npg

ARTICLE

SLC11A1 (formerly NRAMP1) and susceptibility to visceral leishmaniasis in The Sudan

Hiba Salah Mohamed[1,2], Muntaser Eltayeb Ibrahim[1], Elinor Nancy Miller[2], Jacqueline Katie White[2], Heather Jane Cordell[2], Joanna McCammond McGill Howson[2], Christopher Sean Peacock[2], Eltahir Awad Gasim Khalil[1], Ahmed Mohamed El Hassan[1] and Jenefer Mary Blackwell*[,2]

[1]*Institute of Endemic Diseases, University of Khartoum, Khartoum, Sudan;* [2]*Cambridge Institute for Medical Research, Wellcome Trust/MRC Building, University of Cambridge School of Clinical Medicine, Addenbrooke's Hospital, Hills Road, Cambridge CB2 2XY, UK*

Genetic susceptibility to visceral leishmaniasis (VL) is indicated by differences in incidence and clinical phenotypes between ethnic groups in Sudan. In mice, innate susceptibility to *Leishmania donovani*, the etiological agent of VL, is controlled by S*lc11a1* (formerly *Nramp1*). We therefore examined polymorphisms at *SLC11A1* in 59 multicase families of VL from the high-incidence Masalit tribe in Sudan. Multipoint nonparametric analysis in ALLEGRO shows a significant linkage across *SLC11A1* (Z_{lr} scores 2.38–2.55; $0.008 \leqslant P \leqslant 0.012$; information content 0.88). The extended transmission disequilibrium test shows biased transmission of alleles at 5′ polymorphisms in the promoter ($P = 0.0145$), exon 3 ($P = 0.0037$) and intron 4 ($P = 0.0049$), and haplotypes formed by them ($P = 0.0089$), but not for 3′ polymorphisms at exon 15 or the 3′UTR. Stepwise logistic regression analysis using a case/pseudo-control data set derived from the 59 families was consistent with main effects contributed by the intron 4 469 + 14G/C polymorphism. Although the two alleles for 469 + 14G/C lie on haplotypes carrying different alleles for the functional promoter GT$_n$ polymorphism, the latter did not itself contribute separate main effects. Sequence analysis of 36 individuals failed to identify new putative functional polymorphisms in the coding region, intron 1, intron/exon boundaries, intron 4/exon 4a, or in the 3′UTR. One novel promoter polymorphism (−86G/A) was located within a putative nuclear factor kappa B binding site that could be functional. Further work will determine whether additional polymorphisms occur upstream in the promoter, which could be in linkage disequilibrium with the intron 4 polymorphism. These studies contribute to knowledge of the role of *SLC11A1* in infectious disease.

Introduction

The current outbreak of kala-azar or clinical visceral leishmaniasis (VL) caused by *Leishmania donovani sensu strictu* in eastern and southern Sudan has taken its toll in an

*Correspondence: Prof. JM Blackwell. Tel: + 44 1223 336947;
Fax: + 44 1223 331206; E-mail: jennie.blackwell@cimr.cam.ac.uk
Received 13 March 2003; accepted 1 August 2003

impoverished, war-stricken population, killing tens of thousands and depopulating vast areas of southern Sudan. The calamity brought kala-azar to the forefront as one of the greatest epidemics of the 20th century.[1,2] Longitudinal studies in Sudan show marked differences in incidence of clinical disease between adjacent villages inhabited by different ethnic groups.[3,4] Furthermore, when members of the different ethnic groups share the same immediate

environment and exposure, certain ethnic groups remain at a higher risk of developing VL.[5] These observations support the hypothesis that the host genotype plays an important role in disease susceptibility.

VL caused by *L. donovani sensu strictu* in the mouse provides a classical model of major single gene control of innate resistance to infection.[6–10] The gene, designated *Lsh, Ity* or *Bcg*, was also shown to influence innate resistance to *Salmonella typhimurium*,[11–13] *Mycobacterium bovis* BCG,[14] *M. lepraemurium*[15,16] and *M. intracellulare*.[17] Following its identification by positional cloning,[18] the gene was renamed the natural resistance-associated macrophage protein (*Nramp1*). This is now superseded by the functional designation solute carrier family 11a (proton-coupled divalent metal ion transporters) member 1 or *Slc11a1*, consistent with our formal demonstration that the proteins encoded by *Slc11a1/SLC11A1* function as proton/divalent cation (Fe^{2+}, Zn^{2+} and Mn^{2+}) antiporters.[19] Mice carrying the natural gly-to-asp mutation at position 169 in putative membrane-spanning domain number 4 are as susceptible to infection with all the three groups of pathogen as gene disrupted ('knockout') mice.[20] The protein localises to the late endosomal/lysosomal compartment of macrophages,[21,22] but not to early endosomes.[21] In man, *SLC11A1* has been linked to genetic susceptibility to leprosy in Vietnam[23] and to tuberculosis in Brazil[24] and in aboriginal Canadians.[25] Indeed, *SLC11A1* appears to be globally associated with tuberculosis,[24–29] with evidence that both 5′ and 3′ polymorphisms contribute independently to susceptibility.[26] Linkage or allelic association has also been demonstrated for HIV[30] and for a wide range of autoimmune diseases in man.[31–37] Here, we present data demonstrating a role for *SLC11A1* in determining susceptibility to VL in the highly susceptible Masalit tribe in Sudan.

Subjects and methods
Ascertainment of families
The study was carried out on the Nilosaharan-speaking Masalit population who migrated from western Sudan in the early–mid 1980s to occupy villages along the Rahad River in the heart of the endemic area in eastern Sudan. Epidemiological and demographic details relating to the study site are described in detail elsewhere.[3,4,38,39] Multicase families with VL were ascertained from epidemiological and medical records of the Institute of Endemic Diseases. Diagnosis was made on the basis of clinical, parasitological and serological criteria, as described.[3,39] Ethical approval for this study was obtained from the Ethical Committee of the University of Khartoum. Buccal swab DNA was prepared from 312 individuals from 59 extended multicase families (67 nuclear families) with one to six affected offspring per nuclear family (Table 1). Nuclear families with one case were always part of an extended multicase pedigree.

Table 1 Family structures for the 59 Masalit families collected from El Rugab and Um Salala villages in eastern Sudan

Families structure	Number
No. of families	59
No. of nuclear families	67
Nuclear families with one affected sib	9
Nuclear families with two affected sibs	28
Nuclear families with three affected sibs	20
Nuclear families with four affected sibs	8
Nuclear families with five affected sibs	1
Nuclear families with six affected sibs	1
No. of affected offspring	168
No. of affected parents	9
Total no. of affected individuals	177
Total no. of individuals	312

Nuclear families with a single affected offspring were always part of an extended multicase pedigree.

Genotyping
Three SNPs within *SLC11A1* (274C/T, 469 + 14G/C and D543N) were genotyped using the PCR primers, restriction enzymes and conditions described.[40] The microsatellite *D2S1471*[41] and the $(GT)_n$ repeat in the promoter region of *SLC11A1*[42] were PCR amplified using NED- and FAM-labelled forward primers, and PCR products analysed by electrophoresis on 6% polyacrylamide gels using an automated sequencer (model ABI377, Applied Biosystems) as previously described.[37] Two insertion/deletion (IN/DEL) polymorphisms in the 3′UTR of *SLC11A1*[40,43] were also typed using the automated sequencer. The 3′UTR TGTG polymorphism[40] was amplified using FAM-labelled (5′ TAC CTG CAG TAG GGC CA 3′) and unlabelled (5′ AAA CAG CAG GTC CCT AAA GC) primers. This yielded allele sizes of 393 bp (allele 1) and 397 bp (allele 2). The 3′UTR CAAA polymorphism[43] was amplified using HEX-labelled (5′CTC CAG TTT GGA GCC TGT GT 3′) and unlabelled (5′ CTA GCG CAG CCA TGT GAT TA 3′) primers. This yielded allele sizes of 245 bp (allele 1) and 250 bp (allele 2). Microsatellites and IN/DELs were analysed using the computer software genescan and genotyper (Applied Biosystems). Allele frequencies for the polymorphisms used in this study are shown in Table 2.

Linkage analysis
Nonparametric linkage analysis was performed in ALLEGRO[44] using the S_{pairs} scoring function with 0.5 weighting to take account of differences in family size. ALLEGRO reports allele-sharing LOD scores and maximum Z scores for the likelihood ratio (Z_{lr}). One-sided P-values associated with LOD and Z_{lr} scores are used throughout. For multipoint analysis, genetic distances between markers were entered in cM calculated on the basis of the physical map distance (1 Mb = 1 cM). The fraction of the total inheritance information extracted by the available marker data is indicated by the 'information content'. Simulations

Table 2 Allele frequencies for *SLC11A1* markers and *D2S1471*

Marker	No. of alleles	Heterozygosity	Allele frequencies
(GT)$_n$	4	0.394	0.008; 0.246; 0.738; 0.008
274C/T	2	0.322	0.201; 0.799
469+14G/C	2	0.212	0.880; 0.120
D543N	2	0.211	0.120; 0.880
3′UTR TGTG	2	0.305	0.188; 0.812
3′UTR CAAA	2	0.450	0.501; 0.499
D2S1471	17	0.888	0.130; 0.008; 0.031; 0.199; 0.145; 0.078; 0.033; 0.122; 0.012; 0.052; 0.033; 0.082; 0.023; 0.036; 0.012; 0.004; 0.008

performed in ALLEGRO using *SLC11A1* marker information from this study demonstrated that the 59 families had 100% power to detect linkage up to a critical value equivalent to an allele-sharing LOD score >3.95 ($P<0.00001$) for a gene with penetrances 7–61%,[45] controlling the underlying susceptibility to VL.

Allelic association testing
Family-based allelic association tests were performed using the TDT[46] correcting for clustering at the nuclear family level and nonindependence between sibs using a robust sandwich estimator for the variance and the Wald χ^2-test. Single-point and haplotype TDT were also performed using the ETDT[47] implemented within TDTPHASE.[48] Allelic and genotype associations, and relative risk estimates, were obtained by creating a 'case/pseudo-control' study, where the 'cases' comprise the genotypes of the affected offspring, and the 'controls' are the one to three other genotypes (depending on whether phase is known or inferred) which the affected offspring might have received from the parents.[49] The relative risks were estimated using conditional logistic regression analysis, again employing robust variance estimates to control for family clustering and a score test to indicate the overall significance for allelic or genotype associations. A stepwise logistic-regression procedure[49] was used to evaluate the relative importance of variants at the different sites within *SLC11A1*. Score tests were used to compare models in which the main effects for both loci are modelled with one in which the main effects at the primary locus only are included. Robust TDT and case/pseudo-control statistical tests implemented within Stata were developed by Heather Cordell and David Clayton at the Cambridge Institute for Medical Research, and are available at http://www-gene.cimr.cam.ac.uk/clayton/software/.

Linkage disequilibrium
Linkage disequilibrium between pairs of markers across *SLC11A1* was determined using Hedrick's definition of Lewontin's D' statistic.[50]

Sequence analysis
Direct cycle sequence analysis was performed on DNA samples isolated from 29 cases (25 for exon 4a) from 27 families, and seven unaffected (two children, five adults), used in the linkage and association studies. The affected individuals were selected from a subset of families that showed the highest LOD scores for linkage to *SLC11A1*. Separate PCR assays were designed to amplify the 15 exons, intron 1 and all the intron/exon boundaries of *SLC11A1*, the region of intron 4 (nt 3046–3503) around and including alternatively spliced exon 4a[51] (nt 3191–3264), ~600 bp of sequence upstream of the transcription start site and 819 bp of sequence downstream of the TAG stop codon. The TOPO TA Cloning® kit (Invitrogen, living science) was used to clone PCR-amplified products to facilitate sequence analysis of the 58 individual chromosomes, especially through the promoter region repeats. However, it was not possible to determine whether each new SNP was in *cis* or *trans* to SNPs identified on other PCR products in the same individual.

Results
Linkage analysis
A breakdown of family structures for the 59 multicase families used in this study is given in Table 1. Allele frequencies for the *SLC11A1* and *D2S1471* loci derived from genetically independent individuals in the families are shown in Table 2. Multipoint nonparametric linkage analysis in ALLEGRO (Table 3) provides evidence ($Z_{lr_{pairs}}$ scores 2.55–2.38; $0.008 \leqslant P \leqslant 0.012$; information content 0.88) for linkage between VL and markers across *SLC11A1*, dropping away at *D2S1471* ($Z_{lr_{pairs}}$ score 1.99; $P=0.028$;

Table 3 Multipoint nonparametric linkage analysis using the S_{pairs} function in ALLEGRO[44] to examine linkage beween the VL phenotype and markers across the *SLC11A1* and adjacent *D2S1471* region

Marker	LOD score	Z_{lr} score	P	Information content
(GT)$_n$	1.41	2.55	0.008	0.88
274C/T	1.40	2.54	0.008	0.88
469+14G/C	1.40	2.54	0.008	0.87
D543N	1.24	2.39	0.012	0.88
3′UTR TGTG	1.23	2.38	0.012	0.88
3′UTR CAAA	1.23	2.38	0.012	0.88
D2S1471	0.86	1.99	0.028	0.88

Allele-sharing LOD and Z_{lr} scores, with associated *P*-values and information content, are shown.

information content 0.88) lying 68 kb distal (telomeric) to the 3′ end of SLC11A1.

Family-based allelic association

Extended transmission disequilibrium testing (ETDT)[47] for single markers implemented within TDTPHASE[48] showed significant global associations (Table 4a) for the 5′ promoter (GT_n) and intragenic single-nucleotide polymorphisms (SNPs) at 274C/T and 469$+$14G/C, but not for the 3′ exon 15 SNP D543N, the 3′UTR insertion/deletion (IN/DELs) polymorphisms, or for the D2S1471 microsatellite lying $\sim$68 kb distal to SLC11A1. Single-point

TDT P-values at 274C/T and 469$+$14G/C retained significance after correcting for family clustering using TDT with a robust sandwich estimator for the variance (Wald χ^2 with 1 df $=$ 5.59, $P=0.0171$ for 274C/T; Wald χ^2 with 1 df $=$ 5.56, $P=0.0184$ for 469$+$14G/C), that is, markers within SLC11A1 show true allelic association with disease in the presence of linkage. Significant associations for 5′ markers at SLC11A1 were supported by the case/pseudo-control logistic regression analysis (Table 5a) that showed significant allelic associations at the GT_n repeat, as well as for the 274C/T and 469$+$14G/C biallelic markers. A borderline significant χ^2 test comparing allele-wise (1 df) and

Table 4 (a) Global allelewise significance levels (P-values) for family-based SLC11A1 allelic and haplotype associations with VL obtained using ETDT[47] implemented within TDTPHASE.[48] (b) Specific haplotype associations for the three 5′ markers (GT_n−274C/T−469+14C/G) showing significant individual marker and haplotype global associations between SLC11A1 and VL in Sudan

(a)

No. of markers	GT_n	274C/T	469+14C/G	D543N	3′UTRTGTG	3′UTRCAAA	D2S1471
One	**0.0145**	**0.0037**	**0.0049**	0.1972	0.5310	0.8785	0.0816
Two	**0.0112**						
		0.0115					
			0.0367				
				0.1655			
					0.2007		
Three	**0.0089**						
		0.0035					
			0.1465				
				0.2007			
Four	**0.0062**						
		0.0838					
			0.0203				
Five	**0.0309**						
		0.0329					
Six	**0.0240**						

(b)

Haplotypes	T	NT	%T	$\chi2(1)$	P
Disease associated					
3-2-1	64	34	65	8.80	**0.003**
4-1-2[a]	6	1	86	3.57	**0.058**
Disease protective					
2-1-2	8	24	25	7.93	**0.005**

Underlining in (a) indicates which groups of two, three, four, five or six markers were being haplotyped. Bold indicates P values that are significant ($<$0.05). $T=$ number of transmitted haplotypes; NT$=$ number of nontransmitted haplotypes. %T$=$ per cent transmitted. $\chi^2_{(1)}=$ Chi-squared with 1 degree of freedom; $P=$ P-value. [a]Allele 4 of GT_n – all transmissions from one extended pedigree.

Table 5 'Case/pseudo-control' allelic association testing[49] for *SLC11A1* markers and VL. (a) Shows *P*-values for conditional logistic regression analysis for allelic (1 df tests) and genotype (2 df tests) associations at the six polymorphic markers. A significant χ^2 test comparing the difference between 1 and 2 df tests indicates dominance rather than a multiplicative model. (b) Shows the results of score tests of the main effects of a locus in a forward stepwise regression procedure. Robust variance estimates to control for family clustering were used throughout

(a)

Test	GT_n allele 2	GT_n allele 3	GT_n[a] allele 4	274C/T[b]	469+14G/C[b]	D543N[b]	3'UTR TGTG[b]	3'UTR CAAA
rclogit								
1 df	0.1176	0.1273	**0.0119**	**0.0148**	**0.0039**	0.3219	0.4820	0.5208
2 df	0.2955	0.1239	—	—	—	—	—	0.6986
Dominance test	0.4294	**0.0535**	—	—	—	—	—	0.7728

(b)

Null model	Alternative model	χ^2	Test statistic df	P
	GT_n a3[c]	3.76	1	**0.0525**
	274[c]	6.06	1	**0.0138**
	469[c]	9.03	1	**0.0027**
GT_n a3	GT_n a3+274	5.56	1	**0.0184**
274	274+GT_n a3	2.45	1	0.1176
GT_n a3	GT_n a3+469	7.60	1	**0.0058**
469	469+GT_n a3	1.70	1	0.1919
274	274+469	4.29	1	**0.0383**
469	469+274	0.11	1	0.7450

[a]Allele 4 of GT_n found in heterozygous condition only, hence only 1 df test valid. [b]Only two different genotypes observed in cases, hence only 1 df test valid. Dominance effect only observed with the GT_n. [c]These test statistics differ slightly from the 1 df test data in Table 3a, because only individuals for whom all the three markers were fully typed could be included in this stepwise analysis. Bold indicates significant *P* value ($\leqslant 0.05$).

genotype-wise (2 df) tests for GT_n allele 3 indicates dominance rather than a simple multiplicative (ie both alleles contributing equally) model. The genotype-relative risks for carrying one or two copies of allele 3 at the GT_n were 9.63 (95% CI 1.07–86.96; $P=0.044$) and 10.20 (95% CI 1.06–98.16; $P=0.044$), respectively. The allele-wise relative risk associated with allele 4 at the GT_n was 6.00 (95% CI 1.50–23.99; $P=0.011$), with allele 2 at 274C/T was 2.00 (95% CI 1.16–3.44; $P=0.012$), and with allele 2 at 469+14G/C was 0.38 (95% CI 0.20–0.71; $P=0.003$). No dominance effects were observed for 274C/T and 469+14G/C. TDTPHASE (Table 4a) showed significant ($0.011\leqslant P\leqslant0.0089$) global associations for 5' (ie GT_n-274C/T-469+14G/C) intragenic haplotypes and for haplotypes extending from 5' to 3' markers within *SLC11A1* ($0.006\leqslant P\leqslant0.033$), but not for haplotypes involving only the 3' markers (ie D543N-TGTG-CAAA). The 5' haplotype associations involved a significant bias in the transmission of haplotype 3-2-1 ($P=0.003$) to the affected offspring, and significant protection ($P=0.005$) associated with haplotype 2-1-2 (Table 4b). Bias (borderline significant) in transmission of the haplotype 4-1-2 to the affected offspring occurred in a single extended pedigree.

Of the markers associated with VL in the 5' region of *SLC11A1*, only the promoter GT_n is known to be functional in regulating the expression of SLC11A1[52] and yet alleles at this locus were associated with only borderline significance. To determine whether the GT_n contributed significant main effects to the haplotype associations observed, we carried out tests (Table 5b) to determine the main effects at each locus in a forward stepwise logistic regression procedure.[49] For this stepwise analysis, we compared the contribution of the main disease-associated allele 3 at the GT_n with the other two biallelic markers. At the first stage (rows 1–3), the GT_n (borderline, $P=0.05$), 274C/T ($P=0.01$) and 469+14G/C ($P=0.003$) are all significant when included in the model on their own, that is, without accounting for effects at other loci. Once the GT_n is included in the model, both the 274C/T (row 4) and the 469+14G/C (row 6) add significant main effects. Once the 274C/T is included, only the 469+14G/C (row 8) and not the GT_n (row 5) has a main effect. Once the 469+14G/C is included, neither the GT_n (row 7) nor the 274C/T (row 9) have main effects. Overall, this stepwise analysis indicates that all of the association between *SLC11A1* and the disease can be accounted for by the 469+14G/C polymorphism. This is consistent with the observation (Table 6) that the 469+14G/C is in strong linkage disequilibrium with both the GT_n ($D'=0.7136$) and the 274C/T ($D'=0.8919$), even though the latter are not in strong linkage disequilibrium with each other.

Table 6 Pairwise linkage disequilibrium between markers across *SLC11A1* calculated using the Lewontin's D' [50]

	GT$_n$	274C/T	469+14G/C	D543N	3′UTR TGTG	3′UTR CAAA
3′UTR CAAA						0
3′UTR TGTG					0	0.7353
D543N				0	0.9140	1.000
469+14G/C			0	0.6436	0.6833	0.6631
274C/T		0	**0.8919**	0.4717	0.3547	0.0767
GT$_n$	0	0.2755	**0.7136**	0.4400	0.4960	0.1227

Bold indicates 5′ markers in strong linkage disequilibrium.

Table 7 Results of sequence analysis across ∼600 bp upstream of the transcription start site, all 15 exons, intron 1 and all intron/exon boundaries, and 819 bp downstream of the TAG stop codon in 29 cases from 27 of the families and seven unaffected individuals, used in the linkage and association studies

Name	Nucleotide/amino-acid change	Allele frequency for affected individuals (n = 29)
(GT)$_n$	Microsatellite, 5′ to exon 1	Allele 1 = 0 Allele 2 = 0.274 Allele 3 = 0.694 Allele 4 = 0.032
−86G/A	G or A at nucleotide 86 5′ of transcription start site	G = 0.93, A = 0.07
263C/T	C or T at nucleotide 263; CTG or TTG (Leu) at codon 39 in exon 2	C = 0.69, T = 0.31
299−11C/T	C or T at nucleotide −11 of intron 2	C = 0.84, T = 0.16
346C/T[a]	C or T at nucleotide 346; TTC or TTT (Phe) at codon 66 in exon 3	C = 0.85, T = 0.15
649−4T/G	T or G at nucleotide −4 of intron 5	T = 0.98, G = 0.02
787+22C/T	C or T at nucleotide +22 of intron 7	C = 0.79, T = 0.21
895C/T[b]	C or T at nucleotide 895; GGC or GGT (Gly) at codon 249 in exon 8	C = 0.69, T = 0.31
1192+9G/A	G or A at nucleotide +9 of intron 10	G = 0.98, A = 0.02
1192+18G/T	G or T at nucleotide +18 of intron 10	G = 0.98, T = 0.02
1801+55del4[c]	Deletion of TGTG in the 3′UTR, 55 nucleotides 3′ of the stop codon	In = 0.69, Del = 0.31
1801+263del4[d]	Deletion of CAAA in the 3′UTR, 267 nucleotides 3′ of the stop codon	In = 0.69, Del = 0.31
1801+417T/C	T or C in the 3′UTR, 417 nucleotides 3′ of the stop codon	T = 0.94, C = 0.06

Details of novel SNPs are shown, together with allele frequencies in the 29 cases for the previously identified[40,42,43] polymorphic markers. To conform with standard notation, all the previously identified and new SNPs are named in relation to the transcription start site located 148 bp 5′ of the methionine start codon.[42] Previous notations[40,43] were assigned relative to an arbitrary site 76 bp upstream of the methionine start codon. [a]Previously 274C/T.[40] [b]Previously 823C/T.[40] [c]Previously 1729+55del4.[40] [d]Previously 1992 (CAAA)n.[43]

Sequence analysis

The results of stepwise regression analysis for markers across the 5′ region of *SLC11A1* suggested that all of the association with VL is accounted for by the polymorphism at 469+14G/C. Although the functional GT$_n$ might contribute to the overall 469+14G/C effect, the results suggest that other functional polymorphisms in linkage disequilibrium with 469+14G/C must be present and contribute. Since the intronic 469+14G/C SNP is itself unlikely to influence function, further sequence analysis was required to determine whether novel polymorphisms occur within the Masalit populations of eastern Sudan, which could account for this association. To this end, we sequenced ∼600 bp upstream of the transcription start site, intron 1, all 15 exons and intron/exon boundaries, the region of intron 4 containing the alternatively spliced exon 4a, and 819 bp downstream of the TAG stop codon in 29 cases from 27 families and seven unaffected individuals used in the linkage and association studies. Table 7 shows the new SNPs identified, together with allele frequencies in the 29 cases for the previously identified[40,42,43] polymorphic markers that were included in the sequence

analysis. To conform with standard notation, all previously identified and new SNPs are named in relation to the transcription start site located 148 bp 5′ of the methionine start codon.[42] Previous notations[40,43] were assigned relative to an arbitrary site 76 bp upstream of the methionine start codon. Eight new SNPs were identified (Table 7), only one of which occurred in the coding region at codon 39 in exon 2. This was a silent mutation, a C-to-T substitution that resulted in a synonymous amino-acid substitution, which would not cause a functional change in the SLC11A1 protein. Five other SNPs occurred at the intron/exon boundaries, and one in the 3′UTR. All of these were again unlikely to mediate functional changes. Sequence analysis within intron 4 confirmed the presence of exon 4a in this Sudanese population, but no variants that might relate through linkage disequilibrium to the significance associated with the intron 4 variant 469+14G/C were found. One new SNP 86 nucleotide 5′ of the transcription start site (ie position −86 bp) was located within a putative nuclear factor kappa B (NFκB)-binding site, which could be functional in the regulation of gene expression. The wild-type allele shared seven out of 10 nucleotides in the NFκB

known consensus GGGRHTYYCC (where according to the universal genetic code, R = G or A; H = C, A or T and Y = C or T),[53] while the mutant shared only six from 10 nucleotides. This variant occurred in only four (0.06) of the 58 chromosomes sequenced from four affected individuals (ie heterozygous in all), and so was unlikely to account for disease susceptibility on its own. The variant allele occurred in *cis* with allele 3 of the previously identified GT_n polymorphism in one affected individual and in *cis* with allele 2 for three affected individuals.

Discussion

The data presented here provide evidence (maximum multipoint $Z_{lr} = 2.55$; $P = 0.008$) for linkage between *SLC11A1* and susceptibility to VL in members of the Masalit tribe in eastern Sudan, thus replicating the recent report[54] of linkage (maximum multipoint LOD score = 1.08; $P = 0.01$) between *SLC11A1* and VL in the related Aringa ethnic group to the north of our study population in Sudan. Our study extended this finding using family-based TDT and logistic regression analysis to demonstrate the allelic association with 5′ (GT_n, 274C/T, 469 + 14G/C) but not 3′ (D543N, 3′UTR TGTG, 3′UTR CAAA) markers within *SLC11A1*. In a previous case–control analysis, Bellamy *et al*[26] found a significant association between four intragenic *SLC11A1* polymorphisms ($CA_n = GT_n$, INT4 = 469 + 14G/C, D543N and 3′UTR TGTG) and susceptibility to pulmonary tuberculosis in The Gambia. As with our study in Sudan, the two polymorphic markers in the 5′ region of the gene (CA_n and INT4) were in strong linkage disequilibrium, as were the two markers in the 3′ region (D543N and 3′UTR TGTG), but markers in the 5′ region were not in linkage disequilibrium with the markers in the 3′ region. Since they had observed significant associations at all four markers, they concluded that separate polymorphisms in the 5′ and 3′ regions of the gene acted independently to control susceptibility to tuberculosis.[26] In Sudan, we have failed to find any evidence either for association with markers in the 3′ region of *SLC11A1* and VL, or any novel change of function mutations across this region. Nor did we find evidence for any novel (functional) variant in or around the alternatively spliced exon 4a that might account, through linkage disequilibrium, for the main effects observed at 469 + G/C using stepwise logistic regression analysis. Previous studies[51] have shown that exon 4a, encoded by an Alu element within intron 4, is transcribed *in vivo* but would introduce a termination codon in exon V, resulting in a truncated and hence nonfunctional SLC11A1 protein. At the mRNA level, the ratio of transcripts with/without exon 4a was 1:5 in macrophage cell lines.[51] Hence, we hypothesised that any variant (eg elimination of the splice acceptor site for transcription of the alternatively spliced product)

that caused a change in this ratio might influence the amount of functional SLC11A1 protein expressed. If such a variant occurs, it is not encoded at the site of exon 4a. However, variants in the promoter in linkage disequilibrium with 469 + 14G/C could be responsible for regulating the level of normal *SLC11A1* transcribed, and/or the ratio of normal to alternatively spliced product. In our study, there was no novel change of function mutations in the 5′ coding region or intron 1 of *SLC11A1* that would, on its own, account for the association with 469 + 14G/C. This suggests that there are further functional polymorphisms upstream of the GT_n in linkage disequilibrium with 469 + G/C that might contribute to these main effects. This could include the potentially functional variant at –86G/A, although the low frequency of this novel allele means that it is unlikely to account for all of the association with 469 + G/C that we have observed. It is also possible that linkage disequilibrium might extend 230 kb upstream to the gene (IL8RA) encoding the receptor for interleukin 8, although confirmation of the role of *Slc11a1* in susceptibility to VL using gene knockout in mice[20] makes this the most likely candidate for our associations with disease in man. Work is in progress to sequence a larger region of upstream sequence in the *SLC11A1* promoter, and to carry out functional reporter gene analysis of the –86G/A variant. For the moment, we conclude that the previously identified[40,42] promoter region GT_n polymorphism remains the only functional[52] variant identified to date at *SLC11A1*. Since it is in strong linkage disequilibrium with 469 + 14G/C, it likely makes a significant contribution to the haplotype associations observed and some discussion of its role is therefore of interest.

Firstly, it was of interest that the main disease-associated haplotype carried allele 3 at the GT_n, given the previous demonstration[52] that this allele drives high levels of reporter gene (and hence SLC11A1) expression. These reporter gene constructs carried the four allelic variants of the GT_n on constructs that incorporated 571–587 bp of promoter sequence 5′ of the ATG initiation codon.[52] Although these constructs differed only at the GT_n repeat, it is possible that other variants upstream of the GT_n may modulate the promoter activity of allele 3 in macrophages *in vivo*. Nevertheless, in our previous studies[31,37] examining the role of this polymorphism in determining susceptibility to autoimmune *vs* infectious disease susceptibility, we proposed that the high-expressing allele 3 would be associated with proinflammatory responses and autoimmune disease, while the low-expressing allele 2 would be associated with low antimicrobial activity and susceptibility to infectious disease. For those studies in which the promoter GT_n has been studied, this hypothesis has held true in relation to susceptibility to pulmonary tuberculosis associated with allele 2[24,26,28] and autoimmune diseases associated with allele 3.[31,33,36,37,55] In relation to infectious disease, the one exception to date is HIV, where disease was

associated with allele 3.[30] Given the propensity of tumour necrosis factor α (TNFα) to drive high levels of HIV expression via NFκB signalling,[56] association with a high *SLC11A1*-expressing allele is not surprising. Similarly, VL is known to be associated with a high proinflammatory TNFα response and cachexia.[57] In The Sudan, it is possible that high levels of expression of SLC11A1 driven by allele 3 have a detrimental effect on VL susceptibility by causing a high proinflammatory response. Conversely, in a recent study in The Gambia,[58] we demonstrated that the tuberculosis-associated allele 2 at the GT_n repeat was associated with high anti-inflammatory interleukin 10 responses. Hence, the delicate balance between the positive requirement for macrophage activation and antimicrobial activity and the negative effect of proinflammatory TNFα, both of which are pleiotropic effects of *SLC11A1*,[59,60] appear to occur across the spectrum of infectious diseases associated with polymorphism *SLC11A1* and not just in comparing infectious *vs* autoimmune disease susceptibility. Interestingly, in one family, disease was associated with the low-expressing[52] allele 4 of the *SLC11A1* GT_n polymorphism, even though it was carried on an otherwise protective 1–2 haplotype for 274C/T-469 + 14G/C. Whether, in this case, individuals are susceptible to VL because they fail to express SLC11A1, or because allele 4 was on a haplotype with other functional polymorphisms in the *SLC11A1* promoter, will require further investigation. Studies are in progress to examine SLC11A1 expression at the RNA and protein levels in individuals carrying the different *SLC11A1* haplotypes that we have seen associated with disease in this population. We will also determine whether the –86G/A variant, or other variants upstream in the promoter region, modulate the function of the GT_n repeat polymorphism.

Overall, our results have shown that *SLC11A1* is associated with susceptibility to VL in eastern Sudan. Further work is underway to determine the precise functional basis to disease-promoting variants in the *SLC11A1* gene, and to determine which of the many pleiotropic effects of the gene most affects disease phenotype in man. This we hope will contribute to a better understanding of the molecular basis to disease susceptibility, and the development of therapies appropriate to the susceptible individuals in this high-incidence population.

References
1 De Beer P, Harith AE, van Grootheest M, Winkler A: Outbreak of kala-azar in The Sudan. *Lancet* 1990; **335**: 224.

2 Perea WA, Moren A, Ancelle T, Sondorp E: Epidemic of visceral leishmaniasis in southern Sudan. *Lancet* 1989; **ii**: 1222–1223.

3 Zijlstra EE, El Hassan AM, Ismael A, Ghalib HW: Endemic kala-azar in Eastern Sudan, a longitudinal study on the incidence of clinical and subclinical infection and post-kala-azar dermal leishmaniasis. *Am J Trop Med Hyg* 1994; **51**: 826–836.

4 El Hassan AM, Zilijstra EE, Ismael A, Ghalib HW: Recent observations on the epidemiology of kala-azar in the eastern and central states of The Sudan. *Trop Geograph Med* 1995; **47**: 151–156.

5 Ibrahim ME, Lambson B, Yousif AO *et al*: Kala-azar in a high transmission focus: an ethnic and geographic dimension. *Am J Trop Med Hyg* 1999; **61**: 941–944.

6 Bradley DJ, Kirkley J: Variation in susceptibility of mouse strains to *Leishmania donovani* infection. *Trans R Soc Trop Med Hyg* 1972; **66**: 527–528.

7 Bradley DJ: Genetic control of natural resistance to *Leishmania donovani*. *Nature* 1974; **250**: 353–354.

8 Bradley DJ, Kirkley J: Regulation of *Leishmania* populations within the host. I. The variable course of *Leishmania donovani* infections in mice. *Clin Exp Immunol* 1977; **30**: 119–129.

9 Bradley DJ: Regulation of *Leishmania* populations within the host. II. Genetic control of acute susceptibility of mice to *Leishmania donovani* infection. *Clin Exp Immunol* 1977; **30**: 130–140.

10 Bradley DJ, Taylor BA, Blackwell J, Evans EP, Freeman J: Regulation of *Leishmania* populations within the host. III. Mapping of the locus controlling susceptibility to visceral leishmaniasis in the mouse. *Clin Exp Immunol* 1979; **37**: 7–14.

11 Plant J, Glynn AA: Natural resistance to *Salmonella* infection, delayed hypersensitivity and Ir genes in different strains of mice. *Nature* 1974; **248**: 345–347.

12 Plant J, Glynn AA: Genetics of resistance to infection with *Salmonella typhimurium* in mice. *J Infect Dis* 1976; **133**: 72–78.

13 Plant J, Glynn AA: Locating salmonella resistance gene on mouse Chromosome 1. *Clin Exp Immunol* 1979; **37**: 1–6.

14 Gros P, Skamene E, Forget A: Genetic control of natural resistance to *Mycobacterium bovis* (BCG) in mice. *J Immunol* 1981; **127**: 2417–2421.

15 Brown IN, Glynn AA, Plant J: Inbred mouse strain resistance to *Mycobacterium lepraemurium* follows the *Ity/Lsh* pattern. *Immunology* 1982; **47**: 149–156.

16 Skamene E, Gros P, Forget A, Patel PJ, Nesbitt MN: Regulation of resistance to leprosy by chromosome 1 locus in the mouse. *Immunogenetics* 1984; **19**: 117–124.

17 Goto Y, Buschman E, Skamene E: Regulation of host resistance to *Mycobacterium intracellulare in vivo* and *in vitro* by the Bcg gene. *Immunogenetics* 1989; **30**: 218–221.

18 Vidal SM, Malo D, Vogan K, Skamene E, Gros P: Natural resistance to infection with intracellular parasites: isolation of a candidate for *Bcg*. *Cell* 1993; **73**: 469–485.

19 Goswami T, Bhattacharjee A, Babal P *et al*: Natural-resistance-associated macrophage protein 1 is an H+/bivalent cation antiporter. *Biochem J* 2001; **354**: 511–519.

20 Vidal S, Tremblay ML, Govoni G *et al*: The *Ity/Lsh/Bcg* locus: natural resistance to infection with intracellular parasites is abrogated by disruption of the *Nramp1* gene. *J Exp Med* 1995; **182**: 655–666.

21 Gruenheid S, Pinner E, Desjardins M, Gros P: Natural resistance to infections with intracellular pathogens: The *Nramp1* protein is recruited to the membrane of the phagosome. *J Exp Med* 1997; **185**: 717–730.

22 Searle S, Bright NA, Roach TIA *et al*: Localisation of Nramp1 in macrophages: modulation with activation and infection. *J Cell Sci* 1998; **111**: 2855–2866.

23 Abel L, Sanchez FO, Oberti J *et al*: Susceptibility to leprosy is linked to the human NRAMP1 gene. *J Infect Dis* 1998; **177**: 133–145.

24 Shaw M-A, Collins A, Peacock CS *et al*: Evidence that genetic susceptibility to *Mycobacterium tuberculosis* in a Brazilian population is under oligogenic control: linkage study of the

candidate genes NRAMP1 and TNFA. *Tuberc Lung Dis* 1997; **78**: 35–45.

25 Greenwood CM, Fujiwara TM, Boothroyd LJ *et al*: Linkage of tuberculosis to chromosome 2q35 loci, including NRAMP1, in a large aboriginal Canadian family. *Am J Hum Genet* 2000; **67**: 405–416.

26 Bellamy R, Ruwende C, Corrah T, McAdam KPWJ, Whittle HC, Hill AVS: Variation in the NRAMP1 gene is associated with susceptibility to tuberculosis in West Africans. *N Engl J Med* 1998; **338**: 640–644.

27 Ryu S, Park YK, Bai GH, Kim SJ, Park SN, Kang S: 3′UTR polymorphisms in the NRAMP1 gene are associated with susceptibility to tuberculosis in Koreans. *Int J Tuberc Lung Dis* 2000; **4**: 577–580.

28 Gao PS, Fujishima S, Mao XQ *et al*: Genetic variants of NRAMP1 and active tuberculosis in Japanese populations. International Tuberculosis Genetics Team. *Clin Genet* 2000; **58**: 74–76.

29 Cervino AC, Lakiss S, Sow O, Hill AV: Allelic association between the NRAMP1 gene and susceptibility to tuberculosis in Guinea-Conakry. *Ann Hum Genet* 2000; **64**: 507–512.

30 Marquet S, Sanchez FO, Arias M *et al*: Variants of the human NRAMP1 gene and altered human immunodeficiency virus infection susceptibility. *J Infect Dis* 1999; **180**: 1521–1525.

31 Shaw M-A, Clayton D, Atkinson SE *et al*: Linkage of rheumatoid arthritis to the candidate gene NRAMP1 on 2q35. *J Med Genet* 1996; **33**: 672–677.

32 Hofmeister A, Neibergs HL, Pokorny RM, Galanduik S: The natural resistance-associated macrophage protein gene is associated with Crohn's disease. *Surgery* 1997; **122**: 173–179.

33 Esposito L, Hill NJ, Pritchard LE *et al*: Genetic analysis of chromosome 2 in type 1 diabetes: analysis of putative loci IDDM7, IDDM12, and IDDM13 and candidate genes NRAMP1 and IA-2 and the interleukin-1 gene cluster. IMDIAB Group. *Diabetes* 1998; **47**: 1797–1799.

34 Singal DP, Li J, Zhu Y, Zhang G: NRAMP1 gene polymorphisms in patients with rheumatoid arthritis. *Tissue Antigens* 2000; **55**: 44–47.

35 Yang YS, Kim SJ, Kim JW, Koh EM: NRAMP1 gene polymorphisms in patients with rheumatoid arthritis in Koreans. *J Korean Med Sci* 2000; **15**: 83–87.

36 Maliarik MJ, Chen KM, Sheffer RG *et al*: The natural resistance-associated macrophage protein gene in African Americans with sarcoidosis. *Am J Respir Cell Mol Biol* 2000; **22**: 672–675.

37 Sanjeevi CB, Miller EN, Dabadghao P *et al*: Polymorphism at NRAMP1 and D2S1471 loci associated with juvenile rheumatoid arthritis. *Arthritis Rheum* 2000; **43**: 1397–1404.

38 Zijlstra EE, el-Hassan AM: Leishmaniasis in Sudan. Visceral leishmaniasis. *Trans R Soc Trop Med Hyg* 2001; **95** (Suppl 1): S27–S58.

39 Zijlstra EE, el-Hassan AM: Leishmaniasis in Sudan. Post kala-azar dermal leishmaniasis. *Trans R Soc Trop Med Hyg* 2001; **95** (Suppl 1): S59–S76.

40 Liu J, Fujiwara TM, Buu NT *et al*: Identification of polymorphisms and sequence variants in human homologue of the mouse natural resistance-associated macrophage protein gene. *Am J Hum Genet* 1995; **56**: 845–853.

41 Copeman JB, Cucca C, Hearne CM *et al*: Linkage disequilibrium mapping of a type 1 diabetes susceptibility gene (IDDM7) to chromosome 2q31–q33. *Nat Genet* 1995; **9**: 80–85.

42 Blackwell JM, Barton CH, White JK *et al*: Genomic organization and sequence of the human NRAMP gene: identification and mapping of a promoter region polymorphism. *Mol Med* 1995; **1**: 194–205.

43 Buu NT, Cellier M, Gros P, Schurr E: Identification of a highly polymorphic length variant in the 3′UTR of NRAMP1. *Immunogenetics* 1995; **42**: 428–429.

44 Gudbjartsson DF, Jonasson K, Frigge ML, Kong A: Allegro, a new computer program for multipoint linkage analysis. *Nat Genet* 2000; **25**: 12–13.

45 Peacock CS, Collins A, Shaw MA *et al*: Genetic epidemiology of visceral leishmaniasis in northeastern Brazil. *Genet Epidemiol* 2001; **20**: 383–396.

46 Spielman RS, McGinnis RE, Ewens WJ: Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet* 1993; **52**: 506–516.

47 Sham PC, Curtis D: An extended transmission/disequilibrium test (TDT) for multi-allele marker loci. *Ann Hum Genet* 1995; **59**: 323–336.

48 Dudbridge F, Koeleman BP, Todd JA, Clayton DG: Unbiased application of the transmission disequilibrium test to multilocus haplotypes. *Am J Hum Genet* 2000; **66**: 2009–2012.

49 Cordell HJ, Clayton DG: A unified stepwise regression procedure for evaluating the relative effects of polymorphisms within a gene using case/control or family data: application to HLA in type 1 diabetes. *Am J Hum Genet* 2002; **70**: 124–141.

50 Hedrick PW: Gametic disequilibrium measures: proceed with caution. *Genetics* 1987; **117**: 331–341.

51 Cellier M, Govoni G, Vidal S *et al*: Human natural resistance-associated macrophage protein: cDNA cloning chromosomal mapping, genomic organization, and tissue-specific expression. *J Exp Med* 1994; **180**: 1741–1752.

52 Searle S, Blackwell JM: Evidence for a functional repeat polymorphism in the promoter of the human NRAMP1 gene that correlates with autoimmune *vs* infectious disease susceptibility. *J Med Genet* 1999; **36**: 295–299.

53 Lenardo MJ, Baltimore D: NF-kB: a pleiotropic mediator of inducible and tissue-specific gene control. *Cell* 1989; **58**: 227–229.

54 Bucheton B, Abel L, Kheir MM *et al*: Genetic control of visceral leishmaniasis in a Sudanese population: candidate gene testing indicates a linkage to the NRAMP1 region. *Genes Immun* 2003; **4**: 104–109.

55 Kotze MJ, de Villiers JN, Rooney RN *et al*: Analysis of the NRAMP1 gene implicated in iron transport: association with multiple sclerosis and age effects. *Blood Cells Mol Dis* 2001; **27**: 44–53.

56 Swingler S, Morris A, Easton A: Tumour necrosis factor alpha and interleukin-1 beta induce specific subunits of NFKB to bind the HIV-1 enhancer: characterisation of transcription factors controlling human immunodeficiency virus type 1 gene expression in neural cells. *Biochem Biophys Res Commun* 1994; **203**: 623–630.

57 Barral-Netto M, Badaro R, Barral A *et al*: Tumor necrosis factor (cachectin) in human visceral leishmaniasis. *J Infect Dis* 1991; **163**: 853–857.

58 Awomoyi AA, Marchant A, Howson JM, McAdam KP, Blackwell JM, Newport MJ: Interleukin-10, polymorphism in SLC11A1 (formerly NRAMP1), and susceptibility to tuberculosis. *J Infect Dis* 2002; **186**: 1804–1814.

59 Blackwell JM, Goswami T, Evans CAW *et al*: SLC11A1 (formerly NRAMP1) and disease. *Cell Microbiol* 2001; **3**: 773–784.

60 Blackwell JM, Searle S, Goswami T, Miller EN: Understanding the multiple functions of Nramp1. *Microbes Infect* 2000; **2**: 317–321.