Seminars in Immunology xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Seminars in Immunology

journal homepage: www.elsevier.com/locate/ysmim

Review

Host genetic studies in adult pulmonary tuberculosis

Christian G. Meyer*, Thorsten Thye

Bernhard Nocht Institute for Tropical Medicine, Bernhard Nocht Str. 74, 20359 Hamburg, Germany

ARTICLE INFO

Keywords:
Tuberculosis
Genetic epidemiology
Candidate gene studies
Genome-wide association studies
Selection
Missing heritability

ABSTRACT

Early observations, candidate gene studies and, more recently, genome-wide association studies have shown that susceptibility to tuberculosis has a host genetic component. Because the value of candidate gene studies has been doubted due to major limitations such as lack of sufficient power and small study groups, lack of reproducibility in independent groups and, often, ambiguous or even contrasting results in attempts of replication, much hope and expectancy has been put on the progress the genome-wide association approach has created. However, much less than initially expected became clear by the results obtained in genome-wide studies, emphasizing the need of increasing sample sizes, e.g. through meta-analyses, and of increasing the density of genetic variants studied across the human genome. A further reason why a rather low number of associated genetic variants were identified to date in infectious diseases in general and tuberculosis in particular might be the fact that selection acts strongly in diseases that affect the reproductive success. As in most genome-wide association studies performed so far, significant signals, often most likely surrogate marker only, have been found in non-coding regions of genomes, the identification of truly causative genetic variation and of the functionality of associated factors needs urgent attention. In the following we briefly discuss genetic studies in tuberculosis and describe new technologies that are currently employed in the search for responsible genetic elements involved in tuberculosis susceptibility.

1. Introduction

Infectious diseases which affect the biological fitness and which are fatal before the reproductive age are the most potent selective forces shaping the genetic architecture of human populations [1]. For the last decades, an aspect of growing interest in infectious disease research has been to decrypt and decode the host-genetic background of susceptibility to these diseases [2]. New technologies and statistical tools in the analysis of the human genome now allow for more detailed resolution of genes involved and the assessment of the functional relevance of genetic variation. The aim of this research is to identify genes involved in pathways that are relevant to infections in order to design targeted interventions. The salient question is: Why is a proportion of individuals susceptible to an infection with a pathogenic organism and to the disease caused by that pathogen, while others are resistant to infection and are protected from the disease?

2. Genetic epidemiology: host genetic factors in tuberculosis

Valuable instruments to respond to the question of genetic susceptibility are provided by the means of genetic epidemiology. As early as 1982, genetic epidemiology has been addressed as a scientific discipline aiming at describing the interplay of the etiology, occurrence, and control of diseases among related individuals and to understand disease heritability in larger communities and in populations [3].

The theory of inheritance of infectious disease susceptibility is based on early observations made in family clusters and in adults, their biological children and their adoptees [4]. Heritability as well as socio-environmental effects on mortality due to infectious diseases, including tuberculosis in particular have been studied. The individual risks of fatal outcomes due to the same infectious cause were significantly higher among biological children than among adoptees (odds ratio [OR] 5.81). These findings were considered strong indicators that a genetic context of infection phenotypes exists [4].

The postulate of heritability of tuberculosis susceptibility was corroborated by a vaccination accident in Lübeck, Germany in 1928, which became notorious. A virulent strain of *Mycobacterium*

* Corresponding author. Tel.: +49 40 42818 501; fax: +49 40 42818 512.
E-mail address: c.g.meyer@bni.uni-hamburg.de (C.G. Meyer).

http://dx.doi.org/10.1016/j.smim.2014.09.005
1044-5323/

tuberculosis was erroneously dispensed and given as a tuberculosis vaccine to newborns. Out of 251 vaccinated infants, manifest tuberculosis occurred in 212 of the children and 77 newborns died from the disease, while 39 of them did not develop any clinical symptoms of tuberculosis [5]. Notably, all children had received equal doses of the virulent mycobacteria. This vaccine disaster contributed also to the idea that genetically controlled innate and adaptive immune responsiveness might control the development of the tuberculosis infection phenotype.

Subsequently, in the attempt to more precisely define the genetic fraction of tuberculosis phenotypes, comparisons of monozygotic with dizygotic twins were conducted. The authors of these early studies were well aware of the necessity of very similar or, if possible, identical backgrounds of exposure to the pathogen as well as comparable environmental and social conditions. In monozygotic twins, the concordance of tuberculosis was with 66% substantially higher than among pairs of dizygotic twins, where the concordance was 23% only. Concordance rates among non-twin siblings were comparably high as those of dizygotic twin pairs [6–8]. Notably, the requirements of the study designs were at that time not as stringent as they are today, giving rise to substantial flaws of these early studies. These flaws were largely addressed and overcome by re-analyses of the twin study data in the "Prophit Survey" [9,10]. In addition to the confirmation of significant differences in the concordance rates between mono- and dizygotic twins, environmental and other factors (common residence of monozygotic twins, higher proportion of female twins and of positive sputum results of index cases, higher tuberculosis rates among parents of monozygotic twins) were also confirmed as factors contributing to the causality. Thus, the concept of tuberculosis as a partly "genetic" disease was reinforced and later supported by the observation of ethnic differences in tuberculosis susceptibility [11]. Following these early observations and studies, the field of genetic epidemiology and statistical genetics developed rapidly [12].

3. Candidate gene studies in tuberculosis

In studies of host-genetic factors in tuberculosis a major focus was laid primarily on candidate gene studies, rather than on family-based genome-wide linkage analyses which had proven more useful in the exploration of gene variants exhibiting strong effects [13,14]. In contrast to single Mendelian disorders which, due to the strong effects exerted by genetic variants, are well accessible to linkage analyses, in complex diseases rather modest and low effects of variants on the phenotypes studied are expected. Thus, the statistical power and the resolution potential of linkage studies applying a limited set of markers only is clearly of limited value in polygenic conditions [15–18]. Another approach to dissect susceptibility to mycobacterial diseases is based on the analysis of mendelian susceptibility to mycobacterial disease (MSMD) in candidate gene studies of individuals susceptible to generalized infections such as BCGitis and other intramacrophagic pathogens. Although interesting insights were gained through theses studies which largely, but not exclusively, concentrated on detailed analyses of genes involved in the interferon-gamma receptor (IFNGR) pathway, they did not contribute much to the resolution of population-wide susceptibility to tuberculosis. Reasons for the lack of convincing evidence might be that only a limited number of selected pedigrees were studied and that the genetic variants identified to be important in MSMD, in particular rare mutations, were not looked at on a large scale [19]. Most importantly, however, is the fact that genetic variability of the IFNGR pathway might indeed not be of major relevance in susceptibility to tuberculosis (Thye T, Meyer CG, personal communication of yet unpublished data).

Reasons for the preference of candidate gene studies were, among others, the inherent need of several multi-case tuberculosis families or a large number of affected sib pairs in linkage analyses, and methodological requirements. The approach of candidate gene studies has been pursued for many years both in communicable and non-communicable diseases, and numerous studies of tuberculosis candidate genes have been published to date. A largely comprehensive compilation of candidate gene studies performed in tuberculosis is given in Ref. [12], indicating ORs, significance levels and the sizes of study groups. Although many phenotypes such as pulmonary, extrapulmonary, childhood tuberculosis, primary and reactivated tuberculosis exist, most genetic studies have focused on adult pulmonary tuberculosis. Candidate gene studies mostly aim at relating a defined disease phenotype to bi-allelic variation of genes, consisting of single nucleotide polymorphisms (SNPs), but also to other types of mutations such as deletions, insertions and copy number variation. In most of these studies, variants in coding sections, the promoters of genes and in splice sites have been examined.

Candidate gene studies are based on the knowledge or, at least, a conceivable assumption of the function of the gene in the phenotype investigated. Thus, candidate gene studies are largely driven by an a priori hypothesis [20]. This type of study includes validation and confirmation of results obtained in previous candidate gene studies. Further confirmation of a significant result in a candidate gene study may also be attempted among other ethnic groups, in related phenotypes or even with related pathogens.

After an association of the mouse *Nramp1* gene with susceptibility to mycobacterial disease had been shown [21], *NRAMP1* (today: solute carrier family 11, *SLC11A1*, encoding the natural resistance-associated macrophage protein 1) was the first tuberculosis candidate gene that was examined in more detail [22]. SLC11A1 is involved in macrophage activation, the function of neutrophils and in other features of innate immunity. Many groups tried to confirm the initially observed association of tuberculosis susceptibility with *SLC11A1* variability. For example, it was shown that tuberculosis was linked to chromosome 2q35, including *SLC11A1*, in a large aboriginal Canadian family [23]. A meta-analysis looked at a series of studies and concluded that *SLC11A1* polymorphisms might indeed have a role in tuberculosis susceptibility [24]. In particular, the 3′-untranslated region (UTR) variant (rs17235416), the allelic variants D543N (rs17235409), INT4 (rs3731865) and the 5′-(GT)n variant (rs17235416) were analyzed and ORs were between 1.23 and 1.35, confirming the role of these variants as susceptibility factors. Stratification for different ethnic groups, however, rejected most of the associations observed. The immanent problem of small sample sizes of most of the study groups, as encountered in many candidate-gene studies, applied also to the studies included in the *SLC11A1* meta-analysis. Later studies have confirmed that *Nramp1* is of limited relevance in resistance to tuberculosis in mice [25].

Due to their high diversity and their role of antigen presentation and the subsequent induction of specific immune responses, the HLA molecules encoded by the major histocompatibility complex (MHC) have reasonably been considered as candidates for associations with infection phenotypes. A meta-analysis examined 22 studies of HLA antigens [26]. The B13 allele of the HLA class I B locus was in nearly all original studies and in the meta-analysis significantly associated with pulmonary tuberculosis ($P \leq 0.0001$). HLA DR3 and DR7 conferred relative protection from tuberculosis ($P = 0.002$ and $P \leq 0.0001$) and DR8 conferred an increased risk of tuberculosis ($P = 0.003$).

If a distinct finding in a candidate gene study is difficult or impossible to confirm or reject, e.g. due to the unavailability of a suitable replication group, results of candidate gene studies can be supplemented and further explained by functional experiments

Please cite this article in press as: Meyer CG, Thye T. Host genetic studies in adult pulmonary tuberculosis. Semin Immunol (2014), http://dx.doi.org/10.1016/j.smim.2014.09.005