GENETIC SUSCEPTIBILITY TO LEPROSY

Andrea Alter

Division of Experimental Medicine
McGill University, Montreal

Submitted June 2010

A thesis submitted to McGill University in partial fulfilment of the requirements of the degree of Doctor of Philosophy

To my mother.

Table of contents

Acknowledgements 7

Abstract 9

Résumé 11

Publications 13

Contributions of authors 14

Doctoral thesis objectives 16

Chapter 1 │ Introduction 17

PART 1 – AN OVERVIEW OF LEPROSY 18

1.1 Etymology 18

1.2 Hansen's disease 18

1.3 Koch's postulates and leprosy 19

1.4 Modern genomics, ancient origins 21

1.5 Epidemiology: an eliminated disease? 23

1.6 Transmission theories 25

1.7 Classification of the leprosy disease spectrum 26

1.8 The host immune response to *Mycobacterium leprae*: a Th1-Th2 paradigm 28

1.8.1 Tuberculoid leprosy: inside the granuloma 28

1.8.2 Lepromatous leprosy: a defective *Mycobacterium leprae*-specific T-cell response 29

1.9 Leprosy complications 30

1.9.1 Reversal reaction (type 1 reaction) 30

1.9.2 Erythema nodosum leprosum (type 2 reaction) 31

1.10 Step-wise infection of Schwann cells by *Mycobacterium leprae* 32

1.11 Diagnosis and multi-drug therapy 34

PART 2 – GENETIC MAPPING OF COMPLEX BINARY TRAITS 37

2.1 Complex traits 37

2.2 Uncovering a genetic component 37

2.3 Linkage analysis 38

2.3.1 Model-based (classical lod-score method) 38

2.3.2. Model-free (affected sib-pair method) 39

2.3.3. Applications and limitations of linkage analysis 40

2.4 Association studies 41

2.4.1 Population- and family-based association studies 41

2.4.2 Applications of association studies 42

2.4.3 Interpretation of a positive association 43

2.5 Linkage disequilibrium 44

2.5.1 Measures of linkage disequilibrium: D and r^2 45

2.5.2 Applications of linkage disequilibrium 46

2.6 Summary 47

PART 3 – GENETIC SUSCEPTIBILITY TO LEPROSY 48

3.1 Human genetics of infectious diseases 48

3.2 Uncovering a genetic component in leprosy: twin studies and complex segregation analyses 51

3.3 Genome-wide linkage analyses: from linked regions to associated SNPs 52

3.4 A genome-wide association study 56

3.5 Candidate chromosome region approach 57

3.6 Candidate gene approach 57

3.6.1 HLA complex genes 57

3.6.1.1 Classical HLA class I genes 58

3.6.1.2 Classical HLA class II genes 59

3.6.1.3 Tumor necrosis factor (*TNF*) 63

3.6.2 Non-HLA genes 67

3.6.2.1 Solute carrier family 11, member 1 (*SLC11A1*, formerly *NRAMP1*) 67
3.6.2.2 Vitamin D (1,25- dihydroxyvitamin D3) receptor (*VDR*) 70
3.6.2.3 Interleukin 10 (*IL10*) 72
3.6.2.4 Toll-like receptors (TLRs) 75
3.7 Additional genes 80
3.8 Genetic susceptibility to leprosy: a two-stage model 84
Chapter 2 │ Stepwise replication identifies a low-producing lymphotoxin-α allele as a major risk factor for early-onset leprosy 85
Chapter 3 │ Two HLA class I region SNPs are associated with leprosy susceptibility in Vietnam and India 118
Chapter 4 │ Genetic and functional analysis of common *MRC1* exon 7 polymorphisms in leprosy susceptibility 151
Chapter 5 │ Discussion 191
PART 1 – TOWARDS UNDERSTANDING THE COMPLEX HUMAN-*M. leprae* INTERACTION 192
1.1 Causative versus risk factors 192
1.2 Lymphotoxin-alpha and mycobacterial infection 196
1.3 Mannose receptor and mycobacterial infection 198
1.4 HLA class I region SNPs 200
1.5 Common disease common variant hypothesis 201
PART 2 – LEPROSY AS A GENETIC MODEL FOR SUSCEPTIBILITY TO COMMON INFECTIOUS DISEASES 203
2.1 Refining the methodological approach to genetic mapping of common infectious diseases 203
2.2 Single linkage peak – multiple associated SNPs 205
2.3 Leprosy: a good disease to study? 206
CONCLUDING STATEMENT 208

References (Chapter 1, Introduction and Chapter 5, Discussion) 210

Appendix i │ Manuscripts 239

Acknowledgements

The non-scientists:

My mother and father – Maria and Amos Alter – for your unconditional support through grades pre-K to 24.

My husband – Mikey Gutfeld – for accepting that *"I'm almost finished"* in scientific research means 3-5 years, for allowing me to sleep an extra hour every morning and for my NESPRESSO® machine.

My son – Eli Gutfeld – for reminding me what is important (you).

Eli's nanny – Nonette Gamboa – for caring for my son every day as though he were your own.

The 4PM – 6PM babysitters – my mother, Elayn Farr, John-Patrick Mancini, Sharon Alter and Arie Alter – for giving me two extra hours every day to write Chapter 1.

The scientists:

The Natural Sciences and Engineering Research Council of Canada for two consecutive post-graduate scholarships.

The L1-413 group past and present – Dr. Caroline Gallant, Manon Girard, Dr. Louis de Léséleuc (also for translating my abstract into French), Dr. Marcelo Mira, Dr. Marianna Orlova, Tania di Pietrantonio, Dr. Audrey Poon and Annie Verville – for your technical assistance and creating a pleasant work environment.

The I.T. department – Leah Simkin – for your unrelenting efforts to facilitate data management and data analysis.

The sequencing and genotyping teams at the McGill University and Genome Québec Innovation Centre – particularly Dr. Pierre Lepage (Director of the Sequencing Platform), Dr. Alexandre Montpetit (Director of the Genotyping Platform) and Alexandre Bélisle (genotyping technician) – for your invaluable service to this project.

My PhD committee members past and present – Dr. Thomas Hudson, Dr. Danielle Malo and Dr. Robert Sladek – for your time and constructive suggestions.

My graduate academic advisor – Dr. Patricia Tonin – for your efforts to keep me on track and encourage me to continue when I was ready to hang up my lab coat.

My "unofficial" supervisor – Dr. Alexandre Alcaïs – for your fantastic sense of humour, incessant optimism, calming nature and for not blocking my email address (and your statistical expertise).

My official supervisor – Dr. Erwin Schurr – for the education, the experience and eliciting your passion for science in me, thank you.

Abstract

Leprosy (Hansen's disease) is a human infectious disease that can be effectively treated with 6-12 month administration of multi-drug therapy. In 2008, approximately 250,000 new cases were reported to the World Health Organization. The etiological agent, *Mycobacterium leprae*, was identified by G.H.A. Hansen in the nineteenth century. Subsequent to exposure, epidemiological studies maintain the importance of host genetics in leprosy susceptibility. A previous model-free genome-wide linkage scan in multi-case families from Vietnam detected linkage between chromosome region 6p21 (lod = 2.62) and leprosy *per se* and between 10p13 (lod = 1.98) and the paucibacillary sub-type. A 'first-round' high-density association scan (307 SNPs) of a 10.4 Mb target interval on 6p21 (*BAT1→CCND3*) in 194 single-case families from Vietnam led to the identification of a SNP in lymphotoxin-alpha (*LTA*) – *LTA*+80 – as major risk factor for early-onset leprosy ($P = 4.0\times10^{-5}$). The association was replicated in 364 cases and 371 controls from North India ($P = 0.006$) and 104 single-case families from Vietnam (combined Vietnam $P = 4.0\times10^{-7}$). A 'second-round' ultra-high-density association scan (682 SNPs) of a 1.9 Mb sub-interval on 6p21 (*HCG27→HLA-DPA3*) in 198 single-case families from Vietnam led to the identification of two intergenic SNPs in the HLA class I region – rs2394885 ($P = 0.0063$) and rs2922997 ($P = 0.0094$) – as risk factors for leprosy. The associations were replicated in 292 single-case leprosy families from Vietnam ($P = 8.8\times10^{-5}$ and $P = 0.0037$, respectively) and the population-based sample from North India ($P = 3.0\times10^{-8}$ and $P = 2.0\times10^{-5}$, respectively). Finally, mannose receptor, C type 1 (*MRC1*) was selected as a candidate paucibacillary gene in the 10p13 region. In 490 single-case and 90 multi-case families from Vietnam a non-synonymous SNP in exon 7 – rs1926736 (G396S) – was associated with leprosy ($P = 0.035$) but

not the paucibacillary sub-type. The association was replicated in 384 cases and 399 controls from Rio de Janiero, Brazil ($P = 0.016$). The process of identifying the susceptibility variants provided valuable insight into the replication of genetic effects. As such, these studies served to improve our understanding of leprosy pathogenesis by implicating novel biological pathways while simultaneously providing a genetic model for common infectious diseases.

Résumé

La lèpre (maladie de Hansen) est une maladie infectieuse des humains qui peut être traitée avec succès par une antibiothérapie d'une durée de 6-12 mois. En 2008, environ 250 000 nouveaux cas ont été rapportés à l'Organisation Mondiale de la Santé. L'agent étiologique, *Mycobacterium leprae*, fut identifié par G.H.A. Hansen au XIXe siècle. Suite à l'exposition, des études épidémiologiques continuent de souligner l'importance de la génétique de l'hôte dans la susceptibilité à la lèpre. Une analyse génomique de liaison accomplie préalablement à partir de familles Vietnamiennes à cas multiples a détecté une liaison entre la région chromosomique 6p21 (lod = 2,62) et la lèpre *per se* et entre 10p13 (lod = 1,98) et le sous-type paucibacillaire. Une analyse primaire d'association de haute densité (307 SNP) d'une intervalle cible de 10,4 Mb dans le 6p21 (*BAT1→CCND3*) chez 192 familles à cas unique a mené à l'identification d'un SNP dans le gène de la lymphotoxine alpha (*LTA*) – *LTA*+80 – un facteur de risque majeur de la lèpre à début précoce ($P = 4{,}0\times10^{-5}$). L'association a été répliquée chez 364 cas et 371 témoins du nord de l'Inde ($P = 0{,}006$) et 104 famille Vietnamiennes à cas unique ($P = 4.0\times10^{-7}$ en combinant tout le Vietnam). Une analyse secondaire d'association à ultra-haute densité (682 SNP) d'une sous-intervalle de 1,9 Mb dans 6p21 (*HCG27→HLA-DPA3*) chez 198 familles Vietnamiennes à cas unique a permis l'identification de deux SNP intergéniques dans la région HLA de classe I – rs2394885 ($P = 0{,}0063$) et rs2922997 ($P = 0{,}0094$) – en tant que facteurs de risque à la lèpre. Les associations ont été répliquées chez 292 familles Vietnamiennes à cas unique ($P = 8{,}8\times10^{-5}$ et $P = 0{,}0037$, respectivement) et chez l'échantillon de population du nord de l'Inde ($P = 3{,}0\times10^{-8}$ et $P = 2{,}0\times10$-5, respectivement). Finalement, le récepteur du mannose, C type 1 (*MRC1*) a été sélectionné comme gène candidat de susceptibilité à la lèpre paucibacillaire dans la région

10p13. Chez 490 familles Vietnamiennes à cas uniques et 90 à cas multiples, un SNP non synonyme dans l'exon 7 – rs1926736 (G396S) – fut associé à la lèpre (P = 0,035) mais non au sous-type paucibacillaire. L'association fut répliquée chez 384 cas et 399 témoins provenant de Rio de Janeiro, Brésil (P = 0,016). Le processus d'identification des variantes de susceptibilité a enrichi notre compréhension de la réplication des effets génétiques. Ainsi, ces études ont permis d'améliorer nos connaissances de la pathogénie de la lèpre en révélant de nouvelles voies biologiques tout en fournissant un modèle génétique de maladies infectieuses communes.

Publications

Chapters 1 and 5 (select portions)

Alter A, Alcaïs A, Abel L, Schurr E. **Leprosy as a genetic model for susceptibility to common infectious diseases.** Hum. Genet. 2008 Apr;123(3):227-35.

Chapter 2

Alcaïs A, Alter A, Antoni G, Orlova M, Nguyen VT, Singh M, Vanderborght PR, Katoch K, Mira MT, Vu HT, Ngyuen TH, Nguyen NB, Moraes M, Mehra N, Schurr E, Abel L. **Stepwise replication identifies a low-producing lymphotoxin-alpha allele as a major risk factor for early-onset leprosy.** Nat. Genet. 2007 Apr;39(4):517-22.

Chapter 3

Alter A, Huong NT, Singh M, Orlova M, Van Thuc N, Katoch K, Thai VH, Ba NN, Abel L, Mehra N, Alcaïs A, Schurr E. **Two HLA class I region SNPs are associated with leprosy susceptibility in Vietnam and India.** (Submitted for publication).

Chapter 4

Alter A, de Léséleuc L, Van Thuc N, Thai VH, Huong NT, Ba NN, Cardoso CC, Grant AV, Abel L, Moraes MO, Alcaïs A, Schurr E. **Genetic and functional analysis of common *MRC1* exon 7 polymorphisms in leprosy susceptibility.** Hum. Genet. 2009 Dec 25. [Epub ahead of print]

Contributions of authors

The work presented in Chapters 1-5 was done in collaboration with co-authors as follows:

Chapters 1 and 5

A. Alter, A. Alcaïs, L. Abel and E. Schurr contributed to the writing of the manuscript.

Chapter 2

SNP selection was done by A. Alcaïs, A. Alter and M.T. Mira; genotyping was performed by technicians at the McGill University and Genome Québec Innovation Centre, A. Alter (Vietnam-1, India and Brazil samples), M. Orlova (Vietnam-2 sample), M. Singh (*HLA-DRB1* in India sample), and M.T. Mira (Vietnam-1 sample); statistical analyses were performed by A. Alcaïs, A. Alter and G. Antoni; sample ascertainment – including diagnosis, phenotype assessment and DNA extraction – was a collective effort by M. Orlova, N.V. Thuc, V.H. Thai., N. Thu Huong and N.N. Ba (Vietnam-1 and -2 samples), K. Katoch (India sample) and P.R. Vanderborght and M. Moraes (Brazil sample); N. Mehra organized the Indian component of the study; A. Alcaïs, A. Alter, E. Schurr and L. Abel contributed to the writing of the manuscript; A. Alcaïs, A. Alter and G. Antoni share first authorship (listed alphabetically); E. Schurr and L. Abel share senior authorship.

Chapter 3

SNP selection was done by A. Alter and A. Alcaïs; genotyping was performed by technicians at the McGill University and Genome Québec Innovation Centre; M. Singh (*HLA-DRB1* in India sample); statistical analyses were performed by A. Alter and A. Alcaïs; sample ascertainment –

including diagnosis, phenotype assessment and DNA extraction – was a collective effort by N. Thu Huong, M. Orlova, N.V. Thuc, V.H. Thai. and N.N. Ba (Vietnam-1 and -2 samples) and K. Katoch (India sample); N. Mehra organized the Indian component of the study; A. Alter, L. Abel, A. Alcaïs and E. Schurr contributed to the writing of the manuscript; A. Alcaïs and E. Schurr share senior authorship.

Chapter 4

SNP selection was done by A. Alter; genotyping and sequencing was performed by technicians at the McGill University and Genome Québec Innovation Centre and A. Alter; statistical analyses were performed by A. Alter, A.V. Grant and A. Alcaïs; sample ascertainment – including diagnosis, phenotype assessment and DNA extraction – was a collective effort by N.V. Thuc, V.H. Thai, N. Thu Huong and N.N. Ba (Vietnam sample) and C.C. Cardoso and M.O. Moraes (Brazil sample); L. de Léséleuc performed all functional assays; A. Alter, L. de Léséleuc, L. Abel, A. Alcaïs and E. Schurr contributed to the writing of the manuscript; A. Alter and L. de Léséleuc share first authorship.

Doctoral Thesis Objectives

1. Identify and replicate the host genetic risk (or protective) factor(s) for leprosy *per se* (i.e., irrespective of clinical sub-type) and/or sub-type (i.e., multibacillary or paucibacillary) that underlies the previously reported linkage peak on chromosome region 6p21 (i.e., the HLA region).

2. Identify and replicate the host genetic risk (or protective) factor(s) for leprosy *per se* (i.e., irrespective of clinical sub-type) and/or sub-type (i.e., multibacillary or paucibacillary) that underlies the previously reported linkage peak on chromosome region 10p13.

Chapter 1 | Introduction

PART 1 – AN OVERVIEW OF LEPROSY

1.1 Etymology

Leper: from Greek *lepra* noun use of *lepros* (scaly) {Collins English Dictionary 2003}

1.2 Hansen's disease

The prevalence of leprosy in Norway in the mid-19th century was two cases per thousand, in the city of Bergen the prevalence was 25 cases per thousand. It was here that Daniel Cornelius Danielssen (1815-1894), physician-in-chief of the St. Jørgen and Lungegaarden leprosy hospitals, collaborated with Carl Wilhelm Boeck (1808-1875) and published the influential book *On Leprosy* in 1847. They concluded that leprosy was a "hereditary dyscrasia". At 27, Gerhard Henrik Armauer Hansen (1841-1912) was appointed to both hospitals. A young and inexperienced physician, Hansen boldly challenged the theory held by his superiors, stating to Danielssen (his future father-in-law): *"...your opinions about leprosy are completely wrong. You believe that the disease is hereditary but not infectious. The truth is that it is infectious but not hereditary."* In 1873, after two years of examining leprosy families in western Norway, Hansen focused on the microscopic investigation of "brown bodies" from leprous nodules (a term coined by Danielssen and Boeck). His seminal discovery of bacilli or *"...staff-like bodies, much resembling bacteria..."* substantiated his idea of a bacterial etiology for leprosy {Hansen 1875}. The observation earned him the position of Chief Medical Officer for Leprosy in Norway in 1875, but he, and others, struggled to prove causality leaving Danielssen unconvinced of the infectious nature of leprosy. One attempt by Hansen to demonstrate disease transmissibility ended with a criminal charge. On November 3, 1879, Hansen inoculated the eye of an un-consenting 33-year-old female with material from a leprous nodule. No disease ensued but

Hansen was convicted of misusing his official position and relieved of his physician duties. Hansen retained his position as Chief Medical Officer and wrote the 'Norwegian Leprosy Act' (first law passed in 1877) mandating patient isolation based on his yet unproven theory of leprosy transmission {Marmor 2002}. Owing to his outstanding scientific discovery, leprosy is also called Hansen's disease.

1.3 Koch's postulates and leprosy

In 1890, Hansen's contemporary Robert Koch (1843-1910) stated his four famous criteria to establish the microbial etiology of a disease. Ironically, despite being the first bacterium identified to cause a human disease, *Mycobacterium leprae* (*M. leprae*) still does not satisfy all of Koch's postulates:

1. A specific microorganism can always be found associated with a given disease.

Microscopically detectable *M. leprae* – an acid-fast bacillus – in slit-skin smears is a cardinal sign of leprosy (i.e., ~100% specific) but is not sensitive as merely 10-50% of leprosy patients (multibacillary cases) are positive {International Leprosy Association 2002}. In contrast, molecular detection of *M. leprae* in nasal swabs (PCR amplification of species-specific *pra* gene) is not specific. In two leprosy endemic Indonesian villages, 7.8% of the population were PCR+, compared to 0.99% who were diagnosed with leprosy, indicative of asymptomatic carriers {Klatser, van Beers et al. 1993}. Nasal carriage was also transient – without ensuing disease – as 90 PCR+ individuals were PCR- two years later {Hatta, van Beers et al. 1995}.

2. The microorganism can be isolated and grown in pure culture in the laboratory.

A history of failed attempts to culture *M. leprae in vitro* likely reflects its extreme reductive evolution – this obligate intracellular parasite has retained the minimal gene set required for survival in its natural host(s). The circular genome of *M. leprae* (TN strain, Tamil Nadu, India) was sequenced in 2001 (Leproma, http://genolist.pasteur.fr/Leproma/) {Cole, Eiglmeier et al. 2001; Jones, Moszer et al. 2001}. Comparative analysis with the genome and proteome of *Mycobacterium tuberculosis* (*M. tuberculosis*) revealed extensive gene loss across all functional categories. Merely 49.5% of the 3.3 megabase (Mb) *M. leprae* genome is protein coding (1,604 genes) in contrast to 90.8% of the 4.4 Mb *M. tuberculosis* genome (3,959 genes). Nevertheless, it was discovered that *Dasypus novemcinctus* (nine-banded armadillo) {Kirchheimer and Storrs 1971} and footpads of the athymic Nude mouse ($Foxn1^{nu}/Foxn1^{nu}$) are susceptible to *M. leprae* infection. The armadillo and T-cell deficient mouse are invaluable to leprosy research, serving as experimental models of leprosy and biological sources of abundant *M. leprae*.

3. The pure culture of the microorganism will produce the disease when injected into a susceptible animal.

A century after the discovery of *M. leprae* – and long after the deaths of Hansen and Danielssen – data emerged to reconcile their scientific differences. In 1960, Charles C. Shepard was able to reproducibly induce granulomas containing acid-fast bacilli in the foot-pads of mice after inoculation with bacteria harvested from nasal passages (22/22 “takes”) or biopsies (12/16 “takes”) of human leprosy cases {Shepard 1960}.

4. It is possible to recover the injected microorganism from the experimentally infected animal.

Shepard observed the same histopathology after injection of passage material from the foot-pads demonstrating transmissibility of the granulomatous phenotype {Shepard 1960}.

1.4 Modern genomics, ancient origins

Microbial phylogeography – historical geographic distribution of microbes – has emerged from technological advances in comparative genomics. The complete genome sequences of four *M. leprae* strains – TN, Br4923 (Brazil), Thai-53 (Thailand) and NHDP63 (USA) – were 99.995% identical. Combined with a low estimated SNP frequency (1 SNP per 28.4 kb) as compared to the *M. tuberculosis* complex (1 SNP per ~3 kb), the limited diversity indicates that *M. leprae* bacilli are highly clonal and that gene decay occurred prior to global distribution {Monot, Honore et al. 2005; Monot, Honore et al. 2009}. Among seven *M. leprae* strains, 84 informative polymorphisms were identified (78 SNPs, 4 homopolymeric tracts and 2 indels). The 84 markers were genotyped in 400 *M. leprae* samples from 28 geographical regions and 12 ancient *M. leprae* DNA samples from skeletal remains discovered in Europe and Egypt (4^{th} century-16^{th} century AD). The markers defined four phylogenetic groups correlated with geographic origin (SNP-types 1-4) and 16 sub-types (SNP-type 1 A-D, SNP-type 2 E-H, SNP-type 3 I-M and SNP-type 4 N-P) {Monot, Honore et al. 2009}.

The results support a prehistoric African-origin distribution model that follows human migration routes: the progenitor strain – rare SNP-type 2 – originated in East Africa and gave rise to both SNP-type 1 that disseminated eastward into Asia and SNP-type 3 that disseminated westward into the Middle-East and Europe. The latter gave rise to SNP-type 4 in West Africa. The model implies that: (i) *M. leprae* was twice introduced into Asia via a southern route (presence of SNP-type 1 in the Indian subcontinent, Indonesia and the Philippines) and a northern route that follows the Silk Road (1^{st} century, presence of SNP-type 3K in China, Korea and Japan); (ii) *M. leprae* was introduced into West Africa (SNP-type 4) by infected explorers, traders or

colonialists from Europe or North Africa (related SNP-type 3) and not by migrants from East Africa (progenitor SNP-type 2). From West Africa, the slave trade (18th century) disseminated *M. leprae* to the Caribbean islands and South America; and (iii) *M. leprae* was introduced into North, Central and South America (SNP-type 3I) – infecting humans and subsequently armadillos (SNP-type 3) – by infected colonialists and immigrants from Europe (SNP-type 3) (**figure 1**) {Monot, Honore et al. 2005; Monot, Honore et al. 2009}.

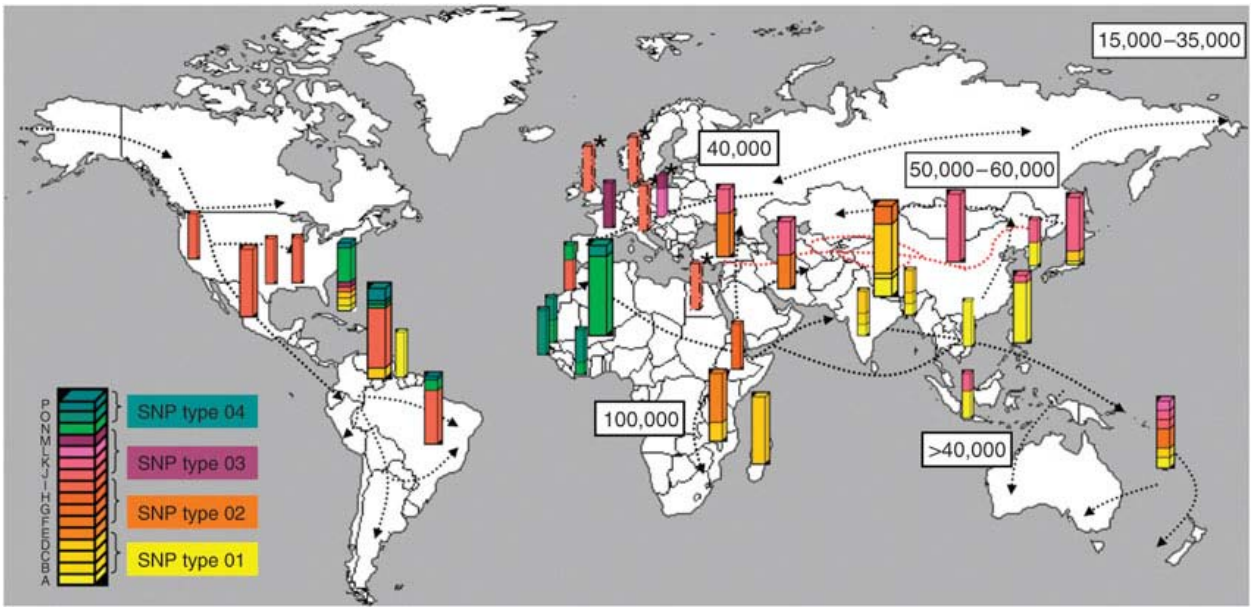


Figure 1 | Prehistoric African-origin distribution model for *M. leprae*

Pillar thickness corresponds to the number of *M. leprae* samples (1–5, thin; 6–29, intermediate; >30, broad). The gray arrows indicate human migration routes with the estimated time of migration shown in years. The red dots indicate the location of the Silk Road (1st century) and * denotes results obtained from ancient *M. leprae* DNA. **Reprinted by permission from Macmillan Publishers Ltd: Nature Genetics {Monot, Honore et al. 2009}, copyright 2009.**

This model is in contrast to the standard evolutionarily recent Indian-origin distribution model that follows urbanization: the progenitor strain originated in India and disseminated to the

Middle East and East Africa via trade routes during the Indus civilization (3rd millennium BCE), westward to Europe by the Greek soldiers of Alexander the Great returning from the invasion of India (4th century BCE) and eastward to China, Japan and the Pacific Islands. Proponents of this model argue that dense populations were required for the dissemination of *M. leprae*, a condition not present in the earlier part of the prehistoric African-origin model {Pinhasi, Foley et al. 2006; Robbins, Tripathy et al. 2009}.

1.5 Epidemiology: an eliminated disease?

At the beginning of 2008, the global prevalence of leprosy (i.e., all patients receiving multi-drug therapy (MDT) on December 31, 2007) was 212,802 and the number of new cases detected in 2007 was 254,525. Among new cases, 55% (140,174) were of the multibacillary (MB) sub-type (see section 1.7) and 9% (22,450) occurred in children. India and Brazil accounted for 15% (137,685) and 54% (39,125) of new cases, respectively. Among countries with a population >1 million, Brazil (2.045 cases per 10,000), Nepal (1.572 cases per 10,000) and East Timor (1.723 cases per 10,000) had not eliminated leprosy {WHO 2008}.

In May 1991, the World Health Assembly passed a resolution to 'eliminate' leprosy as a public health problem by 2000 as defined by a global reduction in prevalence rate to <1 case per 10,000. Simplified diagnosis and the free global distribution of highly effective short-course MDT (see section 1.11) facilitated by The Nippon Foundation of Japan (1995-1999, 50 million USD) and Novartis and the Novartis Foundation for Sustainable Development (2000-2010, 54.5-64.5 million USD) largely contributed to this achievement. However, the elimination of leprosy is a point of contention {Lockwood and Suneetha 2005; Fine 2007}. The definition of 'elimination' used by the World Health Assembly is different from that outlined at the Workshop

on the Eradication of Infectious Diseases: *"reduction to zero of the incidence of a specified disease..."* {Dowdle 1998}. By this definition, elimination of leprosy is precluded by a general lack of knowledge regarding *M. leprae* transmission (see section 1.6).

Compared to 5.35 million cases in 1985, initiatives led by the World Health Organization (WHO) have undeniably reduced the prevalence of leprosy. However, the value of this measure is unclear as prevalence – all patients receiving MDT on December 31 – is greatly impacted by treatment duration. For example, in 1997, the WHO shortened MB-MDT from 24 to 12 months, halving the prevalence of MB cases, and endorsed a single dose of MDT for single-lesion paucibacillary (PB) cases, excluding these patients from prevalence data {WHO Expert Committee on Leprosy 1997}. Similarly, the six month duration of PB-MDT excludes patients initiating treatment between January-June.

In contrast to prevalence, the impact of leprosy control efforts on case detection – an approximation of incidence – is not obvious. No declining trend was observed from 1994-2001, rather global case detection numbers increased {WHO 2002; WHO 2008}, possibly reflecting improved detection efforts. Noticeable decreases since then (e.g., 67% global decrease from 2001-2008) may indicate reduced detection efforts since the "elimination" of leprosy in 2000 as evident by the increased proportion of new cases in India with grade 2 disabilities (i.e., visible deformity) – 1.4% in 2003 and 2.5% in 2007. Moreover, a persistent number of MB (i.e., purportedly contagious cases) and pediatric cases suggest a limited effect of MDT on *M. leprae* transmission {Lockwood and Suneetha 2005}. Nevertheless, SIMLEP – a leprosy modelling software – projected a 2-12% annual decrease in incidence after 2000 {Meima, Smith et al. 2004}.

1.6 Transmission theories

Definitive means of *M. leprae* transmission are largely unknown. Humans are the assumed primary reservoir and the conventional mode of transmission is inhalation of expelled nasal secretions containing *M. leprae* {Britton and Lockwood 2004}. Early experiments demonstrated that one-third of immunologically suppressed mice exposed to *M. leprae* aerosols had quantifiable bacilli in >1 site (ear, footpad, nose, lung). There were no detectable bacilli in the nose immediately following exposure suggesting subsequent recovery from this site resulted from systemic dissemination {Rees 1976}. In a cohort from Indonesia, the adjusted hazard ratio for household contacts (n = 4,903) of MB patients with *M. leprae* DNA in nasal swabs was 9.36 {Bakker, Hatta et al. 2006}, substantiating the high transmission potential of MB cases and the ability of nasal secretions to disseminate *M. leprae*. Occurrences of leprosy originating at the site of *M. leprae* skin inoculation are few, but nevertheless noteworthy: for example, in India, 31 women developed leprosy (10-20 year incubation) after receiving cultural tattoos with unsterile needles {Ghorpade 2002}; in Germany, a surgeon developed leprosy (3 year incubation) after the accidental incision of his finger with a scalpel blade used on a leprosy patient {Achilles, Hagel et al. 2004}; and a Norwegian child developed leprosy (4 year incubation) after a knee wound was dressed in an Ethiopian leprosy hospital {Brandsma, Yoder et al. 2005}.

The limited effect of MDT on *M. leprae* transmission (persistent MB and pediatric cases) suggests additional – possibly environmental or animal – *M. leprae* reservoir(s). In Indonesia, 48% (n = 44) of frequented water sources had *M. leprae* DNA (PCR amplified) and usage (bathing and washing) – but not consumption – of contaminated water was significantly associated with leprosy {Matsuoka, Izumi et al. 1999}. In a laboratory environment, *M. leprae*

survived >4 days after ingestion by the pathogenic soil amoeba *Acanthamoeba castellani*. Furthermore, amoeba derived *M. leprae* replicated normally in foot pads of athymic nude mice {Lahiri and Krahenbuhl 2008}. In India, 37.5% of soil samples (n = 80) had *M. leprae* rRNA (RT-PCR amplification of species-specific 16S rRNA), of which 80% was from "patient" areas (i.e., >3 patients in a single street) and 20% from "no-patient" areas (i.e., 0 patients in a single street), and the bacillary load was significantly higher ($P < 0.00001$) in samples from "patient" areas {Lavania, Katoch et al. 2008}. In the Southeastern United States, the armadillo is a natural *M. leprae* reservoir blamed for the zoonotic transmission of *M. leprae* in patients with no history of travel to leprosy endemic countries or known exposure to leprosy patients. From 1981-2008, 18 such patients from Texas {Freiberger and Fudenberg 1981; Lumpkin, Cox et al. 1984}, Louisiana {West, Todd et al. 1988; Mushatt, Wattanamano et al. 1998}, Georgia {Lane, Walsh et al. 2006} and Mississippi {Abide, Webb et al. 2008} related armadillo exposure either directly (contact, handling carcases, consumption) or indirectly (gardening in proximity to armadillo burrows). Two case-control studies in Texas (28 cases, OR = 3.65) and Brazil (506 cases, OR = 2.01) reported a significant increased risk of leprosy associated with direct armadillo exposure (contact, consumption) {Clark, Murray et al. 2008; Deps, Alves et al. 2008}. Although of interest, the contribution of environmental and zoonotic transmission to leprosy incidence is seemingly irrelevant.

1.7 Classification of the leprosy disease spectrum

Individual differences in the host immune response directed against *M. leprae* largely account for the spectrum of clinical and histological phenotypes delimited by the tuberculoid (TT) and lepromatous (LL) sub-types (Ridley-Jopling classification, see below). Tuberculoid cases

present a limited number of hypopigmented, anaesthetic skin lesions (see section 1.10) with no microscopically discernable bacteria. The correlated Th1-cell-mediated immune (CMI) response (IL2, IFN-gamma) promotes the formation of delineated granulomas – central area of infected macrophages, often fused into multinucleate giant cells, surrounded by T-cells – that successfully limit bacterial replication. Conversely, LL cases present numerous sensitive or anaesthetic skin lesions with high bacillary loads. The correlated Th2-antibody response (IL4, IL10) impedes granuloma formation allowing for uncontrolled bacterial replication and continuous infiltration of the skin and nerves. Borderline forms – borderline-tuberculoid (BT), borderline (BB) and borderline-lepromatous (BL) – comprise the majority of cases. These individuals present intermediate clinical and histological phenotypes resulting from immunologically unstable responses (**figure 2**).

Ridley-Jopling classification is largely based on clinical, histological and immunological findings {Ridley and Jopling 1966}. Given limited medical resources in leprosy endemic regions, the WHO implemented a classification system based on the assumption that a *M. leprae* elicited CMI response is inversely correlated with the number of skin lesions: single-lesion paucibacillary leprosy (1 skin lesion), paucibacillary leprosy (2-5 skin lesions) and multibacillary leprosy (>5 skin lesions) {WHO Expert Committee on Leprosy 1997}. With limited accuracy, the Ridley-Jopling and WHO schemes can be reconciled by re-classifying TT and BT sub-types as paucibacillary (PB), and BB, BL and LL forms as multibacillary (MB) (**figure 2**).

Disparate responses in healthy individuals four weeks after the intradermal injection of whole, autoclaved *M. leprae* (Mitsuda lepromin) approximate the polar ends of the Ridley-Jopling disease spectrum. A positive Mistuda reaction (measurable induration) is predictive of one's

ability to mount a CMI response to *M. leprae* infection (i.e., TT sub-type). Simiarly, a negative Mitsuda reaction is predictive of one's ability to mount an antibody response to *M. leprae* infection (i.e., LL sub-type) {Scollard, Adams et al. 2006}.

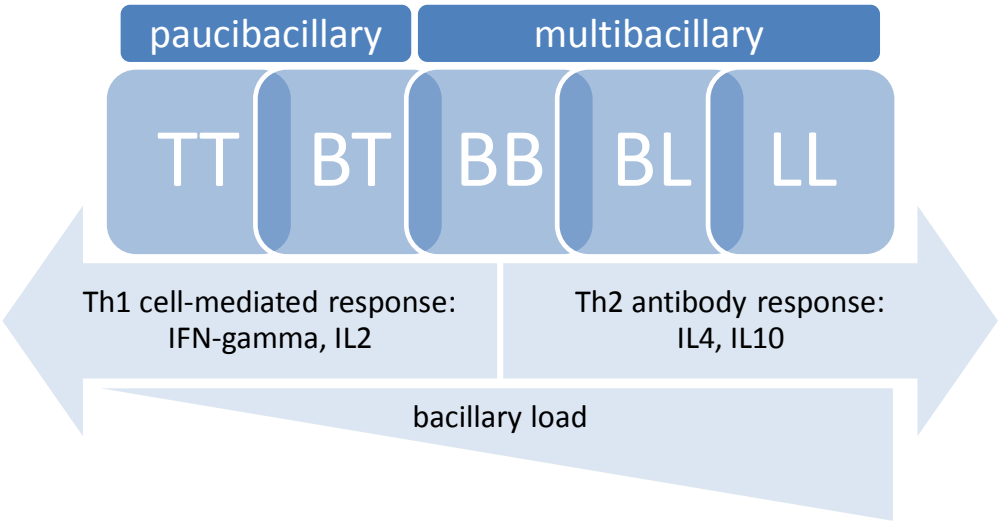


Figure 2 | Clinical and immunological classification of the leprosy disease spectrum

1.8 The host immune response to *Mycobacterium leprae*: a Th1-Th2 paradigm

1.8.1 Tuberculoid leprosy: inside the granuloma

Cellular and molecular effectors underlying the cell-mediated immune (CMI)/granulomatous response in TT skin lesions are emerging. Transcriptional analysis of TT biopsies (n = 4) established a Th1 cytokine mRNA profile – IL2 and IFN-gamma {Yamamura, Wang et al. 1992}. TT biopsies (n = 5) had numerous IL12+ cells and IL12 induced the proliferation of biopsy cultured Th1 T-cells {Sieling, Wang et al. 1994}. Moreover, *M. leprae* up-regulated *IL12RB2* (IL12 receptor, beta subunit) expression on PBMCs (T-cells) from TT cases (n = 9) and

subsequent stimulation with IL12 resulted in IFN-gamma production {Kim, Uyemura et al. 2001}. TT biopsies (*n* = 5) had numerous IL15+ cells and recombinant IL15/*M. leprae* stimulation of PBMCs resulted in the expansion of the CD3+CD56+ T-cell subset predominantly observed in TT biopsies {Jullien, Sieling et al. 1997}. Transcriptional analysis of TT biopsies detected IL18 mRNA and recombinant IL18/*M. leprae* stimulation of PBMCs from TT cases resulted in the expansion of the IFN-gamma producing CD3+CD56+ T-cell subset {Garcia, Uyemura et al. 1999}. CD1a, -b and –c molecules present mycobacterial lipid antigens and were detected in TT biopsies (*n* = 7) on CD83+ dendritic cells clustered at the periphery of granulomas. CD83+ dendritic cells efficiently induced proliferation of a *M. leprae* specific (lipoarabinomannan) CD1-restricted T-cell line, demonstrating their ability to efficiently present antigen {Sieling, Jullien et al. 1999}. Among infiltrating cells in TT biopsies (*n* = 8), 8-15% were positive for granulysin – an antimicrobial protein – and >80% of these were CD4+ T-cells in close proximity to granuloma macrophages. By the granule-exocytosis pathway, *M. leprae*-reactive CD4+ T-cell lines (*n* = 4) from TT cases were able to lyse *M. leprae* pulsed monoctyes {Ochoa, Stenger et al. 2001}.

1.8.2 Lepromatous leprosy: a defective *Mycobacterium leprae*-specific T-cell response

Effectors underlying the selective impairment of the CMI response to *M. leprae* in LL skin lesions is less clear, but attributed to *M. leprae* specific suppressor cells. Transcriptional analysis of LL biopsies (*n* = 4) established a Th2 cytokine mRNA profile – IL4, IL5 and IL10 {Yamamura, Wang et al. 1992}. The majority of PBMCs from MB cases (*n* = 13) did not proliferate or produce IFN-gamma in response to *M. leprae* sonicate or *M. leprae* 35-kDa

protein. Instead, substantial anti-35-kDa immunoglobulin (Ig) G serum levels were detected in these cases. However, proliferative responses and IFN-gamma production was similar to PBMCs from PB cases in response to *M. tuberculosis* antigen (PPD) {Triccas, Roche et al. 1996}. Lepromin antigen (chloroform and ether-treated *M. leprae*) suppressed the proliferation of PBMCs from the majority of LL cases (91%) in response to mitogen (concanavalin A), suggesting the presence of lepromin-induced suppressor cells {Mehra, Mason et al. 1979}. Similarly, PBMCs from LL cases (n = 13) were unresponsive to *M. leprae*, but the proliferative response was restored with the addition of T-cell conditioned medium (i.e., IL2). As before, the proliferative response to *M. tuberculosis* antigen (PPD) was intact {Haregewoin, Godal et al. 1983}. Interestingly, *M. leprae* non-responsive (T-suppressor) and responsive (T-helper) cells were isolated from a single borderline lepromatous patient. The T-suppressor cells suppressed the proliferative response of autologous T-helper cells to *M. leprae* but not to mitogen (phytohaemagglutinin) or herpes simplex virus. Suppression was not mediated by lysis of T-helper cells or antigen-presenting cells {Ottenhoff, Elferink et al. 1986}. Immunostaining of LL skin biopsies revealed that the ratio of CD4+ T-helper cells to CD4+ T-suppressor cells was 1.1:1 (14:1 in TT skin lesions) and the majority of CD8+ T cells were of the suppressor phenotype {Modlin, Gersuk et al. 1986}.

1.9 Leprosy complications

1.9.1 Reversal reaction (type 1 reaction)

Approximately 6-67% of leprosy patients (mainly borderline cases) develop a reversal reaction (RR) – acute inflammation in skin lesions and/or peripheral nerves – often after initiating MDT {Ranque, Nguyen et al. 2007}. Associated neuritis can lead to peripheral nerve function

impairment or loss (sensory and/or motor neuropathy) {Walker and Lockwood 2007}. Reversal reaction (RR) is the result of sudden redirection to a CMI (granulomatous) response, reflecting the immunological instability of the borderline sub-types. Transcriptional analysis of LL biopsies (n = 2) before and during a RR established a dramatic shift from a Th2 cytokine mRNA profile (IL4, IL5 and IL10) to a Th1 cytokine mRNA profile (IL2 and IFN-gamma) {Yamamura, Wang et al. 1992}. RR skin lesions (n = 15) revealed extensive T-cell infiltration (CD4+/CD8+), IFN-gamma, IL12 and iNOS (macrophage effector mechanism) {Little, Khanolkar-Young et al. 2001}.

Interestingly, 15 cases of RR are reported to have developed immune reconstitution inflammatory syndrome (IRIS) after initiation of highly active antiretroviral treatment (HAART) for HIV infection {Batista, Porro et al. 2008}. The restorative effect of HAART on CD4+ T-cell numbers, including *M. leprae* specific memory CD4+ T-cells, resulted in effective CMI responses to *M. leprae* antigens from a presumably latent infection.

1.9.2 Erythema nodosum leprosum (type 2 reaction)

Approximately 20% of LL and 10% of BL cases develop erythema nodosum leprosum (ENL) – acute systemic inflammation, fever and painful skin lesions – often after initiating MDT {Walker and Lockwood 2007}. ENL is an immune complex mediated disorder. Perivascular Ig and complement deposition was detected in ten of 17 biopsies from LL cases with ENL, and *M. leprae* antigen was also detected in seven of these ten cases {Wemambu, Turk et al. 1969}. Transcriptional analysis of LL biopsies (n = 2) before and during ENL established maintenance of the Th2 cytokine profile (IL4, IL5 and IL10) and enhanced expression of IL6, IL8 and IL10

{Yamamura, Wang et al. 1992}. PBMCs from active ENL cases (*n* = 13) produced high levels of TNF in response to *M. leprae* sonicate {Barnes, Chatterjee et al. 1992}.

RR and ENL are effectively treated with the corticosteroids (prednisolone) that inhibit transcription of proinflammatory cytokines. ENL can also be effectively treated with thalidomide, however administration is limited to males and post-menopausal women due to its teratogenic effects {WHO Expert Committee on Leprosy 1997}.

1.10 Step-wise infection of Schwann cells by *Mycobacterium leprae*

Sensory and motor function loss result from peripheral nerve infection by *M. leprae*. Specifically, the leprosy bacillus has a predilection for Schwann cells (SCs), which are myelinating and non-myelinating cells that ensheath large and small diameter axons, respectively {Jessen and Mirsky 2005}.

1. Localization

The idea that *M. leprae* binds to injury exposed dermal SCs was challenged by an armadillo model of lepromatous neuritis. *M. leprae* access to SCs in the endoneurial compartment was preceded by endothelial cell (EC) infection of epineurial and perineurial blood vessels, and increasing the bacillary load in the epineurial blood vessels subsequently increased the bacillary load in endoneurial blood vessels. This suggested the sequential extravasation of *M. leprae* across the peripheral nerve endothelia to SCs {Scollard, McCormick et al. 1999}.

2. Attachment and entry

In the endoneurial compartment, SCs are covered by basil lamina comprised in part by laminins (LNs) – heterotrimeric glycoproteins (α1-4, β1-3, and γ1-2 polypeptide chains). LN2 (α2, β1, γ1) is the predominant endoneurial LN isoform and the G-domain (C-terminal) of the α2 chain (LNα2G) was responsible for "bridging" *M. leprae* to SCs {Rambukkana, Salzer et al. 1997}. Dystroglycan (DG) – a heterodimer (peripheral αDG and transmembrane βDG proteins) – is a laminin receptor expressed on SCs that also binds LN2. Recombinant LNα2G (rLNα2G)-coated *M. leprae* co-localized with αDG on SCs {Rambukkana, Yamada et al. 1998}. Therefore, the *M. leprae*-LNα2G-αDG complex likely defines the SC tropism of *M. leprae*. Both a 21-kDa *M. leprae* surface protein (ML-LBP21) and phenolic glycolipid-1 (PGL-1) – a *M. leprae* specific cell wall component – bound LNα2G {Shimoji, Ng et al. 1999; Ng, Zanazzi et al. 2000}. Both rML-LBP21- and PGL-1-coated beads were ingested by SCs (enhanced by LN2), suggesting that multiple bacterial encoded proteins dually mediate SC binding and infection.

3. Schwann cell-*M. leprae* interactions

In SC-neuron co-cultures and sciatic nerves from *Rag1*$^{-/-}$ mice (absent T- and B-lymphoctyes), *M. leprae* cell wall (PGL-1) first induced rapid demyelination (24 hours post-infection) of myelinating SCs (MSCs) via a contact-dependent mechanism (MSCs were initially resistant to infection) and then proliferation (72 hours) {Rambukkana, Zanazzi et al. 2002}. *M. leprae* binding to the ErbB2 SC receptor induced demyelination by initiating the Ras-Raf-MEK-ERK pathway. Interestingly, this was prevented with Herceptin® – a monoclonal ErbB2 antibody used in the treatment of breast cancer {Tapinos, Ohnishi et al. 2006}. In contrast, >75% of non-myelinating SCs (NMSCs) were infected *in vitro* {Rambukkana, Zanazzi et al. 2002}. Intracellular *M. leprae* induced proliferation (15-30 days post-infection) without cell

transformation by initiating a previously-unknown PKCε-Lck-ERK pathway {Tapinos and Rambukkana 2005}. Recently, *M. leprae* was shown to block human SC apoptosis by inducing SC expression of insulin-like growth factor-I (IGF-I) {Rodrigues, da Silva Maeda et al. 2010}. Combined, these results imply the following SC-*M. leprae* interaction model: *M. leprae* infect NMSCs and promote cell survival (IGF-I) and proliferation (ERK); *M. leprae* expand the host-cell population by rendering MSCs permissive to infection (and subsequent proliferation) through demyelination {Noon and Lloyd 2007}. However, the latter part of the model is problematic given that *M. leprae* does not rapidly induce general demyelination (and subsequent death) in patients with systematic infection.

Conversely, SCs shown to express Toll-like receptor 2 (TLR2) – a pattern recognition receptor that binds mycobacterial lipoproteins – underwent apoptosis in the presence of a *M. leprae* 19-kDa lipoprotein. SCs in leprosy skin biopsies expressed TLR2 and apoptosis was evident (i.e., DNA strand breaks) {Oliveira, Ochoa et al. 2003}. In addition, proliferation of CD4+ Th1-like clones by SCs exposed to *M. leprae* (HLA-DR restricted) demonstrated the efficient antigen presenting capabilities of SCs. HLA-DR restricted antigen presentation resulted in SC lysis by the CD4+ Th1-like clones {Spierings, de Boer et al. 2001}. Therefore, the aforementioned model must be reconciled with experimental evidence of both innate (TLR2) and adaptive (CD4+ Th1 cells) immune mediated SC death as a result of *M. leprae* infection, possibly reflecting temporal events in leprosy-related neuropathogenesis.

1.11 Diagnosis and multi-drug therapy

The "gold standard" for leprosy diagnosis – although imperfect – is the histological examination of a full thickness skin biopsy from an active lesion for: (1) histological patterns of the host

response (see section 1.8); (2) cutaneous nerve involvement; and (3) acid-fast bacilli within nerves {Scollard, Adams et al. 2006}. The WHO recommended a simplified clinical diagnosis of leprosy: a "case of leprosy" is a person having one or more of the following features, and who has still to complete a full course of treatment: (1) hypopigmented or reddish skin lesion(s) with definite loss of sensation; (2) involvement of the peripheral nerves, as demonstrated by definite thickening with loss of sensation; and (3) skin-smear positive for acid-fast bacilli {WHO Expert Committee on Leprosy 1997}. Evaluation of the diagnostic criteria in Ethiopia showed that the sensitivity was 97% and the positive predictive value was 98% when all three cardinal signs were used {International Leprosy Association 2002}.

In the 1940s, dapsone (DDS) was discovered to be an effective treatment for leprosy. DDS is a sulfonamide – a structural analog of *para*-aminobenzoic acid (PABA) – that binds the dihydropteroate synthetase (DHPS) PABA binding site inhibiting bacterial tetrahydrofolic acid (THFA) synthesis. The extended DDS regimen impeded patient compliance leading to the emergence of DDS-resistant *M. leprae* strains. Among selected strains, 50% (n = 10) were DDS-resistant owing to mutations in DHPS (*folP*) {Kai, Matsuoka et al. 1999}. In 1981, to combat drug resistance, the WHO implemented a multi-drug therapy (MDT) regimen that combined DDS, rifampicin (RFP) and clofazimine (CLF). RFP targets the DNA-dependent RNA polymerase β-subunit (*rpoB*) inhibiting transcription and CLF preferentially binds mycobacterial DNA inhibiting growth. Single-dose, 6 or 12 month MDT regimens are prescribed for single-lesion PB, PB and MB leprosy, respectively (doses are adjusted for pediatric cases <15 years) (**figure 3**). The high efficacy of MDT is reflected by low relapse rates – 0.1% per year for PB leprosy and 0.06% per year for MB leprosy {WHO http://www.who.int/lep/mdt/effectiveness/en/index.html 2010}

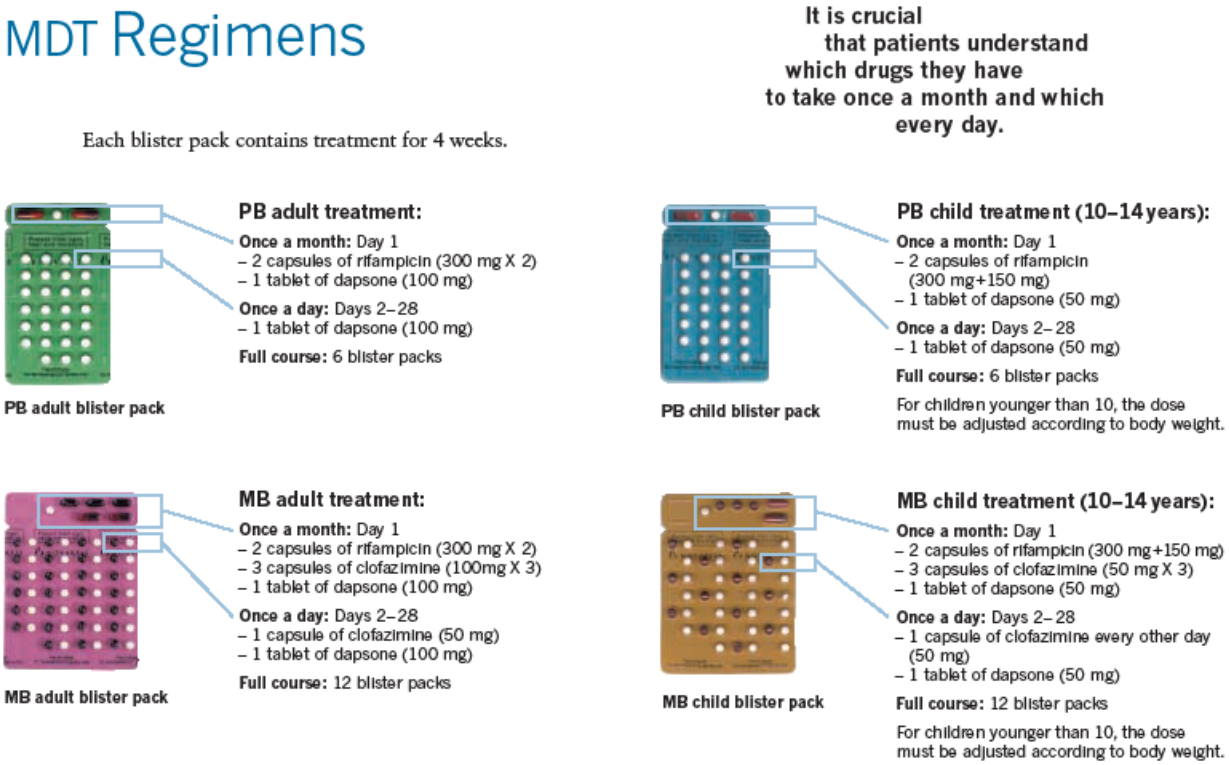


Figure 3 | WHO recommended MDT regimen

Reprinted by permission from the World Health Organization, http://www.who.int/lep/mdt/MDT_Regimens.pdf.

There is no vaccine for leprosy, however studies report varying degrees of protection (20-80%) offered by *Bacille Calmette Guérin* (BCG) – the vaccine for pediatric tuberculosis. A comprehensive review of this topic is presented in *The Continuing Challenges of Leprosy (PREVENTION: THE QUEST FOR A LEPROSY VACCINE),* Scollard D.M. et al., Clin Microbiol Rev, 2006.

PART 2 – GENETIC MAPPING OF COMPLEX BINARY TRAITS

2.1 Complex traits

A complex trait is a non-Mendelian phenotype (binary or quantitative) whose expression is multi-factorial, often involving numerous genetic and/or environmental causes {Lander and Schork 1994; Griffiths, Gelbart et al. 1999}. The hypothetical distribution of the effect size for individual genetic determinants is 'L-shaped' implying the existence of multiple determinants with minimal effects (i.e., minor genes displaying low penetrance of the risk allele) and few determinants with large effects (i.e., major genes displaying increased penetrance of the risk allele) {Wang, Barratt et al. 2005}. The dichotomy requires complementary mapping tools for the identification of major (i.e., linkage analyses) and minor (i.e., association studies) genetic determinants.

2.2 Uncovering a genetic component

The genetic contribution to a complex trait is indicated by several epidemiological observations: for a dichotomous phenotype, increased (>1) relative risk (λ_R = incidence rate in relatives of affecteds (e.g., R = sibling) divided by population incidence rate); for a quantitative phenotype, increased (close to 1) heritability (h^2 = proportion of total phenotypic variance attributable to genetics); increased phenotype concordance between monozygotic twins (genetically identical) compared to dizygotic twins (50% genetically identical); and increased phenotype concordance between adoptees and biological parents compared to adoptive parents {Lander and Schork 1994; Strachan and Read 1999}. Subsequent to establishing a genetic contribution, complex segregation analysis of pedigrees can detect the existence of a major gene and estimate main

parameters accounting for the inheritance pattern (e.g., allele frequency, penetrance) {Lander and Schork 1994; Alcais and Abel 2004}.

2.3 Linkage analysis

The pairing of homologous chromosomes (A1—B1 and A2—B2) during meiosis (i.e., gametogenesis) can result in the exchange of DNA segments creating recombinant chromosomes (A1—B2 and A2—B1). Recombination between two syntenic loci (A1—B1) is inversely proportional to their genetic distance measured by the recombination fraction (proportion of recombinants, 0.01 = 1 CentiMorgan (cM)). For two loci on different chromosomes (i.e., no linkage), the maximum recombination fraction is 0.5 due to independent assortment (Mendel's second law). For two proximal loci that are more rarely separated (i.e., linkage), the recombination fraction is between 0- 0.5 {Strachan and Read 1999}. Linkage analysis exploits this inherent method of maximizing genetic diversity by identifying a genetic marker (frequently microsatellite marker) in sufficiently close proximity to the phenotype locus (i.e., linkage) that they are rarely separated by recombination.

2.3.1 Model-based (classical lod-score method)

In the instance a binary trait (e.g., disease/no disease) has evidence of a major dominant, recessive or additive gene(s), and estimates of allele frequency and penetrance are possible (i.e., specification of a genetic model), model-based linkage analysis is a powerful tool to map the phenotype locus by distinguishing chromosome region(s) that show non-random segregation – or linkage – with the trait {Morton 1955}. Evenly spaced microsatellite markers (highly variable repetitive DNA sequences) are genotyped in multi-case pedigrees (>1 affected child). If the microsatellite marker and phenotype locus are on different chromosomes or separated by a large

physical distance (i.e., no linkage) the recombination fraction is 0.5. For each pedigree, the probability of observing the data is calculated for a series of recombination fractions between 0 and 0.5 (i.e., varying degrees of linkage) to obtain the maximum likelihood estimate of the recombination fraction between the microsatellite marker and the phenotype locus. Next, a likelihood ratio is calculated by dividing the probability of observing the data given linkage (using the estimated recombination fraction) by the probability of observing the data given no linkage (using recombination fraction = 0.5). The 'lod' score is the logarithm of the likelihood ratio and is additive across pedigrees. Lod scores associated with suggestive and significant linkage are 1.9 and 3.3, respectively {Lander and Kruglyak 1995}.

Model-based is the most powerful linkage method if the theorized genetic model is the true model (e.g., Mendelian traits) {Alcais and Abel 2004}. Additionally, the estimated recombination fraction from model-based analyses approximates the physical distance between the genetic marker and the trait locus (0.01 = 1 cM ≈ 1 Mb) {Strachan and Read 1999}.

2.3.2. Model-free (affected sib-pair method)

In the absence of a genetic model for a complex binary trait, model-free linkage analysis is a suitable tool to map the phenotype locus by distinguishing chromosome region(s) excessively shared by related individuals (e.g., sib-pairs) with the same trait {Lander and Schork 1994; Alcais and Abel 2004}. Evenly spaced microsatellite markers are genotyped in affected sibling pairs. If the microsatellite marker and phenotype locus are on different chromosomes or separated by a large physical distance (i.e., no linkage), affected sib-pairs will share 0, 1 or 2 parental alleles with frequency 0.25, 0.5 and 0.25, respectively due to random segregation (Mendel's first law). If the microsatellite marker and phenotype locus are in close proximity

(i.e., linkage), affected sib-pairs will share 1 or 2 alleles with an increased frequency for a dominant or recessive trait, respectively. For each affected sib-pair the number of 0, 1 and 2 shared parental alleles identical by descent (not only by state) is counted and compared to the expected distribution given no linkage using a χ^2 test with 2 degrees of freedom. For genome-wide model-free linkage analyses, commonly accepted *P* values associated with suggestive and significant linkage are 7.4×10^{-4} and 2.2×10^{-5}, respectively {Lander and Kruglyak 1995}.

2.3.3. Applications and limitations of linkage analysis

Linkage analysis is amenable to candidate (e.g., HLA region on chromosome 6p21) or genome-wide studies, the latter offering an unbiased assessment of the genome to identify a chromosome region(s) with suggestive or significant evidence of linkage to the trait irrespective of the biological candidacy of underlying genes (i.e., positional cloning). However, ascertainment of multi-case families (or single-case families, see below) is not trivial, particularly for late-onset traits. Also, the resolution of linkage studies is poor, the resultant co-segregating or shared chromosome region(s) can be extremely large (10-20 cM ≈10-20 Mb) and in the absence of an obvious underlying phenotypes locus, could contain hundreds of genes that require additional fine-mapping by higher resolution methods (see section 2.4.2). More importantly, the identification of minor genetic effects is seemingly beyond the sensitivity of linkage analysis. For a phenotype locus (allele frequency range 0.10-0.50) imparting a genotype relative risk ≥4, 185-297 affected sib-pairs would be required. For a phenotype locus (allele frequency range 0.10-0.50) imparting a genotype relative risk ≤2, unrealistically large samples of >2,500 affected sib-pairs would be required {Risch and Merikangas 1996}.

2.4 Association studies

Genetic association refers to the co-occurrence of an allele or genotype of a genetic marker (frequently single nucleotide polymorphisms (SNPs)) and a trait more often than would be expected from their individual frequencies {Strachan and Read 1999}. For linkage analysis, alleles are irrelevant as different alleles of the same genetic marker will show evidence of linkage among different families owing to the lack of recombination. This can lead to linkage in the absence of detectable association in the corresponding family-based sample.

2.4.1 Population- and family-based association studies

Population-based association studies (the classical epidemiological case – control design) compare the frequency of alleles or genotypes between unrelated individuals expressing (cases) and not-expressing (controls) a trait {Lander and Schork 1994}. SNPs are genotyped in cases and controls and alleles or genotypes are counted and their frequencies compared to the expected distribution given no association (i.e., equal frequencies between cases and controls) using a χ^2 test with 1 or 2 degrees of freedom, respectively. An extension of the genotypic approach is the testing of genetic models (e.g., dominant and recessive). In contrast to multi-case families (or single-case families, see below), population-based samples are easiest to ascertain and most cost-effective as they require the minimum genotyping for the same number of genetic loci. However, a critical limitation is spurious association due to population stratification. This occurs when the targeted population includes a genetically different sub-population (e.g., increased frequency of allele A1) with an increased prevalence of a trait (e.g., disease D). Enrichment of this sub-population in the 'case' group will falsely associate A1 and D {Lander and Schork 1994;

Alcais and Abel 2004}. In addition, for population-based samples, haplotype estimates are less precise and origin of parent effects (e.g., imprinting) are impossible to detect.

Family-based association studies with 'internal controls' (e.g., transmission disequilibrium test (TDT)) are impervious to population stratification {Spielman, McGinnis et al. 1993; Lander and Schork 1994}. Typically, biallelic SNPs are genotyped in single-case families (two parents and one affected child). Considering a heterozygous parent only, the transmitted allele is regarded as the 'case' and the un-transmitted allele as the 'internal control'. If the SNP (alleles A1, A2) is not associated with the trait, the transmission ratio for alleles A1 and A2 from heterozygous parents to affected children is 1:1 due to random segregation (Mendel's first law). If allele A1 is associated with the trait, it will be significantly over-transmitted. The number of transmissions of allele A1 (n_{A1}) and allele A2 (n_{A2}) from heterozygous parents to affected children are counted. The extent to which the observed transmission ratio deviates from the expected is given by the McNemar test statistic = $(n_{A1}- n_{A2})^2/(n_{A1}+ n_{A2})$ that has a χ^2 distribution with 1 degree of freedom. An additional advantage of the family-based approach over case – control is that it lends itself to haplotype analyses as phase can be more easily established.

2.4.2 Applications of association studies

Association studies define three different approaches to the genetic mapping of complex binary traits. The candidate approach tests SNPs in a gene(s) (or putative regulatory region) selected based on a biological understanding of the trait or the results of related experimental models. Alternatively, SNPs are selected in genes underlying a previously identified linkage peak irrespective of their biological candidacy. This 'positional cloning' approach allows for the fine-mapping of a large chromosome region identified by linkage analysis. A comprehensive SNP

association scan of the 95% confidence interval for the underlying trait locus (lod-1 rule) can identify the SNP(s) explaining the linkage signal. Very recently, genome-wide association studies are becoming increasingly more feasible owing to the deposition of >9 million validated SNPs in public databases (e.g., NCBI Entrez SNP), advances in ultra-high throughput genotyping platforms that can assay >1,000,000 SNPs simultaneously and publically available population-specific SNP genotype data (International HapMap project, 1000 Genomes Project) that enable the measure of linkage disequilibrium and selection of tag SNPs (see section 2.5.2) {Hirschhorn and Daly 2005}.

Unlike linkage analysis, the sensitivity of association extends to minor genetic effects. For a phenotype locus (allele frequency range 0.10-0.50) imparting a genotype relative risk = 2, 340-695 single-case families would be required {Risch and Merikangas 1996}. Similarly, for a phenotype locus (allele frequency range 0.20-0.70) imparting a genotype relative risk as low as1.5, ~1500 case – control pairs would be required {Risch 2000}. As such, compared to linkage analysis, association studies may prove to contribute more to the genetic dissection of complex traits (see section 2.1).

2.4.3 Interpretation of a positive association

Despite the level of significance attained, the association of an allele (A1) with a disease (D) requires cautious interpretation pending replication as association defines one of five conditions that are not easily discernable {Lander and Schork 1994}:

1. Functional allele: in this opportune – but rare – instance, the biological impact of allele A1 (e.g., introduction of a premature stop codon, abrogation of a transcription factor binding site) causes D (simple trait) or increases susceptibility to D (complex trait). In the absence of

genetic heterogeneity (multiple independent phenotype loci), non-uniform environmental determinants or important co-variates, the association of A1 with D would be 'easily' replicated in all populations.

2. Natural selection: allele A1 offers a survival advantage to individuals with D resulting in the increased co-occurrence of allele A1 and D. Functional analysis of allele A1 would preclude its contribution to the etiology of D.

3. Population stratification: (see section 2.4.1).

4. Type I error: by chance, allele A1 has a *P* value below the threshold of significance (e.g., 0.05) and the null hypothesis of 'no association' is falsely rejected (i.e., false positive). In this instance, allele A1 would not be replicated.

5. Linkage disequilibrium: The true functional allele emerged on a chromosome with allele A1 and the two variants have not been separated by recombination (i.e., linkage disequilibrium). In an ethnically different population where allele A1 has been separated from the true functional allele, allele A1 will not be replicated (see section 2.5.2).

2.5 Linkage disequilibrium

The emergence of a novel allele resulting in a genetic polymorphism (biallelic SNP-A, allele A1) occurs on a chromosome in the context of a particular extended haplotype (contiguous set of syntenic alleles, B1-C1-D1-A1-E1-F1-G1). Unlike in reconstructed pedigrees with limited generations (i.e., linkage analysis), at the population level, repeated recombination events over many generations will sufficiently reduce the original extended haplotype – eliminating linkage –

to a limited number of non-contiguous alleles (B1-A1-E1). Linkage disequilibrium (LD) is the consequential association of alleles at linked loci (e.g., B1-A1) {Cardon and Bell 2001}.

2.5.1 Measures of linkage disequilibrium: *D* and r^2

The measures of LD consider allelic associations for pair-wise SNP combinations (SNP-A, alleles A1, A2 and SNP-B, alleles B1, B2). In the absence of LD between two SNPs (i.e., linkage equilibrium), the frequency (f) of each possible haplotype (A1—B1, A1—B2, A2—B1, A2—B2) is the product of the corresponding allele frequencies (e.g. f(A1—B1) = f(A1)×f(B1)). In the presence of LD between two SNPs, the distribution of each haplotype deviates from the expected distribution as measured by *D* (*D* = observed f(A1—B1) – f(A1)×f(B1)). Given the dependency of *D* on allele frequencies, *D* is normalized to a value *D*' between 0-1 ($D' = D/D_{max}$, where D_{max} = min[f(A1)×f(B2), f(A2)×f(B1)] if D>0 and D_{max} = max[-f(A1)×f(B1), -f(A2)×f(B2]) {Mueller 2004}. In the instance where one of the four possible haplotypes is not observed in the population, *D*' = 1 {Wang, Barratt et al. 2005}.

More relevant to the design and interpretation of association studies (see section 2.5.2), is the correlation coefficient (r^2) between SNPs in LD. Directly proportional to *D*, $r^2 = D/\sqrt{A1A2B1B2}$ and this measure also takes on a value between 0-1 {Wang, Barratt et al. 2005}. Given the perfect correlation between two SNPs (r^2 = 1 between SNP-A and SNP-B), obtaining the genotype for SNP-A is sufficient to accurately infer the genotype for SNP-B. A concept of critical importance to association studies is that of a 'bin' (see section 2.5.2). A bin is a set of SNPs where at least one SNP – the tag SNP – has a predefined r^2 threshold value with all other SNPs of the bin. SNPs belonging to a bin are not necessarily contiguous and span on average <15 kb {Hinds, Stuve et al. 2005}.

2.5.2 Applications of linkage disequilibrium

The notion of LD – particularly r^2 – bears importance for the interpretation of a positive association, as all SNPs with an $r^2 = 1$ with the associated SNP are statistically indistinguishable. That is to say, in the absence of experimental data, the functional SNP could theoretically be any SNP in a bin constructed utilizing the associated SNP as the tag SNP and an r^2 threshold = 1. As such, the use of high-density association scans to identify the phenotype locus is referred to as 'linkage disequilibrium mapping'. An extension of LD mapping exploits the differential reduction of extended haplotypes in ethnically different populations to identify the functional SNP {Alter, Alcais et al. 2008}. The correlation (r^2) between SNPs varies across ethnic populations thereby changing the physical structure of the bin. Only SNPs that remain correlated with the functional SNP will show consistent evidence for association across ethnic samples. By comparing population-specific bin structures, the number of possible functional SNPs is limited to those that are common among each associated bin (**figure 4**). Inclusion of older populations (e.g., of African origin) is particularly insightful given that the extent of LD is generally reduced owing to increased recombination {De La Vega, Isaac et al. 2005; Hinds, Stuve et al. 2005}.

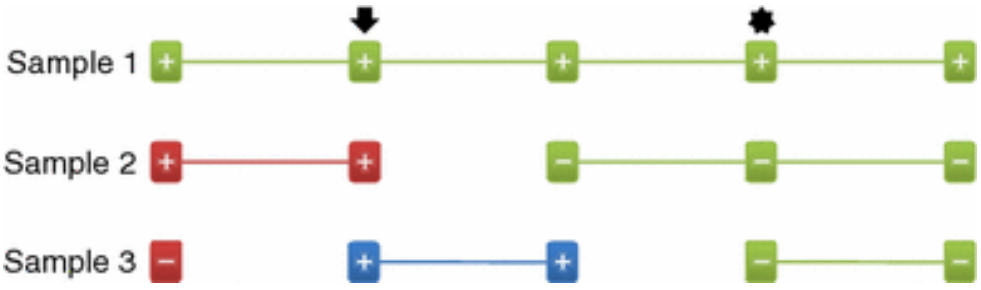


Figure 4 | Reconstructing population-specific bins to identify the functional variant
SNPs identical in color belong to the same bin. The original associated SNP is marked with an asterisk. Positive or absent evidence for association is denoted by (+) or (−), resp ectively. The original SNP is not associated in sample 2 or sample 3. Testing all SNPs from the original bin reveals only SNP 2 (arrow) is associated in all three samples and is therefore the most likely candidate for the functional variant. **With**

kind permission from Springer Science+Business Media: Human Genetics, Leprosy as a genetic model for susceptibility to common infectious diseases, 123, 2008, 227, Alter, Alcaïs, Abel and Schurr, figure 1.

Although the progressive decline in LD facilitates high-resolution genetic mapping of traits, persistent maintenance of LD has permitted the advancement of genome-wide association studies. Owing to the correlation between SNPs, careful selection of tag SNPs maximizes the total genetic variation captured while minimizing the total number of SNPs requiring genotyping {Hirschhorn and Daly 2005}. The recent expansion of the International HapMap project to eleven populations will facilitate the design of population specific SNP panels that more efficiently capture global genomic variation.

2.6 Summary

A flowchart outlining the step-wise approach to genetic mapping of complex traits is provided in **figure 5**.

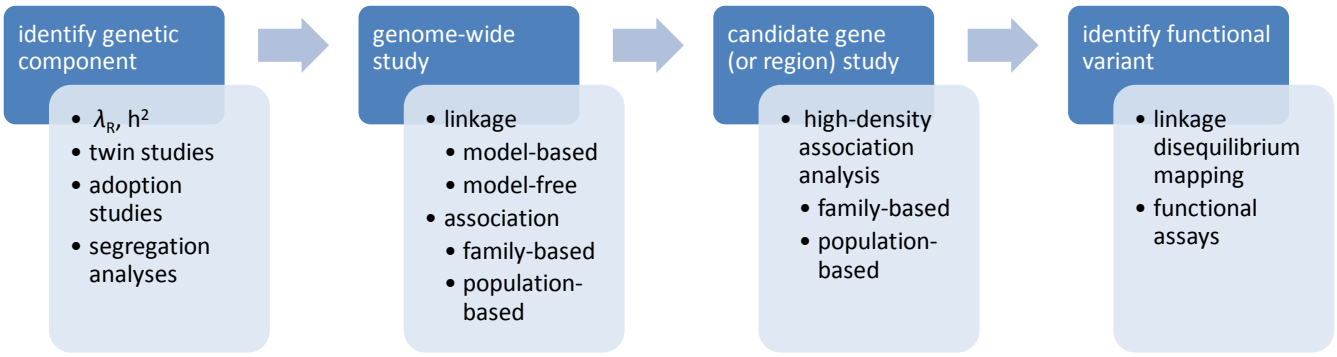


Figure 5 | Step-wise approach to genetic mapping of complex traits

PART 3 – GENETIC SUSCEPTIBILITY TO LEPROSY

3.1 Human genetics of infectious diseases

"The field of human genetics of infectious diseases aims to define the genetic variations accounting for inter-individual variability in the course of human infections" {Casanova and Abel 2007}

The contribution of the host genetic background to infectious disease susceptibility is not as widely acknowledged as for other common diseases (e.g., type-2 diabetes, Crohn's disease, multiple sclerosis) despite the abundance of anecdotal and empirical evidence. A frequently cited account demonstrating the inherent spectrum of resistance (or susceptibility) to infection is the accidental inoculation of 251 neonates with virulent *M. tuberculosis*-contaminated BCG vaccine in Lűbeck, Germany in 1929-1930. Over the course of 12 months, 72 children succumbed to tuberculosis (five children died from causes unrelated to tuberculosis), 61 had severe tuberculosis, 95 had mild tuberculosis and 17 were infected (positive tuberculin test) but asymptomatic, and outcome was independent of socio-economic status {Moegling 1935}. Support of a genetic component to infectious disease susceptibility was provided by a Danish study that followed 960 families and calculated the relative risk (RR) of premature death (age 15-58 years) in adoptees when a biological or adoptive parent died before age 50. The RR of death due to an infectious disease was 5.81 (95% c.i. 2.47-13.7) if a biological parent died of an infectious disease and 1.19 (95% c.i. 0.16-8.99) for cancers. Conversely, the RR of death due to cancers was 5.16 (95% c.i. 1.20-22.2) if an adoptive parent died of a cancer and near unity for infectious diseases {Sorensen, Nielsen et al. 1988}.

Genetic determinants of infection for malaria (*Plasmodium* spp.), HIV and mycobacterial disease are typically recognized examples of genetic susceptibility/resistance to infectious disease. Sickle-cell disease – associated with premature death in developing countries – is caused by a homozygous *β*-globin gene mutation resulting in abnormal haemoglobin. However, it was recognized that the incidence of *Plasmodium* related parasitaemia was 27.9% (n = 43) in African children with sickle-cell trait – heterozygous carriers of the *β*-globin gene mutation – as compared to 45.7% (n = 247) in "non-sicklers". It was concluded that individuals heterozygous for the *β*-globin gene mutation had a selective advantage in malaria-endemic regions {Allison 1954}.

Chemokine receptor 5 (CCR5) is a co-receptor for macrophage-trophic HIV-1 strains, however cell surface expression is abrogated by a 32-basepair deletion in the *CCR5* gene (*CCR5Δ*32). Among 1,343 HIV-1 infected individuals, 0% were homozygous for *CCR5Δ*32 indicating this genotype confers resistance to macrophage-trophic HIV-1 strains (among 612 HIV-1 antibody negative individuals, 3% were homozygous for *CCR5Δ*32). Interestingly, the *CCR5Δ*32 allele is more frequent in the Caucasian population (0.080) as compared to the African (0.017) population suggesting the mutation is evolutionarily recent {Dean, Carrington et al. 1996}.

Mendelian susceptibility to mycobacterial disease (MSMD) is a severe disease caused by BCG vaccines and environmental mycobacteria (and *Salmonella* serotypes) in otherwise healthy individuals. Patients with MSMD carry mutations in genes central to the IL12/23-IFN-gamma circuit that mediates inter-cellular signalling between monocyte/dendritic cells and T/natural killer cells. Mutated genes include *IFNGR1*, *IFNGR2*, *STAT1*, *IL12B* (encodes p40 subunit of IL12 and IL23), *IL12RB1* (encodes beta-1 chain of IL12R and IL23R) and *NEMO*. MSMD

highlights the unique importance of the IL12/23-IFN-gamma circuit in the immune response to mycobacteria (and *Salmonella*) but its redundancy in response to other microorganisms {Filipe-Santos, Bustamante et al. 2006}.

As proposed by Jean-Laurent Casanova and Laurent Abel in 2007, genetic determinants of infectious disease susceptibility occur along a spectrum delimited by monogenic (conventional primary immunodeficiencies) and polygenic models of inheritance, predisposing to multiple infections (one gene → multiple infections) or a single infection (multiple genes → one infection), respectively. Comprising the balance of the spectrum, are monogenic (new primary immunodeficiencies) and major gene(s) models of inheritance, predisposing to a single infection (one gene/major gene(s) → one infection) (**figure 6**) {Casanova and Abel 2007}. Comparable to common diseases, the models of inheritance fitting for common infectious diseases – like leprosy – are the polygenic (minor genes) and major gene(s) models, the latter as determined by complex segregation analyses.

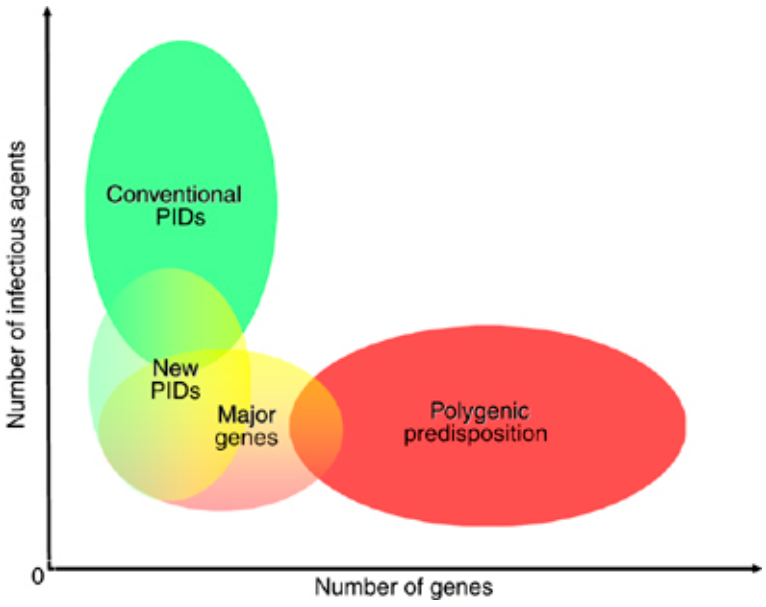


Figure 6 | Spectrum of genetic predisposition to infectious diseases in human patients

PIDs = primary immunodeficiencies. **Reprinted by permission from Macmillan Publishers Ltd: European Molecular Biology Organization Journal {Casanova and Abel 2007}, copyright 2007.**

3.2 Uncovering a genetic component in leprosy: twin studies and complex segregation analyses

A genetic component to leprosy susceptibility – *per se* and sub-type – was established by two twin studies. Comparing 23 monozygotic and 12 dizygotic twin pairs, 19 (82.6%) versus 2 (16.7%) were concordant for disease, respectively. Among the 19 monozygotic concordant twin pairs, 17 (89.5%) were concordant for disease sub-type {Mohamed Ali and Ramanujam 1966}. Similarly, comparing 62 monozygotic and 40 dizygotic twin pairs, 37 (59.7%) versus 8 (20%) were concordant for disease, respectively. Among the 37 monozygotic and 8 dizygotic concordant twin pairs, 32 (86.5%) and 6 (75%) were concordant for disease sub-type {Chakravartti and Vogel 1973}.

Complex segregation analyses have additionally provided clear evidence of a genetic component in so far as the existence of a major gene(s) (displaying Mendelian inheritance) for leprosy *per se* and sub-type, but a consensus on the model of inheritance was not apparent rendering leprosy less suitable to model-based linkage analysis. Among 91 families from the Philippines with ≥1 affected child with lepromatous leprosy, the data were consistent with an autosomal recessive inheritance model for the lepromatous sub-type. Comparing the incidence rate of lepromatous leprosy in brothers of male index cases to that of the population, heritability was estimated at 76.6-84.1% assuming a multifactorial model, suggesting a considerable contribution of multiple genes to susceptibility to the lepromatous sub-type {Smith 1979}. An autosomal recessive

inheritance model was similarly predicted for the tuberculoid sub-type in 72 leprosy families from India {Haile, Iselius et al. 1985}. Among 238 leprosy families from Papua New Guinea, 89 families had >1 affected individual in first and/or second degree relatives, of which 79 showed vertical transmission through multiple generations. The occurrence of familial leprosy (52 families accounted for 301 cases) in a highly communal population suggested underlying genetic susceptibility. However, in a single multigenerational family – 70 unaffected and 19 leprosy affected individuals – it was not possible to discern between a Mendelian (i.e., major gene) or environmental model of inheritance {Shields, Russell et al. 1987}. Among 63 families from Thailand with ≥2 first or second degree relatives with leprosy, the data were most consistent with a dominant inheritance model for leprosy *per se* and, as in India, a recessive inheritance model for the tuberculoid sub-type {Wagener, Schauf et al. 1988}. However, in 27 multigenerational leprosy families from Desirade Island (Caribbean island), although a sporadic model was rejected, a co-dominant or recessive major gene for leprosy *per se* and the non-lepromatous sub-type was detected {Abel and Demenais 1988}. Similarly, a co-dominant and recessive major gene for leprosy *per se* was detected in 285 Vietnamese leprosy families {Abel, Vu et al. 1995} and 1,568 leprosy families from Brazil {Feitosa, Borecki et al. 1995}, respectively. Most recently, among 76 multi-case leprosy families from Brazil, a sporadic model and a single locus model were rejected and a two locus model with a recessive major gene and a recessive modifier gene(s) was predicted for leprosy *per se* {Shaw, Donaldson et al. 2001}.

3.3 Genome-wide linkage analyses: from linked regions to associated SNPs

The first leprosy study employing model-free linkage reported linkage of chromosome region 10p13 (D10S1661 multipoint MLS = 4.09) to paucibacillary leprosy in 224 multi-case families

from Southern India {Siddiqui, Meisner et al. 2001}. Although SNPs in exon 7 of the underlying *MRC1* gene have been suggested to be associated with the paucibacillary sub-type in two review articles, no experimental peer-reviewed data has been provided to date {Hill 2006; Cooke and Hill 2008} (see Chapter 4). In a follow-up study of the regions that showed marginal evidence for linkage, chromosome region 20p12 (D20S115) was modestly linked in families from Tamil Nadu (multipoint MLS = 3.16) but not Andhra Pradesh (multipoint MLS = 0) {Tosh, Meisner et al. 2002}.

An unrelated genome-wide scan in 71 multi-case families from Brazil found chromosome region 6p21 (*HLA-DQA* lod = 3.23) to be weakly linked to leprosy while regions 17q22 (D17S1868 lod = 2.38) and 20p13 (D20S889 lod = 1.51) displayed some suggestive evidence for linkage. Previously in this population, chromosome region 17q – specifically 17q12 and 17q21.33 – had demonstrated suggestive evidence of linkage to the tuberculoid sub-type (D17S250 lod = 2.58) and leprosy *per se* (D17S1795 lod = 2.67), respectively {Jamieson, Miller et al. 2004}. Unlike in the Tamil Nadu sample, the peak at 20p13 in the Belem City sample related almost entirely to the lepromatous and borderline lepromatous families (lod = 1.36) {Miller, Jamieson et al. 2004}. Expanding the genome-wide linkage analysis to quantitative immune response traits, plasma IgG to *M. leprae* soluble antigen (MLSA) was linked to chromosome 8, 17 (17q21) and 21, and IFN-gamma release to MLSA was linked to chromosome 6 (6p21.2 and 6q27), 7, 10, 12 and 14 in a subset of 21 families {Wheeler, Miller et al. 2006}.

A genome-wide scan completed by our group in 86 Vietnamese multi-case families (205 affected offspring) with an unbiased distribution of paucibacillary (44%) and multibacillary (56%) cases, detected linkage to chromosome region 6q25-q26 (multipoint MLB lod = 4.31). In addition,

evidence for linkage between 6p21 microsatellite markers (multipoint MLB lod = 2.62) and leprosy was observed supporting an independent susceptibility locus in the HLA region. Suggestive evidence for linkage of 20p12 (lod = 1.13) and 13q22.1 (lod = 1.68) was noted irrespective of clinical form. Linkage of 13q22.1 was supported by a recent genome-wide association study of leprosy in a case-control sample from China (see section 3.4). SNPs in genes *CCDC122* (13q14.11) and *C13orf31* (13q14.11) were associated with leprosy {Zhang, Huang et al. 2009}. The reported peak at 10p13 in the South Indian study was replicated by testing for linkage in all families with paucibacillary affected sib-pairs (lod = 1.98) {Mira, Alcais et al. 2003}.

A 6.4 Mb interval spanning chromosome region 6q25-q26 (43 genes, *ZDHHC14* → *PACRG*) was selected for a low-density association scan (64 SNPs) in 197 single-case Vietnamese leprosy families. Four of the six SNPs associated with leprosy ($P < 0.05$) localized to the putative *PARK2* promoter overlapping the 5'-region of the adjacent *PACRG* gene. In a subsequent high-density (at the time) association scan (81 polymorphisms) of both genes, 19 SNPs between *PARK2* intron 1 and *PACRG* intron 2 showed evidence for association ($0.03 \geq P \geq 0.0006$). Accounting for the correlation between SNPs, multivariate conditional logistic regression identified two SNPs (*PARK2*_e01(-2599) and rs1040079) that were sufficient to capture all association information. The odds ratio of leprosy for individuals carrying at least one haplotype with both risk alleles (*PARK2*_e01(-2599)-T and rs1040079-C) was 5.28 (95% confidence interval (c.i.) 2.06 – 13.55). Among 13 SNPs tested in an independent sample of 587 cases and 388 controls from Brazil, nine were associated (consistent risk allele) including *PARK2*_e01(-2599) ($P_{corrected} = 0.003$) and rs1040079 ($P_{corrected} = 0.001$) {Mira, Alcais et al. 2004}.

In addition to implicating ubiquitin- and proteasome-dependent processes in the progression of infection with *M. leprae* {Alcais, Mira et al. 2005; Schurr, Alcais et al. 2006}, this study provided a much anticipated "proof of principle" for the genetic analysis of common infectious diseases. This was the first successful attempt to positionally clone genetic risk factors for a common infectious disease, thereby outlining a sound methodology using model-free linkage analysis followed by low- and high-density SNP association scans and successive replication.

In 2005, Bamezai and colleagues published an association study of six *PARK2*/*PACRG* SNPs with leprosy in Northern India. Comparing genotypic frequencies between 286 cases (144 multibacillary and 142 paucibacillary) and 350 controls, the 'TT' genotype for *PARK2*_e01(-2599) was overrepresented in the cases ($P = 0.04$), but insufficiently to withstand an overly conservative Bonferoni correction for multiple testing ($P_{corrected} = 0.6$) {Malhotra, Darvishi et al. 2005}. Clearly, correction in the context of a replication study is an unresolved issue. Significant evidence for association of the same SNP (*PARK2*_e01(-2599)) and risk allele (T-allele) in an ethnically diverse sample should not be underestimated.

Like *M. leprae*, *Salmonella typhi* (typhoid fever) and *Salmonella paratyphi* (paratyphoid fever) are intracellular pathogens. Also, proteasome-mediated degradation of the *Salmonella* protein SopE is a host defence mechanism implicating the proteasome pathway – and possibly PARK2 and PACRG – in enteric fever pathogenesis. As such, in 2006, Ali et al. tested the association of four *PARK2*/*PACRG* SNPs with enteric fever in Indonesia. Comparing allelic frequencies between 115 cases (86 typhoid and 26 paratyphoid) and 259 controls with no known history of enteric fever, the T-allele of *PARK2*_e01(-2599) again showed a weak but significant trend for association with disease ($P = 0.02$, OR=1.58 95% c.i. 1.02-2.36) {Ali, Vollaard et al. 2006}.

Remarkably, this study implicated the same allele in an unrelated phenotype suggesting a broad involvement of PARK2 in intracellular infections.

Interestingly, subsequent to the genetic studies, proteosome function was experimentally implicated in *M. leprae* pathogenesis. An inhibitor of proteasome activity (MG132) reduced *M. leprae* stimulated production of TNF-alpha ($P < 0.05$) and IL10 ($P < 0.001$) from PBMCs, and moreover, inhibited *M. leprae* induced apoptosis of monocytes {Fulco, Lopes et al. 2007}.

3.4 A genome-wide association study

Very recently, the first genome-wide association study of leprosy was published. Initially, 491,883 SNPs (Human610-Quad BeadChip, Illumina®) were analyzed in a "discovery" sample of 706 cases and 1,255 controls from Eastern China. Subsequently, the 93 SNPs with the strongest evidence of association were genotyped in three replication samples (3,254 cases and 5,955 controls combined) from Eastern and Southern China. Among these, 15 SNP across six genes – *HLA-DR-DQ*, *RIPK2*, *TNFSF15*, *CCDC122*, *C13orf31* and *NOD2* – were associated with leprosy (≥2 SNPs in each gene had a $P < 1\times10^{-10}$ for all samples combined). Significant heterogeneity was detected across leprosy sub-types (multibacillary versus paucibacillary) for five SNPs across four genes – *RIPK2*, *C13orf31*, *NOD2* and *LRRK2* – with stronger evidence of association with the multibacillary sub-type. Pathway analysis (Ingenuity Pathways Analysis knowledge database, Ingenuity® Systems) identified a single network (35 genes total) that included five of the seven genes identified in the genome-wide association study (*TNFSF15*, *HLA-DRB1*, *RIPK2*, *NOD2* and *LRRK2*). Remarkably, *PARK2* (see section 3.3) was included in the network as having a known (downstream) direct molecular interaction with *LRRK2* {Zhang, Huang et al. 2009}.

3.5 Candidate chromosome region approach

Based on previous evidence of linkage, association or the harbouring of cytokine genes, eight chromosome regions were tested for linkage to leprosy sub-type in 83 nuclear multi-case leprosy families from North Malawi. Among these, three regions demonstrated suggestive evidence of linkage ($P < 0.05$) and were subsequently tested in 102 extended multi-case leprosy families. Although not statistically significant at the genome-wide level, chromosome region 21q22 showed increased evidence of linkage (from $P = 0.027$ to $P = 0.001$) {Wallace, Fitness et al. 2004}.

3.6 Candidate gene approach

3.6.1 HLA complex genes

The major histocompatibility complex (MHC) or human leukocyte antigen (HLA) complex (~7.6 Mb) on chromosome 6p22.2-p21.32 comprises five contiguous regions: the extended class I (*HIST1H2AA* to *MOG*, 3.9 Mb), the classical class I (*C6orf40* to *MICB*, 1.9 Mb), the classical class III (*PPIP9* to *NOTCH4*, 0.7 Mb), the classical class II (*C6orf10* to *HCG24*, 0.9 Mb) and the extended class II (*COL11A2* to *RPL12P1*, 0.2 Mb) (see MHCmap) {Horton, Wilming et al. 2004}. The HLA complex harbours several gene super-clusters (e.g., HLA class I, olfactory-receptor and tRNA super-clusters) and gene clusters (e.g., HLA class II, tumour necrosis factor and heat shock protein clusters). Among the HLA complex, the classical HLA class I (*HLA-A*, *-B* and *-C*) and classical HLA class II (*HLA-DP*, *-DQ*, *-DR*) genes have been extensively studied in human genetics of disease susceptibility – including infectious disease – owing to their fundamental role in the host immune response {Klein and Sato 2000; Yee and Thursz 2004}. Each classical HLA class I gene encodes a transmembrane alpha-chain that non-covalently

associates with beta-2 microglobulin (*B2M*, 15q21-q22.2). The heterodimeric class I molecule – expressed on all nucleated cells – presents cytosolic pathogen-associated peptide to CD8+ T-cells. In addition, the class I molecules (*HLA-B* and *-C*) engage killer cell immunoglobulin-like receptors (KIR) expressed on natural killer cells {Kulkarni, Martin et al. 2008}. Each classical HLA class II gene encodes a transmembrane alpha (e.g., *HLA-DPA1*) or beta (e.g., *HLA-DPB1*) chain that non-covalently associates (e.g., *HLA-DP*). The heterodimeric class II molecule – expressed on B-cells and antigen-presenting cells – presents extracellular or intravesicular pathogen-associated peptide to CD4+ T-cells (Th1 and Th2) {Klein and Sato 2000; Janeway, Travers et al. 2001}. Of note, the interpretation of linkage and association studies of the HLA complex requires caution owing to long-range LD {De La Vega, Isaac et al. 2005; Lincoln, Montpetit et al. 2005}.

3.6.1.1 Classical HLA class I genes

In a small sample investigating *HLA-B* serotypes, *B**46 was underrepresented (i.e., protective) in multibacillary leprosy cases (frequency = 0.040) compared to controls (frequency = 0.129) from South China (RR = 0.28) {Wang, Kimura et al. 1999}. Among 80 leprosy cases and 120 controls from Turkey, HLA class I serotypes *A**9, *A**10, *A**32, *B**5, *B**21, *Bw**4, *Bw**6, *Cw**1 and *Cw**2 were significantly overrepresented (i.e., risk), and serotypes *A**3, *B**44 and *B**49 were significantly underrepresented in the case group {Kocak, Balci et al. 2002}. In a more comprehensive study, first the distribution of HLA class I serotypes was determined in 103 leprosy cases and 101 controls from Mumbai, India. In the case group, serotypes *A**2, *A**11, *B**40 and *Cw**7 were significantly overrepresented and *A**28, *B**12 and *Cw**3 were significantly underrepresented. Next, the distribution of HLA class I alleles was determined in 32

multibacillary leprosy cases and 67 controls from the same geographic area. In the multibacillary group, alleles *A**0206, *A**1102, *B**1801, *B**5110, *Cw**0407 and *Cw**0703 and haplotypes *A**11─*B**40 and *A**0203─*B**4016─*Cw**0703 were significantly overrepresented and allele *Cw**04011 was significantly underrepresented {Shankarkumar 2004}.

3.6.1.2 Classical HLA class II genes

The study of the HLA region – particularly HLA class II – in leprosy susceptibility is vast and a complete review of the related literature is beyond the scope of this introduction. Moreover, earlier work has been summarized in several reviews {Meyer, May et al. 1998; Fitness, Tosh et al. 2002; Mira 2006}. As such, a limited selection of the most recent publications (i.e., 1993-present) with evidence for association and/or linkage of classical HLA class II genes with leprosy is presented in **table 1**.

Of interest, in the study of 54 tuberculoid leprosy cases and 44 controls from North India, it was noted that for both susceptibility alleles (*DRB1**15 and *DRBl**1404) – in contrast to the risk allele (*DRBl**1301) – an arginine residue occupied amino acid positions 13 (*DRB**15) and 70/71 (*DRB1**1404). Indeed, re-analysis of the data showed that in 87% of cases, arginine occupied positions 13 or 70/71 as compared to 43% of controls, conferring a RR = 8.8 ($P = 5\times10^{-6}$). The side-chains of amino acid residues 13, 70 and 71 line pocket four of the HLA-DR molecule and it was hypothesized by the authors that positively charged residues at these positions may affect peptide binding and/or engagement of the T-cell receptor such that the host immune response favours the development of the tuberculoid sub-type {Zerva, Cizman et al. 1996}.

Table 1. HLA class II association and linkage studies for leprosy (1993-present)

Population	Study design	Sample size	Phenotype	Region or variant (serotype, allele or genotype)	lod, OR, *P* or RR[a]	Reference
North India	case-control	93 cases, 47 controls	MB	*DRB1**15	RR = 16.3	{Rani, Fernandez-Vina et al. 1993}
	case-control	93 cases, 47 controls	TT	*DRB1**15	RR = 5.7	
North India	case-control	54 cases, 44 controls	TT	*DRB1**15 *DRB1**1404 (DR6+) *DRB1**1404 (DR2-) ***DRB1**1301 (DR6+)**	RR = 4.7 RR = 6.0 RR = 7.1 **RR = 0.17**	{Zerva, Cizman et al. 1996}
Belém, Brazil	linkage	73 families (147 sib-pairs)	*per se*	6p21.3 (*DQB1*) 6p21.3 (*DQA1*) 6p21.3 (*DRB1*)	lod = 4.978 lod = 4.870 lod = 5.783	{Shaw, Donaldson et al. 2001}
	TDT	73 families (147 sib-pairs)	*per se*	*DQB1* *DQA1* *DRB1*	*P* = 0.0003 *P* = 0.018 *P* = 0.003	
	TDT	73 families (147 sib-pairs)	TT	*DQB1* *DQA1* *DRB1*	*P* = 0.001 *P* = 0.002 *P* = 0.002	

(Table 1 continued)

Turkey	case-control	80 cases, 120 controls	*per se*	*DQ1* *DQ3* ***DR10*** ***DRw52***	OR = 2.769 OR = 2.727 **OR = 0.937** **OR = 0.886**	{Kocak, Balci et al. 2002}
Tamil Nadu, India	TDT (association and/or linkage)[b]	223 families (230 affected sib-pairs)	PB	*DRB1**15 ***DRB1*09***	*P* = 0.012 ***P* = 0.004**	{Tosh, Ravikumar et al. 2006}
	TDT	223 single-case families	PB	***DRB1*09***	***P* = 0.016**	
Rio de Janeiro, Brazil	case-control	578 cases, 691 controls	*per se*	*DRB1**10 *DRB1**15 ***DRB1*04*** ***DRB1*07*** ***DRB1*12***	*P* = 0.02102 *P* = 0.02288 ***P* = 0.04076** ***P* = 0.04753** ***P* = 0.04399**	{Vanderborght, Pacheco et al. 2007}
Rio de Janeiro, Brazil (Euro-Brazilian)	case-control	N/A	*per se*	***DRB1*04/NN***[c] ***DRB1*07/NN***[c]	**OR = 0.51** **OR = 0.53**	
Rio de Janeiro, Brazil (Afro-Brazilian)	case-control	N/A	*per se*	*DRB1**10/NN[c] *DRB1**15/NN[c]	OR = 3.21 OR = 2.72	
Ho Chi Minh City, Vietnam	TDT	194 single-case families	*per se*	*DRB1**10 ***DRB1*04***	OR = 2.03 **OR = 0.48**	
Rosario city, Argentina	case-control	71 cases, 81 controls	*per se*	*DRB1**1401 *DRB1**1406 ***DRB1*0808*** ***DRB1*1103***	OR = 16.22 OR = 16.22 **OR = 0.1022** **OR = 0.0959**	{Borras, Cotorruelo et al. 2008}

(Table 1 continued)

Paraná, Brazil	case-control	169 cases, 217 controls	*per se*	*DRB1**16 *DQA1**05 **_DQA1_*02** **_DQA1_*04**	OR = 2.52 OR = 1.61 **OR = 0.39** **OR = 0.42**	{da Silva, Mazini et al. 2009}
	case-control	63 cases, 217 controls	LL	*DRB1**1602 *DQA1**05 *DQB1**02 **_DRB1_*04** **_DQA1_*03**	OR = 4.29 OR = 2.38 OR = 2.59 **OR = 0.31** **OR = 0.25**	
	case-control	44 cases, 217 controls	BL	*DRB1**09 *DRB1**1601 **_DQA1_*04**	OR = 4.74 OR = 5.81 **OR = 0.15**	
	case-control	45 cases, 217 controls	TT	*DRB1**1602 *DQA1**05 **_DQA1_*02**	OR = 4.14 OR = 3.85 **OR = 0.23**	
	case-control	63 LL cases, 43 TT cases	LL	*DRB1**08 **_DRB1_*04**	OR = 12.0 **OR = 0.20**	
Chinese Han, China	case-control	305 cases, 527 controls	*per se*	*DRB1**15 **_DRB1_*09**	*P* = 0.002 **_P_ < 0.001**	{Zhang, Liu et al. 2009}
	case-control	141 early-onset cases, 527 controls	*per se* (early-onset)	**_DRB1_*09**	**_P_ = 0.003**	

[a]lod = lod score; OR = odds ratio; RR = relative risk

[b]when all affected siblings are included in the analysis, TDT is a test for association and/or linkage {Tosh, Ravikumar et al. 2006}

[c]NN = all not significantly different alleles collapsed into a unique group (i.e., *DRB1**04, 07, 10, 12 and 15) {Vanderborght, Pacheco et al. 2007}

3.6.1.3 Tumor necrosis factor (*TNF*)

Released in response to infection, TNF is a pleiotropic proinflammatory cytokine whose effects include the upregulation of endothelial adhesion molecules, stimulation of macrophage/neutrophil phagocytic activity and the release of nitric oxide and free oxygen radicals {Knight and Kwiatkowski 1999}. The contribution of *TNF* (chromosome 6p21.3 (HLA class III), 2.8 kb) variants to infectious disease susceptibility (e.g., tuberculosis, cerebral malaria) has been studied, particularly the putative promoter SNP at position -308 relative to the transcription start site (*TNF*-308 allele G = *TNF**1, *TNF*-308 allele A = *TNF**2; OMIM*191160) {Knight and Kwiatkowski 1999}. The G→A transition is thought to alter a transcription factor binding site, and in one study using a 'promoter—luciferase—3'UTR' construct, *TNF**2 (allele A) demonstrated two- to-threefold greater expression in mitogen stimulated cells {Abraham and Kroeger 1999}. Interestingly, the involvement of TNF in the protective host response to *M. leprae* was demonstrated by the relatively rapid development of borderline lepromatous leprosy (1-2 year incubation) in two arthritic individuals from the United States after initiation of Infliximab treatment (humanized monoclonal anti-TNF antibody). Presumably, a pre-existing infection had been sufficiently contained (i.e., sub-clinical leprosy) despite previous administration of immunosuppressive treatments {Scollard, Joyce et al. 2006}. *TNF* association and linkage studies for leprosy are presented in **table 2**. Despite replicated association of *TNF*-308 with leprosy *per se* and sub-type, assignment of the susceptibility allele has been inconsistent.

Table 2. *TNF* association and linkage studies for leprosy

Population	Study design	Sample size	Phenotype	Region or variant (allele or genotype)	lod, OR, *P* or RR[a]	Reference
Calcutta, India	case-control	121 cases, 160 controls	LL	-308 (A)	RR = 3.3	{Roy, McGuire et al. 1997}
Belém, Brazil	linkage	73 families (147 sib-pairs)	*per se*	6p21.3 (-308)	lod = 4.0	{Shaw, Donaldson et al. 2001}
	TDT	73 families (147 sib-pairs)	*per se*	-308 (G)	$P = 1.2\times10^{-5}$	
	TDT	73 families (147 sib-pairs)	LL	-308 (G)	$P = 0.025$	
	TDT	73 families (147 sib-pairs)	TT	-308 (G)	$P = 0.0006$	

(Table 2 continued)

Rio de Janeiro, Brazil	case-control	300 cases, 92 controls	*per se*	**-308 (A)** **-308 (A/G het.)**	***P* < 0.05** ***P* < 0.01**	{Santos, Suffys et al. 2002}
	case-control	210 MB cases, 92 controls	MB	**-308 (A)** **-308 (A/G het.)**	***P* < 0.01** ***P* < 0.01**	
	case-control	210 MB cases, 90 PB cases	PB	-308 (A)	*P* < 0.01	
Ho Chi Minh City, Vietnam	linkage	20 families (59 affecteds)	*per se*	6p21.3 (*TNFa*[b]) 6p21.3 (*TNFb*[b])	*P* = 0.00336 *P* = 0.00438	{Mira, Alcais et al. 2003}
	linkage	12 families (15 LL affecteds, 17 TT affecteds)	sub-type	6p21.3 (*TNFa*[b]) 6p21.3 (*TNFb*[b])	*P* = 0.0033 *P* = 0.0022	
Rio de Janeiro, Brazil	case-control	401 MB cases, 230 PB cases	sub-type	-308/ -238 haplotypes	NS[c]	{Vanderborght, Matos et al. 2004}
Karonga District, Malawi	case-control	243 cases, 283 controls	*per se*	-238	NS[c]	{Fitness, Floyd et al. 2004}
	case-control	216 cases, 258 controls	*per se*	-308	NS[c]	
	case-control	239 cases, 282 controls	*per se*	-376	NS[c]	
	case-control	236 cases, 283 controls	*per se*	-863	NS[c]	

(Table 2 continued)

Tamil Nadu, India	TDT (association and/or linkage)[d]	223 families (230 affected sib-pairs)	PB	-308	NS	{Tosh, Ravikumar et al. 2006}
	TDT (association and/or linkage)[d]	223 families (230 affected sib-pairs)	PB	*TNF*a[b] (10)	*P* = 0.042	
	TDT (association and/or linkage)[d]	223 families (230 affected sib-pairs)	PB	***TNF*a[b] (5)**	***P* = 0.002**	
	TDT	223 single-case families	PB	-308 *TNF*a[b]	NS[c] NS[c]	
Bangkok, Thailand	case-control	37 cases, 140 controls	*per se*	-308 (A/G het.)	OR = 2.69	{Vejbaesya, Mahaisavariya et al. 2007}
	case-control	24 MB cases, 140 controls	MB	-308 (A) -308 (A/G het.)	OR = 2.93 OR = 3.26	
Paraná, Brazil	case-control	165 cases, 239 controls	*per se*	-308 (G) -308 (G/G homo.)	OR = 1.81 OR = 2.06	{Franceschi, Mazini et al. 2009}
	case-control	42 TT cases, 239 controls	TT	-308 (G/G homo.)	OR = 3.33	
	case-control	49 BB cases, 239 controls	BB	-308 (G/G homo.)	OR = 3.08	

[a]lod = lod score; OR = odds ratio; RR = relative risk

[b]microsatellites located in the *TNF* vicinity {Mira, Alcais et al. 2003; Tosh, Ravikumar et al. 2006}

[c]NS = not significant

[d]when all affected siblings are included in the analysis, TDT is a test for association and/or linkage {Tosh, Ravikumar et al. 2006}

3.6.2 Non-HLA genes

3.6.2.1 Solute carrier family 11, member 1 (*SLC11A1*, formerly *NRAMP1*)

Murine susceptibility to *Salmonella typhimurium, Leishmania donovani*, *Mycobacterium bovis* (BCG) and *Mycobacterium lepraemurium* was mapped to a single glycine → aspartic acid substitution (G169D) in *Slc11a1* (formerly *Nramp1*) abrogating its expression {Malo, Vogan et al. 1994; Vidal, Pinner et al. 1996}. Nramp1 – a divalent cation transporter – is recruited to the macrophage phagosome membrane subsequent to phagocytosis of live bacteria (or inert particles) where it was shown to extrude divalent cations – seemingly necessary for microbe function – from the phagosome {Gruenheid and Gros 2000; Jabado, Jankowski et al. 2000}. Additionally, the ability of *M. bovis* to inhibit phagosome-lysosome fusion, which attenuated the acidification of the *M. bovis*-containing phagosome, was diminished by Nramp1 {Hackam, Rotstein et al. 1998}. The presumed conserved function of the human orthologue was substantiated by the observation that SLC11A1 (formerly NRAMP1) promoted phagosome maturation in a human (U-937) monocytic cell line {Gallant, Malik et al. 2007}. The contribution of *SLC11A1* (2q35, 14.9 kb) variants to infectious disease susceptibility (e.g., tuberculosis, buruli ulcer) has been studied (OMIM*600266). *SLC11A1* association and linkage studies for leprosy are presented in **table 3**. Linkage of an extended *SLC11A1* haplotype with leprosy *per se* was substantiated by a single study associating *SLC11A1* (3'UTR TGTG deletion) with the multibacillary sub-type {Abel, Sanchez et al. 1998; Meisner, Mucklow et al. 2001}.

Table 3. *SLC11A1* (formerly *NRAMP1*) association and linkage studies for leprosy

Population	Study design	Sample size	Phenotype	Region or variant (allele or genotype)	OR, *P*[a]	Reference
French Polynesia	linkage	7 multi-case families	*per se*	2q35 (nine intra-*SLC11A1* variants)	NS[b]	{Roger, Levee et al. 1997}
Ho Chi Minh City, Vietnam	linkage	16 families (52 affected sib-pairs)	*per se*	2q35 (*SLC11A1* extended haplotype[c])	$P < 0.005$	{Abel, Sanchez et al. 1998}
Calcutta, India	case-control	227 cases, 165 controls	*per se*	5' promoter microsatellite	NS[b]	{Roy, Frodsham et al. 1999}
	case-control	220 cases, 162 controls	*per se*	intron 4 G→C	NS[b]	
	case-control	222 cases, 154 controls	*per se*	3'UTR TGTG deletion	NS[b]	
Bamako, Mali	case-control	273 cases, 201 controls	*per se*	5' promoter microsatellite intron 4 G→C 3'UTR TGTG deletion	NS[b] NS[b] NS[b]	{Meisner, Mucklow et al. 2001}
	case-control	181 MB cases, 92 PB cases	MB	3'UTR TGTG deletion (het.)	OR = 5.79	

(Table 3 continued)

Karonga District, Malawi	case-control	249 cases, 423 controls	*per se*	5' promoter microsatellite	NS[b]	{Fitness, Floyd et al. 2004}
	case-control	244 cases, 283 controls	*per se*	exon2 9-bp deletion	NS[b]	
	case-control	258 cases, 402 controls	*per se*	3'UTR TGTG deletion	NS[b]	
	case-control	258 cases, 429 controls	*per se*	3'UTR CAAA deletion	NS[b]	
Brazil	case-control	90 cases, 61 household contacts	*per se*	5' promoter microsatellite	NS[b]	{Ferreira, Goulart et al. 2004}
Bangkok, Thailand	case-control	37 cases, 140 controls	*per se*	intron 4 G→C rs17235409 D543N 3'UTR TGTG deletion	NS[b] NS[b] NS[b]	{Vejbaesya, Mahaisavariya et al. 2007}

[a]OR = odds ratio

[b]NS = not significant

[c]*SLC11A1* extended haplotype = six biallelic *SLC11A1* variants, one RFLP located in the *TNP1* gene (*TNP1*-C) and three microsatellite markers (D2S1471, D2S173, D2S104) located in the immediate vicinity of *SLC11A1* {Abel, Sanchez et al. 1998}

3.6.2.2 Vitamin D (1,25- dihydroxyvitamin D3) receptor (*VDR*)

Circulating Vitamin D – 25-hydrox-vitamin D (25D) – is modified by CYP27B1 to produce hormonal 1,25-dihydroxyvitamin D (1,25D), the ligand for VDR – a nuclear receptor. In macrophages, stimulated TLR1/2 (see section 3.6.2.4) induced *CYP27B1* and *VDR* expression, highlighting an immunoregulatory function of 1,25D. The ligand-receptor complex heterodimerizes with retinoid X receptors (RXR) and regulates target gene expression (activation and repression) by binding vitamin D response elements (VDREs). Deficiency in vitamin D has been associated with increased infection (*Helicobacter pylori* and respiratory tract infections) and *VDR* (12q13.11, 63.5 kb) variants have been associated with tuberculosis (albeit inconsistently, OMIM*607948), acute lower respiratory tract infection and HIV {White 2008}. Paradoxically, 1,25D is a documented immunosuppressor, for example, 1,25D suppressed *IL12A* (p35) and *IL12B* (p40) transcription and inhibited dendritic cell maturation into antigen presenting cells {D'Ambrosio, Cippitelli et al. 1998; Piemonti, Monti et al. 2000}. However, 1,25D activated expression of antimicrobial peptides (e.g., cathelicidin antimicrobial peptide), effectors of the innate response that target bacteria, fungus, and viruses {Wang, Nestel et al. 2004}. *VDR* association studies for leprosy are presented in **table 4**. Significant evidence of association for the T/T homozygous genotype for SNP rs731236 (T → C, I352I, AKA *Taq*I) with the lepromatous sub-type was obtained in two studies {Roy, Frodsham et al. 1999; Velarde Felix, Cazarez Salazar et al. 2009}.

Table 4. *VDR* association studies for leprosy

Population	Study design	Sample size	Phenotype	Variant (allele or genotype)	OR[a]	Reference
Calcutta, India	case-control	231 cases, 166 controls	*per se*	**rs731236 I352I *Taq*I (T/C het.)**	**OR = 0.58**	{Roy, Frodsham et al. 1999}
	case-control	107 TT cases, 166 controls	TT	rs731236 I352I *Taq*I (C/C homo.)	OR = 3.22	
	case-control	124 LL cases, 166 controls	LL	rs731236 I352I *Taq*I (T/T homo)	OR = 1.67	
Karonga District, Malawi	case-control	169 cases, 328 controls	*per se*	intron 8 G→T *Apa*I	NS[b]	{Fitness, Floyd et al. 2004}
	case-control	168 cases, 330 controls	*per se*	intron 8 C→T *Bsm*I	NS[b]	
	case-control	247 cases, 398 controsl	*per se*	rs731236 I352I *Taq*I (C/C homo.)	OR = 4.3	
Uberlândia, Brazil	case-control	102 cases, 68 household contracts	*per se*	rs731236 I352I *Taq*I	NS[b]	{Goulart, Ferreira et al. 2006}
Sinaloa, México	case-control	71 LL cases, 144 controls	LL	rs731236 I352I *Taq*I (T/T homo.)	OR = 1.82	{Velarde Felix, Cazarez Salazar et al. 2009}

[a]OR = odds ratio

[b]NS = not significant

3.6.2.3 Interleukin 10 (*IL10*)

IL10 – a Th2 cytokine – is a potent immunoinhibitor and anti-inflammatory, targeting principally macrophages and dendritic cells. IL10 limits antigen presentation by inhibiting HLA class II expression and the upregulation of CD80 and CD86 co-stimulatory molecules, prevents dendritic cell differentiation and maturation and represses the production of proinflammatory cytokines (IL1, IL6, IL12 and TNF), proinflammatory chemokines (CC and CXC) and matrix metalloproteinases {Mosser and Zhang 2008}. *IL10* (1q31-q32, 4.9 kb) variants have been associated with HIV-1, progression of chronic hepatitis B virus infection, tuberculosis (linkage), severe malaria anemia and trachoma (OMIM*124092). Association studies of *IL10* promoter variants for leprosy are presented in **table 5**. Significant evidence of association for the *IL10*-819 T-allele and *IL10*-819 T-allele containing haplotypes with leprosy susceptibility (*per se* and paucibacillary sub-type) was obtained in three studies {Santos, Suffys et al. 2002; Malhotra, Darvishi et al. 2005; Pereira, Brito-de-Souza et al. 2009}. Similarly, evidence of association for *IL10*-819 C-allele containing haplotypes with leprosy protection (*per se* and lepromatous sub-type) was obtained in two studies {Malhotra, Darvishi et al. 2005; Franceschi, Mazini et al. 2009}. In the most recent association study, IL10 production from *M. leprae* stimulated PBMCs was lower in carriers of the *IL10*-819 T-allele as compared to non-carriers {Pereira, Brito-de-Souza et al. 2009}.

Table 5. *IL10* association studies for leprosy

Population	Study design	Sample size	Phenotype	Variant (allele or genotype)	OR, P[a]	Reference
Rio de Janeiro, Brazil	case-control	143 MB cases, 59 PB cases	PB	-819 T→C (T)	OR = 2.28	{Santos, Suffys et al. 2002}
	case-control	202 cases, 62 controls	*per se*	-819 T→C (T/T homo.) -592 C→A	OR = 2.64 NS[b]	
Karonga District, Malawi	case-control	212 cases, 362 controls	*per se*	−1082 A→G	NS[b]	{Fitness, Floyd et al. 2004}
	case-control	215 cases, 353 controls	*per se*	−819 C→T	NS[b]	
	case-control	191 cases, 349 controls	*per se*	−592 C→A	NS[b]	

(Table 5 continued)

Rio de Janeiro, Brazil	case-control	297 cases, 283 controls	*per se*	-3575 -2849 -2763 -1082 -819	NS[b] NS[b] NS[b] NS[b] NS[b]	{Moraes, Pacheco et al. 2004}
	case-control	297 cases, 283 controls	*per se* (and severity)	**-3575A/-2849G/-2763C haplotype**	**OR = 0.35**	
	case-control	297 cases, 283 controls	*per se*	-3575T/-2849A/-2763C haplotype	OR = 2.37	
India	case-control	282 cases, 266 controls	*per se* (and severity)	**-3575T/-2849G/-2763C/-1082A/-819C/-592C haplotype** 3575T/-2849G/-2763C/-1082A/-819T/-592A haplotype	**OR = 0.58** *P* = 0.0002	{Malhotra, Darvishi et al. 2005}
Paraná, Brazil	case-control	165 cases, 240 controls	*per se*	-1082/-819/-592 haplotypes	NS[b]	{Franceschi, Mazini et al. 2009}
	case-control	65 LL cases, 240 9 controls	LL	**-1082G/-819C/-592C haplotype**	***P* = 0.02**	
Brazil	case-control; meta-analysis	374 cases, 380 controls; 2,702 individuals	*per se*	-819T containing haplotype	OR = 1.40	{Pereira, Brito-de-Souza et al. 2009}

[a]OR = odds ratio

[b]NS = not significant

3.6.2.4 Toll-like receptors (TLRs)

TLRs – pluripotent effectors of the innate response – are pattern recognition receptors that bind molecules characterized by their so-called pathogen-associated molecular patterns (PAMPs). For example, among the ten human TLRs, TLR2/TLR6 and TLR2/TLR1 heterodimers bind diacylated and triacylated lipoproteins, respectively. Subsequent to TLR engagement, the cytoplasmic Toll/IL1 receptor (TIR) domain initiates a signalling pathway that culminates in the nuclear translocation of NF-kappaB and transcriptional activation of proinflammatory immunomodulatory genes (cytokines and chemokines). In addition, TLRs were shown to promote phagocytosis, stimulate antimicrobial peptide synthesis, upregulate expression of CD80 and CD86 co-stimulatory molecules and induce apoptosis {McInturff, Modlin et al. 2005}. Although, several *TLR* variants have been associated with infection (e.g., tuberculosis, *Meningococococcus* and malaria) {Misch and Hawn 2008}, very few of them have been clearly replicated and can be considered as convincing.

TLR2 was detected on primary human Schwann cells and a TLR2 ligand – *M. leprae* specific 19-kDa lipopeptide – induced apoptosis. TLR2+ apoptotic Schwann cells were similarly observed in leprosy skin biopsies {Oliveira, Ochoa et al. 2003}. TLR2-dependent IL12p40 production was detected from primary human monocytes and monocyte-derived dendritic cells after stimulation with a 19-kDa or 33-kDa *M. leprae* lipopeptide. In tuberculoid skin biopsies (*n* = 10), 40-50% of granulomatous cells were TLR2+ (predominately TLR2/TLR1 heterodimers) of which 90-95% were of the monocyte/macrophage lineage and 5% were dendritic cells. In contrast, TLR2 and TLR1 expression was relatively much weaker in lepromatous skin biopsies {Krutzik, Ochoa et al. 2003}. Treatment of monocytes with a TLR2/TLR1 ligand (*M.*

tuberculosis 19-kDa lipopeptide) induced differentiation into macrophages (DC-SIGN+CD16+) and dendritic cells (CD1b+DC-SIGN−). Interestingly, TLR2/TLR1 mediated differentiation of peripheral monocytes from lepromatous cases was limited to macrophages (impaired dendritic cell differentiation was not observed in peripheral monocytes from tuberculoid cases) and this was reflected in the absence of CD1b+ cells in lepromatous skin biopsies {Krutzik, Tan et al. 2005}.

TLR1(4p14, 8.5 kb), *TLR2* (4q32, 21.8 kb) and *TLR4* (9q32-q33, 13.3 kb) association studies for leprosy are presented in **table 6**. As for *TNF*-308, despite replicated association of *TLR* variants with leprosy *per se* and sub-type, no consistent susceptibility allele has been identified. Of note however, individuals homozygous for the protective *TLR1* SNP rs5743618 G-allele (602-serine) had lower TLR1 cell surface levels due to impaired cell surface trafficking, suggesting that *M. leprae* benefits from functional TLR1 {Johnson, Lyle et al. 2007}. Although rs5743618 was not associated with leprosy (or sub-type) in Nepal, *TLR1* transfected HEK293 cells with the G-allele (602-serine) had reduced constitutive and stimulated NF-kappaB activity in response to TLR2/TLR1 ligands (whole, irradiated *M. leprae*, *M. leprae* cell wall and tri-acylated lipopeptide). Moreover, TLR2/TLR1 ligand stimulated PBMCs from rs5743618 G/G homozygous individuals had reduced production of IL6, IL1beta and TNF-alpha {Misch, Macdonald et al. 2008}. In the most recent study of *TLR4* variants, *M. leprae* – unlike LPS, the classical TLR4 ligand –did not induce IL1beta, IL6 or IL12p70 production from human monocytes. Additionally, in monocytes and PBMCs, pre-stimulation with *M. leprae* for 24 hours decreased LPS-induced production of IL1beta and IL6, respectively, suggesting that *M. leprae* modulates TLR4 signalling {Bochud, Sinsimer et al. 2009}.

It is important to mention that the reported association of non-synonymous *TLR2* SNP Arg677Trp with lepromatous leprosy in a Korean study has been discredited, and this variant is in fact attributable to a nucleotide substitution in a ~23 kb duplicated region that is 93% homologous to *TLR2* exon 3 {Kang and Chae 2001; Malhotra, Relhan et al. 2005}. This purported SNP was not found in an Ethiopian, Indian or Japanese cohort {Fitness, Floyd et al. 2004; Malhotra, Relhan et al. 2005; Mikita, Kanazawa et al. 2009}. As such, functional studies related to Arg677Trp are of a questionable nature and have not been presented {Kang, Lee et al. 2002; Bochud, Hawn et al. 2003; Kang, Yeum et al. 2004}.

Table 6. *TLR1*, *TLR2* and *TLR4* association studies for leprosy

Population	Study design	Sample size	Phenotype	Gene	Variant (allele, genotype or genetic model)	OR, P[a]	Reference
Turkey	case-control	57 cases, 90 controls	*per se*	*TLR1*	**rs5743618 S602I (G)**	**OR = 0.48**	{Johnson, Lyle et al. 2007}
	case-control	57 cases, 90 controls	*per se*	*TLR1*	rs5743618 S602I (T/T homo.) **rs5743618 S602I (G/G homo.)**	*P* = 0.018 ***P* = 0.017**	
Nepal	case-control	581 MB cases, 343 PB cases	sub-type	*TLR1*	rs5743618 rs5743563 rs5743565 rs5743592 rs5743595 G743A	NS[b] NS[b] NS[b] NS[b] NS[b] NS[b]	{Misch, Macdonald et al. 2008}
Bangladesh	case-control	842 cases, 543 controls	*per se*	*TLR1*	rs5743618 S602I rs4833095 N248S (G/G homo.) **rs4833095 N248S (A/G het.)**	NS[b] OR = 1.34 **OR = 0.78**	{Schuring, Hamann et al. 2009}
Karonga District, Malawi	case-control	210 cases, 379 controls	*per se*	*TLR2*	intron 2 microsatellite	NS[b]	{Fitness, Floyd et al. 2004}

(Table 6 continued)

Ethiopia	case-control	441 cases, 197 controls	*per se*	*TLR2*	rs3804099 N199N rs3804100 S450S **5' microsatellite (290, additive)**	NS[b] NS[b] **OR = 0.62**	{Bochud, Hawn et al. 2008}
	case-control	298 LL cases, 138 TT cases	LL	*TLR2*	5' microsatellite (282, dominant) **5' microsatellite (288, dominant)**	OR = 1.64 **OR = 0.49**	
Karonga District, Malawi	case-control	235 cases, 288 controls	*per se*	*TLR4*	rs4986790 D299G	NS[b]	{Fitness, Floyd et al. 2004}
Ethiopia	case-control	441 cases, 197 controls	*per se*	*TLR4*	**rs4986790 D299G (additive)** **rs4986791 T399I (C dominant)** rs5030719 Q510H M658G	**OR = 0.34** **OR = 0.16** NS[b] NS[b]	{Bochud, Sinsimer et al. 2009}
			per se	*TLR4*	**G/T/G/A haplotype[c]** **G/C/T/A haplotype[c]**	**OR = 0.12** **OR = 0.23**	

[a]OR = odds ratio

[b]NS = not significant

[c]four SNP haplotypes (rs4986790, rs4986791, rs5030719 and M658G) {Bochud, Sinsimer et al. 2009}

3.7 Additional genes

In addition to the presented studies, there are several reported – albeit unreplicated – genetic determinants for leprosy. Those published before 2001 have been reviewed {Fitness, Tosh et al. 2002}. As such, the most recent publications (i.e., 2001-present) of genes with evidence for association with leprosy are presented in **table 7**.

Table 7. Unreplicated positive association studies for leprosy (2001-present)

Population	Study design	Sample size	Phenotype	Gene (protein)	Variant (allele or genotype or genetic model)	OR, P[a]	Reference
Yogyakarta, Indonesia	case-control	27 MB cases, 26 PB cases	PB	*LAMA2* (laminin alpha2)	rs6569605 V2587A (T/C het.)	OR = 6.73	{Wibawa, Soebono et al. 2002}
Goiánia, Brazil	case-control	10 TT cases, 98 controls	TT	*IFNG* (IFN-gamma)	intron 1 microsatellite	*P* = 0.013	{Reynard, Turner et al. 2003}
Karonga District, Malawi	case-control	251 cases, 399 controls	*per se*	*CR1* (complement component (3b/4b) receptor 1)	**rs17047660 K1590E (G/G homo.)**	**OR = 0.3**	{Fitness, Floyd et al. 2004}
	case-control	184 cases, 333 controls	*per se*	*LTA* (lymphotoxin alpha)	5'UTR 3.5-kb microsatellite (105)	OR = 1.6	
Japan	case-control	130 LL cases, 46 TT cases	LL	*IL12RB2* (IL12 receptor, beta 2)	-1035 A→G -1023A→G -650 delG -464 A→G	OR = 3.97 OR = 2.95 OR = 3.74 OR = 3.64	{Ohyama, Ogata et al. 2005}
	case-control	130 LL cases, 46 TT cases	TT	*IL12RB2* (IL12 receptor, beta 2)	-1035A/-1023A/-650G/-464A haplotype	*P* = 0.0002	

(Table 7 continued)

Tamil Nadu, India	TDT (association and/or linkage)[b]	223 families (230 affected sib-pairs)	PB	*MICA* and *MICB* (MHC class I polypeptide-related sequence A and B, respsectively)	*MICA* *MICB*	*P* = 0.01 *P* = 0.003	{Tosh, Ravikumar et al. 2006}
	TDT	223 single-case families	PB	*MICA* and *MICB* (MHC class I polypeptide-related sequence A and B, respsectively)	*MICA* (global) *MICA**5A5.1 ***MICB**CA21**	*P* =0.013 *P* =0.006 ***P*= 0.015**	
New Delhi, India	case-control	80 cases, 89 controls	*per se*	*IL12B* (IL12B)	**3'UTR *Taq*I site (2/2 homo.)**	***P* = 0.002**	{Morahan, Kaur et al. 2007}
Guadalajara, México	case-control	44 LL cases, 51 controls	LL	*IL12B* (IL12B)	3'UTR 1188 A→C (C/C homo.)	*P* < 0.05	{Alvarado-Navarro, Montoya-Buelna et al. 2008}
Curitiba, Brazil	case-control	264 cases, 214 controls	*per se*	*MBL2* (mannose-binding lectin (protein C) 2, soluble	*MBL2**LYPA haplotype	OR = 1.94	{de Messias-Reason, Boldt et al. 2007}
	case-control	150 LL cases, 36 TT cases	LL	*MBL2* (mannose-binding lectin (protein C) 2, soluble	**Low expression of MBL haplotypes**	**OR = 0.56**	

(Table 7 continued)

Paraná, Brazil	case-control	65 LL cases, 42 TT cases	TT	*KIR* genes (killer cell immunoglobulin-like receptor)	*KIR2DS3*	OR = 2.72	{Franceschi, Mazini et al. 2008}
	case-control	45 BB cases, 38 TT cases	BB	*KIR* and HLA class I ligand pairs	***KIR2DL1 – HLA-C* (group C2)**	**OR = 0.30**	
	case-control	38 TT cases, 45 BB cases	TT	*KIR* and HLA class I ligand pairs	*KIR2DL1 – HLA-C* (C2/C2 homo.) ***KIR2DL3 – HLA-C* (C1)** **KIR2DL3 – *HLA-C* (C1/C1 homo.)**	OR = 4.16 **OR = 0.20** **OR = 0.31**	
	case-control	45 BB cases, 60 LL cases	BB	*KIR* and HLA class I ligand pairs	*KIR3DL2 – HLA-A**3/11	OR = 2.67	
Curitiba, Brazil	case-control	158 cases, 210 controls	*per se*	*FCN2* (ficolin 2)	**-986A/-602G/-4A/+6424G**	**OR = 0.10**	{de Messias-Reason, Kremsner et al. 2009}
Sinaloa, México	case-control	46 LL cases, 151 controls	LL	*DEFB1* (defensin, beta 1)	rs1800972 (C dominant) C/G/C haplotype[c]	OR = 3.06 OR = 2.25	{Prado-Montes de Oca, Velarde-Felix et al. 2009}

[a]OR = odds ratio

[b]when all affected siblings are included in the analysis, TDT is a test for association and/or linkage {Tosh, Ravikumar et al. 2006}

[c]three SNP haplotype (rs1800972, rs11362, rs1800971) {Prado-Montes de Oca, Velarde-Felix et al. 2009}

3.8 Genetic susceptibility to leprosy: a two-stage model

Based on the results of complex segregation analyses, linkage studies and association studies, a two-stage model of genetic susceptibility to leprosy was proposed (**figure 7**). In stage one, "group-1" genetic risk factors (e.g., *PARK2*, *LTA*, 13q22.1, 20p12.3) contribute to the establishment of disease by *M. leprae* subsequent to exposure (i.e., leprosy *per se*). Importantly, the majority of individuals (>90%) exposed to *M. leprae* are asymptomatic {Convit, Sampson et al. 1992; Chaudhury, Hazra et al. 1994; Gupte, Vallishayee et al. 1998}. In stage two, "group-2" genetic risk factors (e.g., *HLA-DRB1**15, 10p13) contribute to the manifestation of single-lesion paucibacillary (possibly self-limiting), paucibacillary or multibacillary leprosy (i.e., sub-type).

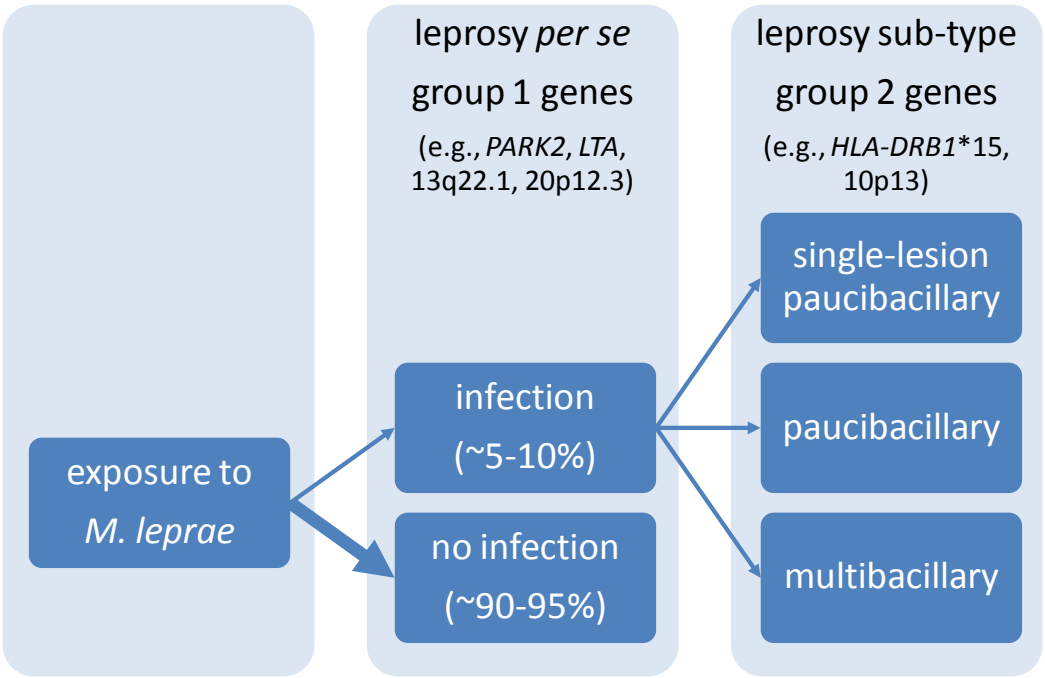


Figure 7 | Two-stage model of genetic susceptibility to leprosy

Chapter 2 │ Stepwise replication identifies a low-producing lymphotoxin-α allele as a major risk factor for early-onset leprosy

In pursuit of novel genetic determinants of leprosy susceptibility, Chapter 2 describes a SNP association scan (307 SNPs) of a 10.4-Mb target interval underlying the previously identified linkage peak on chromosome region 6p21 (i.e., the HLA complex). This approach led to the positional cloning of *LTA*+80 (rs2239704) as a major risk factor for early-onset leprosy in two family-based samples from Vietnam and a population-based sample from north India. Our finding is consistent with the functional role of the *LTA*+80 variant given that the risk allele (A allele) was previously correlated with reduced LTA expression.

Stepwise replication identifies a low-producing lymphotoxin-α allele as a major risk factor for early-onset leprosy

Alexandre Alcaïs[1,2,9], Andrea Alter[3,9], Guillemette Antoni[1,9], Marianna Orlova[3], Nguyen Van Thuc[4], Meenakshi Singh[5], Patrícia R. Vanderborght[6], Kiran Katoch[7], Marcelo T. Mira[8], Vu Hong Thai[4], Ngyuen Thu Huong[4], Nguyen Ngoc Ba[4], Milton Moraes[6], Narinder Mehra[5], Erwin Schurr[3] & Laurent Abel[1,2]

[1]Laboratoire de Génétique Humaine des Maladies Infectieuses, Institut National de la Santé et de la Recherche Médicale, U550, 75015 Paris, France.

[2]Université Paris René Descartes, Faculté Médecine Necker, 75015 Paris, France.

[3]McGill Centre for the Study of Host Resistance and Departments of Human Genetics, Medicine and Biochemistry, McGill University, 1650 Cedar Avenue, Montreal, Quebec H3G1A4, Canada.

[4]Hospital for Dermato-Venereology, Nguyen Thong Street, District 3, Ho Chi Minh City, Vietnam.

[5]Department of Transplant Immunology and Immunogenetics, All India Institute of Medical Sciences, Ansari Nagar, 110029 New Delhi, India.

[6]Leprosy Laboratory, Tropical Medicine Department Oswaldo Cruz Institute, FIOCRUZ, CEP 21040-900 Rio de Janeiro, Brazil.

[7]Central Jalma Institute of Leprosy and Other Infectious Diseases, Taj Ganj, 282001 Agra, India.

[8]Centro de Ciências Biológicas e da Saúde, Pontifícia Universidade Católica do Paraná, Rua Imaculada Conceição, 1155, CEP 80215-901, Curitiba, Paraná, Brasil.

[9]These authors contributed equally to this work and are listed in alphabetical order.

Abstract

Host genetics has an important role in leprosy, and variants in the shared promoter region of *PARK2* and *PACRG* were the first major susceptibility factors identified by positional cloning. Here we report the linkage disequilibrium mapping of the second linkage peak of our previous genome-wide scan, located close to the HLA complex. In both a Vietnamese familial sample and an Indian case-control sample, the low-producing lymphotoxin-α (*LTA*)+80 A allele was significantly associated with an increase in leprosy risk ($P = 0.007$ and $P = 0.01$, respectively). Analysis of an additional case-control sample from Brazil and an additional familial sample from Vietnam showed that the *LTA*+80 effect was much stronger in young individuals. In the combined sample of 298 Vietnamese familial trios, the odds ratio of leprosy for *LTA*+80 AA/AC versus CC subjects was 2.11 ($P = 0.000024$), which increased to 5.63 ($P = 0.0000004$) in the subsample of 121 trios of affected individuals diagnosed before 16 years of age. In addition to identifying *LTA* as a major gene associated with early-onset leprosy, our study highlights the critical role of case- and population-specific factors in the dissection of susceptibility variants in complex diseases.

Leprosy is a chronic infectious disease caused by *Mycobacterium leprae* that still affects an estimated 300,000 new patients each year {WHO 2006}. Susceptibility to leprosy strongly depends on the genetic background of the host {Casanova and Abel 2002; Mira 2006}. Building on previous segregation studies {Abel and Demenais 1988; Abel, Vu et al. 1995} we detected linkage of leprosy susceptibility to chromosome region 6q25-q26 in a genome-wide scan in Vietnamese multiplex leprosy families {Mira, Alcais et al. 2003}. We subsequently identified *PARK2* and *PACRG* variants as strong leprosy susceptibility factors {Mira, Alcais et al. 2004}. The second linkage signal was detected in the 6p21 chromosomal region, and fine mapping placed the peak (lod score = 2.7) close to D6S2427, at approximately 39.5 Mb on build 36 of the human genome sequence map. Based on the 95% confidence interval (c.i.) for the location of the underlying leprosy susceptibility locus {Dupuis and Siegmund 1999}, we selected a chromosomal region extending from 31.6 Mb to 42.0 Mb on the physical map (**Fig. 1**) for association studies in a sample of 194 simplex leprosy families from Vietnam. This interval contains the 108 genes located within the HLA class II and class III regions, and 116 annotated genes centromeric to HLA class II (*PHF1→CCND3*; NCBI build 36). We genotyped 418 SNPs in the target interval, 307 of which were suitable for association analysis (**Fig. 1** and **Supplementary Table 1**).

In univariate analysis, six SNPs were significantly associated with leprosy at the 0.01 level, four of these were located in the HLA class III region (**Fig. 1** and **Supplementary Table 2**). Stepwise multivariate conditional logistic regression analysis with these six SNPs showed that the best model (P = 0.000003) included four SNPs (rs2071590, rs3128961, rs937662 and rs707928) with marginal multivariate P values of 0.001, 0.007, 0.014 and 0.039, respectively. Of note, none of the second-

order interactions between these four SNPs was significant at the 0.05 level. Multivariate analysis also demonstrated the effects of these four SNPs to be independent of the two *PARK2* and *PACRG* SNPs previously associated with leprosy in the same sample {Mira, Alcais et al. 2004} (**Supplementary Table 3**). Thus, the most significant independently associated SNP was rs2071590, also known as *LTA*-293, located in the promoter region of the lymphotoxin-α (*LTA*) gene. Notably, the third most significant SNP in univariate analysis, rs3131628 (univariate $P = 0.0023$), was in linkage disequilibrium (LD) with *LTA*-293 ($r^2 = 0.5$), clearly identifying the *LTA* region as the principal target for high-resolution association mapping.

Over an interval of approximately 90 kb extending from *PPIAP9* to *NCR3* and overlapping the *TNF* and *LTA* genes, we analyzed 29 SNPs, eight of which were in strong LD and associated with leprosy at the 0.05 level (**Fig. 2a**). Seven of these SNPs, including *LTA*-293, belonged to the same bin (as defined in Methods), whereas the eighth (rs3131628) had an r^2 value of 0.5 with *LTA*-293 (**Fig. 2a**). The remaining SNPs were also in strong LD, with more than half (13/21) grouped into four additional bins (**Fig. 2** and **Supplementary Table 4**). The evidence of association with leprosy remained strongest for *LTA*-293, assuming dominant inheritance of the 'T' risk allele ($P_{dom} = 0.0009$; odds ratio (OR) for TT/CT versus CC = 1.97 (95% c.i., 1.30-2.99)). However, exhaustive multivariate conditional logistic regression showed that each of the SNPs in the *LTA*-293 bin was sufficient to account for the observed association of the *LTA* region with leprosy. Therefore, it is not possible to identify the susceptibility variant within the *LTA*-293 bin based on statistical considerations alone.

To more accurately define the boundaries of the *LTA*-293 bin, we first searched the four populations of HapMap release 20 to identify additional SNPs that were in $r^2 >$

0.8 with any of the SNPs in the *LTA*-293 bin. Although we used a 1-Mb window around *LTA*, only four SNPs met this criterion, all of which were located in the intergenic region between *NFKBIL1* and *LTA* and have no known functional role. To delineate the boundaries of the *LTA*-293 bin in our Vietnamese sample, we selected 31 SNPs located on the 31.48 Mb to 32.30 Mb chromosomal segment that, in combination with all previously genotyped HapMap SNPs, capture > 75% of the non-single SNP bins using a 0.8 r^2 threshold (**Supplementary Table 5**). This interval (that is, *MICA*→*NOTCH4*) spans the entire class III region in addition to the class I region centromeric to *HLA-B*. The 31 SNPs were genotyped in 54 unrelated Vietnamese subjects, and none of the 29 informative SNPs displayed an $r^2 > 0.25$ with SNPs in the *LTA*-293 bin. Finally, we also determined HLA class I genotypes in 37 out of the 54 aforementioned subjects, given the close proximity of the *HLA-B* and *HLA-C* loci to *LTA* (200 kb and 300 kb, respectively). Neither the *HLA-A*, the *HLA-B* nor the *HLA-C* alleles resulted in an $r^2 > 0.25$ with *LTA*+80 (**Supplementary Table 6**), confirming the independence of the *LTA* association from class I alleles. These results, which are entirely consistent with the HapMap LD data in all four populations, argue against the extension of the *LTA*-293 bin beyond a 37-kb interval located in the telomeric end of the HLA class III region. Notably, *LTA*+80 (P_{dom} = 0.007; OR for AA/AC versus CC = 1.74 (95% c.i., 1.16-2.60)) is the only SNP in this bin previously shown to have an independent biological function {Knight, Keating et al. 2004}; therefore, it appears to be the most obvious candidate for the susceptibility variant.

To replicate the association of *LTA*-293 bin SNPs with leprosy, we enrolled an additional sample of 364 unrelated individuals with leprosy and 371 unrelated controls from northern India. We tested association with leprosy for the six SNPs at the *LTA* locus (*LTA*-294, *LTA*-293, *LTA*+10, *LTA*+80, *LTA*+252, and *LTA*+368; **Table**

1). Univariate logistic regression indicated that only SNP *LTA*-294 was significantly associated (P_{dom} = 0.0004) with leprosy (**Table 1**). However, multivariate analysis showed that when *LTA*-294 was included in the model, the *LTA*+80 SNP became significant (P_{dom} = 0.026 with allele A being at risk), while *LTA*-293 remained non-significant (P_{dom} = 0.18). These discrepant results between *LTA*+80 and *LTA*-293 were explained by different LD structures between the Indian and the Vietnamese populations (**Table 1**). In particular, in the Indian population, *LTA*-293 did not belong to the same LD bin as *LTA*+80 (r^2 = 0.45) and therefore could be ruled out as a leprosy risk factor.

To follow-up on the observation that *LTA*+80 and *LTA*-294 were independently associated with leprosy susceptibility only among the Indians, we investigated whether an unknown SNP in LD with *LTA*-294 in the Indian population but not in the Vietnamese population could have been missed in our first analysis. We searched HapMap release 20, and we identified four additional SNPs in strong LD ($r^2 > 0.8$) with *LTA*-294 in at least one of the HapMap populations (**Supplementary Table 4**). We genotyped these SNPs in the Vietnamese and Indian samples and observed that three belonged to the same bin as *LTA*-294 in both populations. In the Vietnamese population, none of these four SNPs was associated with leprosy (**Fig. 2a**). In the Indian sample, the association with leprosy was the strongest for the A allele of marker rs2259435 (alias: *MCCD1-NS*), a non-synonymous SNP located in the first exon of the mitochondrial coiled-coil domain 1 (*MCCD1)* gene (P_{add} = 0.0002; OR for AA versus AG = OR for AG versus GG = 1.59 (95% c.i., 1.24-2.04)). In multivariate analysis, the best-fitting model included both *MCCD1-NS* and *LTA*+80 (P = 0.00009 for the global model; **Supplementary Table 7**). With adjustment for *MCCD1-NS*, the *LTA*+80 A allele was found to be significantly more frequent among cases, whereas

LTA-293 again was not associated with leprosy (**Table 1**). Thus, by taking into account the *MCCD1-NS* effect, we were able to detect and replicate association of the *LTA*+80 variant with leprosy in the Indian population (P_{dom} = 0.01; OR for AA/AC versus CC = 1.60 (95% c.i., 1.10-2.33)) with the same risk allele as the one observed in the Vietnamese families.

We then enrolled a third sample of 209 unrelated individuals of European ancestry with leprosy and 192 healthy controls, matched based on self-reported ethnicity, from the city of Rio de Janeiro in Brazil. We studied the six SNPs in the *LTA* promoter region plus six additional SNPs in strong LD with either the *LTA*+80 or the *LTA*-294 marker. None of these SNPs was significantly associated with leprosy in the Brazilian sample in either univariate or multivariate analysis, despite this sample having a posterior power of 84% to detect an effect of the *LTA*+80 A-risk allele assuming a type I error of 0.05 (based on the observed minor allele frequency (MAF) among the unaffected Brazilians of 0.35 and assuming an OR of 1.74). We hypothesized that the association of *LTA*+80 with leprosy might depend on sampling bias, owing to the family-based design of the Vietnamese study. Indeed, the median age at diagnosis was much lower ($P < 10^{-6}$) for affected individuals in the Vietnamese sample (16 years) than for those in the Indian (31 years) and Brazilian (38 years) samples.

Therefore, we carried out further analyses in which we stratified individuals based on age at leprosy diagnosis. The proportion of affected individuals in the four age groups considered is shown in **Figure 3**. In all three ethnic groups studied, we observed that the younger the individual, the stronger the association of *LTA*+80 with leprosy (**Fig. 3**). Notably, the association in the Vietnamese sample related almost entirely to individuals diagnosed before the age of 16 years (P = 0.00004; OR = 5.76 (95% c.i., 2.25-14.78)), and there was significant evidence for heterogeneity of the *LTA*+80

effect between individuals younger than 16 years and those older than 16 years (P = 0.00054; **Fig. 3**). Although not statistically significant (P = 0.07), we also observed a clear enrichment of the *LTA*+80 A allele in the youngest Brazilian age group (16-25 years) with an OR for AA/AC versus CC of 2.76 (95% c.i., 0.74-10.32). In addition, this trend toward a decrease in OR with increasing age was significant in the Brazilian sample (P = 0.04). Finally, we found the same age-at-diagnosis effect on the association between *LTA*+80 and leprosy in the Indian sample, with an OR for AA/AC versus CC of 2.95 (95% c.i., 1.32-6.58) *(*P_{dom} = 0.006) for the 16- to 25-year-old age group versus 1.60 (95% c.i., 1.10-2.33) for the overall sample. Evidence for genetic heterogeneity between the youngest age group (16-25 years) and the older patients (> 25 years) was highly significant (P = 0.003).

These results strongly indicated that the effect of *LTA*+80 on leprosy risk was age dependent. We enrolled a fourth independent sample of 104 simplex leprosy families from Vietnam to validate this conclusion. The median age at diagnosis of the affected individuals in these families was 19 years (the age distribution is shown in **Fig. 3**). Overall, again we observed a significant association between the *LTA*+80 A allele and leprosy risk (P_{dom} = 0.003; OR for AA/AC versus CC = 2.34 (95% c.i., 1.27-4.31)). As observed in the first 194 Vietnamese families studied, the association was strongest for the youngest individuals (age at diagnosis, < 16 years) with an OR of 5.31 (95% c.i., 1.19-23.60), and there was significant evidence (P = 0.04) for heterogeneity of the *LTA*+80 effect between affected individuals younger than 16 years and those older than 16 years (**Fig. 3**, 'Vietnam 2' sample*).* Finally, when we combined both Vietnamese samples, the evidence for association of *LTA*+80A with leprosy in the overall sample and in the group of individuals diagnosed before the age

of 16 years was very strong, with ORs of 2.11 (95% c.i., 1.48-3.01; $P = 0.000024$) and 5.63 (95% c.i., 2.54-12.49; $P = 0.0000004$), respectively.

HLA class II *DRB1* alleles have been repeatedly identified as leprosy susceptibility factors {de Vries, Mehra et al. 1980; Zerva, Cizman et al. 1996; Casanova and Abel 2002}, and the *HLA-DRB1* locus is located about 1 Mb centromeric to *LTA*. Therefore, we therefore investigated in more detail whether the observed associations of leprosy with *LTA*+80 could be due to long-range LD with HLA class II *DRB1* alleles. We first determined *DRB1* genotypes in our initial Vietnamese family sample, and multivariate analysis showed that the effect of *LTA*+80 remained unchanged when we included *DRB1* genotypes in the model both in the overall sample ($P = 0.008$; OR for *LTA*+80 AA/AC versus CC =1.75 (95% c.i., 1.15-1.62)) and the early onset age group ($P = 0.000005$; OR = 6.00 (95% c.i., 2.29-15.7)). Consistent with this result, we did not detect any $r^2 > 0.1$ between *LTA*+80 and *DRB1* alleles in this Vietnamese sample. Next, we genotyped the *HLA-DRB1* locus in unrelated Indian subjects, and again did not find any $r^2 > 0.1$ between alleles of this locus and *LTA*+80. These results demonstrate the independence of the *LTA*+80 effect from the known *DRB1* leprosy susceptibility gene in the Vietnamese and Indian samples.

Following on the heels of the identification of *PARK2* and *PACRG* variants as major leprosy risk factors {Mira, Alcais et al. 2004; Alcais, Mira et al. 2005}, the present results define a new genetic element affecting leprosy susceptibility and shed further light on the complex genetic interplay governing susceptibility to a common infectious disease. We conclude that *LTA*+80 is not the only leprosy risk factor in the 6p21 region. First, the *LTA*+80 effect is clearly independent of *DRB1*, repeatedly identified as an important locus in leprosy control. Second, we detected significant association of the *MCCD1-NS* bin with leprosy in the Indian individuals. Although

we did not observe this association in the Vietnamese and Brazilian samples, this finding suggests that there may be additional leprosy susceptibility genes in the *HLA* region. Finally, it is clear from our results that *LTA*+80 did not have a substantial role in leprosy susceptibility in individuals over the age of 25 years. Additional investigation of candidate genes in the target interval should therefore be carried out, carefully taking the affected individuals' age into account.

The *LTA*+80 polymorphism is an attractive candidate leprosy susceptibility factor, regardless of the other susceptibility variants that may exist. It was localized to a regulatory E2-box motif (CAGCTG) with a 1-bp mismatch (CAGCAG). The C allele of *LTA*+80 results in an additional mismatch (CCGCAG), preventing binding of the activated B-cell factor-1 transcriptional repressor. Therefore, the A allele is associated with significantly lower levels of expression of a downstream reporter than the C allele {Knight, Keating et al. 2004}. Our findings thus indicate that low levels of LTA production are associated with a higher risk of leprosy. This result is consistent with studies in animal models linking disruption of the LTA signaling pathway with an increase in susceptibility to intracellular pathogens {Jacobs, Brown et al. 2000; Bopst, Garcia et al. 2001; Ware 2005}. Irradiated mice with immune system reconstitution from LTA-deficient bone marrow have been shown to be extremely susceptible to virulent *Mycobacterium tuberculosis*, owing to the impaired colocalization of T cells with macrophages in lung granulomas {Roach, Briscoe et al. 2001}. The association of the *LTA*+80 suppressive A allele with a higher risk of leprosy may therefore be mediated by poor lymphocyte recruitment to sites of infection.

Our results obtained in Vietnamese and Indian samples demonstrate that *LTA*+80 is an important risk factor for leprosy that acts in a highly age-dependent manner. This study also provides an instructive example of how a linkage peak in a complex

disease may reflect a complex underlying structure of susceptibility variants. Moreover, the results obtained highlight some of the pitfalls encountered when replicating genetic effects in independent samples and the importance of considering population-specific LD patterns and covariates properly in replication studies. In particular, our study illustrates that genotyping in replication samples cannot be confined to the one SNP with strongest evidence for association in the initial study but needs to encompass all SNPs in the initially associated bin. Overall, our observations should prove useful for the design and interpretation of allelic association replication studies in other complex diseases.

Methods

Affected individuals and controls

We enrolled a total of 2,027 subjects in this study. The 194 Vietnamese simplex leprosy families (Vietnam 1) have been described elsewhere {Mira, Alcais et al. 2003}. The additional 104 Vietnamese families (Vietnam 2) were identified from the records available at the Dermato-Venereology Hospital in Ho Chi Minh City. No particular efforts were made to enroll younger individuals, and the criteria for enrollment were identical to those used for the Vietnam 1 sample: that is, both parents had to be available for genetic analysis in all families of the Vietnam 1 and Vietnam 2 samples. Overall, the mean age at diagnosis (± s.d.) was 19.4 years ± 8.3 years in individuals in the Vietnam 1 sample and 20.6 years ± 7.8 years in individuals in the Vietnam 2 sample. In north India, 364 individuals with leprosy (33.3 years ± 13.3 years) were enrolled from records at the Central JALMA Institute for Leprosy and Other Mycobacterial Diseases at Agra. As a control group, 370 healthy north Indian individuals (32.1 years ± 10.9 years) with no documented history of chronic infectious or inflammatory diseases were enrolled from blood donor clinics. In Brazil, 207 individuals with leprosy (38.8 years ± 16.8 years) of self-reported European ancestry were recruited from the Leprosy Outpatient Clinic at the Oswaldo Cruz Institute in Rio de Janeiro, Brazil. As a control group, 192 control subjects matched based on self-reported ethnicity (33.1 years ± 8.8 years) with no history of chronic infectious, inflammatory or autoimmune disease were recruited from the same geographic area of Rio de Janeiro. The study was approved by institutional review boards and health authorities in Ho Chi Minh City, Vietnam; the Oswaldo Cruz Institute, Rio de Janeiro, Brazil; the All India Institute of Medical Sciences, New Delhi, India and the McGill

University Health Centre, Montreal, and written informed consent was obtained from all participants.

Phenotype definition

The diagnosis of the Vietnamese individuals and the definition of clinical subtypes have been described elsewhere {Mira, Alcais et al. 2003}. The Brazilian and Indian cases were diagnosed and classified on the basis of clinical and histological criteria {Ridley and Jopling 1966}. As in both Vietnamese family samples, the two case-control samples included a balanced mixture of paucibacillary and multibacillary cases. However, the phenotype studied here was leprosy *per se* (leprosy independent of specific clinical manifestations).

Genotyping methods

The 418 SNPs spanning the 10.4-Mb targeted interval on chromosome region 6p21 were selected on the basis of their proximity to or location within known genes in the interval (**Supplementary Table 1**). All SNPs were genotyped on the ultra-high throughput Illumina platform. This platform uses the GoldenGate assay followed by a bead-based technology to resolve individual SNP genotypes {Fan, Oliphant et al. 2003}. A total of 54 SNPs could not be genotyped with this assay, and one SNP could not be placed unambiguously on the sequence map. We excluded 56 of the remaining 363 SNPs from the analysis for the following reasons: 12 showed deviations ($P < 0.05$) from Hardy-Weinberg equilibrium (HWE) among parents, and 44 were noninformative or had a MAF < 5%.

The additional 25 informative SNPs spanning the refined 90-kb target interval in the MHC class III region (**Supplementary Table 4**), as well as the panel of 31 SNPs used

to study LD in the vicinity of the *LTA* gene (**Supplementary Table 5**) were selected based on information regarding bin structure and allelic frequencies publicly available from the International HapMap project (see URL section below). These SNPs were genotyped on one or several of the following platforms: (i) genotyping by direct sequencing on an ABI PRISM 3100 genetic analyzer; (ii) genotyping on the high-throughput GenomeLab SNPstream platform, which uses a single-base pair extension (SBE) method to incorporate fluorescently labeled terminal nucleotides, which are then detected by a specialized imager {Bell, Chaturvedi et al. 2002}; (iii) genotyping on the high-throughput SEQUENOM MassARRAY platform, which uses the iPLEX assay to incorporate mass-modified terminal nucleotides in the SBE step that are then detected by matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectrometry {Griffin and Smith 2000} and (iv) genotyping by the polarized fluorescence TaqMan assay {Lee, Connell et al. 1993}. Five SNPs, including *LTA*+80, were genotyped by two independent methods for all Vietnam 1 samples, and the few individuals (< 0.5%) with discrepant genotypes were eliminated from subsequent analyses. Genotyping of the HLA *DRB1* locus was performed using standard methods {Olerup and Zetterquist 1992}. *HLA-A*, *HLA-B* and *HLA-C* genotyping was performed using a PCR reverse sequence-specific oligonucleotide (PCR-SSO) probe hybridization technique as recommended by the manufacturer (InnoLipa-HLA, Innogenetics).

Statistical methods.

We estimated pairwise LD between SNPs by determining r^2 from parental data for Vietnamese samples and from unaffected controls for the Indian and the Brazilian samples, using the algorithm implemented in Haploview software {Barrett, Fry et al. 2005}. Bins of SNPs were constructed on the basis of the pairwise r^2 between the

SNPs and were defined as a not necessarily contiguous set of SNPs in which at least one SNP has $r^2 > 0.8$ with all the other SNPs of the bin {Hinds, Stuve et al. 2005}.

As the Vietnamese samples consisted of trios with no missing parental data, family-based association studies were performed principally by a classical transmission disequilibrium test, as implemented in FBAT software {Horvath, Xu et al. 2001}. Alleles for which there was evidence of association were also analyzed by conditional logistic regression, as previously described {Schaid and Rowland 1998}. This approach permitted estimation of OR and made it possible to carry out multivariate stepwise regression tests and tests for heterogeneity. We tested for heterogeneity according to age in the Vietnamese samples by carrying out the analysis on the combined sample (for example, 194 trios in the Vietnam 1 sample) and on the following two subsamples: families in which affected individuals were diagnosed before 16 years of age (for example, 86 trios) and families in which affected individuals were diagnosed at or over the age of 16 years (for example, 108 trios). Under the hypothesis of homogeneity of association, twice the difference between the likelihood of the whole sample and the summed likelihoods of the two subsamples is asymptotically distributed as a χ^2 with 1 degree of freedom. Conditional logistic analysis was performed with the PHREG procedure implemented in SAS software v.8.2.

We carried out a population-based association study in the Indian and Brazilian samples, using classical multivariate logistic regression techniques as implemented in the LOGISTIC procedure of SAS software v.8.2. All analyses were adjusted for sex, which has been identified as a classical risk factor for leprosy. To investigate whether the association of leprosy with the *LTA*+80 A allele decreased with age in the Brazilian sample, we tested in the logistic model if the regression coefficient

corresponding to an interaction between *LTA*+80 (coded as 0 or 1, for CC or AC/AA subjects, respectively) and the age classes (coded as 0, 1 or 2 for age ranges 16-25 years, 26-36 years or > 35 years, respectively) was significantly < 0.

We tested for differences in median ages across three populations (Vietnam 1, Brazil and India), using the non parametric Brown-Mood test, as implemented in the NPAR1WAY procedure of SAS v.8.2. This test is equivalent to a one-way analysis of variance with median scores {Conover 1999}.

URLs

HapMap project: http://www.hapmap.org/

Acknowledgments

We thank all affected individuals, their families and population controls for participating in this study. We thank T.J. Hudson, P. Lepage, E.N. Sarno and J.L. Casanova for helpful comments and discussions; A. Montpetit and A. Belisle for assistance with high-throughput genotyping; S. Caillat-Zucman for help in HLA genotyping and L. Simkin for technical assistance. This study was supported by a grant from the Canadian Institutes of Health Research (CIHR) to E.S. A. Alter holds a graduate studentship from the Natural Science and Engineering Research Council of Canada (NSERC). A Alter and M.O. are supported by the CIHR Strategic Training Centre in the Integrative Biology of Infectious Diseases and Autoimmunity. G.A. is supported by Fondation pour la Recherche Médicale. E.S. is a Chercheur National of the Fonds de Recherche en Santé du Québec and an International Research Scholar of the Howard Hughes Medical Institute. A. Alcaïs and L.A. are supported by the Assistance Publique-Hopitaux de Paris, Programme de Recherche Fondamentale en Microbiologie Maladies Infectieuses et Parasitaires and the Agence Nationale de la Recherche of the Ministère Français de l'Education Nationale de la Recherche et de la Technologie. Requests for materials should be addressed L.A. (abel@necker.fr) or E.S. (erwin.schurr@mcgill.ca).

Figure legend

Figure 1 | High-density SNP association scan of a 10.4- Mb region on chromosome 6p21 in 194 Vietnamese trios

Evidence for association with leprosy of 307 SNPs is plotted as $-\log_{10}P$ on the *y*-axis. The location of SNPs, in Mb, on the sequence map (build 36) is indicated on the *x*-axis. The borders of the HLA class II and class III regions and of the studied non-HLA region toward the centromere are indicated by horizontal lines. The thin dotted line indicates the $P = 0.01$ significance threshold. The six SNPs associated with leprosy at $P < 0.01$ are indicated by letters A-F. SNP A corresponds to rs2071590 (also known as *LTA*-293), and SNPs B-F correspond to rs3128961, rs3131628, rs707928, rs805305 and rs937662, respectively. We tested three different genetic models for each SNP (i.e. additive, fully dominant and fully recessive), and the *P* values for the best model are reported here.

Figure 2 | Fine linkage disequilibrium mapping of a 90-kb interval in the HLA class III region in 194 Vietnamese trios

(a) Evidence for the association of 33 SNPs spanning a 90-kb interval in the HLA class III region with leprosy is expressed as $-\log_{10} P$ and plotted against SNP position in Mb. The probability threshold $P = 0.01$ is indicated by the dotted line. The four SNPs represented by triangles were selected based on their linkage disequilibrium with *LTA*-294 in at least one HapMap project population. SNPs identical in color belong to the same bin set (as defined in the Methods). SNPs in black do not belong to any bin set. The SNPs from the primary scan are represented by a filled circle. *LTA*+80 is represented by the red square surrounded by a black oval. The designations and locations of the ten known genes in the 90-kb interval are shown. A

linear view of the bin structure consistent with the color scheme used above is shown immediately below. The dotted line in the red bin set indicates a pairwise $r^2 = 0.5$. Circles: SNPs from primary scan; triangles: HapMap SNPs in LD with *LTA*-294; squares: SNPs selected for fine-mapping of the 90-kb interval. (b) Expanded view of the 5' region of *LTA*. The proximal promoter and the intronic regions are represented by solid lines. The 5' UTR is represented by open boxes and translated exon regions are represented by solid boxes. The relative locations of seven *LTA* SNPs are shown (*LTA*-294, *LTA*-293, *LTA*+10, *LTA*+80, *LTA*+252, *LTA*+368 and rs1041981). The color of each SNP is consistent with the bin set to which it belongs. SNPs A and C correspond to rs2071590 and rs3131628, respectively.

Figure 3 | Age-dependence of the *LTA*+80 A allele effect on leprosy susceptibility

A total of 869 individuals with leprosy from four independent samples (two family-based samples from Vietnam and two population-based case-control samples from India and Brazil) were grouped into four age-at-diagnosis classes (0-15 years, 16-25 years, 26-35 years and older than 35 years). The proportion of cases in each age group is indicated in the lower part of the graph. The total number of cases was 194 for the Vietnam 1 sample, 104 for the Vietnam 2 sample, 207 for the Brazilian sample and 364 for the Indian sample. In the upper part of the figure, the ORs of leprosy for *LTA*+80 AA/AC versus CC subjects (circles) and their 95% c.i. values are plotted on a logarithmic scale for each age group in each ethnic group. In the Indian sample, the ORs are adjusted for the *MCCD1-NS* effect. Red: Vietnam 1; green: Vietnam 2; blue: India; yellow: Brazil). ORs were not calculated for age groups including fewer than 15 affected individuals.

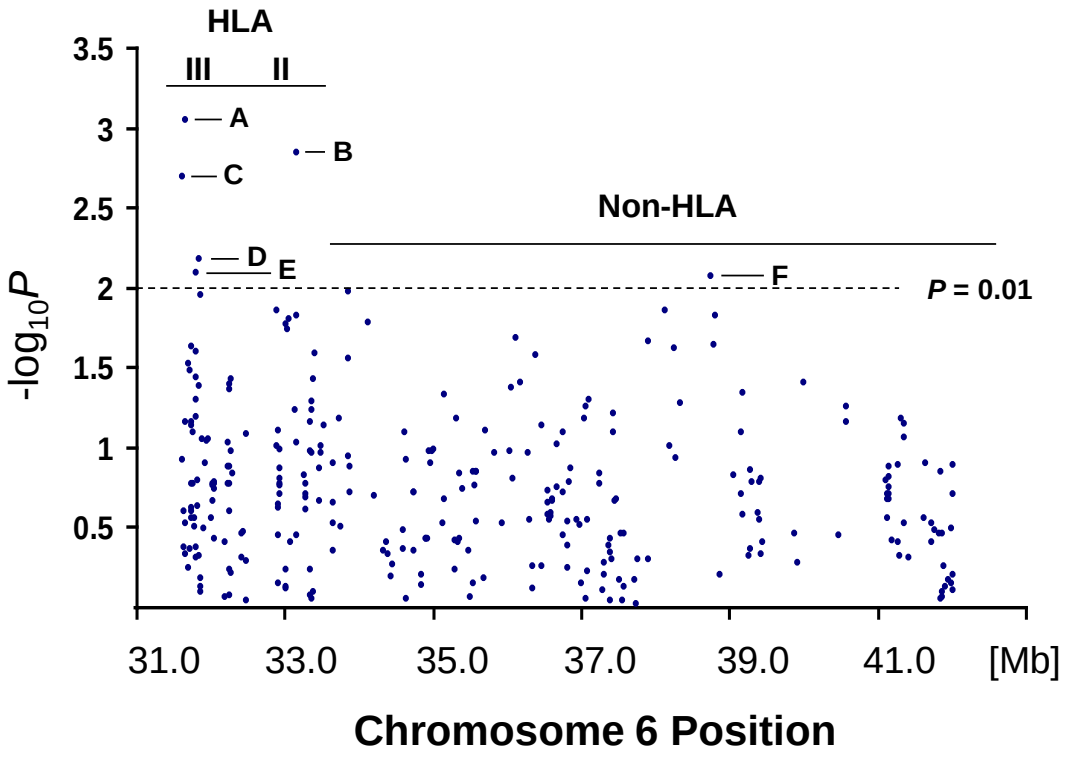

HLA
III
II
A
B
C
D
E
F
Non-HLA
P = 0.01
-log10P
3.5
3
2.5
2
1.5
1
0.5
31.0
33.0
35.0
37.0
39.0
41.0
[Mb]
Chromosome 6 Position

Figure 1

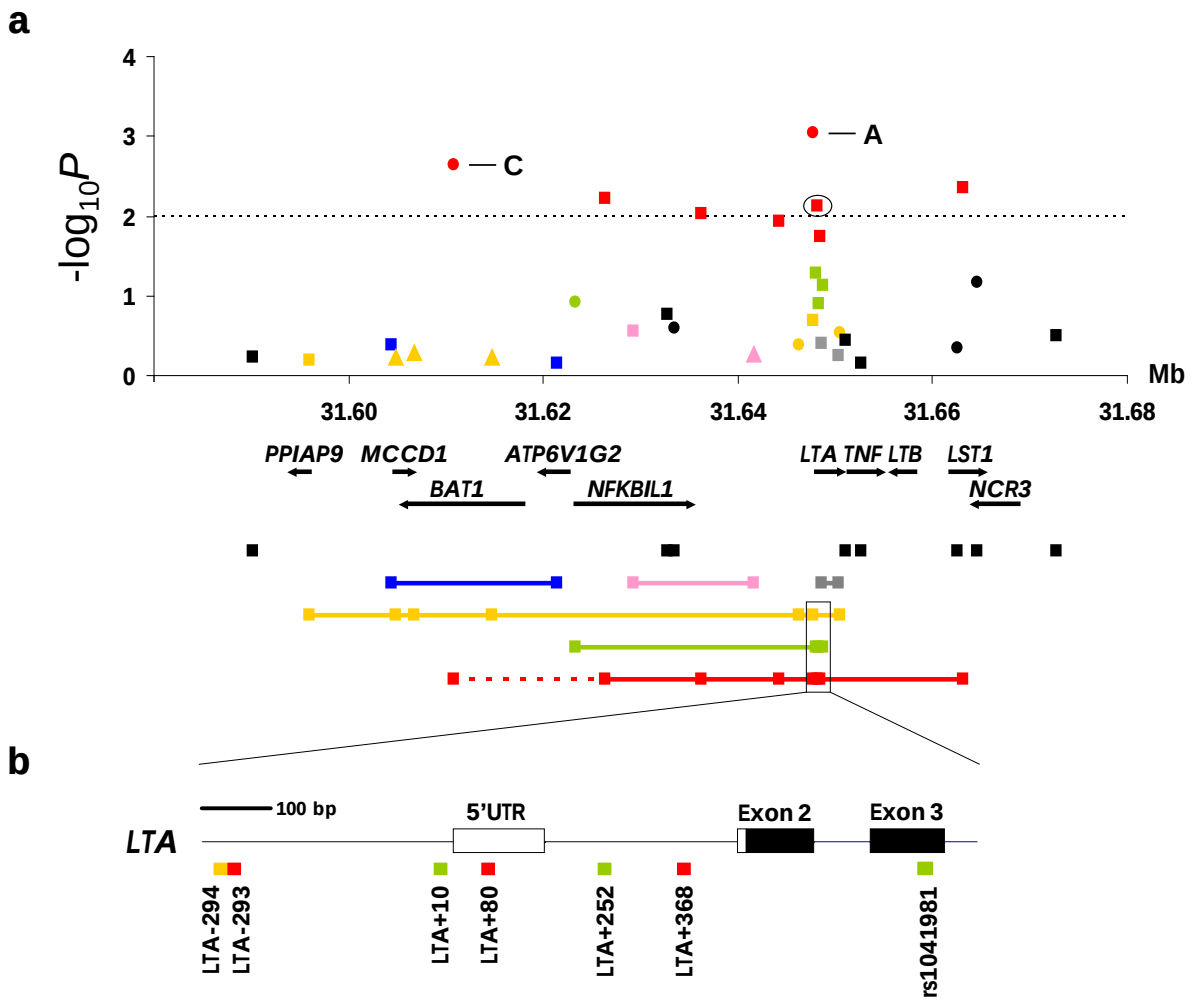

a
4
3
2
1
0
-log10P
C
A
Mb
31.60
31.62
31.64
31.66
31.68
PPIAP9
MCCD1
ATP6V1G2
LTA TNF LTB
LST1
BAT1
NFKBIL1
NCR3
b
100 bp
LTA
5'UTR
Exon 2
Exon 3
LTA-294
LTA-293
LTA+10
LTA+80
LTA+252
LTA+368
rs1041981

Figure 2

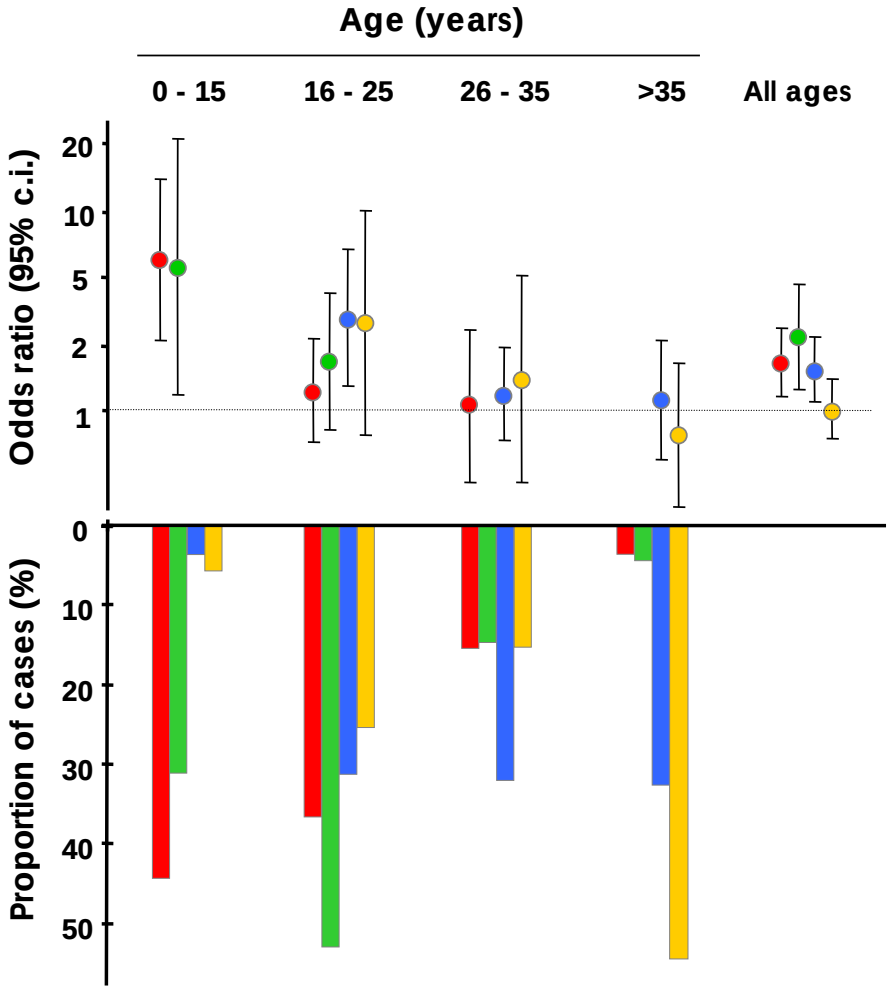

Age (years)
0 - 15
16 - 25
26 - 35
>35
All ages
Odds ratio (95% c.i.)
20
10
5
2
1
0
Proportion of cases (%)
10
20
30
40
50

Figure 3

Table 1. Univariate and multivariate analyses of *MCCD1-NS* and six *LTA* SNPs in the Vietnamese and Indian samples

	MCCD1-NS **rs2259435**	***LTA*-294** **rs2844482**	***LTA*-293** **rs2071950**	***LTA*+10** **rs1800683**	***LTA*+80** **rs2239704**	***LTA*+252** **rs909253**	***LTA*+368** **rs746868**
Vietnamese							
Bin structure							
MAF§	0.16	0.16	0.26	0.50	0.30	0.49	0.30
Univariate OR (95% CI) ***P***	ns	ns	1.97 (1.30-2.99) 0.0009	ns	1.74 (1.16-2.60) 0.007	ns	1.63 (1.09-2.43) 0.02
Multivariate OR (95% CI) ***P***	ns	ns	1.97 (1.30-2.99) 0.0009	ns	ns∗	ns	ns∗
Indian							
Bin structure							
MAF§	0.23	0.22	0.26	0.28	0.42	0.28	0.41
Univariate OR (95% CI) ***P***	1.87 (1.37-2.57) 0.00009	1.78 (1.29-2.45) 0.0004	ns	ns	ns	ns	ns
Multivariate OR (95% CI) ***P***	1.82 (1.38-2.41) 0.00003	ns*	ns	ns	1.60 (1.10-2.33) 0.01	ns	ns *

∗ Non significant when the first SNP of the bin is included in the multivariate analysis. Note that any SNP of the bin is sufficient to capture the observed association between that bin and leprosy

§ MAF (Minor Allelic Frequency) corresponds to the frequency of the risk allele for each of the associated SNPs

Supplementary material

Supplementary tables 1 and **2** are too large for insertion, please see:

http://www.nature.com/ng/journal/v39/n4/suppinfo/ng2000_S1.html

Supplementary Table 3:
rs2071590, rs3128961, rs937662 and rs707928 are associated with leprosy independently of the two *PARK2*/*PACRG* SNPs.

Model	-2Ln likelihood of the model	Test	Number of df [1]	*P* value
rs2071590				
M1: Two *PARK2*/*PACRG* SNPs	302.6			
M2: M1 + rs2071590	295.1	M2 vs. M1 [2]	1df	0.006
M3: M2 + interactions	294	M3 vs. M2 [3]	2df	0.577
rs3128961				
M1: Two *PARK2*/*PACRG* SNPs	309.7			
M2: M1 + rs3128961	300.5	M2 vs. M1	1df	0.002
M3: M2 + interactions	298.1	M3 vs. M2	2df	0.301
rs937662				
M1: Two *PARK2*/*PACRG* SNPs	298.6			
M2: M1 + rs937662	294.1	M2 vs. M1	1df	0.034
M3: M2 + interactions	293.8	M3 vs. M2	2df	0.861
rs707928				
M1: Two *PARK2*/*PACRG* SNPs	307.3			
M2: M1 + rs707928	301.6	M2 vs. M1	1df	0.017
M3: M2 + interactions	301.4	M3 vs. M2	2df	0.905

[1] *df = degree of freedom*

[2] *M2 vs. M1 tests the effect of the LTA SNPs (rs2071590, rs3128961, rs937662 or rs707928) in the presence of the two PARK2/PACRG SNPs*

[3] *M3 vs. M2 tests the interactions between the LTA SNP and each of the two PARK2/PACRG SNPs in the presence of their main effects*

Differences among M1 (-2Ln likelihoods) are due to differences in the number of missing genotypes for each SNP tested.

Supplementary Table 4:
Bin structure of 33 informative SNPs spanning the 90 kb target interval in the HLA class III region analyzed in the Vietnamese sample.

Bin[1]	SNP_ID	Alias	Position[2]	MAF[3]	*P value*
	rs2246986		31590182	0.08	0.6003
	rs2516489		31596017	0.17	0.6479
	rs2516483		31604455	0.11	0.4203
	rs2259435	*MCCD1-NS*	31604894	0.16	0.5959
	rs2516478		31606716	0.16	0.5151
	rs3131628		31610746	0.15	0.0023
	rs2523512		31614780	0.16	0.5959
	rs2239705		31621381	0.11	0.7087
	rs2071592		31623319	0.48	0.1206
	rs2523500		31626333	0.26	0.0061
	rs2255798		31629281	0.11	0.2770
	rs2857605		31632830	0.14	0.1684
	rs2230365		31633427	0.15	0.2519
	rs2516479		31636305	0.30	0.0092
	rs2857708		31641585	0.11	0.5504
	rs2844484		31644203	0.29	0.0118
	rs2009658		31646223	0.16	0.4158
	rs2844482	*LTA*-294	31647746	0.16	0.2014
	rs2071590	*LTA*-293	31647747	0.26	0.0009
	rs1800683	*LTA*+10	31648050	0.50	0.0522
	rs2239704	*LTA*+80	31648120	0.30	0.0076
	rs909253	*LTA*+252	31648292	0.49	0.1259
	rs746868	*LTA*+368	31648408	0.30	0.0182
	rs2857713		31648535	0.21	0.3893
	rs1041981		31648763	0.49	0.0752
	rs1799964	*TNF*-1031	31650287	0.19	0.5563
	rs1800630	*TNF*-863	31650455	0.16	0.2971
	rs1800629	*TNF*-308	31651010	0.08	0.3652
	rs3093664		31652621	0.06	0.7082
	rs7769073		31662586	0.06	0.4609
	rs2256965		31663109	0.28	0.0045
	rs1052248		31664560	0.22	0.0688
	rs2844480		31672800	0.17	0.3181

[1]SNPs identical in color belong to the same bin set (colors are identical to those used for **figure 2**)
this SNP displays a r^2 = 0.5 with SNPs belonging to the red bin
SNPs in black do not belong to any bin set (i.e., they are singletons)
[2]based on NCBI entrez SNP build 125
[3]minor allele frequency

Note: The 33 SNPs include 8 SNPs belonging to the primary map (SNP_ID in italic bold), 4 additional SNPs in strong LD (r^2 > 0.8) with LTA-294 (orange bin), and 21 approximately evenly-spaced gene-centric SNPs

Supplementary Table 5:
Characterisation of 31 SNPs used in the linkage disequilibrium mapping of a 0.82 Mb interval covering the whole HLA Class III region in 54 unrelated Vietnamese individuals

pairwise r^2 with rs2239704	SNP rs#	gene	position [1]	frequent allele	MAF [2]	failed (0=no;	non- (0=no; 1=yes)	MAF <= 0.05 (0=no; 1=yes)	Hardy_Weinberg (0=no; 1=yes)	*P* value for departure Hardy Weinberg
0.01	6934097	*MICA*	31488724	A	0.19	0	0	0	0	
0.08	9357133	*MICA*	31488876	T	0.21	0	0	0	0	
0.00	10223421		31498034	T	0.33	0	0	0	0	
0.08	4713466		31543848	T	0.21	0	0	0	0	
0.10	2844508		31544478	C	0.06	0	0	0	0	
0.05	3828886		31548531	G	0.35	0	0	0	0	
0.19	3094009		31565157	A	0.39	0	0	0	0	
	2534654	*MICB*	31580925	C	0.42	0	0	0	1	6.70E-08
0.18	3131636	*MICB*	31584073	T	0.12	0	0	0	0	
0.23	933208	*BAT1*	31614627	C	0.31	0	0	0	0	
0.01	3117572	*MSH5*	31825671	A	0.27	0	0	0	0	
0.01	4713488	*VARS*	31870290	C	0.05	0	0	1	0	
0.00	2227956	*HSPA1L*	31886251	C	0.23	0	0	0	0	
0.02	3130481	*SLC44A4*	31947735	G	0.37	0	0	0	0	
0.00	589428	*EHMT2*	31956199	A	0.21	0	0	0	0	
0.12	2844458	*EHMT2*	31959448	T	0.29	0	0	0	0	
0.03	7746553	*C2*	32003952	G	0.15	0	0	0	0	
0.05	4151657	*CFB*	32025519	C	0.15	0	0	0	0	
0.07	1009382	*TNXB*	32134085	G	0.41	0	0	0	0	
0.21	408359	*AGPAT1*	32249861	T	0.21	0	0	0	0	
0.02	9469089	*RNF5*	32254635	C	0.07	0	0	0	0	
0.08	2070600	*AGER*	32259421	A	0.20	0	0	0	0	
0.02	204994	*PBX2*	32262976	A	0.16	0	0	0	0	
0.07	2856437	*PBX2*	32265342	T	0.19	0	0	0	0	
0.12	3132935	*NOTCH4*	32279053	G	0.27	0	0	0	0	
	404860	*NOTCH4*	32292323	C	0.49	0	0	0	1	0.0315
0.00	436388	*NOTCH4*	32294242	T	0.41	0	0	0	0	
0.05	394657	*NOTCH4*	32295001	G	0.22	0	0	0	0	
0.01	423023	*NOTCH4*	32296275	G	0.19	0	0	0	0	
0.10	443198	*NOTCH4*	32298384	T	0.42	0	0	0	0	
0.06	2854048	*NOTCH4*	32299096	T	0.12	0	0	0	0	

[1]based on NCBI entrez SNP build 125
[2]MAF = minor allele frequency

Note: ***These 31 SNPs were selected from HapMap data release 21 in order to capture >75% of the variability in the region from 31.48 Mb to 32.30 Mb (covering the whole HLA class III region) using a 0.8 r^2 threshold for bin definition. Note that this is an underestimate of the true coverage since non-HapMap SNPs could not be used in the estimation process. The next gene telomeric to MICA is HLA-B, which was investigated by direct genotyping (see supplementary table 6).***

Supplementary Table 6:
Characterisation of HLA class I alleles in 37 unrelated Vietnamese individuals

pairwise r^2 with rs2239704*	HLA allele(s)	allele frequency
	*HLA-A**01	0.013
0.016	*HLA-A**02, *0203, *0205	0.295
	*HLA-A**03	0.013
0.021	*HLA-A**11	0.192
0.014	*HLA-A**24, *2407, *2410	0.218
0.102	*HLA-A**29	0.051
	*HLA-A**30	0.038
	*HLA-A**31	0.013
	*HLA-A**32	0.013
0.096	*HLA-A**33	0.154
0.186	*HLA-B**07	0.090
	*HLA-B**08	0.013
0.029	*HLA-B**13	0.051
0.208	*HLA-B**15, *1502, *1525	0.282
	*HLA-B**18, *1802	0.026
	*HLA-B**27	0.013
	*HLA-B**35	0.038
	*HLA-B**37	0.013
0.004	*HLA-B**38, *3802	0.090
	*HLA-B**39	0.013
0.019	*HLA-B**40, *4006	0.090
	*HLA-B**44	0.013
0.044	*HLA-B**46	0.077
	*HLA-B**5001	0.013
	*HLA-B**51	0.013
	*HLA-B**54	0.013
	*HLA-B**55	0.026
	*HLA-B**5601, *5604	0.038
0.052	*HLA-B**58	0.090
0.087	*HLA-C**01	0.141
0.116	*HLA-C**03	0.179
0.013	*HLA-C**04	0.154
	*HLA-C**06	0.038
0.010	*HLA-C**07	0.244
0.087	*HLA-C**08	0.141
	*HLA-C**12	0.013
	*HLA-C**1202	0.013
0.157	*HLA-C**15	0.077

*pairwise r^2 was not calculated if the allele frequency was below 5%

Supplementary Table 7:
Multivariate analysis in the Indian sample.

All individuals

Model	-2Ln likelihood of the model	Test	Number of df[1]	*P* value
M1: Sex	885.6			
M2: M1 + MCCD1-NS	873.1	M2 vs. M1	1df	0.0004
M3: M2 + LTA+80	866.9	M3 vs. M2 [2]	1df	0.013
		M3 vs. M1 [3]	2df	0.00009

16-25 years age group

Model	-2Ln likelihood of the model	Test	Number of df[1]	*P* value
M1: Sex	255.5			
M2: M1 + MCCD1-NS	242.4	M2 vs. M1	1df	0.0003
M3: M2 + LTA+80	234.8	M3 vs. M2 [2]	1df	0.006
		M3 vs. M1 [3]	2df	0.00003

[1] *df = degree of freedom*

[2] *M3 vs. M2 tests the effect of the LTA+80 SNP (rs2239704) in the presence of the MCCD1-NS SNP (rs2259435)*

[3] *M3 vs. M1 tests the global effect of the LTA+80 SNP (rs2239704) + the MCCD1-NS SNP (rs2259435)*

References

Abel, L. and F. Demenais (1988). "Detection of major genes for susceptibility to leprosy and its subtypes in a Caribbean island: Desirade island." Am. J. Hum. Genet. **42**(2): 256-66.

Abel, L., D. L. Vu, et al. (1995). "Complex segregation analysis of leprosy in southern Vietnam." Genet. Epidemiol. **12**(1): 63-82.

Alcais, A., M. Mira, et al. (2005). "Genetic dissection of immunity in leprosy." Curr. Opin. Immunol. **17**(1): 44-48.

Barrett, J. C., B. Fry, et al. (2005). "Haploview: analysis and visualization of LD and haplotype maps." Bioinformatics **21**(2): 263-265.

Bell, P., S. Chaturvedi, et al. (2002). "SNPstream UHT: ultra-high throughput SNP genotyping for pharmacogenomics and drug discovery." Biotechniques **Suppl**: 70-2, 74, 76-7.

Bopst, M., I. Garcia, et al. (2001). "Differential effects of TNF and LTalpha in the host defense against M. bovis BCG." Eur. J. Immunol. **31**(6): 1935-43.

Casanova, J. L. and L. Abel (2002). "Genetic dissection of immunity to mycobacteria: the human model." Annu. Rev. Immunol. **20**: 581-620.

Conover, W. J. (1999). Practical Nonparametric Statistics New York, John Wiley & Sons.

de Vries, R. R., N. K. Mehra, et al. (1980). "HLA-linked control of susceptibility to tuberculoid leprosy and association with HLA-DR types." Tissue Antigens **16**(4): 294-304.

Dupuis, J. and D. Siegmund (1999). "Statistical methods for mapping quantitative trait loci from a dense set of markers." Genetics **151**(1): 373-86.

Fan, J., A. Oliphant, et al. (2003). "Highly parallel SNP genotyping." Cold Spring Harb. Symp. Quant. Biol. **68**: 69-78.

Griffin, T. J. and L. M. Smith (2000). "Single-nucleotide polymorphism analysis by MALDI-TOF mass spectrometry." Trends Biotechnol. **18**(2): 77-84.

Hinds, D. A., L. L. Stuve, et al. (2005). "Whole-Genome Patterns of Common DNA Variation in Three Human Populations." Science **307**(5712): 1072-1079.

Horvath, S., X. Xu, et al. (2001). "The family based association test method: strategies for studying general genotype--phenotype associations." Eur. J. Hum. Genet. **9**(4): 301-306.

Jacobs, M., N. Brown, et al. (2000). "Fatal Mycobacterium bovis BCG infection in TNF-LT-alpha-deficient mice." Clin. Immunol. **94**(3): 192-9.

Knight, J. C., B. J. Keating, et al. (2004). "Allele-specific repression of lymphotoxin-[alpha] by activated B cell factor-1." Nat. Genet. **36**(4): 394-399.

Lee, L. G., C. R. Connell, et al. (1993). "Allelic discrimination by nick-translation PCR with fluorogenic probes." Nucleic Acids Res. **21**(16): 3761-6.

Mira, M. T. (2006). "Genetic host resistance and susceptibility to leprosy." Microbes Infect. **8**(4): 1124-1131.

Mira, M. T., A. Alcais, et al. (2004). "Susceptibility to leprosy is associated with PARK2 and PACRG." Nature **427**(6975): 636-640.

Mira, M. T., A. Alcais, et al. (2003). "Chromosome 6q25 is linked to susceptibility to leprosy in a Vietnamese population." Nat. Genet. **33**(3): 412-415.

Olerup, O. and H. Zetterquist (1992). "HLA-DR typing by PCR amplification with sequence-specific primers (PCR-SSP) in 2 hours: an alternative to serological DR typing in clinical practice including donor-recipient matching in cadaveric transplantation." Tissue Antigens **39**(5): 225-35.

Ridley, D. S. and W. H. Jopling (1966). "Classification of leprosy according to immunity. A five-group system." Int. J. Lepr. Other Mycobact. Dis. **34**(3): 255-73.

Roach, D. R., H. Briscoe, et al. (2001). "Secreted Lymphotoxin-{{alpha}} Is Essential for the Control of an Intracellular Bacterial Infection." J. Exp. Med. **193**(2): 239-246.

Schaid, D. J. and C. Rowland (1998). "Use of parents, sibs, and unrelated controls for detection of associations between genetic markers and disease." Am. J. Hum. Genet. **63**(5): 1492-506.

Ware, C. F. (2005). "Network communications: lymphotoxins, LIGHT, and TNF." Annu. Rev. Immunol. **23**: 787-819.

WHO (2006). "Global leprosy situation, 2006." Wkly. Epidemiol. Rec. **81**: 309-316.

Zerva, L., B. Cizman, et al. (1996). "Arginine at positions 13 or 70-71 in pocket 4 of HLA-DRB1 alleles is associated with susceptibility to tuberculoid leprosy." J. Exp. Med. **183**(3): 829-36.

Chapter 3 │ Two HLA class I region SNPs are associated with leprosy susceptibility in Vietnam and India

The results presented in Chapter 2 suggested the existence of additional genetic risk factors for leprosy independent of *LTA*+80. As such, Chapter 3 describes a second higher density association scan (682 SNPs) of a 1.9-Mb sub-interval underlying the previously identified linkage peak on chromosome region 6p21 (i.e., the HLA complex). This effort led to the positional cloning of SNPs rs2394885 and rs2922997 – two intergenic HLA class I region variants – as risk factors for leprosy in two family-based samples from Vietnam and a population-based sample from north India. SNP rs2394885 is a previously reported tag SNP for *HLA-C**0401 in the Han Chinese in Beijing, China population, implicating *HLA-C* and natural killer cell activity in leprosy pathogenesis.

Two HLA class I region SNPs are associated with leprosy susceptibility in Vietnam and India

Andrea Alter[1], Nguyen Thu Huong[2], Meenakshi Singh[3], Marianna Orlova[1], Nguyen Van Thuc[2], Kiran Katoch[4], Vu Hong Thai[2], Nguyen Ngoc Ba[2], Laurent Abel[5,6,7], Narinder Mehra[3], Alexandre Alcaïs[5,6,8] and Erwin Schurr[1,8]

[1]Research Institute of the McGill University Health Centre, McGill Centre for the Study of Host Resistance, Departments of Medicine and Human Genetics, McGill University, Montreal, Quebec, Canada.

[2]Hospital for Dermato-Venereology, Nguyen Thong Street, District 3, Ho Chi Minh City, Vietnam.

[3]Department of Transplant Immunology and Immunogenetics, All-India Institute of Medical Sciences, Ansari Nagar, 110029 New Delhi, India.

[4]Central JALMA Institute of Leprosy and Other Infectious Diseases, Taj Ganj, 282001 Agra, India.

[5]Laboratoire de Génétique des Maladies Infectieuses, Institut National de la Santé et de la Recherche Médicale, U550, 75015 Paris, France.

[6]Université Paris René Descartes, Faculté Médicine Necker, 75015 Paris, France.

[7]Laboratory of Human Genetics of Infectious Diseases, Rockefeller Branch, The Rockefeller University, New York, NY 11065, USA.

[8]These authors share senior authorship.

Abstract

The chromosomal region 6p21 (HLA region) has been linked to leprosy in two independent genome-wide scans. Despite the identification of *HLA-DRB1* (HLA class II) and *LTA* (HLA class III) as underlying leprosy susceptibility genes, experimental evidence suggested the existence of unidentified risk loci in the HLA region. A high density association scan of a 1.9 mega-base (Mb) region in 198 Vietnamese single-case leprosy families identified 59 SNPs associated with leprosy ($P < 0.01$). Genotyping of these SNPs in an independent sample of 292 Vietnamese single-case leprosy families replicated the association of 12 SNPs ($P < 0.01$). Multivariate analysis of these 12 SNPs in the combined Vietnamese sample (490 families) showed that the association information could be captured by a model that included two intergenic HLA class I region SNPs ($P = 9.4\times10^{-9}$) – rs2394885 and rs2922997 (marginal multivariate $P = 2.1\times10^{-7}$ and $P = 0.0016$, respectively). Genotyping of the 12 SNPs in a third sample of 364 leprosy cases and 371 controls from North India confirmed the association of eight SNPs ($P < 0.01$). As in the Vietnamese familial sample, multivariate analysis showed that SNPs rs2394885 and rs2922997 captured all association information (marginal multivariate $P = 6.9\times10^{-8}$ and $P = 0.0038$, respectively). Therefore, in two ethnically different samples, SNP rs2394885, upstream of *HLA-C*, and SNP rs2922997, downstream of *HLA-B*, were independently associated with leprosy. Both leprosy risk factors target the centromeric end of the HLA class I region, specifically the *HLA-C* and *-B* genes, for allele fine-mapping and functional studies.

Introduction

Leprosy is an infectious disease of the skin and peripheral nerves caused by the bacterium *Mycobacterium leprae* (*M. leprae*). Both the development of the disease and its sub-forms (i.e., paucibacillary or multibacillary leprosy) are dependent on genetic factors as demonstrated by the concordance rates among monozygotic twins {Chakravartti and Vogel 1973} and segregation studies {Abel and Demenais 1988; Abel, Vu et al. 1995}. Several candidate genes, including *IL10*, *VDR* and *SLC11A1*, have been associated with leprosy and/or its sub-forms {Casanova and Abel 2002; Mira 2006}, and more recently, positional cloning has successfully identified two risk loci (i.e., *PARK2*/*PACRG* {Mira, Alcais et al. 2004} and *LTA* {Alcais, Alter et al. 2007}). Therefore, forward genetic dissection of the human-*M. leprae* interaction is a powerful strategy to further the understanding of this ancient, yet elusive, affliction {Alcais, Mira et al. 2005}.

Two independent studies have linked chromosome region 6p21 – the HLA region – to leprosy susceptibility, including our own genome-wide scan in 86 Vietnamese multi-case leprosy families (multipoint lod score = 2.62) {Mira, Alcais et al. 2003; Miller, Jamieson et al. 2004}. An association scan of a 10.4 Mb region underlying the linkage peak at 39.48 Mb (build 37.1) indentified a functional SNP in the promoter of the HLA class III gene lymphotoxin-α (*LTA)* as a risk factor for early-onset leprosy in the Vietnamese and North Indian populations {Alcais, Alter et al. 2007}. However, we showed that this association was independent of both *HLA-DRB1*, an established susceptibility gene {Mehra, Rajalingam et al. 1995; Zerva, Cizman et al. 1996; Shaw, Donaldson et al. 2001; Singh, Balamurugan et al. 2007; Vanderborght, Pacheco et al. 2007}, and at least three other potential leprosy-associated SNPs, suggesting that the HLA region harbors multiple unidentified risk loci. By this reasoning, we set to identify additional genetic risk factors for leprosy in this chromosome region.

A high-resolution association scan in 198 single-case leprosy families from Vietnam followed by step-wise replication in an independent family-based sample from Vietnam and a population-based sample from North India, identified two intergenic HLA class I region SNPs – rs2394885 and rs2922997 – as novel leprosy risk factors.

Results

High-density SNP association scan of 6p21 in 198 leprosy families from Vietnam

Based on our own results that provided evidence for the presence of additional leprosy susceptibility genes in the HLA class III region {Alcais, Alter et al. 2007}, and from numerous published studies implicating the HLA class II genes but not the class I genes {Mehra, Rajalingam et al. 1995; Zerva, Cizman et al. 1996; Shaw, Donaldson et al. 2001; Casanova and Abel 2002; Singh, Balamurugan et al. 2007; Vanderborght, Pacheco et al. 2007}, we focused our study on the HLA class II and III regions. However, since a small portion of the class I region fell within the 95% CI of the linkage peak from our previous genome-wide linkage study {Mira, Alcais et al. 2003}, we also selected the corresponding chromosomal segment (*HCG27* → *MICB*) for high SNP coverage. Overall, an interval extending from 31.2 Mb to 33.1 Mb (build 37.1) on chromosome 6 (6p21.33-p21.32) was selected for a high-density association scan in 198 single-case leprosy families from Ho Chi Minh City, Vietnam (Vietnam-1). This 1.9 Mb interval includes 13 genes at the centromeric end of the HLA class I region (*HCG27* → *MIC*B), the 76 genes in the HLA class III region (*PPIAP9* → *NOTCH4*) and the 34 genes in the HLA class II region (*C6orf10* → *HLA-DPA3*). Complementing the target interval, 130 previously determined tag SNPs for classical HLA class I and class II gene alleles in the Han Chinese in Beijing, China (CHB) population were included {de Bakker, McVean et al. 2006}. Although some HLA allele tag SNPs lie telomeric to the main targeted interval (left section of **Figure 1**), these SNPs further increased the overall coverage of the HLA region. In total, 854 SNPs were selected (**Table S1**).

After excluding 172 SNPs (see Methods; **Table S1**), the tagging efficiency (i.e., proportion of tagged "bins" {Hinds, Stuve et al. 2005}, see Methods) of the entire 1.9 Mb interval using an r^2

cutoff = 0.50 was 0.80 in the CHB population (HapMap Data Rel 24). By univariate analysis, 64 of the 682 suitable SNPs were significantly associated with leprosy at the 0.01 level (**Figure 1**). Expectedly, rs2239704 (*LTA*+80), the previously reported *LTA* promoter risk variant {Alcais, Alter et al. 2007}, was among the 64 SNPs with a $P < 0.01$. We included *LTA*+80 to identify and exclude highly correlated SNPs from subsequent replications. Indeed, four SNPs were in strong linkage disequilibrium (LD) ($r^2 \geq 0.80$) with *LTA*+80. As such, only the 59 non-*LTA*+80 correlated SNPs were targeted for replication (**Table S2**).

Of note, when we performed age-stratified analysis of the rs2394885 bin SNPs, we observed that the overall risk effect was exclusively due to individuals > 15 years at time of diagnosis (109 families) (rs2394885 $P_{>15} = 0.006$ and $P_{\leq 15} = 0.371$). This is interesting since we have previously shown that the effect of the *LTA*+80 A-allele on leprosy susceptibility was mainly due to individuals ≤ 15 years at time of diagnosis (89 families) {Alcais, Alter et al. 2007}.

Replication study: 292 leprosy families from Vietnam

We enrolled a second independent sample of 292 single-case leprosy families from Ho Chi Minh City, Vietnam (Vietnam-2) to identify false-positive associations in the Vietnam-1 sample. We genotyped 52 of the 59 non-*LTA*+80 correlated SNPs with a *P* value < 0.01. Seven SNPs were excluded from the panel for the following reasons: three SNPs had a Hardy-Weinberg equilibrium (HWE) *P* value < 0.05 in the Vietnam-1 sample, three SNPs were redundant ($r^2 \geq 0.99$ with an included SNP) and one SNP failed primer design (**Table S2**). By univariate analysis, 12 of the 49 suitable SNPs (three SNPs could not be genotyped) were significantly associated with leprosy at the 0.01 level in the Vietnam-2 sample (with concordant risk alleles) (**Table 1**). Of the 12 replicated SNPs, 11 were located in the centromeric end of the HLA Class I

region (between *HCG27* and *MICA*), and one SNP was located in the adjacent HLA class III region (between *PPIAP9* and *RPL15P4*).

When we combined the Vietnam-1 and Vietnam-2 samples (490 single-case leprosy families), the evidence for association for each of the 12 replicated SNPs was very strong (**Table 1**). However, as not all redundant SNPs were excluded, extensive LD among the 12 SNPs in the combined parental data-set (i.e., 980 unrelated Vietnamese individuals) was still observed (**Table S3**). Specifically, the 12 SNPs could be assigned to four bins using an r^2 cutoff = 0.50 (**Table 1**). Consequently, stepwise multivariate conditional logistic regression analysis of the 12 SNPs in the combined sample showed that the association information could be captured by a model ($P = 9.4\times10^{-9}$) that included two HLA class I intergenic SNPs – rs2394885 and rs2922997 – each with an additive genetic effect and marginal multivariate *P* value of 2.1×10^{-7} (OR for CC versus CG = CG versus GG = 2.32 (95% CI 1.62 – 3.33)) and 0.0016 (OR for CC versus CT = CT versus TT = 1.42 (1.14 – 1.76)), respectively (**Table 2**). Multivariate analysis generated additional models representing different combinations of two SNPs from the two associated bins that were statistically indistinguishable from the one presented here. We choose to present the above two-SNP model since rs2394885 is a reported tag SNP for *HLA-C*0401* in the CHB population (r^2 = 1.0) {de Bakker, McVean et al. 2006}. We have repeatedly compared the LD structure of the CHB population to the Vietnamese population and have consistently found the patterns to be highly comparable (**Table S4**). Of course, this implies that other bin SNPs tag the same *HLA-C* allele equally well.

Validation study: 364 leprosy cases and 371 controls from North India

To validate the observed associations with leprosy in a geographically and ethnically different population, we enrolled a third sample of 364 unrelated individuals with leprosy and 371 unrelated healthy controls from Agra, North India. We successfully genotyped the 12 replicated SNPs. By univariate logistic regression, eight of the 11 suitable SNPs (HWE *P* value for rs9266455 was < 0.01 in controls) were significantly associated with leprosy at the 0.01 level (with concordant risk alleles and genetic models as compared to the combined Vietnamese sample) (**Table 1**). Identical to the combined Vietnamese sample, multivariate analysis of the eight SNPs showed that the best model ($P = 7.6\times10^{-9}$) included the same two SNPs – rs2394885 and rs2922997 – with marginal multivariate *P* values of 6.9×10^{-8} (OR for CC versus CG = CG versus GG = 2.23 (95% CI 1.61 – 3.09)) and 0.0038 (OR for CC versus CT = CT versus TT = 1.45 (1.15 – 1.83)), respectively (**Table 2**).

Risk factors rs2394885 and rs2922997 are independent from *HLA-DRB1* and *LTA*

Given that *HLA-DRB1* (HLA class II) and *LTA* (HLA class III) are known leprosy susceptibility genes in the Vietnamese {Alcais, Alter et al. 2007; Vanderborght, Pacheco et al. 2007} and North Indian {Mehra, Rajalingam et al. 1995; Alcais, Alter et al. 2007} populations, we wanted to exclude the possibility that the observed association of leprosy with SNPs rs2394885 and rs2922997 could be due to long-range LD with *LTA*+80 or *HLA-DRB1* alleles. Multivariate analysis in the Vietnam-1 sample (198 single-case families) showed that there was no significant impact on the risk effect of SNPs rs2394885 (univariate $P = 0.006$) and rs2922997 (univariate $P = 0.009$) when we included *LTA*+80 and *HLA-DRB1* alleles despite the loss of 22 families in the process (adjusted $P = 0.008$ and 0.04 for rs2394885 and rs2922997, respectively). Similarly, multivariate analysis in the North Indian sample showed that there was no noticeable impact on the risk effect of SNPs rs2394885 (univariate $P = 3.1\times10^{-8}$) and rs2922997 (univariate $P =$

2.1×10^{-5}) when we included *LTA*+80 in the multivariate analysis (adjusted $P = 6.9\times10^{-8}$ and 0.0038 for rs2394885 and rs2922997, respectively), nor did we detect any $r^2 > 0.10$ between either SNP and *HLA-DRB1* alleles in the control group. These results demonstrate the independent association of SNPs rs2394885 and rs2922997 from the known HLA leprosy susceptibility genes in these two populations.

Discussion

To date, our step-wise approach to positional cloning has identified three replicated leprosy risk factors in addition to *HLA-DRB1*{Vanderborght, Pacheco et al. 2007} underlying the linkage peak on chromosome region 6p21: *LTA*+80{Alcais, Alter et al. 2007}, rs2394885 and rs2922997. Our experience depicts a complex situation with multiple HLA region alleles impacting on mycobacterial susceptibility. However, these results serve to improve our understanding of linkage peak architecture in the context of a complex disease. Here, we show that in analogy to what has been described for linkage QTLs in experimental models {Legare, Bartlett et al. 2000; Flint and Mackay 2009; Mackay, Stone et al. 2009}, a cluster of multiple susceptibility loci can underlie the linkage peak. The notion of "one linkage peak = multiple risk loci" may prove to be a recurring observation in the genetic analysis of complex disease in humans. A recent genome-wide association study similarly revealed three independent loci within the MHC conferring risk of psoriasis {Feng, Sun et al. 2009}.

As such, despite our extensive coverage of the 1.9 Mb targeted interval on chromosome 6 (tagging efficiency = 0.80), we cannot exclude the possibility that additional leprosy risk factors contributing to the linkage signal may have been overlooked (i.e., untagged SNPs, SNPs not genotyped by HapMap, unreported SNPs and structural variations). The issue of coverage is compounded by the extrapolation of the CHB population LD map to the Vietnamese population. Although we have consistently observed very similar LD patterns (**Table S4**), differences exist. Ideally the process of SNP panel design should be based on the study population, an increasing reality given the addition of seven new populations in HapMap Data Rel 27.

As stated above (see Results), models representing different combinations of two SNPs from the two associated bins are statistically indistinguishable. Analogous to most genetic studies, SNPs rs2394885 and rs2922997 are risk factors that likely tag the causal variants, identification of which requires ultra-fine LD mapping coupled with functional studies. Interestingly however, rs2394885 is a reported tag SNP for *HLA-C*0401* in the closely-related CHB population (r^2 = 1.0) {de Bakker, McVean et al. 2006}. We utilized the HapMap database (Data Rel 24) to screen a 1 Mb region around rs2394885 in the CHB population. Using an r^2 cutoff = 0.40 (MAF > 5%), 27 SNPs were correlated with rs2394885 (**Table S5**). One SNP – rs2233952 – is a non-synonymous coding SNP in *PSORS1C2* (leu → pro) but neither this SNP nor 12 other *PSORS1C2* SNPs were associated with psoriasis {Chang, Tsai et al. 2003}. These data support *HLA-C*0401* as the causative allele, however, it is imperative to type *HLA-C* in select Vietnamese and Indian samples to confirm the strong correlation between the rs2394885 bin and *HLA-C*0401* in both populations. *HLA-C* alleles and promoter variants have previously been associated with several inflammatory {Yen, Moore et al. 2001; Nair, Stuart et al. 2006; Tai Wai, Philip et al. 2007} and infectious diseases {Balamurugan, Sharma et al. 2004; Fellay, Shianna et al. 2007}. The association of *HLA*-C with leprosy susceptibility would implicate Natural Killer (NK) cells – important effectors of the innate response to viruses and intracellular pathogens – in the host response to *M. leprae* infection. This is supported by immunological assays that showed *M. leprae* was able to induce NK cell mediated cytotoxicity {Kaleab, Ottenoff et al. 1990}. A similar analysis of rs2922997 identified eight correlated SNPs, none with any evidence of biological relevance (**Table S5**). However, given that the bin spans a 6.5 kb intergenic region centromeric to *HLA-B*, *HLA-B* typing could be justified.

We have replicated and validated all leprosy susceptibility variants identified from linkage studies, including the four HLA region variants above, in geographically and ethnically different populations. This ease of replication implies that genetic heterogeneity in leprosy susceptibility is very limited. Low genetic heterogeneity is in striking parallel to the low degree of diversity among *M. leprae* strains {Monot, Honore et al. 2005; Monot, Honore et al. 2009} and suggests very general adaptations of the leprosy bacillus to its human host. If this conclusion is correct, one would predict that in infectious diseases, genetic heterogeneity mainly reflects differences in the infectious strains of the pathogen. Interestingly, *Mycobacterium tuberculosis* strain-dependent host genetic effects have been described {Caws, Thwaites et al. 2008}. Together, these data suggest that a careful characterization of pathogens isolated from patients may be an effective way to overcome apparent genetic heterogeneity in common infectious diseases.

Although little attention has been given to the HLA class I region in the context of leprosy, the next logical step in the genetic dissection of leprosy, is to target the classical HLA class I genes – particularly *HLA*-C. In light of our findings and the above referenced genetic studies, the complex role of HLA-C in NK cell biology is an emerging area of research with the potential to expand our understanding of leprosy pathogenesis, inflammation and intracellular infection.

Methods

Patients and controls

We enrolled a total of 2,205 individuals in this study. The 198 single-case Vietnamese leprosy families (Vietnam-1; 52% multibacillary) and the additional 292 single-case Vietnamese leprosy families (Vietnam-2; 54% multibacillary) were identified from records available at the Dermato-Venereology Hospital in Ho Chi Minh City, Vietnam. The criterion for enrolment was the availability of both parents for genetic analysis. The 364 leprosy cases (70% multibacillary) were identified from records available the Central JALMA Institute of Leprosy and Other Infectious Diseases in Agra, India. The 371 healthy controls (i.e., no reported infectious or inflammatory disease) from North India were identified from blood donor clinics. The study was approved by institutional review boards and health authorities in Ho Chi Minh City, Vietnam; the All India Institute of Medical Sciences, New Delhi, India; and the McGill University Health Centre, Montreal, Canada. Written informed consent was obtained from all participants.

The diagnosis of leprosy and the classification of subtype were based on clinical and histological criteria {Ridley and Jopling 1966}. However, the phenotype studied here was leprosy *per se* (i.e., leprosy independent of specific clinical manifestations).

Genotyping methods

The 854 SNPs spanning the 1.9 Mb targeted interval on chromosome 6 (6p21.33-p21.32) were selected based on their proximity to or location within known genes in the interval, and tag SNP information publicly available from the International HapMap project database (www.hapmap.org/). Included in our panel, were 130 previously determined tag SNPs for

classical HLA class I and class II gene alleles in the CHB population {de Bakker, McVean et al. 2006}. These SNPs were genotyped on one of the following platforms: (i) direct sequencing on an ABI PRISM® 3100 genetic analyzer; (ii) polarized fluorescence TaqMan® Assay {Lee, Connell et al. 1993}; (iii) the high-throughput GenomeLab™ SNPstream® platform (formerly Orchid SNPstream UHT), which uses a single-base pair extension (SBE) method to incorporate fluorescently labeled terminal nucleotides, which are then detected by a specialized imager {Bell, Chaturvedi et al. 2002}; (iv) the high-throughput SEQUENOM® MassARRAY® platform, which uses the iPLEX™ assay to incorporate mass-modified terminal nucleotides in the SBE step, which are then detected by MALDI-TOF MS {Griffin and Smith 2000}; (v) the ultra-high throughput Illumina® platform. This platform uses the GoldenGate™ assay followed by a bead-based technology to resolve individual SNP genotypes {Fan, Oliphant et al. 2003}.

After genotyping, we excluded 172 SNPs from the analysis for the following reasons: 39 could not be genotyped, three could not be placed unambiguously on the sequence map, 19 showed deviations ($P < 0.01$) from HWE, 103 were non-informative or had a minor allele frequency (MAF) < 5%, three had > 10 Mendelian errors and five were identified as “problem SNPs” (**Table S1**)

All 52 SNPs genotyped in the Vietnam-2 sample were genotyped on the high-throughput SEQUENOM® MassARRAY® platform {Griffin and Smith 2000} (**Table S2**). All 12 SNPs genotyped in the Indian sample were genotyped on one of the following platforms: (i) direct sequencing on an ABI PRISM® 3100 genetic analyzer; (ii) the high-throughput GenomeLab™ SNPstream® platform (formerly Orchid SNPstream UHT) {Bell, Chaturvedi et al. 2002}; (iii) the high-throughput SEQUENOM® MassARRAY® platform {Griffin and Smith 2000}.

The *HLA DRB1* gene was genotyped in the Vietnam-1 sample and the 371 controls from North India using standard methods {Olerup and Zetterquist 1992}.

Statistical methods

We estimated population allelic frequencies and pairwise LD (r^2) between SNPs from parental or control data using the algorithm implemented in Haploview 4.1 software {Barrett, Fry et al. 2005}. Bins of SNPs were constructed on the basis of the pairwise r^2 and defined as a not necessarily contiguous set of SNPs where at least one SNP has $r^2 > 0.50$ with all the other SNPs of the bin {Hinds, Stuve et al. 2005}. For the 1.9 Mb targeted interval (*HCG27* → *COL11A2*), we uploaded all SNP genotype data for the CHB population from HapMapa Data Rel 24 into Haploview 4.1. We reset the pairwise comparison threshold to 2 Mb, and using the 'Tagger' function, we "force included" our 682 successfully genotyped SNPs and calculated the tagging efficiency by dividing the number of captured bins by the total number of bins (including single-SNP bins) using an r^2 cutoff = 0.50.

Family-based association studies were performed in the Vietnam-1 and -2 samples using principally a classical transmission disequilibrium test, as implemented in FBAT v2.0.3 software {Horvath, Xu et al. 2001}. We carried out a population-based association study in the North Indian sample, using classical multivariate logistic regression techniques as implemented in the LOGISTIC procedure of SAS software v.9.1 (SAS institute, Cary, NC). All population-based analyses were adjusted for sex, which has been identified as a classical risk factor for leprosy.

Acknowledgements

We are grateful to all patients and their families who participated in this study. We thank A. Montpetit and A. Bélisle for assistance with high-throughput genotyping and L. Simkin for assistance with genotyping data management. This study was supported by a grant from the Canadian Institutes of Health Research (CIHR) to E.S. and MAGRALEPRE from l'Ordre de Malte to Al.A. An.A. holds a graduate studentship from the Natural Science and Engineering Research Council of Canada (NSERC). Al.A. and L.A. are supported by the Assistance Publique-Hôpitaux de Paris, Programme de Recherche Fondamentale en Microbiologie Maladies Infectieuses et Parasitaires (PRFMMIP) and the Agence Nationale de la Recherche (ANR) of the Ministère Français de l'Éducation Nationale de la Recherche et de la Technologie. E.S. is a Chercheur National du Fonds de la Recherche en Santé du Québec and an International Research Scholar of the Howard Hughes Medical Institute. The Laboratory of Human Genetics of Infectious Diseases is supported by grants from The Rockefeller University Center for Clinical and Translational Science grant number 5UL1RR024143-03 and The Rockefeller University.

Figure Legend

Figure 1 │ High-density SNP association scan of 6p21 in the Vietnam-1 sample

Evidence for association with leprosy of 682 SNPs in 198 Vietnamese single-case families (Vietnam-1) is plotted as $-\log_{10}P$ on the y-axis. The location of the SNPs, in Mb, is indicated on the x-axis (dbSNP129). The HLA class I, class III and class II regions are indicated by solid lines. The low SNP coverage in the class I regions corresponds to tag SNPs for class I and class II alleles that are located outside the target area. The thin dotted line indicates the $P = 0.01$ significance threshold. Three different genetic models were tested for each SNP (i.e., additive and fully dominant/recessive) and the *P* values for the best model are reported here (P_{min}).

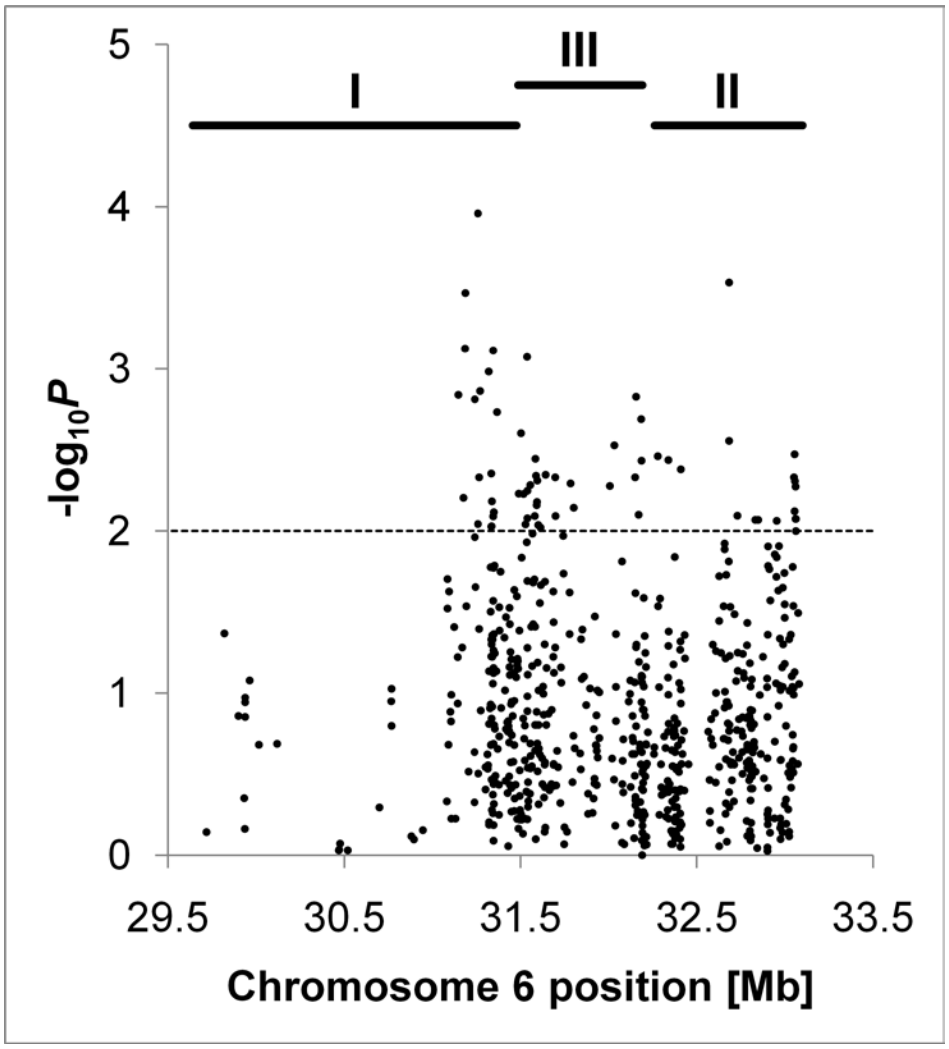

I
III
II
5
4
3
2
1
0
-log10P
29.5
30.5
31.5
32.5
33.5
Chromosome 6 position [Mb]

Figure 1

Table 1. Univariate analysis of 12 HLA SNPs in the Vietnam-1, Vietnam-2, combined Vietnamese and North Indian samples

SNP and bin structure[a]	Position[b]	Risk allele	Vietnam MAF[c]	Vietnam-1 (198 families) P_{min}[d]	Vietnam-2 (292 families) P_{min}[d]	Combined Vietnam (490 families) P_{min}[d]	North India MAF[c]	North India (371 controls, 364 cases) P_{min}[d]
rs4713438	31146846	A (rare)	0.19	0.0015	0.0045	7.9×10^{-5}	0.29	0.0008
rs2394885	31174590	C (rare)	0.11	0.0063	8.8×10^{-5}	1.8×10^{-6}	0.11	3.0×10^{-8}
rs9501522	31184425	T (rare)	0.11	0.0007	3.2×10^{-5}	9.2×10^{-8}	0.10	4.0×10^{-8}
rs9461662	31187103	C (rare)	0.10	0.0003	2.2×10^{-5}	4.7×10^{-8}	0.10	5.0×10^{-8}
rs16899166	31258096	T (rare)	0.10	0.0091	0.0012	3.2×10^{-5}	0.10	6.0×10^{-7}
rs16899203	31266335	C (rare)	0.12	0.0047	0.0002	2.8×10^{-6}	0.11	6.4×10^{-7}
rs2922997	31337281	C (rare)	0.43	0.0094	0.0037	0.0002	0.44	2.0×10^{-5}
rs9266455	31338385	T (rare)	0.35	0.0066	0.0073	0.0040	-[e]	-[e]
rs2507981	31346247	T (rare)	0.42	0.0008	0.0040	0.0020	0.42	8.0×10^{-6}
rs9266722	31349922	G (common)	0.42	0.0077	0.0069	0.0008	0.16	NS[f]
rs2523454	31367865	G (common)	0.39	0.0019	0.0095	0.0003	0.15	NS[f]
rs3093993	31490725	G (rare)	0.20	0.0059	0.0083	0.0002	0.14	NS[f]

[a]SNPs identical in colour belong to the same bin set in the Vietnamese and North Indian populations (r^2 cutoff = 0.50). SNPs in black do not belong to any bin set. Note that rs9266722 and rs2523454 are not in the same bin in the North Indian population.
[b]based on dbSNP 129
[c]minor allele frequency
[d]*P* values for the best model are reported here
[e]HWE $P < 0.01$ in control group
[f]not significant at the 0.01 level

Table 2. Univariate and multivariate analysis of the eight replicated HLA SNPs in the combined Vietnamese and North Indian samples

SNP[a]	rs4713438	rs2394885	rs9501522	rs9461662	rs16899166	rs16899203	rs2922997	rs2507981
				Combined Vietnam (490 families)				
MAF	0.19	0.11	0.11	0.10	0.10	0.12	0.43	0.42
Univariate P_{min}	7.9×10^{-5}	1.8×10^{-6}	9.2×10^{-8}	4.7×10^{-8}	3.2×10^{-5}	2.8×10^{-6}	0.0002	0.0020
Multivariate P[b] OR (95% CI)	-	2.1×10^{-7} 2.32 (1.62 – 3.33)	-	-	-	-	0.0016 1.42 (1.14 – 1.76)	-
				North India (371 controls, 364 cases)				
MAF	0.29	0.11	0.10	0.10	0.10	0.11	0.44	0.42
Univariate P_{min}	0.0008	3.0×10^{-8}	4.0×10^{-8}	5.0×10^{-8}	6.0×10^{-7}	6.4×10^{-7}	2.0×10^{-5}	8.0×10^{-6}
Multivariate P[b] OR (95% CI)	-	6.9×10^{-8} 2.23 (1.61 – 3.09)	-	-	-	-	0.0038 1.45 (1.15 – 1.83)	-

[a]SNPs identical in colour belong to the same bin set in the Vietnamese and North Indian populations (r^2 cutoff = 0.50). rs4713438 does not belong to any bin set.
[b]any SNP in the bin is sufficient to explain the observed association of that bin with leprosy

Supplementary material

Supplementary table 1 is too large for insertion, please see:

http://docs.google.com/fileview?id=0B0I8nribECnnZjYwNGNjYWMtMGY2NS00N2RmLThlYjEtYTFlMDU0ODYwNDVh&hl=en

Supplementary Table 2

Characterisation in Vientam-2 sample of 64 SNPs with P_{min} < 0.01 in Vietnam-1 sample

#	SNP rs#	Position[a]	MAF[b]	Frequent allele	Rare allele	Failed (0=no; 1=yes)	Reason for exclusion from analysis in Vietnam-2 sample
1	rs4713438	31146846	0.20	G	A	0	
2	rs2394885_a	31174590	0.00	-	-	1	
2	rs2394885_b	31174590	0.12	G	C	0	
3	rs9501522	31184425	0.12	C	T	0	
4	rs9461662	31187103	0.12	G	C	0	
5	rs2249742	31240721	0.49	G	A	0	
6	rs16899166	31258096	0.10	C	T	0	
7	rs16899168	31258687	0.18	G	A	0	
8	rs16899203	31266335	0.14	T	C	0	
9	rs9264863	31271530	0.22	T	C	0	HWE *P value* < 0.05 in Vietnam-2 sample
10	rs9266076	31319785	0.13	C	T	0	
11	rs2844575_a	31334945	0.49	G	A	0	
11	rs2844575_b	31334945	0.49	G	A	0	
12	rs9266387	31335216	0.42	T	C	0	
13	rs2922997_a	31337281	0.00	-	-	1	
13	rs2922997_b	31337281	0.44	T	C	0	
14	rs9266455	31338385	0.35	C	T	0	
15	rs2507983	31345794	-	-	-	-	could not design primers
16	rs2507981	31346247	0.42	A	T	0	
17	rs9266722	31349922	0.41	G	A	0	
18	rs2523454	31367865	0.38	G	A	0	

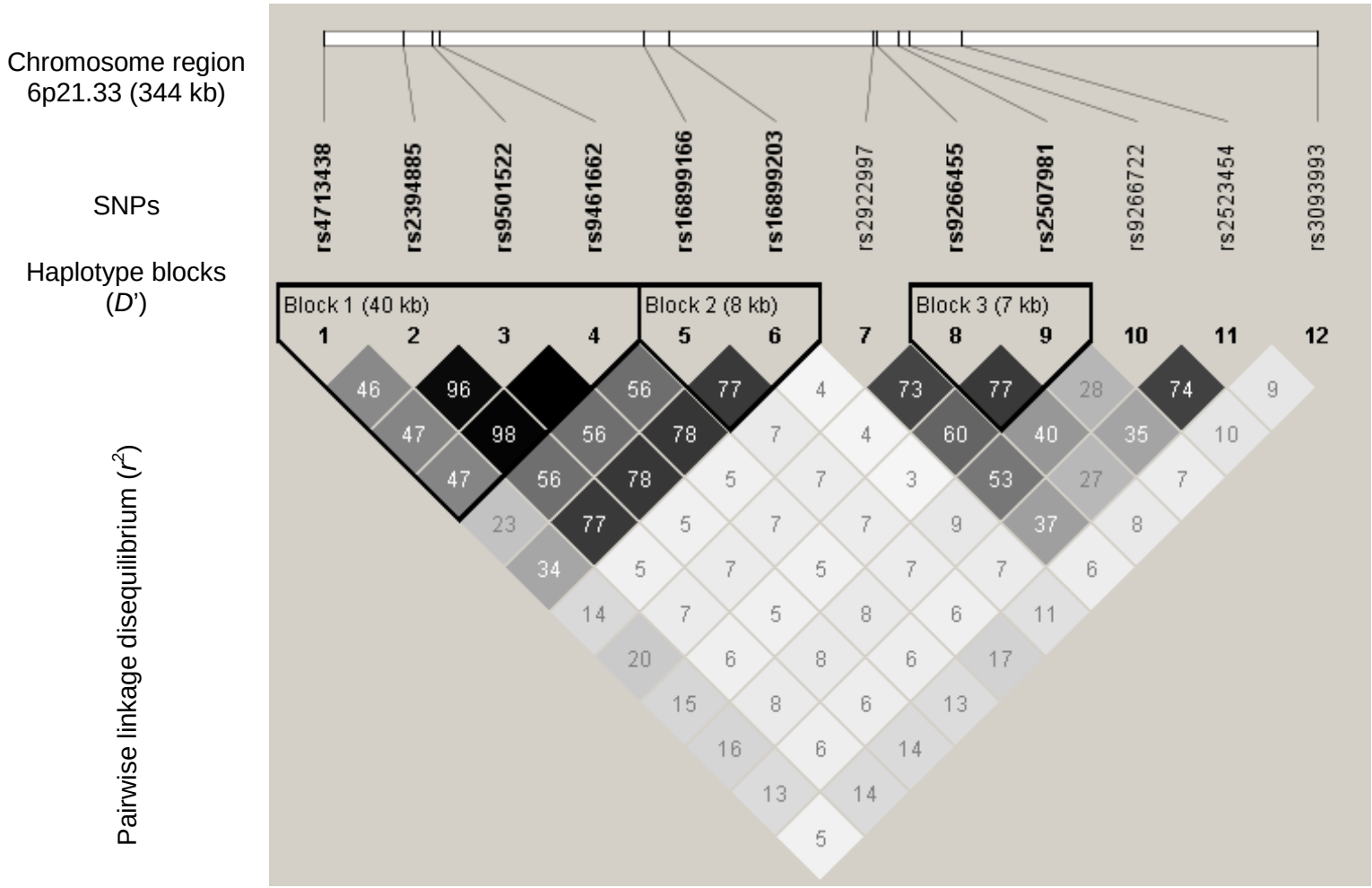
19	rs3093993	31490725	0.21	T	G	0	
20	rs3131628	31502767	0.16	T	C	0	
21	rs2523500	31518354	-	-	-	-	linkage disequilibrium with *LT*A+80 (rs2239704)
22	rs2516479	31528326	-	-	-	-	linkage disequilibrium with *LT*A+80 (rs2239704)
23	rs915654	31538497	0.00	-	-	1	failed
24	rs2071590	31539768	-	-	-	-	linkage disequilibrium with *LT*A+80 (rs2239704)
25	rs2239704	31540141	-	-	-	-	linkage disequilibrium with *LT*A+80 (rs2239704)
26	rs2256965	31555130	-	-	-	-	linkage disequilibrium with *LT*A+80 (rs2239704)
27	rs3132451	31582025	-	-	-	-	linkage disequilibrium with rs3130070 (r^2 = 0.99)
28	rs2857697	31585219	0.00	-	-	1	failed
29	rs2857694	31587870	0.50	A	A	0	
30	rs2857693	31588384	0.50	A	C	0	
31	rs3130070	31591808	0.16	A	G	0	
32	rs2260000	31593476	0.50	A	G	0	
33	rs2736171	31595487	0.48	T	C	0	
34	rs3115663	31601843	-	-	-	-	linkage disequilibrium with rs3130070 (r^2 = 1)
35	rs3130048	31613739	0.00	-	-	1	failed
36	rs9378164	31643522	0.21	G	A	0	
37	rs805305	31697387	0.48	G	C	0	
38	rs2272592	31698352	0.23	C	T	0	
39	rs1043618	31783507	0.49	C	G	0	
40	rs9368699	31802541	0.19	A	G	0	
41	rs6465	32007761	-	-	-	-	HWE *P value* < 0.05 in Vietnam-1 sample
42	rs9469079	32032421	0.17	C	T	0	
43	rs184003	32150296	0.22	C	A	0	
44	rs204993	32155581	0.37	A	G	0	
45	rs2071287	32170433	-	-	-	-	HWE *P value* < 0.05 in Vietnam-1 sample
46	rs404860	32184345	0.32	A	G	0	
47	rs438475	32186245	0.27	C	T	0	
48	rs3864302	32278792	0.38	G	A	0	
49	rs2050190	32339076	0.47	T	C	0	
50	rs3129881	32409484	0.11	C	T	0	

51	rs3104404	32682174	0.43	C	A	0	
52	rs9275602	32682812	0.18	C	A	0	
53	rs2071550	32730940	0.47	C	A	0	
54	rs9276831	32832033	0.23	A	G	0	
55	rs9276900	32848566	0.24	A	C	0	
56	rs206777	32952546	0.45	T	C	0	
57	rs9277378	33050279	0.31	C	T	0	
58	rs3128961	33050742	0.31	A	G	0	
59	rs9277477	33053772	0.31	A	G	0	
60	rs9277489	33053942	0.32	G	A	0	
61	rs3128968	33056253	0.32	A	T	0	
62	rs3128917	33059996	0.44	G	T	0	
63	rs3117223	33060064	0.44	T	C	0	
64	rs3117222	33060949	-	-	-	-	linkage disequilibrium with rs3128917 (r^2 = 1)

[a]based on dbSNP129
[b]minor allele frequency

Supplementary Table 3

Pairwise linkage disequilibrium (r^2) of 12 HLA SNPs in 980 unrelated Vietnamese individuals



Supplementary Table 4
Comparison of pairwise linkage disequilibrium (r^2) of 12 HLA SNPs in the CHB, Vietnamese and North Indian populations

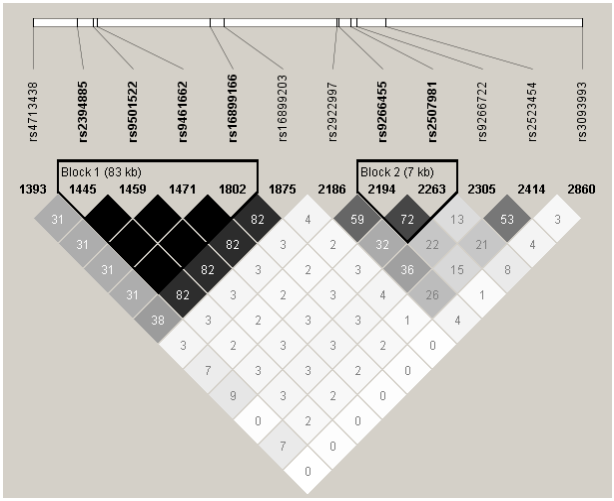
CHB

Chromosome region 6p21.33 (344 kb)

SNPs

Haplotype blocks (*D*')

Pairwise linkage disequilibrium (r^2)



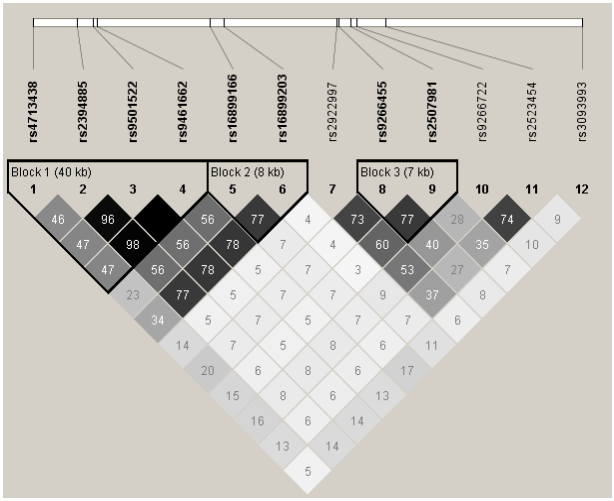
Vietnam

Chromosome region 6p21.33 (344 kb)

SNPs

Haplotype blocks (*D*')

Pairwise linkage disequilibrium (r^2)



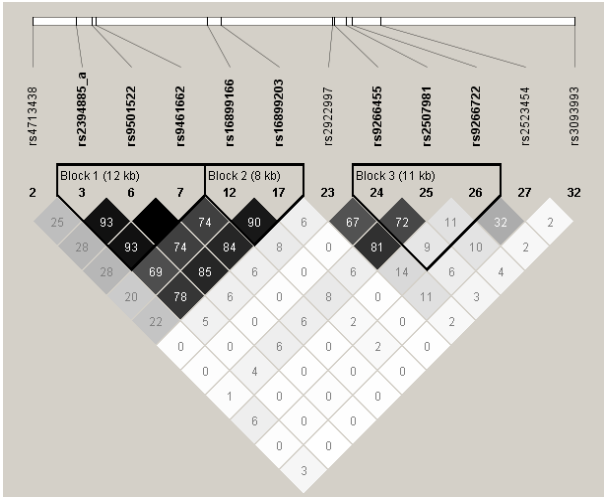
North India

Chromosome region 6p21.33 (344 kb)

SNPs

Haplotype blocks (*D*')

Pairwise linkage disequilibrium (r^2)



Supplementary Table 5
All SNPs (MAF >5%) in a 1 Mb window with r^2 > 0.40 with rs2394885 and rs2922997 in CHB HapMap population (Data Rel 24)

Allele	Best Test	r^2 w/test	Allele	Best Test	r^2 w/test
rs2233952	rs2394885	0.62	rs2523560	rs2922997	0.95
rs9468877	rs2394885	0.47	rs9501374	rs2922997	0.73
rs1052986	rs2394885	0.53	rs9501572	rs2922997	0.74
rs4713440	rs2394885	0.42	rs6457401	rs2922997	0.70
rs2394885	rs2394885	1.00	rs6457402	rs2922997	0.69
rs4516988	rs2394885	0.60	rs7743761	rs2922997	0.74
rs4351302	rs2394885	0.60	rs2523534	rs2922997	1.00
rs9295965	rs2394885	0.60	rs2922997	rs2922997	1.00
rs9295967	rs2394885	0.60	rs13215664	rs2922997	0.74
rs9501522	rs2394885	1.00			
rs9461662	rs2394885	1.00			
rs12662501	rs2394885	0.53			
rs3899471	rs2394885	0.53			
rs6457358	rs2394885	1.00			
rs5009853	rs2394885	0.65			
rs9468913	rs2394885	0.70			
rs16899166	rs2394885	1.00			
rs16899168	rs2394885	0.53			
rs16899170	rs2394885	0.82			
rs16899178	rs2394885	0.82			
rs3906273	rs2394885	0.70			
rs16899203	rs2394885	0.82			
rs3915970	rs2394885	0.82			
rs9264984	rs2394885	0.60			
rs9265420	rs2394885	0.79			
rs7453967	rs2394885	1.00			
rs2523618	rs2394885	1.00			
rs9266076	rs2394885	0.61			

References

Abel, L. and F. Demenais (1988). "Detection of major genes for susceptibility to leprosy and its subtypes in a Caribbean island: Desirade island." Am. J. Hum. Genet. **42**(2): 256-66.

Abel, L., D. L. Vu, et al. (1995). "Complex segregation analysis of leprosy in southern Vietnam." Genet. Epidemiol. **12**(1): 63-82.

Alcais, A., A. Alter, et al. (2007). "Stepwise replication identifies a low-producing lymphotoxin-[alpha] allele as a major risk factor for early-onset leprosy." Nat. Genet. **39**(4): 517-522.

Alcais, A., M. Mira, et al. (2005). "Genetic dissection of immunity in leprosy." Curr. Opin. Immunol. **17**(1): 44-8.

Balamurugan, A., S. K. Sharma, et al. (2004). "Human Leukocyte Antigen Class I Supertypes Influence Susceptibility and Severity of Tuberculosis." J. Infect. Dis. **189**(5): 805-811.

Barrett, J. C., B. Fry, et al. (2005). "Haploview: analysis and visualization of LD and haplotype maps." Bioinformatics **21**(2): 263-265.

Bell, P., S. Chaturvedi, et al. (2002). "SNPstream UHT: ultra-high throughput SNP genotyping for pharmacogenomics and drug discovery." Biotechniques **Suppl**: 70-2, 74, 76-7.

Casanova, J. L. and L. Abel (2002). "Genetic dissection of immunity to mycobacteria: the human model." Annu. Rev. Immunol. **20**: 581-620.

Caws, M., G. Thwaites, et al. (2008). "The influence of host and bacterial genotype on the development of disseminated disease with Mycobacterium tuberculosis." PLoS Pathog. **4**(3): e1000034.

Chakravartti, M. R. and F. A. Vogel (1973). A twin study on leprosy. Topics in human genetics. P. E. Becker, W. Lenz, F. Vogel and G. G. Wendt. Stuttgart, Georg Thieme. **1:** 1-123.

Chang, Y. T., S. F. Tsai, et al. (2003). "SPR1 gene near HLA-C is unlikely to be a psoriasis susceptibility gene." Exp. Dermatol. **12**(3): 307-14.

de Bakker, P. I. W., G. McVean, et al. (2006). "A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC." Nat. Genet. **38**(10): 1166-1172.

Fan, J., A. Oliphant, et al. (2003). "Highly parallel SNP genotyping." Cold Spring Harb. Symp. Quant. Biol. **68**: 69-78.

Fellay, J., K. V. Shianna, et al. (2007). "A Whole-Genome Association Study of Major Determinants for Host Control of HIV-1." Science **317**(5840): 944-947.

Feng, B. J., L. D. Sun, et al. (2009). "Multiple Loci within the major histocompatibility complex confer risk of psoriasis." PLoS Genet. **5**(8): e1000606.

Flint, J. and T. F. C. Mackay (2009). "Genetic architecture of quantitative traits in mice, flies, and humans." Genome Res. **19**(5): 723-733.

Griffin, T. J. and L. M. Smith (2000). "Single-nucleotide polymorphism analysis by MALDI-TOF mass spectrometry." Trends Biotechnol. **18**(2): 77-84.

Hinds, D. A., L. L. Stuve, et al. (2005). "Whole-Genome Patterns of Common DNA Variation in Three Human Populations." Science **307**(5712): 1072-1079.

Horvath, S., X. Xu, et al. (2001). "The family based association test method: strategies for studying general genotype--phenotype associations." Eur. J. Hum. Genet. **9**(4): 301-306.

Kaleab, B., T. Ottenoff, et al. (1990). "Mycobacterial-induced cytotoxic T cells as well as nonspecific killer cells derived from healthy individuals and leprosy patients." Eur. J. Immunol. **20**(12): 2651-9.

Lee, L. G., C. R. Connell, et al. (1993). "Allelic discrimination by nick-translation PCR with fluorogenic probes." Nucleic Acids Res. **21**(16): 3761-6.

Legare, M. E., F. S. Bartlett, et al. (2000). "A Major Effect QTL Determined by Multiple Genes in Epileptic EL Mice." Genome Res. **10**(1): 42-48.

Mackay, T. F. C., E. A. Stone, et al. (2009). "The genetics of quantitative traits: challenges and prospects." Nat. Rev. Genet. **10**(8): 565-577.

Mehra, N. K., R. Rajalingam, et al. (1995). "Variants of HLA-DR2/DR51 group haplotypes and susceptibility to tuberculoid leprosy and pulmonary tuberculosis in Asian Indians." Int. J. Lepr. Other Mycobact. Dis. **63**(2): 241-8.

Miller, E. N., S. E. Jamieson, et al. (2004). "Genome-wide scans for leprosy and tuberculosis susceptibility genes in Brazilians." Genes Immun. **5**(1): 63-67.

Mira, M. T. (2006). "Genetic host resistance and susceptibility to leprosy." Microbes Infect. **8**(4): 1124-31.

Mira, M. T., A. Alcais, et al. (2004). "Susceptibility to leprosy is associated with PARK2 and PACRG." Nature **427**(6975): 636-640.

Mira, M. T., A. Alcais, et al. (2003). "Chromosome 6q25 is linked to susceptibility to leprosy in a Vietnamese population." Nat. Genet. **33**(3): 412-415.

Monot, M., N. Honore, et al. (2005). "On the origin of leprosy." Science **308**(5724): 1040-2.

Monot, M., N. Honore, et al. (2009). "Comparative genomic and phylogeographic analysis of Mycobacterium leprae." Nat. Genet.

Nair, R. P., P. E. Stuart, et al. (2006). "Sequence and Haplotype Analysis Supports HLA-C as the Psoriasis Susceptibility 1 Gene." Am. J. Hum. Genet. **78**(5): 827-851.

Olerup, O. and H. Zetterquist (1992). "HLA-DR typing by PCR amplification with sequence-specific primers (PCR-SSP) in 2 hours: an alternative to serological DR typing in clinical practice including donor-recipient matching in cadaveric transplantation." Tissue Antigens **39**(5): 225-35.

Ridley, D. S. and W. H. Jopling (1966). "Classification of leprosy according to immunity. A five-group system." Int. J. Lepr. Other Mycobact. Dis. **34**(3): 255-73.

Shaw, M. A., I. J. Donaldson, et al. (2001). "Association and linkage of leprosy phenotypes with HLA class II and tumour necrosis factor genes." Genes Immun. **2**(4): 196-204.

Singh, M., A. Balamurugan, et al. (2007). "Immunogenetics of mycobacterial infections in the North Indian population." Tissue Antigens **69 Suppl 1**: 228-30.

Tai Wai, Y., L. D. J. Philip, et al. (2007). "A second major histocompatibility complex susceptibility locus for multiple sclerosis." Ann. Neurol. **61**(3): 228-236.

Vanderborght, P. R., A. G. Pacheco, et al. (2007). "HLA-DRB1[ast]04 and DRB1[ast]10 are associated with resistance and susceptibility, respectively, in Brazilian and Vietnamese leprosy patients." Genes Immun. **8**(4): 320-324.

Yen, J.-H., B. E. Moore, et al. (2001). "Major Histocompatibility Complex Class I-recognizing Receptors Are Disease Risk Genes in Rheumatoid Arthritis." J. Exp. Med. **193**(10): 1159-1168.

Zerva, L., B. Cizman, et al. (1996). "Arginine at positions 13 or 70-71 in pocket 4 of HLA-DRB1 alleles is associated with susceptibility to tuberculoid leprosy." J. Exp. Med. **183**(3): 829-36.

Chapter 4 │ Genetic and functional analysis of common *MRC1* exon 7 polymorphisms in leprosy susceptibility

While Chapters 2 and 3 describe the genetic mapping of the HLA complex, Chapter 4 describes the genetic and functional study of *MRC1* in leprosy, a strong candidate for the risk locus underlying the replicated linkage peak on chromosome region 10p13. Significant – but modest – evidence for association of a non-synonymous SNP in exon 7 (G396S) with leprosy *per se* (i.e., irrespective of clinical sub-type) was shown in a family-based sample from Vietnam and a population-based sample from Brazil. The results establish *MRC1* as a leprosy susceptibility gene but preclude *MRC1* as the basis of the linkage signal which was restricted to the paucibacillary sub-type.

Genetic and functional analysis of common *MRC1* exon 7 polymorphisms in leprosy susceptibility

Andrea Alter[1,7], Louis de Léséleuc[1,7], Nguyen Van Thuc[2], Vu Hong Thai[2], Nguyen Thu Huong[2], Nguyen Ngoc Ba[2], Cynthia Chester Cardoso[3], Audrey Virginia Grant[4,5], Laurent Abel[4,5,6], Milton Ozório Moraes[3], Alexandre Alcaïs[4,5] and Erwin Schurr[1]

[1]Departments of Human Genetics and Medicine, McGill Centre for the Study of Host Resistance, Research Institute of the McGill University Health Centre, McGill University, Montreal, QC, Canada

[2]Hospital for Dermato-Venereology, Nguyen Thong Street, District 3, Ho Chi Minh City, Vietnam

[3]Leprosy Laboratory, Oswaldo Cruz Institute, FIOCRUZ, Rio de Janeiro, Brazil

[4]Laboratoire de Génétique Humaine des Maladies Infectieuses, Institut National de la Santé et de la Recherche Médicale U550, 75015 Paris, France

[5]Faculté Médicine Necker, Université Paris Descartes, 75015 Paris, France

[6]Laboratory of Human Genetics of Infectious Diseases, Rockefeller Branch, The Rockefeller University, New York, NY 11065, USA

[7]These authors contributed equally to this work.

Abstract

The chromosomal region 10p13 has been linked to paucibacillary leprosy in two independent studies. The *MRC1* gene, encoding the human mannose receptor (MR), is located in the 10p13 region and non-synonymous SNPs in exon 7 of the gene have been suggested as leprosy susceptibility factors. We determined that G396S is the only non-synonymous exon 7-encoded polymorphism in 396 unrelated Vietnamese subjects. This SNP was genotyped in 490 simplex and 90 multiplex leprosy families comprising 704 patients (47% paucibacillary; 53% multibacillary). We observed significant under-transmission of the serine allele of the G396S polymorphism with leprosy per se ($P = 0.036$) and multibacillary leprosy ($P = 0.034$). In a sample of 384 Brazilian leprosy cases (51% paucibacillary; 49% multibacillary) and 399 healthy controls, we observed significant association of the glycine allele of the G396S polymorphism with leprosy per se ($P = 0.016$) and multibacillary leprosy ($P = 0.023$). In addition, we observed a significant association of exon 7 encoded amino acid haplotypes with leprosy per se ($P = 0.012$) and multibacillary leprosy ($P = 0.004$). Next, we tested HEK293 cells over-expressing MR constructs (293-MR) with three exon 7 haplotypes of *MRC1* for their ability to bind and internalize ovalbumin and zymosan, two classical MR ligands. No difference in uptake was measured between the variants. In addition, 293-MR failed to bind and internalize viable *Mycobacterium leprae* and BCG. We propose that the MR-*M. leprae* interaction is modulated by an accessory host molecule of unknown identity.

Introduction

Leprosy, caused by the human pathogenic bacterium *Mycobacterium leprae* (*M. leprae*) is a chronic infectious disease that has been recognized for centuries. Leprosy is primarily a disease of the skin and nerves but the mechanisms of pathogenesis that lead to cutaneous spread of bacilli and irreversible peripheral nerve damage are poorly understood. The disease manifests itself in two clinical forms that have been termed "paucibacillary" and "multibacillary" leprosy depending on the number of skin lesions and the number of bacilli found in these lesions. The development of leprosy is strongly influenced by host genetics as illustrated by the high concordance rates among monozygotic twins {Chakravartti and Vogel 1973}. Consequently, a strategy to better understand the disease pathogenesis is to uncover genes that modulate leprosy susceptibility. Past genome-wide linkage analyses have identified several susceptibility loci in chromosome regions 6q25-q26 {Mira, Alcais et al. 2003}, 6p21 {Mira, Alcais et al. 2003; Miller, Jamieson et al. 2004}, 10p13 {Siddiqui, Meisner et al. 2001}, 17q22 {Miller, Jamieson et al. 2004}, 20p12-13 {Tosh, Meisner et al. 2002; Miller, Jamieson et al. 2004} and 21q22 {Wallace, Fitness et al. 2004}. So far, high density linkage disequilibrium (LD) mapping of the linkage peaks on chromosome regions 6q25-q26 and 6p21 has successfully identified two susceptibility loci by positional cloning, *PARK2*/*PACRG* {Mira, Alcais et al. 2004} and *LTA* {Alcais, Alter et al. 2007}, respectively.

The 10p13 chromosomal region carries a replicated susceptibility locus for paucibacillary leprosy {Siddiqui, Meisner et al. 2001; Mira, Alcais et al. 2003}. The *MRC1* gene underlies the corresponding linkage peak and three closely spaced amino acid changes encoded by exon 7 of *MRC1* have been suggested to be associated with altered susceptibility to paucibacillary leprosy {Hill 2006; Cooke and Hill 2008}. *MRC1* encodes the mannose receptor C-type lectin (MR), a

cell surface protein that belongs to a family of receptors for pathogen-associated molecular patterns, or PAMPs, and is part of the innate arm of the immune system. MR is a receptor for mannose, fucose and *N*-acetyl-glucosamine-containing molecules and binds to yeast cell walls {Ezekowitz, Sastry et al. 1990}, to capsular polysaccharides of *Klebsiella pneumonia* {Kabha, Nissimov et al. 1995} as well as to mycobacterial components such as mannose-capped lipoarabinomannans (ManLAM) {Schlesinger, Kaufman et al. 1996}. MR has been shown to be an important mediator between *M. tuberculosis* and the host immune system {Schlesinger 1993; Nigou, Zelle-Rieser et al. 2001; Kang, Azad et al. 2005}. Most notably, cellular entry through MR has been correlated with *M. tuberculosis* virulence {Schlesinger 1993; Schlesinger, Kaufman et al. 1996}. Taken together, these data prompted us to study *MRC1* as a possible leprosy susceptibility gene.

We analyzed the role of *MRC1* exon 7 non-synonymous coding polymorphisms in leprosy susceptibility in a familial sample from Vietnam and a case-control sample from Brazil. We observed in both populations that the G396S polymorphism was significantly associated with leprosy per se and multibacillary leprosy. However, we found in the Brazilian sample that the risk effect of the glycine residue at position 396 depended on the exon 7 haplotype. In functional analysis, we failed to observe an impact of the exon 7 polymorphisms on MR mediated uptake of zymosan and ovalbumin. In addition, HEK293 cells over-expressing MR (293-MR) were unable to bind and ingest viable *M. leprae* and *Bacille Calmette Guerin* (BCG) bacteria. We concluded that exon 7 polymorphisms mediated the interaction of MR with an unknown accessory molecule that is required for MR-mediated uptake of *M. leprae*.

Methods

Patients and controls

In Vietnam, 490 simplex (53% multibacillary) and 90 multiplex (55% multibacillary) leprosy families were identified from the records available at the Dermato-Venereology Hospital in Ho Chi Minh City {Mira, Alcais et al. 2003; Alcais, Alter et al. 2007}. The criterion for enrollment was the availability of both parents for genetic analysis. In Brazil, 384 leprosy patients were recruited from the Leprosy Outpatient Clinic at the Oswaldo Cruz Institute in Rio de Janeiro. The 399 healthy controls (i.e., no reported infectious or inflammatory disease) were recruited from the same geographic area of Rio de Janeiro. Cases and controls were matched based on self-reported ethnicity. In previous experiments, using genomic controls, we failed to observe significant evidence for population stratification in case-control samples matched on self-reported ethnicity from the same geographic area {Mira, Alcais et al. 2004}. The diagnosis of all leprosy patients and the definition of clinical sub-forms were based on clinical and histological criteria {Ridley and Jopling 1966}. Informed consent was obtained from all study participants. The study was approved by institutional review boards and health authorities in Ho Chi Minh City, Vietnam, the Oswaldo Cruz Institute, Rio de Janeiro, Brazil, and the Research Institute of the McGill University Health Centre, Montreal, Canada.

SNP selection

The identification of informative coding SNPs (i.e., minor allele frequency (MAF) >5%) was done by sequencing 23 unrelated Vietnamese individuals for all 30 *MRC1* exons, including both untranslated (UTR) regions, using an ABI PRISM® 3100 genetic analyzer. In addition, exon 7 was sequenced in 396 unrelated Vietnamese individuals (a subset of parents from the 490

Vietnamese simplex families). For subsequent analysis, the six identified informative coding SNPs were designated by their corresponding 'rs' numbers. In total, 75 SNPs spanning *MRC1* (101.8 kb) on chromosome region 10p12.33 were selected based on allelic frequencies publicly available from the NCBI EntrezSNP database (http://www.ncbi.nlm.nih.gov/sites/entrez?db=snp) and the International HapMap project (http://www.hapmap.org/) (**Table S1**). Conflicting annotations between physical maps regarding gene localization and supposed gene duplication (i.e., *MRC1* and *MRC1L1*) posed technical challenges to the genetic analysis of *MRC1*, particularly the resolution of SNP positions. For consistency, SNPs with two positions were mapped to *MRC1* spanning chr10:17,891,368-17,993,183 (HapMap Data Rel 27) and were referred to by their *MRC1L1* 'rs' numbers (NCBI EntrezSNP database Build 130). We analyzed *MRC1* exon 7 chromatograms in 2,423 unrelated individuals (from our Vietnamese and Brazilian samples and the HGDP-CEPH Human Genome Diversity Cell Line Panel) and found no evidence for a common gene duplication event (i.e. biallelic ratio for SNPs was 1:1). The data argues against the duplication of *MRC1* and suggests that *MRC1L1* is an erroneous annotation caused by the presence of a sequence gap and the incorrect assignment of a polymorphic haplotype.

Genotyping methods

The SNPs were genotyped on one or several of the following platforms: (1) genotyping by direct sequencing on an ABI PRISM® 3100 genetic analyzer; (2) genotyping on the high-throughput GenomeLab™ SNPstream® platform (formerly Orchid SNPstream UHT), which uses a single-base pair extension (SBE) method to incorporate fluorescently labeled terminal nucleotides, which are then detected by a specialized imager {Bell, Chaturvedi et al. 2002}; (3) genotyping on the high-throughput SEQUENOM® MassARRAY® platform, which uses the iPLEX™ assay to

incorporate mass-modified terminal nucleotides in the SBE step, which are then detected by MALDI-TOF MS {Griffin and Smith 2000}; (4) genotyping on the ultra-high throughput Illumina® platform. This platform uses the GoldenGate™ assay followed by a bead-based technology to resolve individual SNP genotypes {Fan, Oliphant et al. 2003}. Three SNPs, including rs1926736, were genotyped by two independent methods and the few individuals (<1.7%) with discrepant genotypes were eliminated from subsequent analyses. A total of 15 SNPs could not be successfully genotyped and one SNP could not be placed unambiguously on the sequence map. We excluded 12 of the remaining 59 SNPs from the analysis for the following reasons: 1 showed deviations ($P < 0.01$) from Hardy-Weinberg equilibrium among parents and 11 were non-informative or had a MAF < 5%.

Reagents and antibodies

Dulbecco's modified Eagle Medium, Minimum Essential Medium (MEM) alpha, RPMI-1640, GlutaMAX, penicillin, streptomycin, fetal bovine serum (FBS), AlexaFluor 488-conjugated ovalbumin (OVA488), FITC-conjugated zymosan, rabbit anti-FITC, AlexaFluor 488-conjugated goat anti-mouse antibody and 4',6-diamidino-2-phenylindole (DAPI) were purchased from Invitrogen (Carlsbad, CA, USA). Endothelial cell medium (ECM) was from Sciencell (Carlsbad, CA, USA). Puromycin, yeast mannan, digitonin, paraformaldehyde, polybrene and porcine gelatin were purchased from Sigma (St.-Louis, MO, USA). Anti-MR monoclonal antibody (clone 15-2) was purchased from Cell Sciences (Canton, MA, USA). Cy3-conjugated goat anti-mouse and Cy5-conjugated goat anti-rabbit antibodies were from Jackson ImmunoResearch (West Grove, PA, USA). Permafluor mounting medium was purchased from Thermo Scientific (Waltham, MA, USA). YG fluorescent 1 μm polystyrene microspheres were obtained from Polysciences Inc. (Warrington, PA, USA).

Cell culture

The HEK293 cell line was purchased from the American Type Culture Collection (ATCC, Manassass, VA, USA) and grown in MEM-alpha. Phoenix-Ampho cells were purchased from ATCC with permission of Gary Nolan (Stanford University, Palo Alto, CA, USA). These media were supplemented with 10% heat-inactivated FBS, 100 U/ml penicillin and 100 U/ml streptomycin and 2 mM GlutaMAX. Human dermal microvascular endothelial cells (HDMECs) were purchased from Sciencell and cultured in ECM containing 5% FBS, 100 U/ml penicillin and 100 U/ml streptomycin, in gelatin-coated flasks. All cells were cultured in a humidified incubator at 37°C and 5% CO_2. When mycobacteria were added, cells were cultured under the same conditions except that antibiotics were omitted and in the case of *M. leprae*, the temperature was set to 33°C.

Expression vector cloning

The cDNA encoding the human mannose receptor and cloned in pCDNA3 was a generous gift of Dr. J. J. He (Indiana University, IN, USA). The MR cDNA was amplified by proofreading PCR using primers MR_fwd_SalI, 5'-ATGCGTCGACATGAGGCTACCCCTGCTCC-3', and MR_rev_MfeI, 5'-GCATCAATTGCTAGATGACCGAGTGTTCA-3'. The PCR product was purified, digested with *Sal*I and *Mfe*I and ligated in the pMSCV-puro vector (Clontech, Mountain View, CA, USA) previously cut with *Xho*I and *Eco*RI. A total of 2 µl of the ligation was used to transform STLB2 competent cells (Invitrogen). A positive clone was grown overnight at 30°C in 50 ml of culture broth and the plasmid DNA was extracted using a Perfectprep plasmid midi kit (Eppendorf, Hamburg, Germany). The nucleotide sequence was verified by cycle sequencing in an ABI PRISM® 3100 genetic analyzer (Applied Biosystems, Foster City, CA, USA).

Site-directed mutagenesis

Two polymorphic variants of *MRC1* were generated using the original hMR cDNA as a template for site-directed mutagenesis by overlap extension PCR {Boskovic, Arnold et al. 2006}. MR_fwd_SalI and MR_rev_MfeI were used as the flanking universal primers. Primers 5'-GCAGGAAGGAAGGCAGTGACCTCGCGAGTATCC-3' (sense) and 5'-GGATACTCGCGAGGTCACTGCCTTCCTTCCTGC-3' (antisense) were used for the generation of the MR(SAF) haplotype. Primers 5'-AAGTATCCACACCATCGAGGAATTGGACTTT-'3 (sense) and 5'-ATGGTGTGGATACTTGTGAGGTCACCGCCTT-3' were used for the generation of the MR(GTL) haplotype. All constructs were cloned in pMSCV-puro and their sequence was verified.

Transduction

Amphotropic retroviruses were produced by transfecting the Phoenix-Ampho cell line with pMSCV-puro (empty) or pMSCV-MR constructs using Fugene6 (Roche Diagnostics, Laval, QC, Canada) according to the manufacturer's instructions. Forty-eight hours after transfection, the culture supernatant was harvested, filtered, aliquoted and stored at -80°C. Aliquots showed similar (within margin of error) titers when tested on HEK293 cells. For transductions, HEK293 cells were seeded and grown to 50% confluence, then infected with Phoenix-Ampho supernatants supplemented with 8 μg/ml of polybrene for 24 h. The medium was then removed and fresh medium containing 1 μg/ml of puromycin was added. Stable pools of transductants (293-puro and 293-MR) were collected after 2 weeks of selection.

Immunostaining

Expression of MR was analyzed by indirect immunofluorescence. For microscopy, cells were seeded on gelatin-coated coverslips and fixed with 4% paraformaldehyde (PFA) diluted in PBS for 10 min. Alternatively, single cell suspensions were prepared for cytometry and fixed in 4% PFA. Cells were then incubated for 30 min in SuperBlock blocking buffer (Pierce, Rockford, IL, USA) containing 0.1% Tween-20, then incubated with 1 μg/ml of anti-MR mouse monoclonal antibody for 16 h at 4°C in the same buffer. Cells were rinsed twice in PBS and incubated with either Alexa Fluor 488- or Cy3-conjugated secondary antibodies (1:2000 dilution in SuperBlock) for 60 min at 24°C. For microscopy, coverslips were rinsed twice then mounted onto slides in Permafluor mounting medium containing DAPI. For cytometry, suspensions were processed in a FACSCalibur flow cytometer (BD Biosciences).

Ovalbumin uptake assay

Cell lines were seeded in 12-well plates at a density of 5×10^5 cells/well and left to attach overnight. They were then treated for 1 h with 5 μg/ml of OVA488 in fresh complete medium. Cells were washed twice in PBS, detached in PBS containing 0.2 g/L EDTA and processed for flow cytometry. Fluorescence was measured in the FL1 channel.

Zymosan uptake assay

Cell lines were seeded in 12-well plates at a density of 5×10^5 cells/well and left to attach overnight. They were treated with zymosan-FITC at a ratio of five particles per cell for 16 h in fresh complete medium. Cells were washed twice in PBS then fresh complete medium containing 1 μg/ml anti-FITC antibody was added. Cells were kept at 4°C for 30 min to stop phagocytosis and allow binding of antibodies to surface exposed zymosans. Cells were washed twice in PBS and lysed for 10 minutes in 400 μl lysis buffer (0.01% digitonin in PBS) containing

10^6 YG 1 μm beads (about 1 bead per cell) and anti-rabbit Cy5 diluted 1:200. The lysate was further homogenized in a bath sonicator for 5 minutes. Samples were processed by flow cytometry, surface-bound particles were positive for both the FL1 (green) and FL4 (far red) channels whereas internalized zymosans were scored negative in FL4.

Mycobacteria binding assay

PKH67-stained viable *M. leprae* bacilli grown in *nude* mouse footpads were obtained from Dr R. Truman and Dr L. Adams (Hansen's Disease Program; Baton Rouge, LA, USA). GFP-expressing *M. bovis* BCG strain Pasteur was cultured in Middlebrook 7H9 medium containing ADC enrichment and used fresh to make single-cell suspensions. The 293-puro or 293-MR cells were seeded on coverslips at a density of 2.5×10^5 cells/well, in complete medium without antibiotics then treated with mycobacteria at a multiplicity of infection (MOI) of five for 16 h. In some experiments, cells were pre-incubated for 30 min with yeast mannan or anti-MR monoclonal antibody. Cells were washed three times in PBS then fixed in 4% paraformaldehyde and briefly permeabilized in cold methanol. Nuclei were counterstained with DAPI, then coverslips were mounted and examined by fluorescence microscopy. Images of both green particles and nuclei were acquired using the Northern Eclipse software (Mississauga, ON, USA). Bound bacteria were manually counted in at least 4 fields containing a minimum of 500 cells total.

Statistical methods

We estimated population allelic frequencies from parental data using the algorithm implemented in Haploview 4.1 software {Barrett, Fry et al. 2005}. Family based association studies were performed using principally a classical transmission disequilibrium test, as implemented in FBAT v2.0.3 software {Horvath, Xu et al. 2001}. For the analysis of the multi-case families and the

combined sample, we used the empirical variance–covariance estimator previously advocated {Lake, Blacker et al. 2000}. We carried out a population-based association study in the Brazilian sample, using classical multivariate logistic regression techniques as implemented in the LOGISTIC procedure of SAS software v.9.1. Haplotype analyses were further carried out using the THESIAS software {Tregouet and Garelle 2007}. All population-based analyses were adjusted for sex, which has been identified as a classical risk factor for leprosy. We did not correct for multiple testing given the context of a positive replication utilizing an alternative study design in an ethnically different population, and the reported association of the *MRC1* exon 7 SNPs with leprosy in two reviews {Hill 2006; Cooke and Hill 2008}.

Results

Analysis of exon 7 SNPs in Vietnamese leprosy families

Three non-synonymous exon 7 SNPs were of particular interest since they had been suggested as risk factors for paucibacillary leprosy in a South Indian population {Cooke and Hill 2008}. To investigate the impact of *MRC1* exon 7 SNPs on leprosy susceptibility, we first sequenced exon 7 in 396 unrelated healthy individuals (792 chromosomes) to identify any common population-specific variants. We found two polymorphic SNPs: the non-synonymous rs1926736 [G396S; allele A (S396) MAF = 0.35] and the synonymous rs2437256 [I404I; allele C (I404) MAF = 0.21]. Two additional previously described non-synonymous SNPs, rs2478577 (T399A) and rs2437257 (L407F), were found to be non-polymorphic in the Vietnamese population (MAF = 0). Since the focus of the study was on non-synonymous SNPs, we genotyped the single non-synonymous SNP rs1926736 in 490 simplex families and 90 multiplex families comprising a total of 704 leprosy patients for which both parents were available. These families included approximately even numbers of paucibacillary (n = 325) and multibacillary (n = 374) cases. Under a best dominant model, we observed significant evidence for the association of rs1926736 with leprosy per se [P_{dom} = 0.035, allele A (S396) is protective, OR = 0.76 (95% CI 0.60-0.96)] and multibacillary leprosy [P_{dom} = 0.034, allele A (S396) is protective, OR = 0.71 (0.51-0.99)] (**Table 1**).

To delineate the extent of LD between rs1926736 and additional *MRC1* SNPs, we excluded rare SNPs (i.e., MAF < 5%) given that they cannot be highly correlated with rs1926736 (MAF = 0.35). We selected 69 SNPs from the NCBI EntrezSNP database and the International HapMap project that span the entire 101.8 kb of the *MRC1* gene. To assure inclusion of all common

coding SNPs, we sequenced the remaining 29 *MRC1* exons including the 5'- and 3'-UTR regions in 23 unrelated Vietnamese individuals. According to our criterion, four identified coding SNPs, all previously described in the NCBI EntrezSNP database, were included: rs2253120 (exon 2, T81T), rs2296414 (exon 3, T167I), rs2985837 (exon 4, Q242Q) and rs2477664 (exon 4, I260I) (**Table S2**). Including the two previously described exon 7 SNPs, a total of 75 *MRC1* SNPs were genotyped in 198 simplex families to establish the LD pattern of the *MRC1* gene. Among the 75 *MRC1* SNPs genotyped, 47 were suitable for analysis (**Table S1**). The LD pattern of the entire *MRC1* gene was plotted and revealed that rs1926736 was highly correlated with only two intronic SNPs (rs525830, $r^2 = 0.89$ and rs493862, $r^2 = 0.86$) with no obvious functional implication (**Fig. 1**; **Table S3**). In addition, none of the additional SNPs showed significant evidence for association with leprosy or its sub-forms in the 198 families (data not shown). Finally, we utilized the International HapMap Project database (Data Rel 24) to screen a 2-Mb region around rs1926736 in the Han Chinese in Beijing, China (CHB) samples. No SNPs were correlated with rs1926736 using an $r^2 = 0.5$ cut-off. From these results, we concluded that rs1926736 (G396S) was the most plausible cause of the observed association and that S396 of MR was a leprosy protective factor.

Analysis of exon 7 SNPs in Brazilian leprosy cases

Given that the evidence for association of rs1926736 with leprosy in the Vietnamese samples was weak, we decided to study *MRC1* exon 7 SNPs in a leprosy sample from Brazil. We sequenced exon 7 in 399 healthy controls and observed four SNPs: rs1926736 [G396S; allele A (S396) MAF = 0.32], rs2478577 [T399A; allele A (T399) MAF = 0.21], rs2437256 [I404I; allele C (I404) MAF = 0.21], and rs2437257 [L407F; allele G (L407) MAF = 0.21]. The three non-synonymous exon 7 SNPs could be unambiguously assigned to three amino acid haplotypes, two of which had

been observed in the Vietnamese sample (**Fig. 2**). Next, we sequenced *MRC1* exon 7 in an additional 384 Brazilian leprosy cases and conducted a case-control association analysis of the three non-synonymous SNPs with leprosy. LD analysis showed that SNPs rs2478577 (T399A) and rs2437257 (L407F) were in complete LD ($r^2 = 1$) and therefore statistically redundant. Allele G (G396) of SNP rs1926736 showed evidence for association with susceptibility to leprosy per se [$P_{add} = 0.016$; OR = 1.34 (95% CI 1.06-1.70)], and the multibacillary sub-form [$P_{add} = 0.023$; OR = 1.42 (1.05-1.93)] in a best additive model. Similarly, allele G (L407) of SNP rs2437257 showed borderline evidence for association with resistance to leprosy per se [$P_{dom} = 0.09$; OR = 0.75 (0.54-1.05)] and the multibacillary sub-form [$P_{dom} = 0.04$; OR = 0.63 (0.41-0.97)] in a best dominant model.

Multivariate analysis of the three non-synonymous variants confirmed the correlation between SNPs rs2478577 and rs2437257. The best sex-adjusted model ($P = 0.002$) as determined by multivariate analysis included only rs1926736 and rs2437257. Haplotype analysis of rs1926736 (G396S) and rs2437257 (L407F) identified three haplotypes with G396-F407 being the most common haplotype (**Table 2**). Haplotype G396-F407 had a risk effect while haplotype S396-F407 was protective in the overall sample (**Table 2**). Considering rs2478577 (T399A) and rs2437257 (L407F) are in complete LD, *MRC1* exon 7 therefore encodes two protective amino-acid haplotypes: the borderline protective G396-T399-L407 haplotype and the strongly protective S396-A399-F407 haplotype (**Table 2**). Interestingly, while the G396 residue is a susceptibility factor overall, on the T399-L407 haplotype background, this susceptibility can be compensated and may even be protective. This is concurrent with the result in the Vietnamese leprosy families. Given that in the Vietnamese population rs2478577 (T399A) and rs2437257 (L407F) are non-

polymorphic, the protective S396 allele corresponds unambiguously to the S396-A399-F407 haplotype (**Fig. 2**).

Establishment of an HEK293 cellular model for MR function

The three *MRC1* exon 7 encoded amino acid haplotypes observed in Vietnamese and Brazilian samples were termed MR(GAF), MR(SAF) and MR(GTL) (**Fig. 2**). The underlying three non-synonymous amino acid changes G396S, T399A and L407F map to the CTLD2 segment of the mannose receptor protein, a domain of unknown function. To better understand the functional basis of the observed genetic effect of exon 7 haplotypes, we decided to compare the biological activity of the corresponding MR genetic variants by ectopically expressing the receptor in a non-expressing cell line. A cDNA copy of human *MRC1* in pcDNA3 (a generous gift of Dr. J. J. He, Indiana University, IN, USA) was obtained and subcloned in a MSCV-puro retroviral backbone. Packaged viruses were used to infect the HEK293 cell line, which does not express MR. The resulting pool of transductants (293-MR) was compared to an empty-vector control HEK293 line as well as to HDMEC, which are known to express MR {Groger, Holnthoner et al. 2000}. In addition to MR protein expression, the capacity to internalize fluid-phase (ovalbumin) and particulate (zymosan) ligands was evaluated by microscopy. Immunofluorescence showed that 293-MR and endothelial cells mostly expressed the mannose receptor in an intracellular endocytic compartment as well as on their surface, which overlapped the pattern of ovalbumin uptake (**Fig. 3a**). The control cell line did not stain for MR and was unable to take up ovalbumin. In addition, both 293-MR and HDMECs could bind and internalize yeast particles (zymosans) (**Fig. 3b**). Hence, the 293-MR cell line was considered a good model for the study of ectopically expressed MR since it mirrored the natural activity of the receptor.

Cloning and expression of human mannose receptor variants

The original *MRC1* cDNA sequence used in the generation of the 293-MR cell line corresponded to the MR(GAF) haplotype. This sequence was modified by site-directed mutagenesis at two sites to yield the MR(SAF) and MR(GTL) variants (**Fig. 2**). The DNAs were cloned in a retroviral expression vector, packaged and used to infect HEK293 cells. The entire pools of puromycin-resistant transduced cells were used at passage 3 for further experiments to avoid artifacts due to clonal variations. The proportion and level of MR expression were measured by immunocytometry (**Fig. 4a**). MR(GAF) and MR(SAF) pools had very similar expression levels whereas the MR(GTL) pool was slightly lower, possibly due to a somewhat lower retroviral transduction efficiency during the generation of this cell line. Expression levels were monitored throughout the experiments and were found to be stable. Vector-transduced control cells showed no expression.

MR variants do not differ in their capacity to take up ligands.

The three pools of MR-expressing transductants were then used to compare the functional activity of their corresponding MR proteins by employing ovalbumin and zymosan, two classical mannose-containing ligands. Fluid-phase endocytosis of ovalbumin was measured by flow cytometry. There was no difference in endocytosis between MR(GAF) and MR(SAF) variants and only a slightly lower activity for MR(GTL) that was consistent with a somewhat lower expression (**Fig. 4b**). Vector control cells showed only background level uptake. It has been suggested that the processes of MR-dependent binding and endocytosis of soluble molecules might differ from MR-dependent phagocytosis of particles {Le Cabec, Emorine et al. 2005}. Therefore, we simultaneously measured binding and internalization of fluorescinated zymosans

by flow cytometry. Ingestion efficiency was determined by the ratio of internalized particles over total cell-associated particles. Again, MR(GAF) and MR(SAF) variants were indistinguishable while the lower expressing MR(GTL) variant ingested slightly lower numbers (**Fig. 4b**). Overall, the MR-expressing cells took up 64-83% of bound particles after 16 h of incubation. Vector control cells did not bind sufficient numbers of particles to allow quantification of phagocytosis.

MR expressed by HEK293 cells does not bind to *M. leprae* or BCG

Mycobacteria are thought to bind to MR because of the presence of mannosylated molecules such as ManLAM on their surface {Schlesinger, Kaufman et al. 1996; Kang, Azad et al. 2005}. We therefore evaluated the impact of MR polymorphisms on the interaction with whole viable *M. leprae* and *M. bovis* BCG. Fluorescent mycobacteria were incubated with 293-puro and 293-MR cells for 16 h. Binding was determined by microscopic examination of the monolayer after repeated washing. The presence of MR had no impact on binding of either mycobacteria (**Fig. 5**). Control cells could bind bacteria to the same extent as MR^+ cells, regardless of the variant expressed. These results suggested that 293-MR cells lacked accessory molecules required for efficient uptake of mycobacteria by MR.

Discussion

Linkage of the 10p13 chromosomal region to the paucibacillary sub-form was observed independently in two ethnically different populations from South India and Vietnam {Siddiqui, Meisner et al. 2001; Mira, Alcais et al. 2003}. In several reviews {Hill 2006; Cooke and Hill 2008} the *MRC1* gene was suggested to be the underlying cause of this linkage signal. While experimental details are not yet publicly available, three exon 7 encoded amino acid polymorphisms were reportedly associated with paucibacillary leprosy in South Indian patients. Based on these reports, we decided to study the role of *MRC1* in susceptibility to leprosy and its clinical sub-forms. In a large sample of Vietnamese patients, we detected weak but significant evidence that serine at amino acid position 396 was a protective factor for leprosy per se and multibacillary leprosy. These results strongly argued that *MRC1* is indeed a leprosy susceptibility gene but not the paucibacillary susceptibility gene detected in the genome-wide scan in Vietnamese families {Mira, Alcais et al. 2003}. The results obtained in the Brazilian sample were more complex. In the Brazilian population, we detected three non-synonymous exon 7 SNPs. As suggested for the South Indian patients, an exon 7 encoded S396-A399-F407 haplotype was significantly protective for leprosy. Similar to the results in the Vietnamese families, the protective effect of this haplotype was significant in the overall sample and in the subset of multibacillary patients. Importantly, in both the Vietnamese and Brazilian samples, the effect of the exon 7 polymorphisms is on leprosy per se. Although in subset analysis formal significance of association was only obtained for multibacillary leprosy, there was no significant heterogeneity between paucibacillary and multibacillary cases. The limited effect of *MRC1* detected by our association study is consistent with the absence of linkage (i.e., a major gene effect) detected between the 10p13 region and leprosy per se. These results strongly argue that

MRC1 is indeed a leprosy susceptibility gene but does not explain the observed linkage peak in the South Indian patients {Siddiqui, Meisner et al. 2001} or Vietnamese families {Mira, Alcais et al. 2003}. A high-density association scan of the underlying interval is necessary to identify the paucibacillary risk factor(s).

Although our data pinpoint G396S (rs1926736) as the functional variant, we cannot exclude the possibility that multiple rare coding variants in exons 1-30 contribute to leprosy susceptibility (i.e., allelic heterogeneity). However, the practicality and value of sequencing 30 exons in a large sample of individuals in order to identify these variants is uncertain. Unless a sufficient number of rare variants are identified, a 'meta-SNP' analysis combining the information at each loci would not have sufficient power to detect a genetic effect in the expected range of odds-ratios. Supporting G396S as the functional variant, a recent genetic study found rs1926736 to be associated with asthma risk in 446 Japanese cases (P = 0.011, OR = 0.61 [0.41-0.89]) {Hattori, Konno et al. 2009}. Moreover, the most significant haplotype association was observed for a five-SNP risk haplotype carrying allele G396 (P = 0.00047). The authors concluded that the haplotype association was seemingly driven by rs1926736 in the Japanese population. This study not only substantiates our results but suggests that *MRC1* – particularly G396S –participates in the immune response to mycobacterial infection and the inflammatory process, two highly related immunological pathways.

The results obtained from the Brazilian sample help to better understand the mechanism of susceptibility mediated by the glycine residue at position 396. Closer inspection of the exon 7 haplotype frequencies in Brazilian patients revealed that the G396-A399-F407 haplotype is enriched in cases while the S396-A399-F407 haplotype is enriched in controls (**Table 2**). In contrast, G396 in the context of the G396-T399-L407 haplotype appears neutral and may even be

slightly protective (**Table 2**). Hence, G396 was a risk factor only on the G396-A399-F407 haplotype background. The question is then why G396 was a susceptibility factor on the A399-F407 background, while on the T399-L407 background it was not? Exon 7 polymorphisms in *MRC1* map to the second C-type lectin domain (CTLD2) of the MR protein. CTLD2 has been structurally assigned to a hinge domain linking the C-type lectin region to the fibronectin type II and cysteine-rich domains {Boskovic, Arnold et al. 2006}. While CTLD2 is homologous to C-type lectins, it does not bind to ligands {Taylor, Bezouska et al. 1992}. Of course, it is possible that changes at CTLD2 indirectly impact on MR ligand affinities. However, our results obtained in the 293-MR cells suggest a direct interaction of CTLD2 with an accessory receptor molecule in the MR-*M. leprae* interaction. In this view, interaction of the CTLD2 domain with the accessory molecule would only be sensitive to G396 in the context of the A399-F407 haplotype.

The importance of accessory molecules for MR function was highlighted by the results of our studies in a cellular model of MR function. To assay MR function, we used the non-professional phagocytic human cell line HEK293 as a platform for expression of MR variants. This cellular model mimicked critical features of MR as seen in HDMECs, previously shown to express this receptor {Groger, Holnthoner et al. 2000}. Although HEK293 cells are not professional phagocytes, expression of MR bestowed on these cells the ability to ingest zymosan particles. We then showed that polymorphisms in the *MRC1* associated with leprosy risk did not impact on the core MR protein functions as measured by internalization of soluble or particulate ligands. All genetic variants analyzed showed no deviation from their expected endocytic or phagocytic activity. This implies that the amino acid polymorphisms encoded by exon 7 of *MRC1* have no impact on the ability to directly bind and internalize mannose-containing ligands. Since mycobacterial ManLAM is known to interact with MR through its terminal

mannooligosaccharide in a similar fashion to other ligands {Schlesinger, Hull et al. 1994}, we expect that the binding of this bacterial product would not be affected by the polymorphism under study. Furthermore, 293-MR cells failed to bind and ingest viable BCG and *M. leprae* indicating that MR alone is not sufficient to mediate mycobacterial phagocytosis. These results are consistent with the hypothesis that the interaction of MR with *M. leprae* involves additional molecules and that differential activity of the MR alleles can only be revealed in the context of this partner host molecule. Such an accessory molecule may be present in macrophages, which are considerably more efficient at internalizing zymosans than 293-MR cells despite a much lower expression of the receptor (data not shown). Unfortunately, macrophages cannot be used easily in a study of MR variants because of their endogenous expression of *MRC1*.

Whether the impact of the leprosy-associated *MRC1* alleles is via the alteration of MR ligand interactions or by changed receptor signaling or trafficking is not known. Hence, it is possible that aspects of MR function not directly related to ligand binding and internalization are affected by genetic CTLD2 variations. For instance, MR delivers an intracellular signal that affects the immune status of cells as well as endosome-lysosome fusion {Nigou, Zelle-Rieser et al. 2001; Kang, Azad et al. 2005; Shimada, Takimoto et al. 2006}. However, this signaling is mediated by the cytoplasmic domain of MR and would not be expected to be affected by a remote extracellular site. Our data strongly suggest that exon 7 variants have an indirect effect on overall structure and/or stability of the CTLD2 domain, likely by modulating the interaction with an additional host molecule(s). This heteromeric complex may then impact on *M. leprae* phagocytosis, MR trafficking and/or MR signaling. Clearly, if our hypothesis is correct, identification of the MR-interacting host cell molecule(s) is of highest importance.

Acknowledgements

We are grateful to all patients and their families who participated in this study. We thank J.L. Casanova, M. Behr and K. Morgan for helpful discussions and A. Montpetit and A. Bélisle for assistance with high-throughput genotyping. Fresh nu/nu mouse derived *M. leprae* were supplied by Dr. R. Truman and Dr. L. Adams of the National Hansen's Disease Program Laboratory Research Branch, Baton Rouge, LA. Funding for this service was provided by the American Leprosy Missions. We thank Dr. J.J. He (Indiana University, IN) for providing the *MRC1* cDNA. This study was supported by a grant from the Canadian Institutes of Health Research (CIHR) to E.S. and MAGRALEPRE from l'Ordre de Malte to Al.A. The Laboratory of Human Genetics of Infectious Diseases is supported by grants from The Rockefeller University Center for Clinical and Translational Science grant number 5UL1RR024143-03 and The Rockefeller University. An.A. holds a graduate studentship from the Natural Science and Engineering Research Council of Canada (NSERC). Al.A. and L.A. are supported by the Assistance Publique-Hôpitaux de Paris, Programme de Recherche Fondamentale en Microbiologie Maladies Infectieuses et Parasitaires (PRFMMIP), and the Agence Nationale de la Recherche (ANR) of the Ministère Français de l'Éducation Nationale de la Recherche et de la Technologie. E.S. is a Chercheur National du Fonds de la Recherche en Santé du Québec and an International Research Scholar of the Howard Hughes Medical Institute.

Figure legend

Figure 1 | Pairwise linkage disequilibrium (LD) as defined by r^2 across the *MRC1* gene

The chromosome 10p location of 47 SNPs from rs2947600 (*MRC*1 intron 1) to rs941 (*MRC1* 3'-UTR) is indicated above the LD plot. The r^2 value for each SNP pair is indicated within the corresponding diamond. Increasing depth of black color indicates higher r^2 values. *D'*-defined haplotype blocks are also indicated at the top of the LD plot. There is no evidence for long range LD extending over the entire *MRC1* gene.

Figure 2 | Human *MRC1* exon 7 haplotypes

In the Brazilian sample, four common *MRC1* exon 7 polymorphisms were found, three of which were non-synonymous (boxed). Due to strong LD between SNPs two (rs2478577), three (rs2437256) and four (rs2437257), only three amino acid haplotypes were observed and termed MR(GAF), MR(SAF) and MR(GTL). SNP 1 (rs1926736) gives rise to a G/S amino acid polymorphism that was only observed on the MR(AF) background. Among these three haplotypes, only two – MR(GAF) and MR(SAF) – were observed in the Vietnamese sample.

Figure 3 | Ectopic MR is functional in HEK293 cells

A MR cDNA was transduced into HEK293 cells. The resulting cells were compared to vector-transduced cells and microvascular endothelial cells (HDMEC) which express endogenous MR. **a** Cells were treated with Alexa-488-conjugated ovalbumin (an MR ligand, shown in *green*) for 1 h then immunostained with monoclonal anti-MR followed by Cy3-conjugated anti-mouse (*red*). **b** Cells were treated with fluorescein-conjugated zymosan for 16 h, then washed and stained with rabbit anti-fluorescein followed by Cy3 anti-rabbit antibodies. Ingested zymosans are *green*

while non ingested particles are stained *red* and appear *yellow* in the merged images. Nuclei are stained blue with DAPI.

Figure 4 │ Absence of biological differences among MR variants

HEK293 transductants were analyzed for MR function by flow cytometry. **a** Transduced stable populations of HEK293 cells were immunostained for MR. **b** Populations were incubated for 1 h with 5 μg/ml Alexa488-conjugated ovalbumin and uptake was measured by FACS. Percentage of positive cells is given as well as median fluorescence value (MFV) of the positive population. **c** Populations were incubated with 10^6 zymosan-FITC particles (ratio zymosan:cell 5:1) for 16 h. Extracellular particles were then stained with anti-FITC and Cy5-conjugated secondary antibodies. Cells were lysed and intracellular and extracellular zymosans were gated in FL1 and counted in FL4 channels. Data is expressed as percent of ingested zymosans over total bound, and as number of bound zymosan per cell. Results shown here are representative of three independent experiments.

Figure 5 │ Absence of impact of ectopic MR expression on binding to mycobacteria

HEK293 transductants were incubated for 16 h with PKH67-stained fluorescent *Mycobacterium leprae* (**a)** and GFP-expressing *Mycobacterium bovis* BCG strain Pasteur **(b)** at a MOI of 5. Measurement of bacterial binding was performed by microscopic examination, counting at least 500 cells in four different fields. Results per field were averaged per cell and are given on the left of the graph. Results are given for HEK293 control cells (PURO) and HEK293 cells over-expressing the MR(GAF), MR(SAF) and MR(GTL) expression constructs.

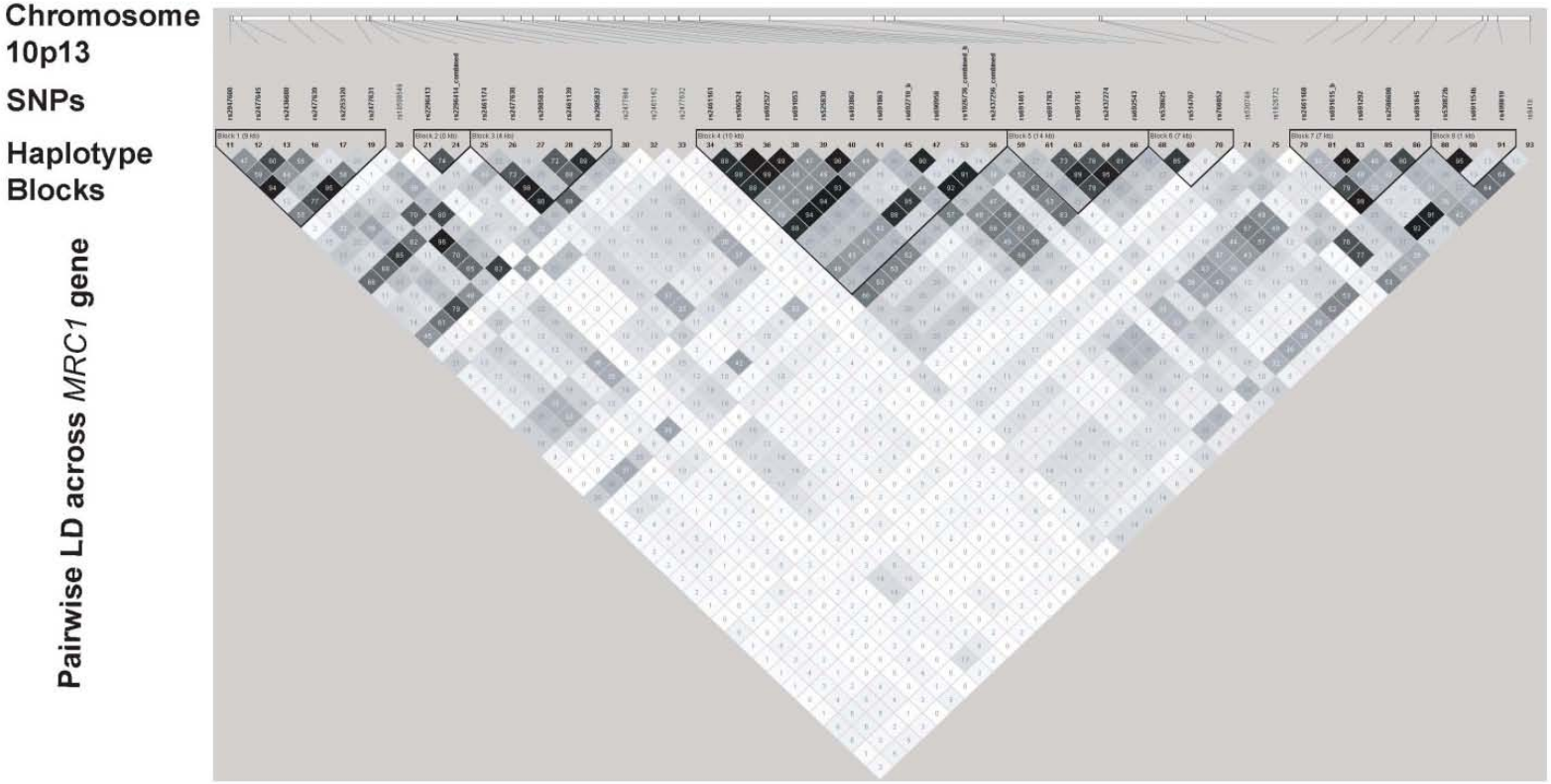

Chromosome 10p13
SNPs
Haplotype Blocks
Pairwise LD across MRC1 gene

Figure 1

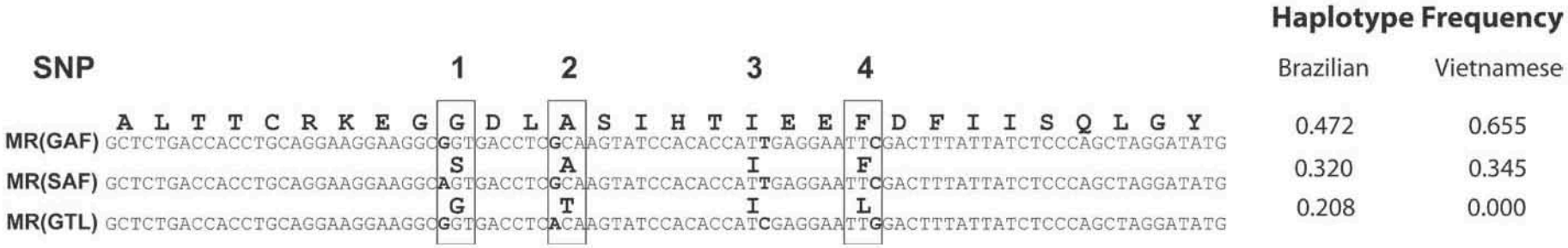

SNP
1
2
3
4
A L T T C R K E G G D L A S I H T I E E F D F I I S Q L G Y
MR(GAF) GCTCTGACCACCTGCAGGAAGGAAGGCGGTGACCTCGCAAGTATCCACACCATTGAGGAATTCGACTTTATTATCTCCCAGCTAGGATATG
MR(SAF) GCTCTGACCACCTGCAGGAAGGAAGGCAGTGACCTCGCAAGTATCCACACCATTGAGGAATTCGACTTTATTATCTCCCAGCTAGGATATG
MR(GTL) GCTCTGACCACCTGCAGGAAGGAAGGCGGTGACCTCACAAGTATCCACACCATCGAGGAATTGGACTTTATTATCTCCCAGCTAGGATATG
S A I F
G T I L
Haplotype Frequency
Brazilian Vietnamese
0.472 0.655
0.320 0.345
0.208 0.000

Figure 2

Figure 3

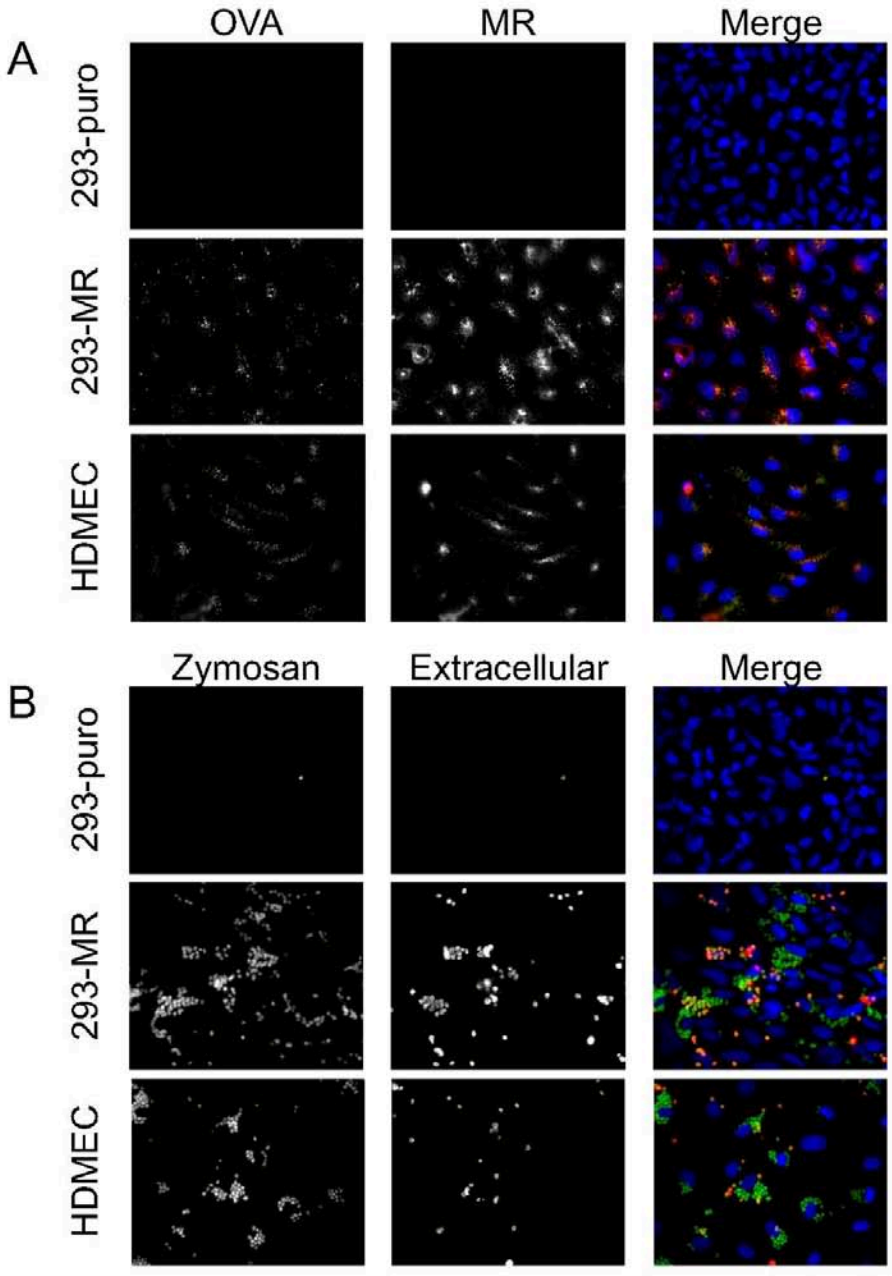

A
OVA
MR
Merge
293-puro
293-MR
HDMEC
B
Zymosan
Extracellular
Merge
293-puro
293-MR
HDMEC

Figure 4

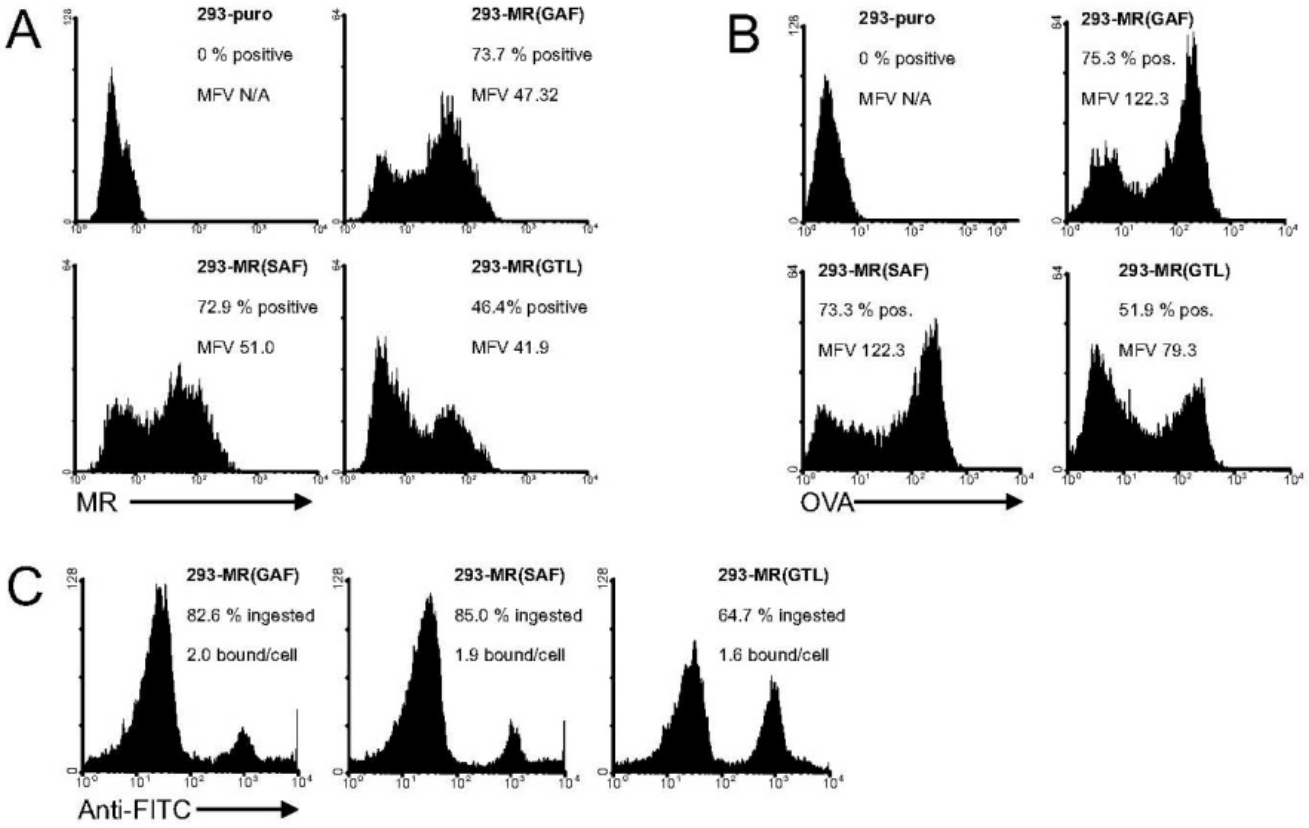

A
293-puro
0 % positive
MFV N/A
293-MR(GAF)
73.7 % positive
MFV 47.32
293-MR(SAF)
72.9 % positive
MFV 51.0
293-MR(GTL)
46.4% positive
MFV 41.9
MR
B
293-puro
0 % positive
MFV N/A
293-MR(GAF)
75.3 % pos.
MFV 122.3
293-MR(SAF)
73.3 % pos.
MFV 122.3
293-MR(GTL)
51.9 % pos.
MFV 79.3
OVA
C
293-MR(GAF)
82.6 % ingested
2.0 bound/cell
293-MR(SAF)
85.0 % ingested
1.9 bound/cell
293-MR(GTL)
64.7 % ingested
1.6 bound/cell
Anti-FITC

Figure 5

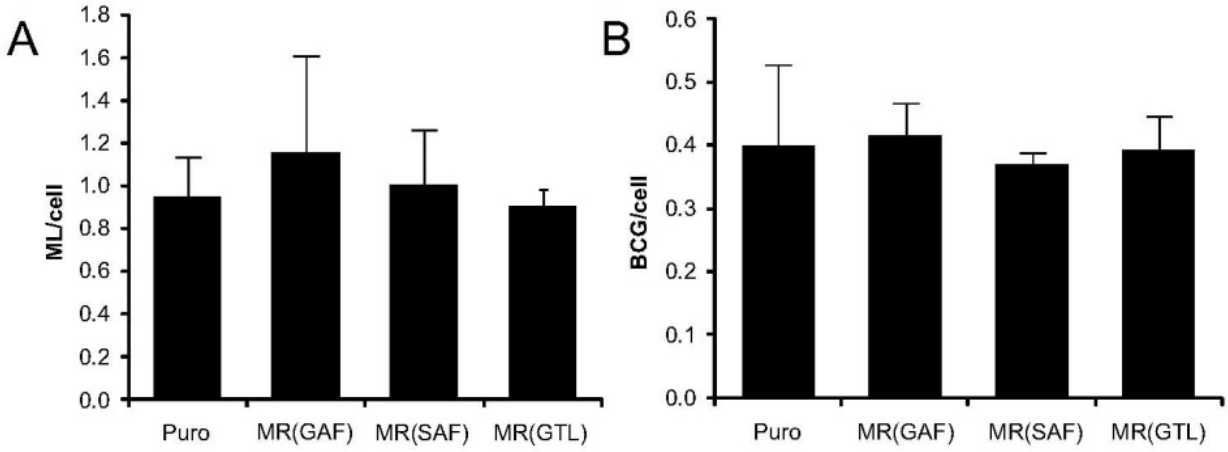

A
ML/cell
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Puro
MR(GAF)
MR(SAF)
MR(GTL)
B
BCG/cell
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Puro
MR(GAF)
MR(SAF)
MR(GTL)

Table 1. Association analysis of *MRC1* exon 7 SNP rs1926736 (G396S) with leprosy and its sub-forms

Phenotype	Number of families	Amino acid (frequency)	*S*[a]	*E(S)*[b]	*Z*[c]	*P* value	OR[d]	95% CI[e]
Per se	339	S (0.370)	202	222	-2.102	0.035	0.76	0.60-0.96
	190	G (0.630)	135	143	-1.069	0.280		
Paucibacillary	172	S (0.369)	106	111.25	-0.823	0.410	0.82	0.58-1.15
	104	G (0.631)	69	73.25	-0.847	0.397		
Multibacillary	186	S (0.371)	100	114.75	-2.125	0.034	0.71	0.51-0.99
	99	G (0.629)	69	73.25	-0.735	0.462		

[a]FBAT statistic

[b]Expected FBAT statistic

[c]Z-score

[d]Odds ratio

[e]95% confidence interval

Table 2. *MRC1* exon7 haplotype analysis in the Brazilian case-control sample

rs1926736–rs2437257 haplotype	Controls[a]	Cases[a]	Global test *P* value[b]	Marginal test *P* value[b]	OR (95% CI)
Per se (343 controls, 329 cases)					
G396–F407	0.47	0.55		0.003	1.41 (1.13-1.76)
G396–L407	0.21	0.18	0.012	0.256	0.85 (0.64-1.12)
S396–F407	0.32	0.27		0.019	0.75 (0.59-0.95)
Paucibacillary (343 controls, 167 cases)					
G396–F407	0.47	0.53		0.068	1.28 (0.98-1.67)
G396–L407	0.21	0.19	0.125	0.479	0.89 (0.64-1.24)
S396–F407	0.32	0.28		0.172	0.82 (0.61-1.09)
Multibacillary (343 controls, 162 cases)					
G396–F407	0.47	0.57		0.001	1.61 (1.21-2.14)
G396–L407	0.21	0.18	0.004	0.174	0.78 (0.55-1.12)
S396–F407	0.32	0.25		0.015	0.68 (0.50-0.93)

[a]proportion of controls and cases

[b]all *P* values adjusted for sex

Supplementary material

Supplementary table 1 is too large for insertion, please see:

http://www.springerlink.com/content/m73vq5275q68kh63/439_2009_Article_775_ESM.html

Supplementary Table 2. Identification of coding SNPs with minor allele frequencies >5% in *MRC1* by direct sequencing in 23 unrelated Vietnamese individuals

Coding SNP	Exon	Position[a]	Allele change	Protein position	Residue change	MAF[b]
rs2253120	2	17905260	ACG → ACA	81	Thr[T] → Thr[T]	0.217
rs2296414	3	17909583	ACC → ATC	167	Thr[T] → Ile[I]	0.348
rs2985837	4	17915768	CAG → CAA	242	Gln[Q] →Gln[Q]	0.260
rs2477664	4	17915822	ATA → ATT	260	Ile[I] → Ile[I]	0.458
rs1926736[c]	7	17931711	GGT → AGT	396	Gly[G] → Ser[S]	0.345
rs2437256[c]	7	17931737	ATT → ATC	404	Ile[I] → Ile[I]	0.207

[a]based on dbSNP129

[b]minor allele frequency calculated from parental genotypes (396 unrelated Vietnamese individuals)

[c]exon 7 sequenced in 396 unrelated Vietnamese individuals

Supplementary Table 3
r^2 values between rs1926736 (G396S) and 47 SNPs used in the association scan of *MRC1*

SNP 1	SNP 2	r^2 value[a]
rs2947600	rs1926736_combined_b	0.200
rs2477645	rs1926736_combined_b	0.301
rs2436680	rs1926736_combined_b	0.375
rs2477639	rs1926736_combined_b	0.204
rs2253120	rs1926736_combined_b	0.068
rs2477631	rs1926736_combined_b	0.399
rs10508549	rs1926736_combined_b	0.018
rs2296413	rs1926736_combined_b	0.135
rs2296414_combined	rs1926736_combined_b	0.166
rs2461174	rs1926736_combined_b	0.107
rs2477630	rs1926736_combined_b	0.425
rs2985835	rs1926736_combined_b	0.072
rs2461139	rs1926736_combined_b	0.107
rs2985837	rs1926736_combined_b	0.097
rs2477664	rs1926736_combined_b	0.333
rs2461162	rs1926736_combined_b	0.034
rs2477632	rs1926736_combined_b	0.021
rs2461161	rs1926736_combined_b	0.483
rs506524	rs1926736_combined_b	0.432
rs692527	rs1926736_combined_b	0.432
rs691053	rs1926736_combined_b	0.428
rs525830	**rs1926736_combined_b**	**0.889**
rs493862	**rs1926736_combined_b**	**0.855**
rs691863	rs1926736_combined_b	0.441
rs692719_b	rs1926736_combined_b	0.135
rs690958	rs1926736_combined_b	0.142
rs1926736_combined_b	**rs1926736_combined_b**	**1.000**
rs2437256_combined	rs1926736_combined_b	0.142

rs691461	rs1926736_combined_b	0.349
rs691783	rs1926736_combined_b	0.098
rs691761	rs1926736_combined_b	0.097
rs2437274	rs1926736_combined_b	0.119
rs692543	rs1926736_combined_b	0.095
rs538625	rs1926736_combined_b	0.055
rs514707	rs1926736_combined_b	0.032
rs766852	rs1926736_combined_b	0.016
rs530749	rs1926736_combined_b	0.138
rs1926732	rs1926736_combined_b	0.017
rs2461168	rs1926736_combined_b	0.051
rs691615_b	rs1926736_combined_b	0.028
rs691292	rs1926736_combined_b	0.028
rs2586698	rs1926736_combined_b	0.056
rs691845	rs1926736_combined_b	0.051
rs530872b	rs1926736_combined_b	0.149
rs691154b	rs1926736_combined_b	0.148
rs499819	rs1926736_combined_b	0.029
rs941b	rs1926736_combined_b	0.094

References

Alcais, A., A. Alter, et al. (2007). "Stepwise replication identifies a low-producing lymphotoxin-[alpha] allele as a major risk factor for early-onset leprosy." Nat. Genet. **39**(4): 517-522.

Barrett, J. C., B. Fry, et al. (2005). "Haploview: analysis and visualization of LD and haplotype maps." Bioinformatics **21**(2): 263-265.

Bell, P., S. Chaturvedi, et al. (2002). "SNPstream UHT: ultra-high throughput SNP genotyping for pharmacogenomics and drug discovery." Biotechniques **Suppl**: 70-2, 74, 76-7.

Boskovic, J., J. N. Arnold, et al. (2006). "Structural Model for the Mannose Receptor Family Uncovered by Electron Microscopy of Endo180 and the Mannose Receptor." J. Biol. Chem. **281**(13): 8780-8787.

Chakravartti, M. and F. Vogel (1973). A twin study on leprosy. Topics in human genetics. P. E. Becker, W. Lenz, F. Vogel and G. G. Wendt. Stuttgart, Georg Thieme. **1:** 1-123.

Cooke, G. and A. Hill (2008). Tuberculosis, Leprosy and Other Mycobacterial Diseases. Genetic Susceptibility to Infectious Diseases. R. A. Kaslow, J. M. McNicholl and A. V. S. Hill. New York, Oxford University Press, Inc.**:** 333-351.

Ezekowitz, R. A., K. Sastry, et al. (1990). "Molecular characterization of the human macrophage mannose receptor: demonstration of multiple carbohydrate recognition-like domains and phagocytosis of yeasts in Cos-1 cells " J. Exp. Med. **172**: 1785-1794.

Fan, J., A. Oliphant, et al. (2003). "Highly parallel SNP genotyping." Cold Spring Harb. Symp. Quant. Biol. **68**: 69-78.

Griffin, T. J. and L. M. Smith (2000). "Single-nucleotide polymorphism analysis by MALDI-TOF mass spectrometry." Trends Biotechnol. **18**(2): 77-84.

Groger, M., W. Holnthoner, et al. (2000). "Dermal Microvascular Endothelial Cells Express the 180-kDa Macrophage Mannose Receptor In Situ and In Vitro." J. Immunol. **165**(10): 5428-5434.

Hattori, T., S. Konno, et al. (2009). "Genetic variants in the mannose receptor gene (MRC1) are associated with asthma in two independent populations." Immunogenetics **61**(11-12): 731-8.

Hill, A. V. (2006). "Aspects of genetic susceptibility to human infectious diseases." Annu. Rev. Genet. **40**: 469-86.

Horvath, S., X. Xu, et al. (2001). "The family based association test method: strategies for studying general genotype--phenotype associations." Eur. J. Hum. Genet. **9**(4): 301-306.

Kabha, K., L. Nissimov, et al. (1995). "Relationships among capsular structure, phagocytosis, and mouse virulence in Klebsiella pneumoniae." Infect. Immun. **63**(3): 847-852.

Kang, P. B., A. K. Azad, et al. (2005). "The human macrophage mannose receptor directs Mycobacterium tuberculosis lipoarabinomannan-mediated phagosome biogenesis." J. Exp. Med. **202**(7): 987-999.

Lake, S. L., D. Blacker, et al. (2000). "Family-Based Tests of Association in the Presence of Linkage." Am. J. Hum. Genet. **67**(6): 1515-1525.

Le Cabec, V., L. J. Emorine, et al. (2005). "The human macrophage mannose receptor is not a professional phagocytic receptor." J. Leukoc. Biol. **77**(6): 934-943.

Miller, E. N., S. E. Jamieson, et al. (2004). "Genome-wide scans for leprosy and tuberculosis susceptibility genes in Brazilians." Genes Immun. **5**(1): 63-67.

Mira, M. T., A. Alcais, et al. (2004). "Susceptibility to leprosy is associated with PARK2 and PACRG." Nature **427**(6975): 636-640.

Mira, M. T., A. Alcais, et al. (2003). "Chromosome 6q25 is linked to susceptibility to leprosy in a Vietnamese population." Nat. Genet. **33**(3): 412-415.

Nigou, J., C. Zelle-Rieser, et al. (2001). "Mannosylated Lipoarabinomannans Inhibit IL-12 Production by Human Dendritic Cells: Evidence for a Negative Signal Delivered Through the Mannose Receptor." J. Immunol. **166**(12): 7477-7485.

Ridley, D. S. and W. H. Jopling (1966). "Classification of leprosy according to immunity. A five-group system." Int. J. Lepr. Other Mycobact. Dis. **34**(3): 255-73.

Schlesinger, L. S. (1993). "Macrophage phagocytosis of virulent but not attenuated strains of Mycobacterium tuberculosis is mediated by mannose receptors in addition to complement receptors." J. Immunol. **150**(7): 2920-2930.

Schlesinger, L. S., S. R. Hull, et al. (1994). "Binding of the terminal mannosyl units of lipoarabinomannan from a virulent strain of Mycobacterium tuberculosis to human macrophages." J. Immunol. **152**(8): 4070-9.

Schlesinger, L. S., T. M. Kaufman, et al. (1996). "Differences in mannose receptor-mediated uptake of lipoarabinomannan from virulent and attenuated strains of Mycobacterium tuberculosis by human macrophages." J. Immunol. **157**(10): 4568-4575.

Shimada, K., H. Takimoto, et al. (2006). "Involvement of mannose receptor in glycopeptidolipid-mediated inhibition of phagosome-lysosome fusion." Microbiol. Immunol. **50**(3): 243-51.

Siddiqui, M. R., S. Meisner, et al. (2001). "A major susceptibility locus for leprosy in India maps to chromosome 10p13." Nat. Genet. **27**(4): 439-441.

Taylor, M. E., K. Bezouska, et al. (1992). "Contribution to ligand binding by multiple carbohydrate-recognition domains in the macrophage mannose receptor." J. Biol. Chem. **267**(3): 1719-1726.

Tosh, K., S. Meisner, et al. (2002). "A Region of Chromosome 20 Is Linked to Leprosy Susceptibility in a South Indian Population." J. Infect. Dis. **186**(8): 1190-1193.

Tregouet, D. A. and V. Garelle (2007). "A new JAVA interface implementation of THESIAS: testing haplotype effects in association studies." Bioinformatics **23**(8): 1038-1039.

Wallace, C., J. Fitness, et al. (2004). "Linkage analysis of susceptibility to leprosy type using an IBD regression method." Genes Immun. **5**(3): 221-225.

Chapter 5 | Discussion

PART 1 – TOWARDS UNDERSTANDING THE COMPLEX HUMAN-*M. leprae* INTERACTION

1.1 Causative versus risk factors

An infectious origin for leprosy was first suggested by G.H.A. Hansen. An intriguing question is to what extent did Hansen's disregard for the role of heredity in the pathogenesis of leprosy impede his efforts to prove causality? Hansen's understanding that *M. leprae* is unequivocally required for disease was correct, but is it sufficient? We now know that the majority of individuals (>90%) exposed to *M. leprae* are asymptomatic {Convit, Sampson et al. 1992; Chaudhury, Hazra et al. 1994; Gupte, Vallishayee et al. 1998}, suggesting that progression from 'exposure' to 'disease' requires additional risk factors. Hansen's subjects might have been successfully inoculated with *M. leprae* but lacked additional risk factors – possibly genetic risk factors – to permit the development of leprosy.

Leprosy, like other infectious diseases, presents a situation where the causal factor (i.e., the infectious agent) is necessary but often insufficient for clinical manifestation of the disease. Disease appearance requires the co-occurrence of risk factors – environmental and/or genetic. For example, herpes simplex virus-1 (HSV-1) encephalitis (HSE) is a rare complication of infection with HSV-1, a ubiquitous virus infecting eighty percent of young adults. HSV-1 is the cause of HSE but given its infrequent occurrence, additional predisposing risk factors must exist. In four unrelated patients, HSE was attributed to either a homozygous mutation in the coding region of *UNC93B1* or a heterozygous mutation in the coding region of *TLR3* {Casrouge, Zhang et al. 2006; Zhang, Jouanguy et al. 2007}. Importantly, *UNC93B1* and *TLR3* mutations do not explain all cases of HSE. Additional genetic risk factors have yet to be identified. Likewise, the

novel genetic determinants identified in the present work are risk factors – not causal factors – for leprosy. Each variant may be uniquely sufficient for the development of leprosy in a sub-set of *M. leprae* exposed individuals, but not necessary for disease in general (see below).

We have identified genetic risk factors for leprosy susceptibility in *LTA* (*LTA*+80), *MRC1* (rs1926736, rs2437257) and the HLA class I region (rs2394885, rs2922997) (**table 1**).

Table 1. Summary of presented association studies for leprosy

Population	Study design	Sample size	Phenotype	Gene	Variant	Allele	OR[a]
Ho Chi Minh City, Vietnam	TDT	298 single-case families	*per se*, early-onset (<16 years)	*LTA*	rs2239704 (*LTA*+80) A → C	A	5.63
North India	case-control	364 cases, 371 controls	*per se*, early-onset (16-25 years)	*LTA*	rs2239704 (*LTA*+80) A → C	A	2.95
Rio de Janeiro, Brazil (Euro-Brazilian)	case-control	209 cases, 192 controls	*per se*, early-onset (16-25 years)	*LTA*	rs2239704 (*LTA*+80) A → C	A	2.76 (NS[b])
Ho Chi Minh City, Vietnam	TDT	490 single-case families	*per se*	*HLA-C*?	rs2394885 C → G	C	2.32
North India	case-control	364 cases, 371 controls	*per se*	*HLA-C*?	rs2394885 C → G	C	2.23
Ho Chi Minh City, Vietnam	TDT	490 single-case families	*per se*	?	rs2922997 C → T	C	1.42
North India	case-control	364 cases, 371 controls	*per se*	?	rs2922997 C → T	C	1.45

(Table 1 continued)

Ho Chi Minh City, Vietnam	TDT	490 single-case families; 90 multi-case families	*per se*	*MRC1*	rs1926736 (G396S)	**S396**	**0.76**
Rio de Janeiro, Brazil	case-control	384 cases, 399 controls	*per se*	*MRC1*	rs1926736 (G396S) rs2437257 (L407F)	G396 L407	1.34 0.75
Rio de Janeiro, Brazil	case-control	329 cases, 343 controls	*per se*	*MRC1*	rs1926736/rs2437257 haplotypes	G396-F407 G396-L407 **S396-F407**	1.41 0.85 **0.75**

[a]OR = odds ratio

[b]NS = not significant

1.2 Lymphotoxin-alpha (LTA) and mycobacterial infection

Lymphotoxin-alpha (LTA) can complex with lymphotoxin-beta (LTB) dimers (LTA_1LTB_2) to form the membrane-bound LTB molecule that signals through the LTB receptor – a critical interaction that precipitates secondary lymphoid organ formation. An important role in the host defence against intracellular infections is emerging for soluble LTA (LTA_3) – secreted by CD4+ T-cells, B-cells and natural killer cells. Normal lymphoid development is perturbed in *Lta* or *Ltb* knockout mice owing to the lack of membrane-bound Lta_1Ltb_2. To overcome the absence of secondary lymphoid organs in *Lta*-/- or *Ltb*-/- mice, *Rag*-/- mice (T- and B-cell deficient) are irradiated and reconstituted with *Lta*-/- (Lta_3 and Lta_1Ltb_2 deficient) or *Ltb*-/- (Lta_1Ltb_2 deficient) bone marrow. The chimeras have normal secondary lymphoid organs but circulating leukocytes deficient in Lta or Ltb.

Lta-/- and *Ltb*-/- chimeras were infected with virulent *Mycobacterium tuberculosis* (*M. tuberculosis*). Despite comparable antigen-specific T-cell and macrophage responses, bacterial replication was less controlled in the *Lta*-/- chimeras and all died within 38 days post-infection (*Ltb*-/- chimeras survived >150 days). Histological examination of the lungs showed that the lesions contained more necrotizing neutrophils and less macrophages. Lymphocytes also failed to extravasate into infected tissue, accumulating in the perivascular and peribronchial regions {Roach, Briscoe et al. 2001}. In a subsequent study using the intracellular bacteria *Listeria monocytogenes*, *Lta*-/- chimeras died 5 to 6 days post-infection (*Ltb*-/- chimeras survived > 14 days). Histological examination of the liver similarly showed that lesions contained more neutrophils and less macrophages and lymphocytes {Roach, Briscoe et al. 2005}. Based on

these observations, soluble Lta mediates the recruitment of macrophages and lymphocytes to sites of infection where they can cooperatively exert their antimicrobial functions.

In human studies, among twelve common polymorphisms (minor allele frequency >10%) at the *LTA* locus, soluble LTA (LTA-homotrimer) production was significantly less from cells carrying the common *LTA*+80A-*LTA*+368C haplotype. A protein-DNA complex, containing activated B-cell factor 1 (ABF-1), was detected when nuclear extracts from mitogen-stimulated cells were incubated specifically with oligonucleotide probes corresponding to the *LTA*+80 A-allele. Basic helix-loop-helix transcription factors like ABF-1 – a transcriptional repressor expressed in lymphoid tissue – bind E2-box consensus motifs (CAGCTG). *LTA*+80 is the second nucleotide in a E2-box motif with a mismatch at position five (CAGC**A**G). *LTA*+80A maintains the 1 bp mismatch in the motif (C**A**GC**A**G) while *LTA*+80C introduces a 2 bp mismatch (C**C**GC**A**G) abrogating ABF-1 binding. As expected, *LTA*+80A-containing constructs co-transfected with ABF-1 in fibroblast cells had significantly lower reporter gene expression than *LTA*+80C-containing constructs {Knight, Keating et al. 2004}. In the Vietnamese, Indian and Brazilian samples, this low-producing *LTA*+80 A-allele was consistently associated with an increased risk of leprosy *per se* in young patients. Extrapolating from the murine models, transcriptional repression of *LTA* may preclude protective immune cells (macrophages and T-cells) from reaching sites of *M. leprae* infection. Age-dependent aspects of adaptive immunity may help to overcome this defect in immune cell recruitment.

Motivated by our genetic finding, the role of LTA in experimental leprosy was recently examined {Hagge, Saunders et al. 2009}. In *Lta*-/- chimeric mice infected with *M. leprae* (6×10^{-3} bacteria per footpad), recovered acid-fast bacilli was significantly higher at 9 months post-

infection (i.e., chronic stage) as compared to controls (chimeric B6 mice) ($P = 0.0006$). Moreover, in *Lta*-/- chimeric mice, the granulomatous response was impaired, with reduced T-cell infiltration. In the draining popliteal LN, the majority of CD4+ T-cells were of the naive phenotype (few effector cells recovered) and transcriptional analysis of the footpads revealed reduced expression of cytokines (*TNF*, *IL12B*, *IL12A*, *IFNG*, *IL2*, *NOS2*), chemokines (*CCL2-5*, *CXCL10*), chemokine receptors (*CXCR3*, *CCR2*) and cellular adhesion molecules (*ICAM*, *VCAM*). These data corroborate our association of the low-producing *LTA*+80 A-allele with increased risk of leprosy. Using *LTA* promoter-reporter constructs, the effect of *LTA*+80 alleles on reporter gene expression (lucierfase) is currently being investigated in a cellular model of BCG and *M. leprae* infection.

1.3 Mannose receptor and mycobacterial infection

Mannose receptor (MR) – expressed on alveolar macrophages, monocyte-derived macrophages (MDMs) and dendritic cells – is a transmembrane protein with an extracellular portion comprised of a cysteine-rich domain, fibronectin type II repeat and eight carbohydrate recognition domains, and a cytoplasmic portion containing a immunoreceptor tyrosine-based activation motif (ITAM). In its ability to bind mannose/fucose to initiate phagocytosis, MR is both a homeostatic receptor – clearing circulating N-linked glycoproteins – and a pattern (pathogen associated molecular pattern) recognition receptor – binding mannose-coated bacteria {Torrelles, Azad et al. 2008}.

MR on MDMs is a receptor for virulent *M. tuberculosis* (Erdman and H37Rv strains), initiating phagocytosis subsequent to binding the mycobacterial cell wall lipoglycan, mannose-capped lipoarabinomannan (ManLAM) {Schlesinger 1993; Schlesinger, Hull et al. 1994}. Anti-MR

serum inhibited CD1b-restricted antigen (i.e., lipoglycan) presentation of mycobacterial LAM (*M. tuberculosis* and *M. leprae*) by antigen presenting cells to LAM-specific T-cells, suggesting MR partakes in the initiation of the adaptive immune response {Prigozy, Sieling et al. 1997}. However, phagocytosis of zymosan by MR – but not the beta-glucan receptor – evaded superoxide generation {Astarie-Dequeker, N'Diaye et al. 1999}. In addition, *M. bovis* BCG ManLAM inhibited LPS-induced IL12 p70 production by human dendritic cells and *M. tuberculosis* ManLAM treated dendritic cells had significantly increased IL10 (anti-inflammatory) and decreased IL12 production {Nigou, Zelle-Rieser et al. 2001; Chieppa, Bianchi et al. 2003}. Moreover, blockade of MR attenuated the inhibition of phagosome-lysosome fusion by *M. tuberculosis* (Erdman strain) in human macrophages {Kang, Azad et al. 2005}, collectively suggesting *Mycobacteria* spp. subvert protective innate immune responses via MR-mediated internalization.

Surprisingly, despite a direct role in mycobacterial pathogenesis, our results attribute only a marginal effect of non-synonymous *MRC1* exon 7 SNPs (G396S, L407F) on susceptibility to leprosy *per se*. Based on our functional data, we also propose that MR requires an accessory molecule to bind *M. leprae* and *M. bovis* BCG that was not expressed in our cellular model (HEK293 cell line). It is possible that MR binding to virulent *M. tuberculosis* similarly requires an accessory molecule that was endogenously present in macrophages used to establish MR as the host cell receptor {Schlesinger 1993}. This unknown accessory molecule could have been overlooked as competitive inhibitors of MR and anti-MR antibody were sufficient to abrogate binding. It is possible that in the context of this hypothetical accessory molecule, the *MRC1* exon 7 leprosy risk haplotype – G396-A399-F407 – modulates the receptor complex favouring the establishment of *M. leprae* infection.

That *MRC1* is a minor gene conferring susceptibility to leprosy *per se* precludes *MRC1* as the major gene underlying the paucibacillary linkage peak on chromosome region 10p13 in the South Indian and Vietnamese populations {Siddiqui, Meisner et al. 2001; Mira, Alcais et al. 2003}. As such, a gene-centric (±3 kb) high-density association scan of the target interval (10p13-p12.1, 9.7 Mb, *NMT2* → *KIAA1217*) has been completed in 198 Vietnamese single-case leprosy families and is currently being analyzed. Tag SNPs (1,155 total) from the Han Chinese in Beijing, China (CHB) population – a very similar population as compared to the Vietnamese population with respect to linkage-disequilibrium structure – were selected using an r^2 cutoff = 0.8 and MAF cutoff = 5% to maximize the common variation captured.

1.4 HLA class I region SNPs

Both rs2394885 (between *HCG27* and *HLA-C*) and rs2922997 (*HLA-B* and *HCP5*) are independent intergenic HLA class I region leprosy risk variants, with no obvious biological function. Interestingly, rs2394885 is a reported tag SNP for *HLA-C**0401 in the CHB population, thereby targeting *HLA-C**0401 as a candidate leprosy susceptibility allele {de Bakker, McVean et al. 2006}. Genotyping of the classical HLA class I alleles (*HLA-A*, *-B*, *-C*) in select Vietnamese individuals is in progress to confirm that rs2394885 is in linkage-disequilibrium with *HLA-C**0401 in our study population. Perhaps, this will similarly reveal LD between rs2922997 and an additional classical HLA class I allele thereby also explaining the association of this variant with leprosy.

HLA-C alleles and promoter variants have previously been associated with several inflammatory and infectious diseases. *HLA-C*04* (OR = 0.39) and *HLA-C*05* (OR = 2.92) were under- and over-represented in 70 patients with rheumatoid arthritis, respectively {Yen, Moore et al. 2001};

HLA-Cw6 was identified as an early-onset risk factor for psoriasis {Nair, Stuart et al. 2006}; and *HLA-C*05* was found to be protective against multiple sclerosis ($P = 3.3\times10^{-5}$) {Yeo, De Jager et al. 2007}. Likewise, the frequency of *HLA-Cw4* was increased in tuberculosis patients (28.9%) as compared to controls (21.1%) {Balamurugan, Sharma et al. 2004}, and a high-producing *HLA-C* promoter variant – rs9264942 – was associated with low viral set point following HIV infection ($P = 3.8 \times 10^{-9}$) {Fellay, Shianna et al. 2007}.

The association of *HLA*-C with leprosy would implicate natural killer (NK) cells – important effectors of the innate response to viruses and intracellular pathogens – in the host response to *M. leprae* infection. NK cell cytolytic activity and cytokine production is regulated by killer cell immunoglobulin-like receptors (KIRs) that bind MHC class I molecules, and HLA-C has emerged as a dominant ligand for KIRs {Kulkarni, Martin et al. 2008}. Immunological assays showed that *M. leprae* was able to induce NK cell mediated cytotoxicity, and there is evidence of altered frequencies of *KIR* genes and *KIR*/*HLA-C* cognate allele pairs in leprosy patients {Kaleab, Ottenoff et al. 1990; Franceschi, Mazini et al. 2008}. Therefore, a logical step in the future genetic dissection of leprosy is to target the *KIR* genes and their epistatic interaction with *HLA-C* alleles – an emerging area of study with much potential to expand our understanding of both leprosy pathogenesis and intracellular infection.

1.5 Common disease common variant hypothesis

In every instance, we have identified *common* susceptibility alleles (frequency range, 0.11-0.44), extrapolating the common disease common variant (CDCV) hypothesis to leprosy: *"...for complex disease traits [...] common alleles at a handful of loci interact to cause disease"* {Smith and Lusis 2002}. However, we cannot discount the contribution of multiple rare alleles (i.e.,

genetic heterogeneity model) as we bias the discovery of common alleles by excluding variants with a MAF <5% due to power limitations imposed by our sample size. Nor can we attest to the biological interaction of risk factors. It is unknown if subsequent to *M. leprae* exposure, a single genetic risk factor is uniquely sufficient for the development of leprosy, or several interacting risk factors are required. It is very interesting that in the recent genome-wide association study of leprosy {Zhang, Huang et al. 2009}, five of seven identified susceptibility genes (*TNFSF15*, *HLA-DRB1*, *RIPK2*, *NOD2* and *LRRK2*), and *PARK2*, encode proteins comprising a single biological pathway. As such, one could speculate that a single genetic risk factor impacting on a critical host response pathway may be uniquely sufficient for the development of leprosy subsequent to *M. leprae* exposure.

PART 2 – LEPROSY AS A GENETIC MODEL FOR SUSCEPTIBILITY TO COMMON INFECTIOUS DISEASES

2.1 Refining the methodological approach to genetic mapping of common infectious diseases

Recently, standard operating protocols for human complex trait analysis have been proposed {Zondervan and Cardon 2007}. Still new studies serve to refine our methodologies as much as our understanding of the human genome. In this context leprosy has proven to be a valuable model for studying the genetic aspect of common infectious diseases. Previous to the present work, the unforeseen discovery of risk factors in the 5'-regulatory region of *PARK2* and *PACRG* confirmed the predicted usefulness of linkage analysis followed by association studies {Mira, Alcais et al. 2004}. Moreover, the process by which *LTA* (*LTA*+80), *MRC1* (rs1926736, rs2437257) and HLA class I region (rs2394885, rs2922997) risk variants were identified highlighted six important considerations for future genetic replication studies irrespective of the method used to initially identify the variant:

1. It is insufficient to genotype only the original associated variant – which is likely to be correlated with the functional risk variant – in the replication sample(s) due to differences in population-specific bin structures. We recommend genotyping all SNPs in the original associated bin (statistically indistinguishable) to facilitate identification of the functional risk variant by process of elimination (**figure 1**). Fortunately, the continual inclusion of SNPs and addition of seven populations (eleven total) to the International HapMap project (Data Rel 27) will enable the reconstruction of population-specific bins *in silico* that are relevant to study samples without additional extensive genotyping efforts.

2. Multivariate analysis is necessary to uncover "hidden" associations (e.g., *LTA*+80 in the North Indian sample) by adjusting for proximal independent risk factors (e.g., *MCCD1-NS* in the North Indian sample).

3. Sample ascertainment (e.g., family-based versus population-based) can introduce biases (e.g., age at diagnosis) that must be accounted for by appropriate adjustments. Our data clearly show that covariates can have a strong effect on replicating previously found associations (e.g., *LTA*+80 in the relatively older North Indian and Brazilian samples).

4. A systematic high-density association scan of a chromosome region with evidence for linkage is superior to a "gene picking" approach. By selectively choosing only strong biological candidates underlying the linkage peak (e.g., *MRC1*), unbiased positional candidates imparting the strongest genetic effects may be overlooked.

5. Haplotype analysis can elucidate subtle complexities in establishing the risk/susceptibility allele for a given risk variant by considering correlated variants (e.g., *MRC1* G396 is a risk *and* borderline protective allele in the context of specific exon 7 haplotypes).

6. We argue that correction for multiple testing (e.g., Bonferroni correction method) is overly conservative in the context of a replication study (not an original study) and may falsely reject true – but weak effect – risk variants (e.g., *MRC1* G396S). A positive replication, particularly one that has utilized a different study design in an ethnically different population(s), should be sufficient to corroborate a previously reported – and presumably corrected – association.

Figure 1 | Identifying *LTA*+80 as the functional variant

SNPs identical in colour belong to the same bin. The original associated SNP (*LTA*-293) is marked with an asterisk. Positive or absent evidence for association is denoted by (+) or (−), respectively. *LTA*-293 was not associated in the Indian sample. Testing all SNPs from the original bin revealed *LTA* + 80 was associated in both samples and is therefore the most likely candidate for the functional variant (supported by experimental evidence). The association of *LTA*-294 in the Indian –but not in the Vietnamese sample – can be explained by a yet unidentified functional variant in the same bin as *LTA*-294 in the Indian population but not in the Vietnamese population. **With kind permission from Springer Science+Business Media: Human Genetics, Leprosy as a genetic model for susceptibility to common infectious diseases, 123, 2008, 227, Alter, Alcaïs, Abel and Schurr, figure 2.**

2.2 Single linkage peak – multiple associated SNPs

At least four independent risk factors for leprosy underlie the linkage peak detected in Vietnamese multi-case families on chromosome region 6p21 – rs2394885 (class I), rs2922997 (class I), *LTA*+80 (class III) and *HLA-DRB1**10 (class II) {Mira, Alcais et al. 2003; Vanderborght, Pacheco et al. 2007}. If multiple independent risk factors span the HLA complex, why was the original evidence for linkage so modest (lod = 2.62)? It is possible that the chromosome region that co-segregated with the disease phenotype harboured multiple risk factors each modestly impacting on leprosy susceptibility (OR range, 1-2). Generally, linkage analysis is insensitive to weak genetic effects (see Chapter 1, section 2.3.3). In the case of 6p21, the occurrence of several minor genes was perhaps sufficient to generate a linkage peak – albeit a weak one. It would be interesting to determine to what extent the evidence of linkage increases

if linkage analysis was done in a sample of early-onset cases only, given *LTA*+80 was determined to be a major risk factor for early-onset leprosy. Nonetheless, our experience in the HLA region justifies a high-density association scan of a target interval with only suggestive evidence of linkage to a phenotype.

2.3 Leprosy: a good disease to study?

The number of replicated susceptibility loci with substantial odds ratios (e.g., *HLA-DRB1*, *LTA*, *PARK2*/*PACRG*) lends the question, is leprosy particularly amenable to genetic studies? Enrolling cases with accurately determined phenotypes is far from trivial but absolutely critical to the success of any genetic study. Unknowingly including "phenocopies" with diverse underlying genetic causes is detrimental to the investigation. The three diagnostic signs of leprosy – the result of more than 100 years of clinical research – have a sensitivity of 97% {Britton and Lockwood 2004}. The reliability of leprosy samples ascertained in endemic countries by experienced leprologists is high, thereby increasing the chances of discovering any existing common risk factor(s). In addition, although leprosy-related morbidity is substantial (e.g., neuropathy), related mortality is negligible. As such, ascertainment of intact nuclear families (i.e., affected(s) and both parents) for genetic studies is feasible. The use of family-based association studies eliminates the occurrence of non-replicable spurious associations due to population-stratification that can arise in poorly matched population-based studies.

Possibly facilitating these efforts is the recent report that differential susceptibility is not confounded by varying strains of *M. leprae*. Among seven geographically diverse samples of the bacterium only three informative SNPs were found in 142 kb of genomic sequence – an unusually low SNP frequency (1 SNP per 28 kb). Moreover, only four SNP combinations (16

sub-types) were observed, each with a limited global distribution (e.g., SNP-type 1 is restricted to Asia, the Pacific region and East Africa) {Monot, Honore et al. 2005; Monot, Honore et al. 2009}.

The considerable progress made to date, could also be a reflection of the intricate human-*M. leprae* relationship. The sequencing of the *M. leprae* genome revealed widespread reductive evolution – only 49.5% of the 3.3 Mb sequence encodes functional proteins. More than 2000 genes – across all functional categories – have been lost since diverging from the last common mycobacterial ancestor {Cole, Eiglmeier et al. 2001}. This extreme case of host adaptation has rendered the bacterium completely dependent on many cellular processes of the host cell, exemplified by its inability to grow *in vitro* and its narrow host range. The viability of this obligate intracellular parasite should then be highly sensitive to host genetic variations affecting these cellular processes – variations that translate into increased disease susceptibility or resistance. For example, the ability of *M. leprae* to neutralize reactive oxygen species (ROS) has been compromised in part by the loss of catalase-peroxidase (*katG* inactivation) {Eiglmeier, Fsihi et al. 1997}. So how does the bacterium survive the hostile macrophage environment? Interestingly, human dopaminergic neuroblastoma cell lines overexpressing PARKIN produced significantly reduced levels of basal ROS and dopamine-induced ROS {Hyun et al. 2002}. If the anti-oxidative role of PARKIN is exploited by *M. leprae* to compensate for its own deficiency, perhaps the *PARK2* susceptibility variant increases host-cell permissibility by enhancing this function.

It is difficult to estimate the individual contribution of genetic and environmental influences on phenotypic variation. As mentioned (see Chapter 1, section 3.1), an adoption study estimated the

relative risk of death due to infectious disease to be 5.81 if a biological parent died of an infectious disease {Sorensen, Nielsen et al. 1988}. Infectious disease susceptibility could therefore be more impervious to environmental factors than other common diseases, relying heavily on the genetic background of the host. Specifically, for persistent diseases such as leprosy and tuberculosis, extensive host-pathogen co-adaptation may have occurred favouring relative strong host genetic effects. This might explain the sizeable odds ratios for leprosy risk factors (*LTA*+80 OR=5.63 95% CI [2.54-12.49]) relative to the most recently identified risk factors for type-2 diabetes (*EXT2* rs11037909 OR=1.27 95% CI [0.97-1.57]), Crohn's disease (*IRGM* rs13361189 OR=1.38 95% CI [1.15-1.66]), multiple sclerosis (*IL7R* rs6897932 OR=0.76 95% CI [0.65-0.90]) and colorectal cancer (*SMAD7* rs4939827 OR=0.86 95% CI [0.79-0.92]) {Broderick, Carvajal-Carmona et al. 2007; Lundmark, Duvefelt et al. 2007; Parkes, Barrett et al. 2007; Sladek, Rocheleau et al. 2007}.

CONCLUDING STATEMENT

Determining the genetic basis of leprosy susceptibility is far from complete. Genome-wide linkage scans have identified major susceptibility loci on chromosome regions 6q25-q26, 6p21, 10p13 and 20p12. To date, only those genetic risk factors underlying 6q25-q26 (*PARK2*/*PACRG*), 6p21 (rs2394885, rs2922997, *LTA*, *HLA-DRB1*) and to some extent 10p13 (*MRC1*), have been discovered. The continued positional cloning of major genes in ethnically diverse family-based and population samples, coupled with functional studies of the implicated proteins is critical to the understanding of leprosy.

It is fitting that the oldest recorded human disease has permitted such remarkable scientific achievements. In 1873 leprosy was the first human disease to be associated with a bacterium and

in 2004 it was the first common infectious disease for which genetic risk factors were positionally cloned. From influencing government policies to modeling complex trait analysis, the value of studying leprosy has proven itself. Despite this progress, disease pathogenesis – particularly disease transmission – continues to elude scientists, precluding the complete eradication of leprosy. Clearly, identifying additional genetic risk factors will implicate new pathways in the complex network of interactions that define this ancient host-pathogen relationship. Likewise, our understanding of non-specific mechanisms underlying the host immune response will improve, as will our ability to identify and ultimately address the inherent weaknesses in the human response to infection.

References (Chapter 1, Introduction and Chapter 5, Discussion)

Abel, L. and F. Demenais (1988). "Detection of major genes for susceptibility to leprosy and its subtypes in a Caribbean island: Desirade island." Am. J. Hum. Genet. **42**(2): 256-66.

Abel, L., F. O. Sanchez, et al. (1998). "Susceptibility to leprosy is linked to the human NRAMP1 gene." J. Infect. Dis. **177**(1): 133-45.

Abel, L., D. L. Vu, et al. (1995). "Complex segregation analysis of leprosy in southern Vietnam." Genet. Epidemiol. **12**(1): 63-82.

Abide, J. M., R. M. Webb, et al. (2008). "Three indigenous cases of leprosy in the Mississippi delta." South. Med. J. **101**(6): 635-8.

Abraham, L. J. and K. M. Kroeger (1999). "Impact of the -308 TNF promoter polymorphism on the transcriptional regulation of the TNF gene: relevance to disease." J. Leukoc. Biol. **66**(4): 562-6.

Achilles, E. G., C. Hagel, et al. (2004). "Leprosy accidentally transmitted from a patient to a surgeon in a nonendemic area." Ann. Intern. Med. **141**(3): W51.

Alcais, A. and L. Abel (2004). Application of genetic epidemiology to dissecting host susceptibility/resistance to infection illustrated with the study of common mycobacterial infections. Susceptibility to Infectious Diseases: The Importance of Host Genetics (Advances in Molecular and Cellular Microbiology). R. Bellamy. Cambridge, Cambridge University Press.

Alcais, A., M. Mira, et al. (2005). "Genetic dissection of immunity in leprosy." Curr. Opin. Immunol. **17**(1): 44-48.

Ali, S., A. Vollaard, et al. (2006). "PARK2/PACRG polymorphisms and susceptibility to typhoid and paratyphoid fever." Clin. Exp. Immunol. **144**(3): 425-431.

Allison, A. C. (1954). "Protection afforded by sickle-cell trait against subtertian malareal infection." Br. Med. J. **1**(4857): 290-4.

Alter, A., A. Alcais, et al. (2008). "Leprosy as a genetic model for susceptibility to common infectious diseases." Hum. Genet. **123**(3): 227-35.

Alvarado-Navarro, A., M. Montoya-Buelna, et al. (2008). "The 3'UTR 1188 A/C polymorphism in the interleukin-12p40 gene (IL-12B) is associated with lepromatous leprosy in the west of Mexico." Immunol. Lett. **118**(2): 148-51.

Astarie-Dequeker, C., E. N. N'Diaye, et al. (1999). "The mannose receptor mediates uptake of pathogenic and nonpathogenic mycobacteria and bypasses bactericidal responses in human macrophages." Infect. Immun. **67**(2): 469-77.

Bakker, M. I., M. Hatta, et al. (2006). "Risk factors for developing leprosy--a population-based cohort study in Indonesia." Lepr. Rev. **77**(1): 48-61.

Balamurugan, A., S. K. Sharma, et al. (2004). "Human leukocyte antigen class I supertypes influence susceptibility and severity of tuberculosis." J. Infect. Dis. **189**(5): 805-11.

Barnes, P. F., D. Chatterjee, et al. (1992). "Tumor necrosis factor production in patients with leprosy." Infect. Immun. **60**(4): 1441-6.

Batista, M. D., A. M. Porro, et al. (2008). "Leprosy reversal reaction as immune reconstitution inflammatory syndrome in patients with AIDS." Clin. Infect. Dis. **46**(6): e56-60.

Bochud, P. Y., T. R. Hawn, et al. (2003). "Cutting edge: a Toll-like receptor 2 polymorphism that is associated with lepromatous leprosy is unable to mediate mycobacterial signaling." J. Immunol. **170**(7): 3451-4.

Bochud, P. Y., T. R. Hawn, et al. (2008). "Toll-like receptor 2 (TLR2) polymorphisms are associated with reversal reaction in leprosy." J. Infect. Dis. **197**(2): 253-61.

Bochud, P. Y., D. Sinsimer, et al. (2009). "Polymorphisms in Toll-like receptor 4 (TLR4) are associated with protection against leprosy." Eur. J. Clin. Microbiol. Infect. Dis. **28**(9): 1055-65.

Borras, S. G., C. Cotorruelo, et al. (2008). "Association of leprosy with HLA-DRB1 in an Argentinean population." Ann. Clin. Biochem. **45**(Pt 1): 96-8.

Brandsma, J. W., L. Yoder, et al. (2005). "Leprosy acquired by inoculation from a knee injury." Lepr. Rev. **76**(2): 175-9.

Britton, W. J. and D. N. Lockwood (2004). "Leprosy." Lancet **363**(9416): 1209-19.

Broderick, P., L. Carvajal-Carmona, et al. (2007). "A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk." Nat. Genet. **39**(11): 1315-1317.

Cardon, L. R. and J. I. Bell (2001). "Association study designs for complex diseases." Nat. Rev. Genet. **2**(2): 91-9.

Casanova, J. L. and L. Abel (2007). "Human genetics of infectious diseases: a unified theory." Embo J. **26**(4): 915-22.

Casrouge, A., S. Y. Zhang, et al. (2006). "Herpes simplex virus encephalitis in human UNC-93B deficiency." Science **314**(5797): 308-12.

Chakravartti, M. and F. Vogel (1973). A twin study on leprosy. Topics in Human Genetics. P. E. Becker, W. Lenz, F. Vogel and G. G. Wendt. Stuttgart, Georg Thieme.

Chaudhury, S., S. K. Hazra, et al. (1994). "An eight-year field trial on antileprosy vaccines among high-risk household contacts in the Calcutta metropolis." Int. J. Lepr. Other Mycobact. Dis. **62**(3): 389-94.

Chieppa, M., G. Bianchi, et al. (2003). "Cross-linking of the mannose receptor on monocyte-derived dendritic cells activates an anti-inflammatory immunosuppressive program." J. Immunol. **171**(9): 4552-60.

Clark, B. M., C. K. Murray, et al. (2008). "Case-control study of armadillo contact and Hansen's disease." Am. J. Trop. Med. Hyg. **78**(6): 962-7.

Cole, S. T., K. Eiglmeier, et al. (2001). "Massive gene decay in the leprosy bacillus." Nature **409**(6823): 1007-11.

Collins English Dictionary (2003). Complete & Unabridged 6th Edition. New York, HarperCollins Publishers.

Convit, J., C. Sampson, et al. (1992). "Immunoprophylactic trial with combined Mycobacterium leprae/BCG vaccine against leprosy: preliminary results." Lancet **339**(8791): 446-50.

Cooke, G. and A. Hill (2008). Tuberculosis, leprosy and other mycobacterial diseases. Genetic Susceptibility to Infectious Diseases. R. A. Kaslow, J. M. McNicholl and A. V. S. Hill. New York, Oxford University Press, Inc.

D'Ambrosio, D., M. Cippitelli, et al. (1998). "Inhibition of IL-12 production by 1,25-dihydroxyvitamin D3. Involvement of NF-kappaB downregulation in transcriptional repression of the p40 gene." J. Clin. Invest. **101**(1): 252-62.

da Silva, S. A., P. S. Mazini, et al. (2009). "HLA-DR and HLA-DQ alleles in patients from the south of Brazil: markers for leprosy susceptibility and resistance." BMC Infect. Dis. **9**: 134.

de Bakker, P. I. W., G. McVean, et al. (2006). "A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC." Nat. Genet. **38**(10): 1166-1172.

De La Vega, F. M., H. Isaac, et al. (2005). "The linkage disequilibrium maps of three human chromosomes across four populations reflect their demographic history and a common underlying recombination pattern." Genome Res. **15**(4): 454-62.

de Messias-Reason, I., P. G. Kremsner, et al. (2009). "Functional haplotypes that produce normal ficolin-2 levels protect against clinical leprosy." J. Infect. Dis. **199**(6): 801-4.

de Messias-Reason, I. J., A. B. Boldt, et al. (2007). "The association between mannan-binding lectin gene polymorphism and clinical leprosy: new insight into an old paradigm." J. Infect. Dis. **196**(9): 1379-85.

Dean, M., M. Carrington, et al. (1996). "Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. Hemophilia Growth and Development Study, Multicenter AIDS Cohort Study, Multicenter Hemophilia Cohort Study, San Francisco City Cohort, ALIVE Study." Science **273**(5283): 1856-62.

Deps, P. D., B. L. Alves, et al. (2008). "Contact with armadillos increases the risk of leprosy in Brazil: a case control study." Indian J. Dermatol. Venereol. Leprol. **74**(4): 338-42.

Dowdle, W. R. (1998). "The principles of disease elimination and eradication." Bull. World Health Organ. **76 Suppl 2**: 22-5.

Eiglmeier, K., H. Fsihi, et al. (1997). "On the catalase-peroxidase gene, katG, of Mycobacterium leprae and the implications for treatment of leprosy with isoniazid." FEMS Microbiol. Lett. **149**(2): 273-278.

Feitosa, M. F., I. Borecki, et al. (1995). "The genetic epidemiology of leprosy in a Brazilian population." Am. J. Hum. Genet. **56**(5): 1179-85.

Fellay, J., K. V. Shianna, et al. (2007). "A whole-genome association study of major determinants for host control of HIV-1." Science **317**(5840): 944-7.

Ferreira, F. R., L. R. Goulart, et al. (2004). "Susceptibility to leprosy may be conditioned by an interaction between the NRAMP1 promoter polymorphisms and the lepromin response." Int. J. Lepr. Other Mycobact. Dis. **72**(4): 457-67.

Filipe-Santos, O., J. Bustamante, et al. (2006). "Inborn errors of IL-12/23- and IFN-gamma-mediated immunity: molecular, cellular, and clinical features." Semin. Immunol. **18**(6): 347-61.

Fine, P. E. (2007). "Leprosy: what is being "eliminated"?" Bull. World Health Organ. **85**(1): 2.

Fitness, J., S. Floyd, et al. (2004). "Large-scale candidate gene study of leprosy susceptibility in the Karonga district of northern Malawi." Am. J. Trop. Med. Hyg. **71**(3): 330-40.

Fitness, J., J. Tosh, et al. (2002). "Genetics of susceptibility to leprosy." Genes Immun. **3**(8): 441-453.

Franceschi, D. S., P. S. Mazini, et al. (2008). "Association between killer-cell immunoglobulin-like receptor genotypes and leprosy in Brazil." Tissue Antigens **72**(5): 478-82.

Franceschi, D. S., P. S. Mazini, et al. (2009). "Influence of TNF and IL10 gene polymorphisms in the immunopathogenesis of leprosy in the south of Brazil." Int. J. Infect. Dis. **13**(4): 493-8.

Freiberger, H. F. and H. H. Fudenberg (1981). "An appetite for armadillo." Hosp. Pract. (Off. Ed.) **16**(6): 137-44.

Fulco, T. O., U. G. Lopes, et al. (2007). "The proteasome function is required for Mycobacterium leprae-induced apoptosis and cytokine secretion." Immunol. Lett. **110**(1): 82-5.

Gallant, C. J., S. Malik, et al. (2007). "Reduced in vitro functional activity of human NRAMP1 (SLC11A1) allele that predisposes to increased risk of pediatric tuberculosis disease." Genes Immun. **8**(8): 691-8.

Garcia, V. E., K. Uyemura, et al. (1999). "IL-18 promotes type 1 cytokine production from NK cells and T cells in human intracellular infection." J. Immunol. **162**(10): 6114-21.

Ghorpade, A. (2002). "Inoculation (tattoo) leprosy: a report of 31 cases." J. Eur. Acad. Dermatol. Venereol. **16**(5): 494-9.

Goulart, L. R., F. R. Ferreira, et al. (2006). "Interaction of TaqI polymorphism at exon 9 of the vitamin D receptor gene with the negative lepromin response may favor the occurrence of leprosy." FEMS Immunol. Med. Microbiol. **48**(1): 91-8.

Griffiths, A. J. F., W. M. Gelbart, et al. (1999). Modern Genetic Analysis. New York, W. H. Freeman & Co.

Gruenheid, S. and P. Gros (2000). "Genetic susceptibility to intracellular infections: Nramp1, macrophage function and divalent cations transport." Curr. Opin. Microbiol. **3**(1): 43-8.

Gupte, M., R. Vallishayee, et al. (1998). "Comparative leprosy vaccine trial in south India." Indian J. Lepr. **70**(4): 369-388.

Hackam, D. J., O. D. Rotstein, et al. (1998). "Host resistance to intracellular infection: mutation of natural resistance-associated macrophage protein 1 (Nramp1) impairs phagosomal acidification." J. Exp. Med. **188**(2): 351-64.

Hagge, D. A., B. M. Saunders, et al. (2009). "Lymphotoxin-alpha and TNF have essential but independent roles in the evolution of the granulomatous response in experimental leprosy." Am. J. Pathol. **174**(4): 1379-89.

Haile, R. W., L. Iselius, et al. (1985). "Segregation and linkage analyses of 72 leprosy pedigrees." Hum. Hered. **35**(1): 43-52.

Hansen, G. A. (1875). "On the etiology of leprosy." Br. Foreign Med.-Chir. Rev. **55**: 459–489.

Haregewoin, A., T. Godal, et al. (1983). "T-cell conditioned media reverse T-cell unresponsiveness in lepromatous leprosy." Nature **303**(5915): 342-4.

Hatta, M., S. M. van Beers, et al. (1995). "Distribution and persistence of Mycobacterium leprae nasal carriage among a population in which leprosy is endemic in Indonesia." Trans. R. Soc. Trop. Med. Hyg. **89**(4): 381-5.

Hill, A. V. (2006). "Aspects of genetic susceptibility to human infectious diseases." Annu. Rev. Genet. **40**: 469-86.

Hinds, D. A., L. L. Stuve, et al. (2005). "Whole-genome patterns of common DNA variation in three human populations." Science **307**(5712): 1072-9.

Hirschhorn, J. N. and M. J. Daly (2005). "Genome-wide association studies for common diseases and complex traits." Nat. Rev. Genet. **6**(2): 95-108.

Horton, R., L. Wilming, et al. (2004). "Gene map of the extended human MHC." Nat. Rev. Genet. **5**(12): 889-99.

International Leprosy Association (2002). "Report of the International Leprosy Association Technical Forum: epidemiology and control."

Jabado, N., A. Jankowski, et al. (2000). "Natural resistance to intracellular infections: natural resistance-associated macrophage protein 1 (Nramp1) functions as a pH-dependent manganese transporter at the phagosomal membrane." J. Exp. Med. **192**(9): 1237-48.

Jamieson, S. E., E. N. Miller, et al. (2004). "Evidence for a cluster of genes on chromosome 17q11-q21 controlling susceptibility to tuberculosis and leprosy in Brazilians." Genes Immun. **5**(1): 46-57.

Janeway, C. A., P. Travers, et al. (2001). Immunobiology. New York and London, Garland Science.

Jessen, K. R. and R. Mirsky (2005). "The origin and development of glial cells in peripheral nerves." Nat. Rev. Neurosci. **6**(9): 671-82.

Johnson, C. M., E. A. Lyle, et al. (2007). "Cutting edge: A common polymorphism impairs cell surface trafficking and functional responses of TLR1 but protects against leprosy." J. Immunol. **178**(12): 7520-4.

Jones, L., I. Moszer, et al. (2001). "Leproma: a Mycobacterium leprae genome browser." Lepr. Rev. **72**(4): 470-7.

Jullien, D., P. A. Sieling, et al. (1997). "IL-15, an immunomodulator of T cell responses in intracellular infection." J. Immunol. **158**(2): 800-6.

Kai, M., M. Matsuoka, et al. (1999). "Diaminodiphenylsulfone resistance of Mycobacterium leprae due to mutations in the dihydropteroate synthase gene." FEMS Microbiol. Lett. **177**(2): 231-5.

Kaleab, B., T. Ottenoff, et al. (1990). "Mycobacterial-induced cytotoxic T cells as well as nonspecific killer cells derived from healthy individuals and leprosy patients." Eur. J. Immunol. **20**(12): 2651-9.

Kang, P. B., A. K. Azad, et al. (2005). "The human macrophage mannose receptor directs Mycobacterium tuberculosis lipoarabinomannan-mediated phagosome biogenesis." J. Exp. Med. **202**(7): 987-99.

Kang, T. J. and G. T. Chae (2001). "Detection of Toll-like receptor 2 (TLR2) mutation in the lepromatous leprosy patients." FEMS Immunol. Med. Microbiol. **31**(1): 53-8.

Kang, T. J., S. B. Lee, et al. (2002). "A polymorphism in the toll-like receptor 2 is associated with IL-12 production from monocyte in lepromatous leprosy." Cytokine **20**(2): 56-62.

Kang, T. J., C. E. Yeum, et al. (2004). "Differential production of interleukin-10 and interleukin-12 in mononuclear cells from leprosy patients with a Toll-like receptor 2 mutation." Immunology **112**(4): 674-80.

Kim, J., K. Uyemura, et al. (2001). "A role for IL-12 receptor expression and signal transduction in host defense in leprosy." J. Immunol. **167**(2): 779-86.

Kirchheimer, W. F. and E. E. Storrs (1971). "Attempts to establish the armadillo (Dasypus novemcinctus Linn.) as a model for the study of leprosy. I. Report of lepromatoid leprosy in an experimentally infected armadillo." Int. J. Lepr. Other Mycobact. Dis. **39**(3): 693-702.

Klatser, P. R., S. van Beers, et al. (1993). "Detection of Mycobacterium leprae nasal carriers in populations for which leprosy is endemic." J. Clin. Microbiol. **31**(11): 2947-51.

Klein, J. and A. Sato (2000). "The HLA system. First of two parts." N. Engl. J. Med. **343**(10): 702-9.

Klein, J. and A. Sato (2000). "The HLA system. Second of two parts." N. Engl. J. Med. **343**(11): 782-6.

Knight, J. C., B. J. Keating, et al. (2004). "Allele-specific repression of lymphotoxin-[alpha] by activated B cell factor-1." Nat. Genet. **36**(4): 394-399.

Knight, J. C. and D. Kwiatkowski (1999). "Inherited variability of tumor necrosis factor production and susceptibility to infectious disease." Proc. Assoc. Am. Physicians **111**(4): 290-8.

Kocak, M., M. Balci, et al. (2002). "Associations between human leukocyte antigens and leprosy in the Turkish population." Clin. Exp. Dermatol. **27**(3): 235-9.

Krutzik, S. R., M. T. Ochoa, et al. (2003). "Activation and regulation of Toll-like receptors 2 and 1 in human leprosy." Nat. Med. **9**(5): 525-32.

Krutzik, S. R., B. Tan, et al. (2005). "TLR activation triggers the rapid differentiation of monocytes into macrophages and dendritic cells." Nat. Med. **11**(6): 653-60.

Kulkarni, S., M. P. Martin, et al. (2008). "The Yin and Yang of HLA and KIR in human disease." Semin. Immunol. **20**(6): 343-52.

Lahiri, R. and J. L. Krahenbuhl (2008). "The role of free-living pathogenic amoeba in the transmission of leprosy: a proof of principle." Lepr. Rev. **79**(4): 401-9.

Lander, E. and L. Kruglyak (1995). "Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results." Nat. Genet. **11**(3): 241-7.

Lander, E. S. and N. J. Schork (1994). "Genetic dissection of complex traits." Science **265**(5181): 2037-48.

Lane, J. E., D. S. Walsh, et al. (2006). "Borderline tuberculoid leprosy in a woman from the state of Georgia with armadillo exposure." J. Am. Acad. Dermatol. **55**(4): 714-6.

Lavania, M., K. Katoch, et al. (2008). "Detection of viable Mycobacterium leprae in soil samples: insights into possible sources of transmission of leprosy." Infect. Genet. Evol. **8**(5): 627-31.

Lincoln, M. R., A. Montpetit, et al. (2005). "A predominant role for the HLA class II region in the association of the MHC region with multiple sclerosis." Nat. Genet. **37**(10): 1108-12.

Little, D., S. Khanolkar-Young, et al. (2001). "Immunohistochemical analysis of cellular infiltrate and gamma interferon, interleukin-12, and inducible nitric oxide synthase expression in leprosy type 1 (reversal) reactions before and during prednisolone treatment." Infect. Immun. **69**(5): 3413-7.

Lockwood, D. N. and S. Suneetha (2005). "Leprosy: too complex a disease for a simple elimination paradigm." Bull. World Health Organ. **83**(3): 230-5.

Lumpkin, L. R., 3rd, G. F. Cox, et al. (1984). "Leprosy in armadillo handlers." J. Am. Acad. Dermatol. **10**(6): 1073.

Lundmark, F., K. Duvefelt, et al. (2007). "Variation in interleukin 7 receptor [alpha] chain (IL7R) influences risk of multiple sclerosis." Nat. Genet. **39**(9): 1108-1113.

Malhotra, D., K. Darvishi, et al. (2005). "Association study of major risk single nucleotide polymorphisms in the common regulatory region of PARK2 and PACRG genes with leprosy in an Indian population." Eur. J. Hum. Genet. **14**(4): 438-442.

Malhotra, D., K. Darvishi, et al. (2005). "IL-10 promoter single nucleotide polymorphisms are significantly associated with resistance to leprosy." Hum. Genet. **118**(2): 295-300.

Malhotra, D., V. Relhan, et al. (2005). "TLR2 Arg677Trp polymorphism in leprosy: revisited." Hum. Genet. **116**(5): 413-5.

Malo, D., K. Vogan, et al. (1994). "Haplotype mapping and sequence analysis of the mouse Nramp gene predict susceptibility to infection with intracellular parasites." Genomics **23**(1): 51-61.

Marmor, M. F. (2002). "The Ophthalmic Trials of G. H. A. Hansen." Surv. Ophthalmol. **47**(3): 275-287.

Matsuoka, M., S. Izumi, et al. (1999). "Mycobacterium leprae DNA in daily using water as a possible source of leprosy infection." Indian J. Lepr. **71**(1): 61-7.

McInturff, J. E., R. L. Modlin, et al. (2005). "The role of toll-like receptors in the pathogenesis and treatment of dermatological disease." J. Invest. Dermatol. **125**(1): 1-8.

Mehra, V., L. H. Mason, et al. (1979). "Lepromin-induced suppressor cells in patients with leprosy." J. Immunol. **123**(4): 1813-7.

Meima, A., W. C. Smith, et al. (2004). "The future incidence of leprosy: a scenario analysis." Bull. World Health Organ. **82**(5): 373-80.

Meisner, S. J., S. Mucklow, et al. (2001). "Association of NRAMP1 polymorphism with leprosy type but not susceptibility to leprosy per se in west Africans." Am. J. Trop. Med. Hyg. **65**(6): 733-5.

Meyer, C. G., J. May, et al. (1998). "Human leukocyte antigens in tuberculosis and leprosy." Trends Microbiol. **6**(4): 148-54.

Mikita, N., N. Kanazawa, et al. (2009). "No involvement of non-synonymous TLR2 polymorphisms in Japanese leprosy patients." J. Dermatol. Sci. **54**(1): 48-9.

Miller, E. N., S. E. Jamieson, et al. (2004). "Genome-wide scans for leprosy and tuberculosis susceptibility genes in Brazilians." Genes Immun. **5**(1): 63-67.

Mira, M. T. (2006). "Genetic host resistance and susceptibility to leprosy." Microbes Infect. **8**(4): 1124-1131.

Mira, M. T., A. Alcais, et al. (2003). "Segregation of HLA/TNF region is linked to leprosy clinical spectrum in families displaying mixed leprosy subtypes." Genes Immun. **4**(1): 67-73.

Mira, M. T., A. Alcais, et al. (2004). "Susceptibility to leprosy is associated with PARK2 and PACRG." Nature **427**(6975): 636-640.

Mira, M. T., A. Alcais, et al. (2003). "Chromosome 6q25 is linked to susceptibility to leprosy in a Vietnamese population." Nat. Genet. **33**(3): 412-415.

Misch, E. A. and T. R. Hawn (2008). "Toll-like receptor polymorphisms and susceptibility to human disease." Clin. Sci. (Lond.) **114**(5): 347-60.

Misch, E. A., M. Macdonald, et al. (2008). "Human TLR1 deficiency is associated with impaired mycobacterial signaling and protection from leprosy reversal reaction." PLoS Negl. Trop. Dis. **2**(5): e231.

Modlin, R. L., G. M. Gersuk, et al. (1986). "T-lymphocyte clones from leprosy skin lesions." Lepr. Rev. **57 Suppl 2**: 143-7.

Moegling, A. (1935). "Die Epidemiologie der Lûbecker Säuglingstuberkulose." Arbeiten a.d.Reichsges.-Amt **69**: 1-24.

Mohamed Ali, P. and K. Ramanujam (1966). "Leprosy in twins." Int. J. Lepr. **34**(4): 405-407.

Monot, M., N. Honore, et al. (2005). "On the origin of leprosy." Science **308**(5724): 1040-2.

Monot, M., N. Honore, et al. (2009). "Comparative genomic and phylogeographic analysis of Mycobacterium leprae." Nat. Genet. **41**(12): 1282-9.

Moraes, M. O., A. G. Pacheco, et al. (2004). "Interleukin-10 promoter single-nucleotide polymorphisms as markers for disease susceptibility and disease severity in leprosy." Genes Immun. **5**(7): 592-5.

Morahan, G., G. Kaur, et al. (2007). "Association of variants in the IL12B gene with leprosy and tuberculosis." Tissue Antigens **69 Suppl 1**: 234-6.

Morton, N. E. (1955). "Sequential tests for the detection of linkage." Am. J. Hum. Genet. **7**(3): 277-318.

Mosser, D. M. and X. Zhang (2008). "Interleukin-10: new perspectives on an old cytokine." Immunol. Rev. **226**: 205-18.

Mueller, J. C. (2004). "Linkage disequilibrium for different scales and applications." Brief. Bioinform. **5**(4): 355-64.

Mushatt, D. M., P. Wattanamano, et al. (1998). "Lepromatous leprosy in a renal transplant recipient." Clin. Infect. Dis. **26**(1): 217-8.

Nair, R. P., P. E. Stuart, et al. (2006). "Sequence and haplotype analysis supports HLA-C as the psoriasis susceptibility 1 gene." Am. J. Hum. Genet. **78**(5): 827-51.

Ng, V., G. Zanazzi, et al. (2000). "Role of the cell wall phenolic glycolipid-1 in the peripheral nerve predilection of Mycobacterium leprae." Cell **103**(3): 511-24.

Nigou, J., C. Zelle-Rieser, et al. (2001). "Mannosylated Lipoarabinomannans Inhibit IL-12 Production by Human Dendritic Cells: Evidence for a Negative Signal Delivered Through the Mannose Receptor." J. Immunol. **166**(12): 7477-7485.

Noon, L. A. and A. C. Lloyd (2007). "Treating leprosy: an Erb-al remedy?" Trends Pharmacol. Sci. **28**(3): 103-5.

Ochoa, M. T., S. Stenger, et al. (2001). "T-cell release of granulysin contributes to host defense in leprosy." Nat. Med. **7**(2): 174-9.

Ohyama, H., K. Ogata, et al. (2005). "Polymorphism of the 5' flanking region of the IL-12 receptor beta2 gene partially determines the clinical types of leprosy through impaired transcriptional activity." J. Clin. Pathol. **58**(7): 740-3.

Oliveira, R. B., M. T. Ochoa, et al. (2003). "Expression of Toll-like receptor 2 on human Schwann cells: a mechanism of nerve damage in leprosy." Infect. Immun. **71**(3): 1427-33.

Ottenhoff, T. H., D. G. Elferink, et al. (1986). "Cloned suppressor T cells from a lepromatous leprosy patient suppress Mycobacterium leprae reactive helper T cells." Nature **322**(6078): 462-4.

Parkes, M., J. C. Barrett, et al. (2007). "Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility." Nat. Genet. **39**(7): 830-832.

Pereira, A. C., V. N. Brito-de-Souza, et al. (2009). "Genetic, epidemiological and biological analysis of interleukin-10 promoter single-nucleotide polymorphisms suggests a definitive role for -819C/T in leprosy susceptibility." Genes Immun. **10**(2): 174-80.

Piemonti, L., P. Monti, et al. (2000). "Vitamin D3 affects differentiation, maturation, and function of human monocyte-derived dendritic cells." J. Immunol. **164**(9): 4443-51.

Pinhasi, R., R. Foley, et al. (2006). "Reconsidering the Antiquity of Leprosy." Science **312**(5775): 846a-.

Prado-Montes de Oca, E., J. S. Velarde-Felix, et al. (2009). "SNP 668C (-44) alters a NF-kappaB1 putative binding site in non-coding strand of human beta-defensin 1 (DEFB1) and is associated with lepromatous leprosy." Infect. Genet. Evol. **9**(4): 617-25.

Prigozy, T. I., P. A. Sieling, et al. (1997). "The mannose receptor delivers lipoglycan antigens to endosomes for presentation to T cells by CD1b molecules." Immunity **6**(2): 187-97.

Rambukkana, A., J. L. Salzer, et al. (1997). "Neural targeting of Mycobacterium leprae mediated by the G domain of the laminin-alpha2 chain." Cell **88**(6): 811-21.

Rambukkana, A., H. Yamada, et al. (1998). "Role of alpha-dystroglycan as a Schwann cell receptor for Mycobacterium leprae." Science **282**(5396): 2076-9.

Rambukkana, A., G. Zanazzi, et al. (2002). "Contact-dependent demyelination by Mycobacterium leprae in the absence of immune cells." Science **296**(5569): 927-31.

Rani, R., M. A. Fernandez-Vina, et al. (1993). "Study of HLA class II alleles by PCR oligotyping in leprosy patients from north India." Tissue Antigens **42**(3): 133-7.

Ranque, B., V. T. Nguyen, et al. (2007). "Age is an important risk factor for onset and sequelae of reversal reactions in Vietnamese patients with leprosy." Clin. Infect. Dis. **44**(1): 33-40.

Rees, R. J. (1976). "Airborne infection with Mycobacterium leprae in mice." Int. J. Lepr. Other Mycobact. Dis. **44**(1-2): 99-103.

Reynard, M. P., D. Turner, et al. (2003). "Allele frequencies for an interferon-gamma microsatellite in a population of Brazilian leprosy patients." Eur. J. Immunogenet. **30**(2): 149-51.

Ridley, D. S. and W. H. Jopling (1966). "Classification of leprosy according to immunity. A five-group system." Int. J. Lepr. Other Mycobact. Dis. **34**(3): 255-73.

Risch, N. and K. Merikangas (1996). "The future of genetic studies of complex human diseases." Science **273**(5281): 1516-7.

Risch, N. J. (2000). "Searching for genetic determinants in the new millennium." Nature **405**(6788): 847-56.

Roach, D., H. Briscoe, et al. (2005). "Independent protective effects for tumor necrosis factor and lymphotoxin alpha in the host response to Listeria monocytogenes infection." Infect. Immun. **73**(8): 4787–4792.

Roach, D. R., H. Briscoe, et al. (2001). "Secreted Lymphotoxin-{{alpha}} Is Essential for the Control of an Intracellular Bacterial Infection." J. Exp. Med. **193**(2): 239-246.

Robbins, G., V. M. Tripathy, et al. (2009). "Ancient skeletal evidence for leprosy in India (2000 B.C.)." PLoS One **4**(5): e5669.

Rodrigues, L. S., E. da Silva Maeda, et al. (2010). "Mycobacterium leprae induces insulin-like growth factor and promotes survival of Schwann cells upon serum withdrawal." Cell. Microbiol. **12**(1): 42-54.

Roger, M., G. Levee, et al. (1997). "No evidence for linkage between leprosy susceptibility and the human natural resistance-associated macrophage protein 1 (NRAMP1) gene in French Polynesia." Int. J. Lepr. Other Mycobact. Dis. **65**(2): 197-202.

Roy, S., A. Frodsham, et al. (1999). "Association of vitamin D receptor genotype with leprosy type." J. Infect. Dis. **179**(1): 187-91.

Roy, S., W. McGuire, et al. (1997). "Tumor necrosis factor promoter polymorphism and susceptibility to lepromatous leprosy." J. Infect. Dis. **176**(2): 530-2.

Santos, A. R., P. N. Suffys, et al. (2002). "Role of tumor necrosis factor-alpha and interleukin-10 promoter gene polymorphisms in leprosy." J. Infect. Dis. **186**(11): 1687-91.

Schlesinger, L. S. (1993). "Macrophage phagocytosis of virulent but not attenuated strains of Mycobacterium tuberculosis is mediated by mannose receptors in addition to complement receptors." J. Immunol. **150**(7): 2920-2930.

Schlesinger, L. S., S. R. Hull, et al. (1994). "Binding of the terminal mannosyl units of lipoarabinomannan from a virulent strain of Mycobacterium tuberculosis to human macrophages." J. Immunol. **152**(8): 4070-9.

Schuring, R. P., L. Hamann, et al. (2009). "Polymorphism N248S in the human Toll-like receptor 1 gene is related to leprosy and leprosy reactions." J. Infect. Dis. **199**(12): 1816-9.

Schurr, E., A. Alcais, et al. (2006). "Genetic predisposition to leprosy: A major gene reveals novel pathways of immunity to Mycobacterium leprae." Semin. Immunol. **18**(6): 404-410.

Scollard, D. M., L. B. Adams, et al. (2006). "The continuing challenges of leprosy." Clin. Microbiol. Rev. **19**(2): 338-81.

Scollard, D. M., M. P. Joyce, et al. (2006). "Development of leprosy and type 1 leprosy reactions after treatment with infliximab: a report of 2 cases." Clin. Infect. Dis. **43**(2): e19-22.

Scollard, D. M., G. McCormick, et al. (1999). "Localization of Mycobacterium leprae to endothelial cells of epineurial and perineurial blood vessels and lymphatics." Am. J. Pathol. **154**(5): 1611-20.

Shankarkumar, U. (2004). "HLA associations in leprosy patients from Mumbai, India." Lepr. Rev. **75**(1): 79-85.

Shaw, M. A., I. J. Donaldson, et al. (2001). "Association and linkage of leprosy phenotypes with HLA class II and tumour necrosis factor genes." Genes Immun. **2**(4): 196-204.

Shepard, C. C. (1960). "The experimental disease that follows the injection of human leprosy bacilli into footpads of mice." J. Exp. Med. **112**(3): 445-454.

Shields, E. D., D. A. Russell, et al. (1987). "Genetic epidemiology of the susceptibility to leprosy." J. Clin. Invest. **79**(4): 1139-43.

Shimoji, Y., V. Ng, et al. (1999). "A 21-kDa surface protein of Mycobacterium leprae binds peripheral nerve laminin-2 and mediates Schwann cell invasion." Proc. Natl. Acad. Sci. U S A **96**(17): 9857-62.

Siddiqui, M. R., S. Meisner, et al. (2001). "A major susceptibility locus for leprosy in India maps to chromosome 10p13." Nat. Genet. **27**(4): 439-441.

Sieling, P. A., D. Jullien, et al. (1999). "CD1 expression by dendritic cells in human leprosy lesions: correlation with effective host immunity." J. Immunol. **162**(3): 1851-8.

Sieling, P. A., X. H. Wang, et al. (1994). "IL-12 regulates T helper type 1 cytokine responses in human infectious disease." J. Immunol. **153**(8): 3639-47.

Sladek, R., G. Rocheleau, et al. (2007). "A genome-wide association study identifies novel risk loci for type 2 diabetes." Nature **445**(7130): 881-885.

Smith, D. G. (1979). "The genetic hypothesis for susceptibility to lepromatous leprosy." Hum. Genet. **50**(2): 163-77.

Smith, D. J. and A. J. Lusis (2002). "The allelic structure of common disease." Hum. Mol. Genet. **11**(20): 2455-61.

Sorensen, T. I., G. G. Nielsen, et al. (1988). "Genetic and environmental influences on premature death in adult adoptees." N. Engl. J. Med. **318**(12): 727-32.

Spielman, R. S., R. E. McGinnis, et al. (1993). "Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM)." Am. J. Hum. Genet. **52**(3): 506-16.

Spierings, E., T. de Boer, et al. (2001). "Mycobacterium leprae-specific, HLA class II-restricted killing of human Schwann cells by CD4+ Th1 cells: a novel immunopathogenic mechanism of nerve damage in leprosy." J. Immunol. **166**(10): 5883-8.

Strachan, T. and A. P. Read (1999). Human Molecular Genetics 2. New York and London, Garland Science.

Tapinos, N., M. Ohnishi, et al. (2006). "ErbB2 receptor tyrosine kinase signaling mediates early demyelination induced by leprosy bacilli." Nat. Med. **12**(8): 961-6.

Tapinos, N. and A. Rambukkana (2005). "Insights into regulation of human Schwann cell proliferation by Erk1/2 via a MEK-independent and p56Lck-dependent pathway from leprosy bacilli." Proc. Natl. Acad. Sci. U S A **102**(26): 9188-93.

Torrelles, J. B., A. K. Azad, et al. (2008). "Role of C-type lectins in mycobacterial infections." Curr. Drug. Targets **9**(2): 102-12.

Tosh, K., S. Meisner, et al. (2002). "A region of chromosome 20 is linked to leprosy susceptibility in a South Indian population." J. Infect. Dis. **186**(8): 1190-1193.

Tosh, K., M. Ravikumar, et al. (2006). "Variation in MICA and MICB genes and enhanced susceptibility to paucibacillary leprosy in South India." Hum. Mol. Genet. **15**(19): 2880-7.

Triccas, J. A., P. W. Roche, et al. (1996). "A 35-kilodalton protein is a major target of the human immune response to Mycobacterium leprae." Infect. Immun. **64**(12): 5171-7.

Vanderborght, P. R., H. J. Matos, et al. (2004). "Single nucleotide polymorphisms (SNPs) at -238 and -308 positions in the TNFalpha promoter: clinical and bacteriological evaluation in leprosy." Int. J. Lepr. Other Mycobact. Dis. **72**(2): 143-8.

Vanderborght, P. R., A. G. Pacheco, et al. (2007). "HLA-DRB1[ast]04 and DRB1[ast]10 are associated with resistance and susceptibility, respectively, in Brazilian and Vietnamese leprosy patients." Genes Immun. **8**(4): 320-324.

Vejbaesya, S., P. Mahaisavariya, et al. (2007). "TNF alpha and NRAMP1 polymorphisms in leprosy." J. Med. Assoc. Thai. **90**(6): 1188-92.

Velarde Felix, J. S., S. G. Cazarez Salazar, et al. (2009). "[Association between the TaqI polymorphism of the vitamin D receptor gene and lepromatous leprosy in a Mexican population sample]." Salud Publica Mex. **51**(1): 59-61.

Vidal, S. M., E. Pinner, et al. (1996). "Natural resistance to intracellular infections: Nramp1 encodes a membrane phosphoglycoprotein absent in macrophages from susceptible (Nramp1 D169) mouse strains." J. Immunol. **157**(8): 3559-68.

Wagener, D. K., V. Schauf, et al. (1988). "Segregation analysis of leprosy in families of northern Thailand." Genet. Epidemiol. **5**(2): 95-105.

Walker, S. L. and D. N. Lockwood (2007). "Leprosy." Clin. Dermatol. **25**(2): 165-72.

Wallace, C., J. Fitness, et al. (2004). "Linkage analysis of susceptibility to leprosy type using an IBD regression method." Genes Immun. **5**(3): 221-5.

Wang, L. M., A. Kimura, et al. (1999). "HLA linked with leprosy in southern China: HLA-linked resistance alleles to leprosy." Int. J. Lepr. Other Mycobact. Dis. **67**(4): 403-8.

Wang, T. T., F. P. Nestel, et al. (2004). "Cutting edge: 1,25-dihydroxyvitamin D3 is a direct inducer of antimicrobial peptide gene expression." J. Immunol. **173**(5): 2909-12.

Wang, W. Y., B. J. Barratt, et al. (2005). "Genome-wide association studies: theoretical and practical concerns." Nat. Rev. Genet. **6**(2): 109-18.

Wemambu, S. N., J. L. Turk, et al. (1969). "Erythema nodosum leprosum: a clinical manifestation of the arthus phenomenon." Lancet **2**(7627): 933-5.

West, B. C., J. R. Todd, et al. (1988). "Leprosy in six isolated residents of northern Louisiana. Time-clustered cases in an essentially nonendemic area." Arch. Intern. Med. **148**(9): 1987-92.

Wheeler, E., E. N. Miller, et al. (2006). "Genome-wide scan for loci influencing quantitative immune response traits in the Belem family study: comparison of methods and summary of results." Ann. Hum. Genet. **70**(Pt 1): 78-97.

White, J. H. (2008). "Vitamin D signaling, infectious diseases, and regulation of innate immunity." Infect. Immun. **76**(9): 3837-43.

WHO (2002). "Leprosy. Global situation." Wkly Epidemiol Rec **77**(1): 1-8.

WHO (2008). "Global leprosy situation, beginning of 2008." Wkly Epidemiol. Rec. **83**(33): 293-300.

WHO Expert Committee on Leprosy (1997). "Seventh report."

WHO http://www.who.int/lep/mdt/effectiveness/en/index.html. (2010).

Wibawa, T., H. Soebono, et al. (2002). "Association of a missense mutation of the laminin alpha2 gene with tuberculoid type of leprosy in Indonesian patients." Trop. Med. Int. Health **7**(7): 631-6.

Yamamura, M., X. H. Wang, et al. (1992). "Cytokine patterns of immunologically mediated tissue damage." J. Immunol. **149**(4): 1470-5.

Yee, L. J. and M. R. Thursz (2004). Genetic diversity in the major histocompatibility complex and the immune response to infectious diseases. Susceptibility to Infectious Diseases: The Importance of Host Genetics (Advances in Molecular and Cellular Microbiology). R. Bellamy. Cambridge, Cambridge University Press.

Yen, J. H., B. E. Moore, et al. (2001). "Major histocompatibility complex class I-recognizing receptors are disease risk genes in rheumatoid arthritis." J. Exp. Med. **193**(10): 1159-67.

Yeo, T. W., P. L. De Jager, et al. (2007). "A second major histocompatibility complex susceptibility locus for multiple sclerosis." Ann. Neurol. **61**(3): 228-36.

Zerva, L., B. Cizman, et al. (1996). "Arginine at positions 13 or 70-71 in pocket 4 of HLA-DRB1 alleles is associated with susceptibility to tuberculoid leprosy." J. Exp. Med. **183**(3): 829-836.

Zhang, F., H. Liu, et al. (2009). "Evidence for an association of HLA-DRB1*15 and DRB1*09 with leprosy and the impact of DRB1*09 on disease onset in a Chinese Han population." BMC Med. Genet. **10**: 133.

Zhang, F. R., W. Huang, et al. (2009). "Genomewide association study of leprosy." N. Engl. J. Med. **361**(27): 2609-18.

Zhang, S.-Y., E. Jouanguy, et al. (2007). "TLR3 Deficiency in Patients with Herpes Simplex Encephalitis." Science **317**(5844): 1522-1527.

Zondervan, K. T. and L. R. Cardon (2007). "Designing candidate gene and genome-wide case-control association studies." Nat. Protocols **2**(10): 2492-2501.

Appendix i │ Manuscripts

nature genetics

Stepwise replication identifies a low-producing lymphotoxin-α allele as a major risk factor for early-onset leprosy

Alexandre Alcaïs[1,2,9], Andrea Alter[3,9], Guillemette Antoni[1,9], Marianna Orlova[3], Nguyen Van Thuc[4], Meenakshi Singh[5], Patrícia R Vanderborght[6], Kiran Katoch[7], Marcelo T Mira[8], Vu Hong Thai[4], Ngyuen Thu Huong[4], Nguyen Ngoc Ba[4], Milton Moraes[6], Narinder Mehra[5], Erwin Schurr[3] & Laurent Abel[1,2]

Host genetics has an important role in leprosy, and variants in the shared promoter region of *PARK2* and *PACRG* were the first major susceptibility factors identified by positional cloning[1,2]. Here we report the linkage disequilibrium mapping of the second linkage peak of our previous genome-wide scan[1], located close to the HLA complex. In both a Vietnamese familial sample and an Indian case-control sample, the low-producing lymphotoxin-α (*LTA*)+80 A allele was significantly associated with an increase in leprosy risk ($P = 0.007$ and $P = 0.01$, respectively). Analysis of an additional case-control sample from Brazil and an additional familial sample from Vietnam showed that the *LTA*+80 effect was much stronger in young individuals. In the combined sample of 298 Vietnamese familial trios, the odds ratio of leprosy for *LTA*+80 AA/AC versus CC subjects was 2.11 ($P = 0.000024$), which increased to 5.63 ($P = 0.0000004$) in the subsample of 121 trios of affected individuals diagnosed before 16 years of age. In addition to identifying *LTA* as a major gene associated with early-onset leprosy, our study highlights the critical role of case- and population-specific factors in the dissection of susceptibility variants in complex diseases.

Leprosy is a chronic infectious disease caused by *Mycobacterium leprae* that still affects an estimated 300,000 new patients each year[3]. Susceptibility to leprosy strongly depends on the genetic background of the host[4,5]. Building on previous segregation studies[6,7] we detected linkage of leprosy susceptibility to chromosome region 6q25–q26 in a genome-wide scan in Vietnamese multiplex leprosy families[1]. We subsequently identified *PARK2* and *PACRG* variants as strong leprosy susceptibility factors[2]. The second linkage signal was detected in the 6p21 chromosomal region, and fine mapping placed the peak (lod score = 2.7) close to D6S2427, at approximately 39.5 Mb on build 36 of the human genome sequence map. Based on the 95% confidence interval (c.i.) for the location of the underlying leprosy susceptibility locus[8], we selected a chromosomal region extending from 31.6 Mb to 42.0 Mb on the physical map (**Fig. 1**) for association studies in a sample of 194 simplex leprosy families from Vietnam. This interval contains the 108 genes located within the HLA class II and class III regions and 116 annotated genes centromeric to HLA class II (*PHF1*→*CCND3*; NCBI build 36). We genotyped 418 SNPs in the target interval, 307 of which were suitable for association analysis (**Fig. 1** and **Supplementary Table 1** online).

In univariate analysis, six SNPs were significantly associated with leprosy at the 0.01 level; four of these were located in the HLA class III region (**Fig. 1** and **Supplementary Table 2** online). Stepwise multivariate conditional logistic regression analysis with these six SNPs showed that the best model ($P = 0.000003$) included four SNPs (rs2071590, rs3128961, rs937662 and rs707928) with marginal multivariate P values of 0.001, 0.007, 0.014 and 0.039, respectively. Of note, none of the second-order interactions between these four SNPs was significant at the 0.05 level. Multivariate analysis also demonstrated the effects of these four SNPs to be independent of the two *PARK2* and *PACRG* SNPs previously associated with leprosy in the same sample[2] (**Supplementary Table 3** online). Thus, the most significant independently associated SNP was rs2071590, also known as *LTA*–293, located in the promoter region of the lymphotoxin-α (*LTA*) gene. Notably, the third most significant SNP in univariate analysis, rs3131628 (univariate $P = 0.0023$), was in linkage disequilibrium (LD) with *LTA*–293 ($r^2 = 0.5$), clearly identifying the *LTA* region as the principal target for high-resolution association mapping.

[1]Laboratoire de Génétique Humaine des Maladies Infectieuses, Institut National de la Santé et de la Recherche Médicale, U550, 75015 Paris, France. [2]Université Paris René Descartes, Faculté Médecine Necker, 75015 Paris, France. [3]McGill Centre for the Study of Host Resistance and Departments of Human Genetics, Medicine and Biochemistry, McGill University, 1650 Cedar Avenue, Montreal, Quebec H3G1A4, Canada. [4]Hospital for Dermato-Venereology, Nguyen Thong Street, District 3, Ho Chi Minh City, Vietnam. [5]Department of Transplant Immunology and Immunogenetics, All-India Institute of Medical Sciences, Ansari Nagar, 110029 New Delhi, India. [6]Leprosy Laboratory, Tropical Medicine Department, Oswaldo Cruz Institute, FIOCRUZ, CEP 21040-900 Rio de Janeiro, Brazil. [7]Central JALMA Institute of Leprosy and Other Infectious Diseases, Taj Ganj, 282001 Agra, India. [8]Centro de Ciências Biológicas e da Saúde, Pontifícia Universidade Católica do Paraná, Rua Imaculada Conceição, 1155, CEP 80215-901, Curitiba, Paraná, Brasil. [9]These authors contributed equally to this work and are listed in alphabetical order. Correspondence should be addressed to L.A. (abel@necker.fr) or E.S. (erwin.schurr@mcgill.ca).

Received 21 December 2006; accepted 8 February 2007; published online 11 March 2007; doi:10.1038/ng2000

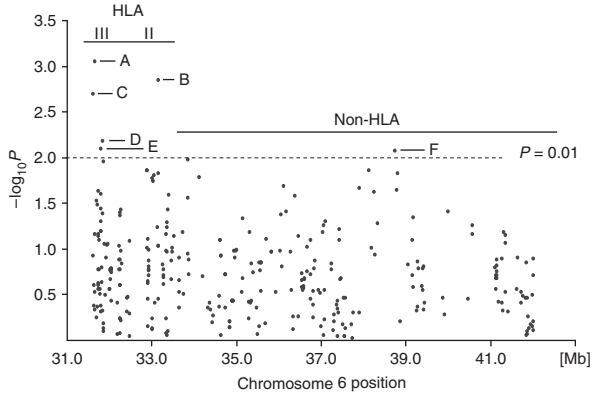


Figure 1 High-density SNP association scan of a 10.4-Mb region on chromosome 6p21 in 194 Vietnamese trios. Evidence for association with leprosy of 307 SNPs is plotted as $-\log_{10}P$ on the *y*-axis. The location of SNPs, in Mb, on the sequence map (build 36) is indicated on the *x*-axis. The borders of the HLA class II and class III regions and of the studied non-HLA region toward the centromere are indicated by horizontal lines. The thin dotted line indicates the $P = 0.01$ significance threshold. The six SNPs associated with leprosy at $P < 0.01$ are indicated by letters A–F. SNP A corresponds to rs2071590 (also known as *LTA*–293), and SNPs B–F correspond to rs3128961, rs3131628, rs707928, rs805305 and rs937662, respectively. We tested three different genetic models for each SNP (additive, fully dominant and fully recessive), and the *P* values for the best model are reported here.

Over an interval of approximately 90 kb extending from *PPIAP9* to *NCR3* and overlapping the *TNF* and *LTA* genes, we analyzed 29 SNPs, eight of which were in strong LD and associated with leprosy at the 0.05 level (**Fig. 2a**). Seven of these SNPs, including *LTA*–293, belonged to the same bin (as defined in Methods), whereas the eighth (rs3131628) had an r^2 value of 0.5 with *LTA*–293 (**Fig. 2a**). The remaining SNPs were also in strong LD, with more than half (13/21) grouped into four additional bins (**Fig. 2** and **Supplementary Table 4** online). The evidence of association with leprosy remained strongest for *LTA*–293, assuming dominant inheritance of the 'T' risk allele ($P_{dom} = 0.0009$; odds ratio (OR) for TT/CT versus CC = 1.97 (95% c.i., 1.30–2.99)). However, exhaustive multivariate conditional logistic regression showed that each of the SNPs in the *LTA*–293 bin was sufficient to account for the observed association of the *LTA* region with leprosy. Therefore, it is not possible to identify the susceptibility variant within the *LTA*–293 bin based on statistical considerations alone.

To more accurately define the boundaries of the *LTA*–293 bin, we first searched the four populations of HapMap release 20 to identify additional SNPs that were in $r^2 > 0.8$ with any of the SNPs in the *LTA*–293 bin. Although we used a 1-Mb window around *LTA*, only four SNPs met this criterion, all of which were located in the intergenic region between *NFKBIL1* and *LTA* and have no known functional role. To delineate the boundaries of the *LTA*–293 bin in our Vietnamese sample, we selected 31 SNPs located on the 31.48 Mb to 32.30 Mb chromosomal segment that, in combination with all previously genotyped HapMap SNPs, capture >75% of the non-single SNP bins using an 0.8 r^2 threshold (**Supplementary Table 5** online). This interval (that is, *MICA*→*NOTCH4*) spans the entire class III region in addition to the class I region centromeric to *HLA-B*. The 31 SNPs were genotyped in 54 unrelated Vietnamese subjects, and none of the 29 informative SNPs showed an $r^2 > 0.25$ with SNPs in the *LTA*–293 bin. Finally, we also determined HLA class I genotypes in 37 out of the 54 aforementioned subjects, given the close proximity of the *HLA-B* and *HLA-C* loci to *LTA* (200 kb and 300 kb, respectively). Neither the *HLA-A*, the *HLA-B* nor the *HLA-C* alleles resulted in an $r^2 > 0.25$ with *LTA*+80 (**Supplementary Table 6** online), confirming the independence of the *LTA* association from class I alleles. These results, which are entirely consistent with the HapMap LD data in all four populations, argue against the extension of the *LTA*–293 bin beyond a 37-kb interval located in the telomeric end of the HLA class III region. Notably, *LTA*+80 ($P_{dom} = 0.007$; OR for AA/AC versus CC = 1.74 (95% c.i., 1.16–2.60)) is the only SNP in this bin previously shown to have an independent biological function[9]; therefore, it appears to be the most obvious candidate for the susceptibility variant.

To replicate the association of *LTA*–293 bin SNPs with leprosy, we enrolled an additional sample of 364 unrelated individuals with leprosy and 371 unrelated controls from northern India. We tested association with leprosy for the six SNPs at the *LTA* locus (*LTA*–294, *LTA*–293, *LTA*+10, *LTA*+80, *LTA*+252 and *LTA*+368; **Table 1**). Univariate logistic regression indicated that only SNP *LTA*–294 was

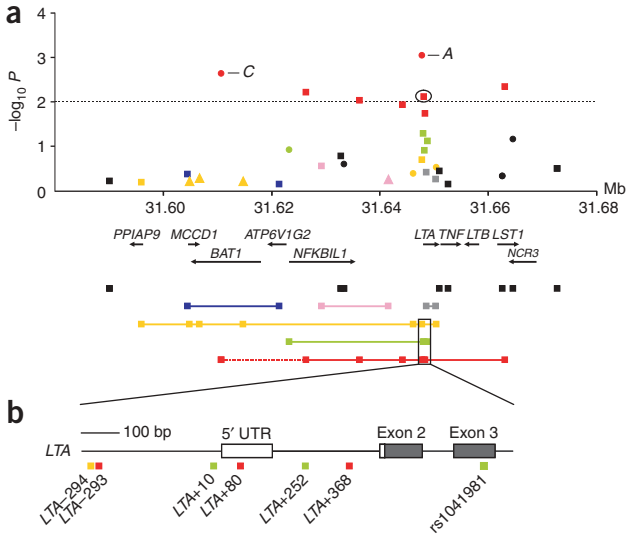


Figure 2 Fine linkage disequilibrium mapping of a 90-kb interval in the HLA class III region in 194 Vietnamese trios. (**a**) Evidence for the association of 33 SNPs spanning a 90-kb interval in the HLA class III region with leprosy is expressed as $-\log_{10} P$ and plotted against SNP position in Mb. The probability threshold $P = 0.01$ is indicated by the dotted line. The four SNPs represented by triangles were selected based on their linkage disequilibrium with *LTA*–294 in at least one HapMap project population. SNPs identical in color belong to the same bin set (as defined in the Methods). SNPs in black do not belong to any bin set. The SNPs from the primary scan are represented by a filled circle. *LTA*+80 is represented by the red square surrounded by a black oval. The designations and locations of the ten known genes in the 90-kb interval are shown. A linear view of the bin structure consistent with the color scheme used above is shown immediately below. The dotted line in the red bin set indicates a pairwise $r^2 = 0.5$. Circles: SNPs from primary scan; triangles, HapMap SNPs in LD with *LTA*-294; squares: SNPs selected for fine-mapping of the 90-kb interval. (**b**) Expanded view of the 5′ region of *LTA*. The proximal promoter and the intronic regions are represented by solid lines. The 5′ UTR is represented by open boxes, and translated exon regions are represented by solid boxes. The relative locations of seven *LTA* SNPs are shown (*LTA*–294, *LTA*–293, *LTA*+10, *LTA*+80, *LTA*+252, *LTA*+368 and rs1041981). The color of each SNP is consistent with the bin set to which it belongs. SNPs A and C correspond to rs2071590 and rs3131628, respectively.

Table 1 Univariate and multivariate analyses of *MCCD1-NS* and six *LTA* SNPs in Vietnamese and Indian samples

	MCCD1-NS rs2259435	*LTA*-294 rs2844482	*LTA*-293 rs2071590	*LTA*+10 rs1800683	*LTA*+80 rs2239704	*LTA*+252 rs909253	*LTA*+368 rs746868
Vietnamese							
Bin structure							
MAF[b]	0.16	0.16	0.26	0.50	0.30	0.49	0.30
Univariate OR (95% CI) *P*	NS	NS	1.97 (1.30–2.99) 0.0009	NS	1.74 (1.16–2.60) 0.007	NS	1.63 (1.09–2.43) 0.02
Multivariate OR (95% CI) *P*	NS	NS	1.97 (1.30–2.99) 0.0009	NS	NS[a]	NS	NS[a]
Indian							
Bin structure							
MAF[b]	0.23	0.22	0.26	0.28	0.42	0.28	0.41
Univariate OR (95% CI) *P*	1.59 (1.24–2.04) 0.0002	1.78 (1.29–2.45) 0.0004	NS	NS	NS	NS	NS
Multivariate OR (95% CI) *P*	1.82 (1.38–2.41) 0.0004	NS[a]	NS	NS	1.60 (1.10–2.33) 0.01	NS	NS[a]

[a]Any SNP in the bin is sufficient to explain the observed association with leprosy. [b]MAF (minor allele frequency) corresponds to the frequency of the risk allele for each of the associated SNPs.

significantly associated (P_{dom} = 0.0004) with leprosy (**Table 1**). However, multivariate analysis showed that when *LTA*–294 was included in the model, the *LTA*+80 SNP became significant (P_{dom} = 0.026 with allele A being at risk), whereas *LTA*–293 remained nonsignificant (P_{dom} = 0.18). These discrepant results between *LTA*+80 and *LTA*–293 were explained by different LD structures between the Indian and the Vietnamese populations (**Table 1**). In particular, in the Indian population, *LTA*–293 did not belong to the same LD bin as *LTA*+80 (r^2 = 0.45) and therefore could be ruled out as a leprosy risk factor.

To follow up on the observation that *LTA*+80 and *LTA*–294 were independently associated with leprosy susceptibility only among the Indians, we investigated whether an unknown SNP in LD with *LTA*–294 in the Indian population but not in the Vietnamese population could have been missed in our first analysis. We searched HapMap release 20, and we identified four additional SNPs in strong LD ($r^2 > 0.8$) with *LTA*–294 in at least one of the HapMap populations (**Supplementary Table 4**). We genotyped these SNPs in the Vietnamese and Indian samples and observed that three belonged to the same bin as *LTA*–294 in both populations. In the Vietnamese population, none of these four SNPs was associated with leprosy (**Fig. 2a**). In the Indian sample, the association with leprosy was the strongest for the A allele of marker rs2259435 (alias: *MCCD1-NS*), a nonsynonymous SNP located in the first exon of the mitochondrial coiled-coil domain 1 (*MCCD1*) gene (P_{add} = 0.0002; OR for AA versus AG = OR for AG versus GG = 1.59 (95% c.i., 1.24–2.04)). In multivariate analysis, the best-fitting model included both *MCCD1-NS* and *LTA*+80 (P = 0.00009 for the global model; **Supplementary Table 7** online). With adjustment for *MCCD1-NS*, the *LTA*+80 A allele was found to be significantly more frequent among cases, whereas *LTA*–293 again was not associated with leprosy (**Table 1**). Thus, by taking into account the *MCCD1-NS* effect, we were able to detect and replicate association of the *LTA*+80 variant with leprosy in the Indian population (P_{dom} = 0.01; OR for AA/AC versus CC = 1.60 (95% c.i., 1.10–2.33)) with the same risk allele as the one observed in the Vietnamese families.

We then enrolled a third sample of 209 unrelated individuals of European ancestry with leprosy and 192 healthy controls, matched based on self-reported ethnicity, from the city of Rio de Janeiro in Brazil. We studied the six SNPs in the *LTA* promoter region plus six additional SNPs in strong LD with either the *LTA*+80 or the *LTA*–294 marker. None of these SNPs was significantly associated with leprosy in the Brazilian sample in either univariate or multivariate analysis, despite this sample having a posterior power of 84% to detect an effect of the *LTA*+80 A risk allele assuming a type I error of 0.05 (based on the observed minor allele frequency (MAF) among the unaffected Brazilians of 0.35 and assuming an OR of 1.74). We hypothesized that the association of *LTA*+80 with leprosy might depend on sampling bias, owing to the family-based design of the Vietnamese study. Indeed, the median age at diagnosis was much lower ($P < 10^{-6}$) for affected individuals in the Vietnamese sample (16 years) than for those in the Indian (31 years) and Brazilian (38 years) samples.

Therefore, we carried out further analyses in which we stratified individuals based on age at leprosy diagnosis. The proportion of affected individuals in the four age groups considered is shown in **Figure 3**. In all three ethnic groups studied, we observed that the younger the individual, the stronger the association of *LTA*+80 with leprosy (**Fig. 3**). Notably, the association in the Vietnamese sample related almost entirely to individuals diagnosed before the age of 16 years (P = 0.00004; OR = 5.76 (95% c.i., 2.25–14.78)), and there was significant evidence for heterogeneity of the *LTA*+80 effect between affected individuals younger than 16 years and those older than 16 years (P = 0.00054; **Fig. 3**). Although not statistically significant (P = 0.07), we also observed a clear enrichment of the *LTA*+80 A allele in the youngest Brazilian age group (16–25 years) with an OR for AA/AC versus CC of 2.76 (95% c.i., 0.74–10.32). In addition, this trend toward a decrease in OR with increasing age was significant in the Brazilian sample (P = 0.04). Finally, we found the same age-at-diagnosis effect on the association between *LTA*+80 and leprosy in the Indian sample, with an OR for AA/AC versus CC of 2.95 (95% c.i., 1.32–6.58) (P_{dom} = 0.006) for the 16- to 25-year-old age group versus 1.60 (95% c.i., 1.10–2.33) for the overall sample. Evidence for genetic heterogeneity between the youngest age group (16–25 years) and the older patients (>25 years) was highly significant (P = 0.003).

These results strongly indicated that the effect of *LTA*+80 on leprosy risk was age dependent. We enrolled a fourth independent sample of 104 simplex leprosy families from Vietnam to validate this conclusion. The median age at diagnosis of the affected individuals in these families was 19 years (the age distribution is shown in **Fig. 3**). Overall, again we observed a significant association between the *LTA*+80 A allele and leprosy risk (P_{dom} = 0.003; OR for AA/AC versus CC = 2.34 (95% c.i., 1.27–4.31)). As observed in the first 194 Vietnamese families studied, the association was strongest for the youngest individuals (age at diagnosis, <16 years) with an OR of 5.31 (95% c.i., 1.19–23.60), and there was significant evidence (P = 0.04) for heterogeneity of the *LTA*+80 effect between affected individuals younger than 16 years and those older than 16 years (**Fig. 3**, 'Vietnam 2' sample). Finally, when

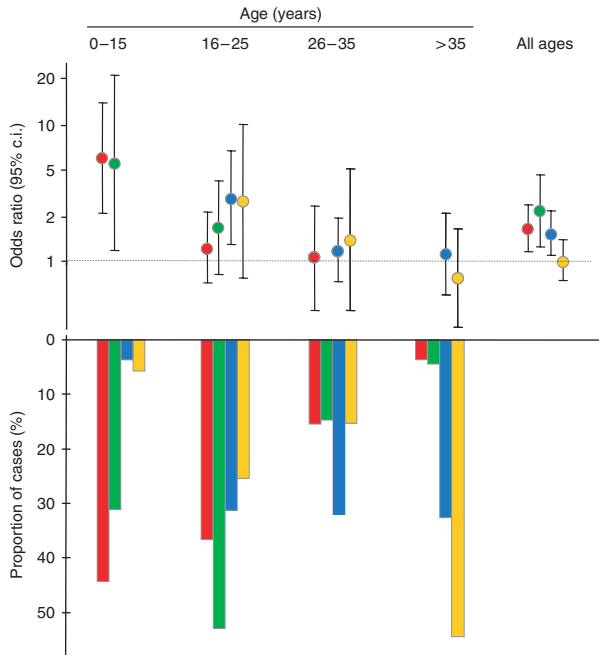


Figure 3 Age-dependence of the *LTA*+80 A allele effect on leprosy susceptibility. A total of 869 individuals with leprosy from four independent samples (two family-based samples from Vietnam and two population-based case-control samples from India and Brazil) were grouped into four age-at-diagnosis classes (0–15 years, 16–25 years, 26–35 years and older than 35 years). The proportion of cases in each age group is indicated in the lower part of the graph. The total number of cases was 194 for the Vietnam 1 sample, 104 for the Vietnam 2 sample, 207 for the Brazilian sample and 364 for the Indian sample. In the upper part of the figure, the ORs of leprosy for *LTA*+80 AA/AC versus CC subjects (circles) and their 95% c.i. values are plotted on a logarithmic scale for each age group in each ethnic group. In the Indian sample, the ORs are adjusted for the *MCCD1-NS* effect. Red: Vietnam 1; green: Vietnam 2; blue: India; yellow: Brazil. ORs were not calculated for age groups including fewer than 15 affected individuals.

we combined both Vietnamese samples, the evidence for association of *LTA*+80A with leprosy in the overall sample and in the group of individuals diagnosed before the age of 16 years was very strong, with ORs of 2.11 (95% c.i., 1.48–3.01; $P = 0.000024$) and 5.63 (95% c.i., 2.54–12.49; $P = 0.0000004$), respectively.

HLA class II *DRB1* alleles have been repeatedly identified as leprosy susceptibility factors[5,10,11], and the *HLA-DRB1* locus is located about 1 Mb centromeric to *LTA*. Therefore, we investigated in more detail whether the observed associations of leprosy with *LTA*+80 could be due to long-range LD with HLA class II *DRB1* alleles. We first determined *DRB1* genotypes in our initial Vietnamese family sample, and multivariate analysis showed that the effect of *LTA*+80 remained unchanged when we included *DRB1* genotypes in the model both in the overall sample ($P = 0.008$; OR for *LTA*+80 AA/AC versus CC = 1.75 (95% c.i., 1.15–1.62)) and the early onset age group ($P = 0.000005$; OR = 6.00 (95% c.i., 2.29–15.7)). Consistent with this result, we did not detect any $r^2 > 0.1$ between *LTA*+80 and *DRB1* alleles in this Vietnamese sample. Next, we genotyped the *HLA-DRB1* locus in unrelated Indian subjects, and again did not find any $r^2 > 0.1$ between alleles of this locus and *LTA*+80. These results demonstrate the independence of the *LTA*+80 effect from the known *DRB1* leprosy susceptibility gene in the Vietnamese and Indian samples.

Following on the heels of the identification of *PARK2* and *PACRG* variants as major leprosy risk factors[2,12], the present results define a new genetic element affecting leprosy susceptibility and shed further light on the complex genetic interplay governing susceptibility to a common infectious disease. We conclude that *LTA*+80 is not the only leprosy risk factor in the 6p21 region. First, the *LTA*+80 effect is clearly independent of *DRB1*, repeatedly identified as an important locus in leprosy control. Second, we detected significant association of the *MCCD1-NS* bin with leprosy in the Indian individuals. Although we did not observe this association in the Vietnamese and Brazilian samples, this finding suggests that there may be additional leprosy susceptibility genes in the *HLA* region. Finally, it is clear from our results that *LTA*+80 did not have a substantial role in leprosy susceptibility in individuals over the age of 25 years. Additional investigation of candidate genes in the target interval should therefore be carried out, carefully taking the affected individuals' ages into account.

The *LTA*+80 polymorphism is an attractive candidate leprosy susceptibility factor, regardless of the other susceptibility variants that may exist. It was localized to a regulatory E2-box motif (CAGCTG) with a 1-bp mismatch (CAGCAG). The C allele of *LTA*+80 results in an additional mismatch (CCGCAG), preventing binding of the activated B-cell factor-1 transcriptional repressor. Therefore, the A allele is associated with significantly lower levels of expression of a downstream reporter than the C allele[9]. Our findings thus indicate that low levels of LTA production are associated with a higher risk of leprosy. This result is consistent with studies in animal models linking disruption of the LTA signaling pathway with an increase in susceptibility to intracellular pathogens[13–15]. Irradiated mice with immune system reconstitution from LTA-deficient bone marrow have been shown to be extremely susceptible to virulent *Mycobacterium tuberculosis*, owing to the impaired colocalization of T cells with macrophages in lung granulomas[16]. The association of the *LTA*+80 suppressive A allele with a higher risk of leprosy may therefore be mediated by poor lymphocyte recruitment to sites of infection.

Our results obtained in Vietnamese and Indian samples demonstrate that *LTA*+80 is an important risk factor for leprosy that acts in a highly age-dependent manner. This study also provides an instructive example of how a linkage peak in a complex disease may reflect a complex underlying structure of susceptibility variants. Moreover, the results obtained highlight some of the pitfalls encountered when replicating genetic effects in independent samples and the importance of considering population-specific LD patterns and covariates properly in replication studies. In particular, our study illustrates that genotyping in replication samples cannot be confined to the one SNP with strongest evidence for association in the initial study but needs to encompass all SNPs in the initially associated bin. Overall, our observations should prove useful for the design and interpretation of allelic association replication studies in other complex diseases.

METHODS

Affected individuals and controls. We enrolled a total of 2,027 subjects in this study. The 194 Vietnamese simplex leprosy families (Vietnam 1) have been described elsewhere[1]. The additional 104 Vietnamese families (Vietnam 2) were identified from the records available at the Dermato-Venereology Hospital in Ho Chi Minh City. No particular efforts were made to enroll younger individuals, and the criteria for enrollment were identical to those used for the Vietnam 1 sample: that is, both parents had to be available for genetic analysis in all families of the Vietnam 1 and Vietnam 2 samples. Overall, the mean age at diagnosis (± s.d.) was 19.4 years ± 8.3 years in individuals in the Vietnam 1 sample and 20.6 years ± 7.8 years in individuals in the Vietnam 2 sample. In north India, 364 individuals with leprosy (33.3 years ± 13.3 years)

were enrolled from records at the Central JALMA Institute for Leprosy and Other Mycobacterial Diseases at Agra. As a control group, 370 healthy north Indian individuals (32.1 years ± 10.9 years) with no documented history of chronic infectious or inflammatory diseases were enrolled from blood donor clinics. In Brazil, 207 individuals with leprosy (38.8 years ± 16.8 years) of self-reported European ancestry were recruited from the Leprosy Outpatient Clinic at the Oswaldo Cruz Institute in Rio de Janeiro, Brazil. As a control group, 192 control subjects matched based on self-reported ethnicity (33.1 years ± 8.8 years) with no history of chronic infectious, inflammatory or autoimmune disease were recruited from the same geographic area of Rio de Janeiro. The study was approved by institutional review boards and health authorities in Ho Chi Minh City, Vietnam; the Oswaldo Cruz Institute, Rio de Janeiro, Brazil; the All India Institute of Medical Sciences, New Delhi, India and the McGill University Health Centre, Montreal, and written informed consent was obtained from all participants.

Phenotype definition. The diagnosis of the Vietnamese individuals and the definition of clinical subtypes have been described elsewhere[1]. The Brazilian and Indian cases were diagnosed and classified on the basis of clinical and histological criteria[17]. As in both Vietnamese family samples, the two case-control samples included a balanced mixture of paucibacillary and multibacillary cases. However, the phenotype studied here was leprosy *per se* (leprosy independent of specific clinical manifestations).

Genotyping methods. The 418 SNPs spanning the 10.4-Mb targeted interval on chromosome region 6p21 were selected on the basis of their proximity to or location within known genes in the interval (**Supplementary Table 1**). All SNPs were genotyped on the ultra–high-throughput Illumina platform. This platform uses the GoldenGate assay followed by a bead-based technology to resolve individual SNP genotypes[18]. A total of 54 SNPs could not be genotyped with this assay, and one SNP could not be placed unambiguously on the sequence map. We excluded 56 of the remaining 363 SNPs from the analysis for the following reasons: 12 showed deviations ($P < 0.05$) from Hardy-Weinberg equilibrium (HWE) among parents, and 44 were noninformative or had a MAF <5%.

The additional 25 informative SNPs spanning the refined 90-kb target interval in the MHC class III region (**Supplementary Table 4** online), as well as the panel of 31 SNPs used to study LD in the vicinity of the *LTA* gene (**Supplementary Table 5** online) were selected based on information regarding bin structure and allelic frequencies publicly available from the International HapMap project (see URL section below). These SNPs were genotyped on one or several of the following platforms: (i) genotyping by direct sequencing on an ABI PRISM 3100 genetic analyzer; (ii) genotyping on the high-throughput GenomeLab SNPstream platform, which uses a single–base pair extension (SBE) method to incorporate fluorescently labeled terminal nucleotides, which are then detected by a specialized imager[19]; (iii) genotyping on the high-throughput SEQUENOM MassARRAY platform, which uses the iPLEX assay to incorporate mass-modified terminal nucleotides in the SBE step that are then detected by matrix-assisted laser desorption/ionization–time of flight (MALDI–TOF) mass spectrometry[20] and (iv) genotyping by the polarized fluorescence TaqMan assay[21]. Five SNPs, including *LTA*+80, were genotyped by two independent methods for all Vietnam 1 samples, and the few individuals (<0.5%) with discrepant genotypes were eliminated from subsequent analyses. Genotyping of the HLA *DRB1* locus was performed using standard methods[22]. *HLA-A*, *HLA-B* and *HLA-C* genotyping was performed using a PCR reverse sequence–specific oligonucleotide (PCR-SSO) probe hybridization technique as recommended by the manufacturer (InnoLipa-HLA, Innogenetics).

Statistical methods. We estimated pairwise LD between SNPs by determining r^2 from parental data for Vietnamese samples and from unaffected controls for the Indian and the Brazilian samples, using the algorithm implemented in Haploview software[23]. Bins of SNPs were constructed on the basis of the pairwise r^2 between the SNPs and were defined as a not necessarily contiguous set of SNPs in which at least one SNP has $r^2 > 0.8$ with all the other SNPs of the bin[24].

As the Vietnamese samples consisted of trios with no missing parental data, family-based association studies were performed principally by a classical transmission disequilibrium test, as implemented in FBAT software[25]. Alleles for which there was evidence of association were also analyzed by conditional logistic regression, as previously described[26]. This approach permitted estimation of OR and made it possible to carry out multivariate stepwise regression tests and tests for heterogeneity. We tested for heterogeneity according to age in the Vietnamese samples by carrying out the analysis on the combined sample (for example, 194 trios in the Vietnam 1 sample) and on the following two subsamples: families in which affected individuals were diagnosed before 16 years of age (for example, 86 trios) and families in which affected individuals were diagnosed at or over the age of 16 years (for example, 108 trios). Under the hypothesis of homogeneity of association, twice the difference between the likelihood of the whole sample and the summed likelihoods of the two subsamples is asymptotically distributed as a χ^2 with 1 degree of freedom. Conditional logistic analysis was performed with the PHREG procedure implemented in SAS software v.8.2.

We carried out a population-based association study in the Indian and Brazilian samples, using classical multivariate logistic regression techniques as implemented in the LOGISTIC procedure of SAS software v.8.2. All analyses were adjusted for sex, which has been identified as a classical risk factor for leprosy. To investigate whether the association of leprosy with the *LTA*+80 A allele decreased with age in the Brazilian sample, we tested in the logistic model if the regression coefficient corresponding to an interaction between *LTA*+80 (coded as 0 or 1, for CC or AC/AA subjects, respectively) and the age classes (coded as 0, 1 or 2 for age ranges 16–25 years, 26–36 years or >35 years, respectively) was significantly <0.

We tested for differences in median ages across three populations (Vietnam 1, Brazil and India), using the nonparametric Brown-Mood test, as implemented in the NPAR1WAY procedure of SAS 8.2. This test is equivalent to a one-way analysis of variance with median scores[27].

URLs. HapMap project: http://www.hapmap.org/

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank all affected individuals, their families and population controls for participating in this study. We thank T.J. Hudson, P. Lepage, E.N. Sarno and J.L. Casanova for helpful comments and discussions; A. Montpetit and A. Belisle for assistance with high-throughput genotyping; S. Caillat-Zucman for help in HLA genotyping and L. Simkin for technical assistance. This study was supported by a grant from the Canadian Institutes of Health Research (CIHR) to E.S. A. Alter holds a graduate studentship from the Natural Science and Engineering Research Council of Canada (NSERC). A. Alter and M.O. are supported by the CIHR Strategic Training Centre in the Integrative Biology of Infectious Diseases and Autoimmunity. G.A. is supported by Fondation pour la Recherche Médicale. E.S. is a Chercheur National of the Fonds de Recherche en Santé du Québec and an International Research Scholar of the Howard Hughes Medical Institute. A. Alcaïs and L.A. are supported by the Assistance Publique-Hopitaux de Paris, Programme de Recherche Fondamentale en Microbiologie Maladies Infectieuses et Parasitaires and the Agence Nationale de la Recherche of the Ministère Français de l'Education Nationale de la Recherche et de la Technologie. Requests for materials should be addressed to L.A. (abel@necker.fr) or E.S. (erwin.schurr@mcgill.ca).

AUTHOR CONTRIBUTIONS

Genotyping and SNP selection were performed by A. Alter, G.A., M.O., M.S. P.R.V. and M.T.M. Statistical analysis was performed by A. Alcais, A. Alter and G.A. Clinical analyses and phenotype assessment were performed by N.V.T., K.K., V.H.T., N.T.H., N.N.B. and M.M. A. Alcais, M.M., N.M., E.S. and L.A. designed the study. A. Alcais, A. Alter, E.S. and L.A. wrote the paper. E.S. and L.A. share senior authorship of this work.

COMPETING INTERESTS STATEMENT

The authors declare no competing financial interests.

Published online at http://www.nature.com/naturegenetics
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions

1. Mira, M.T. *et al.* Chromosome 6q25 is linked to susceptibility to leprosy in a Vietnamese population. *Nat. Genet.* **33**, 412–415 (2003).

2. Mira, M.T. *et al.* Susceptibility to leprosy is associated with PARK2 and PACRG. *Nature* **427**, 636–640 (2004).
3. WHO. Global leprosy situation, 2006. *Wkly. Epidemiol. Rec.* **81**, 309–316 (2006).
4. Mira, M.T. Genetic host resistance and susceptibility to leprosy. *Microbes Infect.* **8**, 1124–1131 (2006).
5. Casanova, J.L. & Abel, L. Genetic dissection of immunity to mycobacteria: The human model. *Annu. Rev. Immunol.* **20**, 581–620 (2002).
6. Abel, L. & Demenais, F. Detection of major genes for susceptibility to leprosy and its subtypes in a Caribbean island: Desirade island. *Am. J. Hum. Genet.* **42**, 256–266 (1988).
7. Abel, L. *et al.* Complex segregation analysis of leprosy in Southern Vietnam. *Genet. Epidemiol.* **12**, 63–85 (1995).
8. Dupuis, J. & Siegmund, D. Statistical methods for mapping quantitative trait loci from a dense set of markers. *Genetics* **151**, 373–386 (1999).
9. Knight, J.C., Keating, B.J. & Kwiatkowski, D.P. Allele-specific repression of lymphotoxin-alpha by activated B cell factor-1. *Nat. Genet.* **36**, 394–399 (2004).
10. de Vries, R.R. *et al.* HLA-linked control of susceptibility to tuberculoid leprosy and association with HLA-DR types. *Tissue Antigens* **16**, 294–304 (1980).
11. Zerva, L. *et al.* Arginine at positions 13 or 70–71 in pocket 4 of HLA-DRB1 alleles is associated with susceptibility to tuberculoid leprosy. *J. Exp. Med.* **183**, 829–836 (1996).
12. Alcais, A., Mira, M., Casanova, J.L., Schurr, E. & Abel, L. Genetic dissection of immunity in leprosy. *Curr. Opin. Immunol.* **17**, 44–48 (2005).
13. Bopst, M. *et al.* Differential effects of TNF and LTalpha in the host defense against M. bovis BCG. *Eur. J. Immunol.* **31**, 1935–1943 (2001).
14. Jacobs, M., Brown, N., Allie, N. & Ryffel, B. Fatal Mycobacterium bovis BCG infection in TNF-LT-alpha-deficient mice. *Clin. Immunol.* **94**, 192–199 (2000).
15. Ware, C.F. Network communications: lymphotoxins, LIGHT, and TNF. *Annu. Rev. Immunol.* **23**, 787–819 (2005).
16. Roach, D.R. *et al.* Secreted lymphotoxin-alpha is essential for the control of an intracellular bacterial infection. *J. Exp. Med.* **193**, 239–246 (2001).
17. Ridley, D.S. & Jopling, W.H. Classification of leprosy according to immunity. A five-group system. *Int. J. Lepr. Other Mycobact. Dis.* **34**, 255–273 (1966).
18. Fan, J.B. *et al.* Highly parallel SNP genotyping. *Cold Spring Harb. Symp. Quant. Biol.* **68**, 69–78 (2003).
19. Bell, P.A. *et al.* SNPstream UHT: ultra-high throughput SNP genotyping for pharmacogenomics and drug discovery. *Biotechniques* (Suppl.), 70–77 (2002).
20. Griffin, T.J. & Smith, L.M. Single-nucleotide polymorphism analysis by MALDI-TOF mass spectrometry. *Trends Biotechnol.* **18**, 77–84 (2000).
21. Lee, L.G., Connell, C.R. & Bloch, W. Allelic discrimination by nick-translation PCR with fluorogenic probes. *Nucleic Acids Res.* **21**, 3761–3766 (1993).
22. Olerup, O. & Zetterquist, H. HLA-DR typing by PCR amplification with sequence-specific primers (PCR-SSP) in 2 hours: an alternative to serological DR typing in clinical practice including donor-recipient matching in cadaveric transplantation. *Tissue Antigens* **39**, 225–325 (1992).
23. Barrett, J.C., Fry, B., Maller, J. & Daly, M.J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–265 (2005).
24. Hinds, D.A. *et al.* Whole-genome patterns of common DNA variation in three human populations. *Science* **307**, 1072–1079 (2005).
25. Horvath, S., Xu, X. & Laird, N.M. The family based association test method: strategies for studying general genotype–phenotype associations. *Eur. J. Hum. Genet.* **9**, 301–306 (2001).
26. Schaid, D.J. & Rowland, C. Use of parents, sibs, and unrelated controls for detection of associations between genetic markers and disease. *Am. J. Hum. Genet.* **63**, 1492–1506 (1998).
27. Conover, W.J. *Practical Nonparametric Statistics* (John Wiley & Sons, New York, 1999).

Hum Genet
DOI 10.1007/s00439-009-0775-x

ORIGINAL INVESTIGATION

Genetic and functional analysis of common *MRC1* exon 7 polymorphisms in leprosy susceptibility

Andrea Alter · Louis de Léséleuc · Nguyen Van Thuc · Vu Hong Thai · Nguyen Thu Huong · Nguyen Ngoc Ba · Cynthia Chester Cardoso · Audrey Virginia Grant · Laurent Abel · Milton Ozório Moraes · Alexandre Alcaïs · Erwin Schurr

Received: 22 September 2009 / Accepted: 4 December 2009

Abstract The chromosomal region 10p13 has been linked to paucibacillary leprosy in two independent studies. The *MRC1* gene, encoding the human mannose receptor (MR), is located in the 10p13 region and non-synonymous SNPs in exon 7 of the gene have been suggested as leprosy susceptibility factors. We determined that G396S is the only non-synonymous exon 7-encoded polymorphism in 396 unrelated Vietnamese subjects. This SNP was genotyped in 490 simplex and 90 multiplex leprosy families comprising 704 patients (47% paucibacillary; 53% multibacillary). We observed significant under-transmission of the serine allele of the G396S polymorphism with leprosy per se ($P = 0.036$) and multibacillary leprosy ($P = 0.034$). In a sample of 384 Brazilian leprosy cases (51% paucibacillary; 49% multibacillary) and 399 healthy controls, we observed significant association of the glycine allele of the G396S polymorphism with leprosy per se ($P = 0.016$) and multibacillary leprosy ($P = 0.023$). In addition, we observed a significant association of exon 7 encoded amino acid haplotypes with leprosy per se ($P = 0.012$) and multibacillary leprosy ($P = 0.004$). Next, we tested HEK293 cells over-expressing MR constructs (293-MR) with three exon 7 haplotypes of *MRC1* for their ability to bind and internalize ovalbumin and zymosan, two classical MR ligands. No difference in uptake was measured between the variants. In addition, 293-MR failed to

A. Alter, L. de Léséleuc contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s00439-009-0775-x) contains supplementary material, which is available to authorized users.

A. Alter · L. de Léséleuc · E. Schurr
Department of Medicine, McGill Centre for the Study of Host Resistance, Research Institute of the McGill University Health Centre, McGill University, Montreal, QC, Canada

A. Alter · L. de Léséleuc · E. Schurr
Department of Human Genetics, McGill Centre for the Study of Host Resistance, Research Institute of the McGill University Health Centre, McGill University, Montreal, QC, Canada

N. Van Thuc · V. H. Thai · N. T. Huong · N. N. Ba
Hospital for Dermato-Venereology, Nguyen Thong Street, District 3, Ho Chi Minh City, Vietnam

C. C. Cardoso · M. O. Moraes
Leprosy Laboratory, Oswaldo Cruz Institute, FIOCRUZ, Rio de Janeiro, Brazil

A. V. Grant · L. Abel · A. Alcaïs
Laboratoire de Génétique Humaine des Maladies Infectieuses, Institut National de la Santé et de la Recherche Médicale U550, 75015 Paris, France

A. V. Grant · L. Abel · A. Alcaïs
Faculté Médicine Necker, Université Paris Descartes, 75015 Paris, France

L. Abel
Laboratory of Human Genetics of Infectious Diseases, Rockefeller Branch, The Rockefeller University, New York, NY 11065, USA

E. Schurr (✉)
Montreal General Hospital Research Institute, Rm. L11-521, 1650 Cedar Avenue, Montreal, QC H3G 1A4, Canada
e-mail: erwin.schurr@mcgill.ca

Published online: 25 December 2009

bind and internalize viable *Mycobacterium leprae* and BCG. We propose that the MR–*M. leprae* interaction is modulated by an accessory host molecule of unknown identity.

Introduction

Leprosy, caused by the human pathogenic bacterium *Mycobacterium leprae (M. leprae*) is a chronic infectious disease that has been recognized for centuries. Leprosy is primarily a disease of the skin and nerves but the mechanisms of pathogenesis that lead to cutaneous spread of bacilli and irreversible peripheral nerve damage are poorly understood. The disease manifests itself in two clinical forms that have been termed "paucibacillary" and "multibacillary" leprosy depending on the number of skin lesions and the number of bacilli found in these lesions. The development of leprosy is strongly influenced by host genetics as illustrated by the high concordance rates among monozygotic twins (Chakravartti and Vogel 1973). Consequently, a strategy to better understand the disease pathogenesis is to uncover genes that modulate leprosy susceptibility. Past genome-wide linkage analyses have identified several susceptibility loci in chromosome regions 6q25–q26 (Mira et al. 2003), 6p21 (Miller et al. 2004; Mira et al. 2003), 10p13 (Siddiqui et al. 2001), 17q22 (Miller et al. 2004), 20p12–13 (Miller et al. 2004; Tosh et al. 2002) and 21q22 (Wallace et al. 2004). So far, high density linkage disequilibrium (LD) mapping of the linkage peaks on chromosome regions 6q25–q26 and 6p21 has successfully identified two susceptibility loci by positional cloning, *PARK2/PACRG* (Mira et al. 2004) and *LTA* (Alcais et al. 2007), respectively.

The 10p13 chromosomal region carries a replicated susceptibility locus for paucibacillary leprosy (Mira et al. 2003; Siddiqui et al. 2001). The *MRC1* gene underlies the corresponding linkage peak and three closely spaced amino acid changes encoded by exon 7 of *MRC1* have been suggested to be associated with altered susceptibility to paucibacillary leprosy (Cooke and Hill 2008; Hill 2006). *MRC1* encodes the mannose receptor C-type lectin (MR), a cell surface protein that belongs to a family of receptors for pathogen-associated molecular patterns, or PAMPs, and is part of the innate arm of the immune system. MR is a receptor for mannose, fucose and *N*-acetyl-glucosamine-containing molecules and binds to yeast cell walls (Ezekowitz et al. 1990), to capsular polysaccharides of *Klebsiella pneumonia* (Kabha et al. 1995) as well as to mycobacterial components such as mannose-capped lipoarabinomannans (ManLAM) (Schlesinger et al. 1996). MR has been shown to be an important mediator between *M. tuberculosis* and the host immune system (Kang et al. 2005; Nigou et al. 2001; Schlesinger 1993). Most notably, cellular entry through MR has been correlated with *M. tuberculosis* virulence (Schlesinger 1993; Schlesinger et al. 1996). Taken together, these data prompted us to study *MRC1* as a possible leprosy susceptibility gene.

We analyzed the role of *MRC1* exon 7 non-synonymous coding polymorphisms in leprosy susceptibility in a familial sample from Vietnam and a case–control sample from Brazil. We observed in both populations that the G396S polymorphism was significantly associated with leprosy per se and multibacillary leprosy. However, we found in the Brazilian sample that the risk effect of the glycine residue at position 396 depended on the exon 7 haplotype. In functional analysis, we failed to observe an impact of the exon 7 polymorphisms on MR mediated uptake of zymosan and ovalbumin. In addition, HEK293 cells over-expressing MR (293-MR) were unable to bind and ingest viable *M. leprae* and *Bacille Calmette Guerin* (BCG) bacteria. We concluded that exon 7 polymorphisms mediated the interaction of MR with an unknown accessory molecule that is required for MR-mediated uptake of *M. leprae.*

Methods

Patients and controls

In Vietnam, 490 simplex (53% multibacillary) and 90 multiplex (55% multibacillary) leprosy families were identified from the records available at the Dermato-Venereology Hospital in Ho Chi Minh City (Alcais et al. 2007; Mira et al. 2003). The criterion for enrollment was the availability of both parents for genetic analysis. In Brazil, 384 leprosy patients were recruited from the Leprosy Outpatient Clinic at the Oswaldo Cruz Institute in Rio de Janeiro. The 399 healthy controls (i.e., no reported infectious or inflammatory disease) were recruited from the same geographic area of Rio de Janeiro. Cases and controls were matched based on self-reported ethnicity. In previous experiments, using genomic controls, we failed to observe significant evidence for population stratification in case–control samples matched on self-reported ethnicity from the same geographic area (Mira et al. 2004). The diagnosis of all leprosy patients and the definition of clinical subforms were based on clinical and histological criteria (Ridley and Jopling 1966). Informed consent was obtained from all study participants. The study was approved by institutional review boards and health authorities in Ho Chi Minh City, Vietnam, the Oswaldo Cruz Institute, Rio de Janeiro, Brazil, and the Research Institute of the McGill University Health Centre, Montreal, Canada.

SNP selection

The identification of informative coding SNPs (i.e., minor allele frequency (MAF) >5%) was done by sequencing 23 unrelated Vietnamese individuals for all 30 *MRC1* exons, including both untranslated (UTR) regions, using an ABI PRISM® 3100 genetic analyzer. In addition, exon 7 was sequenced in 396 unrelated Vietnamese individuals (a subset of parents from the 490 Vietnamese simplex families). For subsequent analysis, the six identified informative coding SNPs were designated by their corresponding 'rs' numbers. In total, 75 SNPs spanning *MRC1* (101.8 kb) on chromosome region 10p12.33 were selected based on allelic frequencies publicly available from the NCBI EntrezSNP database (http://www.ncbi.nlm.nih.gov/sites/entrez?db=snp) and the International HapMap project (http://www.hapmap.org/) (Table S1). Conflicting annotations between physical maps regarding gene localization and supposed gene duplication (i.e., *MRC1* and *MRC1L1*) posed technical challenges to the genetic analysis of *MRC1*, particularly the resolution of SNP positions. For consistency, SNPs with two positions were mapped to *MRC1* spanning chr10:17,891,368–17,993,183 (HapMap Data Rel 27) and were referred to by their *MRC1L1* 'rs' numbers (NCBI EntrezSNP database Build 130). We analyzed *MRC1* exon 7 chromatograms in 2,423 unrelated individuals (from our Vietnamese and Brazilian samples and the HGDP-CEPH Human Genome Diversity Cell Line Panel) and found no evidence for a common gene duplication event (i.e., biallelic ratio for SNPs was 1:1). The data argues against the duplication of *MRC1* and suggests that *MRC1L1* is an erroneous annotation caused by the presence of a sequence gap and the incorrect assignment of a polymorphic haplotype.

Genotyping methods

The SNPs were genotyped on one or several of the following platforms (1) genotyping by direct sequencing on an ABI PRISM® 3100 genetic analyzer; (2) genotyping on the high-throughput GenomeLabB™ SNPstream® platform (formerly Orchid SNPstream UHT), which uses a single-base pair extension (SBE) method to incorporate fluorescently labeled terminal nucleotides, which are then detected by a specialized imager (Bell et al. 2002); (3) genotyping on the high-throughput SEQUENOM® MassARRAY® platform, which uses the iPLEX™ assay to incorporate mass-modified terminal nucleotides in the SBE step, which are then detected by MALDI-TOF MS (Griffin and Smith 2000); (4) genotyping on the ultra-high throughput Illumina® platform. This platform uses the GoldenGate™ assay followed by a bead-based technology to resolve individual SNP genotypes (Fan et al. 2003). Three SNPs, including rs1926736, were genotyped by two independent methods and the few individuals (<1.7%) with discrepant genotypes were eliminated from subsequent analyses. A total of 15 SNPs could not be successfully genotyped and one SNP could not be placed unambiguously on the sequence map. We excluded 12 of the remaining 59 SNPs from the analysis for the following reasons: 1 showed deviations ($P < 0.01$) from Hardy–Weinberg equilibrium among parents and 11 were non-informative or had a MAF < 5%.

Reagents and antibodies

Dulbecco's modified Eagle Medium, Minimum Essential Medium (MEM) alpha, RPMI-1640, GlutaMAX, penicillin, streptomycin, fetal bovine serum (FBS), AlexaFluor 488-conjugated ovalbumin (OVA488), FITC-conjugated zymosan, rabbit anti-FITC, AlexaFluor 488-conjugated goat anti-mouse antibody and 4',6-diamidino-2-phenylindole (DAPI) were purchased from Invitrogen (Carlsbad, CA, USA). Endothelial cell medium (ECM) was from Sciencell (Carlsbad, CA, USA). Puromycin, yeast mannan, digitonin, paraformaldehyde, polybrene and porcine gelatin were purchased from Sigma (St. Louis, MO, USA). Anti-MR monoclonal antibody (clone 15-2) was purchased from Cell Sciences (Canton, MA, USA). Cy3-conjugated goat anti-mouse and Cy5-conjugated goat anti-rabbit antibodies were from Jackson ImmunoResearch (West Grove, PA, USA). Permafluor mounting medium was purchased from Thermo Scientific (Waltham, MA, USA). YG fluorescent 1 μm polystyrene microspheres were obtained from Polysciences Inc. (Warrington, PA, USA).

Cell culture

The HEK293 cell line was purchased from the American Type Culture Collection (ATCC, Manassas, VA, USA) and grown in MEM-alpha. Phoenix-Ampho cells were purchased from ATCC with permission of Gary Nolan (Stanford University, Palo Alto, CA, USA). These media were supplemented with 10% heat-inactivated FBS, 100 U/ml penicillin and 100 U/ml streptomycin and 2 mM GlutaMAX. Human dermal microvascular endothelial cells (HDMECs) were purchased from Sciencell and cultured in ECM containing 5% FBS, 100 penicillin and 100 U/ml streptomycin, in gelatin-coated flasks. All cells were cultured in a humidified incubator at 37°C and 5% CO_2. When mycobacteria were added, cells were cultured under the same conditions except that antibiotics were omitted and in the case of *M. leprae*, the temperature was set to 33°C.

Expression vector cloning

The cDNA encoding the human mannose receptor and cloned in pCDNA3 was a generous gift of Dr. J. J. He (Indiana University, IN, USA). The MR cDNA was amplified by proofreading PCR using primers MR_fwd_SalI, 5′-ATGCGTCGACATGAGGCTACCCCT GCTCC-3′, and MR_rev_MfeI, 5′-GCATCAATTGCTA GATGACCGAGTGTTCA-3′. The PCR product was purified, digested with *Sal*I and *Mfe*I and ligated in the pMSCV-puro vector (Clontech, Mountain View, CA) previously cut with *Xho*I and *Eco*RI. A total of 2 µl of the ligation was used to transform STLB2 competent cells (Invitrogen). A positive clone was grown overnight at 30°C in 50 ml of culture broth and the plasmid DNA was extracted using a Perfectprep plasmid midi kit (Eppendorf, Hamburg, Germany). The nucleotide sequence was verified by cycle sequencing in an ABI PRISM® 3100 genetic analyzer (Applied Biosystems, Foster City, CA, USA).

Site-directed mutagenesis

Two polymorphic variants of *MRC1* were generated using the original hMR cDNA as a template for site-directed mutagenesis by overlap extension PCR (Boskovic et al. 2006). MR_fwd_SalI and MR_rev_MfeI were used as the flanking universal primers. Primers 5′-GCAGGAAG GAAGGCAGTGACCTCGCGAGTATCC-3′ (sense) and 5′-GGATACTCGCGAGGTCACTGCCTTCCTTCCTGC-3′ (antisense) were used for the generation of the MR(SAF) haplotype. Primers 5′-AAGTATCCACACCATCGAGG AATTGGACTTT-3′ (sense) and 5′-ATGGTGTGGATAC TTGTGAGGTCACCGCCTT-3′ were used for the generation of the MR (GTL) haplotype. All constructs were cloned in pMSCV-puro and their sequence was verified.

Transduction

Amphotropic retroviruses were produced by transfecting the Phoenix-Ampho cell line with pMSCV-puro (empty) or pMSCV-MR constructs using Fugene6 (Roche Diagnostics, Laval, QC, Canada) according to the manufacturer's instructions. Forty-eight hours after transfection, the culture supernatant was harvested, filtered, aliquoted and stored at −80°C. Aliquots showed similar (within margin of error) titers when tested on HEK293 cells. For transductions, HEK293 cells were seeded and grown to 50% confluence, then infected with Phoenix-Ampho supernatants supplemented with 8 µg/ml of polybrene for 24 h. The medium was then removed and fresh medium containing 1 µg/ml of puromycin was added. Stable pools of transductants (293-puro and 293-MR) were collected after 2 weeks of selection.

Immunostaining

Expression of MR was analyzed by indirect immunofluorescence. For microscopy, cells were seeded on gelatin-coated coverslips and fixed with 4% paraformaldehyde (PFA) diluted in PBS for 10 min. Alternatively, single cell suspensions were prepared for cytometry and fixed in 4% PFA. Cells were then incubated for 30 min in SuperBlock blocking buffer (Pierce, Rockford, IL, USA) containing 0.1% Tween-20, then incubated with 1 µg/ml of anti-MR mouse monoclonal antibody for 16 h at 4°C in the same buffer. Cells were rinsed twice in PBS and incubated with either Alexa Fluor 488- or Cy3-conjugated secondary antibodies (1:2,000 dilution in SuperBlock) for 60 min at 24°C. For microscopy, coverslips were rinsed twice then mounted onto slides in Permafluor mounting medium containing DAPI. For cytometry, suspensions were processed in a FACSCalibur flow cytometer (BD Biosciences).

Ovalbumin uptake assay

Cell lines were seeded in 12-well plates at a density of 5×10^5 cells/well and left to attach overnight. They were then treated for 1 h with 5 µg/ml of OVA488 in fresh complete medium. Cells were washed twice in PBS, detached in PBS containing 0.2 g/L EDTA and processed for flow cytometry. Fluorescence was measured in the FL1 channel.

Zymosan uptake assay

Cell lines were seeded in 12-well plates at a density of 5×10^5 cells/well and left to attach overnight. They were treated with zymosan-FITC at a ratio of five particles per cell for 16 h in fresh complete medium. Cells were washed twice in PBS then fresh complete medium containing 1 µg/ml anti-FITC antibody was added. Cells were kept at 4°C for 30 min to stop phagocytosis and allow binding of antibodies to surface exposed zymosans. Cells were washed twice in PBS and lysed for 10 min in 400 µl lysis buffer (0.01% digitonin in PBS) containing 10^6 YG 1 µm beads (about 1 bead per cell) and anti-rabbit Cy5 diluted 1:200. The lysate was further homogenized in a bath sonicator for 5 min. Samples were processed by flow cytometry, surface-bound particles were positive for both the FL1 (green) and FL4 (far red) channels whereas internalized zymosans were scored negative in FL4.

Mycobacteria binding assay

PKH67-stained viable *M. leprae* bacilli grown in *nude* mouse footpads were obtained from Dr R. Truman and

Dr L. Adams (Hansen's Disease Program; Baton Rouge, LA, USA). GFP-expressing *M. bovis* BCG strain Pasteur was cultured in Middlebrook 7H9 medium containing ADC enrichment and used fresh to make single-cell suspensions. The 293-puro or 293-MR cells were seeded on coverslips at a density of 2.5×10^5 cells/well, in complete medium without antibiotics then treated with mycobacteria at a multiplicity of infection (MOI) of five for 16 h. In some experiments, cells were pre-incubated for 30 min with yeast mannan or anti-MR monoclonal antibody. Cells were washed three times in PBS then fixed in 4% paraformaldehyde and briefly permeabilized in cold methanol. Nuclei were counterstained with DAPI, then coverslips were mounted and examined by fluorescence microscopy. Images of both green particles and nuclei were acquired using the Northern Eclipse software (Mississauga, ON, USA). Bound bacteria were manually counted in at least 4 fields containing a minimum of 500 cells total.

Statistical methods

We estimated population allelic frequencies from parental data using the algorithm implemented in Haploview 4.1 software (Barrett et al. 2005). Family based association studies were performed using principally a classical transmission disequilibrium test, as implemented in FBAT v2.0.3 software (Horvath et al. 2001). For the analysis of the multi-case families and the combined sample, we used the empirical variance–covariance estimator previously advocated (Lake et al. 2000). We carried out a population-based association study in the Brazilian sample, using classical multivariate logistic regression techniques as implemented in the LOGISTIC procedure of SAS software v.9.1. Haplotype analyses were further carried out using the THESIAS software (Tregouet and Garelle 2007). All population-based analyses were adjusted for sex, which has been identified as a classical risk factor for leprosy. We did not correct for multiple testing given the context of a positive replication utilizing an alternative study design in an ethnically different population, and the reported association of the *MRC1* exon 7 SNPs with leprosy in two reviews (Cooke and Hill 2008; Hill 2006).

Results

Analysis of exon 7 SNPs in Vietnamese leprosy families

Three non-synonymous exon 7 SNPs were of particular interest since they had been suggested as risk factors for paucibacillary leprosy in a South Indian population (Cooke and Hill 2008). To investigate the impact of *MRC1* exon 7 SNPs on leprosy susceptibility, we first sequenced exon 7 in 396 unrelated healthy individuals (792 chromosomes) to identify any common population-specific variants. We found two polymorphic SNPs: the non-synonymous rs1926736 [G396S; allele A (S396) MAF = 0.35] and the synonymous rs2437256 [I404I; allele C (I404) MAF = 0.21]. Two additional previously described non-synonymous SNPs, rs2478577 (T399A) and rs2437257 (L407F), were found to be non-polymorphic in the Vietnamese population (MAF = 0). Since the focus of the study was on non-synonymous SNPs, we genotyped the single non-synonymous SNP rs1926736 in 490 simplex families and 90 multiplex families comprising a total of 704 leprosy patients for which both parents were available. These families included approximately even numbers of paucibacillary ($n = 325$) and multibacillary ($n = 374$) cases. Under a best dominant model, we observed significant evidence for the association of rs1926736 with leprosy per se [$P_{dom} = 0.035$, allele A (S396) is protective, OR = 0.76 (95% CI 0.60–0.96)] and multibacillary leprosy [$P_{dom} = 0.034$, allele A (S396) is protective, OR = 0.71 (0.51–0.99)] (Table 1).

To delineate the extent of LD between rs1926736 and additional *MRC1* SNPs, we excluded rare SNPs (i.e., MAF < 5%) given that they cannot be highly correlated with rs1926736 (MAF = 0.35). We selected 69 SNPs from the NCBI EntrezSNP database and the International HapMap project that span the entire 101.8 kb of the *MRC1* gene. To assure inclusion of all common coding SNPs, we sequenced the remaining 29 *MRC1* exons including the 5′- and 3′-UTR regions in 23 unrelated Vietnamese individuals. According to our criterion, four identified coding SNPs, all previously described in the NCBI EntrezSNP database, were included: rs2253120 (exon 2, T81T), rs2296414 (exon 3, T167I), rs2985837 (exon 4, Q242Q) and rs2477664 (exon 4, I260I) (Table S2). Including the two previously described exon 7 SNPs, a total of 75 *MRC1* SNPs were genotyped in 198 simplex families to establish the LD pattern of the *MRC1* gene. Among the 75 *MRC1* SNPs genotyped, 47 were suitable for analysis (Table S1). The LD pattern of the entire *MRC1* gene was plotted and revealed that rs1926736 was highly correlated with only two intronic SNPs (rs525830, $r^2 = 0.89$ and rs493862, $r^2 = 0.86$) with no obvious functional implication (Fig. 1; Table S3). In addition, none of the additional SNPs showed significant evidence for association with leprosy or its subforms in the 198 families (data not shown). Finally, we utilized the International HapMap Project database (Data Rel 24) to screen a 2-Mb region around rs1926736 in the Han Chinese in Beijing, China (CHB) samples. No SNPs were correlated with rs1926736 using an $r^2 = 0.5$ cut-off. From these results, we concluded that rs1926736 (G396S) was the most plausible cause of the observed association and that S396 of MR was a leprosy protective factor.

Table 1 Association analysis of *MRC1* exon 7 SNP rs1926736 (G396S) with leprosy and its sub-forms

Phenotype	Number of families	Amino acid (frequency)	S^a	$E(S)^b$	Z^c	P value	OR[d]	95% CI[e]
Per se	339	S (0.370)	202	222	−2.102	0.035	0.76	0.60–0.96
	190	G (0.630)	135	143	−1.069	0.280		
Paucibacillary	172	S (0.369)	106	111.25	−0.823	0.410	0.82	0.58–1.15
	104	G (0.631)	69	3.25	−0.847	0.397		
Multibacillary	186	S (0.371)	100	114.75	−2.125	0.034	0.71	0.51–0.99
	99	G (0.629)	69	73.25	−0.735	0.462		

[a] FBAT statistic

[b] Expected FBAT statistic

[c] Z score

[d] Odds ratio

[e] 95% confidence interval

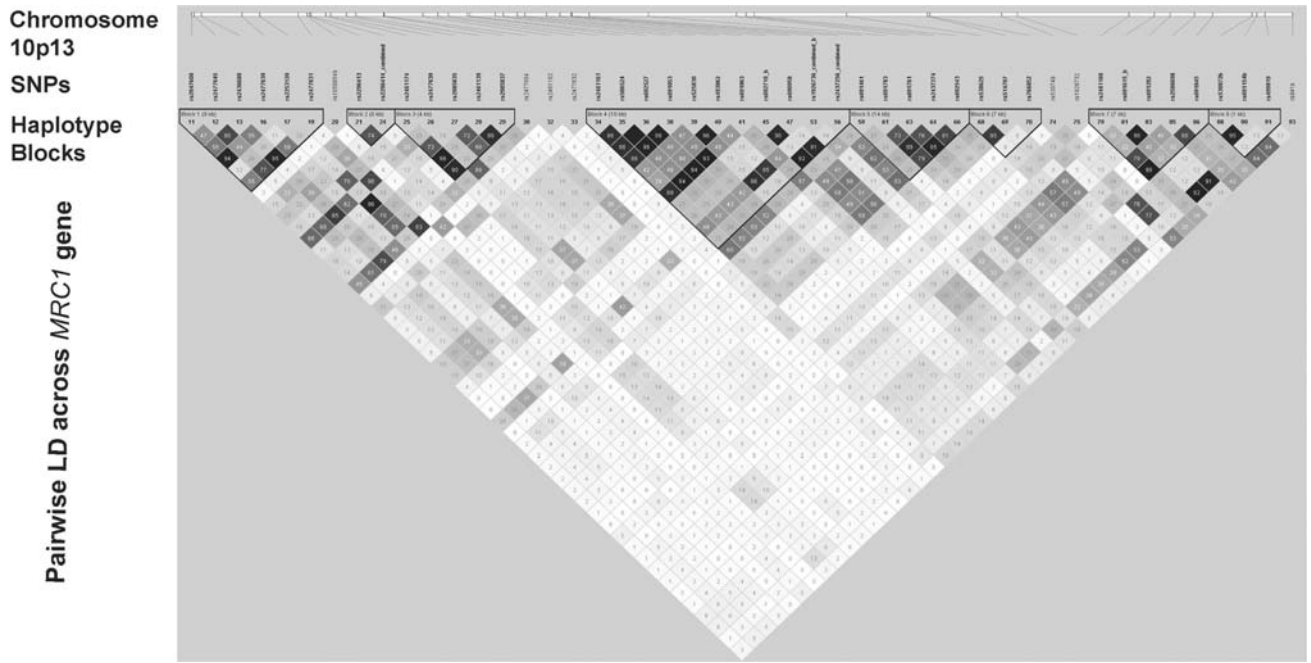


Fig. 1 Pairwise linkage disequilibrium (LD) as defined by r^2 across the *MRC1* gene. The chromosome 10p location of 47 SNPs from rs2947600 (*MRC*1 intron 1) to rs941 (*MRC1* 3′-UTR) is indicated above the LD plot. The r^2 value for each SNP pair is indicated within the corresponding diamond. Increasing depth of black color indicates higher r^2 values. *D'-defined haplotype blocks* are also indicated at the top of the LD plot. There is no evidence for long range LD extending over the entire *MRC1* gene

Analysis of exon 7 SNPs in Brazilian leprosy cases

Given that the evidence for association of rs1926736 with leprosy in the Vietnamese samples was weak, we decided to study *MRC1* exon 7 SNPs in a leprosy sample from Brazil. We sequenced exon 7 in 399 healthy controls and observed four SNPs: rs1926736 [G396S; allele A (S396) MAF = 0.32], rs2478577 [T399A; allele A (T399) MAF = 0.21], rs2437256 [I404I; allele C (I404) MAF = 0.21], and rs2437257 [L407F; allele G (L407) MAF = 0.21]. The three non-synonymous exon 7 SNPs could be unambiguously assigned to three amino acid haplotypes, two of which had been observed in the Vietnamese sample (Fig. 2). Next, we sequenced *MRC1* exon 7 in an additional 384 Brazilian leprosy cases and conducted a case–control association analysis of the three non-synonymous SNPs with leprosy. LD analysis showed that SNPs rs2478577 (T399A) and rs2437257 (L407F) were in complete LD ($r^2 = 1$) and therefore statistically redundant. Allele G (G396) of SNP rs1926736 showed evidence for association with susceptibility to leprosy per se [$P_{add} = 0.016$; OR = 1.34 (95% CI 1.06–1.70)], and the multibacillary sub-form [$P_{add} = 0.023$; OR = 1.42 (1.05–1.93)] in a

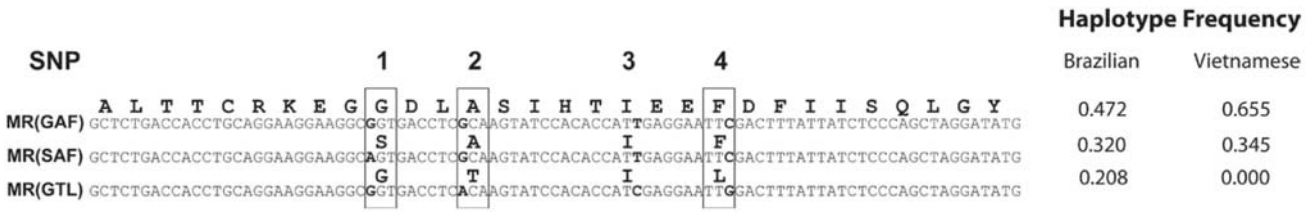


Fig. 2 Human *MRC1* exon 7 haplotypes. In the Brazilian sample, four common *MRC1* exon7 polymorphisms were found, three of which were non-synonymous (boxed). Due to strong LD between SNPs two (rs2478577), three (rs2437256) and four (rs2437257), only three amino acid haplotypes were observed and termed MR (GAF), MR (SAF) and MR (GTL). SNP 1 (rs1926736) gives rise to a G/S amino acid polymorphism that was only observed on the MR(AF) background. Among these three haplotypes, only two–MR(GAF) and MR(SAF)–were observed in the Vietnamese sample

Table 2 *MRC1* exon7 haplotype analysis in the Brazilian case–control sample

rs1926736–rs2437257 haplotype	Controls[a]	Cases[a]	Global test *P* value[b]	Marginal test *P* value[b]	OR (95% CI)
Per se (343 controls, 329 cases)					
G396–F407	0.47	0.55	0.012	0.003	1.41 (1.13–1.76)
G396–L407	0.21	0.18		0.256	0.85 (0.64–1.12)
S396–F407	0.32	0.27		0.019	0.75 (0.59–0.95)
Paucibacillary (343 controls, 167 cases)					
G396–F407	0.47	0.53	0.125	0.068	1.28 (0.98–1.67)
G396–L407	0.21	0.19		0.479	0.89 (0.64–1.24)
S396–F407	0.32	0.28		0.172	0.82 (0.61–1.09)
Multibacillary (343 controls, 162 cases)					
G396–F407	0.47	0.57	0.004	0.001	1.61 (1.21–2.14)
G396–L407	0.21	0.18		0.174	0.78 (0.55–1.12)
S396–F407	0.32	0.25		0.015	0.68 (0.50–0.93)

[a] Proportion of controls and cases

[b] All *P* values adjusted for sex

best additive model. Similarly, allele G (L407) of SNP rs2437257 showed borderline evidence for association with resistance to leprosy per se [$P_{dom} = 0.09$; OR = 0.75 (0.54–1.05)] and the multibacillary sub-form [$P_{dom} = 0.04$; OR = 0.63 (0.41–0.97)] in a best dominant model.

Multivariate analysis of the three non-synonymous variants confirmed the correlation between SNPs rs2478577 and rs2437257. The best sex-adjusted model ($P = 0.002$) as determined by multivariate analysis included only rs1926736 and rs2437257. Haplotype analysis of rs1926736 (G396S) and rs2437257 (L407F) identified three haplotypes with G396-F407 being the most common haplotype (Table 2). Haplotype G396-F407 had a risk effect while haplotype S396-F407 was protective in the overall sample (Table 2). Considering rs2478577 (T399A) and rs2437257 (L407F) are in complete LD, *MRC1* exon 7 therefore encodes two protective amino acid haplotypes: the borderline protective G396-T399-L407 haplotype and the strongly protective S396-A399-F407 haplotype (Table 2). Interestingly, while the G396 residue is a susceptibility factor overall, on the T399-L407 haplotype background, this susceptibility can be compensated and may even be protective. This is concurrent with the result in the Vietnamese leprosy families. Given that in the Vietnamese population rs2478577 (T399A) and rs2437257 (L407F) are non-polymorphic, the protective S396 allele corresponds unambiguously to the S396-A399-F407 haplotype (Fig. 2).

Establishment of an HEK293 cellular model for MR function

The three *MRC1* exon 7 encoded amino acid haplotypes observed in Vietnamese and Brazilian samples were termed MR(GAF), MR(SAF) and MR(GTL) (Fig. 2). The underlying three non-synonymous amino acid changes G396S, T399A and L407F map to the CTLD2 segment of the mannose receptor protein, a domain of unknown function. To better understand the functional basis of the observed genetic effect of exon 7 haplotypes, we decided to compare the biological activity of the corresponding MR genetic variants by ectopically expressing the receptor in a

non-expressing cell line. A cDNA copy of human *MRC1* in pcDNA3 (a generous gift of Dr. J. J. He, Indiana University, IN, USA) was obtained and subcloned in a MSCV-puro retroviral backbone. Packaged viruses were used to infect the HEK293 cell line, which does not express MR. The resulting pool of transductants (293-MR) was compared to an empty-vector control HEK293 line as well as to HDMEC, which are known to express MR (Groger et al. 2000). In addition to MR protein expression, the capacity to internalize fluid-phase (ovalbumin) and particulate (zymosan) ligands was evaluated by microscopy. Immunofluorescence showed that 293-MR and endothelial cells mostly expressed the mannose receptor in an intracellular endocytic compartment as well as on their surface, which overlapped the pattern of ovalbumin uptake (Fig. 3a). The control cell line did not stain for MR and was unable to take up ovalbumin. In addition, both 293-MR and HDMECs could bind and internalize yeast particles (zymosans) (Fig. 3b). Hence, the 293-MR cell line was considered a good model for the study of ectopically expressed MR since it mirrored the natural activity of the receptor.

Cloning and expression of human mannose receptor variants

The original *MRC1* cDNA sequence used in the generation of the 293-MR cell line corresponded to the MR(GAF) haplotype. This sequence was modified by site-directed mutagenesis at two sites to yield the MR(SAF) and MR(GTL) variants (Fig. 2). The DNAs were cloned in a retroviral expression vector, packaged and used to infect HEK293 cells. The entire pools of puromycin-resistant transduced cells were used at passage 3 for further experiments to avoid artifacts due to clonal variations. The proportion and level of MR expression were measured by immunocytometry (Fig. 4a). MR(GAF) and MR(SAF) pools had very similar expression levels whereas the MR(GTL) pool was slightly lower, possibly due to a somewhat lower retroviral transduction efficiency during the generation of this cell line. Expression levels were monitored throughout the experiments and were found to be stable. Vector-transduced control cells showed no expression.

MR variants do not differ in their capacity to take up ligands

The three pools of MR-expressing transductants were then used to compare the functional activity of their corresponding MR proteins by employing ovalbumin and zymosan, two classical mannose-containing ligands. Fluid-phase endocytosis of ovalbumin was measured by flow cytometry. There was no difference in endocytosis between

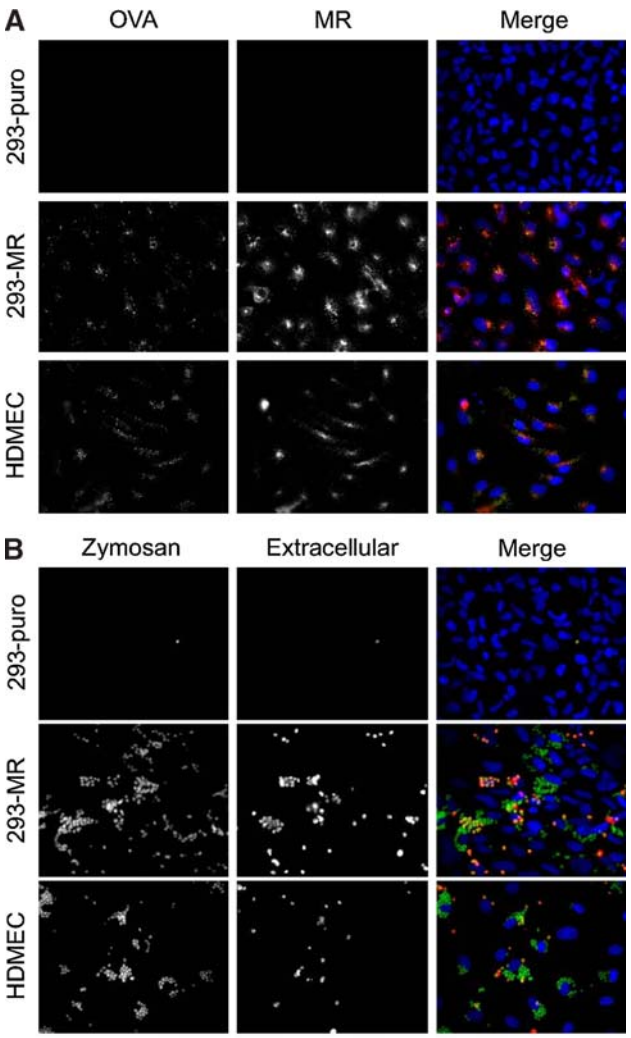


Fig. 3 Ectopic MR is functional in HEK293 cells. A MR cDNA was transduced into HEK293 cells. The resulting cells were compared to vector-transduced cells and microvascular endothelial cells (HDMEC) which express endogenous MR. **a** Cells were treated with Alexa-488-conjugated ovalbumin (an MR ligand, shown in *green*) for 1 h then immunostained with monoclonal anti-MR followed by Cy3-conjugated anti-mouse (*red*). **b** Cells were treated with fluorescein-conjugated zymosan for 16 h, then washed and stained with rabbit anti-fluorescein followed by Cy3 anti-rabbit antibodies. Ingested zymosans are *green* while non ingested particles are stained *red* and appear *yellow* in the merged images. Nuclei are stained blue with DAPI

MR(GAF) and MR(SAF) variants and only a slightly lower activity for MR(GTL) that was consistent with a somewhat lower expression (Fig. 4b). Vector control cells showed only background level uptake. It has been suggested that the processes of MR-dependent binding and endocytosis of soluble molecules might differ from MR-dependent phagocytosis of particles (Le Cabec et al. 2005). Therefore, we simultaneously measured binding and internalization of fluorescinated zymosans by flow cytometry. Ingestion efficiency was determined by the ratio of internalized

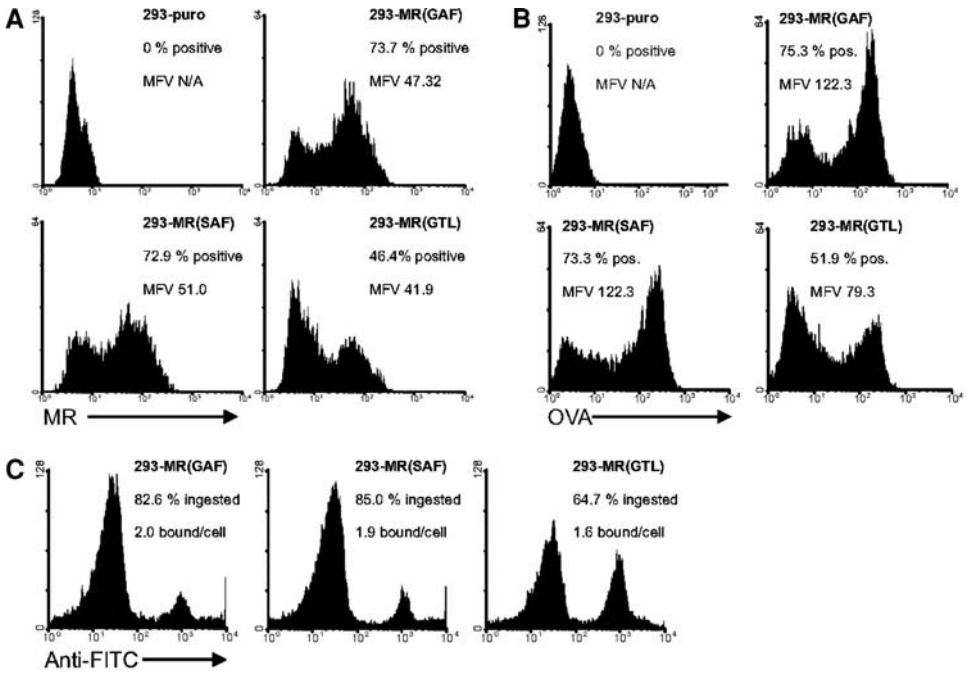


Fig. 4 Absence of biological differences among MR variants. HEK293 transductants were analyzed for MR function by flow cytometry. **a** Transduced stable populations of HEK293 cells were immunostained for MR. **b** Populations were incubated for 1 h with 5 μg/ml Alexa488-conjugated ovalbumin and uptake was measured by FACS. Percentage of positive cells is given as well as median fluorescence value (MFV) of the positive population. **c** Populations were incubated with 10^6 zymosan-FITC particles (ratio zymosan:cell 5:1) for 16 h. Extracellular particles were then stained with anti-FITC and Cy5-conjugated secondary antibodies. Cells were lysed and intracellular and extracellular zymosans were gated in FL1 and counted in FL4 channels. Data is expressed as percent of ingested zymosans over total bound, and as number of bound zymosan per cell. Results shown here are representative of three independent experiments

particles over total cell-associated particles. Again, MR(GAF) and MR(SAF) variants were indistinguishable while the lower expressing MR(GTL) variant ingested slightly lower numbers (Fig. 4b). Overall, the MR-expressing cells took up 64–83% of bound particles after 16 h of incubation. Vector control cells did not bind sufficient numbers of particles to allow quantification of phagocytosis.

MR expressed by HEK293 cells does not bind to *M. leprae* or BCG

Mycobacteria are thought to bind to MR because of the presence of mannosylated molecules such as ManLAM on their surface (Kang et al. 2005; Schlesinger et al. 1996). We therefore evaluated the impact of MR polymorphisms on the interaction with whole viable *M. leprae* and *M. bovis* BCG. Fluorescent mycobacteria were incubated with 293-puro and 293-MR cells for 16 h. Binding was determined by microscopic examination of the monolayer after repeated washing. The presence of MR had no impact on binding of either mycobacteria (Fig. 5). Control cells could bind bacteria to the same extent as MR^+ cells, regardless of the variant expressed. These results suggested that 293-MR cells lacked accessory molecules required for efficient uptake of mycobacteria by MR.

Discussion

Linkage of the 10p13 chromosomal region to the paucibacillary sub-form was observed independently in two ethnically different populations from South India and Vietnam (Mira et al. 2003; Siddiqui et al. 2001). In several reviews (Cooke and Hill 2008; Hill 2006), the *MRC1* gene was suggested to be the underlying cause of this linkage signal. While experimental details are not yet publicly available, three exon 7 encoded amino acid polymorphisms were reportedly associated with paucibacillary leprosy in South Indian patients. Based on these reports, we decided to study the role of *MRC1* in susceptibility to leprosy and its clinical sub-forms. In a large sample of Vietnamese patients, we detected weak but significant evidence that serine at amino acid position 396 was a protective factor for leprosy per se and multibacillary leprosy. These results strongly argued that *MRC1* is indeed a leprosy susceptibility gene but not the paucibacillary susceptibility gene detected in the genome-wide scan in Vietnamese families

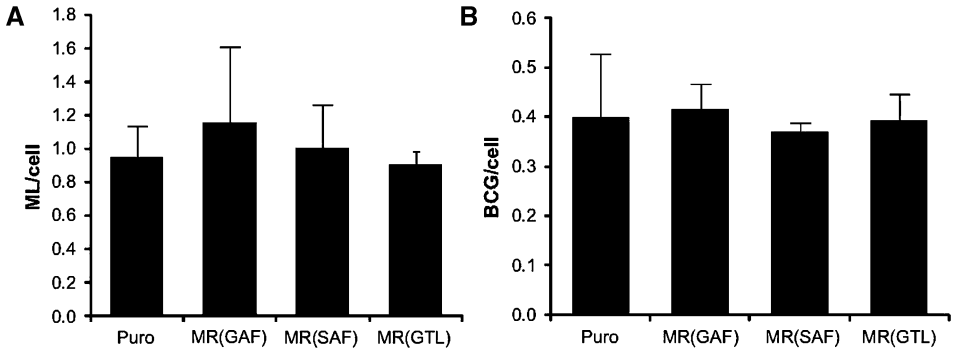


Fig. 5 Absence of impact of ectopic MR expression on binding to mycobacteria. HEK293 transductants were incubated for 16 h with PKH67-stained fluorescent *Mycobacterium leprae* (**a**) and GFP-expressing *Mycobacterium bovis* BCG strain Pasteur (**b**) at a MOI of 5. Measurement of bacterial binding was performed by microscopic examination, counting at least 500 cells in four different fields. Results per field were averaged per cell and are given on the left of the graph. Results are given for HEK293 control cells (PURO) and HEK293 cells over-expressing the MR(GAF), MR(SAF) and MR(GTL) expression constructs

(Mira et al. 2003). The results obtained in the Brazilian sample were more complex. In the Brazilian population, we detected three non-synonymous exon 7 SNPs. As suggested for the South Indian patients, an exon 7 encoded S396-A399-F407 haplotype was significantly protective for leprosy. Similar to the results in the Vietnamese families, the protective effect of this haplotype was significant in the overall sample and in the subset of multibacillary patients. Importantly, in both the Vietnamese and Brazilian samples, the effect of the exon 7 polymorphisms is on leprosy per se. Although in subset analysis formal significance of association was only obtained for multibacillary leprosy, there was no significant heterogeneity between paucibacillary and multibacillary cases. The limited effect of *MRC1* detected by our association study is consistent with the absence of linkage (i.e., a major gene effect) detected between the 10p13 region and leprosy per se. These results strongly argue that *MRC1* is indeed a leprosy susceptibility gene but does not explain the observed linkage peak in the South Indian patients (Siddiqui et al. 2001) or Vietnamese families (Mira et al. 2003). A high-density association scan of the underlying interval is necessary to identify the paucibacillary risk factor(s).

Although our data pinpoint G396S (rs1926736) as the causal variant, we cannot exclude the possibility that multiple rare coding variants in exons 1–30 contribute to leprosy susceptibility (i.e., allelic heterogeneity). However, the practicality and value of sequencing 30 exons in a large sample of individuals in order to identify these variants is uncertain. Unless a sufficient number of rare variants are identified, a 'meta-SNP' analysis combining the information at each loci would not have sufficient power to detect a genetic effect in the expected range of odds-ratios. Supporting G396S as the causal variant, a recent genetic study found rs1926736 to be associated with asthma risk in 446 Japanese cases ($P = 0.011$, OR = 0.61 [0.41–0.89]) (Hattori et al. 2009). Moreover, the most significant haplotype association was observed for a five-SNP risk haplotype carrying allele G396 ($P = 0.00047$). The authors concluded that the haplotype association was seemingly driven by rs1926736 in the Japanese population. This study not only substantiates our results but suggests that *MRC1*—particularly G396S—participates in the immune response to mycobacterial infection and the inflammatory process, two highly related immunological pathways.

The results obtained from the Brazilian sample help to better understand the mechanism of susceptibility mediated by the glycine residue at position 396. Closer inspection of the exon 7 haplotype frequencies in Brazilian patients revealed that the G396-A399-F407 haplotype is enriched in cases while the S396-A399-F407 haplotype is enriched in controls (Table 2). In contrast, G396 in the context of the G396-T399-L407 haplotype appears neutral and may even be slightly protective (Table 2). Hence, G396 was a risk factor only on the G396-A399-F407 haplotype background. The question is then why G396 was a susceptibility factor on the A399-F407 background, while on the T399-L407 background it was not? Exon 7 polymorphisms in *MRC1* map to the second C-type lectin domain (CTLD2) of the MR protein. CTLD2 has been structurally assigned to a hinge domain linking the C-type lectin region to the fibronectin type II and cysteine-rich domains (Boskovic et al. 2006). While CTLD2 is homologous to C-type lectins, it does not bind to ligands (Taylor et al. 1992). Of course, it is possible that changes at CTLD2 indirectly impact on MR ligand affinities. However, our results obtained in the 293-MR cells suggest a direct interaction of CTLD2 with an accessory receptor molecule in the MR–*M. leprae* interaction. In this view, interaction of the CTLD2 domain with the accessory molecule would only be sensitive to G396 in the context of the A399-F407 haplotype.

The importance of accessory molecules for MR function was highlighted by the results of our studies in a cellular model of MR function. To assay MR function, we used the non-professional phagocytic human cell line HEK293 as a platform for expression of MR variants. This cellular model mimicked critical features of MR as seen in HDMECs, previously shown to express this receptor (Groger et al. 2000). Although HEK293 cells are not professional phagocytes, expression of MR bestowed on these cells the ability to ingest zymosan particles. We then showed that polymorphisms in the *MRC1* associated with leprosy risk did not impact on the core MR protein functions as measured by internalization of soluble or particulate ligands. All genetic variants analyzed showed no deviation from their expected endocytic or phagocytic activity. This implies that the amino acid polymorphisms encoded by exon 7 of *MRC1* have no impact on the ability to directly bind and internalize mannose-containing ligands. Since mycobacterial ManLAM is known to interact with MR through its terminal mannooligosaccharide in a similar fashion to other ligands (Schlesinger et al. 1994), we expect that the binding of this bacterial product would not be affected by the polymorphism under study. Furthermore, 293-MR cells failed to bind and ingest viable BCG and *M. leprae* indicating that MR alone is not sufficient to mediate mycobacterial phagocytosis. These results are consistent with the hypothesis that the interaction of MR with *M. leprae* involves additional molecules and that differential activity of the MR alleles can only be revealed in the context of this partner host molecule. Such an accessory molecule may be present in macrophages, which are considerably more efficient at internalizing zymosans than 293-MR cells despite a much lower expression of the receptor (data not shown). Unfortunately, macrophages cannot be used easily in a study of MR variants because of their endogenous expression of *MRC1*.

Whether the impact of the leprosy-associated *MRC1* alleles is via the alteration of MR ligand interactions or by changed receptor signaling or trafficking is not known. Hence, it is possible that aspects of MR function not directly related to ligand binding and internalization are affected by genetic CTLD2 variations. For instance, MR delivers an intracellular signal that affects the immune status of cells as well as endosome–lysosome fusion (Kang et al. 2005; Nigou et al. 2001; Shimada et al. 2006). However, this signaling is mediated by the cytoplasmic domain of MR and would not be expected to be affected by a remote extracellular site. Our data strongly suggest that exon 7 variants have an indirect effect on overall structure and/or stability of the CTLD2 domain, likely by modulating the interaction with an additional host molecule(s). This heteromeric complex may then impact on *M. leprae* phagocytosis, MR trafficking and/or MR signaling. Clearly, if our hypothesis is correct, identification of the MR-interacting host cell molecule(s) is of highest importance.

Acknowledgments We are grateful to all patients and their families who participated in this study. We thank J.L. Casanova, M. Behr and K. Morgan for helpful discussions and A. Montpetit and A. Bélisle for assistance with high-throughput genotyping. Fresh nu/nu mouse derived *M. leprae* were supplied by Dr. R. Truman and Dr. L. Adams of the National Hansen's Disease Program Laboratory Research Branch, Baton Rouge, LA. Funding for this service was provided by the American Leprosy Missions. We thank Dr. J.J. He (Indiana University, IN) for providing the *MRC1* cDNA. This study was supported by a grant from the Canadian Institutes of Health Research (CIHR) to E.S. and MAGRALEPRE from l'Ordre de Malte to Al.A. The Laboratory of Human Genetics of Infectious Diseases is supported by grants from The Rockefeller University Center for Clinical and Translational Science grant number 5UL1RR024143-03 and The Rockefeller University. An.A. holds a graduate studentship from the Natural Science and Engineering Research Council of Canada (NSERC). Al.A. and L.A. are supported by the Assistance Publique-Hôpitaux de Paris, Programme de Recherche Fondamentale en Microbiologie Maladies Infectieuses et Parasitaires (PRFMMIP), and the Agence Nationale de la Recherche (ANR) of the Ministère Français de l'Éducation Nationale de la Recherche et de la Technologie. E.S. is a Chercheur National du Fonds de la Recherche en Santé du Québec and an International Research Scholar of the Howard Hughes Medical Institute.

References

Alcais A, Alter A, Antoni G, Orlova M, Van Thuc N, Singh M, Vanderborght PR, Katoch K, Mira MT, Thai VH, Huong NT, Ba NN, Moraes M, Mehra N, Schurr E, Abel L (2007) Stepwise replication identifies a low-producing lymphotoxin-[alpha] allele as a major risk factor for early-onset leprosy. Nat Genet 39:517–522

Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21:263–265

Bell P, Chaturvedi S, Gelfand C, Huang C, Kochersperger M, Kopla R, Modica F, Pohl M, Varde S, Zhao R, Zhao X, Boyce-Jacino M, Yassen A (2002) SNPstream UHT: ultra-high throughput SNP genotyping for pharmacogenomics and drug discovery. Biotechniques (Suppl 70–2, 74, 76–7)

Boskovic J, Arnold JN, Stilion R, Gordon S, Sim RB, Rivera-Calzada A, Wienke D, Isacke CM, Martinez-Pomares L, Llorca O (2006) Structural model for the mannose receptor family uncovered by electron microscopy of Endo180 and the mannose receptor. J Biol Chem 281:8780–8787

Chakravartti M, Vogel F (1973) A twin study on leprosy. In: Becker PE, Lenz W, Vogel F, Wendt GG (eds) Topics in human genetics, vol 1. Georg Thieme, Stuttgart, pp 1–123

Cooke G, Hill A (2008) Tuberculosis, leprosy and other mycobacterial diseases. In: Kaslow RA, McNicholl JM, Hill AVS (eds) Genetic susceptibility to infectious diseases. Oxford University Press, New York, pp 333–351

Ezekowitz RA, Sastry K, Bailly P, Warner A (1990) Molecular characterization of the human macrophage mannose receptor: demonstration of multiple carbohydrate recognition-like domains and phagocytosis of yeasts in Cos-1 cells. J Exp Med 172:1785–1794

Fan J, Oliphant A, Shen R, Kermani B, Garcia F, Gunderson K, Hansen M, Steemers F, Butler S, Deloukas P, Galver L, Hunt S, McBride C, Bibikova M, Rubano T, Chen J, Wickham E, Doucet D, Chang W, Campbell D, Zhang B, Kruglyak S, Bentley D, Haas J, Rigault P, Zhou L, Stuelpnagel J, Chee M (2003) Highly parallel SNP genotyping. Cold Spring Harb Symp Quant Biol 68:69–78

Griffin TJ, Smith LM (2000) Single-nucleotide polymorphism analysis by MALDI-TOF mass spectrometry. Trends Biotechnol 18:77–84

Groger M, Holnthoner W, Maurer D, Lechleitner S, Wolff K, Mayr BB, Lubitz W, Petzelbauer P (2000) Dermal microvascular endothelial cells express the 180-kDa macrophage mannose receptor in situ and in vitro. J Immunol 165:5428–5434

Hattori T, Konno S, Hizawa N, Isada A, Takahashi A, Shimizu K, Shimizu K, Gao P, Beaty TH, Barnes KC, Huang SK, Nishimura M (2009) Genetic variants in the mannose receptor gene (MRC1) are associated with asthma in two independent populations. Immunogenetics. doi:10.1007/s00251-009-0403-x

Hill AV (2006) Aspects of genetic susceptibility to human infectious diseases. Annu Rev Genet 40:469–486

Horvath S, Xu X, Laird NM (2001) The family based association test method: strategies for studying general genotype–phenotype associations. Eur J Hum Genet 9:301–306

Kabha K, Nissimov L, Athamna A, Keisari Y, Parolis H, Parolis LA, Grue RM, Schlepper-Schafer J, Ezekowitz AR, Ohman DE (1995) Relationships among capsular structure, phagocytosis, and mouse virulence in Klebsiella pneumoniae. Infect Immun 63:847–852

Kang PB, Azad AK, Torrelles JB, Kaufman TM, Beharka A, Tibesar E, DesJardin LE, Schlesinger LS (2005) The human macrophage mannose receptor directs *Mycobacterium tuberculosis* lipoarabinomannan-mediated phagosome biogenesis. J Exp Med 202:987–999

Lake SL, Blacker D, Laird NM (2000) Family-based tests of association in the presence of linkage. Am J Hum Genet 67:1515–1525

Le Cabec V, Emorine LJ, Toesca I, Cougoule C, Maridonneau-Parini I (2005) The human macrophage mannose receptor is not a professional phagocytic receptor. J Leukoc Biol 77:934–943

Miller EN, Jamieson SE, Joberty C, Fakiola M, Hudson D, Peacock CS, Cordell HJ, Shaw MA, Lins-Lainson Z, Shaw JJ, Ramos F, Silveira F, Blackwell JM (2004) Genome-wide scans for leprosy and tuberculosis susceptibility genes in Brazilians. Genes Immun 5:63–67

Mira MT, Alcais A, Van Thuc N, Thai VH, Huong NT, Ba NN, Verner A, Hudson TJ, Abel L, Schurr E (2003) Chromosome 6q25 is linked to susceptibility to leprosy in a Vietnamese population. Nat Genet 33:412–415

Mira MT, Alcais A, Van Thuc N, Moraes MO, Di Flumeri C, Hong Thai V, Chi Phuong M, Thu Huong N, Ngoc Ba N, Xuan Khoa P, Sarno EN, Alter A, Montpetit A, Moraes ME, Moraes JR, Dore C, Gallant CJ, Lepage P, Verner A, van de Vosse E, Hudson TJ, Abel L, Schurr E (2004) Susceptibility to leprosy is associated with PARK2 and PACRG. Nature 427:636–640

Nigou J, Zelle-Rieser C, Gilleron M, Thurnher M, Puzo G (2001) Mannosylated lipoarabinomannans inhibit IL-12 production by human dendritic cells: evidence for a negative signal delivered through the mannose receptor. J Immunol 166:7477–7485

Ridley DS, Jopling WH (1966) Classification of leprosy according to immunity. A five-group system. Int J Lepr Other Mycobact Dis 34:255–273

Schlesinger LS (1993) Macrophage phagocytosis of virulent but not attenuated strains of *Mycobacterium tuberculosis* is mediated by mannose receptors in addition to complement receptors. J Immunol 150:2920–2930

Schlesinger LS, Hull SR, Kaufman TM (1994) Binding of the terminal mannosyl units of lipoarabinomannan from a virulent strain of *Mycobacterium tuberculosis* to human macrophages. J Immunol 152:4070–4079

Schlesinger LS, Kaufman TM, Iyer S, Hull SR, Marchiando LK (1996) Differences in mannose receptor-mediated uptake of lipoarabinomannan from virulent and attenuated strains of *Mycobacterium tuberculosis* by human macrophages. J Immunol 157:4568–4575

Shimada K, Takimoto H, Yano I, Kumazawa Y (2006) Involvement of mannose receptor in glycopeptidolipid-mediated inhibition of phagosome-lysosome fusion. Microbiol Immunol 50:243–251

Siddiqui MR, Meisner S, Tosh K, Balakrishnan K, Ghei S, Fisher SE, Golding M, Shanker Narayan NP, Sitaraman T, Sengupta U, Pitchappan R, Hill AVS (2001) A major susceptibility locus for leprosy in India maps to chromosome 10p13. Nat Genet 27:439–441

Taylor ME, Bezouska K, Drickamer K (1992) Contribution to ligand binding by multiple carbohydrate-recognition domains in the macrophage mannose receptor. J Biol Chem 267:1719–1726

Tosh K, Meisner S, Siddiqui MÂR, Balakrishnan K, Ghei S, Golding M, Sengupta U, Pitchappan Ramasamy ÂM, Hill Adrian ÂVÂS (2002) A region of chromosome 20 is linked to leprosy susceptibility in a south Indian population. J Infect Dis 186:1190–1193

Tregouet DA, Garelle V (2007) A new JAVA interface implementation of THESIAS: testing haplotype effects in association studies. Bioinformatics 23:1038–1039

Wallace C, Fitness J, Hennig B, Sichali L, Mwaungulu L, Ponnighaus JM, Warndorff DK, Clayton D, Fine PEM, Hill AVS (2004) Linkage analysis of susceptibility to leprosy type using an IBD regression method. Genes Immun 5:221–225

Hum Genet (2008) 123:227–235
DOI 10.1007/s00439-008-0474-z

REVIEW ARTICLE

Leprosy as a genetic model for susceptibility to common infectious diseases

Andrea Alter · Alexandre Alcaïs · Laurent Abel · Erwin Schurr

Received: 7 December 2007 / Accepted: 21 January 2008 / Published online: 5 February 2008

Abstract Leprosy (Hansen's disease) is a human infectious disease that can be effectively treated with long-term administration of multi-drug therapy. In 2006, over 250,000 new cases were reported to the World Health Organization. In the nineteenth century, disagreement among leprologists regarding the hereditary or infectious nature of leprosy was resolved with the identification of the etiological agent, *Mycobacterium leprae*. However, epidemiological studies maintain the importance of host genetics in leprosy susceptibility. A model free genome-wide linkage scan in multicase families from Vietnam led to the positional cloning of global genetic risk factors in the *PARK2*/*PACRG* and *LTA* genes. The process of identifying the susceptibility variants provided invaluable insight into the replication of genetic effects, particularly the importance of considering population-specific linkage-disequilibrium structure. As such, these studies serve to improve our understanding of leprosy pathogenesis by implicating novel biological pathways while simultaneously providing a genetic model for common infectious diseases.

> '...your opinions about leprosy are completely wrong. You believe that the disease is hereditary but not infectious. The truth is that it is infectious but not hereditary.' (Marmor in Surv Ophthalmol 47:275–287, 2002)
>
> Gerhard Henrik Armauer Hansen (1841–1912)
>
> 'I want to say that, if I formerly may have doubted the theory about the heredity of the disease, I have now no longer any doubts about it.' (Pandya in Int J Lepr Other Mycobact Dis 66:374–384, 1998)
>
> Carl Wilhelm Boeck (1808–1875)

A. Alter · E. Schurr (✉)
McGill Centre for the Study of Host Resistance,
The Research Institute of the McGill University Health Centre,
1650 Cedar Avenue, Montreal H3G 1A4, QC, Canada
e-mail: erwin.schurr@mcgill.ca

A. Alter · E. Schurr
McGill Department of Medicine, McGill University,
McIntyre Medical Building, 3655 Promenade Sir William Osler,
Montreal H3G 1Y6, QC, Canada

A. Alcaïs · L. Abel
Laboratoire de Génétique Humaine des Maladies Infectieuses,
Institut National de la Santé et de la Recherche Médicale U550,
156 Rue de Vaugirard, 75015 Paris, France

A. Alcaïs · L. Abel (✉)
Faculté Médecine Necker, Université Paris René Descartes,
156 Rue de Vaugirard, 75730 Paris, France
e-mail: abel@necker.fr

E. Schurr
McGill Departments of Human Genetics and Biochemistry,
Stewart Biology Building, Room N5-13,
1205 Dr Penfield Avenue, Montreal H3A 1B1, QC, Canada

Brief introduction to leprosy

Leprosy is an infectious disease caused by *Mycobacterium leprae* (*M. leprae*) and can consequently be treated with antibiotics. Freely available long-term multi-drug therapy that combines rifampicin, clofazimine and dapsone effectively targets the bacteria while minimizing the development of drug-resistant strains. The World Health Organization (WHO) currently recommends a 6- and 12-month treatment regimen for paucibacillary leprosy and multibacillary leprosy, respectively. Humans are the only known hosts relevant to the transmission of leprosy; however, there is insufficient research to discount zoonotic transmission from

naturally infected armadillos and non-human primates. The 1991 WHO initiative to eliminate leprosy as a public health problem dramatically reduced the global disease prevalence but had only a modest impact on leprosy incidence, suggesting a persisting unknown reservoir. In 2006, 259,017 new cases were reported, 54% of which occurred in India—1 of 118 countries that has achieved elimination status (defined as prevalence < 1 case/10,000 population). Brazil, Democratic Republic of the Congo, Mozambique and Nepal have not achieved elimination and account for 23% of new cases (WHO 2007).

Individual differences in the host immune response directed against the bacilli largely account for the spectrum of clinical phenotypes delimited by the tuberculoid and lepromatous forms (Ridley–Jopling classification) (Ridley and Jopling 1966). Tuberculoid cases present a limited number of hypopigmented and anaesthetic skin lesions with no microscopically discernable bacteria. The correlated Th1-cell-mediated immune response (IFNG, IL2) promotes the formation of delineated granulomas that successfully limit bacterial replication. Conversely, lepromatous cases present numerous sensitive or anaesthetic skin lesions with high-bacillary loads. The correlated Th2-antibody response (IL4, IL10) impedes granuloma formation allowing for uncontrolled bacterial replication and continuous infiltration of the skin and nerves. Borderline forms—borderline-tuberculoid, borderline and borderline-lepromatous—comprise the majority of cases. These individuals present intermediate clinical and histological phenotypes resulting from immunologically unstable responses. A simpler dichotomized classification based on the number of skin lesions includes paucibacillary ($\leq$5) and multibacillary ($\geq$6 lesions) forms (WHO classification). With limited accuracy, the Ridley–Jopling and WHO schemes can be reconciled by re-classifying tuberculoid and borderline-tuberculoid forms as paucibacillary, and borderline, borderline-lepromatous and lepromatous forms as multibacillary (Britton and Lockwood 2004).

Historical perspective—cause versus risk

An infectious origin for leprosy was first suggested by the Norwegian physician Gerhard Henrik Armauer Hansen. A young and inexperienced Hansen boldly challenged the theory held by his superiors, Daniel Cornelius Danielssen, Hansen's father-in-law, and Carl Wilhelm Boeck, that leprosy was a 'hereditary dyscrasia'. Hansen's seminal discovery in 1873 of bacilli in leprous nodules substantiated his idea of a bacterial etiology for leprosy, but he struggled throughout his career to prove causality. One attempt to demonstrate disease transmissibility ended with a criminal charge. On November 3, 1879 Hansen inoculated the eye of an unwilling female subject with material from a leprous nodule. No disease ensued but Hansen was convicted of misusing his official position to his own advantage and relieved of his physician duties. Hansen retained his position as Medical Officer of Health for Leprosy in Norway and wrote the 'Norwegian Leprosy Act' (first law passed in 1877) mandating patient isolation based on his yet unproven theory of leprosy transmission (Marmor 2002).

An intriguing question is to what extent did Hansen's disregard for the role of heredity in the pathogenesis of leprosy impede his efforts to prove causality? Hansen's understanding that *M. leprae* is unequivocally required for disease was correct, but is it sufficient? We now know that the majority of individuals (>90%) exposed to *M. leprae* are asymptomatic (Chaudhury et al. 1994; Convit et al. 1992; Gupte et al. 1998), suggesting that progression from 'exposure' to 'disease' requires additional risk factors. Hansen's subjects might have been successfully inoculated with *M. leprae* but lacked the additional factors—possibly genetic risk factors—to permit the development of leprosy.

Leprosy, like other infectious diseases, presents a situation where the causal factor (i.e. the infectious agent) is necessary but often insufficient for clinical manifestation of the disease. Disease appearance requires the co-occurrence of risk factors—environmental and/or genetic—that may be uniquely sufficient but not necessary. For example, herpes simplex virus-1 (HSV-1) encephalitis (HSE) is a rare complication of infection with HSV-1, a ubiquitous virus infecting eighty percent of young adults. HSV-1 is the cause of HSE but given its infrequent occurrence, additional predisposing risk factors must exist. In four unrelated patients, HSE was attributed to either a homozygous mutation in the coding region of *UNC93B1* or a heterozygous mutation in the coding region of *TLR3* (Casrouge et al. 2006; Zhang et al. 2007). Importantly, *UNC93B1* and *TLR3* mutations do not explain all cases of HSE (i.e. uniquely sufficient but not necessary). Additional risk factors have yet to be identified.

Genetic risk factors for leprosy

A century after the discovery of *M. leprae*—and long after the deaths of Hansen, Danielssen and Boeck—data emerged to reconcile their scientific differences. Charles C. Shepard was able to reproducibly induce granulomas containing acid-fast bacilli in the foot-pads of mice after injection with bacteria from the nasal passages or skin lesions of human leprosy cases. Furthermore, the same histopathology was observed after injection of passage material from the foot-pads demonstrating transmissibility of the granulomatous phenotype—a partial fulfillment of Koch's postulates (Shepard 1960). Soon after, Chakravartti and Vogel

showed that the concordance of leprosy and the resulting clinical forms was higher among genetically identical monozygotic twins (60–85%) as compared to dizygotic twins (5–20%) (Chakravartti and Vogel 1973). Additional studies demonstrated familial clustering of the disease (Shields et al. 1987) and segregation analyses supported a polygenic model of inheritance that includes major susceptibility genes (Abel and Demenais 1988; Abel et al. 1995). Collectively, these landmark publications stressed the indispensability of both the pathogen and host genetics to leprosy pathogenesis.

Ongoing efforts to identify genetic risk factors for leprosy include candidate gene association studies and genome-wide linkage analysis in samples from endemic countries. Possibly facilitating these efforts is the recent report that differential susceptibility is not confounded by varying strains of *M. leprae*. Among seven geographically diverse samples of the bacterium only five informative single-nucleotide polymorphisms (SNPs) were found in 142 kilo-base (kb) of genomic sequence—an unusually low SNP frequency (1 SNP per 28 kb). Moreover, only four SNP combinations were observed, each with a limited global distribution (e.g. SNP-type 1 is restricted to Asia, the Pacific region and East Africa) (Monot et al. 2005).

Candidate gene association studies

The candidate gene approach tests genetic variants (e.g. SNPs, copy-number variants) for association with a phenotype. Variants are selected based on a biological understanding of the trait (e.g. *INS* variants in Type 2 diabetes), or the results of related experimental models (e.g. *NRAMP1* in tuberculosis). In a family-based design, an associated variant is significantly over- or under-transmitted from heterozygous parents to affected progeny (transmission-disequilibrium test). In a population-based sample, the frequency or genotypic distribution of an associated variant differs significantly between cases and ethnically matched controls. Unlike population-based samples, family based samples are protected from spurious associations due to population stratification. Given an adequate family- or population-based sample size, association studies can identify even minor genetic effects that are beyond the limit of detection for model-free linkage studies. Associated variants are confirmed by replication in independent samples and ultimately by functional experiments. Numerous candidate gene studies for leprosy and related phenotypes (e.g. clinical form, Mitsuda reaction etc.) have been published and are reviewed elsewhere (Alcais et al. 2005; Fernando and Britton 2006; Fitness et al. 2002; Mira 2006). The most consistent association with leprosy has been with *HLA-DRB1* alleles (Shaw et al. 2001; Vanderborght et al. 2007; Zerva et al. 1996).

Genome-wide linkage analysis

Model-free genome-wide linkage analysis is an unbiased method to identify chromosomal regions with an underlying trait-predisposing locus. Evenly spaced microsatellite markers (highly variable repetitive DNA sequences) or common SNPs are genotyped in multi-case families. The trait locus and flanking regions will be shared by affected siblings, reflected by increased parental allele sharing (sib-pair analysis). Genome-wide linkage analysis detects major genetic effects. However, linked regions are large—on the order of several mega-basepairs (Mbp)—and often contain numerous candidate genes.

The first leprosy study employing this approach reported linkage of chromosome region 10p13 (D10S1661 multipoint MLS = 4.09) to paucibacillary leprosy in 224 multi-case families from Southern India (Siddiqui et al. 2001). In a follow-up study of the regions that showed marginal evidence for linkage, chromosome region 20p12 (D20S115) was modestly linked in families from Tamil Nadu (multipoint MLS = 3.16) but not Andhra Pradesh (multipoint MLS = 0) (Tosh et al. 2002). An unrelated genome-wide scan in 71 multi-case families from Brazil found chromosome region 6p21 (HLA-DQA lod = 3.23) to be weakly linked to leprosy while regions 17q22 (D17S1868 lod = 2.38) and 20p13 (D20S889 lod = 1.51) displayed some suggestive evidence for linkage. Unlike in the Tamil Nadu sample, the peak at 20p13 in the Belem City sample related almost entirely to the lepromatous and borderline lepromatous families (lod = 1.36) (Miller et al. 2004).

A genome-wide scan completed by our group in 86 Vietnamese multi-case families (205 affected offspring) with an unbiased distribution of paucibacillary (44%) and multibacillary (56%) cases, detected linkage to chromosome region 6q25–q26 (multipoint MLB lod = 4.31). In addition, evidence for linkage between 6p21 microsatellite markers (multipoint MLB lod = 2.62) and leprosy was observed supporting an independent susceptibility locus in the HLA region. Suggestive evidence for linkage of 20p12 (lod = 1.13) was noted irrespective of clinical form. The reported peak at 10p13 in the South Indian study was replicated by testing for linkage in all families with paucibacillary affected sib-pairs (lod = 1.98) (Mira et al. 2003). Collectively, these studies pointed to chromosome regions 6q25–q26, 6p21, 10p13 and 20p12–p13, as primary targets for further investigation.

An update on *PARK2/PACRG*

A 6.4-Mb interval spanning chromosome region 6q25–q26 (43 genes, *ZDHHC14* → *PACRG*) was selected for a low-density association scan (64 SNPs) in 197 single-case

Vietnamese leprosy families. Four of the six SNPs associated with leprosy ($P < 0.05$) localized to the putative *PARK2* promoter overlapping the 5′-region of the adjacent *PACRG* gene. In a subsequent high-density association scan (81 polymorphisms) of both genes, 19 SNPs between *PARK2* intron 1 and *PACRG* intron 2 showed evidence for association ($0.03 \geq P \geq 0.0006$). Accounting for the correlation between SNPs, multivariate conditional logistic regression identified two SNPs (*PARK2*_e01(-2599) and rs1040079) that were sufficient to capture all association information. The odds ratio of leprosy for individuals carrying at least one haplotype with both risk alleles (*PARK2*_e01(-2599) -T and rs1040079-C) was 5.28 (95% CI [2.06–13.55]). Among 13 SNPs tested in an independent sample of 587 cases and 388 controls from Brazil, nine were associated (consistent risk allele) including *PARK2*_e01(-2599) ($P_{corrected} = 0.003$) and rs1040079 ($P_{corrected} = 0.001$) (Mira et al. 2004).

In addition to implicating ubiquitin- and proteasome-dependent processes in the progression of infection with *M. leprae* (Alcais et al. 2005; Schurr et al. 2006), this study provided a much anticipated 'proof of principle' for the genetic analysis of common infectious diseases. This was the first successful attempt to positionally clone genetic risk factors for a common infectious disease, thereby outlining a sound methodology using model-free linkage analysis followed by low- and high-density SNP association scans and successive replication.

In 2005, Bamezai and colleagues published an association study of six *PARK2*/*PACRG* SNPs with leprosy in Northern India. Comparing genotypic frequencies between 286 cases (144 multibacillary and 142 paucibacillary) and 350 controls, the TT genotype for *PARK2*_e01(-2599) was overrepresented in the cases ($P = 0.04$), but insufficiently to withstand an overly conservative Bonferoni correction for multiple testing ($P_{corrected} = 0.6$) (Malhotra et al. 2005). Clearly, correction in the context of a replication study is an unresolved issue. Significant evidence for association of the same SNP (*PARK2*_e01(-2599)) and risk allele (T-allele) in an ethnically diverse sample should not be underestimated.

Like *M. leprae*, *Salmonella typhi* (typhoid fever) and *Salmonella paratyphi* (paratyphoid fever) are intracellular pathogens. Also, proteasome-mediated degradation of the *Salmonella* protein SopE is a host defence mechanism implicating the proteasome pathway—and possibly PARK2 and PACRG—in enteric fever pathogenesis. As such, in 2006, Ali et al. tested the association of four *PARK2*/*PACRG* SNPs with enteric fever in Indonesia. Comparing allelic frequencies between 115 cases (86 typhoid and 26 paratyphoid) and 259 controls with no known history of enteric fever, the T-allele of *PARK2*_e01(-2599) again showed a weak but significant trend for association with disease ($P = 0.02$, OR = 1.58, 95% CI [1.02–2.36]) (Ali et al. 2006). Remarkably, this study implicated the same allele in an unrelated phenotype suggesting a broad involvement of PARK2 in intracellular infections.

Undoubtedly, the data warrant the functional assessment of *PARK2* variants in the host response to intracellular bacteria, but which variants? All three studies attribute an increased risk of infection to the T-allele of *PARK2*_e01 (-2599), but this SNP is not necessarily the 'functional' risk factor that directly impacts the biological process critical to disease development. As a result of linkage-disequilibrium, *PARK2*_e01(-2599) is simply a 'tag SNP' representing a non-contiguous group of highly correlated SNPs, referred to as a 'bin'. Each SNP in the bin shows comparable evidence for association with leprosy and enteric fever. The most direct approach to determine which SNP is the functional variant requires assessing each SNP in the bin for a biological effect in the context of infection—an enormous undertaking.

An alternative method is to exploit the fact that the pairwise correlation (measure of linkage-disequilibrium expressed as r^2) between SNPs varies across ethnic populations thereby changing the physical structure of the bin. Only SNPs that remain correlated with the functional risk factor will show consistent evidence for association across ethnic groups. By comparing population-specific bin structures, the number of possible functional SNPs is limited to those that are common among each associated bin (Fig. 1). Unfortunately, not enough SNPs in the *PARK2*_e01(-2599) bin were genotyped in the Indian and Indonesian samples to undertake this 'reductive' approach. High-density linkage-disequilibrium mapping of *PARK2* in ethnically diverse samples should be productive in identifying strong functional candidates for further evaluation. A leprosy sample of African origin would be particularly insightful given that the extent of linkage-disequilibrium is generally reduced in older populations owing to increased recombination. The consequential reduction in bin size can facilitate high-resolution positional cloning by further limiting the number of functional candidates.

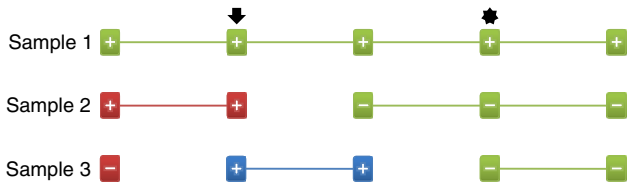


Fig. 1 Reconstructing population-specific bins to identify the functional variant. SNPs identical in color belong to the same bin. The original associated SNP is marked with an asterisk. Positive or absent evidence for association is denoted by (+) or (−), respectively. The original SNP is not associated in sample 2 or sample 3. Testing all SNPs from the original bin reveals only SNP 2 (*arrow*) is associated in all three samples and is therefore the most likely candidate for the functional variant

Lymphotoxin-alpha (LTA)

LTA as a risk factor for early-onset leprosy

The strategy used to study chromosome region 6q25–q26 was adapted to a 10.4-Mb region including the HLA class III and class II genes on chromosome region 6p21 (224 genes, *BAT1* → *CCND3*). A low-density primary association scan (307 SNPs) in 194 single-case leprosy families identified *LTA*-293 in the *LTA* promoter as the most significant independently associated SNP ($P = 0.0009$). In a high-density association scan (29 SNPs) of a 90 kb interval around *LTA* (*PPIAP9* → *NCR3*) eight SNP alleles were associated with leprosy ($P < 0.05$). Seven of these SNPs comprised a single bin (pair-wise r^2 with tag SNP > 0.8) represented by *LTA*-293 (tag SNP). Although statistically each bin-SNP was an equally probable candidate, we hypothesized that *LTA* + 80 ($P = 0.007$) was the functional determinant based on previously published data attributing a biological role to this variant (Knight et al. 2004). Association of the *LTA* + 80-containing bin with leprosy was confirmed in a second sample of 104 single-case families from Vietnam ($P = 0.003$; combined OR = 2.11, 95% CI [1.48–3.01], combined $P = 0.000024$) (Alcais et al. 2007). Interestingly, the effect of *LTA* + 80 was shown to be independent and additive to the associated *PARK2/PACRG* promoter variants.

To replicate the association of *LTA* + 80 with leprosy in an ethnically diverse population, six *LTA* SNPs, including three from the originally associated bin (*LTA*-293, *LTA* + 80 and *LTA* + 368), were tested in a third sample of 364 leprosy cases and 371 controls from Northern India. Unexpectedly, only *LTA*-294—a SNP one base pair upstream of *LTA*-293—was associated in this replication sample ($P = 0.0004$). However, after multivariate analysis to adjust for *LTA*-294, *LTA* + 80 became significant ($P = 0.026$) but *LTA*-293 ($P = 0.18$) did not, highlighting the interest of such multivariate analysis when considering SNPs in incomplete LD. Why was *LTA*-293 not associated with leprosy in this new sample? Reconstructing the population-specific bins revealed that *LTA*-293 was not correlated with *LTA* + 80 in the Indian sample ($r^2 = 0.45$) and was therefore excluded from the associated bin, precluding its candidacy as the functional variant and corroborating our hypothesis regarding *LTA* + 80 (Fig. 2).

In a fourth sample of 209 leprosy cases and 192 controls from Brazil, none of the six *LTA* SNPs tested showed evidence for association. Given that the median age at leprosy diagnosis was highest in the Brazilian sample, we hypothesized that *LTA* + 80 was a risk factor for early-onset leprosy. To test this hypothesis, cases in each sample were stratified into four groups based on age at diagnosis (0–15, 16–25, 26–35 and > 35 years). Indeed, in the combined

Fig. 2 Identifying LTA + 80 as the functional variant. SNPs identical in color belong to the same bin. The original associated SNP (*LTA*-293) is marked with an asterisk. Positive or absent evidence for association is denoted by (+) or (−), respectively. *LTA*-293 was not associated in the Indian sample. Testing all SNPs from the original bin revealed *LTA* + 80 was associated in both samples and is therefore the most likely candidate for the functional variant (supported by experimental evidence). The association of *LTA*-294 in the Indian—but not in the Vietnamese sample—can be explained by a yet unidentified functional variant in the same bin as *LTA*-294 in the Indian population but not in the Vietnamese population

Vietnamese sample (298 families) the association of *LTA* + 80 could mostly be attributed to the youngest cases (0–15 years of age group combined OR = 5.63, 95% CI [2.54–12.49], combined $P = 0.0000004$). The same age at diagnosis effect was observed in the Indian sample (16–25 years of age group OR = 2.95, 95% CI [1.32–6.58] $P = 0.006$). Furthermore, the *LTA* + 80 risk allele was notably enriched among the youngest Brazilian cases nearly reaching statistical significance (16–25 years of age group $P = 0.07$). Genetic heterogeneity testing confirmed a highly significant differential role for the *LTA* + 80 variant in the pathogenesis of early- versus late-onset leprosy in the Vietnamese and Indian cases (Alcais et al. 2007).

Additional 6p21 risk factors and genetic heterogeneity

The proximity of *LTA* to the classical HLA class I genes (*HLA-A*, *-B*, and *-C*) and *HLA-DRB1* necessitated addressing the possibility that the increased risk of leprosy associated with *LTA* + 80 was due to long-range linkage-disequilibrium with these known immunoregulatory genes. HLA class I genotypes were determined in 37 unrelated Vietnamese individuals and *HLA-DRB1* genotypes in the 194 Vietnamese single-case leprosy families. None of the tested class I and class II alleles in this population was in linkage-disequilibrium with *LTA* + 80 beyond $r^2 = 0.25$, confirming that *LTA* + 80 is an independent leprosy risk factor. The density and proximity of infectious disease susceptibility gene candidates in the 6p21 region compels such extensive analysis to establish the independence of any identified genetic risk factor.

Although *LTA*-293 was the most significant independently associated SNP in the primary association scan, three other independently associated SNPs (rs3128961, rs937662 and rs707928) suggested the existence of additional risk factors underlying the 6p21 linkage peak. To follow-up, a third association scan of the region in the 194 single-case Vietnamese families is in progress. The panel

was designed to maximize the genetic variation captured by tag SNPs selected from SNP genotype data publically available through the International HapMap project (http://www.hapmap.org). In the HLA class III and class II regions, >90% of all common SNPs in the Chinese (CHB) population will have been tagged ($r^2 > 0.8$) enabling a comprehensive analysis of these regions for leprosy risk factors.

The association of *LTA*-294 in the Indian sample ($P = 0.0004$)—but not in the Vietnamese sample—raises an intriguing question: is this another independent common risk factor or genetic heterogeneity between two ethnic populations? The latter is an obvious interpretation of the observation but the former explanation is equally plausible based on our understanding of dynamic bin structures. In this situation, the functional risk factor would be common to both the Vietnamese and Indian samples but would only be correlated with *LTA*-294 in the Indian population, thereby explaining the significant association of this SNP in this population only (Fig. 2). Analogous situations may account for many failed replication studies in the literature that have been attributed to genetic heterogeneity. Unfortunately, the Indian population is not well represented by the HapMap samples of European ancestry (CEU), otherwise it would be straightforward to limit the functional candidates to those correlated with *LTA*-294 in the CEU population but not in the CHB population. As such all SNPs associated in the Vietnamese families need to be tested in the Indian sample for both association with leprosy and correlation with *LTA*-294.

LTA in mycobacterial disease and *LTA* + 80

LTA can complex with LTB dimers (LTA_1LTB_2) to form the membrane-bound LTB molecule that signals through the LTB receptor—a critical interaction that precipitates secondary lymphoid organ formation. An important role in the host defence against intracellular infections is emerging for soluble LTA (LTA_3)—secreted by CD4+ T cells, B cells and NK cells. Normal lymphoid development is perturbed in *Lta* or *Ltb* knockout mice owing to the lack of membrane-bound Lta_1Ltb_2. To overcome the absence of secondary lymphoid organs in *Lta*−/− or *Ltb*−/− mice, *Rag*−/− mice (T and B cell deficient) are irradiated and reconstituted with *Lta*−/− (Lta_3 and Lta_1Ltb_2 deficient) or *Ltb*−/− (Lta_1Ltb_2 deficient) bone marrow. The chimeras have normal secondary lymphoid organs but circulating leukocytes deficient in Lta or Ltb.

Lta−/− and *Ltb*−/− chimeras were infected with virulent *M. tuberculosis*, a second human pathogenic mycobacterium. Despite comparable antigen-specific T-cell and macrophage responses, bacterial replication was less controlled in the *Lta*−/− chimeras and all died within 38 days post-infection (*Ltb*−/− chimeras survived > 150 days). Histological examination of the lungs showed that the lesions contained more necrotizing neutrophils and less macrophages. Lymphocytes also failed to extravasate into infected tissue, accumulating in the perivascular and peribronchial regions (Roach et al. 2001). In a subsequent study using the intracellular bacteria *Lysteria monocytogenes*, *Lta*−/− chimeras died 5–6 days post-infection (*Ltb*−/− chimeras survived > 14 days). Histological examination of the liver similarly showed that lesions contained more neutrophils and less macrophages and lymphocytes (Roach et al. 2005). Based on these observations, soluble Lta mediates the recruitment of macrophages and lymphocytes to sites of infection where they can cooperatively exert their antimicrobial functions.

In human studies, among twelve common polymorphisms (minor allele frequency > 10%) at the *LTA* locus, soluble LTA (LTA-homotrimer) production was significantly less from cells carrying the common *LTA* + 80A-*LTA* + 368C haplotype. A protein–DNA complex, containing activated B-cell factor 1 (ABF-1), was detected when nuclear extracts from mitogen-stimulated cells were incubated specifically with oligonucleotide probes corresponding to the *LTA* + 80A allele. Basic helix-loop-helix transcription factors like ABF-1—a transcriptional repressor expressed in lymphoid tissue—bind E2-box consensus motifs (CAGCTG). *LTA* + 80 is the second nucleotide in a E2-box motif with a mismatch at position five (CAGC**A**G). *LTA* + 80A maintains the 1-bp mismatch in the motif (C**A**GC**A**G) while *LTA* + 80C introduces a 2-bp mismatch (C**C**GC**A**G) abrogating ABF-1 binding. As expected, *LTA* + 80A-containing constructs co-transfected with ABF-1 in fibroblast cells had significantly lower reporter gene expression than *LTA* + 80C-containing constructs (Knight et al. 2004). In the Vietnamese, Indian and Brazilian samples, this low-producing *LTA* + 80 A-allele was consistently associated with an increased risk of leprosy in young patients. Extrapolating from the murine models, transcriptional repression of *LTA* may preclude protective immune cells (macrophages and T-cells) from reaching sites of *M. leprae* infection. Age-dependent aspects of adaptive immunity may help to overcome this defect in immune cell recruitment.

Leprosy: a good disease to study?

The number of replicated susceptibility loci with substantial odds ratios (e.g. *HLA-DRB1*, *LTA*, *PARK2*/*PACRG*) lends the question, is leprosy particularly amenable to genetic studies? Enrolling cases with accurately determined phenotypes is far from trivial but absolutely critical to the success of any genetic study. Unknowingly including 'phenocopies' with diverse underlying genetic causes is detri-

mental to the investigation. The three diagnostic signs of leprosy—the result of more than 100 years of clinical research—have a sensitivity of 97% (Britton and Lockwood 2004). The reliability of leprosy samples ascertained in endemic countries by experienced leprologists is high, thereby increasing the chances of discovering any existing common risk factor(s).

The considerable progress made to date, could also be a reflection of the intricate human–*M. leprae* relationship. The sequencing of the *M. leprae* genome revealed widespread reductive evolution—only 49.5% of the 3.3-Mb sequence encodes functional proteins. More than 2,000 genes—across all functional categories—have been lost since diverging from the last common mycobacterial ancestor (Cole et al. 2001). This extreme case of host adaptation has rendered the bacterium completely dependent on many cellular processes of the host cell, exemplified by its inability to grow in vitro and its narrow host range. The viability of this obligate intracellular parasite should then be highly sensitive to host genetic variations affecting these cellular processes—variations that translate into increased disease susceptibility or resistance. For example, the ability of *M. leprae* to neutralize reactive oxygen species (ROS) has been compromised in part by the loss of catalase–peroxidase (katG inactivation) (Eiglmeier et al. 1997). So how does the bacterium survive the hostile macrophage environment? Interestingly, human dopaminergic neuroblastoma cell lines overexpressing PARKIN produced significantly reduced levels of basal ROS and dopamine-induced ROS (Hyun et al. 2002). If the anti-oxidative role of PARKIN is exploited by *M. leprae* to compensate for its own deficiency, perhaps the *PARK2* susceptibility variant increases host-cell permissibility by enhancing this function.

It is difficult to estimate the individual contribution of genetic and environmental influences on phenotypic variation. A Danish study followed 960 families and calculated the relative risk (RR) of premature death (age 15–58 years) in adoptees when a biological or adoptive parent died before age 50. The RR of death due to an infectious disease was 5.81 (95% CI [2.47–13.7]) if a biological parent died of an infectious disease and 1.19 (95% CI [0.16–8.99]) for cancers. Conversely, the RR of death due to cancers was 5.16 (95% CI [1.20–22.2]) if an adoptive parent died of a cancer and near unity for infectious diseases (Sørensen et al. 1988). Infectious disease susceptibility could therefore be more impervious to environmental factors than other common diseases, relying heavily on the genetic background of the host. Specifically, for persistent diseases such as leprosy and tuberculosis, extensive host–pathogen co-adaptation may have occurred favouring relative strong host genetic effects. This might explain the sizeable odds ratios for leprosy risk factors (*LTA* + 80 OR = 5.63, 95% CI [2.54–12.49]) relative to the most recently identified risk factors for type-2 diabetes (*EXT2* rs11037909 OR = 1.27, 95% CI [0.97–1.57]), Crohn's disease (*IRGM* rs13361189 OR = 1.38, 95% CI [1.15–1.66]), multiple sclerosis (*IL7R* rs6897932 OR = 0.76, 95% CI [0.65–0.90]) and colorectal cancer (*SMAD7* rs4939827 OR = 0.86, 95% CI [0.79–0.92]) (Broderick et al. 2007; Lundmark et al. 2007; Parkes et al. 2007; Sladek et al. 2007).

Concluding statement

Recently, standard operating protocols for human complex trait analysis have been proposed (Zondervan and Cardon 2007). Still new studies serve to refine our methodologies as much as our understanding of the human genome. In this context leprosy has proven to be a valuable model for studying the genetic aspect of common infectious diseases. The unforeseen discovery of risk factors in the 5′-regulatory region of *PARK2* and *PACRG* confirmed the predicted usefulness of linkage analysis followed by association studies—the efficiency of which has been markedly improved by the International HapMap project. It is now reasonable to capture all the genetic variation in a target region with a minimal number of SNPs. Also, the process by which *LTA* + 80 was identified highlighted three important considerations for future genetic replication studies irrespective of the method used to initially identify the variant: (1) it is insufficient to genotype only the original associated variant(s) in the replication sample due to differences in population-specific bin structures. Currently, the transferability of HapMap data is limited and many populations are 'intermediate' to the CEU, CHB, JPT and YRI populations. Generating HapMap-like data for more populations will expedite the process of reconstructing bins; (2) multivariate analysis is necessary to uncover "hidden" associations by adjusting for proximal independent risk factors; (3) sample ascertainment (e.g. family-based versus population-based) can introduce biases (e.g. age at diagnosis) that must be accounted for by appropriate adjustments. Our data clearly show that covariates can have a strong effect on replicating previously found associations.

Determining the genetic basis of leprosy susceptibility is far from complete. Genome-wide linkage scans have identified major susceptibility loci on chromosome regions 6q25–q26, 6p21, 10p13 and 20p12. To date, only those genetic risk factors underlying 6q25–q26 (*PARK2*/*PACRG*) and 6p21 (*HLA-DRB1* and *LTA*) have been discovered. Furthermore, our data suggests the existence of additional genetic risk factors in the HLA region. The continued positional cloning of major genes in ethnically diverse family-based and population samples, coupled with functional studies of the implicated proteins is critical to the understanding of leprosy.

It is fitting that the oldest recorded human disease has permitted such remarkable scientific achievements. In 1873 leprosy was the first human disease to be associated with a bacterium and in 2004 it was the first common infectious disease for which genetic risk factors were positionally cloned. From influencing government policies to modeling complex trait analysis, the value of studying leprosy has proven itself. Despite this progress, disease pathogenesis—particularly disease transmission—continues to elude scientists, precluding the complete eradication of leprosy. However, identifying additional genetic risk factors will implicate new pathways—like the ubiquitin-dependent pathway—in the complex network of interactions that define this ancient host–pathogen relationship. Likewise, our understanding of non-specific mechanisms underlying the host immune response will improve, as will our ability to identify and ultimately address the inherent weaknesses in the human response to infection.

Acknowledgments AnA is supported by the CIHR Strategic Training Centre in the Integrative Biology of Infectious Diseases and Autoimmunity. ES is a *Chercheur National* of the *Fonds de Recherche en Santé du Québec* (FRSQ) and an International Research Scholar of the Howard Hughes Medical Institute. AlA and LA are supported by the Assistance Publique-Hopitaux de Paris, *Programme de Recherche Fondamentale en Microbiologie Maladies Infectieuses et Parasitaires* (PRFMMIP), and the *Agence Nationale de la Recherche* (ANR) of the *Ministère Français de l'Education Nationale de la Recherche et de la Technologie*.

References

Abel L, Demenais F (1988) Detection of major genes for susceptibility to leprosy and its subtypes in a Caribbean island: Desirade island. Am J Hum Genetics 42:256–266

Abel L, Vu D, Oberti J, Nguyen V, Van V, Guilloud-Bataille M, Schurr E, Lagrange P (1995) Complex segregation analysis of leprosy in southern Vietnam. Genetic Epidemiol 12:63–82

Alcais A, Alter A, Antoni G, Orlova M, Van Thuc N, Singh M, Vanderborght PR, Katoch K, Mira MT, Thai VH, Huong NT, Ba NN, Moraes M, Mehra N, Schurr E, Abel L (2007) Stepwise replication identifies a low-producing lymphotoxin-[alpha] allele as a major risk factor for early-onset leprosy. Nat Genet 39:517–522

Alcais A, Mira M, Casanova J-L, Schurr E, Abel L (2005) Genetic dissection of immunity in leprosy. Curr Opin Immunol 17:44–48

Ali S, Vollaard A, Widjaja S, Surjadi C, van de Vosse E, van Dissel J (2006) PARK2/PACRG polymorphisms and susceptibility to typhoid and paratyphoid fever. Clin Exp Immunol 144:425–431

Britton WJ, Lockwood DNJ (2004) Leprosy. Lancet 363:1209–1219

Broderick P, Carvajal-Carmona L, Pittman AM, Webb E, Howarth K, Rowan A, Lubbe S, Spain S, Sullivan K, Fielding S, Jaeger E, Vijayakrishnan J, Kemp Z, Gorman M, Chandler I, Papaemmanuil E, Penegar S, Wood W, Sellick G, Qureshi M, Teixeira A, Domingo E, Barclay E, Martin L, Sieber O, Kerr D, Gray R, Peto J, Cazier J-B, Tomlinson I, Houlston RS (2007) A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk. Nat Genet 39:1315–1317

Casrouge A, Zhang S-Y, Eidenschenk C, Jouanguy E, Puel A, Yang K, Alcais A, Picard C, Mahfoufi N, Nicolas N, Lorenzo L, Plancoulaine S, Senechal B, Geissmann F, Tabeta K, Hoebe K, Du X, Miller RL, Heron B, Mignot C, de Villemeur TB, Lebon P, Dulac O, Rozenberg F, Beutler B, Tardieu M, Abel L, Casanova J-L (2006) Herpes Simplex Virus Encephalitis in Human UNC-93B Deficiency. Science 314:308–312

Chakravartti M, Vogel F (1973) A twin study on leprosy. In: Becker PE, Lenz W, Vogel F, Wendt GG (eds) Topics in human genetics, vol 1. Georg Thieme, Stuttgart, pp 1–123

Chaudhury S, Hazra S, Saha B, Mazumder B, Biswas P, Chattopadhya D, Saha K (1994) An eight-year field trial on antileprosy vaccines among high-risk household contacts in the Calcutta metropolis. Int J Lepr Other Mycobact Dis 62:389–394

Cole ST, Eiglmeier K, Parkhill J, James KD, Thomson NR, Wheeler PR, Honore N, Garnier T, Churcher C, Harris D, Mungall K, Basham D, Brown D, Chillingworth T, Connor R, Davies RM, Devlin K, Duthoy S, Feltwell T, Fraser A, Hamlin N, Holroyd S, Hornsby T, Jagels K, Lacroix C, Maclean J, Moule S, Murphy L, Oliver K, Quail MA, Rajandream MA, Rutherford KM, Rutter S, Seeger K, Simon S, Simmonds M, Skelton J, Squares R, Squares S, Stevens K, Taylor K, Whitehead S, Woodward JR, Barrell BG (2001) Massive gene decay in the leprosy bacillus. Nature 409:1007–1011

Convit J, Sampson C, Zuniga M, Plata J, Silva J, Molina J, Pinardi ME, Salgado A, Smith PG, Bloom BR (1992) Immunoprophylactic trial with combined *Mycobacterium leprae*/BCG vaccine against leprosy: preliminary results. The Lancet 339:446–450

Eiglmeier K, Fsihi H, Heym B, Cole ST (1997) On the catalase–peroxidase gene, katG, of *Mycobacterium leprae* and the implications for treatment of leprosy with isoniazid. FEMS Microbiol Lett 149:273–278

Fernando SL, Britton WJ (2006) Genetic susceptibility to mycobacterial disease in humans. Immunol Cell Biol 84:125–137

Fitness J, Tosh J, Hill AVS (2002) Genetics of susceptibility to leprosy. Genes Immunol 3:441–453

Gupte M, Vallishayee R, Anantharaman D, Nagaraju B, Sreevatsa., Balasubramanyam S, de Britto R, Elango N, Uthayakumaran N, Mahalingam V, Lourdusamy G, Ramalingam A, Kannan S, Arokiasamy J (1998) Comparative leprosy vaccine trial in south India. Indian J Lepr 70:369–388

Hyun D, Lee M, Hattori N, Kubo S, Mizuno Y, Halliwell B, Jenner P (2002) Effect of wild-type or mutant Parkin on oxidative damage, nitric oxide, antioxidant defenses, and the proteasome. J Biol Chem 277:28572–28577

Knight JC, Keating BJ, Kwiatkowski DP (2004) Allele-specific repression of lymphotoxin-[alpha] by activated B cell factor-1. Nat Genet 36:394–399

Lundmark F, Duvefelt K, Iacobaeus E, Kockum I, Wallstrom E, Khademi M, Oturai A, Ryder LP, Saarela J, Harbo HF, Celius EG, Salter H, Olsson T, Hillert J (2007) Variation in interleukin 7 receptor [alpha] chain (IL7R) influences risk of multiple sclerosis. Nat Genet 39:1108–1113

Malhotra D, Darvishi K, Lohra M, Kumar H, Grover C, Sood S, Reddy BSN, Bamezai RNK (2005) Association study of major risk single nucleotide polymorphisms in the common regulatory region of PARK2 and PACRG genes with leprosy in an Indian population. Eur J Hum Genet 14:438–442

Marmor MF (2002) The ophthalmic trials of G. H. A. Hansen. Surv Ophthalmol 47:275–287

Miller EN, Jamieson SE, Joberty C, Fakiola M, Hudson D, Peacock CS, Cordell HJ, Shaw MA, Lins-Lainson Z, Shaw JJ, Ramos F, Silveira F, Blackwell JM (2004) Genome-wide scans for leprosy and tuberculosis susceptibility genes in Brazilians. Genes Immun 5:63–67

Mira MT (2006) Genetic host resistance and susceptibility to leprosy. Microbes Infect 8:1124–1131

Mira MT, Alcais A, Van Thuc N, Moraes MO, Di Flumeri C, Hong Thai V, Chi Phuong M, Thu Huong N, Ngoc Ba N, Xuan Khoa P,

Sarno EN, Alter A, Montpetit A, Moraes ME, Moraes JR, Dore C, Gallant CJ, Lepage P, Verner A, van de Vosse E, Hudson TJ, Abel L, Schurr E (2004) Susceptibility to leprosy is associated with PARK2 and PACRG. Nature 427:636–640

Mira MT, Alcais A, Van Thuc N, Thai VH, Huong NT, Ba NN, Verner A, Hudson TJ, Abel L, Schurr E (2003) Chromosome 6q25 is linked to susceptibility to leprosy in a Vietnamese population. Nat Genet 33:412–415

Monot M, Honore N, Garnier T, Araoz R, Coppee J-Y, Lacroix C, Sow S, Spencer JS, Truman RW, Williams DL, Gelber R, Virmond M, Flageul B, Cho S-N, Ji B, Paniz-Mondolfi A, Convit J, Young S, Fine PE, Rasolofo V, Brennan PJ, Cole ST (2005) On the origin of leprosy. Science 308:1040–1042

Pandya S (1998) Anti-contagionism in leprosy, 1844–1897. Int J Lepr Other Mycobact Dis 66:374–384

Parkes M, Barrett JC, Prescott NJ, Tremelling M, Anderson CA, Fisher SA, Roberts RG, Nimmo ER, Cummings FR, Soars D, Drummond H, Lees CW, Khawaja SA, Bagnall R, Burke DA, Todhunter CE, Ahmad T, Onnie CM, McArdle W, Strachan D, Bethel G, Bryan C, Lewis CM, Deloukas P, Forbes A, Sanderson J, Jewell DP, Satsangi J, Mansfield JC, Cardon L, Mathew CG (2007) Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility. Nat Genet 39:830–832

Ridley D, Jopling W (1966) Classification of leprosy according to immunity. A five-group system. Int J Lepr Other Mycobact Dis 34:255–273

Roach D, Briscoe H, Saunders B, Britton W (2005) Independent protective effects for tumor necrosis factor and lymphotoxin alpha in the host response to *Listeria monocytogenes* infection. Infect Immunity 73:4787–4792

Roach DR, Briscoe H, Saunders B, France MP, Riminton S, Britton WJ (2001) Secreted lymphotoxin-{{alpha}} is essential for the control of an intracellular bacterial infection. J Exp Med 193:239–246

Schurr E, Alcais A, de Leseleuc L, Abel L (2006) Genetic predisposition to leprosy: a major gene reveals novel pathways of immunity to *Mycobacterium leprae*. Semin Immunol 18:404–410

Shaw M, Donaldson I, Collins A, Peacock C, Lins-Lainson Z, Shaw J, Ramos F, Silveira F, Blackwell J (2001) Association and linkage of leprosy phenotypes with HLA class II and tumour necrosis factor genes. Genes Immun. 2:196–204

Shepard CC (1960) The experimental disease that follows the injection of human leprosy bacilli into footpads of mice. J Exp Med 112:445–454

Shields E, Russell D, Pericak-Vance MA (1987) Genetic epidemiology of the susceptibility to leprosy. J Clin Invest 79:1139–1143

Siddiqui MR, Meisner S, Tosh K, Balakrishnan K, Ghei S, Fisher SE, Golding M, Shanker Narayan NP, Sitaraman T, Sengupta U, Pitchappan R, Hill AVS (2001) A major susceptibility locus for leprosy in India maps to chromosome 10p13. Nat Genet 27:439–441

Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D, Boutin P, Vincent D, Belisle A, Hadjadj S, Balkau B, Heude B, Charpentier G, Hudson TJ, Montpetit A, Pshezhetsky AV, Prentki M, Posner BI, Balding DJ, Meyre D, Polychronakos C, Froguel P (2007) A genome-wide association study identifies novel risk loci for type 2 diabetes. Nature 445:881–885

Sørensen T, Nielsen G, Andersen P, Teasdale T (1988) Genetic and environmental influences on premature death in adult adoptees. New Engl J Med 318:727–732

Tosh K, Meisner S, Siddiqui M, Balakrishnan K, Ghei S, Golding M, Sengupta U, Pitchappan R, Hill A (2002) A region of chromosome 20 is linked to leprosy susceptibility in a South Indian population. J Infectious Dis 186:1190–1193

Vanderborght PR, Pacheco AG, Moraes ME, Antoni G, Romero M, Verville A, Thai VH, Huong NT, Ba NN, Schurr E, Sarno EN, Moraes MO (2007) HLA-DRB1[ast]04 and DRB1[ast]10 are associated with resistance and susceptibility, respectively, in Brazilian and Vietnamese leprosy patients. Genes Immun 8:320–324

WHO (2007) Global leprosy situation, 2007. Weekly Epidemiol Record 82:225–232

Zerva L, Cizman B, Mehra NK, Alahari SK, Murali R, Zmijewski CM, Kamoun M, Monos DS (1996) Arginine at positions 13 or 70–71 in pocket 4 of HLA-DRB1 alleles is associated with susceptibility to tuberculoid leprosy. J Exp Med 183:829–836

Zhang S-Y, Jouanguy E, Ugolini S, Smahi A, Elain G, Romero P, Segal D, Sancho-Shimizu V, Lorenzo L, Puel A, Picard C, Chapgier A, Plancoulaine S, Titeux M, Cognet C, von Bernuth H, Ku C-L, Casrouge A, Zhang X-X, Barreiro L, Leonard J, Hamilton C, Lebon P, Heron B, Vallee L, Quintana-Murci L, Hovnanian A, Rozenberg F, Vivier E, Geissmann F, Tardieu M, Abel L, Casanova J-L (2007) TLR3 deficiency in patients with herpes simplex encephalitis. Science 317:1522–1527

Zondervan KT, Cardon LR (2007) Designing candidate gene and genome-wide case-control association studies. Nat Protocols 2:2492–2501