**The contribution of host genetics to TB disease**

Melanie J. Newport

Division of Clinical Medicine

Brighton and Sussex Medical School

Falmer

BN1 9PS

UK

Tel:+44 (0) 1273 877882

Fax: +44 (0) 1273 877884

Email:m.j.newport@bsms.ac.uk

Running title: Host genetics and TB

**Abstract**

There is robust epidemiological evidence that susceptibility to tuberculosis is in part heritable. This has driven the use of genetics to try to find the genes and pathways involved that could in the longer term contribute towards the development of new therapies and a better vaccine for this major global health problem. This paper reviews the progress made in the field to date, and discusses the challenges inherent in undertaking genetics studies on a complex disease with clinically diverse phenotypes, that affects many genetically different populations and which is further complicated by the presence of a pathogen which has a genome too.

**Key words**

Genetic linkage, genome wide association studies, immune response genes, Mendelian susceptibility to mycobacterial disease, single nucleotide polymorphisms, tuberculosis

**Introduction**

In 1997 Dye *et al.* estimated that there were approximately 8 million new cases of tuberculosis (TB) globally and between 1.4 and 2.8 million deaths [[1](#_ENREF_1)]. At this time, TB rates were on the rise, partly driven by the human immunodeficiency virus (HIV) pandemic, to peak in around 2003 [[2](#_ENREF_2)]. The efforts of global TB control initiatives such as the World Health Organization STOP TB partnership have helped reverse this trend such that by 2011 the statistics were similar to those for 1997 [[3](#_ENREF_3)] [[4](#_ENREF_4)]. Even so, TB remains one of the most important global public health challenges of our time. Dye *et al.* also estimated that 32% of the world’s population at the time was infected with *Mycobacterium tuberculosis* (MTB), the bacterium that causes TB. Thus, although TB is one of the commonest infectious diseases in the world, the vast majority of people infected with MTB do not go on to develop clinical disease. Increased understanding of the host factors that are associated with resistance (or susceptibility) to TB could be key to the development of improved TB control strategies that require the development of new treatments and a better vaccine.

Many factors are recognised to predispose to TB. Poverty, and its knock-on effects including overcrowding, living or working in poorly ventilated environments, lack of access to health care and poor nutrition, is globally the most important risk factor for the development of TB [[5-7](#_ENREF_5)] Immunosuppression, whether secondary to co-infection with HIV [[8](#_ENREF_8)], recent measles infection [[9](#_ENREF_9)] or iatrogenic causes (for example anti-tumour necrosis factor- treatments for autoimmune disorders [[10](#_ENREF_10)]) is also a recognised risk factor. However, even allowing for environmental and pathogen factors that contribute to disease there is clear evidence that host genetic factors are also important determinants of the outcome of the encounter between an individual and MTB. The year 2013 harbours two significant genetics anniversaries: it is 60 years since the structure of DNA was elucidated [[11](#_ENREF_11)] and 10 years since the Human Genome Project (HGP) was completed [[12](#_ENREF_12)]. It is therefore timely to review advances in our understanding of the contribution of host genetics to TB in the ‘post-genome era’.

**Evidence that susceptibility to TB is genetically regulated**

It has been observed that susceptibility to TB varies between populations of different genetic origin [[13](#_ENREF_13)]. Accidental inoculation of 251 children with virulent MTB in 1926 (which contaminated the BCG vaccine they received to protect them against TB) provided early evidence of variation in host susceptibility to disease once infected with MTB. All infants were given the same oral dose and strain of infection yet the outcome varied considerably. Seventy-seven infants died, 72 of whom had confirmed TB at autopsy, 47 remained healthy without evidence of TB and 127 had radiological evidence of TB [[14](#_ENREF_14)]. More direct evidence that this variation in susceptibility has a genetic basis comes from twin studies which have shown higher concordance rates in genetically identical monozygous twins than in dizygous twins who on average shared only 50% on their genes, suggesting a heritable component to TB [[15](#_ENREF_15), [16](#_ENREF_16)]. Twin studies have also been used to assess the genetic contribution to immune responses relevant to TB such as cellular responses to purified protein derivative (PPD) used for tuberculin skin testing and to killed MTB. These studies reported heritabilities of between 39-71% [[17](#_ENREF_17), [18](#_ENREF_18)]. Finally, a number of monogenic disorders, reviewed below, have also indicated that genetic variants predispose to mycobacterial infections including TB. Here, the spectrum and severity of the clinical phenotypes correlate with the functional impact of the various mutations suggesting a model of susceptibility that can be extrapolated to the outbred population. Single mutations with extreme effects are lethal and tend to cluster in families, but combinations of more subtle genetic variants in the same genes (for example, those that affect gene regulation not protein function) may contributing towards TB susceptibility at the population level [[19](#_ENREF_19)].

However, identification of the molecular mechanisms that underlie these observations - i.e. elucidation of the genes and the proteins they encode – is not straightforward. The genetic model for inheritance of susceptibility to TB is not understood, beyond it being a multi-factorial trait in which multiple genes interact with each other, with environmental factors and with the pathogen which has a genome of its own. Indeed, with a generation time of around 20 hours MTB evolves far more quickly than its human counterpart and has developed a range of mechanisms to survive through subversion of host immune responses [[20](#_ENREF_20)]. It is also not understood whether there are several genes with small additive effects or a few major genes whose effects are modified by other genes, whether the same gene variant functions differently in different environments (e.g. according to micronutrient levels) or when interacting with different strains of MTB (of which there are many) and whether different genes are involved depending on the clinical phenotype or the population under investigation.

**Approaches towards identifying TB susceptibility genes**

Approaches towards the identification of human TB susceptibility genes include: the extrapolation of studies done in other organisms, most commonly mice; learning from studying human families with rare single gene disorders that predispose to mycobacterial infection, also known as Mendelian susceptibility to mycobacterial disease (MSMD); studies of multi-case families where genetic similarities are correlated with phenotypic similarities (resistant or susceptible to TB) within families, known as linkage studies; and population studies where genetic differences are correlated between those who have or do not have the disease, known as association studies that usually follow a case control study design.

Linkage and association studies can focus on individual (candidate) genes or take a genome wide perspective. Given the importance of a competent immune response in controlling infection with MTB, most candidate gene studies have focused on immune response genes, particularly those involved in innate and cell mediated immunity which deal with intracellular pathogens such as MTB. The key players include macrophages which phagocytose mycobacteria, triggering a range of secondary responses that upregulate macrophage function to kill the organism, and the activation of other components of the immune system through secretion of critical cytokines such as interleukin (IL)-12 and IL-18. These cytokines act on lymphocytes to induce interferon- (IFN-) production which further upregulates macrophage function and enhances antigen presentation by these cells ultimately to stimulate antigen-specific responses, typically but not exclusively, by CD4 type 1 helper T-lymphocytes. However, it has become clear that this is a highly simplistic model. There is evidence that other lymphocyte subsets including CD8 T-lymphocytes and B –lymphocytes are involved, as are neutrophils that previously have not been thought to be important in immunity to mycobacteria [[21](#_ENREF_21)]. Clearly the number of molecules and therefore potential candidate genes involved is enormous. Readers are referred to more detailed reviews of the immune response to TB published elsewhere that include helpful figures [[22-24](#_ENREF_22)].

Before the HGP, the main shortcoming of candidate gene studies was that studies were limited to genes that had already been discovered. Publication of the human genome sequence and its gene content changed this, but more importantly paved the way for other genome projects such as International HapMap and the 1000 genomes project that characterised the variation within the human genome in population specific ways that were required to map disease genes [[25](#_ENREF_25)] [[26](#_ENREF_26)]. Concurrent technological advances that allowed high-throughput low-cost genotyping enabled systematic ‘hypothesis-free’ interrogation of the genome and the focus switched from candidate gene studies to genome-wide studies in much larger, statistically more powerful population samples.

**Animal studies**

Animal studies have contributed much to our understanding of the functional components of the immune system and their role in immunity to mycobacterial infection. Differential susceptibility to infection with *M. bovis* BCG, *Salmonella typhimurium* and *Leishmania donovani* in strains of inbred laboratory mice led to the discovery of *the Bcg/Ity/Lsh* gene, later renamed *Nramp1* (natural resistance associated macrophage protein and then *Slc11a1* (solute carrier family 11a member 1)as an infection susceptibility gene [[27](#_ENREF_27)] [[28](#_ENREF_28)]*.* The human homologue *NRAMP1/SLC11A1* has been extensively studied as a human TB susceptibility gene with data both supporting and refuting an association between variation in the *NRAMP1* gene and TB. A recent meta-analysis of all published human data concluded that there is a role for this gene is TB susceptibility [[29](#_ENREF_29)].

As gene disruption techniques became available, various ‘knockout’ mice which lacked functional copies of specific genes were shown to be more susceptible to mycobacterial infection. This approach identified several genes that when disrupted altered murine susceptibility to TB [[30](#_ENREF_30)]. These genes are intrinsic to innate and adaptive pathways as well as for granuloma formation, the hallmark of TB pathology, underlining the complexity of the immune response s to MTB infection. However, few of these genes have been convincing as human susceptibility genes. Even when there is effectively an equivalent human ‘knockout’ (i.e. people with MSMD, described in more detail below) the effects of null mutations (i.e. no gene function) can be less severe in humans suggesting the possibility of redundancy in humans for some molecules such as IL-12 [[31](#_ENREF_31)].

More recently, elegant studies in a zebrafish model led to the identification of a variant in the gene that encodes leukotriene A(4) hydrolase (*lta4h*) that predisposed fish to disease caused by its natural pathogen *M. marinum*, a close relative of MTB [[32](#_ENREF_32)]. This enzyme is involved in the synthesis of the chemoattractant molecule leukotriene B(4) and two intronic variants in the human *LTA4H* gene were found to be associated with protection from tuberculosis and leprosy in Vietnamese and Nepalese population respectively [[32](#_ENREF_32)]. Being heterozygous for both polymorphisms conferred protection against TB compared to the homozygous states. In the zebrafish model, dysfunction of the *lta4h* gene led both to the build up of an anti-inflammatory agent lipoxin A4 and a proinflammatory state due to independent interaction with the tumour-necrosis factor - pathway. Demonstrating how studies on host genetics can potentially lead to therapeutic benefits, a functional polymorphism was identified in the promoter region of the *LTA4H* gene in humans that was associated with inflammatory cell recruitment, patient survival and response to dexamethasone therapy in TB meningitis [[33](#_ENREF_33)]. The dichotomous situation, where both hypo- and hyper-inflammatory states contributed to poor outcome was highlighted, leading to the suggestion that *LTA4H* genotype could predict outcome and response to treatment in TB. Further clinical studies will be required to test this hypothesis. The association between *LTA4H* variants and TB was not replicated in a study of over 9000 Russians [[34](#_ENREF_34)]. The reasons for inconsistencies between study results, which occur frequently in association studies, are discussed in more detail later.

**Mendelian disorders associated with increased susceptibility to mycobacterial infection (Mendelian Susceptibility to Mycobacterial Disease, MSMD)**

Genetic analysis of primary human immunodeficiency disorders has shed light on the critical pathways and genes required to control mycobacterial infections in humans. These rare ‘experiments of nature’ have especially highlighted the role of the IFN-/IL-12/23 pathway in immunity to mycobacteria. The first mutation leading to disseminated mycobacterial infection was reported in a consanguineous Maltese family, where a point mutation causing a premature stop codon in the gene encoding the IFN- receptor ligand binding chain (IFNGR1) was identified [[35](#_ENREF_35)]. The affected children were homozygous for the same mutation and did not express the receptor on their immune cells leading to a severe clinical phenotype. Three out of four children died and the fourth survived as a result of a bone marrow transplant. A different mutation was identified in the same gene to explain disseminated BCG-osis in a vaccinated infant whose parents were first cousins [[36](#_ENREF_36)]. These families defined a new clinical syndrome: an inherited immunodeficiency that predisposed to mycobacterial infection, coined MSMD. However, mutation in *IFNGR1* did not explain all cases of MSMD. Further investigation of other children with severe mycobacterial infections led to the discovery of mutations in six other genes within the IFN-/IL-12/23 pathway reviewed in more detail elsewhere [[37-39](#_ENREF_37)]. These genes encode the signal transducing chain of the IFN- receptor (*IFNGR2*), the p40 subunit of IL-12 (*IL12B*), the beta subunit of the IL-12 receptor (*IL12RB1* which is shared with the IL-23 receptor), the signal transducing and activator of transcription molecule 1 (*STAT1*), nuclear factor-κB-essential modulator (*NEMO*) and tyrosine kinase 2 (*TYK2*). MSMD caused by mutations in NEMO is X-linked, while the other genes are autosomal, but may have recessive or dominant effects depending on the nature of the mutation. There are many different mutations described for most of these genes leading to a spectrum of clinical presentation and disease severity [[40](#_ENREF_40)]. These genes became obvious candidate TB susceptibility genes in outbred populations.

**Linkage studies**

Linkage studies were originally developed to map the genes responsible for single gene Mendelian traits such as cystic fibrosis. The general principles are based on the fact that the closer together two loci are on a given chromosome the less likely they are to be separated during recombination, when paternal and maternal copies of the chromosome exchange genetic material before the hybrid chromosomes are transmitted to the next generation. If the loci are both polymorphic it is possible to track them through families: when the same variant of a genetic marker is co-inherited with the disease in a family the two are said to be linked - i.e. in close physical proximity on the chromosome. Knowing the genomic location of the markers used allows mapping of the disease gene.

This methodology was adapted for use in the study of multi-factorial traits and the affected sibling pair study design became popular. According to Mendelian rules, siblings are expected to inherit the same copy of a gene (or genetic marker) from their parents 50% of the time. Thus if siblings who both have a disease such as TB are also inherited specific variants of a genetic marker from their parents more often than expected by chance, then that marker is linked to the trait. This approach led to the identification of Nucleotide-binding Oligomerization Domain containing 2 (*NOD2*) as a biologically plausible Crohn’s disease susceptibility gene that has subsequently been confirmed in other studies [[41](#_ENREF_41)].

Regarding TB, linkage studies on populations from South Africa, The Gambia, Malawi, Uganda, Brazil, Morocco and Thailand have to date identified linkage with regions on chromosomes 5, 6, 7, 8, 10, 11, 15, 17, 20 and the X chromosome [[42-47](#_ENREF_42)]. Only one of these was identified independently in two populations (chromosome 20). Fine mapping studies on some of these regions have been undertaken, but have not identified any functional mutations that could be implicated in TB susceptibility [[48](#_ENREF_48)]. However, linkage studies were another source of candidate genes that could be tested in association studies [[49](#_ENREF_49), [50](#_ENREF_50)].

**Association studies**

Association studies investigating the role of candidate genes have been the commonest studies undertaken towards identifying TB susceptibility genes. For diseases such as TB it is much easier to collect cases and controls than multi-case families and there were many candidates to test based on what was known about immunity in TB and from the animal, MSMD and linkage studies described above. Candidates tested included genes encoding innate immune response proteins (e.g. toll-like receptors 2 and 9, and toll-interleukin 1 receptor domain containing adaptor protein TIRAP), cytokines and their receptors (including IL-10, IFN- and IL-12), proteins involved in phagocytosis and intracellular killing of mycobacteria (e.g. *NRAMP1*, mannose receptor, complement 3 receptor, purinergic receptor P2X and nitric oxide synthase) and the vitamin D receptor. Summarising the results of hundreds of reported studies, with a few exceptions such as *NRAMP1* and the class II human leukocyte antigen (HLA) locus where there is consistent association across populations [[29](#_ENREF_29)] [[51](#_ENREF_51)], few genes that have been found to be associated with TB by one group have been confirmed convincingly by others. Reasons for this are multiple and include statistically underpowered studies, a lack of rigorous phenotyping (which for a disease such as TB is critical given that the clinical manifestations are so diverse), genetically heterogenous populations and publication bias. Comprehensive reviews of the results of candidate gene association studies in TB, including summary tables, have been published elsewhere [[52-54](#_ENREF_52)].

Once the variation in the human genome has been extensively characterised and the technology became available to allow hundreds of thousands of single nucleotide polymorphisms (SNPs) to be genotyped per individual in one experiment, attention turned to genome-wide association studies (GWAS). In GWAS, SNP variants across the genome are tested for association with the disease of interest. Large numbers of SNPs need to be tested to get adequate coverage of the genome (typically between 0.5 and 1 million) and large sample sizes are required, because gene effects are likely to be small, multiple loci are likely to contribute and statistical correction is required given the large numbers of tests undertaken. Proof of principle for this approach was demonstrated by the Wellcome Trust Case Control Consortium (WTCCC) which investigated seven common diseases in Caucasian populations [[55](#_ENREF_55)] Known associations, for example between the HLA region and type 1 diabetes, and *NOD2* and Crohn’s disease were reassuringly identified in addition to several new loci for many of the diseases. The WTCCC extended its studies to GWAS for malaria and TB in a Gambian population sample. The initial analysis for the Gambian population did not reveal any significant associations but when the data were combined with genome-wide data from a Ghanaian population a TB susceptibility locus was identified on chromosome 18 [[56](#_ENREF_56)]. The effect was small (odd ratio of 1.19, 95% CI= 1.13-1.27) and the associated SNP was located in a gene desert. Interestingly, none of the previous associations (e.g. with *NRAMP1* or HLA) were unequivocally detected in this study. However, the malaria GWAS in a Gambian population did not identify the expected strong association between the sickle polymorphism in the -globin gene and malaria despite its known protective effect. When this region of chromosome 11 was sequenced in the study population and population-specific SNPs tested the association signal was much stronger. This highlights the challenges of undertaking GWAS in African populations which are older, have smaller blocks of DNA markers in linkage disequilibrium and harbour more genetic diversity than the Caucasian populations in which the GWAS tools used in these studies had been developed. This can be overcome by including more SNPs and by using population specific variants [[57](#_ENREF_57)]. Indeed, as such data became available through the 1000 genomes project, a statistical technique known as imputing was used to supplement the GWAS data from the Ghana study and a new locus on chromosome 11 was found to be associated with TB and replicated in the Gambian data as well as in Russian and Indonesian samples [[58](#_ENREF_58)]. One of the SNPs on chromosome 18 associated with susceptibility to TB in 11425 Africans was replicated in one Chinese population sample of 2280 people, though interestingly the SNP had a protective effect [[59](#_ENREF_59)]. The association was not identified in a second study in China, though at 1218 individuals the samples size was smaller and therefore an effect may have been missed [[60](#_ENREF_60)].

GWAS studies have been undertaken in other populations where TB is endemic. No association was found in a mixed Japanese and Thai population, but when the data were analysed by age, an locus on chromosome 20q 12 was identified [[61](#_ENREF_61)]. In Indonesia, a small GWAS using 259 samples led to the identification of 2453 promising SNPs (p value of <0.05) that were genotyped in a larger Indonesian sample of 1189 subjects. From this study 251 SNPs with suggestive association (p<0.05) were typed in 3760 Russian samples. Eight loci were identified with suggestive association (p values ranging from 0.0004-0.0067) [[62](#_ENREF_62)]. Selecting a threshold for statistical significance is an area of much debate in GWAS studies. Given the large number of markers being tested (100,000 to over 1 million) in samples sizes that may include over 10,000 individuals, many argue that a rigorous approach is required to avoid false positive results arising from multiple testing. On the other hand, given the usually small effect of each genetic variant and the ‘hypothesis-generating’ nature of the GWAS approach it may be reasonable to make the case for less stringent cut-offs so that potential real loci are not discarded in under-powered studies and data suggestive of association can be further investigated in larger samples [[63](#_ENREF_63)].

**Discussion and future directions**

In summary, a lot of studies have been undertaken in order to identify the molecular basis of the epidemiologically well-defined contribution of host genetics to susceptibility to TB. However, progress remains slow, especially when compared to other complex traits. There are a number of reasons for this, some of which have already been discussed. Starting with the basics, any genetic study must begin with a clearly defined phenotype. TB has a reputation for its myriad of clinical presentations and even pulmonary TB can be diverse in its clinical presentation. If, as is quite possible, different genes regulate the different clinical phenotypes there is plenty of scope for results to be confounded if meticulous phenotyping is not undertaken when enrolling cases to studies.

Many genetic candidate gene association studies were underpowered to detect effects, or the small sample size led to spurious false positive results. The advent of the GWAS was expected to overcome this issue but for TB the results so far have not been as promising. This may be due to a number of reasons. The majority of TB cases live in low-and middle income countries where development of genomics capacity has not been a priority [[64](#_ENREF_64)]. The need for population specific data to underpin the choice and number of SNPs to be typed has already been highlighted and this is now being addressed through initiatives such as the African Genome Variation Project (<http://www.sanger.ac.uk/research/initiatives/globalhealth/research/africangenome.html>) and the H3 Africa programme (http://www.h3africa.org).

Secondly, the genetic architecture of TB is poorly understood. In other words, it is not known whether disease results from a cumulative effect of a few common variants or from the cumulative effects of many rare variants [[65](#_ENREF_65), [66](#_ENREF_66)]. This is important because different experimental designs are required. To be successful, the GWAS approach requires that disease is due to the former and common variants will be detected because they are in linkage disequilibrium with the SNPs included in the SNP array Thus, the SNP array does not need to include every one of the roughly 10 million SNPs in the human genome because, due to the phenomenon of linkage disequilibrium, one SNP can be used to represent a block of SNPs located in the same chromosomal region. However, if disease is due to the effects of rare variants, these will not be represented on the SNP array due to a lack of linkage disequilibrium and a sequencing approach to identify the rare variants is required.

These matters could be overcome either through GWAS using much larger numbers of SNPs to increase coverage across the genome and thereby detect more of the common variants (if common variants are important), or alternatively through sequencing the exons across the whole genomes of cases and controls (the exome i.e. the sum of all parts of the genome that encode amino acids) which is where the rare functional variants implicated in the second model are thought likely to be found. These variants could then be typed in much larger case control association studies. Whole genome sequencing would be a step up from whole exome sequencing and ensure all potential rare variants (i.e. include non-coding functional SNPs) are captured but the main limitation here is cost. Neither approach has been applied in TB yet but one proof of principle study supporting the exome sequencing approach identified a higher frequency of amino-acid altering variants in TB cases than controls when the exons were sequenced from Toll-like receptor genes [[67](#_ENREF_67)].

In the studies described above, family and population genetic studies are used as tools to identify critical pathways as targets for new therapies and vaccines. Advances in technology in other molecular areas have allowed the development of other approaches to address this aim and to bypass the genetics stage, moving downstream to the molecules that the genes encode. For example, whole genome studies of RNA expression at the cellular level in the context of MTB infection have been revealing. Gene expression profiling in macrophages (the cells which MTB preferentially infects and which orchestrate the immune response once infected) infected with MTB from healthy individuals with different clinical manifestations of TB were undertaken in a Vietnamese population [[68](#_ENREF_68)]. Expression of 16 genes was significantly different in cells from individuals with latent, pulmonary and meninigeal TB (studied after successful completion of treatment for the latter two groups). Replication studies in larger sample of individuals identified a single gene that was differentially expressed when compared between the different clinical phenotypes, the chemokine (C-C motif) ligand 1 gene *CCL1*. SNPs in this gene were typed in a case control genetic study and found to be associated with susceptibility to TB.

In another study, macrophages were transfected *in vitro* with small interference RNA (siRNA) molecules that bind to mRNA molecules and prevent their translation into functional proteins [[69](#_ENREF_69)]. The effect of silencing these genes on MTB growth was measured. In the initial screen, silencing of 1138 of 18,174 targeted genes resulted in either an increase or a decrease in the mycobacterial load within the cells. An iterative process that became more stringent with each round led to the identification of 275 validated, functionally-associated proteins organised into networks that were either engaged by or suppressed by MTB infection. This work will contribute to the understanding the mechanisms of persistence of mycobacteria in latent infection and has identified potential therapeutic targets.

TB is an excellent example of how the host response to a pathogen determines the clinical outcome in infectious diseases. At one end of the spectrum a weak response allows uncontrolled replication and dissemination of infection, while at the other end of the spectrum, vigorous but ineffective responses lead to pathology and death if untreated. Whilst host genetics plays a role there is also good evidence that genetic variation in the pathogen can modulate this response and given the co-evolution of MTB and humans, such that specific MTB lineages are associated with specific human populations [[70](#_ENREF_70)],. it is imperative that both partners are considered. Studies have demonstrated a correlation between clinical phenotype and MTB strains [[71](#_ENREF_71)] , host genotype and MTB strain in individuals with confirmed pulmonary TB [[72](#_ENREF_72)], ethnicity, MTB strain and clinical phenotype [[73](#_ENREF_73)] and a relationship between host and bacterial genetics and clinical phenotype [[74](#_ENREF_74)]. The siRNA study described above also demonstrated that the host response varied according to the strain of infecting MTB, highlighting the importance of considering host and pathogen factors in unison.

**Conclusions**

Whilst there has not been a major breakthrough in TB genetics equivalent to the discovery of a deletion mutation in the chemokine CCR5 receptor that protects against HIV infection and that led to the development of a new class of anti-retroviral drugs [[75](#_ENREF_75)], much has been learnt over the years in the field of host genetics and TB. We have certainly learnt that it is proving to be more complicated than perhaps originally anticipated and when compared to other complex disease such as Crohn’s disease. As technology develops and the knowledge gained so far from various disciplines is integrated it is likely that a large-scale, multidisciplinary effort involving epidemiology, clinical medicine, host and pathogen genetics, systems biology and bioinformatics, will ultimate bear fruit towards control of a major global health problem.

**Acknowledgements**

The author’s has received support from the Wellcome Trust, the MRC, the National Centre for the Replacement, Refinement and Reduction of Animals in Research (NC3Rs) and the British Lung Foundation for studies on the genetics of TB.

**List of abbreviations**

BCG Bacille Calmette Guerin

CCL1 Chemokine (C-C motif) Ligand 1

CCR5 Chemokine (C-C motif) Receptor 5

GWAS Genome Wide Association Study

HIV Human Immunodeficiency Virus

HGP Human Genome Project

HLA Human Leukocyte Antigen

IFN- Interferon-

IFNGR1 IFN- Receptor 1 (ligand binding chain)

IFNGR2 IFN- Receptor 2 (signal transducing chain)

IL Interleukin

IL12B Interleukin-12 p40 subunit

IL12RB Interleukin 12 Receptor Beta subunit

LTA4H Leukotriene A(4) hydrolase

MSMD Mendelian susceptibility to mycobacterial disease

MTB *Mycobacterium tuberculosis*

NEMO Nuclear Factor-κB-Essential Modulator

NOD2 Nucleotide-binding Oligomerization Domain containing 2

NRAMP1 Natural Resistance Associated Macrophage Protein

PPD Purified Protein Derivative

siRNA small interference RNA

SLC11A1 Solute Carrier family 11A member 1

SNP Single Nucleotide Polymorphism

STAT1 Signal Transducing and Activator of Transcription molecule 1

TB Tuberculosis

TIRAP Toll-Interleukin 1 Receptor domain containing Adaptor Protein

TYK2 Tyrosine Kinase 2

WTCCC Wellcome Trust Case Control Consortium

**Conflict of interest**

None

**References**

1. Dye C, S Scheele, P Dolin, P V. and MC Raviglione, Consensus statement. Global burden of tuberculosis:estimated incidence, prevalence, and mortality by country. WHO global surveillance and monitoring project. JAMA, 1999. 282: p. 677-686.

2. Zumla A, M Raviglione, R Hafner and CF von Reyn, Tuberculosis. N Engl J Med, 2013. 368(8): p. 745-55.

3. Glaziou P, D Falzon, K Floyd and M Raviglione, Global epidemiology of tuberculosis. Semin Respir Crit Care Med, 2013. 34(1): p. 3-16.

4. Stop TB Partnership and World Health Organization, *Global Plan to Stop TB 2006-2015*. 2006 World Health Organization: Geneva.

5. Escombe AR, L Huaroto, E Ticona, et al., Tuberculosis transmission risk and infection control in a hospital emergency department in Lima, Peru. Int J Tuberc Lung Dis, 2010. 14(9): p. 1120-6.

6. Andrews JR, C Morrow and R Wood, Modeling the role of public transportation in sustaining tuberculosis transmission in South Africa. Am J Epidemiol, 2013. 177(6): p. 556-61.

7. Benatar SR and R Upshur, Tuberculosis and poverty: what could (and should) be done? Int J Tuberc Lung Dis, 2010. 14(10): p. 1215-21.

8. Pitchenik AE and MA Fischl, Disseminated tuberculosis and the acquired immunodeficiency syndrome. Ann Intern Med, 1983. 98(1): p. 112.

9. Starr S and S Berkovich, The Effect of Measles, Gamma Globulin Modified Measles, and Attenuated Measles Vaccine on the Course of Treated Tuberculosis in Children. Pediatrics, 1965. 35: p. 97-102.

10. Keane J, S Gershon, RP Wise, et al., Tuberculosis associated with infliximab, a tumor necrosis factor alpha-neutralizing agent. N. Engl. J. Med., 2001. 345: p. 1098-1104.

11. Watson JD and FH Crick, Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature, 1953. 171(4356): p. 737-8.

12. Collins FS, M Morgan and A Patrinos, The Human Genome Project: lessons from large-scale biology. Science, 2003. 300(5617): p. 286-90.

13. Stead WW, JW Senner, WT Reddick and JP Lofgren, Racial differences in susceptibility to infection by *Mycobacterium tuberculosis*. N. Engl. J. Med., 1990. 322: p. 422-427.

14. Reider HL, Die Abklärung der Lübecker Säuglingstuberkulose (Clarification of the Lubeck infact tuberculosis). Pneumologie, 2003. 57: p. 402-405.

15. Kallman FJ and D Reisner, Twin studies on genetic variations in resistance to tuberculosis. J. Hered., 1943. 34: p. 293-301.

16. Comstock G, Tuberculosis in Twins: A Re-analysis of the Prophit Survey. Am. Rev. Respir. Dis., 1978. 117: p. 621-624.

17. Newport MJ, T Goetghebuer, HA Weiss, et al., Genetic regulation of immune responses to vaccines in early life. Genes Immun., 2004. 5: p. 122-129.

18. Wiart A, A Jepson, W Banya, et al., Quantitative association tests of immune responses to antigens of Mycobacterium tuberculosis. Twin Res., 2004. 7: p. 578-588.

19. Ottenhoff THM, D Kumararatne and J-L Casanova, Novel human immunodeficiencies reveal the essential role of type-1 cytokines in immunity to intracellular bacteria. Immunol. Today, 1998. 19: p. 491-494.

20. Gupta A, A Kaul, AG Tsolaki, U Kishore and S Bhakta, Mycobacterium tuberculosis: immune evasion, latency and reactivation. Immunobiology, 2012. 217(3): p. 363-74.

21. Lowe DM, PS Redford, RJ Wilkinson, A O'Garra and AR Martineau, Neutrophils in tuberculosis: friend or foe? Trends Immunol, 2012. 33(1): p. 14-25.

22. Ernst JD, The immunological life cycle of tuberculosis. Nat Rev Immunol, 2012. 12(8): p. 581-91.

23. Ottenhoff TH, New pathways of protective and pathological host defense to mycobacteria. Trends Microbiol, 2012. 20(9): p. 419-28.

24. Dheda K, SK Schwander, B Zhu, RN van Zyl-Smit and Y Zhang, The immunology of tuberculosis: from bench to bedside. Respirology, 2010. 15(3): p. 433-50.

25. The International HapMap Consortium, The international HapMap project. Nature, 2003. 426: p. 789-796.

26. Abecasis GR, D Altshuler, A Auton, et al., A map of human genome variation from population-scale sequencing. Nature, 2010. 467(7319): p. 1061-73.

27. Blackwell JM, The macrophage resistance gene, *Lsh/Ity/Bcg*. Res. Immunol., 1989. 140: p. 767-828.

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

29. Li X, Y Yang, F Zhou, et al., SLC11A1 (NRAMP1) polymorphisms and tuberculosis susceptibility: updated systematic review and meta-analysis. PLoS One, 2011. 6(1): p. e15831.

30. Di Pietrantonio T and E Schurr, Mouse models for the genetic study of tuberculosis susceptibility. Brief Funct Genomic Proteomic, 2005. 4(3): p. 277-92.

31. Fieschi C and JL Casanova, The role of interleukin-12 in human infectious diseases: only a faint signature. Eur J Immunol, 2003. 33(6): p. 1461-4.

32. Tobin DM, JC Vary, JP Ray, et al., The lta4h locus modulates susceptibility to mycobacterial infection in zebrafish and humans. Cell, 2010. 140: p. 717-730.

33. Tobin DM, FJ Roca, SF Oh, et al., Host genotype-specific therapies can optimize the inflammatory response to mycobacterial infections. Cell, 2012. 148(3): p. 434-46.

34. Curtis J, L Kopanitsa, E Stebbings, et al., Association analysis of the LTA4H gene polymorphisms and pulmonary tuberculosis in 9115 subjects. Tuberculosis (Edinb), 2011. 91(1): p. 22-5.

35. Newport MJ, CM Huxley, S Huston, et al., A mutation in the interferon-gamma receptor gene and susceptibility to mycobacterial infections in man. N. Engl. J. Med., 1996. 335: p. 1941-1949.

36. Jouanguy E, F Altare, S Lamhamedi, et al., Interferon-gamma-receptor deficiency in an infant with fatal bacille Calmette-Guerin infection. N. Engl. J. Med., 1996. 335: p. 1956-1959.

37. Newport MJ, SM Holland, M Levin and J-L Casanova, *Inherited disorders of the interleukin-12-interferon gamma axis*, in *Primary Immunodeficiency Diseases. A Molecular and Genetic Approach*, H.D. Ochs, E. Smith, and J.M. Puck, Editors. 2006, Oxford University Press: Oxford. p. 390-401.

38. Patel SY, R Doffinger, G Barcenas-Morales and DS Kumararatne, Genetically determined susceptibility to mycobacterial infection. J Clin Pathol, 2008. 61(9): p. 1006-12.

39. Cottle LE, Mendelian susceptibility to mycobacterial disease. Clin Genet, 2011. 79(1): p. 17-22.

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

41. Hugot J-P, M Chamaillard, H Zouali, et al., Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. Nature, 2001. 411: p. 599-603.

42. Bellamy R, N Beyers, KPWJ McAdam, et al., Genetic susceptibility to tuberculosis in Africans: a genome-wide scan. Proc. Natl. Acad. Sci. USA, 2000. 97: p. 8005-8009.

43. Stein CM, S Zalwango, LL Malone, et al., Genome scan of M. tuberculosis infection and disease in Ugandans. PLoS One, 2008. 3: p. e4095.

44. Miller EN, SE Jamieson, C Joberty, et al., Genome-wide scans for leprosy and tuberculosis susceptibility genes in Brazilians. Genes Immun., 2004. 5: p. 63-67.

45. Baghdadi JE, M Orlova, A Alter, et al., An autosomal dominant major gene confers predisposition to pulmonary tuberculosis in adults. J. Exp. Med., 2006. 203: p. 1679-1684.

46. Cooke GS, SJ Campbell, S Bennett, et al., Mapping of a novel susceptibility locus suggests a role for MC3R and CTSZ in human tuberculosis. Am J Respir Crit Care Med, 2008. 178(2): p. 203-7.

47. Mahasirimongkol S, H Yanai, N Nishida, et al., Genome-wide SNP-based linkage analysis of tuberculosis in Thais. Genes Immun., 2009. 10: p. 77-83.

48. Cervino ACL, S Lakiss, O Sow, et al., Fine mapping of a putative tuberculosis-susceptibility locus on chromosome 15q11-13 in African families. Hum. Mol. Genet., 2002. 11: p. 1599-1603.

49. Adams LA, M Moller, A Nebel, et al., Polymorphisms in MC3R promoter and CTSZ 3'UTR are associated with tuberculosis susceptibility. Eur J Hum Genet, 2011. 19(6): p. 676-81.

50. Moller M, A Nebel, R Valentonyte, et al., Investigation of chromosome 17 candidate genes in susceptibility to TB in a South African population. Tuberculosis (Edinb), 2009. 89(2): p. 189-94.

51. Newport MJ and S Nejentsev, Genetic susceptibility to tuberculosis. Monaldi Arch. Chest Dis., 2004. 61: p. 102-111.

52. Moller M and EG Hoal, Current findings, challenges and novel approaches in human genetic susceptibility to tuberculosis. Tuberculosis (Edinb), 2010. 90(2): p. 71-83.

53. Azad AK, W Sadee and LS Schlesinger, Innate immune gene polymorphisms in tuberculosis. Infect Immun, 2012. 80(10): p. 3343-59.

54. Qu HQ, SP Fisher-Hoch and JB McCormick, Knowledge gaining by human genetic studies on tuberculosis susceptibility. J Hum Genet, 2011. 56(3): p. 177-82.

55. Wellcome Trust Case Control Consortium, Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature, 2007. 447: p. 661-678.

56. Thye T, FO Vannberg, SH Wong, et al., Combined analysis of tuberculosis genome-wide association studies reveals a novel susceptibility locus on chromosome 8q11.2. Nat. Genet., 2010. 42: p. 739-741.

57. Teo YY, KS Small and DP Kwiatkowski, Methodological challenges of genome wide association analysis in Africa. Nat. Rev. Genet., 2010. 11: p. 249-160.

58. Thye T, E Owusu-Dabo, F Vannberg, et al., A novel susceptibility locus for tuberculosis on chromosome 11p13 downstream the WT1 gene. Nat. Genet., 2012. 44: p. 257-259.

59. Wang X, NL Tang, CC Leung, et al., Association of polymorphisms in the Chr18q11.2 locus with tuberculosis in Chinese population. Hum Genet, 2013.

60. Ji LD, PF Chai, BB Zhou, et al., Lack of association between polymorphisms from genome-wide association studies and tuberculosis in the Chinese population. Scand J Infect Dis, 2013. 45(4): p. 310-4.

61. Mahasirimongkol S, H Yanai, T Mushiroda, et al., Genome-wide association studies of tuberculosis in Asians identify distinct at-risk locus for young tuberculosis. J Hum Genet, 2012. 57(6): p. 363-7.

62. Png E, B Alisjahbana, E Sahiratmadja, et al., A genome wide association study of pulmonary tuberculosis susceptibility in Indonesians. BMC Med Genet, 2012. 13: p. 5.

63. Dudbridge F, A Gusnanto and BP Koeleman, Detecting multiple associations in genome-wide studies. Hum Genomics, 2006. 2(5): p. 310-7.

64. Newport MJ and CN Rotimi, Reducing the Global Genomic Inequity Gap: Development of an African Genome Project. Public Health Genomics, 2009. 12: p. 251-252.

65. Bodmer W and C Bonilla, Common and rare variants in multifactorial susceptibility to common diseases. 2008. 40: p. 695-701.

66. Iyengar SK and RC Elston, The genetic basis of complex traits: rare variants or "common gene, common disease"? Methods Mol Biol, 2007. 376: p. 71-84.

67. Ma X, Y Liu, BB Gowen, et al., Full-exon resequencing reveals toll-like receptor variants contribute to human susceptibility to tuberculosis disease. PLoS One, 2007. 2(12): p. e1318.

68. Thuong NT, SJ Dunstan, TT Chau, et al., Identification of tuberculosis susceptibility genes with human macrophage gene expression profiles. PLoS Pathog., 2008. 4: p. e1000229.

69. Kumar D, L Nath, MA Kamal, et al., Genome-wide analysis of the host intracellular network that regulates survival of Mycobacterium tuberculosis. Cell, 2010. 140(5): p. 731-43.

70. Gagneux S, K DeRiemer, T Van, et al., Variable host-pathogen compatibility in Mycobacterium tuberculosis. Proc Natl Acad Sci U S A, 2006. 103(8): p. 2869-73.

71. Thwaites G, M Caws, TT Chau, et al., Relationship between Mycobacterium tuberculosis genotype and the clinical phenotype of pulmonary and meningeal tuberculosis. J. Clin. Microbiol., 2008. 46: p. 1363-1368.

72. Intemann CD, T Thye, S Niemann, et al., Autophagy gene variant IRGM -261T contributes to protection from tuberculosis caused by Mycobacterium tuberculosis but not by M. africanum strains. PLoS Pathog, 2009. 5(9): p. e1000577.

73. Pareek M, J Evans, J Innes, et al., Ethnicity and mycobacterial lineage as determinants of tuberculosis disease phenotype. Thorax, 2013. 68(3): p. 221-9.

74. Caws M, G Thwaites, S Dunstan, et al., The influence of host and bacterial genotype on the development of disseminated disease with Mycobacterium tuberculosis. PLoS Pathog., 2008. 4: p. e1000034.

75. Westby M and E van der Ryst, CCR5 antagonists: host-targeted antivirals for the treatment of HIV infection. Antivir Chem Chemother, 2005. 16(6): p. 339-54.