

Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Khor CC, Vannberg FO, Chapman SJ, et al. *CISH* and susceptibility to infectious diseases. *N Engl J Med* 2010;362:2092-101. DOI: 10.1056/NEJMoa0905606.

Supplementary information

Patient collections

Kenyan bacteraemia study (KB)

Kenyan children (<13 years old) with invasive bacteraemia were recruited from consecutive acute admissions to the Kilifi District hospital between 1998 and 2002. This genetic study was part of a larger clinical study assessing the prevalence of bacteraemia in Kilifi, bacteraemia and protein-energy malnutrition, bacteraemia and HIV, as well as bacteraemia and antibiotic resistance. The bacteraemia cases comprised patients with isolated Gram-positive or Gram-negative infections, diagnosed using standard blood culture techniques. The most frequent organisms isolated were *Streptococcus pneumoniae* (25.5%), Non-typhoidal *Salmonellae* (16.6%), *Haemophilus influenzae* (14.6%) and *Escherichia coli* (8.7%). The community controls were matched to the cases on the basis of time (recruited within 14 days), location of homestead, age, and gender. Only children with complete data for HIV, malnutrition, and malaria status were included for analysis. This study comprised 770 bacteraemic cases and 560 controls. Ethical approval was granted by the Kenya Medical Research Institute (KEMRI) National Scientific Steering and Research Committees.

Malawian tuberculosis study (MTB)

For the Malawian tuberculosis (TB) study, TB cases recruited into the study were ascertained by screening individuals attending health facilities. Sputum samples and lymph node aspirates were examined (fluorescence and Ziehl-Neelsen stain) and cultured at project headquarters in Chilumba, with species confirmation by the United Kingdom Public Health Laboratory Service Mycobacterial Reference Unit (Dulwich, United Kingdom). Inclusion in this study required confirmation of TB by culture, smear, or histology (excluding those whose only evidence of TB was a single scanty smear with fewer than 10 bacilli). Recruitment for this study began in 1996 and includes only HIV-negative TB cases. Randomly selected community controls were frequency matched by age, gender and district zone of homestead and had no history or evidence of TB. Analyses reported here include

cases and controls recruited up to the end of September 2001. This study comprised of 335 tuberculosis cases and 450 controls. The work was approved by the Malawi National Health Sciences Research Committee and the Ethics Committee of the London School of Hygiene and Tropical Medicine.

Hong Kong Tuberculosis study (HKTb)

The genetic study of predisposition to TB in Hong Kong has been approved by clinical research ethical committees of The Chinese University of Hong Kong and Department of Health in 2002 and comprised 907 patients with TB and 784 controls. Patients attending the territory-wide Chest Clinic of the Tuberculosis and Chest Service in Hong Kong with confirmed TB were invited to participate in the study. The diagnosis of TB was made by: (1) smear positive for *Mycobacterium tuberculosis* (M. TB), and/or (2) culture positive for M. TB and/or (3) TB diagnosed on clinical, radiological and / or histological grounds together with an appropriate response to anti-TB treatment. Patients with other predisposing conditions were excluded, such as (1) patients with HIV, (2) patients with uncontrolled diabetes or other causes of immunodeficiency. The controls comprised of 784 anonymous archived sample of umbilical cord blood which had been collected after delivery by a mother of Chinese ethnicity.

Gambian Tuberculosis study (GTb)

Samples for the Gambian TB study comprised of 1309 pulmonary TB cases and 1427 controls. Suspected cases were identified by field workers on the basis of clinical criteria including cough (or more than 3 weeks duration), fever, weight loss, night sweats, chest pain, haemoptysis and PPD positive skin response and referred for clinical examination. Cases confirmed by sputum positivity on smear or culture for *Mycobacterium tuberculosis* were included in the study. TB cases that were smear positive but culture negative all had radiographic evidence of TB. The controls were recruited from routine births at local health clinics. The majority (>95%) of the TB cohort was screened for

HIV-1, with positive cases excluded from the study. This study has been reviewed and approved by the Gambian Government / MRC Joint Ethical Committee.

Gambian malaria Study (GM)

For the Gambian malaria study, children aged 1 to 10 years were enrolled between August 1988 and November 1990 at the Royal Victoria Hospital of Banjul, and the Medical Research Council hospital of Fajara, The Gambia. Malaria was diagnosed if parasitaemia of $>2500/\mu\text{l}$ of blood was detected. Children were classified as severe malaria (N = 485) if they suffered from either cerebral malaria or severe malarial anaemia. Children were scored to have mild malaria if they had an uncomplicated febrile illness with *P. falciparum* parasitaemia and no other apparent cause of fever. The controls (N = 210) were recruited at both hospitals and health centres in the study area, and were frequency matched to the cases for age and area of residence. These children had mild, mostly infectious diseases (e.g. mild cough and cold) and illness that did not require hospital admission and did not have parasitaemia on blood films. These children were not matched specifically for ethnic groups. Instead, correction for the potential confounding effect of ethnicity was performed within a logistic regression framework. Ethical approval was granted by the Gambian Government / MRC Joint Ethical Committee.

Kenyan malaria Study (KM)

For the Kenyan malaria study, children admitted to the high dependency unit of the Kilifi District Hospital with severe malaria were recruited as cases between 1992 and 1995 (N = 685). Malaria was diagnosed if parasitaemia of $>2500/\mu\text{l}$ of blood was detected. Children were classified as severe malaria if they suffered from cerebral malaria, severe malarial anaemia, with concurrent acidosis or respiratory distress. The control group was the same as that used in the bacteraemia study (N = 560). Ethical approval was given by the Kenya Medical Research Institute (KEMRI) National Scientific Steering and Research Committees.

Vietnamese malaria study (VM)

Cases (N = 375) were adults suffering from severe malaria at the Centre for Tropical Diseases, Ho Chi Minh City, Vietnam recruited between 1991 and 1996. Malaria was diagnosed if parasitaemia of $>2500/\mu\text{l}$ of blood was detected. The adults were classified as severe malaria if they suffered from either cerebral malaria or severe malarial anaemia. Controls (N = 105) were ethnically matched adult non-malaria cases recruited from other departments of the same hospital. Subjects with mild malaria were not included in this study. Ethical approval was granted by the Ethical and Scientific committee of the Centre for Tropical Diseases in Ho Chi Minh City, Vietnam.

For all patient collections, ethnicity was determined by self-reporting.

CISH was one of twenty candidate genes previously assessed in detail using these sample collections.

Laboratory methods

Sample genotyping

In order to identify novel polymorphisms in *CISH*, direct sequencing of the gene and its flanking 5' and 3' regions was performed with the BigDye terminator mix v 3.1 (Applied Biosystems) in the first study group, comprising 48 Kenyan children (24 invasive bacterial disease cases and 24 controls). This provides 99% power to detect polymorphisms with a minor allele frequency (MAF) of $>5\%$, and 80% power for polymorphisms with $\text{MAF} >1\%$. All detected *CISH* polymorphisms with a minor allele frequency of $>5\%$ were then genotyped in the bacteremia, malaria and tuberculosis patient cohorts using the Sequenom MALDI-TOF primer extension assay, with the exception of *CISH* -163 which was genotyped using direct sequencing. In addition, the genotypes of *CISH* -292 and -639 were further confirmed via direct sequencing, and -292 genotypes were confirmed by a third assay, an independent restriction-fragment-length polymorphism (RFLP) assay (with *Bsr*BI, described in supplementary information). The call rates demonstrated complete concordance between the different assays. All control

genotype frequencies were not significantly deviated from Hardy-Weinberg equilibrium ($P>0.1$).

Healthy volunteers for cell stimulation studies

Ethical approval was obtained to enroll 400 healthy volunteers of Chinese descent in Singapore. We obtained written informed consent from each subject as approved by the National Institutional Review Boards (IRB) at the National University Hospital, and the National Healthcare Group in Singapore.

Genomic DNA purification from volunteers

After buccal swab samples (MicroRheologics) had been obtained, genomic DNA was extracted from them using the QIAamp Blood mini kit (Qiagen), according to the manufacturer's protocol. The DNA samples were genotyped to determine their SNPs at *CIS1* -639, -292 and -163. Five volunteers from each of the following genotype groups were then randomly recalled to obtain a blood sample for peripheral blood mononuclear cell (PBMC) isolation: *CIS1* -292 A/A, *CIS1* -292 A/T, *CIS1* -292 T/T, *CIS1* -163 C/C, and *CIS1* -163 C/G (total n=25). As there was only one volunteer with the variant *CIS1* -163 G/G, this genotype was not studied.

Isolation and stimulation of cells from volunteers

A sample of peripheral blood was taken from each volunteer (n=25). Human PBMCs were derived using the Ficoll gradient density centrifugation method. The blood, which was collected in sodium heparin vacutainers (BD Biosciences), was diluted with an equal volume of phosphate buffered saline (PBS) at room temperature. The diluted blood was layered on top of Ficoll-Plaque plus (GE healthcare) and centrifuged at room temperature at 400g for 30mins. The cells were removed from the plasma-Ficoll interface, washed twice with PBS and centrifuged at room temperature at 250g for 10 minutes. The cells were then re-suspended in RPMI medium +10% fetal bovine serum (FBS) at a

concentration of 2×10^6 cells/ml. The cells were then aliquoted into 5 mL tubes (500 μ L of cells per tube) and incubated in 5% CO₂ at 37°C for 2 hours. A baseline sample was then harvested (time point = 0 hours) while the rest of the samples were either stimulated with IL2 or IL3 (final concentration of 100 U/ml of IL2 or IL3 (BD Pharmingen) with specific activities of 0.5×10^8 U/mg and 0.1×10^8 U/mg, respectively). At 0.5h, 1h and 2h after addition, the relevant cells were harvested. At the time of harvesting, all cells were immediately pelleted at 300g for 4mins and then snap-frozen in dry ice/ethanol bath for 2 minutes. Cell pellets were stored at -80°C for future RNA extraction. These experiments were performed in duplicate and by a person (H.G) who was blinded to the genotypes.

RNA extraction and cDNA synthesis

Total RNA was extracted using RNeasy Mini kit (Qiagen) according to the manufacturer's protocol. 100ng of total RNA was used for cDNA synthesis using high capacity RNA-to-cDNA master mix kit (Applied Biosystems) according to manufacturer's protocol. The synthesized cDNA was used for real time PCR.

Real time PCR

SYBR green real time PCR was performed in an ABI7300 Sequence Detection System (Perkin Elmer, Applied Biosystems) with 1 μ L of the first-strand cDNA synthesis reaction, SYBR green PCR master mix (PE, Applied Biosystems), and 400nM of the appropriate primers. The primers for the target gene (human *CIS1*) were: forward 5'-CTG TGC ATA GCC AAG ACC TTC TC-3' and reverse 5'-CTG GCA TCT TCT GCA GGT GTT -3'. The primers for the endogenous control gene (human *beta-2 microglobulin*) were: forward 5'-TTG CCG ACA GGA TGC AGA AGG A -3' and reverse 5'-AGG TGG ACA GCG AGG CCA GGA T-3'. The cycling conditions were 50°C for 2 mins, 94°C for 10 mins, followed by 40 cycles 94°C for 15s and 60°C for 60s.

Statistical analysis

All *P* values reported are two-tailed. When an allelic association was assessed in multiple populations, pooled analysis stratifying for the different populations was done using the Mantel-Haenszel χ^2 test (SPSS), together with the Breslow-Day test for homogeneity of the odds ratio (OR) across populations. Logistic regression was also used to assess for potential confounding by patient covariates in each study group. The patient covariates included for logistic regression are: HIV status (KB study), anthropometrically and clinically defined severe malnutrition (KB study), and age (as a continuous variable, all studies), gender and ethnic group (all studies). The r^2 algorithm within HaploView v3.2 was utilized for analysis of pair-wise linkage disequilibrium (LD) between SNPs, as this method does not adjust for minor allele frequency, and as such is a direct estimate of LD. The 28 additional SNPs genotyped for exclusion mapping were selected from the HapMap database and were of varying minor allele frequencies (between 5% to 40%) to better capture haplotypic diversity across the 2 Mb region.

Descriptions of genetic modes of inheritance:

a) General allelic: Wild-type allele (W) vs variant allele (V); this is a general model of inheritance in which individuals carrying one copy of the variant allele are at increased risk of disease, and it follows that individuals carrying two copies are at even higher relative risk. The majority of complex trait genetics follow the general allelic model, although notable exceptions exist for certain phenotypes (e.g. Recessive for CCR5 delta32 and HIV; heterozygous model for sickle cell trait and G6PD with malaria).

b) Dominant: Wild-type genotype (WW) vs presence of variant allele (WV+VV); For this model of inheritance, the presence of a single copy of the variant allele is sufficient to account for disease susceptibility. The presence of the second variant allele does not confer any additional disease risk over and above that of the first allele. Disease examples include osteogenesis imperfecta and hypertrophic cardiomyopathy.

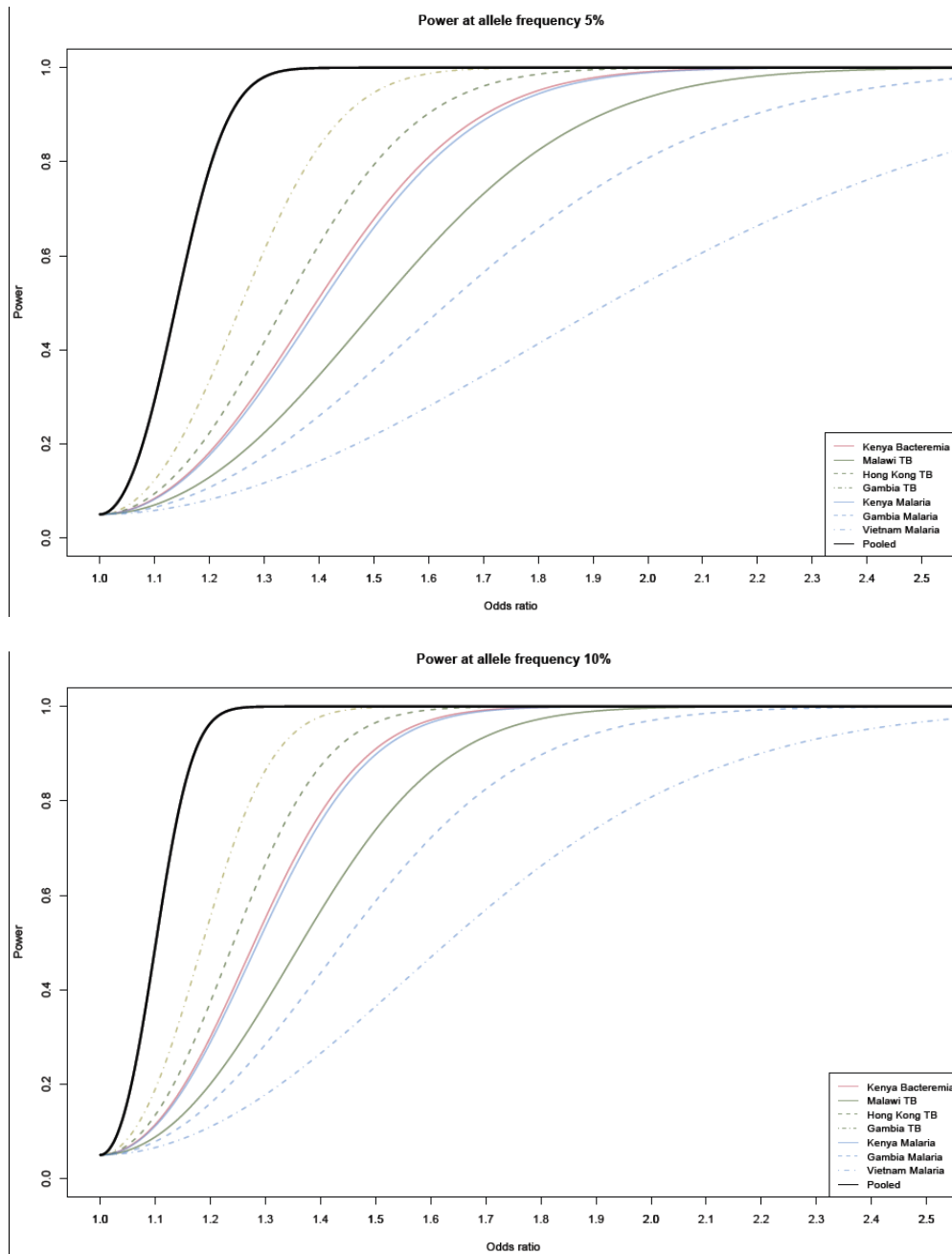
c) Recessive: Variant genotype (VV) vs presence of wild-type allele (VM+WW). The recessive model of inheritance require both copies of the variant allele to be present in a given individual. Individuals carrying only one copy of the variant allele are often phenotypically indistinguishable from wild-type individuals. Many enzymopathies have a recessive mode of inheritance (e.g. maple syrup urine disease, hypertyrosinemia).

Statistical analysis for gene expression data

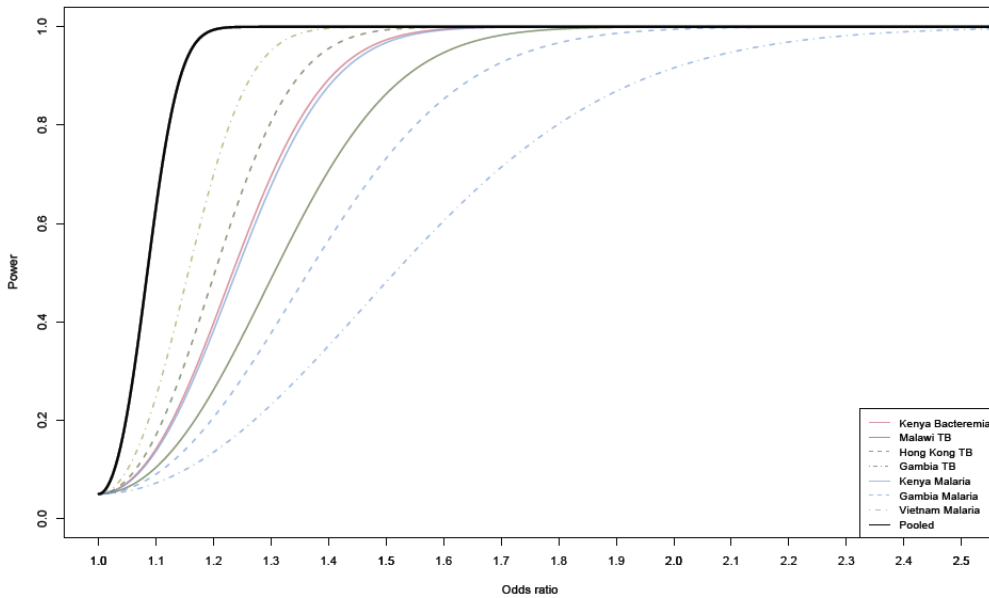
The expression of *CIS1* mRNA relative to *beta-2 microglobulin* was calculated using the comparative Ct ($\Delta\Delta C_T$) method as described in the manufacturer's protocol (Applied Biosystems; ABI prism). All $\Delta\Delta C_T$ s were then normalized to the average $\Delta\Delta CT$ of the wild type genotype at time 0. The fold change ($2^{-\Delta\Delta CT}$) and area under the curve (using the Trapezoid rule) were computed. The area under the curve (AUC) represented the total amount of mRNA produced over the 2 hour period. The fold change at each time point, as well as the AUCs was compared between each genotype group using the student's t-test (SPSS version 16.0). This analysis was done using each of following genetic modes of inheritance; Allelic: Wild-type allele (W) vs Mutant allele (M); Recessive: [Wild-type genotype (W/W) + carriers (W/M)] vs homozygous mutant (M/M); Dominant: WW vs (W/M + M/M). For the fitting of the allelic model, each individual was classified as having 0, 1 or 2 mutant alleles. The model denotes that the odds ratio comparing 1 vs 0 mutant alleles is similar as for 2 vs 1 mutant alleles (additive effect).

Statistical power calculations

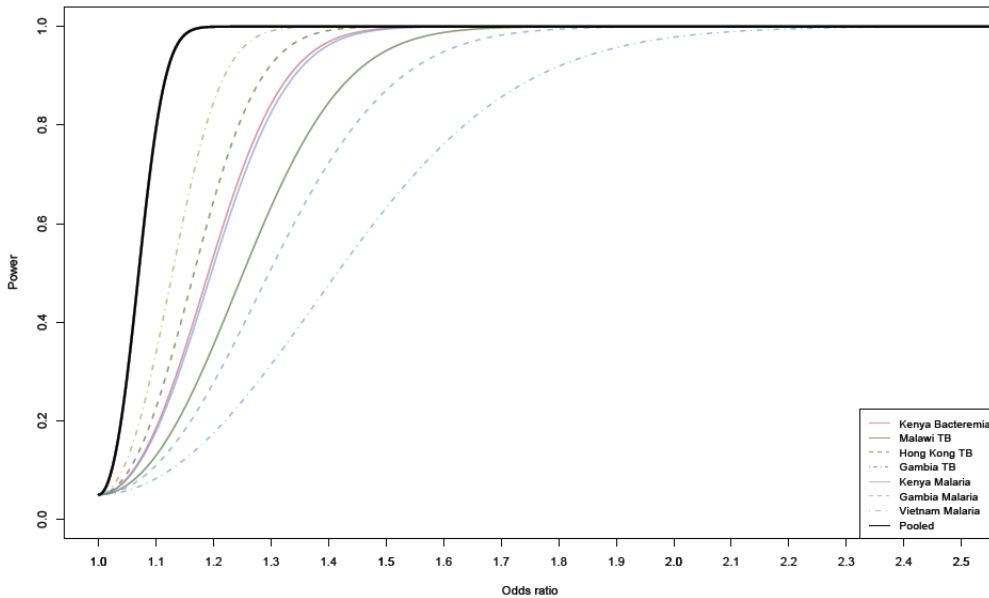
Supplementary Figure 1: Statistical power to detect association of genetic variants at $P = 0.05$ for each study cohort and the combined sample at different minor allele frequencies.



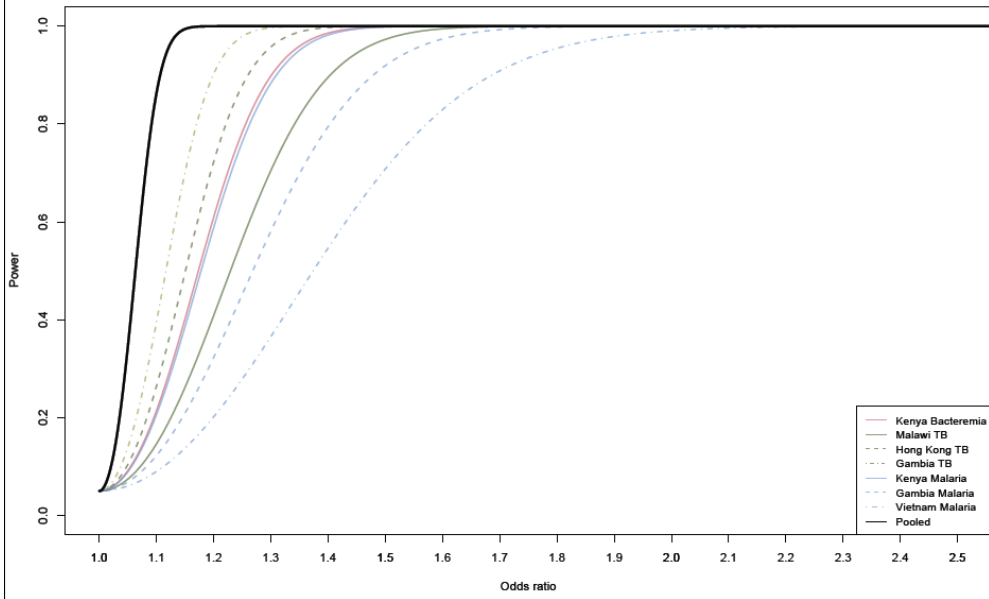
Power at allele frequency 15%



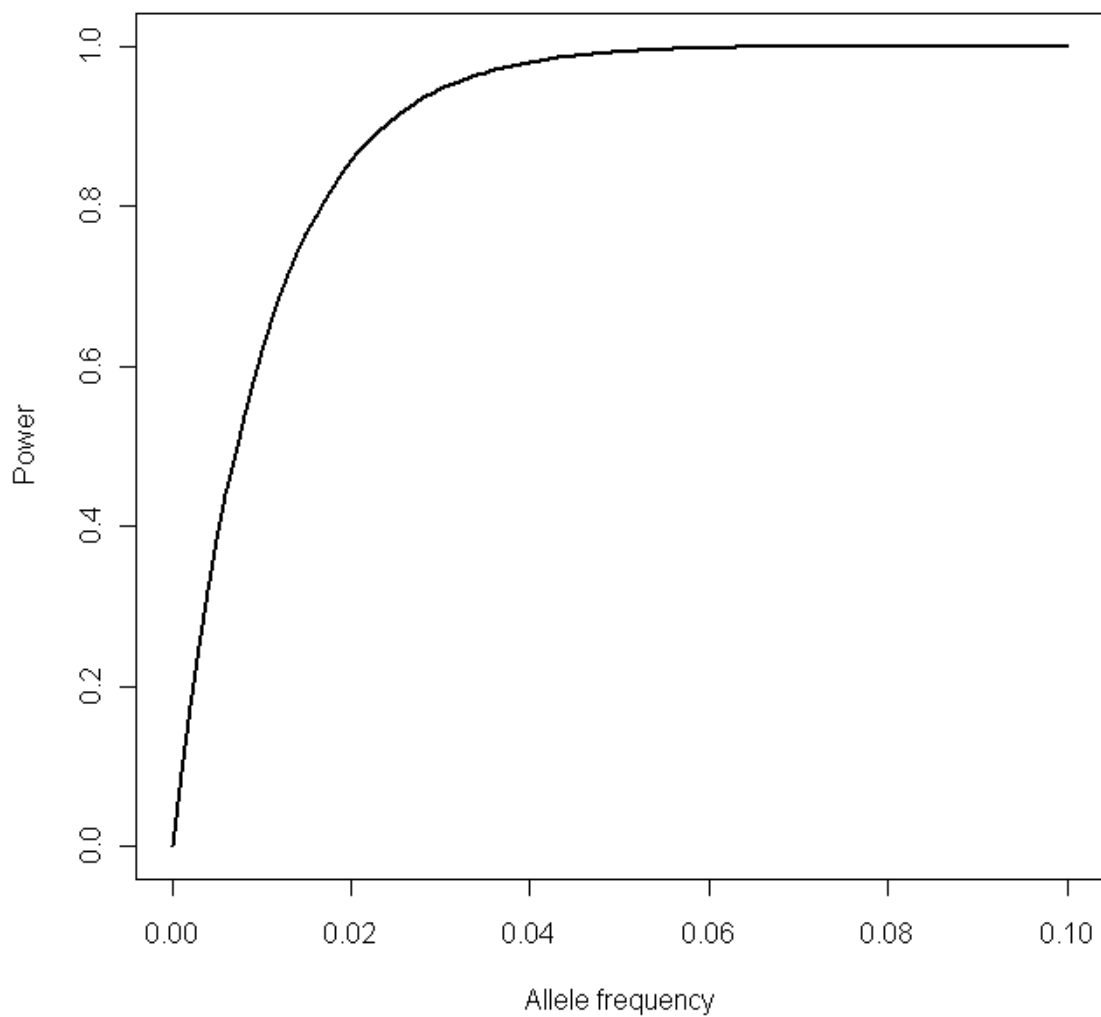
Power at allele frequency 25%



Power at allele frequency 35%



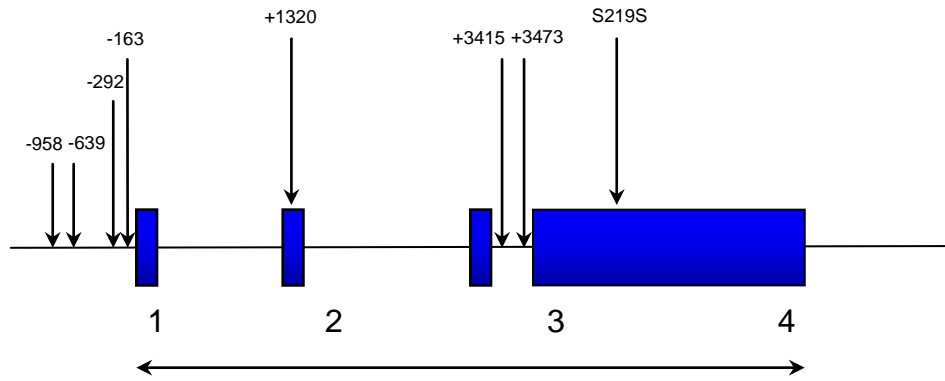
Supplementary Figure 2: Power to detect SNPs of a given minor allele frequency (MAF) via direct sequencing of 48 individuals.



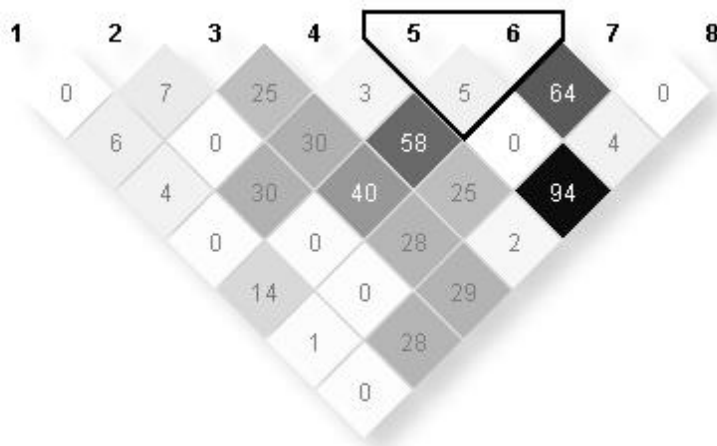
Statistical power to detect SNPs of MAF=0.03 is 94.63% and MAF=0.05 is 99.27%.

Supplementary Figure 3

a)



b)



Analysis of linkage disequilibrium between pairs of markers using the R-squared (r^2) algorithm on Haploview v3.2. Numbers within the diamonds indicate the r^2 value expressed as a percentile. Increasing degrees of r^2 value is denoted by squares of a darker shade. The SNPs shown from 5' to 3' (corresponding to numerals 1 – 8) are: *CISH* -958, -639, -292, -163, +1320 +3415, +3473, and S219S. Exons 1 to 4 span 5283 base pairs.

Supplementary Table 1

Multi-SNP analysis of *CISH* -639, -292, -163, +1320 and +3415 in all study cohorts. Trend tests were performed for the increasing number of *CISH* risk alleles (0, 1, 2, 3, or ≥ 4) carried and increasing odds of disease susceptibility for each population.

Population specific disease odds ratios										
No. of risk alleles carried	KB N=1198	MTB N=605	HKTB N=1609	GTB N=2688	All TB N=4902	GM N=647	KM N=1020	VM N=470	All Malaria N=2137	All cohorts N=8237
0	1	1	1	1	1	1	1	1	1	1
1	1.79	1	0.9	1.28	1.12	1.24	0.64	1.58	1.01	1.18
2	1.37	0.88	1.22	1.28	1.2	2.13	0.74	1.65	1.32	1.26
3	2.08	1.54	1.2	1.28	1.29	1.92	1.34	1.38	1.5	1.46
≥ 4	2.36	1.77	1.43	1.28	1.4	4.19	1.7	-	2.6	1.81
Overall Trend <i>P</i>	5.1×10^{-5}	0.03	0.011	0.034	8.7×10^{-4}	4.7×10^{-3}	1.7×10^{-3}	0.13	1.96×10^{-6}	3.8×10^{-11}

KB: Kenyan Bacteraemia study

MTB: Malawian tuberculosis study

HKTB: Hong Kong tuberculosis study

GTB: Gambian tuberculosis study

GM: Gambian malaria study

KM: Kenyan malaria study

VM: Vietnam malaria study

Supplementary Table 2: The twenty genes (listed within broad gene groups) and their accompanying variants tested in one or more of our patient collections.

Gene Group	Number of genes	Number of SNPs analyzed
Toll-like receptor pathway	6	44
NFKB signaling pathway	5	67
Cytokines and receptors	5	68
Miscellaneous genes	4	8
Total	20	187

Accession numbers

CISH; full name: cytokine inducible SH2-containing protein; other names: *CIS*, *CIS1*, *SOCS*, *G18*; gene ID: 1154; location: 3p21.3; MIM: 602441; SNP ID: *CISH* -292 (rs414171), *CISH* -163 (rs6768330), *CISH* +1320 (rs2239751), *CISH* +3415 (rs622502).

Author contributions

The authors wish particularly to thank the following individuals for assistance with the original studies: Dolly Jackson-Sillah, Sarah Campbell, Richard Bellamy, Keith McAdam, Tumani Corrah, Nick Anstey and Brian Greenwood. C.C.K., F.O.V., S.J.C., and S.W., performed genotyping, analyzed data, and wrote the article. H.G., and K.Y.T. carried out functional experiments on *CISH*. S.W. and D.V. performed the analysis statistical power calculations. A.J.W., D.V., A.R., T.M., K-C.C, K-MK, A.C.C., C.C.L., C.M.T., C-Y.C., J.J-Y.S., S.K.H.T., D.K., C.L., N.P.D., N.P., K.Marsh, K.Maitland, J.A.S., T.N.W., J.A.B., S.F., N.L.S.T., P.E.M.F., D.L.M.G. and AVSH contributed to the experimental design of the genetic studies. K-C.C, K-MK, A.C.C., C.C.L., C.M.T., C-Y.C., J.J-Y.S., D.K., C.L., N.P.D., N.P., K.Mar., K.Mai., J. A.S., T.N.W., J.A.B., S.F., N.L.S.T., and P.E.M.F contributed to the design and collection of the case-control studies. All authors critically reviewed and contributed intellectual input to the final manuscript. C.C.K., D.L.M.G., and A.V.S.H. led the research effort and vouch for the data and analysis. The authors collectively decided to publish this paper; this paper was published with permission of the director of the Kenya Medical Research Institute (KEMRI).