Chapter 1

General Introduction
Introduction

Tuberculosis (TB) exerts a strong selection pressure on human evolution. As a causative agent, Mycobacterium tuberculosis has been hailed to be co-evolving with the current human race since its origin from Africa (Hirsh et al., 2004). The advent of chemotherapy in the early 1980’s has done wonders for tuberculosis afflicted human race and has significantly brought down the mortality rate. Yet an increase in number of reported cases in the recent times combined with the emergence of HIV and multiple drug resistance tuberculosis (MDR-TB) was a serious setback for the worldwide efforts to eradicate this highly successful pathogen. Despite such close association and selection pressure not the entire human race is infected with the bacteria but as many as one-third of the world population is latently infected with the bacillus (Raviglione et al., 1995) and of these only 5-10% have the lifetime risk of progressing to active disease (Kaufmann et al., 2006).

What could explain this inter-individual variation? We could reason that apart from natural selective factors and environment the genetic makeup of the host could play a significant role in determining the outcome of infection. Variations in the genetic makeup in the form of polymorphisms could lead to such inter-individual variation. More specifically the varying degree of immune response central to counter mycobacteria, might get affected due to changes in the immune response genes by means of genetic polymorphisms and contribute to a scenario where by the virtue of their genetic makeup most of the individuals mount an effective immune response and are able to either clear or contain the mycobacteria and a certain few fail to do so. This line of thought brought us to host genetic bias in multifactorial diseases such as tuberculosis. Evidence in support of the notion comes from clustering of with higher concordance of tuberculosis in monozygotic versus dizygotic twins (Comstock, 1978), the ethnic clustering of the disease with higher prevalence of tuberculosis in individuals of recent African descent (Stead et al., 1990), as well as the demonstration of both common polymorphisms and rare mutations which confer susceptibility to mycobacterial species in humans (Doffinger et al., 2002). These studies suggest that in addition to unique environment and natural selective factors host genetic factors are associated with TB which is perhaps ethnically governed.

Recent times have seen a multitude of studies emerging on genetic pretext of tuberculosis due mainly to two reasons: 1) sequencing of the entire human genome (Venter et al., 2001) led to identification of sequence polymorphisms which were previously unknown and this information led to generation of
databases such as dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP) and Hapmap (http://hapmap.ncbi.nlm.nih.gov) which have facilitated the selection and evaluation of yet unexplored variants, 2) the advent of high throughput genotyping technologies such as Sequenom massArray platform (Sequenom Inc., USA) and Illumina platform for sequencing of these variants in a short span of time. Keeping the above information in perspective has facilitated the identification of a range of genetic variations which include variations in the innate and acquired immune factors capable of identifying persons who are genetically prone from those who are not. The studies evaluating the involvement of innate immune response have focussed on receptors on macrophages such as toll-like receptors including (TLR2, TLR4, TLR8), P2X7 receptor, VDR, SP110, SLC11A (formerly NRAMP1), IRGM, NOS2 and DC-SIGN (CD209) (Möller and Hoal, 2010) and the adaptive immune response characterized by mainly cytokine and certain chemokine gene polymorphisms. The cytokines of note that have been studied include variants of IFNG, TNFA, IL1B, IL1RA, IL18, IL8, IL12, IL8 and TNFB (Yim and Selvaraj, 2010).

The role of genetic variants has also been exemplified in other associated respiratory ailments including asthma (Smolonska et al., 2009), and chronic obstructive pulmonary disease (COPD) (Haukim et al., 2002). The variants studied in COPD are limited and restricted to majorly inflammatory process. While the variants of asthma are also been related to inflammatory pathways these studies have been more extensive. A closer look reveals a similar selection of inflammatory gene variants for tuberculosis also; one putative reason might be the involvement of an initial inflammatory phase in all the three diseases. The Indian population also has had its share of genetic studies on tuberculosis. Majority of them are from south India (Yim and Selvaraj, 2010) and very limited studies from north India (Abhimanyu et al., 2011).

Recently a database called the Indian genome variation database (IGVDB) (Indian genome variation consortium, 2008) has been developed on the lines of Hapmap database (which includes the variation frequencies typed in five world populations but that does not include the Indian population) which have identified some sequence polymorphisms and typed them in samples from all over India and facilitated the classification of the Indian population into structured subgroups. This database has shown that the gene pools of north and south Indians differ significantly in their genetic makeup which means variations posing a risk in south Indian studies need not be valid for north India. This argument is further substantiated by the fact that the genetic susceptibility has an ethnicity bias (Stead et al., 1990).
Taking into context the abovementioned facts we planned this study to account for the apparent void due to the lack of the north Indian representation in the field of genetics of tuberculosis susceptibility. We adopted a population based case-control association of candidate genes study design for this purpose. Genetic studies can be linkage based analysis or candidate gene based association studies which further can be family based, population based or genome-wide study. All these approaches have their advantages and disadvantages (discussed in review of literature). To talk about our study design, one of the major reasons for selecting this study design was its feasibility, ease of patient recruitment (as compared to linkage, family based or a genome-wide scale) and cost–effectiveness in a limited setting. Also, such studies have greater power than linkage studies and can detect effects of minor variation in a given population. One of the major criticisms of this study design is of achieving false/confounding associations due to either population stratification or multiple testing corrections (Abhimanyu and Bose, 2012). We in the present study have accounted for each of these points before interpreting the results thus abrogating any possibility of confounding association.

We selected two most frequent forms of the tuberculosis disease i.e. the pulmonary and the lymph node tuberculosis as disease groups. The role of genetic polymorphisms in LNTB had not been addressed from any part of the world and ours is a first attempt at it.

The genetic variants encompassing both innate and adaptive arms of immunity to tuberculosis were selected. We attempted at maximum and meaningful representation of the genes and their variants, so as to attempt to understand their apparent genetic effect in governing susceptibility to tuberculosis and the variant manifestation as pulmonary TB and LNTB.

We also mapped the serum cytokine network and tried to identify biomarkers for the current population and especially for the management of LNTB where diagnosis is very cumbersome and involves primarily a high index of suspicion. A biomarker would aid in speedy diagnosis in such cases.

A panel of 25 genes with 112 SNPs in them have been addressed in the current study. Twelve genes belonged to the adaptive panel of cytokine genetic variants accounting for 64 SNPs. We also correlated the serum cytokine levels with the genetic polymorphisms in them. In the present study we were able to identify 11 novel variants implicated in tuberculosis susceptibility. We also examined 25 variants of *SP110* gene for susceptibility to TB and could for the first time identify a variant for LNTB susceptibility in north Indians. The rest 23 variants were from the relatively well mapped genes including toll-like receptors including (*TLR2, TLR4, TLR8, P2RX*, receptor, *VDR, SP110, SLC11A* (formerly*Nramp1*), *IRGM* (Immunity related GTPase M), *NOS2A* (Nitric oxide synthase), *LTA4H*)
Chapter 1

(leukotriene 4 hydroxylase) and DC-SIGN (CD209).

Overall, in the proposed study a comprehensive attempt was made to understand the influence of genetic makeup of north Indians on development of tuberculosis in general and differential manifestation as pulmonary TB and lymph node TB in particular. The panel of innate and adaptive immune response genes were explored for variations at sequence level and a comparative analysis of the respective cytokine profile vis–a–vis genetic variation was taken up to understand the influence of the genetic variations (genotypes) on the cytokine profile of the subjects.
AIMS AND OBJECTIVES

1. Selection of a panel of sequence variations in the genes related to innate and adaptive immune response.

2. To study the sequence variation in the panel of genes in the study groups comprising of healthy volunteers, pulmonary (PTB) and lymph node tuberculosis (LNTB) patients.

3. Profiling of serum levels of the cytokines relevant to the immune response genes included in the study.

4. To examine the correlation of the serum cytokine levels with the corresponding genotypes within the study population.

5. Analysis of the sequence variants and genotype of the study groups for possible relevance to the disease type.
The presented thesis deals in a very systematic manner with the quest for mapping the prevalence of genetic variants in tuberculosis in the north Indian population. A few studies were available from north India addressing the genetic aspect of tuberculosis immunity when the study was planned. Most of the earlier works dealt with the study of HLA genes and effect of their variants. Now, that reports have started emerging, it is a very encouraging sign and very exciting to be a part of it. Through our work we have tried to cover and address the importance of as much genes as possible to clearly investigate and elucidate the role of the genetic variants in various components of tuberculosis immunity.

To start with Chapter 1 introduces the topic in a precise manner stating the aims and objectives, presented as introduction. Chapter 2 deals with the nuisances of tuberculosis immunopathogenesis, population genetic studies, the dos and don’ts for such studies, specifying the available methods and our choice of method and other relevant details in order to provide a clear understanding of the topic, presented as review of literature. In chapter 3 lays out detailed description of the methods used from the choice of the genes and SNPs (Objective 1) to the details of and primers used for genotyping and genotyping methods used. The inclusion and exclusion criteria of patients, the sample collection and processing (Objective 2) is also explained along with the detailed genetic and statistical methods used. The results presented in Chapter 4 deals with the effects of the innate immune variants and their relevance in pulmonary and extrapulmonary tuberculosis, identifying risk alleles for north Indians. Continuing our quest in Chapter 5, we report here the variants in the adaptive immune panel comprising of 64 variants in the 12 cytokine genes and determining their relevance to tuberculosis and identification of the risk alleles. Most of the investigated variants included here have been studied for the first time in tuberculosis susceptibility (Objective 5). Chapter 6 deals with profiling of 12
General Introduction

serum cytokines in pulmonary and lymph node tuberculosis cases as compared to healthy controls in order to identify serum biosignatures to aid the diagnosis of tuberculosis, especially lymph node TB where the only reliable diagnostic method is cytology based which can be sometimes indeterminate delaying the treatment of these patients. The result presented in this chapter identifies elevated mean serum levels of IL-10, TNF-β and IL-8 as biomarkers for LNTB and elevated mean serum level of IL-6 as biomarker for PTB (Objective 3). In chapter 7 we tried to correlate the serum cytokine level obtained in the preceding chapter to the corresponding genotypes in pulmonary and lymph node TB. By this comparison we aimed at establishing the effect of genotypes on the varying serum cytokine levels so as to demonstrate the reason for varying cytokine response of the individuals to tuberculosis. In the results presented, we could indeed detect that certain varying genotypes correlated well with the change in serum cytokine levels which show that genotypes can indeed influence the serum cytokine levels (Objective 4). In the following chapter 8 we explored the effect of gene – gene interaction among cytokine genes and their effect on overall risk of tuberculosis.

This is followed by a conclusion and summary of the whole thesis with principal findings to conclude our investigations and also discussing a way forward.