., Jia, X., McLaren, P. J., de Bakker, P. I. W., Walker, B. D., et al. (2010). The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. *Science*, *330*(6010), 1551–1557. http://doi.org/10.1126/science.1195271

International Multiple Sclerosis Genetics Consortium (IMSGC), Beecham, A. H., Patsopoulos, N. A., Xifara, D. K., Davis, M. F., Kemppinen, A., et al. (2013). Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. *Nature Genetics*, *45*(11), 1353–1360. http://doi.org/10.1038/ng.2770

Jabado, N., Jankowski, A., Dougaparsad, S., Picard, V., Grinstein, S., & Gros, P. (2000). Natural resistance to intracellular infections: natural resistance-associated macrophage protein 1 (Nramp1) functions as a pH-dependent manganese transporter at the phagosomal membrane. *The Journal of Experimental Medicine*, *192*(9), 1237–1248.

Jallow, M., Teo, Y. Y., Small, K. S., Rockett, K. A., Deloukas, P., Clark, T. G., et al. (2009). Genome-wide and fine-resolution association analysis of malaria in West Africa. *Nature Genetics*, *41*(6), 657–665. http://doi.org/10.1038/ng.388

Janssen, R., van Wengen, A., Hoeve, M. A., Dam, ten, M., van der Burg, M., van Dongen, J., et al. (2004). The same IkappaBalpha mutation in two related individuals leads to completely different clinical syndromes. *The Journal of Experimental Medicine*, *200*(5), 559–568. http://doi.org/10.1084/jem.20040773

Johnston, R. B., Newman, S. L., & Struth, A. G. (1973). An abnormality of the alternate pathway of complement activation in sickle-cell disease. *The New England Journal of Medicine*, *288*(16), 803–808. http://doi.org/10.1056/NEJM197304192881601

Joiner, K. A., Hammer, C. H., Brown, E. J., & Frank, M. M. (1982a). Studies on the mechanism of bacterial resistance to complement-mediated killing. II. C8 and C9 release C5b67 from the surface of Salmonella minnesota S218 because the terminal complex does not insert into the bacterial outer membrane. *The Journal of Experimental Medicine*, *155*(3), 809–819.

Joiner, K. A., Hammer, C. H., Brown, E. J., Cole, R. J., & Frank, M. M. (1982b). Studies on the mechanism of bacterial resistance to complement-mediated killing. I. Terminal complement components are deposited and released from Salmonella minnesota S218 without causing bacterial death. *The Journal of Experimental Medicine*, *155*(3), 797–808.

Jong, R. D. (1998). Severe Mycobacterial and Salmonella Infections in Interleukin-12 Receptor-Deficient Patients. *Science*, *280*(5368), 1435–1438. http://doi.org/10.1126/science.280.5368.1435

Jostins, L., Ripke, S., Weersma, R. K., Duerr, R. H., McGovern, D. P., Hui, K. Y., et al. (2012). Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature*, *491*(7422), 119–124. http://doi.org/10.1038/nature11582

Jouanguy, E., Dupuis, S., Pallier, A., Doffinger, R., Fondanèche, M. C., Fieschi, C., et al. (2000). In a novel form of IFN-gamma receptor 1 deficiency, cell surface receptors fail to bind IFN-gamma. *Journal of Clinical Investigation*, *105*(10), 1429–1436. http://doi.org/10.1172/JCI9166

Kalonji, L. M., Post, A., Phoba, M.-F., Falay, D., Ngbonda, D., Muyembe, J.-J., et al. (2015). Invasive Salmonella Infections at Multiple Surveillance Sites in the Democratic Republic of the Congo, 2011-2014. *Clinical Infectious Diseases : an Official Publication of the Infectious Diseases Society of America*, *61 Suppl 4*(suppl 4), S346–53. http://doi.org/10.1093/cid/civ713

Kariuki, S., Revathi, G., Gakuya, F., Yamo, V., Muyodi, J., & Hart, C. A. (2002). Lack of clonal relationship between non-typhi Salmonella strain types from humans and those isolated from animals living in close contact. *FEMS Immunology and Medical Microbiology*, *33*(3), 165–171.

Kariuki, S., Revathi, G., Kariuki, N., Kiiru, J., Mwituria, J., Muyodi, J., et al. (2006). Invasive multidrug-resistant non-typhoidal Salmonella infections in Africa: zoonotic or anthroponotic transmission? *Journal of Medical Microbiology*, *55*(Pt 5), 585–591. http://doi.org/10.1099/jmm.0.46375-0

Kawai, T., & Akira, S. (2011). Toll-like receptors and their crosstalk with other innate receptors in infection and immunity. *Immunity*, *34*(5), 637–650. http://doi.org/10.1016/j.immuni.2011.05.006

Kawai, T., Adachi, O., Ogawa, T., Takeda, K., & Akira, S. (1999). Unresponsiveness of MyD88-deficient mice to endotoxin. *Immunity*, *11*(1), 115–122.

Kayagaki, N., Wong, M. T., Stowe, I. B., Ramani, S. R., Gonzalez, L. C., Akashi-Takamura, S., et al. (2013). Noncanonical inflammasome activation by intracellular LPS independent of TLR4. *Science*, *341*(6151), 1246–1249. http://doi.org/10.1126/science.1240248

Keddy, K. H., Sooka, A., Musekiwa, A., Smith, A. M., Ismail, H., Tau, N. P., et al. (2015). Clinical and Microbiological Features of Salmonella Meningitis in a South African Population, 2003-2013. *Clinical Infectious Diseases : an Official Publication of the Infectious Diseases Society of America*, *61 Suppl 4*(suppl 4), S272–82. http://doi.org/10.1093/cid/civ685

Kenyan Bacteraemia Study Group, Wellcome Trust Case Control Consortium 2 (WTCCC2), Rautanen, A., Pirinen, M., Mills, T. C., Rockett, K. A., et al. (2016). Polymorphism in a lincRNA Associates with a Doubled Risk of Pneumococcal Bacteremia in Kenyan Children. *American Journal of Human Genetics*. http://doi.org/10.1016/j.ajhg.2016.03.025

Khan, R. T., Yuki, K. E., & Malo, D. (2014). Fine-mapping and phenotypic analysis of the ity3 salmonella susceptibility locus identify a complex genetic structure. *PLoS ONE*, *9*(2), e88009. http://doi.org/10.1371/journal.pone.0088009

Khor, C. C., Chapman, S. J., Vannberg, F. O., Dunne, A., Murphy, C., Ling, E. Y., et al. (2007). A Mal functional variant is associated with protection against invasive pneumococcal disease, bacteremia, malaria and tuberculosis. *Nature Genetics*, *39*(4), 523–528. http://doi.org/10.1038/ng1976

Khor, C. C., Vannberg, F. O., Chapman, S. J., Guo, H., Wong, S. H., Walley, A. J., et al. (2010). CISH and susceptibility to infectious diseases. *The New England Journal of Medicine*, *362*(22), 2092–2101. http://doi.org/10.1056/NEJMoa0905606

Kingsley, R. A., Msefula, C. L., Thomson, N. R., Kariuki, S., Holt, K. E., Gordon, M. A., et al. (2009). Epidemic multiple drug resistant Salmonella Typhimurium causing invasive disease in sub-Saharan Africa have a distinct genotype. *Genome Research*, *19*(12), 2279–2287. http://doi.org/10.1101/gr.091017.109

Kintz, E., Davies, M. R., Hammarlöf, D. L., Canals, R., Hinton, J. C. D., & van der Woude, M. W. (2015). A BTP1 prophage gene present in invasive non-typhoidal Salmonella determines composition and length of the O-antigen of the lipopolysaccharide. *Molecular Microbiology*, *96*(2), 263–275. http://doi.org/10.1111/mmi.12933

Klein, C., Lisowska-Grospierre, B., LeDeist, F., Fischer, A., & Griscelli, C. (1993). Major histocompatibility complex class II deficiency: Clinical manifestations, immunologic features, and outcome. *The Journal of Pediatrics*, *123*(6), 921–928. http://doi.org/10.1016/S0022-3476(05)80388-9

Klemm, E. J., Gkrania-Klotsas, E., Hadfield, J., Forbester, J. L., Harris, S. R., Hale, C., et al. (2016). Emergence of host-adapted Salmonella Enteritidis through rapid evolution in an immunocompromised host. *Nature Microbiology*, *1*(3), 15023. http://doi.org/10.1038/nmicrobiol.2015.23

Ko, D. C., Gamazon, E. R., Shukla, K. P., Pfuetzner, R. A., Whittington, D., Holden, T. D., et al. (2012). Functional genetic screen of human diversity reveals that a methionine salvage enzyme regulates inflammatory cell death. *Proceedings of the National Academy of Sciences*, *109*(35), E2343–E2352. http://doi.org/10.1073/pnas.1206701109

Kubori, T., Matsushima, Y., Nakamura, D., Uralil, J., Lara-Tejero, M., Sukhan, A., et al. (1998). Supramolecular structure of the Salmonella typhimurium type III protein secretion system. *Science*, *280*(5363), 602–605.

Kupz, A., Scott, T. A., Belz, G. T., Andrews, D. M., Greyer, M., Lew, A. M., et al. (2013). Contribution of Thy1+ NK cells to protective IFN-γ production during Salmonella typhimurium infections. *Proceedings of the National Academy of Sciences*, *110*(6), 2252–2257. http://doi.org/10.1073/pnas.1222047110

Lane, C. R., LeBaigue, S., Esan, O. B., Awofisyo, A. A., Adams, N. L., Fisher, I. S. T., et al. (2014). Salmonella enterica serovar Enteritidis, England and Wales, 1945-2011. *Emerging Infectious Diseases*, *20*(7), 1097–1104. http://doi.org/10.3201/eid2007.121850

Lee, P. H., O'Dushlaine, C., Thomas, B., & Purcell, S. M. (2012a). INRICH: interval-based enrichment analysis for genome-wide association studies. *Bioinformatics (Oxford, England)*, *28*(13), 1797–1799. http://doi.org/10.1093/bioinformatics/bts191

Lee, S.-J., Dunmire, S., & McSorley, S. J. (2012b). MHC class-I-restricted CD8 T cells play a protective role during primary Salmonella infection. *Immunology Letters*, *148*(2), 138–143. http://doi.org/10.1016/j.imlet.2012.10.009

Leekitcharoenphon, P., Friis, C., Zankari, E., Svendsen, C. A., Price, L. B., Rahmani, M., et al. (2013). Genomics of an emerging clone of Salmonella serovar Typhimurium ST313 from Nigeria and the Democratic Republic of Congo. *Journal of Infection in Developing Countries*, *7*(10), 696–706. http://doi.org/10.3855/jidc.3328

Lehmann, J., Bellmann, S., Werner, C., Schröder, R., Schütze, N., & Alber, G. (2001). IL-12p40-dependent agonistic effects on the development of protective innate and adaptive immunity against Salmonella enteritidis. *Journal of Immunology (Baltimore, Md. : 1950)*, *167*(9), 5304–5315.

Levy, J., Espanol-Boren, T., Thomas, C., & Fischer, A. (1997). Clinical spectrum of X-linked hyper-IgM syndrome. *The Journal of Pediatrics*, *131*(1), 47-54.

Liang-Takasaki, C. J., Saxén, H., Mäkelä, P. H., & Leive, L. (1983). Complement activation by polysaccharide of lipopolysaccharide: an important virulence determinant of salmonellae. *Infection and Immunity*, *41*(2), 563–569.

Liu, J. Z., Almarri, M. A., Gaffney, D. J., Mells, G. F., Jostins, L., Cordell, H. J., et al. (2012). Dense fine-mapping study identifies new susceptibility loci for primary biliary cirrhosis. *Nature Genetics*, *44*(10), 1137–1141. http://doi.org/10.1038/ng.2395

Löber, S., Jäckel, D., Kaiser, N., & Hensel, M. (2006). Regulation of Salmonella pathogenicity island 2 genes by independent environmental signals. *International Journal of Medical Microbiology : IJMM*, *296*(7), 435–447. http://doi.org/10.1016/j.ijmm.2006.05.001

Mackenzie, G., Ceesay, S. J., Hill, P. C., Walther, M., Bojang, K. A., Satoguina, J., et al. (2010). A decline in the incidence of invasive non-typhoidal Salmonella infection in The Gambia temporally associated with a decline in malaria infection. *PLoS ONE*, *5*(5), e10568. http://doi.org/10.1371/journal.pone.0010568

MacLennan, C. A., Gilchrist, J. J., Gordon, M. A., Cunningham, A. F., Cobbold, M., Goodall, M., et al. (2010). Dysregulated Humoral Immunity to Nontyphoidal Salmonella in HIV-Infected African Adults. *Science*, *328*(5977), 508–512. http://doi.org/10.1126/science.1180346

MacLennan, C. A., Gondwe, E. N., Msefula, C. L., Kingsley, R. A., Thomson, N. R., White, S. A., et al. (2008). The neglected role of antibody in protection against bacteremia caused by nontyphoidal strains of Salmonella in African children. *Journal of Clinical Investigation*, *118*(4), 1553–1562. http://doi.org/10.1172/JCI33998

MacLennan, C. A., Martin, L. B., & Micoli, F. (2014). Vaccines against invasive Salmonella disease: Current status and future directions. *Human Vaccines & Immunotherapeutics*, *10*(6), 1478–1493. http://doi.org/10.4161/hv.29054

MacLennan, C., Fieschi, C., Lammas, D. A., Picard, C., Dorman, S. E., Sanal, O., et al. (2004). Interleukin (IL)-12 and IL-23 are key cytokines for immunity against Salmonella in humans. *The Journal of Infectious Diseases*, *190*(10), 1755–1757. http://doi.org/10.1086/425021

MacMicking, J. D. (2012). Interferon-inducible effector mechanisms in cell-autonomous immunity. *Nature Reviews Immunology*, *12*(5), 367–382. http://doi.org/10.1038/nri3210

Majowicz, S. E., Musto, J., Scallan, E., Angulo, F. J., Kirk, M., O'Brien, S. J., et al. (2010). The global burden of nontyphoidal Salmonella gastroenteritis. *Clinical Infectious Diseases : an Official Publication of the Infectious Diseases Society of America*, *50*(6), 882–889. http://doi.org/10.1086/650733

Malaria Genomic Epidemiology Network. (2014). Reappraisal of known malaria resistance loci in a large multicenter study. *Nature Genetics*, *46*(11), 1197–1204. http://doi.org/10.1038/ng.3107

Mandala, W. L., MacLennan, J. M., Gondwe, E. N., Ward, S. A., Molyneux, M. E., & MacLennan, C. A. (2010). Lymphocyte subsets in healthy Malawians: implications for immunologic assessment of HIV infection in Africa. *The Journal of Allergy and Clinical Immunology*, *125*(1), 203–208. http://doi.org/10.1016/j.jaci.2009.10.010

Mandomando, I., Bassat, Q., Sigaúque, B., Massora, S., Quintó, L., Ácacio, S., et al. (2015). Invasive Salmonella Infections Among Children From Rural Mozambique, 2001-2014. *Clinical Infectious Diseases : an Official Publication of the Infectious Diseases Society of America*, *61 Suppl 4*(suppl 4), S339–45. http://doi.org/10.1093/cid/civ712

Marchini, J., & Howie, B. (2010). Genotype imputation for genome-wide association studies. *Nature Reviews Genetics*, *11*(7), 499–511. http://doi.org/10.1038/nrg2796

Marchini, J., Howie, B., Myers, S., McVean, G., & Donnelly, P. (2007). A new multipoint method for genome-wide association studies by imputation of genotypes. *Nature Genetics*, *39*(7), 906–913. http://doi.org/10.1038/ng2088

Mastroeni, P., Arena, A., Costa, G. B., Liberto, M. C., Bonina, L., & Hormaeche, C. E. (1991). Serum TNF alpha in mouse typhoid and enhancement of a Salmonella infection by anti-TNF alpha antibodies. *Microbial Pathogenesis*, *11*(1), 33–38.

Mastroeni, P., Grant, A., Restif, O., & Maskell, D. (2009). A dynamic view of the spread and intracellular distribution of Salmonella enterica. *Nature Reviews. Microbiology*, *7*(1), 73–80. http://doi.org/10.1038/nrmicro2034

Mastroeni, P., Harrison, J. A., Robinson, J. H., Clare, S., Khan, S., Maskell, D. J., et al. (1998). Interleukin-12 is required for control of the growth of attenuated aromatic-compound-dependent salmonellae in BALB/c mice: role of gamma interferon and macrophage activation. *Infection and Immunity*, *66*(10), 4767–4776.

Mastroeni, P., Simmons, C., Fowler, R., Hormaeche, C. E., & Dougan, G. (2000a). Igh-6(-/-) (B-cell-deficient) mice fail to mount solid acquired resistance to oral challenge with virulent Salmonella enterica serovar typhimurium and show impaired Th1 T-cell responses to Salmonella antigens. *Infection and Immunity*, *68*(1), 46–53.

Mastroeni, P., Vazquez-Torres, A., Fang, F. C., Xu, Y., Khan, S., Hormaeche, C. E., & Dougan, G. (2000b). Antimicrobial actions of the NADPH phagocyte oxidase and inducible nitric oxide synthase in experimental salmonellosis. II. Effects on microbial proliferation and host survival in vivo. *The Journal of Experimental Medicine*, *192*(2), 237–248.

Mastroeni, P., Villarreal-Ramos, B., & Hormaeche, C. E. (1992). Role of T cells, TNF alpha and IFN gamma in recall of immunity to oral challenge with virulent salmonellae in mice vaccinated with live attenuated aro- Salmonella vaccines. *Microbial Pathogenesis*, *13*(6), 477–491.

Mastroeni, P., Villarreal-Ramos, B., & Hormaeche, C. E. (1993). Adoptive transfer of immunity to oral challenge with virulent salmonellae in innately susceptible BALB/c mice requires both immune serum and T cells. *Infection and Immunity*, *61*(9), 3981–3984.

Maurano, M. T., Humbert, R., Rynes, E., Thurman, R. E., Haugen, E., Wang, H., et al. (2012). Systematic localization of common disease-associated variation in regulatory DNA. *Science*, *337*(6099), 1190–1195. http://doi.org/10.1126/science.1222794

Mazurkiewicz, P., Thomas, J., Thompson, J. A., Liu, M., Arbibe, L., Sansonetti, P., & Holden, D. W. (2008). SpvC is a Salmonella effector with phosphothreonine lyase activity on host mitogen-activated protein kinases. *Molecular Microbiology*, *67*(6), 1371–1383. http://doi.org/10.1111/j.1365-2958.2008.06134.x

McCarthy, M. I., Abecasis, G. R., Cardon, L. R., Goldstein, D. B., Little, J., Ioannidis, J. P. A., & Hirschhorn, J. N. (2008). Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nature Reviews Genetics*, *9*(5), 356–369. http://doi.org/10.1038/nrg2344

McGourty, K., Thurston, T. L., Matthews, S. A., Pinaud, L., Mota, L. J., & Holden, D. W. (2012). Salmonella inhibits retrograde trafficking of mannose-6-phosphate receptors and lysosome function. *Science*, *338*(6109), 963–967. http://doi.org/10.1126/science.1227037

Miller, S. I., Kukral, A. M., & Mekalanos, J. J. (1989). A two-component regulatory system (phoP phoQ) controls Salmonella typhimurium virulence. *Proceedings of the National Academy of Sciences of the United States of America*, *86*(13), 5054–5058.

Minegishi, Y., Saito, M., Morio, T., Watanabe, K., Agematsu, K., Tsuchiya, S., et al. (2006). Human Tyrosine Kinase 2 Deficiency Reveals Its Requisite Roles in Multiple Cytokine Signals Involved in Innate and Acquired Immunity. *Immunity*, *25*(5), 745–755. http://doi.org/10.1016/j.immuni.2006.09.009

Molyneux, E. M., Mankhambo, L. A., Phiri, A., Graham, S. M., Forsyth, H., Phiri, A., et al. (2009). The outcome of non-typhoidal salmonella meningitis in Malawian children, 1997-2006. *Annals of Tropical Paediatrics*, *29*(1), 13–22. http://doi.org/10.1179/146532809X401980

Molyneux, E. M., Walsh, A. L., Forsyth, H., Tembo, M., Mwenechanya, J., Kayira, K., et al. (2002). Dexamethasone treatment in childhood bacterial meningitis in Malawi: a randomised controlled trial. *Lancet*, *360*(9328), 211–218.

Moriwaki, K., Bertin, J., Gough, P. J., & Chan, F. K.-M. (2015). A RIPK3-caspase 8 complex mediates atypical pro-IL-1β processing. *The Journal of Immunology*, *194*(4), 1938–1944. http://doi.org/10.4049/jimmunol.1402167

Msefula, C. L., Kingsley, R. A., Gordon, M. A., Molyneux, E., Molyneux, M. E., MacLennan, C. A., et al. (2012). Genotypic homogeneity of multidrug resistant S. Typhimurium infecting distinct adult and childhood susceptibility groups in Blantyre, Malawi. *PLoS ONE*, *7*(7), e42085. http://doi.org/10.1371/journal.pone.0042085

Mtove, G., Amos, B., Nadjm, B., Hendriksen, I. C. E., Dondorp, A. M., Mwambuli, A., et al. (2011). Decreasing incidence of severe malaria and community-acquired bacteraemia among hospitalized children in Muheza, north-eastern Tanzania, 2006-2010. *Malaria Journal*, *10*(1), 320. http://doi.org/10.1186/1475-2875-10-320

Murray, G. L., Attridge, S. R., & Morona, R. (2003). Regulation of Salmonella typhimurium lipopolysaccharide O antigen chain length is required for virulence; identification of FepE as a second Wzz. *Molecular Microbiology*, *47*(5), 1395–1406.

Murray, G. L., Attridge, S. R., & Morona, R. (2005). Inducible serum resistance in Salmonella typhimurium is dependent on wzz(fepE)-regulated very long O antigen chains. *Microbes and Infection*, *7*(13), 1296–1304. http://doi.org/10.1016/j.micinf.2005.04.015

Muthumbi, E., Morpeth, S. C., Ooko, M., Mwanzu, A., Mwarumba, S., Mturi, N., et al. (2015). Invasive Salmonellosis in Kilifi, Kenya. *Clinical Infectious Diseases : an Official Publication of the Infectious Diseases Society of America*, *61 Suppl 4*(suppl 4), S290–301. http://doi.org/10.1093/cid/civ737

Nairz, M., Theurl, I., Schroll, A., Theurl, M., Fritsche, G., Lindner, E., et al. (2009). Absence of functional Hfe protects mice from invasive Salmonella enterica Serovar Typhimurium infection via induction of lipocalin-2. *Blood*, *114*(17), 3642–3651. http://doi.org/10.1182/blood-2009-05-223354

Nanton, M. R., Way, S. S., Shlomchik, M. J., & McSorley, S. J. (2012). Cutting edge: B cells are essential for protective immunity against Salmonella independent of antibody secretion. *The Journal of Immunology*, *189*(12), 5503–5507. http://doi.org/10.4049/jimmunol.1201413

Naranbhai, V., Fairfax, B. P., Makino, S., Humburg, P., Wong, D., Ng, E., et al. (2015). Genomic modulators of gene expression in human neutrophils. *Nature Communications*, *6*, 7545. http://doi.org/10.1038/ncomms8545

Newport, M. J., Huxley, C. M., Huston, S., Hawrylowicz, C. M., Oostra, B. A., Williamson, R., & Levin, M. (1996). A mutation in the interferon-gamma-receptor gene and susceptibility to mycobacterial infection. *The New England Journal of Medicine*, *335*(26), 1941–1949. http://doi.org/10.1056/NEJM199612263352602

North, B. V., Curtis, D., & Sham, P. C. (2002). A note on the calculation of empirical P values from Monte Carlo procedures. *The American Journal of Human Genetics*, *71*(2), 439–441. http://doi.org/10.1086/341527

North, B. V., Curtis, D., & Sham, P. C. (2003). A note on calculation of empirical P values from Monte Carlo procedure. *The American Journal of Human Genetics*, *72*(2), 498–499.

Nyakoe, N. K., Taylor, R. P., Makumi, J. N., & Waitumbi, J. N. (2009). Complement consumption in children with Plasmodium falciparum malaria. *Malaria Journal*, *8*(1), 7. http://doi.org/10.1186/1475-2875-8-7

Ohmori, Y., Schreiber, R. D., & Hamilton, T. A. (1997). Synergy between interferon-gamma and tumor necrosis factor-alpha in transcriptional activation is mediated by cooperation between signal transducer and activator of transcription 1 and nuclear factor kappaB. *The Journal of Biological Chemistry*, *272*(23), 14899–14907.

Okoro, C. K., Barquist, L., Connor, T. R., Harris, S. R., Clare, S., Stevens, M. P., et al. (2015). Signatures of adaptation in human invasive Salmonella Typhimurium ST313 populations from sub-Saharan Africa. *PLoS Neglected Tropical Diseases*, *9*(3), e0003611. http://doi.org/10.1371/journal.pntd.0003611

Okoro, C. K., Kingsley, R. A., Connor, T. R., Harris, S. R., Parry, C. M., Al-Mashhadani, M. N., et al. (2012a). Intracontinental spread of human invasive Salmonella Typhimurium pathovariants in sub-Saharan Africa. *Nature Genetics*, *44*(11), 1215–1221. http://doi.org/10.1038/ng.2423

Okoro, C. K., Kingsley, R. A., Quail, M. A., Kankwatira, A. M., Feasey, N. A., Parkhill, J., et al. (2012b). High-resolution single nucleotide polymorphism analysis distinguishes recrudescence and reinfection in recurrent invasive nontyphoidal Salmonella typhimurium disease. *Clinical Infectious Diseases : an Official Publication of the Infectious Diseases Society of America*, *54*(7), 955–963. http://doi.org/10.1093/cid/cir1032

Old, I. G., Phillips, S. E., Stockley, P. G., & Saint Girons, I. (1991). Regulation of methionine biosynthesis in the Enterobacteriaceae. *Progress in Biophysics and Molecular Biology*, *56*(3), 145–185.

Oneko, M., Kariuki, S., Muturi-Kioi, V., Otieno, K., Otieno, V. O., Williamson, J. M., et al. (2015). Emergence of Community-Acquired, Multidrug-Resistant Invasive Nontyphoidal Salmonella Disease in Rural Western Kenya, 2009-2013. *Clinical Infectious Diseases : an Official Publication of the Infectious Diseases Society of America*, *61 Suppl 4*(suppl 4), S310–6. http://doi.org/10.1093/cid/civ674

Onengut-Gumuscu, S., Chen, W.-M., Burren, O., Cooper, N. J., Quinlan, A. R., Mychaleckyj, J. C., et al. (2015). Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. *Nature Genetics*, *47*(4), 381–386. http://doi.org/10.1038/ng.3245

Paglietti, B., Falchi, G., Mason, P., Chitsatso, O., Nair, S., Gwanzura, L., et al. (2013). Diversity among human non-typhoidal salmonellae isolates from Zimbabwe. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *107*(8), 487–492. http://doi.org/10.1093/trstmh/trt046

Parkhill, J., Wren, B. W., Thomson, N. R., Titball, R. W., Holden, M. T., Prentice, M. B., et al. (2001). Genome sequence of Yersinia pestis, the causative agent of plague. *Nature*, *413*(6855), 523–527. http://doi.org/10.1038/35097083

Parry, C. M., Hien, T. T., Dougan, G., White, N. J., & Farrar, J. J. (2002). Typhoid fever. *The New England Journal of Medicine*, *347*(22), 1770–1782. http://doi.org/10.1056/NEJMra020201

Parsons, B. N., Humphrey, S., Salisbury, A. M., Mikoleit, J., Hinton, J. C. D., Gordon, M. A., & Wigley, P. (2013). Invasive non-typhoidal Salmonella typhimurium ST313 are not host-restricted and have an invasive phenotype in experimentally infected chickens. *PLoS Neglected Tropical Diseases*, *7*(10), e2487. http://doi.org/10.1371/journal.pntd.0002487

Pearson, H. A., Spencer, R. P., & Cornelius, E. A. (1969). Functional asplenia in sickle-cell anemia. *The New England Journal of Medicine*, *281*(17), 923–926. http://doi.org/10.1056/NEJM196910232811703

Peltonen, L., & McKusick, V. A. (2001). Genomics and medicine. Dissecting human disease in the postgenomic era. *Science*, *291*(5507), 1224–1229.

Peña-Sagredo, J. L., Fariñas, M. C., Perez-Zafrilla, B., Cruz-Valenciano, A., Crespo, M., Joven-Ibañez, B., et al. (2009). Non-typhi Salmonella infection in patients with rheumatic diseases on TNF-alpha antagonist therapy. *Clinical and Experimental Rheumatology*, *27*(6), 920–925.

Pfeiffer, R. M., Gail, M. H., & Pee, D. (2009). On Combining Data From Genome-Wide Association Studies to Discover Disease-Associated SNPs. *Statistical Science*, *24*(4), 547–560. http://doi.org/10.1214/09-STS286

Picard, C., Casanova, J. L., & Puel, A. (2011). Infectious Diseases in Patients with IRAK-4, MyD88, NEMO, or I B Deficiency. *Clinical Microbiology Reviews*, *24*(3), 490–497. http://doi.org/10.1128/CMR.00001-11

Poltorak, A., He, X., Smirnova, I., Liu, M. Y., Van Huffel, C., Du, X., et al. (1998). Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: mutations in Tlr4 gene. *Science*, *282*(5396), 2085–2088. http://doi.org/10.1126/science.282.5396.2085

Prando, C., Samarina, A., Bustamante, J., Boisson-Dupuis, S., Cobat, A., Picard, C., et al. (2013). Inherited IL-12p40 Deficiency. *Medicine*, *92*(2), 109–122. http://doi.org/10.1097/MD.0b013e31828a01f9

Prendergast, A. J. (2015). Malnutrition and vaccination in developing countries. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1671), 20140141–20140141. http://doi.org/10.1098/rstb.2014.0141

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, *38*(8), 904–909. http://doi.org/10.1038/ng1847

Prost, L. R., Daley, M. E., Le Sage, V., Bader, M. W., Le Moual, H., Klevit, R. E., & Miller, S. I. (2007). Activation of the bacterial sensor kinase PhoQ by acidic pH. *Molecular Cell*, *26*(2), 165–174. http://doi.org/10.1016/j.molcel.2007.03.008

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, *81*(3), 559–575. http://doi.org/10.1086/519795

Raffatellu, M., Santos, R. L., Verhoeven, D. E., George, M. D., Wilson, R. P., Winter, S. E., et al. (2008). Simian immunodeficiency virus-induced mucosal interleukin-17 deficiency promotes Salmonella dissemination from the gut. *Nature Medicine*, *14*(4), 421–428. http://doi.org/10.1038/nm1743

Raj, T., Rothamel, K., Mostafavi, S., Ye, C., Lee, M. N., Replogle, J. M., et al. (2014). Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes. *Science*, *344*(6183), 519–523. http://doi.org/10.1126/science.1249547

Ramachandran, G., Aheto, K., Shirtliff, M. E., & Tennant, S. M. (2016). Poor biofilm forming ability and long-term survival of invasive Salmonella Typhimurium ST313. *Pathogens and Disease*, *74*(5), ftw049. http://doi.org/10.1093/femspd/ftw049

Raupach, B., Peuschel, S. K., Monack, D. M., & Zychlinsky, A. (2006). Caspase-1-Mediated Activation of Interleukin-1 (IL-1) and IL-18 Contributes to Innate Immune Defenses against Salmonella enterica Serovar Typhimurium Infection. *Infection and Immunity*, *74*(8), 4922–4926. http://doi.org/10.1128/IAI.00417-06

Reddy, E. A., Shaw, A. V., & Crump, J. A. (2010). Community-acquired bloodstream infections in Africa:a systematic review and meta-analysis. *The Lancet Infectious Diseases*, *10*(6), 417–432. http://doi.org/10.1016/S1473-3099(10)70072-4

Rikimaru, T., Taniguchi, K., Yartey, J. E., Kennedy, D. O., & Nkrumah, F. K. (1998). Humoral and cell-mediated immunity in malnourished children in Ghana. *European Journal of Clinical Nutrition*, *52*(5), 344–350.

Risch, N. (1990). Linkage strategies for genetically complex traits. I. Multilocus models. *The American Journal of Human Genetics*, *46*(2), 222–228.

Robson, H. G., & Vas, S. I. (1972). Resistance of inbred mice to Salmonella typhimurium. *The Journal of Infectious Diseases*, *126*(4), 378–386.

Ross, S. C., & Densen, P. (1984). Complement deficiency states and infection: epidemiology, pathogenesis and consequences of neisserial and other infections in an immune deficiency. *Medicine*, *63*(5), 243–273.

Roy, M.-F., Riendeau, N., Bédard, C., Hélie, P., Min-Oo, G., Turcotte, K., et al. (2007). Pyruvate kinase deficiency confers susceptibility to Salmonella typhimurium infection in mice. *Journal of Experimental Medicine*, *204*(12), 2949–2961. http://doi.org/10.1084/jem.20062606

Rytter, M. J. H., Kolte, L., Briend, A., Friis, H., & Christensen, V. B. (2014). The immune system in children with malnutrition--a systematic review. *PLoS ONE*, *9*(8), e105017. http://doi.org/10.1371/journal.pone.0105017

Santos, R. L., Zhang, S., Tsolis, R. M., Kingsley, R. A., Adams, L. G., & Bäumler, A. J. (2001). Animal models of Salmonella infections: enteritis versus typhoid fever. *Microbes and Infection*, *3*(14-15), 1335–1344.

Schopfer, K., & Douglas, S. D. (1976). Neutrophil function in children with kwashiorkor. *The Journal of Laboratory and Clinical Medicine*, *88*(3), 450–461.

Schwarzer, E., Turrini, F., Ulliers, D., Giribaldi, G., Ginsburg, H., & Arese, P. (1992). Impairment of macrophage functions after ingestion of Plasmodium falciparum-infected erythrocytes or isolated malarial pigment. *The Journal of Experimental Medicine*, *176*(4), 1033–1041.

Scott, J. A. G., Berkley, J. A., Mwangi, I., Ochola, L., Uyoga, S., Macharia, A., et al. (2011). Relation between falciparum malaria and bacteraemia in Kenyan children: a population-based, case-control study and a longitudinal study. *Lancet*, *378*(9799), 1316–1323. http://doi.org/10.1016/S0140-6736(11)60888-X

Segal, B. H., Leto, T. L., Gallin, J. I., Malech, H. L., & Holland, S. M. (2000). Genetic, biochemical, and clinical features of chronic granulomatous disease. *Medicine*, *79*(3), 170–200.

Sham, P. C., & Purcell, S. M. (2014). Statistical power and significance testing in large-scale genetic studies. *Nature Reviews Genetics*, *15*(5), 335–346. http://doi.org/10.1038/nrg3706

Simonson, M. A., McQueen, M. B., & Keller, M. C. (2014). Whole-genome pathway analysis on 132,497 individuals identifies novel gene-sets associated with body mass index. *PLoS ONE*, *9*(1), e78546. http://doi.org/10.1371/journal.pone.0078546

Singletary, L. A., Karlinsey, J. E., Libby, S. J., Mooney, J. P., Lokken, K. L., Tsolis, R. M., et al. (2016). Loss of Multicellular Behavior in Epidemic African Nontyphoidal Salmonella enterica Serovar Typhimurium ST313 Strain D23580. *mBio*, *7*(2). http://doi.org/10.1128/mBio.02265-15

Sinha, K., Mastroeni, P., Harrison, J., de Hormaeche, R. D., & Hormaeche, C. E. (1997). Salmonella typhimurium aroA, htrA, and aroD htrA mutants cause progressive infections in athymic (nu/nu) BALB/c mice. *Infection and Immunity*, *65*(4), 1566–1569.

Sivick, K. E., Arpaia, N., Reiner, G. L., Lee, B. L., Russell, B. R., & Barton, G. M. (2014). Toll-like receptor-deficient mice reveal how innate immune signaling influences Salmonella virulence strategies. *Cell Host and Microbe*, *15*(2), 203–213. http://doi.org/10.1016/j.chom.2014.01.013

Smith, A. C., Heo, W. D., Braun, V., Jiang, X., Macrae, C., Casanova, J. E., et al. (2007). A network of Rab GTPases controls phagosome maturation and is modulated by Salmonella enterica serovar Typhimurium. *The Journal of Cell Biology*, *176*(3), 263–268. http://doi.org/10.1083/jcb.200611056

Stephens, M., & Balding, D. J. (2009). Bayesian statistical methods for genetic association studies. *Nature Reviews Genetics*, *10*(10), 681–690. http://doi.org/10.1038/nrg2615

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(43), 15545–15550. http://doi.org/10.1073/pnas.0506580102

Sørensen, T. I., Nielsen, G. G., Andersen, P. K., & Teasdale, T. W. (1988). Genetic and environmental influences on premature death in adult adoptees. *The New England Journal of Medicine*, *318*(12), 727–732. http://doi.org/10.1056/NEJM198803243181202

Tabu, C., Breiman, R. F., Ochieng, B., Aura, B., Cosmas, L., Audi, A., et al. (2012). Differing burden and epidemiology of non-Typhi Salmonella bacteremia in rural and urban Kenya, 2006-2009. *PLoS ONE*, *7*(2), e31237. http://doi.org/10.1371/journal.pone.0031237

Teo, Y. Y., Small, K. S., & Kwiatkowski, D. P. (2010). Methodological challenges of genome-wide association analysis in Africa. *Nature Reviews Genetics*, *11*(2), 149–160. http://doi.org/10.1038/nrg2731

Thomson, N. R., Clayton, D. J., Windhorst, D., Vernikos, G., Davidson, S., Churcher, C., et al. (2008). Comparative genome analysis of Salmonella Enteritidis PT4 and Salmonella Gallinarum 287/91 provides insights into evolutionary and host adaptation pathways. *Genome Research*, *18*(10), 1624–1637. http://doi.org/10.1101/gr.077404.108

Timmann, C., Thye, T., Vens, M., Evans, J., May, J., Ehmen, C., et al. (2013). Genome-wide association study indicates two novelresistance loci for severe malaria. *Nature*, *489*(7416), 443–446. http://doi.org/10.1038/nature11334

Trent, M. S., Pabich, W., Raetz, C. R., & Miller, S. I. (2001). A PhoP/PhoQ-induced Lipase (PagL) that catalyzes 3-O-deacylation of lipid A precursors in membranes of Salmonella typhimurium. *The Journal of Biological Chemistry*, *276*(12), 9083–9092. http://doi.org/10.1074/jbc.M010730200

Trynka, G., Hunt, K. A., Bockett, N. A., Romanos, J., Mistry, V., Szperl, A., et al. (2011). Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. *Nature Genetics*, *43*(12), 1193–1201. http://doi.org/10.1038/ng.998

Tsoi, L. C., Spain, S. L., Knight, J., Ellinghaus, E., Stuart, P. E., Capon, F., et al. (2012). Identification of 15 new psoriasis susceptibility loci highlights the role of innate immunity. *Nature Genetics*, *44*(12), 1341–1348. http://doi.org/10.1038/ng.2467

Uchiya, K., Barbieri, M. A., Funato, K., Shah, A. H., Stahl, P. D., & Groisman, E. A. (1999). A Salmonella virulence protein that inhibits cellular trafficking. *The EMBO Journal*, *18*(14), 3924–3933. http://doi.org/10.1093/emboj/18.14.3924

Uyoga, S., Ndila, C. M., Macharia, A. W., Nyutu, G., Shah, S., Peshu, N., et al. (2015). Glucose-6-phosphate dehydrogenase deficiency and the risk of malaria and other diseases in children in Kenya: a case-control and a cohort study. *The Lancet. Haematology*, *2*(10), e437–44. http://doi.org/10.1016/S2352-3026(15)00152-0

Vazquez-Torres, A., Jones-Carson, J., Bäumler, A. J., Falkow, S., Valdivia, R., Brown, W., et al. (1999). Extraintestinal dissemination of Salmonella by CD18-expressing phagocytes. *Nature*, *401*(6755), 804–808. http://doi.org/10.1038/44593

Vazquez-Torres, A., Xu, Y., Jones-Carson, J., Holden, D. W., Lucia, S. M., Dinauer, M. C., et al. (2000). Salmonella pathogenicity island 2-dependent evasion of the phagocyte NADPH oxidase. *Science*, *287*(5458), 1655–1658.

Verani, J. R., Toroitich, S., Auko, J., Kiplang'at, S., Cosmas, L., Audi, A., et al. (2015). Burden of Invasive Nontyphoidal Salmonella Disease in a Rural and Urban Site in Kenya, 2009-2014. *Clinical Infectious Diseases : an Official Publication of the Infectious Diseases Society of America*, *61 Suppl 4*(suppl 4), S302–9. http://doi.org/10.1093/cid/civ728

Vidal, S., Tremblay, M. L., Govoni, G., Gauthier, S., Sebastiani, G., Malo, D., et al. (1995). The Ity/Lsh/Bcg locus: natural resistance to infection with intracellular parasites is abrogated by disruption of the Nramp1 gene. *The Journal of Experimental Medicine*, *182*(3), 655–666.

Wacholder, S., Chanock, S., Garcia-Closas, M., Ghormli, El, L., & Rothman, N. (2004). Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. *Journal of the National Cancer Institute*, *96*(6), 434–442.

Wakefield, J. (2009). Bayes factors for genome-wide association studies: comparison with P-values. *Genetic Epidemiology*, *33*(1), 79–86. http://doi.org/10.1002/gepi.20359

Watford, W. T., Hissong, B. D., Bream, J. H., Kanno, Y., Muul, L., & O'Shea, J. J. (2004). Signaling by IL-12 and IL-23 and the immunoregulatory roles of STAT4. *Immunological Reviews*, *202*(1), 139–156. http://doi.org/10.1111/j.0105-2896.2004.00211.x

Weir, B. S., Anderson, A. D., & Hepler, A. B. (2006). Genetic relatedness analysis: modern data and new challenges. *Nature Reviews Genetics*, *7*(10), 771–780. http://doi.org/10.1038/nrg1960

Weiss, D. S., Raupach, B., Takeda, K., Akira, S., & Zychlinsky, A. (2004). Toll-like receptors are temporally involved in host defense. *Journal of Immunology (Baltimore, Md. : 1950)*, *172*(7), 4463–4469.

Williams, T. N., Uyoga, S., Macharia, A., Ndila, C., McAuley, C. F., Opi, D. H., et al. (2009). Bacteraemia in Kenyan children with sickle-cell anaemia: a retrospective cohort and case-control study. *Lancet*, *374*(9698), 1364–1370. http://doi.org/10.1016/S0140-6736(09)61374-X

Winkelstein, J. A., Marino, M. C., Johnston, R. B., Boyle, J., Curnutte, J., Gallin, J. I., et al. (2000). Chronic granulomatous disease. Report on a national registry of 368 patients. *Medicine*, *79*(3), 155–169.

Winkelstein, J. A., Marino, M. C., Lederman, H. M., Jones, S. M., Sullivan, K., Burks, A. W., et al. (2006). X-linked agammaglobulinemia: report on a United States registry of 201 patients. *Medicine*, *85*(4), 193–202. http://doi.org/10.1097/01.md.0000229482.27398.ad

Wu, C., Ferrante, J., Gately, M. K., & Magram, J. (1997). Characterization of IL-12 receptor beta1 chain (IL-12Rbeta1)-deficient mice: IL-12Rbeta1 is an essential component of the functional mouse IL-12 receptor. *Journal of Immunology (Baltimore, Md. : 1950)*, *159*(4), 1658–1665.

Yang, J., Barrila, J., Roland, K. L., Kilbourne, J., Ott, C. M., Forsyth, R. J., & Nickerson, C. A. (2015). Characterization of the Invasive, Multidrug Resistant Non-typhoidal Salmonella Strain D23580 in a Murine Model of Infection. *PLoS Neglected Tropical Diseases*, *9*(6), e0003839. http://doi.org/10.1371/journal.pntd.0003839

Yang, L., Neale, B. M., Liu, L., Lee, S. H., Wray, N. R., Ji, N., et al. (2013). Polygenic transmission and complex neuro developmental network for attention deficit hyperactivity disorder: genome-wide association study of both common and rare variants. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics : the Official Publication of the International Society of Psychiatric Genetics*, *162B*(5), 419–430. http://doi.org/10.1002/ajmg.b.32169

Yoshimura, S.-I., Egerer, J., Fuchs, E., Haas, A. K., & Barr, F. A. (2007). Functional dissection of Rab GTPases involved in primary cilium formation. *The Journal of Cell Biology*, *178*(3), 363–369. http://doi.org/10.1083/jcb.200703047

Yuki, K. E., Eva, M. M., Richer, E., Chung, D., Paquet, M., Cellier, M., et al. (2013). Suppression of hepcidin expression and iron overload mediate Salmonella susceptibility in ankyrin 1 ENU-induced mutant. *PLoS ONE*, *8*(2), e55331. http://doi.org/10.1371/journal.pone.0055331

Zhang, F.-R., Huang, W., Chen, S.-M., Sun, L.-D., Liu, H., Li, Y., et al. (2009). Genomewide Association Study of Leprosy. *The New England Journal of Medicine*, *361*(27), 2609–2618. http://doi.org/10.1056/NEJMoa0903753

Zhou, D., Mooseker, M. S., & Galán, J. E. (1999). Role of the S. typhimurium actin-binding protein SipA in bacterial internalization. *Science*, *283*(5410), 2092–2095.