Genetic variants of \textit{SLC11A1} are associated with both autoimmune and infectious diseases: systematic review and meta-analysis.

Nicholas S. Archer\textsuperscript{1,a}, Najah T. Nassif\textsuperscript{1}, Bronwyn A. O’Brien\textsuperscript{1}*  

\textsuperscript{1} School of Medical and Molecular Biosciences, University of Technology Sydney, PO Box 123, Broadway, NSW 2007, Australia.  
\textsuperscript{a} Present address: CSIRO, North Ryde, NSW, Australia.  

Corresponding author – Bronwyn A. O’Brien  
Corresponding author. Tel.: +61 2 9514 4104; fax: +61 2 9514 8206  
E-mail address: bronwyn.obrien@uts.edu.au (B.A. O’Brien).  

Keywords- SLC11A1, NRAMP1, meta-analysis, infectious disease, autoimmune disease
ABSTRACT

A systematic review and meta-analyses were undertaken to investigate the association of \textit{SLC11A1} genetic variants with disease occurrence. Literature searching indentified 109 publications to include in the meta-analyses assessing the association of 11 \textit{SLC11A1} variants with autoimmune and infectious disease. The (GT)\textsubscript{n} promoter alleles 2 and 3 (rs534448891), which alter \textit{SLC11A1} expression, were significantly associated with tuberculosis [OR=1.47 (1.30-1.66), OR=0.76 (0.65-0.89), respectively] and infectious disease [OR=1.25 (1.10-1.42), OR=0.83 (0.74-0.93), respectively]. However, while no association was observed with autoimmune disease, a modest significant association was observed with Type 1 diabetes [allele 2 OR=0.94 (0.89-0.98)]. Based on the stronger association of (GT)\textsubscript{n}, allele 2 with tuberculosis, compared to the protective effect of allele 3, we hypothesise that allele 2 is likely the disease causing variant influencing disease susceptibility. Significant associations were observed between the 469+14G/C polymorphism (rs3731865) and autoimmune disease [OR=1.30 (1.04-1.64)] and rheumatoid arthritis [OR=1.60 (1.20-2.13)] and between the -237C/T polymorphism (rs7573065) and inflammatory bowel disease [OR=0.60 (0.43-0.84)]. Further, significant associations were identified between the 469+14G/C, 1730G/A and 1729+55del4 polymorphisms (rs3731865, rs17235409 and rs17235416, respectively) and both infectious disease \textit{per se} and tuberculosis. These findings show a clear association between variants in the \textit{SLC11A1} locus and autoimmune and infectious disease susceptibility.
INTRODUCTION

Solute Carrier Family 11A Member 1 (SLC11A1), formerly NRAMP1, plays an immunomodulatory role in influencing macrophage activation status and the T helper 1/T helper 2 bias. SLC11A1 appears to have multiple functions, playing a role in both the resolution of infections and erythrophagocytosis. Localised to the endosomal/lysosomal compartment of macrophages, SLC11A1 functions as a divalent cation symporter which, when recruited to the phagosomal membrane, transports ions out of the phagosome along the proton gradient. SLC11A1 elicits a range of pleiotropic effects on macrophage function, including increased expression of pro-inflammatory cytokines (interleukin [IL]-1β and tumour necrosis factor [TNF]-α), production of pro-inflammatory effector molecules (increased inducible nitric oxide synthase (iNOS) expression, resulting in increased L-arginine flux, and subsequent production of nitric oxide (NO) and oxidative burst), and modulation of an adaptive immune response (increased MHC Class II expression and enhanced antigen presentation to T cells). How divalent cation transport by SLC11A1 mediates these pleiotropic effects is currently unknown (i.e. a direct effect or secondary result of SLC11A1 activity). However, these pleiotropic effects are essential in the resolution of infection and also in the initiation and perpetuation of Th1 mediated autoimmune diseases.

Due to the immunomodulatory capabilities of SLC11A1, the encoding gene is a strong candidate influencing autoimmune and infectious disease susceptibility. Infectious and autoimmune diseases are complex multi-factorial diseases with multiple genetic (both host and pathogen) and environmental factors playing an aetiological role. An understanding of the host genetic factors involved in these complex diseases will help to develop new preventative and therapeutic strategies. While murine models show a strong correlation between the expression of functional Slc11a1 and both resistance to macrophage-tropic
4 pathogens and susceptibility to autoimmune disease,\textsuperscript{1; 5; 16-18} familial and case control association studies analysing the association of \textit{SLC11A1} variants with disease incidence in humans have produced inconsistent results.

Of the most commonly assessed \textit{SLC11A1} variants, the polymorphic (GT),, microsatellite repeat has been shown to alter the level of \textit{SLC11A1} expression,\textsuperscript{19; 20} and is therefore a strong candidate for influencing disease incidence. Several alleles of different repeat length have been identified, with (GT),, allele 2 conferring lower \textit{SLC11A1} expression compared to the more commonly occurring (GT),, allele 3. It has therefore been hypothesised that allele 3 would provide protection against infectious disease by driving high \textit{SLC11A1} expression and a resultant Th1 mediated immune response. However, allele 3 would also be associated with an increased susceptibility to Th1-mediated autoimmune diseases.\textsuperscript{19} Other \textit{SLC11A1} variants, including the -237C/T promoter and 1730G/A (D543N) polymorphisms, have also been suggested to modulate expression or alter the functional capacity of SLC11A1 to transport divalent cations, respectively.\textsuperscript{20; 21} While several meta-analyses assessing the association of \textit{SLC11A1} polymorphisms with the incidence of tuberculosis [(GT),, repeat, 1730G/A and 2 additional variants]\textsuperscript{22-26} and autoimmune disease [(GT),, repeat only]\textsuperscript{27; 28} have been completed, no study to date has systematically reviewed the literature and completed meta-analyses for all \textit{SLC11A1} polymorphisms (Figure 1). The objective of this study was to systematically review the literature to identify all case-control association studies and where possible complete meta-analyses to determine if \textit{SLC11A1} variants are associated with autoimmune and infectious disease occurrence.

The current meta-analysis was undertaken for a number of reasons. Firstly, there has been a doubling in the number of case control association studies completed since the most current
meta-analysis of the association of the (GT)$_n$ promoter polymorphism with autoimmune disease incidence was completed.\textsuperscript{27,28} Secondly, the current meta-analysis is more inclusive than all other meta-analyses, including all infectious diseases (excluding viruses). Previous meta-analyses have only assessed pulmonary tuberculosis publications.\textsuperscript{22-26} Finally, this meta-analysis assessed a number of polymorphisms within \textit{SLC11A1} for which meta-analyses to determine disease association had not been previously performed due to insufficient numbers of published studies. Specifically, we present novel findings of the association of 17 \textit{SLC11A1} variants with autoimmune and infectious diseases.

Overall, the present study constitutes the largest and most inclusive meta-analysis examining the association of \textit{SLC11A1} polymorphisms with the incidence of infectious and autoimmune diseases conducted to date. Furthermore, based on the findings, inferences about possible functional variants responsible for the identified associations are presented.
RESULTS

A total of 131 case control studies were identified through literature searches and cross-referencing, of which 117 publications were included in the meta-analysis as they assessed the association of \textit{SLC11A1} variants with autoimmune or infectious disease (Figure 2, Table S1-S3). A further 8 publications were excluded from the analysis due to duplicate reporting of identical data. From the 36 identified publications covering autoimmune disease, 11 \textit{SLC11A1} polymorphisms had been investigated in a sufficient number of association studies to allow completion of a meta-analysis (a total of 160 associations) (Table 1). Of the 84 publications investigating infectious disease, 10 \textit{SLC11A1} variants had been examined in a sufficient number of case control studies to perform meta-analyses (274 associations in total) (Table 1).

Associations of the (GT)$_n$ promoter variants with the incidence of autoimmune disease.

Meta-analyses assessing the association of \textit{SLC11A1} (GT)$_n$ alleles 2 and 3 with autoimmune disease (28 datasets) yielded non-significant pooled OR estimates of 0.93 (CI:0.83-1.05) and 1.07 (0.94-1.22) (Table 2, S4 and S5). Analysis of the funnel plots from the meta-analyses did not indicate bias within the datasets. Further analysis of the association of (GT)$_n$ allele 3 with individual autoimmune diseases found a significant association with the incidence of Type 1 diabetes [pooled OR estimates 1.07 (1.01-1.12)] (Table 2). Conversely, the protective effect [OR = 0.94 (CI: 0.89-0.98)] (Table 2). No association was observed between either (GT)$_n$ allele 2 or 3 and the occurrence of, specifically, inflammatory bowel disease, rheumatoid arthritis, multiple sclerosis and sarcoidosis (Table 2). When stratified according to ethnicity, a significant association was observed between both alleles 3 [OR = 1.75 (1.19-2.59)] and 2 [protective effect OR = 0.58 (0.35-0.96)] and autoimmune disease incidence in
the African population, however, similar findings were not observed in either of the Asian, European or Mediterranean populations (Table 3).

125 The 469+14G/C (INT4) variant is significantly associated with the incidence of autoimmune disease and **Rheumatoid Arthritis**

Prior to this study, the (GT)$_n$ promoter polymorphism had been the only $SLC11A1$ genetic variant to be analysed for association with autoimmune disease, as there were insufficient association studies on other $SLC11A1$ variants to enable meta-analyses to be completed.$^{28}$ In addition to the (GT)$_n$ repeat polymorphism (Table 2, S6-S15), we report, for the first time, the results of meta-analyses assessing the associations of 10 additional $SLC11A1$ variants with autoimmune disease. Analysis of the 469+14G/C (INT4) polymorphism identified that the less frequent C variant was significantly associated with the occurrence of autoimmune disease [OR = 1.30 (1.04-1.64)] (Table 2, S8). Surprisingly, the observed association of the C variant with disease occurrence is in opposition to the significant protective effect identified in the large sample size (n = 8787 cases, 10611 controls) of the study of Yang and co-workers.$^{29}$ Re-analysis in the absence of this large study did not alter the observed association [OR = 1.39 (1.24-1.56)]. Further analysis of the 469+14G/C polymorphism identified a significant association between the less frequent C variant and the occurrence of rheumatoid arthritis [OR = 1.60 (1.20-2.13)], but not sarcoidosis (Table 2).

No significant associations were identified between the $SLC11A1$ polymorphisms, -237C/T, 274C/T, 577-18G/A, 823C/T, 1029C/T, 1465-85G/A, 1730G/A, 1729+55del4 and 1729+271del4, and the incidence of autoimmune disease (Table 2). However, while the -237C/T polymorphism was not associated with autoimmune disease as a whole, further analysis of the -237C/T polymorphism found that the less frequent T variant exerted a
putative protective effect over the onset of inflammatory bowel disease (combined Crohn’s
disease and ulcerative colitis) [OR = 0.60 (0.43-0.84)].

Associations of the (GT)$_n$ promoter variants with the incidence of infectious disease

The meta-analyses of the association of (GT)$_n$ alleles 2 and 3 with the incidence of infectious
disease included 19 and 25 datasets, respectively (Table 1, S16 and S17). The meta-analyses
showed that (GT)$_n$ allele 2 was significantly associated with the incidence of infectious
disease [OR = 1.25 (1.10-1.42)], while (GT)$_n$ allele 3 was shown to be protective against the
occurrence of infectious disease [OR = 0.83 (0.74-0.93)] (Table 4). An analysis of the funnel
plots indicated the presence of bias within the datasets (See Table S16 and S17). While the
trim and fill method was previously used to adjust for bias, use of the trim and fill method
in the current analysis was not required, since, if the funnel plots did not show bias (i.e. the
"missing" studies were filled in), they would be located in a position that would strengthen
the pooled OR estimate.

Further analysis of the association of (GT)$_n$ alleles 2 and 3 with the incidence of tuberculosis
alone, revealed a stronger association than those observed with the occurrence of infectious
disease per se, with fixed and random-effects pooled ORs of 1.47 (1.30-1.66) and 0.75 (0.69-
0.82), respectively (Table 4). A meta-analysis assessing the association of (GT)$_n$ allele 2 with
the occurrence of infectious disease or tuberculosis alone has not been completed prior to the
current study. Previous meta-analyses, and case control association studies have focused
primarily on the association of allele 3 with infectious disease, and have not investigated
allele 2 in this context. However, the results of the current meta-analysis show that the
association of (GT)$_n$ allele 2 with the incidence of tuberculosis alone is more significant than
the protective effect putatively exerted by (GT)$_n$ allele 3. No association was identified between (GT)$_n$ allele 3 and the incidence of Leprosy (Table 4).

Stratification of the data based on ethnicity found that (GT)$_n$ allele 2 was significantly associated with infectious disease susceptibility in the African population, with a susceptibility trend that failed to reach significance among the Asian and European populations (Table 3). Furthermore, no association was found in the South American population. Allele 3 was found to be significantly associated with resistance to infectious disease in the African and Asian populations, however no association was found among the European and South American populations (Table 3). While the lack of association of both (GT)$_n$ alleles 2 and 3 with the occurrence of infectious disease in the South American population may be due to the small numbers of publications completed to date (n=2), conflicting results were observed with the association of the (GT)$_n$ alleles with infectious disease in the European population. The results from the European population indicate that allele 2 may be associated with the incidence of infectious disease (OR=1.24), while allele 3 appears to play no role in affording disease protection (OR=1.01), suggesting allele 2 exerts a greater influence over infectious disease susceptibility in the European population, compared to allele 3.

The 469+14G/C, 1730G/A and 1729+55del4 polymorphisms are associated with the incidence of infectious disease

Meta-analyses assessing the association of the 469+14G/C, 1730G/A and 1729+55del4 polymorphisms with the incidence of infectious disease included 47, 54 and 52 datasets, respectively (Table 1, S20, S24 and S25). The meta-analyses revealed that the presence of the less frequent variant for each polymorphism was significantly associated with the incidence
of infectious disease, with random effects pooled OR estimates of 1.27 (1.12-1.43), 1.23
(1.08-1.40) and 1.25 (1.13-1.38) for the 469+14G/C, 1730G/A and 1729+55del4
polymorphisms, respectively (Table 4). Furthermore, analysis of the association of the
469+14G/C, 1730G/A and 1729+55del4 polymorphisms with the incidence of tuberculosis
alone identified a significant association consistent with previous meta-analyses,\textsuperscript{23, 24, 26} with
OR estimates of 1.31, 1.24 and 1.31, respectively (Table 4). Significant heterogeneity, as
determined by the Cochran Q value, was identified within the datasets of the meta-analyses
assessing both infectious disease and tuberculosis alone for all three polymorphisms (Table
4). No association between the occurrence of the 1729+55del4 polymorphism and the
incidence of leprosy was identified (Table 4). No asymmetry was identified in the data from
the analysis of the funnel plots for the 469+14G/C, 1730G/A and 1729+55del4
polymorphisms.

Analysis of the association of the 469+14G/C, 1730G/A and 1729+55del4 polymorphisms
with the occurrence of infectious disease among different ethnicities identified a trend in
which the less frequent variant for each polymorphism was associated with the incidence of
infectious disease (Table 3). In particular, a significant association was identified between
each polymorphism and the incidence of infectious disease in the Asian population. The
469+14C/C and 17291730G/A+55del4 polymorphisms were significantly associated with the
incidence of infectious disease in the African population. However, a protective effect
appeared to be conferred by the less frequent 1730 A variant in the Mediterranean population
(Table 6). However, this analysis incorporated only two publications, suggesting that the
observed association may be largely attributable to random variation.
No significant association was identified between the occurrence of the -237C/T, 274C/T, 577-18G/A, 823C/T, 1485-85G/A and 1729+271del4 polymorphisms and the incidence of infectious disease or tuberculosis alone (Table 4, S18, S19, S21-23 and S26). The association of the -237C/T polymorphism with infectious disease incidence and tuberculosis failed to reach statistical significance and this is likely attributable to the small number of publications that have been completed to date. The results suggest that the -237C/T promoter polymorphism may be associated with the occurrence of infectious disease and tuberculosis, however more association studies are required to confirm such an observation.
The current study aimed to determine the association of genetic variants throughout the 
SLC11A1 locus with the occurrence of infectious and autoimmune disease. This meta-
analysis incorporates the largest number of publications (120 publications, 23 individual 
meta-analyses) and the largest number of SLC11A1 polymorphisms investigated to date, with 
21 SLC11A1 polymorphisms analysed with the occurrence of autoimmune (10/21) and 
infectious (11/21) disease, respectively, 17 of which have not been previously analysed. The 
results of the current meta-analyses have shown that genetic variants throughout SLC11A1 
are associated with the incidence of both infectious and autoimmune disease (Figure 3). Of 
the 17 new SLC11A1 variants assessed, this meta-analysis has identified a significant 
association between the 469+14G/C polymorphism and the incidence of autoimmune disease 
as a whole and rheumatoid arthritis in particular, and the -273C/T polymorphism with the 
occurrence of inflammatory bowel disease. Similar to previous meta-analyses, the current 
analysis did not identify a significant association between either (GT)$_n$ allele 2 or 3 with a 
reduced or increased incidence of autoimmune disease, respectively. However, stratification 
according to disease did identify a significant association with Type 1 diabetes incidence, 
suggesting that the (GT)$_n$ polymorphism may exert a minor effect on some autoimmune 
diseases.

The 469+14G/C, 1730G/A and 1729+55del4 polymorphisms were significantly associated 
with the incidence of infectious disease as a whole and with tuberculosis in particular, with 
pooled OR estimates determined in the current analyses being similar to previously reported 
OR estimates. Similarly, consistent with previous reports, (GT)$_n$ allele 3 was found to be 
significantly protective of infectious disease and tuberculosis, while, for the first time, a
significant association between (GT)$_n$ allele 2 and an increased susceptibility to infectious
disease and tuberculosis was shown to exist.

A meta-analysis assessing the association of (GT)$_n$ allele 2 with the occurrence of infectious
disease or tuberculosis alone has not been completed prior to the current study. Previous
meta-analyses,$^{23; 24; 26}$ and case control association studies, have focused primarily on the
association of allele 3 with infectious disease,$^{30-35}$ and associations of allele 2 with infectious
disease incidence have not been investigated. However, the results of the current meta-
analysis show that the association of (GT)$_n$ allele 2 with the incidence of tuberculosis alone is
more significant than the protective effect putatively exerted by (GT)$_n$ allele 3. This data
suggests that allele 2 may exert a greater influence on the incidence of infectious disease than
the previously thought (GT)$_n$ allele 3.$^{40}$

Reporter studies have shown that different lengths of the (GT)$_n$ promoter microsatellite repeat
alter $SLC11A1$ expression levels, with (GT)$_n$ allele 3 driving higher expression than (GT)$_n$
allele 2. Due to the important role $SLC11A1$ plays in initiating and perpetuating a Th1
immune response, it was hypothesised that over expression of $SLC11A1$, driven by (GT)$_n$ allele 3 would result in a heightened Th1 immune response and a subsequent
"chronic hyperactivation of macrophages" (i.e. classical activation).$^{19; 39}$ This chronic
hyperactivation of macrophages would confer resistance to infectious disease, but also
susceptibility to autoimmune diseases. While the current analysis shows an association
between the (GT)$_n$ alleles and infectious disease (in particular tuberculosis), no association
was evident with autoimmune disease per se, however a minor effect was observed with
Type 1 diabetes.
The current analyses identified that allele 2 of the (GT)$_n$ repeat had a stronger association with tuberculosis susceptibility than the protective effect afforded by allele 3. This was highlighted in the European population, where allele 2 showed a trend for increased susceptibility to tuberculosis, however, allele 3 showed no protective effect. Additionally, the (GT)$_n$ allele 2 dataset was found to be homogenous ($X^2 = 12.23, p = 0.27$), however, heterogeneity was identified within the (GT)$_n$ allele 3 dataset, as well as all other variants associated with tuberculosis (Table 4). It is envisaged that a sequence variant which alters the propensity of an individual to contract an infectious disease like tuberculosis (i.e., the variant provides a selective advantage or disadvantage to the carrier) would be common to all studies irrespective of other factors responsible for heterogeneity (like ethnicity or nutritional status). In such a case, the ORs for the individual studies in the meta-analysis would be expected to be homogenous, as is observed with the meta-analysis examining the association of allele 2 with the incidence of tuberculosis. Therefore, the data suggests that allele 2 may exert a greater influence on the incidence of infectious disease than the previously thought (GT)$_n$ allele 3. Due to this stronger association, we hypothesise that (GT)$_n$ allele 2, and not allele 3, is the disease causing variant at the (GT)$_n$ microsatellite, which exerts the selective pressure at the SLC11A1 locus to influence infectious disease susceptibility.

The question then arises as to how might (GT)$_n$ allele 2 function to alter infectious and autoimmune disease susceptibility? Reporter studies show different SLC11A1 expression levels in the presence of different (GT)$_n$ alleles, with (GT)$_n$ allele 2 driving lower expression than (GT)$_n$ allele 3 (refs). The (GT)$_n$ microsatellite has endogenous transcriptional enhancer activity due to the ability of the repetitive GT units to form Z-DNA. Furthermore, Alleles 2 and 3 which differ by a single 2bp GT repeat are reported to influence transcription through altered transcription factor binding to the SLC11A1 promoter. Specifically, the transcription
factors HIF-1α and ATF-3/JunB have been shown to bind within and adjacent to the (GT)n repeat, respectively.\textsuperscript{36-38} Thus altered transcription factor binding, in the presence of the different repeat lengths may alter SLC11A1 expression to influence macrophage phenotype and susceptibility to infectious and autoimmune disease. Indeed, murine studies show modest reductions in Slc11a1 expression result in significant phenotypic consequences.\textsuperscript{2; 4; 16} suggesting a similar reduction in SLC11A1 promoter activity with (GT)n allele 2 will also result in an altered cellular phenotype to influence disease susceptibility. Consistent with this hypothesis is the observation that allele 2 carriers have increased expression of the anti-inflammatory cytokine IL-10, compared to individuals who do not carry allele 2.\textsuperscript{39} and murine macrophages which lack functional Slc11a1 show higher IL-10 expression after infectious challenge.\textsuperscript{11; 18; 40-43}

Human and murine studies suggest that (GT)n allele 2 may alter disease susceptibility through higher expression of the anti-inflammatory cytokine, IL-10. Macrophages or dendritic cells isolated from mice which lack functional Slc11a1 show higher IL-10 expression after infectious challenge, or induction of a model of autoimmune disease, compared to macrophages/dendritic cells containing functional Slc11a1.\textsuperscript{11; 18; 40-43} While the loss of functional Slc11a1 in the murine model does not correlate with the observed phenotypic changes in SLC11A1 expression occurring with the different (GT)n repeat alleles in humans (i.e. a reduced level of SLC11A1 expression rather than loss of function), a human-based study has identified allele 2 carriers to have increased expression of the anti-inflammatory cytokine IL-10, compared to individuals who do not carry allele 2.\textsuperscript{44} It is therefore hypothesised that allele 2 is the disease causing variant at the (GT)n microsatellite repeat driving low SLC11A1 expression and a subsequent increase in IL-10 expression. The
increased IL-10 expression would produce a heightened anti-inflammatory immune response, inhibiting the production of an adequate Th1 pro-inflammatory immune response.

Specifically, IL-10 has been shown to inhibit innate macrophage anti-microbial molecules involved in a pro-inflammatory immune response and has also been shown to reduce antigen processing, antigen presentation and T-cell activation.\textsuperscript{44-49} Thus, the inhibition of a Th1 pro-inflammatory immune response, in the presence of allele 2 (conferring lower $SLC11A1$ expression), would confer susceptibility to infectious disease, while in the presence of (GT)$_n$ allele 3 an adequate level of $SLC11A1$ expression would exist, high enough to produce a Th1 pro-inflammatory immune response to allow efficient resolution of infectious disease. This could possibly explain why meta-analyses show significant associations between variants at the (GT)$_n$ polymorphism with incidence of infectious disease with only very modest associations with autoimmune disease, specifically Type 1 diabetes. Future work should aim to explore further the role of (GT)$_n$ allele 2 in infectious disease occurrence.

The current meta-analysis identified positive associations between polymorphisms within the 5' region of $SLC11A1$, but not within the 3' region, and the incidence of autoimmune disease, while polymorphisms located in the 5' and 3' regions of $SLC11A1$ were associated with the incidence of infectious disease (Figure 4). Previous publications have identified the existence of significant linkage disequilibrium (LD) between the (GT)$_n$, -237C/T, 274C/T and 469+14G/C variants and markers 110kb upstream of the $SLC11A1$ locus, including the IL8Rb locus (termed 5'LD haplotype end). Furthermore, significant LD has been observed between the 823C/T, 1465-85G/A, 1730G/A and 1729+55del4 variants and markers 110kb downstream of the $SLC11A1$ locus (termed 3'LD haplotype end). However, LD is not
observed between variants located in the 5' and 3' LD haplotype ends of the SLC11A1 locus (Figure 4).44

The SLC11A1 polymorphisms identified to be significantly associated with disease incidence in the current analysis may be the functional cause of the association(s), or, alternatively, the associations observed may be due to the particular polymorphism being either positively or negatively selected because it is in LD with the true disease causing variant. In the latter case, a genetic variant which alters disease incidence provides either a positive or negative selective pressure for the inheritance of all of the neutral variants within that LD block (hitchhiker effect).45 Due to the complex LD pattern which exists at the SLC11A1 locus,44,46-48 the findings suggest that at least one functional polymorphism exists within the 5' LD region of SLC11A1, which alters the cellular phenotype to influence autoimmune disease susceptibility, while at least two functional polymorphisms, one in the 5' region and a second in the 3' region, influence the occurrence of infectious disease (Figure 4). Thus polymorphisms in LD with the significantly associated SLC11A1 polymorphisms should also be considered as potential functional candidates for disease susceptibility. However, the observed associations with infectious and autoimmune disease are most likely mediated by a polymorphism(s) within the SLC11A1 locus given the role SLC11A1 plays in the activation of a Th1 (pro-inflammatory) immune response, and not due to variants located in the LD regions but outside of the SLC11A1 locus.

Of the SLC11A1 variants significantly associated with infectious disease, the (GT)n and the 1730G/A polymorphisms are putative candidates for the alteration of disease incidence observed at the 5' and 3' LD ends, respectively. These two polymorphisms are likely candidates as they have putative functional effects, being able to either influence the level of
SLC11A1 expressed\textsuperscript{19, 20} or alter the ability of SLC11A1 to transport divalent cations,\textsuperscript{21, 24} respectively. These putative functional effects result in an altered phenotype, which may explain the reason for the associations with infectious disease identified in this study.

Of all polymorphisms examined, the 469+14G/C is the only variant to show an association with the incidence of both autoimmune and infectious disease and is therefore another potential disease causing variant within the 5' LD block of SLC11A1. Surprisingly, the C variant was associated with increased risk of developing both infectious and autoimmune disease. The 469+14G/C polymorphism is located in intron 4 of SLC11A1, near an alternatively spliced exon designated 4a, that produces a truncated transcript and non-functional protein. It has been suggested that the 469+14G/C polymorphism may alter the ratio of truncated to functional transcripts (which is normally relatively low at approximately 1:5).\textsuperscript{50} However, Yang and co-workers did not identify any difference in SLC11A1 expression or the ratio of truncated to functional transcripts between differing genotypes of the 469+14G/C polymorphism,\textsuperscript{29} suggesting that the 469+14G/C polymorphism may influence SLC11A1 function through an as yet unidentified mechanism. Further functional tests are required to identify the polymorphic variants that may result in an altered cellular phenotype to influence infectious/autoimmune disease susceptibility.

Future association studies should ideally analyse cases and controls through haplotype analyses, rather than adopting a narrow binomial approach of analysing only single polymorphisms. For example, while the current meta-analyses suggest an association between the (GT)\textsubscript{n} repeat and the incidence of infectious disease, the (GT)\textsubscript{n} repeat does not function independently to alter SLC11A1 expression, as reporter studies show that both the (GT)\textsubscript{n} and -237C/T polymorphisms function synergistically to determine SLC11A1
expression levels. Therefore, association studies which analyse the effect of the (GT)$_n$ repeat and -237C/T polymorphisms independently will not be able to assess the complex interaction that determines the level of SLC11A1 expressed. Additionally, there are other polymorphisms within SLC11A1 that putatively exert phenotypic effects to alter SLC11A1 expression/function (e.g. 1730G/A). Therefore, an individual’s propensity to develop disease would be determined by a summation of the effects of each of the individual polymorphisms within the SLC11A1 locus. Testament to this, association studies which have assessed SLC11A1 haplotypes have identified more robust associations.

Additionally, while some polymorphisms have been assessed in a large number of association studies to allow the completion of a meaningful meta-analysis, there were still insufficient association studies completed for several polymorphisms which showed a trend with disease incidence, however, the pooled OR estimates did not reach significance. It is possible that the existence of more association studies may have allowed statistical significance to be attained. This includes, for example, analyses of the association of the -237C/T and 1029C/T (A318V) polymorphisms with the incidence of tuberculosis and autoimmune disease, respectively. Both of these polymorphisms may exert effects on SLC11A1 expression/function and show a significant trend with disease incidence, but in the absence of sufficient numbers of studies, the existence of significant associations cannot be determined. Furthermore, the current case/control literature has focused solely on the effect of SLC11A1 on pro-inflammatory (M1) macrophages with disease occurrence and it is unclear the effect that SLC11A1 variation may have on M2 macrophages and disease. For example, given the role SLC11A1 plays in erythrophagocytosis, could SLC11A1 variants influence iron homeostasis and anaemia.
The aim of this work was to determine, based on previously published case control association studies, the association of \textit{SLC11A1} polymorphisms with the incidence of infectious and autoimmune disease. Of the 23 datasets covering 11 \textit{SLC11A1} variants, associations were found for 9, with 4 of the 23 datasets investigated showing trends, possibly due to the low numbers of association studies available. Based on the findings of the current meta-analyses, the \textit{SLC11A1} locus appears to play a role in influencing susceptibility to both infectious and autoimmune diseases. The findings of this meta-analysis are significant in helping to determine the multiple host genetic factors involved in complex diseases. Identification of these host genetic factors will help to prevent, control and treat these complex diseases.
MATERIALS AND METHODS

**Literature Search and Inclusion Criteria**

Publications included in the meta-analysis were identified by searching literature databases (PubMed, Medline/Ovid, Chinese National Knowledge Infrastructure (CNKI) and Asia/China on demand) using the search terms “SLC11A1”, “NRAMP1”, “autoimmunity”, “tuberculosis” and “infection”, individually and in combination (from 1996-2012). Additional papers were sourced by cross-referencing original and review publications. Inclusion criteria for the meta-analysis were that studies assessed SLC11A1 polymorphisms in patients diagnosed with a specific autoimmune or infectious disease and used non-familial subjects as controls. Studies analysing cancer, viral infections or pathology due to infection were excluded. Furthermore, all publications included in the meta-analyses had to assess HIV negative cases and controls. When duplicate association studies were encountered, studies published in English or containing the more informative data were included in the analyses.

**Data Collection**

Information regarding the disease studied, the population analysed and the study findings were extracted from all publications meeting the inclusion criteria. Total study numbers (individuals and alleles) and allelic frequencies (numbers and percentages) were also tabulated for all relevant datasets within a publication. When a publication contained several datasets/associations for a single polymorphism, each dataset was assessed as an individual association when the populations/diseases were different between the datasets. Alternatively, data was pooled if the same population/disease was analysed. Allele frequencies were inferred from genotype frequencies when reported. In the few cases where carrier frequencies were reported, the genotype frequencies were first determined and then allele frequencies.
were inferred. Corresponding authors were contacted by email if the information to determine
the odds ratio (OR) was unavailable or if the published data was ambiguous. When
publications assessed specific SLC11A1 polymorphisms, but concluded that an analysis was
not completed due to a low frequency of the less commonly occurring variant, the data was
omitted from the analysis. The data extracted from all publications satisfying the inclusion
criteria for the meta-analysis was reanalysed to ensure that the extracted data was correct.

Statistical Analyses

Statistical analyses were completed using the Rmeta package in the program R. Using the
relevant data sets, the OR and 95% confidence intervals (CI) were determined for each
individual association included in each of the meta-analyses. Associations which contained
zero observations for both cases and controls were excluded from analyses, while the
reciprocal of the opposite treatment size method was used to allow studies with a zero
observation in either case or control groups to be included.

The association of a polymorphism with disease incidence, from the individual associations,
was completed by the determination of the fixed-effects pooled OR estimate (Mantel-
Haenszel method). The Cochran Q test was utilised to determine whether heterogeneity was
present in the analysed data set. If the Cochran Q test identified the presence of heterogeneity
within the dataset, the random-effects pooled OR estimate (DerSimonian-Laird method) was
determined. Funnel plots were assessed to determine the presence of publication bias.

Only polymorphisms that had been investigated in three or more individual association
studies were included in the analysis. Where a large number of datasets were available for a
particular polymorphism, smaller meta-analyses were completed, where possible, analysing
the association of individual diseases (for example Type 1 diabetes, tuberculosis), or
geographical location, with the SLC11A1 polymorphisms. In these cases, analyses were
performed from as many as two association studies.

Although nine alleles of a polymorphic SLC11A1 promoter (GT)_n microsatellite repeat
(rs534448891) have been identified to date, seven of these alleles (alleles 1 and 4-9) occur at
low frequencies. Therefore, association studies have focused on the association of the most
common alleles 2 and 3 with disease occurrence. Meta-analyses of both (GT)_n allele 3 and
allele 2 were completed to determine the association of these alleles with the incidence of
autoimmune and infectious disease. For the analysis of allele 3, the frequency data for alleles
1, 2 and 4-9 were pooled and compared against the frequency of allele 3. Likewise, for the
analysis of allele 2, the frequencies of alleles 1 and 3-9 were pooled and compared against the
frequency of allele 2.
SUPPLEMENTARY INFORMATION

Supplementary information is available at the Genes and Immunity website. Supplementary information includes 26 tables, forest and funnel plots of each meta-analysis and references for all publications.
ACKNOWLEDGEMENTS

The authors wish to thank Jenefer M. Blackwell, Anna Dubaniewicz, A M Graham, Leonardo A Sechi, Lee E Sieswerda, Maria Gazouli, Margje Haverkamp, Linda Wicker, Jennie Yang, Eileen Hoal, Timothy Sterling and Alison Motsinger-Reif for supplying additional population data. Nicholas Archer was funded through an Australian Postgraduate Award. The authors declare no conflicts of interest.

CONFLICT OF INTEREST

The authors declare no conflict of interest.
REFERENCES


FIGURE LEGENDS

Figure 1
Location of SLC11A1 polymorphisms analysed in the meta-analysis. Associations between
the occurrence of these polymorphisms and the incidence of autoimmune and infectious
disease were analysed using meta-analyses. The 15 exons of the gene are shown as black
boxes with their respective numbers. The corresponding scale above indicates the length (kb)
of the gene. The grey boxes indicate the 3’ and 5’ untranslated regions and the introns and
flanking regions are represented by a thin line. The arrows indicate the position of sequence
variants. Below each polymorphism is the reference SNP (rs#) identification number. Genetic
variants shown in italics are those for which meta-analyses have previously been performed.

Figure 2
Results of the search strategy showing the number of case control publications identified and
excluded from the meta-analyses.

Figure 3
Summary of the results from the meta-analyses (pooled OR estimates and 95% CI interval)
assessing the association of the different SLC11A1 polymorphisms with the incidence of
autoimmune disease, infectious disease and tuberculosis alone.

Figure 4
Linkage disequilibrium at the SLC11A1 locus and location of polymorphisms associated with
the incidence of autoimmune and infectious disease. (A) Genomic organisation of SLC11A1
and location of studied sequence variants. The 15 exons of the gene are shown as black boxes
with their respective numbers and the corresponding scale above indicates the length (kb) of
the gene. The grey boxes indicate the 3’ and 5’ untranslated regions and the introns and
flanking regions are represented by a thin line. The arrows indicate the position of sequence
variants. (B) LD located within the SLC11A1 locus. The blue circles indicate the location of
the SLC11A1 polymorphisms, with the thin line representing the flanking DNA regions. The
two LD blocks (termed 5’ LD haplotype end and 3’ LD haplotype end) are shown, with the
double dashed line designating the weak LD observed between 5’ and 3’ SLC11A1 regions.
(C) Polymorphisms within the 5’ LD haplotype end but not the 3’ end are associated with the
incidence of autoimmune disease (red circles indicate an association, while white circles
indicate no association). (D) Polymorphisms in both the 5’ and 3’ LD haplotype blocks were
found to be associated with infectious disease. The (GT)n and 469+14G/C; and 1730G/A are
candidate polymorphisms in the SLC11A1 locus influencing autoimmune and infectious
disease susceptibility at the 5’ and 3’ LD haplotype ends, respectively (arrows).
### Table 1: Summary of Identified Publications, Datasets Analysed and Numbers of Cases and Controls.

<table>
<thead>
<tr>
<th>Polymorphism</th>
<th>Autoimmune Disease</th>
<th>Infectious Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Publications¹</td>
<td>Datasets¹</td>
</tr>
<tr>
<td>$(GT)_n$ allele 3</td>
<td>29</td>
<td>31</td>
</tr>
<tr>
<td>$(GT)_n$ allele 2</td>
<td>29</td>
<td>31</td>
</tr>
<tr>
<td>-237C/T</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>274C/T</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>469+14G/C</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>577-18G/A</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>823C/T</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>1029C/T</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>1465-85G/A</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>1730G/A</td>
<td>17</td>
<td>16</td>
</tr>
<tr>
<td>1729+55del4</td>
<td>18</td>
<td>17</td>
</tr>
<tr>
<td>1729+271del4</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

¹ Total number of published studies identified from the literature search meeting the inclusion criteria of the meta-analysis.
² Total number of datasets from the identified publications for inclusion into the meta-analysis.
³ The number of datasets analysed in the meta-analysis after the removal of datasets containing zero observations for both cases and controls and when data to determine OR was not forthcoming from corresponding authors.
Table 2: Pooled OR estimates of the association of SLC11A1 polymorphisms with the incidence of autoimmune disease.

<table>
<thead>
<tr>
<th>Polymorphism Association</th>
<th>Test of heterogeneity X² (P-value)</th>
<th>Pooled OR estimate (CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(GT)n allele 3</td>
<td>93.77 (p &lt; 0.01)</td>
<td>1.07 (0.94-1.22)²</td>
</tr>
<tr>
<td>IBD</td>
<td>17.71 (p = 0.01)</td>
<td>1.05 (0.81-1.37)²</td>
</tr>
<tr>
<td>MS</td>
<td>12.80 (p &lt; 0.01)</td>
<td>1.22 (0.80-1.85)²</td>
</tr>
<tr>
<td>RA</td>
<td>18.08 (p &lt; 0.01)</td>
<td>1.06 (0.75-1.51)²</td>
</tr>
<tr>
<td>SA</td>
<td>30.06 (p &lt; 0.01)</td>
<td>1.16 (0.59-2.28)²</td>
</tr>
<tr>
<td>T1D</td>
<td>1.68 (p = 0.64)</td>
<td>1.07 (1.01-1.12)¹</td>
</tr>
<tr>
<td>(GT)n allele 2</td>
<td>73.35 (p &lt; 0.01)</td>
<td>0.93 (0.83-1.05)²</td>
</tr>
<tr>
<td>IBD</td>
<td>4.70 (p = 0.70)</td>
<td>0.91 (0.78-1.06)</td>
</tr>
<tr>
<td>MS</td>
<td>14.58 (p &lt; 0.01)</td>
<td>0.84 (0.53-1.33)²</td>
</tr>
<tr>
<td>RA</td>
<td>15.48 (p &lt; 0.01)</td>
<td>0.91 (0.65-1.26)²</td>
</tr>
<tr>
<td>SA</td>
<td>24.43 (p &lt; 0.01)</td>
<td>0.96 (0.52-1.80)²</td>
</tr>
<tr>
<td>T1D</td>
<td>3.91 (p = 0.27)</td>
<td>0.94 (0.89-0.98)¹</td>
</tr>
<tr>
<td>-237C/T</td>
<td>12.43 (p = 0.13)</td>
<td>0.92 (0.83-1.02)</td>
</tr>
<tr>
<td>IBD</td>
<td>5.82 (p = 0.32)</td>
<td>0.60 (0.43-0.84)¹</td>
</tr>
<tr>
<td>274C/T</td>
<td>18.41 (p = 0.01)</td>
<td>1.16 (0.96-1.40)²</td>
</tr>
<tr>
<td>469+14G/C</td>
<td>86.50 (p &lt; 0.01)</td>
<td>1.30 (1.04-1.64)¹²</td>
</tr>
<tr>
<td>RA</td>
<td>1.82 (p = 0.61)</td>
<td>1.60 (1.20-2.13)¹</td>
</tr>
<tr>
<td>SA</td>
<td>21.17 (p &lt; 0.01)</td>
<td>1.07 (0.53-2.18)²</td>
</tr>
<tr>
<td>577-18G/A</td>
<td>2.87 (p = 0.58)</td>
<td>0.74 (0.50-1.09)</td>
</tr>
<tr>
<td>823C/T</td>
<td>23.71 (p &lt; 0.01)</td>
<td>1.02 (0.67-1.56)²</td>
</tr>
<tr>
<td>1029C/T</td>
<td>1.57 (p = 0.67)</td>
<td>0.48 (0.21-1.11)</td>
</tr>
<tr>
<td>1465-85G/A</td>
<td>10.98 (p = 0.14)</td>
<td>0.98 (0.93-1.03)</td>
</tr>
<tr>
<td>1730G/A</td>
<td>46.45 (p &lt; 0.01)</td>
<td>1.14 (0.86-1.51)²</td>
</tr>
<tr>
<td>RA</td>
<td>14.45 (p &lt; 0.01)</td>
<td>1.29 (0.62-2.68)²</td>
</tr>
<tr>
<td>1729+55del4</td>
<td>34.75 (p &lt; 0.01)</td>
<td>1.21 (0.96-1.54)²</td>
</tr>
<tr>
<td>RA</td>
<td>19.09 (p &lt; 0.01)</td>
<td>1.52 (0.67-3.44)²</td>
</tr>
<tr>
<td>1729+271del4</td>
<td>1.79 (p = 0.41)</td>
<td>0.98 (0.80-1.22)</td>
</tr>
</tbody>
</table>

IBD: inflammatory bowel disease, MS: multiple sclerosis, RA: rheumatoid arthritis, SA: sarcoidosis, T1D: Type 1 diabetes.

¹ Statistically significant (p < 0.05, bold).
² Random-effects pooled OR estimate.
### Table 3: Pooled OR estimates of the association of *SLC11A1* variants and disease occurrence stratified by ethnicity.

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>Autoimmune disease</th>
<th>Infectious disease</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(GT)_n allele 3</td>
<td>(GT)_n allele 2</td>
</tr>
<tr>
<td>African</td>
<td>1.75 (1.19-2.59)</td>
<td>0.58 (0.35-0.96)</td>
</tr>
<tr>
<td>Asian</td>
<td>0.85 (0.69-1.03)</td>
<td>0.86 (0.68-1.09)</td>
</tr>
<tr>
<td>European</td>
<td>1.17 (0.97-1.42)</td>
<td>0.84 (0.70-1.01)</td>
</tr>
<tr>
<td>Mediterranean</td>
<td>0.97 (0.74-1.30)</td>
<td>1.14 (0.89-1.45)</td>
</tr>
</tbody>
</table>

1 Statistically significant (*p* < 0.05, bold).
2 Random-effects pooled OR estimate.
Table 4: Pooled OR estimates of the association of *SLC11A1* polymorphisms with the incidence of infectious disease.

<table>
<thead>
<tr>
<th>Polymorphism Association</th>
<th>Test of heterogeneity $\chi^2$ (P-value)</th>
<th>Pooled OR estimate (CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(GT)$_n$ allele 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mycobacterium spp.</td>
<td>61.93 ($p &lt; 0.01$)</td>
<td>0.83 (0.74-0.93)$^{1,2}$</td>
</tr>
<tr>
<td>Tuberculosis</td>
<td>58.73 ($p &lt; 0.01$)</td>
<td>0.82 (0.71-0.95)$^{1,2}$</td>
</tr>
<tr>
<td>Leprosy</td>
<td>40.54 ($p &lt; 0.01$)</td>
<td>0.76 (0.65-0.89)$^{1,2}$</td>
</tr>
<tr>
<td>(GT)$_n$ allele 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mycobacterium spp.</td>
<td>30.77 ($p = 0.03$)</td>
<td>1.25 (1.10-1.42)$^1$</td>
</tr>
<tr>
<td>Tuberculosis</td>
<td>20.80 ($p = 0.07$)</td>
<td>1.37 (1.23-1.53)$^1$</td>
</tr>
<tr>
<td>-237C/T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tuberculosis</td>
<td>6.33 ($p = 0.28$)</td>
<td>1.03 (0.83-1.29)</td>
</tr>
<tr>
<td>274C/T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tuberculosis</td>
<td>17.33 ($p = 0.18$)</td>
<td>1.01 (0.92-1.11)</td>
</tr>
<tr>
<td>469+14G/C</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mycobacterium spp.</td>
<td>115.8 ($p &lt; 0.01$)</td>
<td>1.27 (1.12-1.43)$^{1,2}$</td>
</tr>
<tr>
<td>Tuberculosis</td>
<td>109.16 ($p &lt; 0.01$)</td>
<td>1.30 (1.13-1.49)$^{1,2}$</td>
</tr>
<tr>
<td>Leprosy</td>
<td>21.17 ($p &lt; 0.01$)</td>
<td>1.31 (1.12-1.54)$^{1,2}$</td>
</tr>
<tr>
<td>577-18G/A$^3$</td>
<td>1.28 ($p = 0.53$)</td>
<td>0.96 (0.60-1.55)</td>
</tr>
<tr>
<td>823C/T$^3$</td>
<td>7.63 ($p = 0.02$)</td>
<td>0.67 (0.29-1.53)$^2$</td>
</tr>
<tr>
<td>1465-85G/A</td>
<td>3.40 ($p = 0.76$)</td>
<td>1.00 (0.90-1.11)</td>
</tr>
<tr>
<td>1730G/A</td>
<td>2.85 (P=0.28)</td>
<td>1.05 (0.88-1.26)</td>
</tr>
<tr>
<td>1729+55del4</td>
<td>128.81 ($p &lt; 0.01$)</td>
<td>1.23 (1.08-1.40)$^{1,2}$</td>
</tr>
<tr>
<td>Mycobacterium spp.</td>
<td>125.59 ($p &lt; 0.01$)</td>
<td>1.26 (1.09-1.46)$^{1,2}$</td>
</tr>
<tr>
<td>Tuberculosis</td>
<td>121.93 ($p &lt; 0.01$)</td>
<td>1.24 (1.07-1.44)$^{1,2}$</td>
</tr>
<tr>
<td>1729+271del4</td>
<td>112.86 ($p &lt; 0.01$)</td>
<td>1.25 (1.13-1.38)$^{1,2}$</td>
</tr>
<tr>
<td>Mycobacterium spp.</td>
<td>109.47 ($p &lt; 0.01$)</td>
<td>1.27 (1.14-1.41)$^{1,2}$</td>
</tr>
<tr>
<td>Tuberculosis</td>
<td>79.11 ($p &lt; 0.01$)</td>
<td>1.31 (1.18-1.46)$^{1,2}$</td>
</tr>
<tr>
<td>Leprosy</td>
<td>1.63 ($p = 0.80$)</td>
<td>1.06 (0.89-1.26)</td>
</tr>
<tr>
<td>1729+271del4</td>
<td>4.53 ($p = 0.61$)</td>
<td>1.00 (0.91-1.11)</td>
</tr>
<tr>
<td>Tuberculosis</td>
<td>2.12 ($p = 0.71$)</td>
<td>1.02 (0.87-1.19)</td>
</tr>
</tbody>
</table>

$^1$ Statistically significant ($p < 0.05$, bold).
$^2$ Random-effects pooled OR estimate.
$^3$ Publications only analyse tuberculosis.