Published in final edited form as:

Hum Genet. 2009 February ; 125(1): 105–109. doi:10.1007/s00439-008-0597-2.

A genetic association study in The Gambia using tagging polymorphisms in the major histocompatibility complex (MHC) class III region implicates a HLA-B associated transcript 2 (BAT2) polymorphism in severe malaria susceptibility

Mahamadou Diakite^{1,2,5}, Taane G. Clark^{1,3,5,*}, Sarah Auburn³, Susana G Campino³, Andrew E Fry¹, Angela Green¹, Andrew P. Morris¹, Anna Richardson¹, Muminatou Jallow⁴, Fatou Sisay-Joof⁴, Margaret Pinder⁴, Dominic P. Kwiatkowski^{1,3}, and Kirk A. Rockett¹

¹Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, OX3 7BN, UK

²Malaria Research and Training Centre, Faculty of Medicine, Pharmacy and Odontostomatology, University of Bamako, BP 1805, Bamako, Mali

³Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, CB10 1SA, UK

⁴Medical Research Council Laboratories, Banjul, The Gambia

Abstract

The tumour necrosis factor (TNF) gene and other genes flanking it in the major histocompatibility complex (MHC) class III region are potentially important mediators of both immunity and pathogenesis of malaria. We investigated the association of severe malaria with eleven haplotype tagging-polymorphisms for eleven MHC class III candidate genes, including TNF, lymphotoxin alpha (LTA), allograft inflammatory factor 1 (AIF1), and HLA-B associated transcript 2 (BAT2). An analysis of 2162 case-controls demonstrated the first evidence of association between a BAT2 polymorphism (rs1046089) and severe malaria.

Keywords

Genetic association; haplotype analysis; malaria; MHC class III region; AIF1; BAT2; TNF

INTRODUCTION

Malaria caused by *Plasmodium falciparum* parasites is responsible for 1-2 million deaths per annum, mainly afflicting children in sub-Saharan Africa. It is a complex disease with many genetic and environmental determinants influencing the observed variation in response to infection, progression and severity. In order to develop effective vaccines and anti-malaria therapies, it is crucial to understand the mechanisms of protective immunity against malaria as well as its immunopathology. The human major histocompatibility complex (MHC) class III region contains many genes that are relevant to immunity and inflammation. Specifically, it contains the tumour necrosis factor (TNF)- α genomic region (6p21.3) that includes a

*Corresponding Author: Dr Taane G Clark, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, OX3 7BN, UK, tgc@well.ox.ac.uk.

⁵Contributed equally to this work

Competing interests

The authors declare that no competing interests exist.

number of interesting associations with malaria susceptibility and protection (reviewed in (Clark et al. in press; Kwiatkowski 2005)). This region spans approximately 150kb and includes the genes: (i) MHC class I polypeptide-related sequence B (MICB), (ii) HLA-B associated transcript 1 (BAT1), (iii) ATPase, H⁺-transporting, lysosomal, subunit g, isoform 2 (ATP6VIG2), (iv) nuclear factor of kappa light-chain gene enhancer in B cells inhibitor-like 1 (NFKBIL1), (v) lymphotoxin alpha (LTA), (vi) TNF, (vii) lymphotoxin beta (LTB), (viii) leucocyte-specific transcript 1 (LST1), (ix) natural cytotoxicity triggering receptor 3 (NCR3), (x) allograft inflammatory factor 1 (AIF1), and (xi) HLA-B associated transcript 2 (BAT2) (MHC-Consortium 1999). The LTA+80 polymorphism has been found to be functional, but the evidence of functionality of other polymorphisms in this region is less clear (Knight et al. 2004) (see (Clark et al. in press) for a review of potentially functional TNF promoter SNPs e.g. TNF-376).

We investigated associations between SNPs in the TNF- α region and severe malaria in The Gambia, but were concerned with several mapping issues. First, the MHC class III region has been found to have a complex haplotype structure (Ackerman et al. 2003). Second, there are regions in the genome, for example the haemoglobin-beta (HBB) locus, where there are linkage disequilibrium (LD) differences between the mixed ethnic Gambian population and the potential reference HapMap Yoruba panel (MalariaGEN/WTCCC in press). Therefore, we used a set of Gambian trios (affected children and their parents) to identify haplotype-tagging SNPs (htSNPs) for our population. The resulting htSNPs were tested for association with severe malaria in a Gambian case-control cohort consisting of 2162 individuals, adjusting for the potentially confounding effects of population structure.

METHODS

Participants and malaria phenotypes

Patient samples were collected as part of ongoing epidemiological studies of severe malaria at the Royal Victoria Hospital, Banjul, The Gambia (1534 severe malaria cases; 628 cord blood controls). All DNA samples were collected and genotyped following approval from the relevant research ethics committees and informed consent from participants. All cases were children (median age 3.5 years; age range 1 month to 13 years) admitted to hospital with evidence of *P. falciparum* on blood film and clinical features of severe malaria (Marsh et al. 1995). Subjects were defined as having had cerebral malaria (CM) if their Blantyre coma score was less than or equal to 2 on presentation or early during admission. A second phenotypic subset of severe malarial anaemia (SMA) was defined as those subjects having had a haemoglobin concentration of less than 5g/dl or a haematocrit less than 15%. Participants with co-existing severe or chronic medical conditions (e.g. bacterial pneumonia, kwashiorkor) unrelated to a severe malarial infection were excluded. The overall severe malaria case mix includes: CM only (63.8%), SMA (19.1%), both CM and SMA (11.3%) and other (mostly respiratory distress) (5.8%). Controls were cord blood samples obtained from birth clinics in the same locality as the cases, and thought to approximate a random sample of the population, thus reflecting the true allele frequency. The majority (85.7%) of participants are from four ethnic groups (Mandinka (37.2%), Jola (18.8%), Fula (16.7%), Wolof (13.0%), other (14.3%)), consistent with the ethno-demographics of The Gambia. Limited statistical power due to sample size meant we did not consider individual analyses of each ethnic group.

Selection of SNPs and genotyping

Initially we identified 200 SNPs in publicly available databases (including ensembl, NCBI, HapMap) in the TNF- α genomic region (6p21.3; see Figure 1a) spanning 150kb and the genes: (i) MICB, (ii) BAT1, (iii) ATP6VIG2, (iv) NFKBIL1, (v) LTA, (vi) TNF, (vii) LTB,

(viii) LST1, (ix) NCR3, (x) AIF1, and (xi) BAT2. All selected SNPs were genotyped using MALDI-TOF mass-spectrometry on samples from 128 unrelated parental chromosomes (from 32 Gambian affected child-parental trios) to confirm their existence at a useful frequency, LD and to construct haplotypes. These trios are representative of the same ethnic groups as described above. Using the 128 unrelated parental chromosomes, fifty two SNPs were found to be polymorphic and did not deviate from Hardy-Weinberg equilibrium (HWE) (one degree of freedom chi-square test $P > 0.001$). Extreme deviations from HWE (e.g. $P < 0.001$) can be indicative of genotyping error (Teo et al. 2007). Application of a maximum entropy procedure (<http://www.well.ox.ac.uk/~rmott/SNPS>) selected eleven htSNPs (see Table 1 for the list) to be genotyped in all cases and controls using MALDI-TOF mass-spectrometry. These SNPs captured 98% of the entropy (a measure of haplotypic diversity) in the sample. The potentially functional TNF-376 polymorphism is in tight linkage ($D' = 0.98$) with the htSNP TNF-238 (Clark et al. in press). We also genotyped the HbS polymorphism (rs334), as it is known to be strongly associated with severe malaria (Kwiatkowski 2005).

Statistical methods

Case-control association analysis using SNP alleles, genotypes and haplotypes was undertaken using logistic regression, adjusting for HbS and self-reported ethnicity. Adjustment for self-reported ethnicity in the Gambian population has been found to be a robust approach to controlling the potentially confounding effects of population structure (MalariaGEN/WTCCC in press). Haplotypes were estimated using an expectation-maximisation algorithm, and score tests were applied to assess the level of evidence of both global and individual haplotype associations using the *haplo.stats* R library (<http://www.r-project.org>). Multiple independent tests can lead to an inflation of the false positive error rate and spurious associations. By definition there is low correlation between htSNPs, and a Bonferroni correction assuming twelve independent statistical tests and a false positive rate of 5%, leads to a significance threshold of $P = 0.004$.

RESULTS

There is low inter-SNP r^2 LD in the region (see Figure 1b), as expected from our tagging-based strategy that minimises the association between htSNPs. Specifically, all pairwise r^2 estimates are less than 0.42, except between rs928815 and LTA+80 ($r^2 = 0.94$). Two polymorphisms in the MICB gene deviated extremely from HWE in the 628 controls ($P < 10^{-10}$) (see Table 1) and cases ($P < 10^{-15}$), indicative of genotyping error (Teo et al. 2007), and were not considered for further analysis. Allelic tests of association (Table 1) suggest that TNF-238 (A vs G OR=1.33, $P=0.04$), AIF1(i) (C vs A OR=0.57, $P=0.004$), and BAT2(i) (G vs A OR=0.73, $P < 10^{-6}$) may be associated with severe malaria ($P < 0.05$), but only AIF1(i) and BAT2(i) are robust to a Bonferroni correction. No association was found for the functional LTA+80 polymorphism ($P=0.80$). Genotypic analysis suggest a heterozygous genotypic effect (AG vs AA/GG) for TNF-238 (OR = 1.58, $P=0.005$), a dominant C-allele effect (CC/CA vs AA) for AIF1(i) (OR = 0.61, $P=0.02$), and a recessive (G allele) effect (GG vs GA/AA) for BAT2(i) (OR = 0.48, $P < 10^{-6}$), but only the latter result is robust to a Bonferroni correction. Restricting the analysis to CM cases yields similar allele-based (TNF-238 OR=1.33, $P=0.05$; AIF1(i) OR = 0.53, $P=0.003$; BAT2(i) OR = 0.70, $P < 10^{-7}$) and genotype-based (TNF-238 OR = 0.75, $P=0.03$; AIF1(i) OR = 0.64, $P=0.04$; BAT2(i) OR = 0.48, $P < 10^{-7}$) results.

We investigated multi-SNP (haplotypic) effects on severe malaria (Table 2). Because of the fragmented inter-SNP LD (D') structure (Figure 1b) we considered the haplotypes in two sub-regions: (a) rs928815, LTA+80, TNF-1031, TNF-308, and TNF-238 (D' range: 0.943 – 0.999); (b) AIF1(i), AIF1(ii) and BAT2(i) (D' range: 0.853–1.000). A haplotype analysis of

the contiguous SNPs rs928815, LTA+80, TNF-1031, TNF-308 and TNF-238 indicated no strong association signals (global-statistic = 9.356, degrees of freedom (df) = 5, $P=0.096$). The only haplotype with the TNF-238-A allele (CGCGA) had an increased risk of severe malaria ($P=0.037$), consistent with the allelic-based tests. Restricting the analysis to cases with CM resulted in no evidence of haplotypic effects (global-statistic = 7.818, df = 5, $P=0.167$). An analysis of the contiguous AIF1(i), AIF1(ii) and BAT2(i) polymorphisms indicated a strong signal of overall association (global-statistic = 29.214, df = 4, $P<10^{-5}$). The haplotypes containing the BAT2(i)-G allele (AGG, CGG) have a reduced risk of severe malaria ($P=0.0002$), and the inclusion of the AIF1(i)-C allele (CGG) leads to a 40% reduced risk over that from AGG ($P=0.022$). Restricting the analysis to CM cases ($n=1152$) did not alter these results. We also confirmed the known protective effect (~90% reduced risk) of the HbS AS genotype (Table 1).

DISCUSSION

The TNF- α genomic region is dense with genes that may be mediators of malaria pathogenesis. Our data demonstrate the first evidence that alleles at the BAT2-rs1046089 polymorphism may be associated with severe malaria in the Gambian population. The BAT2 gene has microsatellite repeats which are associated with the age-at-onset of insulin-dependent diabetes mellitus (IDDM) and potentially involved with the inflammatory process of pancreatic beta-cell destruction during the development of IDDM (Hashimoto et al. 1999). The inflammatory role of BAT2 is especially interesting for malaria research given that a crucial balance between the induction of an inflammatory response against infection and inflammo-pathology appears to be key to severe disease progression (reviewed in (Langhorne et al. 2008)). We also found weaker association evidence for two other candidates with inflammatory roles: the AIF1-rs2259571 (Harney et al. 2008) and TNF-238 (Clark et al. in press) polymorphisms. However, neither of these associations is statistically significant after correction for multiple testing.

The AIF1-rs2259571 and the BAT2-rs1046089 polymorphisms are in high LD (D'), and it remains possible that the observed haplotypic associations with severe malaria arise from a functional variant in high linkage with these SNPs, especially as the MHC class III region is dense with malaria candidates. In addition, long-range LD patterns (Ackerman et al. 2003) and population-specific haplotypes (MalariaGEN/WTCCC in press) further complicate the picture. In order to define an optimal haplotype tagging SNP set for the case-control study we used a Gambian trio panel, as using another reference population (e.g. HapMap Yoruba panel) may have reduced the power of our analysis (MalariaGEN/WTCCC in press). Our initial set of SNPs is derived from publicly available databases, but many more SNPs are yet to be discovered. Future work should involve resequencing the BAT2 gene and the surrounding MHC class III region in the Gambian and other African study populations. The identification of potentially population-specific functional variants will assist the design of subsequent large-scale epidemiological and immunological functional studies of malaria.

Acknowledgments

We thank the patients from the Gambian study population, as well as the many investigators involved in the original study for their contributions. We thank Bronwyn MacInnis and Daniel Alcock for providing useful comments on an earlier version of this manuscript. This work was funded by the UK Medical Research Council, Wellcome Trust, Bill and Melinda Gates Foundation and Grand Challenges in Global Health and a PhD training fellowship from the International Atomic Energy Agency (MD).

REFERENCES

- Ackerman HC, Ribas G, Jallow M, Mott R, Neville M, Sisay-Joof F, Pinder M, Campbell RD, Kwiatkowski DP. Complex haplotypic structure of the central MHC region flanking TNF in a West African population. *Genes Immun.* 2003; 4:476–86. [PubMed: 14551600]
- Clark T, Diakite M, Auburn S, Campino S, Fry A, Green A, Richardson A, Small K, Teo Y, Wilson J, Jallow M, Sisay-Joof F, Pinder M, Griffiths M, Peshu N, Williams T, Marsh K, Molyneux M, Taylor T, Rockett K, Kwiatkowski D. Tumour necrosis factor and lymphotoxin alpha polymorphisms and severe malaria in African populations. *J Infect Dis.* in press.
- Harney SM, Vilarino-Guell C, Adamopoulos IE, Sims AM, Lawrence RW, Cardon LR, Newton JL, Meisel C, Pointon JJ, Darke C, Athanasou N, Wordsworth BP, Brown MA. Fine mapping of the MHC Class III region demonstrates association of AIF1 and rheumatoid arthritis. *Rheumatology (Oxford).* 2008
- Hashimoto M, Nakamura N, Obayashi H, Kimura F, Moriwaki A, Hasegawa G, Shigeta H, Kitagawa Y, Nakano K, Kondo M, Ohta M, Nishimura M. Genetic contribution of the BAT2 gene microsatellite polymorphism to the age-at-onset of insulin-dependent diabetes mellitus. *Hum Genet.* 1999; 105:197–9. [PubMed: 10987645]
- Knight JC, Keating BJ, Kwiatkowski DP. Allele-specific repression of lymphotoxin-alpha by activated B cell factor-1. *Nat Genet.* 2004; 36:394–9. [PubMed: 15052269]
- Kwiatkowski DP. How malaria has affected the human genome and what human genetics can teach us about malaria. *Am J Hum Genet.* 2005; 77:171–92. [PubMed: 16001361]
- Langhorne J, Ndungu FM, Sponaas AM, Marsh K. Immunity to malaria: more questions than answers. *Nat Immunol.* 2008; 9:725–32. [PubMed: 18563083]
- MalariaGEN/WTCCC. Genomic signals of association with malaria and population structure in West Africa. *Nature Genetics.* in press.
- Marsh K, Forster D, Waruiru C, Mwangi I, Winstanley M, Marsh V, Newton C, Winstanley P, Warn P, Peshu N, et al. Indicators of life-threatening malaria in African children. *N Engl J Med.* 1995; 332:1399–404. [PubMed: 7723795]
- MHC-Consortium. Complete sequence and gene map of a human major histocompatibility complex. The MHC sequencing consortium. *Nature.* 1999; 401:921–3. [PubMed: 10553908]
- Singal DP, Li J, Zhu Y. HLA class III region and susceptibility to rheumatoid arthritis. *Clin Exp Rheumatol.* 2000; 18:485–91. [PubMed: 10949724]
- Teo YY, Fry AE, Clark TG, Tai ES, Seielstad M. On the usage of HWE for identifying genotyping errors. *Ann Hum Genet.* 2007; 71:701–3. author reply 704. [PubMed: 17388941]

Figure 1a. A total of 150 kb segment of the MHC class III region (chromosome 6) encompassing tumour necrosis factor. Grey boxes denote the genes* found within this region

* MHC class I polypeptide-related sequence B (MICB), HLA-B associated transcript 1 (BAT1), ATPase, H⁺-transporting, lysosomal, subunit g, isoform 2 (ATP6V1G2), nuclear factor of kappa light-chain gene enhancer in B cells inhibitor-like 1 (NFKBIL1), lymphotoxin alpha (LTA), tumour necrosis factor (TNF), lymphotoxin beta (LTB), (viii) leucocyte-specific transcript 1 (LST1), natural cytotoxicity triggering receptor 3 (NCR3), allograft inflammatory factor 1 (AIF1), and HLA-B associated transcript 2 (BAT2)

Figure 1b. Linkage disequilibrium for the 628 Gambian controls

* the (NCBI build 36) coordinates on chromosome 6 for the polymorphisms from MICB(i) to BAT2(i) are 31570224, 31570300, 31614603, 31639194, 31648120, 31650287, 31651010, 31651080, 31691806, 31691910, and 31710946.

Table 1

Allelic frequencies and association for the Gambian case-control study

SNPs	RS	Function	Alleles [*]	Minor allele freq. Controls (628)	Cases (1534)	HWE ^{**} P-value	Allelic-based tests			Genotype-based tests		
							OR	95% CI	P-value	genotypes	OR	95% CI
HbS	rs334	C,Sy	A,S	0.080	0.012	0.54	0.14	0.10	0.21	AS v AA	0.07	0.04
MICB(i)	rs2534685	NA	A,G	0.471	0.496	<0.0001	1.10	0.96	1.26	AG v AA/GG	1.42	1.09
MICB(ii)	rs2516412	NA	C,A	0.450	0.436	<0.0001	0.94	0.82	1.08	AC v AA/CC	1.07	0.87
BAT1(i)	rs1129640	C,Sy	G,A	0.283	0.291	0.14	1.04	0.89	1.22	AG v AA/GG	1.18	0.94
rs928815	rs928815	NA	C,A	0.417	0.419	0.68	1.01	0.88	1.15	AA/AC v CC	0.97	0.78
LTA+80	rs2239704	UTR	G,T	0.419	0.416	0.77	0.99	0.86	1.13	GG/GT v TT	0.82	0.65
TNF-1031	rs1799964	P	T,C	0.128	0.137	0.15	1.08	0.89	1.32	CT v CC/TT	1.24	0.97
TNF-308	rs1800629	P	G,A	0.150	0.167	0.52	1.14	0.95	1.36	AA/AG v GG	1.12	0.91
TNF-238	rs361525	P	G,A	0.058	0.075	0.03	1.33	1.01	1.75	AG v GG/AA	1.58	1.15
AIF1(i)	rs2259571	INS	A,C	0.038	0.022	0.02	0.57	0.39	0.84	CC/AC v AA	0.61	0.40
AIF1(ii)	rs2269475	C,NS	G,A	0.181	0.205	0.38	1.17	0.98	1.38	AA/AG v GG	1.12	0.91
BAT2(i)	rs1046089	C,NS	A,G	0.475	0.398	0.15	0.73	0.64	0.84	GG v AA/AG	0.48	0.37
												<0.0001

OR = odds ratios, CI = confidence interval, NA = not known, C = coding, I = intron, NS = non-synonymous, Sy = synonymous, P = promoter, () = sample size

^{*} alleles = major, minor alleles^{**} one degree of freedom chi-square test of Hardy-Weinberg equilibrium applied to the 628 controls^{***} adjusted for ethnicity and HbS^{****} we performed additive, dominant, recessive and heterozygous advantage genotypic tests, adjusted for ethnicity and HbS, but only the most statistically significant result is presented; in bold P 0.004

Table 2

Haplotype association analysis

Sub-region	* Haplotype	** Control Freq.	Case Freq.	OR	95% CI	*** P-value
(a)	ATTGG	0.410	0.413	1.00		0.842
	CGTGG	0.302	0.273	0.90	0.76 1.06	0.055
	CGTAG	0.147	0.165	1.11	0.91 1.36	0.159
	CGCGG	0.067	0.058	0.87	0.66 1.15	0.285
	CGCGA	0.057	0.075	1.31	0.98 1.74	0.037
(b)	AGG	0.438	0.381	1.00		0.0002
	AGA	0.344	0.390	1.40	1.18 1.66	0.002
	AAA	0.181	0.205	1.35	1.12 1.63	0.059
	CGG	0.035	0.017	0.60	0.39 0.93	0.0002

* (a) rs928815, LTA+80, TNF-1031, TNF-308, and TNF-238; (b) AIF1(i), AIF1(ii) and BAT2(i)

** Haplotypes with frequency >1% in either cases or controls. In region (a) these accounted for 98.3% and 98.4% of haplotypes in controls and cases respectively, and in region (b) accounted for 99.8% and 99.3% of haplotypes in controls and cases respectively; OR = odds ratio, CI = confidence interval

*** based on a score test comparing this haplotype to the others; **in bold P 0.004**