Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA

SCIENZE PNEUMO-CARDIO-TORACICHE DI INTERESSE MEDICO E CHIRURGICO

Ciclo XXI

Settore scientifico disciplinari di afferenza**:** MED 10

TITOLO TESI:

IMMUNOGENETICS OF GRANULOMATOUS LUNG DISEASE:

A PHENOTYPIC APPROACH TO SUSCEPTIBILITY

Presentata da: Silvia Contini

Coordinatore Dottorato	**Relatore**
Prof. S. Mattioli	**Prof. C. Saltini**

Esame finale anno 2009

BACKGROUND

Immunogenetics

Immunogenetics is a scientific discipline that uses immunological methods to study the inheritance of traits. Traditionally, immunogenetics has been concerned with moieties that elicit immune response, that is, with antigens (antigenic determinants). It has now broadened its scope to study also the genetic control of the individual's ability to respond to an antigen. The immunological methods used in immunogenetics are of two principal kinds, serological and histogenetical. In serological methods, antibodies are used to detect antigens, either in solution or on a cell surface. In histogenetical methods, immune cells (lymphocytes) are used to detect antigens on the surface of other cells. In modern immunogenetics research, the serological and histogenetical methods are combined with molecular methods in which the researcher isolates and works with the genes that code for the traits. This approach of going back and forth from classical to molecular methods has proved to be very successful and has led to the elucidation of several complex genetic systems.

Granulomatous Lung Disease

Granulomatous lung diseases, such as sarcoidosis, hypersensitivity pneumonitis, Wegener's granulomatosis, and chronic beryllium disease, along with granulomatous diseases of known infectious etiologies, such as tuberculosis, are major causes of morbidity and mortality throughout the world. Clinical manifestations of these diseases are highly heterogeneous, and the determinants of disease susceptibility and clinical course (e.g., resolution vs. chronic, progressive fibrosis) are largely unknown. The underlying pathogenic mechanisms of these diseases also remain poorly understood. Within this context, these diseases have been approached using genomic and proteomic technologies to allow us to identify patterns of gene/protein expression that track with clinical disease or to identify new pathways involved in disease pathogenesis. The results from these initial studies highlight the potential for these "-omics" approaches to reveal novel insights into the pathogenesis of granulomatous lung disease and provide new tools to improve diagnosis, clinical classification, course prediction, and response to therapy. Realizing this potential will require collaboration among multidisciplinary groups with expertise in the respective technologies, bioinformatics, and clinical medicine for these complex diseases.

IMMUNOGENETICS OF TUBERCULOSIS

EVALUATION OF A MULTI-ANTIGEN TEST BASED ON B-CELL EPITOPES PEPTIDES FOR THE SERODIAGNOSIS OF PULMONARY TUBERCULOSIS

Abstract

SETTING: Two sample panels: (i) twenty pulmonary tuberculosis (TBp) patients and ten healthy subjects (HS) from a country with low incidence of TB (Italy) and (ii) forty-seven TBp patients and 26 HS from a country with high incidence of TB (Morocco).

OBJECTIVE: To identify a combination of *Mycobacterium tuberculosis* (*Mtb*) peptides useful for the serodiagnosis of active TBp.

METHODS: Fifty-seven B-cell epitopes peptides of *Mtb* were evaluated by immunoenzymatic assay and the data were analyzed using logistic regression analysis and Random Forest method.

RESULTS: The best discriminating peptide between TBp patients and HS from the sample of the country with low incidence of TB has been the 23 amino acid peptide of the Rv3878 protein. Thus, the sensitivity and specificity was 65 % and 100 %, respectively. In contrast, the same peptide showed 47% and 100% as sensitivity and specificity respectively in the country with high incidence of TB. In addition, the best peptides combination was a pool of nine peptides which showed a sensitivity of 70.2% and a specificity of 100% in the country with high incidence of TB.

CONCLUSIONS: The 9-peptides pool can be useful in identifying patients with active TBp

Keywords : Tuberculosis, peptides, serodiagnosis, B-cell epitope, ELISA

Introduction

Tuberculosis (TB) remains a major public health problem in most developing countries. The global annual incidence of tuberculosis has been estimated at 8 to 10 million cases, with approximately 1.7 million deaths (1). The control of the disease depends largely on early detection and the treatment of active cases.

At present, efforts at developing immunological tests are directed towards the identification of novel antigens that are associated with active disease. Thus the major objective of current immunological studies in TB is the identification of species-specific antigens and determination of the significance of corresponding immune responses. In this context, it is reasonable to hypothesize that proteins preferentially expressed by *Mtb* in models of intracellular growth and infection might be ideal targets for the design of new, highly sensitive diagnostic tests of *Mtb* infection and, eventually, for vaccine design (2, 3).

Serodiagnosis tests like enzyme linked immuno-sorbent assay (ELISA) are promising in view of their ease of performance and cost-effectiveness, many studies have been carried out by using quite complex antigens, such as whole bacteria, culture filtrates, bacterial extracts, and tuberculin or their purified protein derivatives (PPD) (4-6). However, a large variability in diagnostic accuracy has been reported, depending on the antigen employed and on the heterogeneity of the antibody response in TB patients(7).

The use of biochemical purification methods and recombinant DNA techniques of natural and recombinant proteins to obtain mycobacterial antigens is a challenging task with high cost and low specificities of the assays based on these antigens (8, 9). Therefore, we used synthetic peptides to assess their antibody responses in tuberculosis patients. Furthermore, the *in silico* approach based on screening of *Mtb* coding sequences from the proteins databases by using the prediction programs algorithms to identify B-cell epitopes is promising, since it reduce the cost of synthesizing peptides by reducing the number of peptides required for experimental evaluation for antibody reactivity (10, 11).

In this study, we assessed the potential use of the in-silico selection approach of M. tuberculosis-specific synthetic peptides in the serodiagnosis of active tuberculosis.

Materials and methods

Study population

To identify the immunological relevant peptides between the *in-silico* selected peptides, a first panel of sera from a country with low incidence of tuberculosis (Italy) were analysed. Specifically, 10 sera from skin test negative (TST-negative) healthy subjects without any history of TB exposure and 20 sera from active TB patients enrolled at Division of Respiratory Medicine of the University of Rome "Tor Vergata", at the "L. Spallanzani" Institute (Roma, Italy). The diagnosis of active TB was microbiologically confirmed by culture isolation of *Mtb* in all TB cases. Patient's details are shown in Table 1.

A second panel of sera from a country with high level of TB incidence has been used to evaluate the relevant peptide and the pool of peptides selected with the first study population. To this end, 47 patients with pulmonary active tuberculosis from Morocco were evaluated (Table 1). All patients underwent to smear and culture examination for *Mtb* according to standard procedures. Specifically, 31 Smear and culture positive (S+C+) patients and 16 smear-negative, culture-positive (S-C+) TB patients were evaluated in the study (table 1). Further, sera from twenty-six skin test positive healthy voluntary donors were collected from the Blood transfusion Regional Center (CRTS) of Rabat, Morocco.

All study subjects were found to be HIV negative. Sera from all patients were collected before anti-tuberculous treatment and were stored at -20°C until use.

The B-cell epitopes selection

Fifty-seven synthetic peptides containing potential B-cell epitopes were selected by *in-silico* strategy from 25 proteins of *Mtb*. Briefly, the sequence of *Mtb* proteins belonging to RD1 genomic region (12) and of proteins expressed during *Mtb* growth in human macrophage (2, 3), was obtained by the Swiss-Prot database (Table 2). Each single protein was screened for potential B-cell epitopes by using three computer algorithms of antigenicity: the Jameson-Wolf index, the Hopp-Woods method and the Parker index. These algorithms, available with Genrunner software [Hastings Software, Inc., Hastings, N.Y.] and Antheprot software (13-15), predict antigenic regions based on probability of surface exposure, local hydrophobicity, beta turns amino acid sequence, atomic flexibility and experimental HPLC retention times of synthetic peptides. Only protein regions identified as containing B-cell epitopes by at least two algoritms were selected.

Finally, the sequence of each epitope was screened for homology search of all known protein sequences by BLAST (http://www.ebi.ac.uk/blast/) through the selection of peptides that are uniquely restricted to the *Mtb*-complex (homology lower than 80% compared to other bacteria).

The sequence of all the selected peptides is reported on Table 2.

Peptides

The peptides were synthesized by Fmoc chemistry as free amino acid termini by ABI (http://www.abi.it/). All peptides were purified by reverse-phase chromatography (RPC). The purity of the purified peptides was higher than 90%. Sequence and purity were confirmed by mass spectrometry and analytical RPC. Lyophilized peptides were diluted in DMSO at stock concentrations of 10 mg/ml for each peptide and stored at – 80°C until use.

Peptide-based ELISA

Ninety-six wells high-binding capacity flat-bottom microplate (Greiner) was coated with 1 µg per well of each peptide in a Phosphate Buffered saline (PBS) buffer (pH 7.4) and kept at 4°C overnight. Then the plate was washed with 0.05% (v/v) Tween-20 in PBS and blocked with 3 % of bovine serum albumin (BSA, Sigma) for 1 hour. Each sample was assayed at a 1:50 dilution (2 hours, at room temperature). A horseradish peroxidase (HRP) conjugated anti-human IgG was used as the second antibody. The reaction was revealed using Ortho-Phenyl-Diamine (OPD)/H_2O_2 (Sigma) and quantified using an ELISA reader at 450 nm after stopping the reaction with 0.5M H_2S0_4. Blank wells were included as negative control in this assay.

Statistical analysis

Test results are presented as means ± standard deviation (SD) of the mean. ELISA tests were scored positive when individual readings were greater than two standards deviations above the optical density average of healthy subjects. Comparisons between different groups were made by Mann-Whitney test.

The best peptides discriminating between TB patients and controls sera were identified by using univariate logistic regression model using *Statistical Package for Social Science*™ (SPSS version 15), the Random Forests Method [http://genesrf.bioinfo.cnio.es/] (16) and stepwise multivariate logistic regression model using SPSS program.

The Receiver Operating Characteristic (ROC) curve was used to evaluate the performance of the ELISA tests with the two categories (TB patients and controls) ; the Area Under the Receiver Operating Characteristic curve (AUC) was used as a measure of diagnostic quality. The cut-off value used here is the one that leads to a ≥ 50% probability of being a case.

Results

In order to identify peptides capable of being recognized by TB patients, a first panel of TB and control sera from a low incidence country for TB was tested. Figure 1 shows the IgG response against each one of the fifty-seven selected peptides containing potential B–cell epitope (Figure 1). The antibody responses against the peptides were variable (figure.1). Univariate logistic regression method showed that the best discrimination score value between TB and healthy controls was 12.128 obtained with the peptide Pep51 of the Rv3878 protein (see for sequence table 2). With a 2SD cut-off value, the sensitivity and specificity was 65 % and 100% respectively and the area-under-the-ROC curve was 0.912 (Figure 2).

When multivariate logistic regression method and random Forest method were used, the best peptides combination identified was: Pep5 (Rv0747), Pep11 (Rv1114), Pep14 (Rv1434), Pep26 (Rv1979c), Pep42 (Rv3736), Pep48 (Rv3874 also known as CFP10), Pep51 (Rv3878), Pep55 (Rv3883) and Pep57 (Rv3883).

The peptides identified were then used in a second population from a country with high incidence of TB in order to evaluate their performances in the diagnosis of active TB.

The results of the ELISA based on the peptide Pep51 alone among the forty-seven TB patients and the twenty-six healthy controls showed a sensitivity and specificity of 47% and 100% respectively with a cut-off value at 2SD of the controls and the area-under-the-ROC curve was 0.842 (Figure 3). However, the results among the same population but with the ELISA based on the selected 9-peptides showed a higher value (0.938) of the area-under-the-ROC curve (P value=0.04) (Figure 4). Further, figure 5 shows the box plots of the antibody titer obtained with the second panel sera of TB and healthy controls. With a cut-off value at 2SD of the controls, the sensitivity and specificity was 70.2 % and 100% respectively (Figure4).

Discussion

The availability of mycobacterial genome sequences, the development of chemical synthesis techniques of peptides and the development of algorithms for the identification of immunodominant B-cell epitopes, might determine the rapid identification and testing of epitopes with diagnostic potential for active TB with the advantage of reducing the number of experimental evaluations for antibody reactivity. In fact, the use of ELISA test based on synthetic peptides would in principle circumvent the technical complexity and potential antigen variability associated with antigen purification. Further, synthetic peptides used as antigens have the added advantage of avoiding contamination with low-level impurities derived from cloning vectors required to produce recombinant proteins. Such impurities could decrease the specificity of the assay if they were cross-reactive with *M. tuberculosis* antigens. Finally, using synthetic peptides as the diagnostic antigen permit the uniformity and standardization of antigen preparations (17).

However, the main problem in setting immunodiagnostic systems is the identification of immunodominant antigen(s) specific of the active disease phase. In this context, previous studies (2, 3), have shown that *Mtb* is capable of expressing differential transcription programs in response to different environments. In particular, *Mtb* can express different gene sets in synthetic media compared to macrophage cultures. For instance, RD1 proteins are expressed both in early and late phase cultures, are variably expressed in *in vivo* in murine models, but are expressed at very low level in macrophage cultures (2, 3, 18-20).

With this background, we hypothesized that the *Mtb* proteins overexpressed in a model of *Mtb* infection in macrophage cultures could be the specific target of a strong immune response in active TB subjects. Thus, we combined the use of relevant *Mtb* antigen in a model of *Mtb* infection with *in silico* approach to identify B-cell epitopes to evaluate their potential use in TB serodiagnosis.

We were able to identify a peptide belonging to the *Mtb* protein Rv3878, as the best discriminating peptide between TB patients and healthy controls among a population of low exposure level of TB.

Rv3878 is one the proteins encoded by RD1, a region present in *Mtb* and virulent *M. bovis* genomes but missing from the DNA of all substrains of *M. bovis* Bacillus Calmette-Guerin (BCG) (12). Although our findings are not in agreement with previous observation by Brusasca et al (21) who reported that only 3–7% of human TB patients responded to RD1 region proteins, including Rv3878, it is well known that all RD1 antigens elicited a high antibody response in guinea-pigs infected with *Mtb* (21). Further, in a recent study (22), it has been shown that the Rv3878 protein elicited strong humoral immune response in TB patients in Indian population, supporting the notion

that the antibody response to a specific antigen often depends on the geographical location and ethnic background of the population being studied (23).

Several studies indicate that combinations of mycobacterial antigens are likely to have better diagnostic value (24, 25). Using both multivariate logistic regression analysis and random Forest method, we were able to identify a pool of 9-peptides beloning to the proteins Rv0747, Rv1114, Rv1434, Rv1979c, Rv3874, Rv3736, Rv3878 and Rv3883 as the best discriminating peptides combination between TB patients and healthy controls.

Recent reports indicate that the antigen Rv3874 is frequently recognized by antibodies in patients with TB, from regions where TB is not endemic (21) and from regions where TB is endemic (26). Further, Rv3883c or Mycosin-1 is an extra-cellular protein that is membrane- and cell wall-associated, and is shed into the culture supernatant. The protein is expressed after infection of Macrophages (27), but Rv3883C is not expressed in the attenuated *M. bovis* strain BCG, although the gene for mycosin-1 is present in the genome of this organism. This gene was found to be situated 3700 bp from the RD1 deletion region (28).

Moreover, recombinant Rv1979c has been identifies as an antigen capable to induce IFNγ in correlation with extent of lesions at necropsy (29). Finally, no information is available in the literature about the immunoreactivity of the other proteins identified as relevant in discriminating TB patient sera.

Both genetic and environmental factors also influence the antibody response. For example, the varied antibody response to *Mtb* is managed by HLA types. That fact, therefore, that our pool of peptides identified in an population has strong diagnostic value also for another population with a different genetic background, suggest that these peptides may form the basis of a test that can be used independent of location.

Our results, in accordance with those obtained by other investigators, confirmed higher sensitivity for the smear-positive patients than for the smear-negative patients ($P<0.001$), due probably to a greater exposure to antigens in patients with high bacillary loads (30, 31). However, in developing countries such as Morocco where TB disease is prevalent, serological tests might play a major role in the setting of the diagnosis of patients with suspect TB where acid-fast bacilli examination is negative. Our results showed that with the ELISA based on the peptide pool, TB can be diagnosed with 44% sensitivity in acid fast bacilli smear-negative TB patients. This results is interesting given the facts that the are no simple fast alternatives (e.g. culture takes 6 weeks) in these clinical settings, and even thought that x-ray is a standard tool of TB diagnosis it could be quit questionable in some cases (32).

Finally, it is worth to consider that in the study population with high TB prevalence we evaluated, all controls were TST-positive. Thus, the peptide based ELISA was able to distinguish between active and inactive tuberculosis infection, while other immunological based tests such as TST and gamma–interferon based assays are not capable (33, 34). Nevertheless it would be important to explore how the ELISA based on the peptide pool performs in recently exposed individuals as well as in subjects with cured tuberculosis or under chemotherapy treatment that are known to present high anti-*Mtb* IgG levels (35).

In conclusion, with the limitation of the small study population evaluated, this study indicates the applicability of the combination of nine well-defined synthetic peptides to the serodiagnosis of active TB. Their use in an ELISA would be important because of their stability, reproducibility and capability to identify strong humoral immune responses in patients with active pulmonary TB. The *Mtb*-complex specificity of these peptides strongly encourages further evaluation of this pool of peptides as a reagent for TB-specific immunodiagnostic assays on larger study populations including other disease groups such as patients with pulmonary disease other than TB, or seropositive patients for HIV.

REFERENCES

1. OMS. 2006. Global tuberculosis control: Surveillance, planning, financing. Genève.
2. Cappelli, G., E. Volpe, M. Grassi, B. Liseo, V. Colizzi, and F. Mariani. 2006. Profiling of Mycobacterium tuberculosis gene expression during human macrophage infection: upregulation of the alternative sigma factor G, a group of transcriptional regulators, and proteins with unknown function. *Research in microbiology* 157:445-455.
3. Mariani, F., G. Cappelli, G. Riccardi, and V. Colizzi. 2000. Mycobacterium tuberculosis H37Rv comparative gene-expression analysis in synthetic medium and human macrophage. *Gene* 253:281-291.
4. Chiang, I. H., J. Suo, K. J. Bai, T. P. Lin, K. T. Luh, C. J. Yu, and P. C. Yang. 1997. Serodiagnosis of tuberculosis. A study comparing three specific mycobacterial antigens. *Am J Respir Crit Care Med* 156:906-911.
5. Harboe, M., H. G. Wiker, and S. Nagai. 1992. Protein antigens of mycobacteria studied by quantitative immunologic techniques. *Clin Infect Dis* 14:313-319.
6. Raja, A., U. D. Ranganathan, R. Bethunaickan, and V. Dharmalingam. 2001. Serologic response to a secreted and a cytosolic antigen of Mycobacterium tuberculosis in childhood tuberculosis. *Pediatr Infect Dis J* 20:1161-1164.
7. Lyashchenko, K., R. Colangeli, M. Houde, H. Al Jahdali, D. Menzies, and M. L. Gennaro. 1998. Heterogeneous antibody responses in tuberculosis. *Infection and immunity* 66:3936-3940.
8. Mustafa, A. 2005. Progress towards the development of new antituberculosisvaccines. In *Focus on Tuberculosis Research*. e. Smithe LT, ed. Nova Publishers, New York. p. 47-76.
9. Mustafa, A. S. 2002. Development of new vaccines and diagnostic reagents against tuberculosis. *Mol Immunol* 39:113-119.
10. Hagmann, M. 2000. Computers aid vaccine design. *Science (New York, N.Y* 290:80-82.
11. Zerpa, N. C., A. Wide, J. Noda, H. Bermudez, R. Pabon, and O. O. Noya. 2006. Immunogenicity of synthetic peptides derived from Plasmodium falciparum proteins. *Experimental parasitology* 113:227-234.
12. Mahairas, G. G., P. J. Sabo, M. J. Hickey, D. C. Singh, and C. K. Stover. 1996. Molecular analysis of genetic differences between Mycobacterium bovis BCG and virulent M. bovis. *Journal of bacteriology* 178:1274-1282.
13. Deleage, G., F. F. Clerc, and B. Roux. 1989. ANTHEPROT: IBM PC and Apple Macintosh versions. *Comput Appl Biosci* 5:159-160.
14. Deleage, G., F. F. Clerc, B. Roux, and D. C. Gautheron. 1988. ANTHEPROT: a package for protein sequence analysis using a microcomputer. *Comput Appl Biosci* 4:351-356.
15. Deleage, G., C. Combet, C. Blanchet, and C. Geourjon. 2001. ANTHEPROT: an integrated protein sequence analysis software with client/server capabilities. *Comput Biol Med* 31:259-267.
16. Heidema, A. G., J. M. Boer, N. Nagelkerke, E. C. Mariman, A. D. van der, and E. J. Feskens. 2006. The challenge for genetic epidemiologists: how to analyze large numbers of SNPs in relation to complex diseases. *BMC Genet* 7:23.
17. Liang, F. T., A. C. Steere, A. R. Marques, B. J. Johnson, J. N. Miller, and M. T. Philipp. 1999. Sensitive and specific serodiagnosis of Lyme disease by enzyme-linked immunosorbent assay with a peptide based on an immunodominant conserved region of Borrelia burgdorferi vlsE. *Journal of clinical microbiology* 37:3990-3996.
18. Gao, Q., K. E. Kripke, A. J. Saldanha, W. Yan, S. Holmes, and P. M. Small. 2005. Gene expression diversity among Mycobacterium tuberculosis clinical isolates. *Microbiology (Reading, England)* 151:5-14.

19. Rogerson, B. J., Y. J. Jung, R. LaCourse, L. Ryan, N. Enright, and R. J. North. 2006. Expression levels of Mycobacterium tuberculosis antigen-encoding genes versus production levels of antigen-specific T cells during stationary level lung infection in mice. *Immunology* 118:195-201.
20. Shi, L., R. North, and M. L. Gennaro. 2004. Effect of growth state on transcription levels of genes encoding major secreted antigens of Mycobacterium tuberculosis in the mouse lung. *Infection and immunity* 72:2420-2424.
21. Brusasca, P. N., R. Colangeli, K. P. Lyashchenko, X. Zhao, M. Vogelstein, J. S. Spencer, D. N. McMurray, and M. L. Gennaro. 2001. Immunological characterization of antigens encoded by the RD1 region of the Mycobacterium tuberculosis genome. *Scand J Immunol* 54:448-452.
22. Mukherjee, P., M. Dutta, P. Datta, A. Dasgupta, R. Pradhan, M. Pradhan, M. Kundu, J. Basu, and P. Chakrabarti. 2007. The RD1-encoded antigen Rv3872 of Mycobacterium tuberculosis as a potential candidate for serodiagnosis of tuberculosis. *Clin Microbiol Infect* 13:146-152.
23. Weldingh, K., I. Rosenkrands, L. M. Okkels, T. M. Doherty, and P. Andersen. 2005. Assessing the serodiagnostic potential of 35 Mycobacterium tuberculosis proteins and identification of four novel serological antigens. *Journal of clinical microbiology* 43:57-65.
24. Gennaro, M. L. 2000. Immunologic diagnosis of tuberculosis. *Clin Infect Dis* 30 Suppl 3:S243-246.
25. Houghton, R. L., M. J. Lodes, D. C. Dillon, L. D. Reynolds, C. H. Day, P. D. McNeill, R. C. Hendrickson, Y. A. Skeiky, D. P. Sampaio, R. Badaro, K. P. Lyashchenko, and S. G. Reed. 2002. Use of multiepitope polyproteins in serodiagnosis of active tuberculosis. *Clinical and diagnostic laboratory immunology* 9:883-891.
26. Greenaway, C., C. Lienhardt, R. Adegbola, P. Brusasca, K. McAdam, and D. Menzies. 2005. Humoral response to Mycobacterium tuberculosis antigens in patients with tuberculosis in the Gambia. *Int J Tuberc Lung Dis* 9:1112-1119.
27. Dave, J. A., N. C. Gey van Pittius, A. D. Beyers, M. R. Ehlers, and G. D. Brown. 2002. Mycosin-1, a subtilisin-like serine protease of Mycobacterium tuberculosis, is cell wall-associated and expressed during infection of macrophages. *BMC microbiology* 2:30.
28. Brown, G. D., J. A. Dave, N. C. Gey van Pittius, L. Stevens, M. R. Ehlers, and A. D. Beyers. 2000. The mycosins of Mycobacterium tuberculosis H37Rv: a family of subtilisin-like serine proteases. *Gene* 254:147-155.
29. Cockle, P. J., S. V. Gordon, A. Lalvani, B. M. Buddle, R. G. Hewinson, and H. M. Vordermeier. 2002. Identification of novel Mycobacterium tuberculosis antigens with potential as diagnostic reagents or subunit vaccine candidates by comparative genomics. *Infection and immunity* 70:6996-7003.
30. Bartoloni, A., M. Strohmeyer, F. Bartalesi, D. Messeri, E. Tortoli, A. Farese, F. Leoncini, S. Nutini, R. Righi, A. Gabbuti, F. Mazzotta, and F. Paradisi. 2003. Evaluation of a rapid immunochromatographic test for the serologic diagnosis of tuberculosis in Italy. *Clin Microbiol Infect* 9:632-639.
31. Pottumarthy, S., V. C. Wells, and A. J. Morris. 2000. A comparison of seven tests for serological diagnosis of tuberculosis. *Journal of clinical microbiology* 38:2227-2231.
32. Linh, N. N., G. B. Marks, and A. B. Crawford. 2007. Radiographic predictors of subsequent reactivation of tuberculosis. *Int J Tuberc Lung Dis* 11:1136-1142.
33. Miguel, G., Madariaga, J. Ziba, and S. Susan. 2007. Clinical Utility of Interferon Gamma Assay in the Diagnosis of Tuberculosis. *J Am Board Fam Med* 20:540-547.
34. Ravn, P., M. E. Munk, A. B. Andersen, B. Lundgren, J. D. Lundgren, L. N. Nielsen, A. Kok-Jensen, P. Andersen, and K. Weldingh. 2005. Prospective evaluation of a whole-blood

test using Mycobacterium tuberculosis-specific antigens ESAT-6 and CFP-10 for diagnosis of active tuberculosis. *Clinical and diagnostic laboratory immunology* 12:491-496.

35. Amicosante, M., S. Barnini, V. Corsini, G. Paone, C. A. Read, Jr., P. L. Tartoni, M. Singh, C. Albera, A. Bisetti, S. Senesi, and et al. 1995. Evaluation of a novel tuberculosis complex-specific 34 kDa protein in the serological diagnosis of tuberculosis. *Eur Respir J* 8:2008-2014.

Table 1: Demographic characteristics of the study subjects

Study group		Number	Mean Age (years)	Sex-ratio (M/F)	Ethnicity
First panel	Pulmonary TB patients	20	39 (33-53)	1.22	Caucasians
	Healthy subjects (TST-)	10	30 (26-42)	1.00	Caucasians
Second panel	Pulmonary TB patients TPM+ TPM0	 31 16	 35 (25-45) 39 (28-51)	 2.1 2.2	North Africans
	Healthy subjects (TST+)	26	39 (34-48)	2.5	North Africans

TP = Pulmonary tuberculosis, TPM+ = Smear positive pulmonary tuberculosis, TPM0 = Smear positive pulmonary tuberculosis, M/F = male /female, TST+ = Tuberculin Skin Test positive, TST- = Tuberculin Skin Test negative.

Table 2: List of the peptides used in this study.

*Peptide epitopes starting from amino acid position 1 in the protein sequence where two Alanine were added to the N-terminus of the peptide in order to increase binding affinity to the epitope.

Protein	Peptide Number	Starting amino acid	Amino acid sequence
Rv0418	Pep1	at 1*	AAMVNKSRMMPAV
	Pep2	at 216	IPVVSVTKSVGFQLRGQSGPTTVK
	Pep3	at 377	AGIERTFVAYLKMAGKTAQDT
Rv0747	Pep4	at 1	MSWVMVSPELVVAAAADLAG
	Pep5	at 385	HTLQQDVINMVNDPFQTLTGRPLIG
Rv0755c	Pep6	at 107	VVHPAVVQANRVRTWLLAVSNVFGQ
Rv0934 (PAP)	Pep7	at 140	EHLKLNGKVLAAMYQ
Rv0956	Pep8	at 12	ARLVVLASGTGSLLR
	Pep9	at 176	ETLHERIKVTERRLLVAAVAALATH
Rv1114	Pep10	at 46	RDVVLDLLANLYQFPVV
	Pep11	at 60	FPVVTHDEVLRLVGRRRLWGRG
Rv1382	Pep12	at 8	GSLIFAAVLVMLIAVLARLMMRGWR
	Pep13	at 120	RGVAGKVVAGIGILAIRWRLPSGT
Rv1434	Pep14	at 1*	AAMRASPAERVDGAYAGAGPHTQSV
Rv1582c (RFO-10)	Pep15	at 143	SIYPIVQRELARQTGF
	Pep16	at 254	AIWRRIRVVPFEVVIPADEQD
Rv1866	Pep17	at 91	DRYRHLVALSITDFGAAGPRSSWRA
	Pep18	at 149	AAVQAAWAVLVAYFNRLRCGTGDY
	Pep19	at 222	DGYVRFCVMAPRQWRGLRRWLGEP
	Pep20	at 372	YPFEGLRILDLGIIVAGGELS
	Pep21	at 502	TRMGYGPLVRAATGVTRVWTSDEAQ
	Pep22	at 538	TTIFPDHVVGRVGALLALAALIHRD
	Pep23	at 568	AHVHISQAEVVVNQLDTMFVAE
Rv1979c	Pep24	at 138	LTFLGFIGVLLAINLFGNRAIKWAN
	Pep25	at 413	DKVLPLVAIVVSVGLAVSYD
	Pep26	at 466	HYYRIIRRVGDRPSTR
Rv1983	Pep27	at 348	NGIVTAPTAVNVVLLSIPTSPFAI
Rv1986	Pep28	at 36	LPVVALCTVSDIVLIAAGIAGFGA
	Pep29	at 102	TPVRLAEVLVTCA
	Pep30	at 165	GRLRGLFTNPGSW
Rv2994	Pep31	at 13	IMIVSLGVTASSFLFINGVAFLIPR
	Pep32	at 56	PSWGLVVTMFAWGYLLDHVGERMV
	Pep33	at 98	HSLLWIGVFLFLGGMAAGGCNS
	Pep34	at 204	SPYRGSSILWRIHAASALLMMP
	Pep35	at 228	VTVTFMLVWLINHHGWSVAQ
Rv3042c	Pep36	at 137	LELRVSVPPGCVGPLQIALTKVAAE
Rv3219	Pep37	at 68	RALKRRNARTKAR
Rv3619c	Pep38	at 28	QAIISDVLTASDF
	Pep39	at 72	QKVQAAGNNMAQT
Rv3736 (38 kD)	Pep40	at 24	TLLRAAGVRDQDV
	Pep41	at 34	QDVGNYDAFISIRAAIRAIESA
	Pep42	at 145	VALGVIRLLLGADYAPLAVH
	Pep43	at 226	GIVESVRTIVRQLLPTGAATLNV
	Pep44	at 269	TFVILVDRVRKDVADRYLR
Rv3872	Pep45	at 56	EGIQLLASNASAQ
Rv3873	Pep46	at 68	TPMVVWLQTASTQAKTR
Rv3874 (CFP 10)	Pep47	at 54	AVVRFQEAANKQK
	Pep48	at 74	TNIRQAGVQYSRA
Rv3875 (ESAT-6)	Pep49	at 2	QQWNFAGIEAAASAIQGNVTSIHSL
	Pep50	at 65	NALQNLARTISEA
Rv3878	Pep51	at 139	PRVVATVPQLVQLAPHAVQMSQN
	Pep52	at 216	GDVQPAEVVAAAR
Rv3879 (UF)	Pep53	at 135	TAINSLVTATHGA
	Pep54	at 325	PHVKPAALAEQPG
Rv3883c (MycP1)	Pep55	at 219	IDYAVNVKGVVVVVAAGNTGGDC
	Pep56	at 337	DDLVGAGVIDAVA
	Pep57	at 372	APYNVRRLPPPVVEP

Figure 1 : Graphical value representation of the optical density (OD)

obtained for each peptide and sera from the first panel by ELISA showing the heterogeneity of the antibody responses in TB patients and healthy subjects against *Mycobacterium tuberculosis* peptides.
Antibody level (OD): OD<0.100 ; 0.100≤OD<0.200 and 200≤OD.
The right column shows the score value for each peptide obtained by unvariate regression analysis.

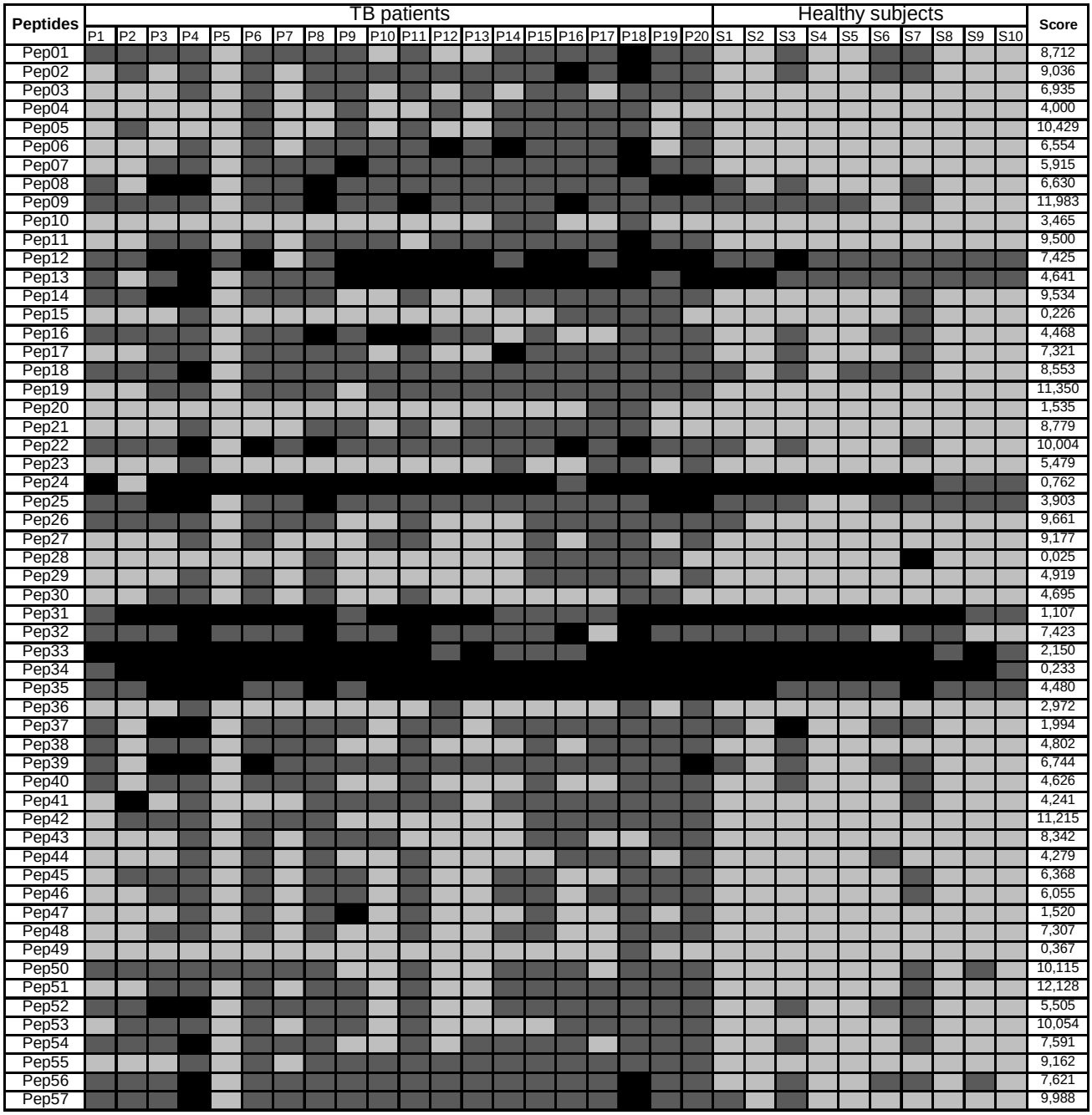


Figure 2: The ROC curve of the peptide Pep51 evaluated with the panel of TB and control sera from a low incidence country for TB. (AUC = Area under the curve)

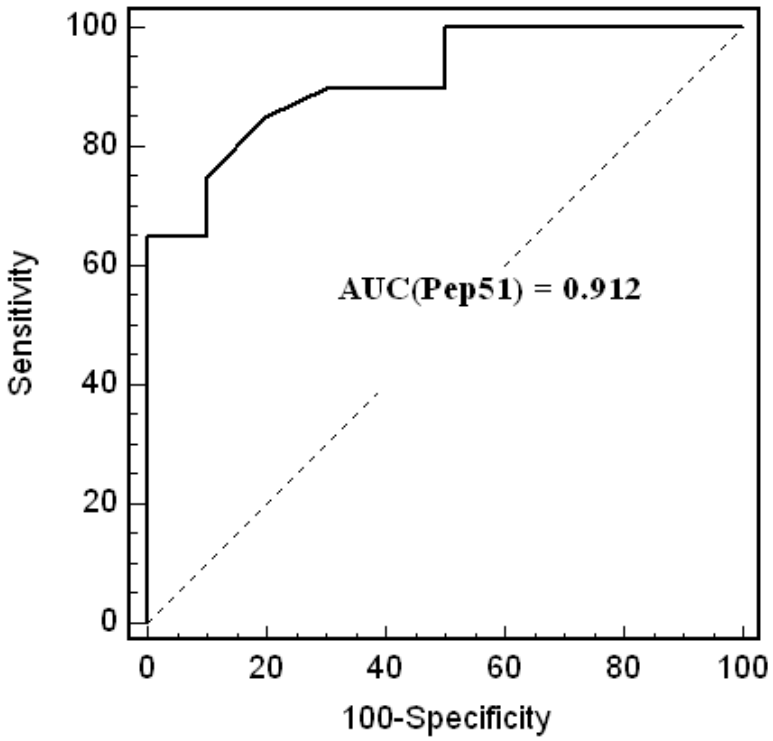


Figure 3: The ROC curves of the peptide Pep51 and the 9-peptides pool evaluated with the panel of TB and control sera from a high incidence country for TB. (AUC = Area under the curve)

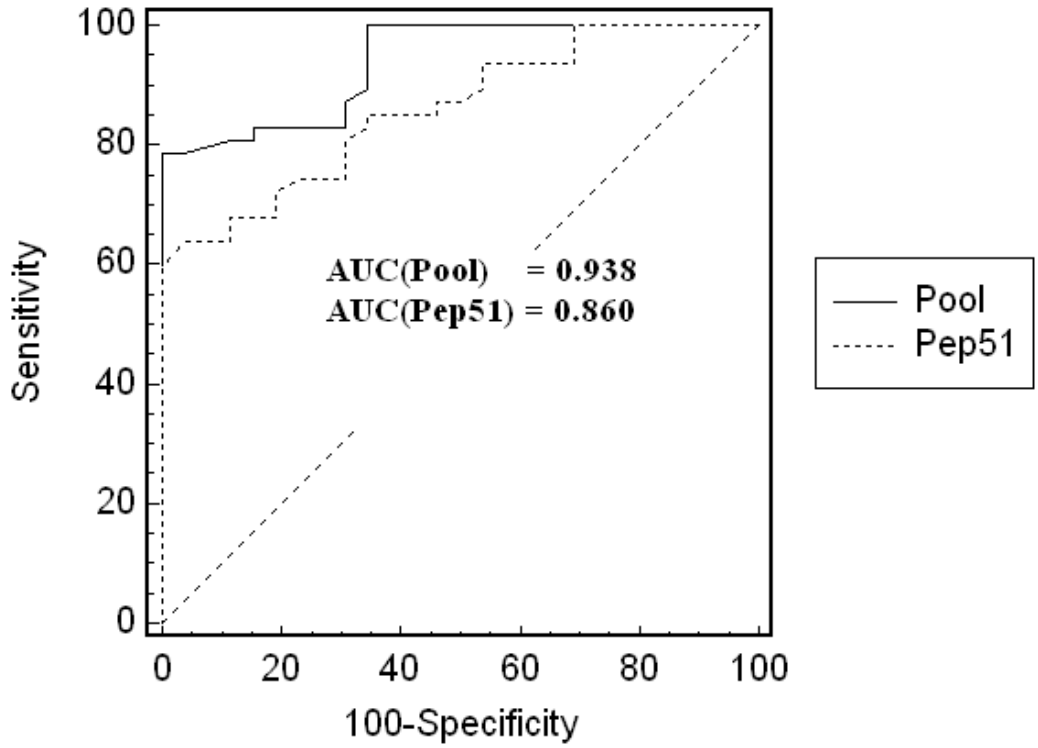
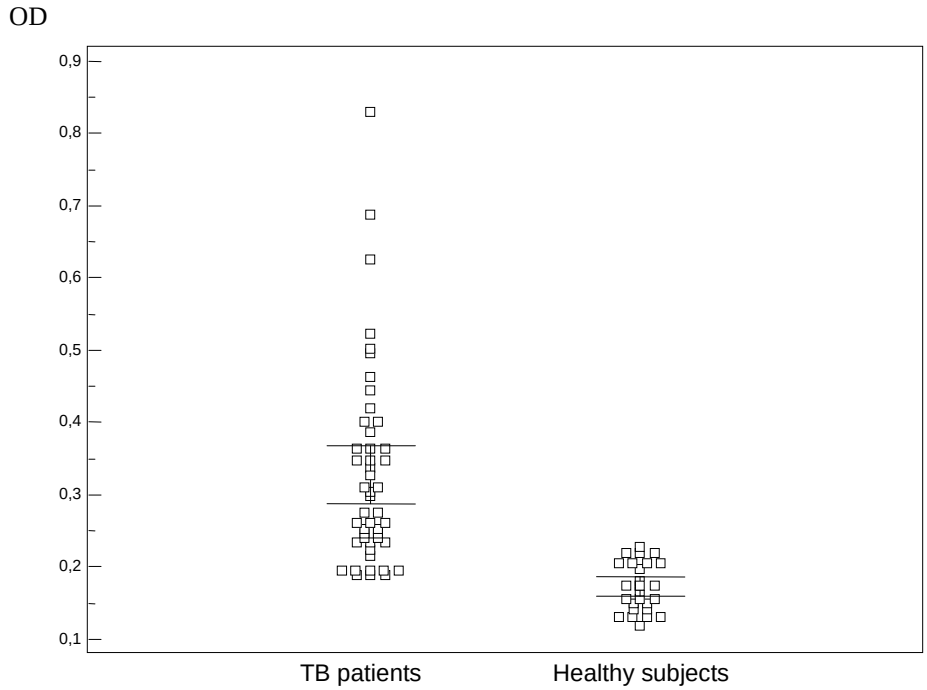


Figure 4: Box plot representation of the distribution of levels of antibodies (IgG) to the 9-peptides pool of *Mycobacterium tuberculosis* in TB patients and healthy subjects originated from a high incidence country for TB. (Error bars : 95% confidence interval for mean).



A MODEL OF PHENOTYPIC SUSCEPTIBILITY TO TUBERCULOSIS: DEFICIENT IN SILICO SELECTION OF MYCOBACTERIUM TUBERCULOSIS EPITOPES BY HLA ALLELES

Abstract.

HLA-DR allelic variants have been associated with tuberculosis (TB) susceptibility in different populations with risk ratios of 3.7 to 7.2.We hypothesized that the genetic susceptibility to TB depends upon the reduced capability of HLA-class II alleles of TB patients to bind and select peptide antigen from the Mycobacterium tuberculosis (MTB) expressed genome. To test this hypothesis, we developed a software that can predict HLA-DR restricted epitopes within the whole MTB genome based on quantitative peptide binding matrices.We analyzed the number of MTB epitopes recognized in two previously described populations of TB patients and matched controls and in a control population comprised of individuals affected by a sarcoid-like granuloma induced by beryllium and by healthy exposed controls. The number of putative epitopes within the whole MTB genome which could be bound by any HLA-DR allele (HLA-DR immunome of MTB) was 405,422 out of 1,304,277 possible 9-mers i.e., 31.08% of the global capability, instead of the expected 35%. When tested at an affinity level equivalent of the 1% of the best binder peptides, the HLA-DR alleles (HLA-DRB1*0801, *0802, *1401, *1501 and *1502) associated with TB susceptibility recognized a significantly lower mean number of MTB-epitopes (7,862±4,258) than the MTB-epitopes recognized by HLA-DR alleles (HLA-DRB1*0301, *0701, *1101, *1102, *1301 and *1302) negatively associated with TB (11,376±1,984, $p<0.032$). The number of epitopes bound at high affinity out of the whole MTB genome by the combination of the two HLA-DR alleles carried by each individual was lower in TB patients [TB-population 1: 11,341±908 (mean+SEM); TB-population 2: 15,303±657] than in matched healthy controls (CTR-population 1: 13,587±605, $p<0.03$ vs TB-population 1; CTR-population 2: 1,6841±555, $p<0.04$ vs TB-population 2). No difference was seen in individuals with the sarcoid-like granuloma induced by beryllium compared to the exposed healthy (beryllium-hypersensitivity: 17,593+447; controls 18,014±421; $p=0.57$). The data suggest that HLA-DR alleles associated with susceptibility to tuberculosis may be endowed with a reduced capability to bind at high affinity T-cell epitopes and select them for antigen presentation. The same alleles may contribute to determine the reaction to mycobacteria in non tuberculous granulomatous disorders. (Sarcoidosis Vasc Diffuse Lung Dis 2008; 25: 21-28).

Key words: tuberculosis, susceptibility, HLA, epitope prediction, T-cell response, sarcoid-like granulomas

Introduction

Classic tuberculosis (TB) epidemiologic studies indicate that the genetic background may play an important role in susceptibility to TB (1).With the expansion of molecular genetics studies, a large number of genes have been associated with TB, leading to think that susceptibility might be the result of an imbalance between the effects of susceptibility genes and of protective genes such as the natural resistance- associated macrophage protein Nramp1, the IFN-α receptor 1, the IL-12 receptor β1 genes and the HLA genes. In this regard, a number of allelic variants of the HLA locus which have been associated with TB risk in population studies indicating DQB1*05, *06 and DRB1*08, *14, *15 and *16 as susceptibility genes with risk ratios ranging from 3.7 to 7.2, and HLA-DRB1*03, *07, *11 and *13 as "protective" genes (2-10). HLA genes code for surface receptors that are known to play a pivotal role in the generation of antimicrobial immunity (11). HLA class II proteins (HLA-DP, -DQ and -DR) bind peptides derived from the digestion of microbes in the phago-lysosome of antigen presenting cells, carry them to the cell surface and present them to cytokine-producing CD4 T-cells. Structure-function studies of HLA class II molecules have indicated that the selection of antigenic peptides by the HLA receptors is dictated by the chemico-physical interaction between the amino acid side chains lining receptor-like pockets on the floor of the HLA antigen binding groove and the the agretopes i.e., the aminoacid side chains of the antigenic peptides (12, 13). As most of the polymorphisms generating allelic variability of the HLA molecules code for aminoacid changes in the peptide binding groove's pockets, each HLA allele will bind a unique set of aminoacid side chains hence selecting a discrete set of antigenic peptides for antigen presentation (14, 15). As a consequence, the ability of HLA alleles to select a peptide antigen repertoire from a given microbe for antigen presentation, hence to induce a protective immune response, may vary widely, leading to greater susceptibility to infection of the subjects carrying the HLA alleles less efficiently binding and presenting antigens (16). In this context though, since each individual coexpresses at least 2 HLA-DR molecules on the cell surface of the APC, it is reasonable to think that individual susceptibility to infection shall be determined by the capability of each subject's two HLA-DR molecules combined to recognize microbial antigen epitopes, rather than the carriage of a single susceptibility allele. Moreover, among the HLA-DR alleles, there is a predominant expression of HLA-DRB1 alleles, being expressed at a level five times higher than its paralogues DRB3, DRB4 and DRB5 (17-19). The assessment of HLA-associated susceptibility to TB, as model for susceptibility to granulomatous disorder mediated by mycobacteria, might thus require an analysis of disease associated phenotypes instead of disease associated alleles. To assess this hypothesis, we took advantage of bioinformatics tools allowing the

identification of antigenic peptides in whole microbial genomes by quantitative peptide binding motifs analysis for the HLA alleles (20). We developed software that can predict HLA-DR restricted epitopes in the whole MTB genome based on quantitative implemented peptide binding matrices and used this tool to determine the number of epitopes potentially recognized in the MTB genome in two already described populations of TB patients and matched healthy controls (21, 22). In addition, a population of patients affected by the sarcoid-like granulomatous reaction induced by beryllium and matched beryllium-exposed subjects (23).

Methods

Patients' characteristics: The study populations was composed by the TB patients and matched controls of two already described separate reports on the genetic susceptibility to TB in which HLA-DR high resolution typing was available for all study subjects (21,22). A population of patients with beryllium hypersensitivity and matched beryllium-exposed unaffected subjects were used as disease control population (23). They were 160 patients with tuberculosis (TBpopulation 1), and 200 controls (CTR-population 1) (see table 1) included in the Kim HS et al. study (22) and the 127 patients with tuberculosis (TB-population 2) and 120 matched controls (CTR-population 2) included in the Vejbaesya S et al. study (21) (see 22 A model of phenotypic susceptibility to Tuberculosis (table 1). The disease control population included 74 subjects with beryllium hypersensitivity (BeH) and 86 beryllium exposed matched controls (Be-CTR) from the Amicosante et al. study (23). The quantitative implemented peptide binding motifs are available only for 52 over more than 300 HLA-DRB1 alleles (24, 25). All together they may cover, with at least one allele, about 90% of the HLA-DR variability of different human populations (26). For the purpose of this study, only the subjects carrying both HLA-DR alleles with an available HLA-DR binding motif were used (see table 1). Specifically, 106 out of 320 (33%) control subjects and 85 out of 287 (30%) TB patients could be subjected to the analysis, as 168 (52%) controls and 151 (53%) TB patients had only one HLA-DR allele with an available binding motif, while 48 (15%) controls and 51 (18%) TB patients had both HLADR alleles without an available binding motif. Consequently, the alleles which were analyzed in the study populations were: HLA-DRB1* 0101, 0102, 0301, 0401, 0402, 0404, 0405, 0408, 0410, 0701, 0802, 0806, 1101, 1102, 1104, 1106, 1301, 1302, 1307, 1401, 1501, 1502. The selected subgroups did not differ for demographical characteristics from the subgroups of subjects excluded from the study for having one or both HLA-DR alleles without a known binding motif (data not shown). Genomes The genome of Mycobacterium tuberculosis H37Rv strain (NC 000962.2) composed of 4048 genes transcribed into 3,989 proteins and the genome of Escherichia coli K12 (NC 000913.2) were used in this study for the immuno-informatic analysis. Software for the identification and enumeration of epitopes in whole genomes To enumerate the T-cell epitopes present in data sets of proteins as large as a microbial genome, we eveloped a software for the identification and enumeration of peptide binding epitopes to HLA-DR molecules. This software was developed on LabView platform (National Instruments, US) using a graphic language. Basically, it is an open system in which two different databases are uploaded and crossed. Briefly, the first input database is represented by the protein sequences that can be uploaded from a file in FASTA format. The second database is represented by matrice(s) describing

the peptide binding capabilities of the HLA-DR alleles under analysis. The present version of the software is equipped with a set of 52 additional matrices for HLA-DR peptide binding profiles (15, 27).

The matrices database includes also the threshold values for different affinity levels as reported in the

original packages. The distribution of the matrix results of all the 209 possible peptides that can theoretically bound HLA-DR molecule is automatically generated as well as theoretical affinity thresholds. The analysis can be customized by selecting single/multiple proteins and HLA alleles among the set of data loaded, threshold and other parameters. The software generates all the possible nonamer peptides in a protein sequence and analyse them on the HLA-peptide binding matrices in analysis. For each protein and each HLA allele, all the peptides presenting a permissive 23 aminoacid (AA) in relative position P1 are stored in memory together with its matrix score and relative position in the protein. For the purpose of this work, two analysis were implemented in the software. A first analysis is represented by the identification of the epitopes recognised, in each single protein of the data set at the affinity threshold applied for each HLA-DR allele in analysis. This allows the enumeration of the epitopes in the whole data set and the identification and enumeration of proteins that present a defined number of epitopes, such as the proteins that are putatively not recognised by the HLA-DR allele at the affinity threshold used, as they present zero epitopes. A second specific analysis for the identification of the epitopes recognised in a set of proteins by 2 HLA-allelestogether has been developed to mime the situation of the HLA-DR recognition in a single subject. In this analysis, the software identify and enumerate the common and the different peptide epitopes recognised by the two HLA-DR alleles of the subject under evaluation by the position in the test set of proteins at the threshold of affinity applied. The whole MTB and E. coli genomes were analysed for the HLA-DR alleles negatively and positively associated to TB and for the enumeration of epitopes recognized by single subjects in analysis at the different thresholds of affinities equivalent to the 1%, 2%, 3%, 4% and 5% of the best binding natural peptides for HLA-DR alleles (27). The Threshold of affinity is a preselected numerical value used to differentiate between binders and non binders, any peptide frame scoring higher than this value is predicted as binder or vice versa; it correlates with the peptide score (15) and therefore with HLA-ligand interaction, therefore it is an indicator for the likelihood that predicted peptide is capable of binding to a given HLA-molecule. To express the results, we have chosen, the percentage of 1% in order to lower the false positive rate. The number of epitopes recognized by each study subject has been evaluated both as absolute number and as a relative number respect the amount of MTB epitopes in the whole genome. Statistical analysis All the data are expressed as

mean + standard deviation of the mean (SD). Comparisons between groups are made by Student's t test. Results 1. The immunome of MTB H37Rv HLA-DR molecules bind a core of nine aminoacids long protein fragments, when they carry in the relative position 1 (hereafter named P1) non-polar residues (I, L, M, F, W, Y, V) i.e., 7 out of 20 aminoacids or 35% of all the aminoacids. Analysis of the genome of M. tuberculosis H37Rv strain (NC_000962), which comprises 4,048 genes that can be transcribed into 3,989 proteins, allowed to estimate that, independently of the relative affinity of the epitopes for the HLA-DR molecules, the number of putative epitopes capable of binding any HLA-DR allele in the whole MTB genome was 405,422 out of 1,304,277 possible 9-mers, that we define as the HLA-DR MTB-immunome. This indicates that the HLA-DR MTB-immunome encompasses 31.08% of the whole MTB nonamer population. This number is equivalent to the 88% of the expected theoretical recognition that is equivalent to 35% of the all the peptides, thus suggesting that MTB encompasses a lower number of epitopes than expected if its genome presented a normal aminoacid distribution. Differently from MTB, the number of putative eepitopes capable of binding any HLA-DR allele in the E.coli genome was 537,294 out of 1,566,080 nonamers (34.30%). This is equivalent to 98% of the expected theoretical recognition, a fraction that is significantly higher than that of the MTB HLA-DR immunome ($p<0.0001$). 2. Impaired MTB proteins recognition by HLA-DR alleles associated with TB susceptibility. HLA-DR alleles HLA-DRB1*0801, *0802, *1401, *1501, *1502, which have been associated with TB susceptibility in previous studies, recognized significantly lower number of MTB-epitopes (7,862+4,258) than the HLA-DR alleles HLADRB1* 0301, *0701, *1101, *1102, *1301 and *1302, associated with TB resistance [11,376+1,984 ($P<0.032$)], at the affinity level of 1%. Consequently, there was a significantly higher number of MTB proteins (1,268+686) which could not be recognized by the HLA-DR alleles associated with TB susceptibility compared to HLA-DR alleles negatively associated with TB (776+232; $p<0.001$). 24 A model of phenotypic susceptibility to Tuberculosis 25 3. Phenotypic analysis of the ability of HLA-DR alleles of TB patients to bind whole MTB genome peptides. When this analysis was applied to those individual subjects carrying a pair of HLA-DR alleles with known peptide-binding motifs, the number of MTB epitopes recognized by TB patients in both populations was significantly lower than that recognized by controls [TB-population 1: 15,303±657, CTR-population 1: 16,841±555; $p=0.038$ compared to TB population 1; TB-population 2: 11,341±908, CTR-population 2: 13,587±605, $p=0.035$ compared to TB-population 2 (Figure 1 panel A)]. Interestingly, when a population of individuals affected by the sarcoid-like granulomatous reaction to beryllium was analyzed as a control, no differences where observed between granuloma-affected and unaffected subjects in their recognition ability of MTB genome peptides (beryllium-hypersensitivity: 17,593±447; beryllium-exposed controls 18,014±421;

p=0.608). Finally, when the the three population groups were tested for their ability to recognize E. coli genome peptides, no differences were seen between TB and their matched controls nor between Be-hypersensitive subjects and their matched controls (figure 1 panel B), suggesting that the epitope binding defect seen in TB patients was restricted to, or more pronounced for, the MTB genome.

Discussion

The binding of antigenic peptides by the host HLA proteins expressed by antigen presenting cells is thought to represents a limiting step in the development of an antimicrobial immune response. In the

context of the importance of HLA-DR genes in the immune response to MTB (28, 29) and of the observations positively or negatively linking different alleles of the HLA-DR, HLA-DQ, HLA-DP, and

HLA class I genes to TB susceptibility in HLA association studies (10), it is reasonable to think that altered peptide binding by susceptible HLA molecules may be the cause of susceptibility to disease (11).

With this as a background and in the context that TB susceptibility is increased in homozygous, compared to heterozygous twins (30), it is reasonable to hypothesize that a deficient antigen recognition capability of the immune system might be at the basis of the inefficacious response to MTB in susceptible individuals, and that the determinant of susceptibility to MTB infection and disease may be the combination of the recognition abilities of both HLA-DR alleles expressed by each subject, i.e., the HLA-DR phenotype. The finding of this in silico model that the number of MTB epitopes ecognized by the combination of the two HLA-DR alleles by TB subject was significantly lower than the matched Fig. 1. Total number of epitopes recognized by M. tuberculosis and E. coli. The number of M. tuberculosis (panel A) and E. coli (panel B) epitopes recognised in silico by each study subject with the combination of the two carried HLA-DR alleles at the threshold affinity of 1% in the two genomes. The three populations evaluated (TB-population 1, TB-population 2, and Be-exposed control population) are presented separately. Open circles, control subjects; closed circles, TB patients; grey circles Be-exposed control population. p value has been determined by Student's t-test. 26 S. Contini, M. Pallante, S. Vejbaesya, M. Hee Park, N. Chierakul, et al controls in both study populations suggest that TB patients present may be affected by a deficient capability of recognition of the MTB proteome compared to the non affected subjects in the population. Interestingly, Delgado et al. have recently reported a highly significant association between progressive pulmonary TB and homozygosity for HLADQ beta57-Asp alleles where a single polymorphism in the HLA-DQ beta chain played a critical role in the binding of ESAT-6, a highly immunogenic MTB protein and in the ensuing CD4+ T-cells immune response. Although they do not explain the mechanism of susceptibility to the development active TB, these data provided a functional link between an HLA polymorphism and susceptibility to progressive tuberculosis infection (31). In contrast to infection, current concepts are that in hypersensitivity and

autoimmune diseases susceptibility is associated with excessive HLA binding or the binding of specific (neo)-antigens by the HLA allele associated to the diseases (32).

In this context, it has been shown that the immune reaction to non-tuberculous mycobacteria can be characterized by an exaggerated reaction leading to the hypersensitivity pneumonias of the hot tub

lung (33), to the metalworking fluid-associated hypersensitivity pneumonitis (34), or to the formation of granulomas within the bronchial walls leading to the formation of bronchiectasis (35), a condition that's been associated with HLA-DR 6, i.e., the alleles 13 and 14 (36). Mycobacteria have also been implicated in sarcoidosis, where acid fast rods have been seen in affected tissues (37), wall-deficient form (l-form) of mycobacteria have been isolated (38), and mycobacterial DNA has been detected (39). Interestingly, a reaction to mycobacterial antigens has been described in sarcoidosis as T-cell and antibody responses to MTB ESAT-6 and KatG protein (40), HSP70 (41), and superoxide dismutase (SodA) (42) have been observed in patients with sarcoidosis and some antigens such as the Heat Shock Protein (HSP) and Catalase-Peroxidase (KatG) have been detected by immunohistochemistry in sarcoid tissues (43, 44). It is worth noticing in this regard, that the same HLA allelic variants which have been associated with susceptibility or resistance to tuberculosis have also been implicated in susceptibility to sarcoidosis. The HLA-DRB1*03 alleles, which have been negatively associated with TB, being therefore dubbed as "resistance" genes (45), have been associated with acute, self limiting, sarcoidosis (46). In contrast, the HLA-DRB1*15 alleles, which have been associated with the susceptibility to develop active TB (45), have been associated with stage III, or chronic sarcoidosis (47). Thus, one might hypothesize that the lack of HLA binding and presentation of mycobacterial antigens could determine disease progression in tuberculosis as well as in sarcoidosis, the type of reaction –extensively necrotizing versus non-necrotizing, being possibly driven by the expression of allelic variants of the host's genes of the innate immune response (48) or by variants expressed in the infected organism of the parasite virulence genes (49). The virtual approximation to truth made by in silico models, together with the limitation imposed by the reduced number of HLA-DR phenotypes that can be analyzed in the different populations, require that in silico results be confirmed by studies using conventional biological techniques. With this caveats, is counceivable that the use of immuno-informatic tools for the prediction of T-cells epitopes also on other HLA class I and II alleles on data sets as large as an the entire MTB genome, will help with generating mechanistic hypothesis on the determinants of HLA-associated genetic susceptibility to TB, as well as to other granulomatous disorders caused, or induced, by mycobacteria.

References

1. Grigg ER: The Arcana of Tuberculosis. Am. Rev. Tuberc. 1958, 78:426-53.
2. Vidal SM, Malo D, Vogan K, Skamene E, Gros P: Natural resistance to infection with intracellular parasites: isolation of a candidate for Bcg. Cell 1993, 73: 469-85.
3. Bellamy R, Ruwende C, Corrah T, McAdam KP, Whittle HC, Hill AV: Variations in the NRAMP1 gene and susceptibility to tuberculosis in West Africans. N Engl J Med 1998, 338: 640-4.
4. Greenwood CM, Fujiwara TM, Boothroyd LJ, Miller MA, Frappier D, Fanning EA, Schurr E, Morgan K: Linkage of tuberculosis to chromosome 2q35 loci, including NRAMP1, in a large aboriginal Canadian family. Am J Hum Genet 2000, 67: 405-16.
5. NewportMJ, Huxley CM, Huston S, Hawrylowicz CM, Oostra BA, Williamson R, Levin M: A mutation in the interferon-gamma-receptor gene and susceptibility to mycobacterial infection. N Engl J Med 1996, 335: 1941-9.
6. Jouanguy E, Altare F, Lamhamedi S, Revy P, Emile JF, Newport M, Levin M, Blanche S, Seboun E, Fischer A, Casanova JL: Interferongamma-receptor deficiency in an infant with fatal bacille Calmette-Guerin infection. N Engl J Med 1996, 335: 1956-61.
7. Altare F, Durandy A, Lammas D, Emile JF, Lamhamedi S, Le Deist A model of phenotypic susceptibility to Tuberculosis 27 F, Drysdale P, Jouanguy E, Doffinger R, Bernaudin F, Jeppsson O, Gollob JA,Meinl E, Segal AW, Fischer A, Kumararatne D, CasanovaJL: Impairment of mycobacterial immunity in human interleukin-12 receptor deficiency. Science 1998, 280: 1432-5.
8. de Jong R, Altare F, Haagen IA, Elferink DG, Boer T, van Breda Vriesman PJ, Kabel PJ, Draaisma JM, van Dissel JT, Kroon FP, Casanova JL, Ottenhoff TH: Severe mycobacterial and Salmonellainfections in interleukin-12 receptor-deficient patients. Science 1998, 280: 1435-8.
9. Baghdadi JE, Orlova M, Alter A, Ranque B, Chentoufi M, Lazrak F, Archane MI, Casanova JL, Benslimane A, Schurr E, Abel L: An autosomaldominant major gene confers predisposition to pulmonary tuberculosis in adults. J Exp Med 2006, 203: 1679 - 1684.
10. Kettaneh A, Seng L,Tiev KP,Toledano C, Fabre B, Cabane J: Human leukocyte antigens and susceptibility to tuberculosis: a meta-analysis of case-control studies. Int J Tuberc Lung Dis 2006, 10: 717-25.
11. Ghodke Y, Joshi K, Chopra A, Patwardhan B: HLA and disease. Eur J Epidemiol. 2005, 20: 475-88.
12. Davenport MP, Quinn CL, Valsasnini P, Sinigaglia F, Hill AV, Bell JI: Analysis of peptide-binding motifs for two disease associated HLA-DR13 alleles using an M13 phage display library. Immunology 1996, 88: 482-486.
13. Mallios RR: Predicting class II MHC/peptide multi-level binding with an iterative stepwise discriminant analysis meta-algorithm. Bioinformatics 2001,17: 942-8.
14. Germain RN: MHC-dependent antigen processing and peptide presentation: providing ligands for T lymphocyte activation. Cell 1994, 76: 287-99.
15. Sturniolo T, Bono E, Ding J, Raddrizzani L, Tuereci O, Sahin U, Braxenthaler M, Gallazzi F, Protti MP, Sinigaglia F, Hammer J: Generation of tissue-specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices. Nat Biotechnol 1999, 17: 555-61.
16. Scherer A, Frater J, Oxenius A, Agudelo J, Price DA, Gunthard HF, Barnardo M, Perrin L, Hirschel B, Phillips RE, McLean AR; Swiss HIV: Quantifiable cytotoxic T lymphocyte responses and HLA-related risk of progression to AIDS. Proc Natl Acad Sci U S A 2004,101: 12266-70.

17. Stunz L, Karr RW, Anderson R: HLA-DRB1 and -DRB4 genes are differentially regulated at the transcriptional level. J Immunol 1989, 143: 3081– 86.
18. Berdoz J, Gorski J,Termijtelen A, Dayer J, Irle C, Schendel D, Mach B: Constitutive and induced expression of the individual HLA-DR h and a chain loci in different cell types. J Immunol 1987, 139:1336–41.
19. Cotner T, Charbonneau H, Mellins E, Pious D: mRNA abundance, rather than differences in subunit assembly, determine expression of HLA-DRh1 and DRh3 molecules. J Biol Chem 1996, 264: 11107 –11.
20. De Groot AS, Bosma A, Chinai N, Frost J, Jesdale BM, Gonzalez MA, Martin W, Saint-Aubin C: From genome to vaccine: in silico predictions, ex vivo verification. Vaccine 2001, 19: 4385-95.
21. Vejbaesya S, Chierakul N, Luangtrakool K, Srinak D, Stephens HA: Associations of HLA class II alleles with pulmonary tuberculosis inThais. Eur J Immunogenet 2002, 29: 431-4.
22. Kim HS, Park MH, Song EY, Park H, Kwon SY, Han SK, Shim YS:Association of HLA-DR and HLA-DQ genes with susceptibility topulmonary tuberculosis in Koreans: preliminary evidence of associations with drug resistance, disease severity, and disease recurrence.Hum Immunol 2005, 66: 1074-81.
23. Amicosante M, Berretta F, Rossman M, Butler RH, Rogliani P, van den Berg-Loonen E, Saltini C: Identification of HLA-DRPhebeta47 as the susceptibility marker of hypersensitivity to beryllium in individuals lacking the berylliosis-associated supratypic marker HLADPGlubeta69. Respir Res 2005, 6: 94.
24. Marsh SGE: Nomenclature for factors of the HLA system, update January 1998. Tissue Antigens 1998, 51: 582-3.
25. Marsh SG:WHO Nomenclature Committee for Factors of the HLA System. Nomenclature for factors of the HLA system, update September 2006. Tissue Antigens 2006, 68: 540-2.
26. Wilson CC, Palmer B, Southwood S, Sidney J, Higashimoto Y, Appella E, Chesnut R, Sette A, Livingston BD: Identification and antigenicity of broadly cross-reactive and conserved human immunodeficiency virus type 1-derived helper T-lymphocyte epitopes. J Virol 2001, 75: 4195-207.
27. Singh H, Raghava GP: ProPred: prediction of HLA-DR binding sites. Bioinformatics 2001, 17: 1236-7.
28. Uma H, Selvaraj P, Reetha AM, Xavier T, Prabhakar R, Narayanan PR: Influence of HLA-DR antigens on lymphocyte response to Mycobacterium tuberculosis culture filtrate antigens and mitogens in pulmonary tuberculosis. Tuber Lung Dis 1999, 79: 199-206.
29. Sriram U, Selvaraj P, Kurian SM, Reetha AM,Narayanan PR: HLADR2 subtypes & immune responses in pulmonary tuberculosis. Indian J Med Res 2001, 113: 117-24.
30. Comstock GW: Tuberculosis in twins: a re-analysis of the Prophit survey. Am Rev Respir Dis 1978, 117: 621-4.
31. Delgado JC, Baena A, Thim S, Goldfeld AE: Aspartic acid homozygosity at codon 57 of HLA-DQ beta is associated with susceptibility to pulmonary tuberculosis in Cambodia. J Immunol 2006, 176: 1090-7.
32. Czaja AJ,Donaldson PT. Genetic susceptibilities for immune expression and liver cell injury in autoimmune hepatitis. Immunol Rev. 2000; 174: 250-9.
33. Embil J,Warren P, Yakrus M, Stark R, Corne S, Forrest D, Hershfield E: Pulmonary Illness Associated With Exposure to Mycobacterium-avium Complex in Hot Tub Water: Hypersensitivity Pneumonitis or Infection? Chest 1997; 111: 813-6.
34. Khan IUH, Selvaraju SB, Yadav JS: Occurrence and characterization of multiple novel genotypes of Mycobacterium immunogenum and Mycobacterium chelonae in metalworking fluids. FEMS Microbiology Ecology 2005, 54: 329–38.

35. Fujita J, Ohtsuki Y, Suemitsu I, Shigeto E, Yamadori I, Obayashi Y,Miyawaki H, Dobashi N, Matsushima T,Takahara J: Pathological and adiological changes in resected lung specimens in Mycobacterium Avium intracellulare complex disease. Eur Respir J 1999; 13: 535-40.
36. Kubo K, Yamazaki Y, Hanaoka M, Nomura H, Fujimoto K, Honda T, Ota M, Kamijou Y. Analysis of HLA antigens in Mycobacterium avium-intracellulare pulmonary infection. Am J Respir Crit CareMed 2000, 161: 1368-71.
37. Van?k J, Schwarz J: Demonstration of acid-fast rods in sarcoidosis. Am Rev Respir Dis 1970, 101: 395-400.
38. Almenoff PL, Johnson A, Lesser M, Mattman LH: Growth of acid fast L forms from the blood of patients with sarcoidosis. Thorax 1996, 51: 530-3.
39. Mangiapan G, Hance AJ.Mycobacteria and sarcoidosis: an overview and summary of recent molecular biological data. Sarcoidosis 1995, 12: 20-37.
40. Drake WP, Dhason MS, Nadaf M, Shepherd BE, Vadivelu S, Hajizadeh R, Newman LS, Kalams SA: Cellular recognition of Mycobacterium tuberculosis ESAT-6 and KatG peptides in systemic sarcoidosis. Infect Immun 2007, 75: 527-30.
41. Dubaniewicz A, Trzonkowski P, Dubaniewicz-Wybieralska M, Dubaniewicz A, Singh M, MyÊliwski A: Mycobacterial heat shock protein-induced blood T lymphocytes subsets and cytokine pattern: comparison of sarcoidosis with tuberculosis and healthy controls. Respirology 2007, 12: 346-54.
42. Allen SS, Evans W, Carlisle J, Hajizadeh R, Nadaf M, Shepherd BE, Pride DT, Johnson JE, Drake WP: Superoxide dismutase A antigens derived from molecular analysis of sarcoidosis granulomas elicit systemic Th-1 immune responses. Respir Res 2008; 9: 36.
43. S. Contini, M. Pallante, S. Vejbaesya, M. Hee Park, N. Chierakul, et al 43. Dubaniewicz A, Dubaniewicz-Wybieralska M, Sternau A, Zwolska Z, Izycka-Swieszewska E, Augustynowicz-Kopec E, Skokowski J, Singh M, Zimnoch L: Mycobacterium tuberculosis complex and mycobacterialheat shock proteins in lymphnode tissue from patients with pulmonary sarcoidosis. J Clin Microbiol 2006, 44: 3448-51.
44. Song Z, Marzilli L, Greenlee BM, Chen ES, Silver RF, Askin FB, Teirstein AS, Zhang Y, Cotter RJ, Moller DR: Mycobacterial catalase- peroxidase is a tissue antigen and target of the adaptive immune response in systemic sarcoidosis. J Exp Med 2005, 201: 755-67.
45. Kettaneh A, Seng L,Tiev KP,Toledano C, Fabre B, Cabane J: Human leukocyte antigens and susceptibility to tuberculosis: a meta-analysis of case-control studies. Int J Tuberc Lung Dis 2006, 10: 717-25.
46. Seitzer U, Gerdes J, Müller-Quernheim J: Evidence for disease phenotype associated haplotypes (DR.TNF) in sarcoidosis. Sarcoidosis. Vasc Diffuse Lung Dis 2001, 18: 279-83.
47. Rutherford RM, Brutsche MH, Kearns M, Bourke M, Stevens F, Gilmartin JJ: HLA-DR2 predicts susceptibility and disease chronicity in Irish sarcoidosis patients. Sarcoidosis Vasc Diffuse Lung Dis2004, 21: 191-8.
48. Burgner D, Jamieson SE, Blackwell JM: Genetic susceptibility to infectious diseases: big is beautiful, but will bigger be even better? Lancet Infect Dis 2006, 6: 653-63.
49. Rollenhagen C, Sörensen M, Rizos K, Hurvitz R, Bumann D: Antigen selection based on expression levels during infection facilitates vaccine development for an intracellular pathogen. Proc Natl Acad Sci U S A 2004, 101: 8739-44.

DESIGN OF IMMUNOGENIC PEPTIDES FROM M. TUBERCULOSIS GENES EXPRESSED DURING MACROPHAGE INFECTION

Abstract

In vitro diagnosis of MTB-infection uses MTB proteins coded for by genes of the region of differentiation 1 (RD1) of the MTB genome. This study wants to test if proteins preferentially expressed by during MTB intracellular growth might provide new targets for the diagnosis of MTB infection.

To this end Seventy-five multiepitopic HLA-promiscuous MTB-peptides were designed by quantitative implemented peptide binding motif analysis from 3 MTB-protein genes expressed in activated human macrophages (MA), 4 genes expressed during growth in non-activated human macrophages (MN-A), 12 housekeeping genes (HKG) and 6 genes of the RD1 region (RD1) as control. ELISpot for IFN- was performed to measure the responses of PBMCs deriving from 45 patients affected by active Tuberculosis and 34 controls. In active TB patients, the mean response to RD1 derived peptides was higher than that to either MA ($p<0.01$), MN-A ($p<0.008$) or HKG ($p<0.01$) derived peptides. In TST-positive subjects all selected peptides elicited significant IFN- T-cell responses ($p<0.02$ compared to TST-negatives), but without differences between the subgroups. Further, T-cell responses to RD1 peptides were lower in the 23 active-TB treated patients than in the untreated ones ($p<0.01$). The response to MA peptides in treated active-TB was higher than when untreated ($p<0.01$). These results demonstrate that the use of in vitro models of MTB-intracellular infection to select MTB gene products for further in silico and in vitro assessment of their immunogenicity has the potential to identify novel antigens amenable to the design of new tools for diagnosis and monitoring of tuberculosis.

KEYWORDS:
Tuberculosis, peptide binding motifs, ELISpot, macrophage-induced mycobacterial genes.

INTRODUCTION

Mycobacterium tuberculosis (MTB) is a facultative intracellular pathogen endowed with the ability to adapt to different growth milieus, either in cell culture or synthetic media, by altering its transcriptional profile[1,2,3]. The ability of MTB to evoke a strong type-1 immune response, hence allowing the rapid detection of MTB-specific interferon-gamma (IFN-γ producing T-cells, has been exploited to design *in vitro* diagnostic blood tests for MTB latent infection (LTBI)[4,5,6]. Following the demonstration that proteins coded for by genes of the region of differentiation 1 (RD1) of the MTB genome i.e., the early secreted antigenic target 6-kDa protein (ESAT-6) and the culture filtrate protein (CFP-10), can stimulate IFN- γ release by blood T-cells in 81 to 97% of tuberculosis (TB) affected individuals, IFN- γ -based blood tests have been designed for the diagnosis of LTBI[4,5]. In this context, it is reasonable to hypothesize that proteins preferentially expressed by MTB in models of intracellular growth and infection might be ideal targets for the design of new, highly sensitive diagnostic tests of MTB infection and, possibly, for vaccine design[1,2]. However, a strong limitation in evaluating the efficacy of T-cell response-based assays is represented by the restriction of the host HLA haplotypes[7]. The "reverse immune-genetics" approach by using the quantitative implemented HLA peptide-binding motifs has been successfully applied for the identification and selection of HLA-restricted peptide epitopes[8,9]. It is worth noticing that the peptide binding motifs are now available for a large number of HLA alleles covering more than 90% of the HLA haplotypes present in different human populations[4,10,11,12]. The present study was designed to determine whether the selection of MTB proteins upon their expression in different conditions of intracellular growth in *in vitro* models might help with the identification of immunogenic epitopes amenable to *in vitro* diagnosis of LTBI. To this end, a panel of HLA-restricted, promiscuous multiepitopic MTB-peptides, identified by quantitative implemented HLA peptide-binding motifs, has been evaluated in a study population including subjects of MTB exposed and not-exposed controls, BGC and not-BCG vaccinated, and patients with active TB.

MATERIALS AND METHODS

Study design

To measure the *ex vivo* immune response to the products of MTB genes expressed within activated and non activated human macrophages, or belonging to RD1 proteins, as control, a set of antigenic peptides were designed by bioinformatic analysis and tested in patients with active tuberculosis as well as in subjects infected and not with MTB. These responses were also measured in active tuberculosis during standard chemotherapy.

Study population

The study population included 22 patients with newly diagnosed, untreated active pulmonary TB and 23 TB patients undergoing standard treatment for 2 to 4 months, 10 of whom were evaluated both at presentation and after 2 months of treatment. Thirty four healthy individuals, 9 tuberculin skin test (TST)-negative subjects without any history of TB exposure (hereafter referred to as TB unexposed controls) and 25 TST-positive health-care workers, with a history of previous or current professional exposure (hereafter referred to as TBexposed controls) were also evaluated. Subjects were recruited at the Division of Respiratory Diseases of the University of Rome “Tor Vergata” at the Spallanzani Hospital of Roma

(Italy), the Forlanini Hospital of Rome (Italy) and at the National Center of Infectious and Parasitic Diseases (University Hospital for Lung Diseases "St. Sofia"; Sofia, Bulgaria) after informed consent. The diagnosis of active TB was confirmed by *M. tuberculosis* culture isolation in all cases. A detailed analysis of the main demographic features, including the BCG vaccination status and the TST results, of each study group is shown in Table 1. An EDTA-peripheral venous blood sample was collected from all the participants of the study. Epitope prediction and design of antigenic peptides Epitope prediction and design of antigenic peptides was performed by quantitative implemented HLA peptide-binding motif analysis as previously described [8] with minor modifications. Briefly, the sequence of MTB proteins belonging to the RD1 genomic region[13] and of proteins expressed during MTB growth in human macrophage[1,2], was obtained using the Swiss-Prot database. Each protein sequence was then analyzed by quantitative implemented HLA peptide-binding motif analysis using the ProPred database (http://www.imtech.res.in/raghava/propred/page2.html)[10], which allowed the identification and scoring of all putative epitopes for any of the 51 considered HLA-class II alleles. Finally, to verify the specificity to MTB, the sequence of each pre-selected peptide was screened for homology of all known protein sequences using BLAST (http://www.ebi.ac.uk/blast/) through the selection of peptides that are restricted to the MTB complex with an identity higher than 85% respect to other mycobacteria* and higher than 75% respect to

other microbes. Table 2 shows the list of the peptides evaluated in the study. Each selected peptide included: (i) one or more epitopes able to bind at least four different HLA-DR alleles, with a putative binding capability equivalent to that of the 3% of the best binding peptides for any allele belonging to a given HLA-class II specificity and (ii) a flanking region of 2 aminoacids at the N- and C-terminus to increase the peptide binding affinity[14].

Peptides

Designed peptides were synthesized by Fmoc chemistry as free amino acid termini with ABI technology (http://www.abi.it/). All peptides were purified by reverse-phase chromatography * At the time of the selection analysis the *Mycobacterium marinum, Mycobacterium ulcerans* and *Mycobacterium smegmatis* genomes were not available yet. However, the homolgy data presented in table 2 report also the identity to these mycobacteria although for some peptides the homology is higher than the 85% used as threshold during the selection steps. (RPC) to >90% purity. The sequence and purity were confirmed by mass spectrometry and analytical RPC. Lyophylized peptides were diluted in DMSO at stock concentrations of 10 mg/mL for each peptide and stored at – 80°C till use.

Cell isolation and culture conditions

Peripheral blood mononuclear cells (PBMCs) were obtained by Ficoll separation and resuspended in complete medium [RPMI, 10% heat inactivated fetal calf serum (FCS), 10 mM HEPES, 2 mM L-glutamine and 10U/ml penicillin-streptomycin, all from Euroclone Ltd, United Kingdom]. Then, $2x10^5$ PBMCs per well were incubated in duplicate in a 96-well plate and stimulated with: (i) 10μg/ml PPD (batch RT47; Statens Serum Institut, Copenhagen, Denmark); (ii) 5μg/ml Phytohemagglutinin (PHA) (Sigma, St Louis, MO, USA); and (iii) 18 pools of the 75 selected peptides. Briefly, to assess the T-cell response to a given peptide, 18 pools was prepared in a square matrix fashion set up with 9 rows and 9 columns pools. Each pool was prepared to contain 8 or 9 peptides (a part pool 9 that contained only 2 peptides) so that each single peptide (final concentration of 5μg/ml) was tested in two different pools, once in a row-grouped pool and once in a column grouped pool in order to determine the specific contribution to the T cell response of each single peptide as previously described[15]. Both unstimulated and DMSO-treated (at the same concentration present in the peptide solution) PBMCs were included as negative controls. Data were rejected if the sum of the peptide pool result in in columns and in rows of the square matrix presented a difference above 10%. Evaluation of peptide specific single T-cell IFN-γ release. The enumeration of peptide specific T-cell producing IFN-γ was performed at the single cell level using a ELISpot assay, as previously described[8]. The number of spot-forming cells (SFCs) was measured in each well using an EliSpot reader (AID EliSpot reader system ELR03; Autoimmune Diagnostika

GmbH, Strassberg, Germany). Evaluated spots had a size >15U (1U=500μm^2). The average number of SFCs of duplicate wells was then calculated and the value of appropriate controls subtracted from the SFCs of the stimuli. The absolute value of the SFCs for a single peptide was obtained by resolving the square matrix with the tested peptide pools as previously described[16]. Further, in some subjects the response to single peptides were tested individually to confirm the results of the matrix analysis method. An overall concordance above 90% between the methods were observed.

Statistical analysis Data are expressed as the mean + standard deviation (SD). Comparison between groups was made using Mann-Whitney tests. A *p* value below 0.05 was considered significant. All tests

were performed using the GraphPad Prism 4.0 (Graphpad software, San Diego, CA, USA) software package.

RESULTS

Peptide selection

Out of 25 MTB proteins, three expressed during growth in activated human macrophages (hereafter grouped as MA), four expressed during growth in non-activated human macrophages (hereafter grouped as MN-A), twelve coded for by constitutively expressed housekeeping genes (hereafter grouped as HKG) and six RD1 region proteins (hereafter grouped as RD1, which were used as comparators), 75 peptides were selected and used in this study (table 2). Specifically, 8 peptides were selected from the MA, 5 peptides from the MN-A, 47 from the HKG and 15 from the RD1 proteins respectively (Table 2). Each peptide contain one or more HLA-DR restricted epitopes (Table 2). Each peptide, in its overall sequence, is predicted to bind HLA-DR alleles belonging at least to three different serological groups of HLA-DR1, 3, 4, 7, 8, 11, 13 and 15, with an affinity equivalent or higher than the 3% of to the best binder epitopes known for the HLA-DR alleles themselves (Table 2). Further, BLAST analysis for these multiepitopic peptides reveals a sequence identity with other mycobacteria varing from 100% to lower than 50% with an identity respect to other bacteria and viruses lower than 75%. Analysis of human T-cell responses to *in silico* selected peptides To evaluate the cell-mediated immune response to *in silico* selected peptides, the number of IFN-γ producing peptide-specific T-cell precursors in PBMCs of subjects with active pulmonary TB, in comparison with MTB-exposed and non-exposed healthy controls were determined by ELISpot. Further, the effect of standard chemotherapy on the response against these peptides were assessed. In its overall, 16 out of 75 peptides (n: 6, 7, 21, 25, 26, 28, 29, 33, 34, 38, 39, 40, 41, 49, 59 and 66) did not show any significant difference of response between the three study groups. For the remaining 59 peptides, patients with active TB and TB-exposed controls presented a mean number of SFU/million PBMC significantly higher than the TST-negative controls (Figure 1). Analysis of T-cell responses to selected peptides in active TBpatients as positive control. The T-cell responses in active TB subjects presented a greater intensity of response compared to MTB-unexposed controls [TST-negative 1.4+2.1 mean spots per peptide/million PBMCs; active-TB 12.6+9.8; $p<0.006$ (Figure 1, panel A)]. Further, in active TB patients, the average intensity of the responses to either MA (7.1+3.5), MN-A (6.9+1.1) or HKG (8.2+3.2) derived peptides was lower that that of RD1 derived peptides (21.3+10.8; $p<0.01$ all comparisons) (Figure 1, panel A). Analysis of T-cell responses to selected peptides in TB-exposed subjects. TST-positive controls, elicited a significantly higher IFN-γT-cell responses against the *in silico* selected peptides (MA 7.6+3.5; MN-A 8.9+3.5; HKG 7.9+4.5; RD1 12.9+6.8) than TST-negative ($p<0.02$; Figure 1, panel B). As already shown with ESAT-6 and CFP-10 multiepitopic promiscuous peptides[15,17] the average

intensity of the T-cell responses to RD1 derived peptides was lower in TB-exposed TST-positive subjects (12.9+6.8) i.e., in the LTBIaffected, compared to active-TB patients (21.3+10.8; $p<0.003$). To the contrary, there was no difference in the response to MA, MN-A and HKG derived peptides in the TST-positive compared to the active TB patients affected ($p>0.05$, all comparisons). Analysis of T-cell responses to selected peptides during the course of active TB. In the context of the observation that the response to ESAT-6 and CFP-10 multiepitopic promiscuous peptides in active TB patients declines during chemotherapy[18], T-cell responses to the MA, MN-A and HKG derived peptides were also analysed in subgroups of untreated and of treated active-TB patients, as well as in a selected group of active TB patients which were tested before and after two months of standard chemotherapy. As expected, the T-cell responses to RD1-derived peptides were higher in the 22 untreated active-TB treated patients (mean total RD1 peptides SFC 251+116) compare to the 23 chemotherapy treated ones (mean total RD1 peptides SFC 155+129, $p<0.01$ compared to untreated patients). Strikingly, we found an increased response to MA peptides in chemotherapy treated patients (mean total MA peptides SFC 121.8+44.6) compared to the untreated (mean total MA peptides SFC 52+29; $p<0.01$), while no differences were found for the MN-A (untreated active-TB mean total MNA peptides SFC 38+12; active-TB treated 33+9; $p>0.05$) and HKG- derived peptides (untreated active-TB mean total HKG peptides SFC 389+45; active-TB treated 306+78; $p>0.05$). This observation was confirmed in the ten active TB patients which were followed up during the two months of chemotherapy treatment. In eight out of ten patients IFN-γ release in response to RD1-derived peptides markedly declined over time (figure 2 panel A, $p<0.02$). Instead, it did strikingly augment in 9 out of ten (figure 2 panel B, $p<0.01$) in response to MA-derived peptides. As above, no change was seen with MN-A or HKG-derived peptides ($p>0.05$, all comparisons).

DISCUSSION

The finding that the measure of IFN-γ-release as marker of type-1 immune response, in response to MTB-proteins is capable of identifying subjects with MTB-infection, either latent or active, has led to the production of a number of tests such as the RD1-derived antigen based ELISA "Quantiferon-TB" and the ELISpot "T-spot TB" tests. These tests, that are endowed with sensitivities ranging from 81 to 97% in the identification of patients with active MTB proliferation in tissue lesions[19], have spurred microbial immunology and molecular biology research into MTB immunogenic proteins. However, critical aspects in this context are represented by the choice of immunodominant and specific antigens and, for T-cell response evaluation, in the restriction imposed by the patients' HLA background. In this context, the use of synthetic peptides in TB has been successfully demonstrated for diagnostic applications in humans[20,21,22]. Overlapping synthetic peptides covering the entire sequence of a protein have been used to identify the peptides recognized by T-cells[20,23]. However, Since Th1 cells recognize mycobacterial antigens and epitopes in association with MHC class II molecules[24] an alternative approach has been to screen proteins for identification of regions that can associate with MHC molecules[25] and then test the peptides predicted to bind the MHC molecules for T-cell reactivity. Such approach has been shown to reduce drastically the number of peptides to be testes and thus the cost to screen for T-cell reactivity. Although several prediction programs have been proposed to identify peptides capable of binding to MHC class II molecules, recently a virtual matrix-based prediction server (ProPred) has been described, which can predict the binding capability of peptides to 51 HLA–DR alleles and has been shown to identify epitopes with a confidence of 95%[10]. Although these prediction systems are far to be fully precise, the result of this study confirm that the *in silico* selected HLA-promiscous multiepitopic area of proteins can be the region presenting the highest capability of immuno response[17], as shown comparing the result of this study with classic epitope mapping for some of the RD1 proteins such as ESAT-6, CFP-10, Rv3873 and Rv3878[26,27,28,29]. Furthermore, previous studies from our laboratories have shown that MTB is capable of expressing differential transcription programs in response to different environments. In particular, MTB can express starkly different gene sets in synthetic media compared to macrophage cultures. RD1 proteins are expressed both in early and late phase cultures, are variably expressed in *in vivo* in murine models, but are expressed at very low level in macrophage cultures[1,2,30,31,32,33].

With this background, we hypothesized that the MTB proteins expressed at the highest level in macrophage cultures could induce a stronger T-cell response in MTB infected subjects. The result of this study show that a group of peptides derived from a panel of MTB proteins selected for their expression in macrophage cultures are highly immunogenic, as they seem capable of eliciting Th1

T-cell responses in LTBI affected subjects that are as high as those elicited by RD1 derived peptides i.e., the present day diagnostic standard. A number of studies have analyzed the *in vivo* or *ex vivo* mycobacterial gene expression profile[17]. However, the inter-relationship between RNA and protein expression profiles, immunogenicity, either *in silico* or *in vitro* of the expressed proteins and the induction of Tcell responses has not been yet investigated in sufficient depth. Further, this study confirm the immunogenicity of the RD1 proteins in their overall and not only confined to ESAT-6 and CFP-10. In fact, only 4 out of the 15 RD1 derived peptides were designed on ESAT-6 and CFP10 and these peptides account only for 19-43% of the response to the whole RD1 peptides (data not shown). Furthermore, this study identified 19 proteins which elicit a strong T-cell response in LTBI and TB affected subjects as high as in RD1 peptides. However, few of these proteins have been characterised so far. Specifically, none of the MA proteins has been immunologically characterized. The MA proteins: Rv3042 and Rv1114; the M-NA proteins: Rv0198c, Rv1382 and Rv0755; the HKG proteins Rv1582, Rv0418, Rv1434, Rv1866, Rv0956 and Rv3219 have not yet been cloned and expressed for biochemical and functional studies[34]. The MA protein Rv3883c (MycP1) is a membrane-anchored mycosin serine protease[35] and Rv3042 (serB2) is a phosphoserine phosphatase[34]. The only characterized M-NA protein is Rv0125 (PepA), another protein belonging to the MTB serine protease gene family, is also known as MTB32A and has been found in culture filtrates. Consistent with our data, PepA is immunogenic as it has been shown to elicit memory T-cell responses in latent and active TB[36,37]. Among the 12 HKG proteins evaluated in the study, three are known to be immunogenic in an animal model and two in an *in silico* model. Rv1979c and Rv1986, that's belong to the RD2 gene region and Rv1769 that's belong to the RD14 gene region, have been shown to elicit effector T-cell responses in 59-86% of *M. bovis*-infected and/or BCG-vaccinated cattles[38]. Interestingly, two other HKG gene products, Rv1983 and Rv0747, belong to the PE-PGRS gene family[34] i.e., a multigene family that, together with the PPE gene family, makes up about 10% of the MTB genome and whose gene products comprise an extraordinarily high number of peptides displaying high-affinity for HLA class I alleles *in silico*[39] supporting the hypothesis of their immunogenicity and potential role in immunodiagnosis as suggested by our data by the reports on other proteins belonging to these multigene families[10,40,41]. Further, the comparison between the time course of the responses elicited by the MA, MN-A and the HKG derived peptides and that against RD1-derived peptides in active TB patients, whereby the response to RD1 peptides declines with the reduction of active lesion subsequent to treatment, may suggest that the IFN-γ response to certain RD1 peptides may be mycobacterial load, hence antigen concentration, dependent and may wane with the its rapid mycobacterial load reduction achieved by standard chemotherapy[15,18]. Interestingly, the response to MA, MN-A and the HKG derived

peptides seems to be more stable or may increase with time, as shown for MA-derived peptides. In conclusion, the results of our study suggest that promiscuous peptides capable of Th1 cell reactivity in HLA heterogeneous population can be predicted from the in silico HLA-DR
binding analysis of complete proteins overexpressed in model M. tuberculosis infections in macrophage. It has been predicted that the number of putative T-cell epitopes in M tuberculosis are about 1,500,000, which can be reduced effieciently reduced to few hundreds peptides by identifying proteins overexpressed during MTB infection in human and prescreening the coding sequences for HLA-DR binding using the prediction programs. This has the potential to identify novel antigens eventually be amenable to the design of new diagnostic tools.

REFERENCES

1. Cappelli G, Volpe E, Grassi M, Liseo B, Colizzi V, Mariani F. Profiling of Mycobacterium tuberculosis gene expression changes during human macrophages infection: up-regulation of the alternative sigma factor G, of a group of transcriptional regulators and of proteins with no known function. *Res. Microbiol* 2006;157:445-455.
2. Mariani F, Cappelli G, Riccardi G, Colizzi V. Mycobacterium tuberculosis H37Rv comparative gene-expression analysis in synthetic medium and human macrophage. *Gene* 2000;253:281-91.
3. McMurry J, Sbai H, Gennaro ML, Carter EJ, Martin W, De Groot AS. Analyzing Mycobacterium tuberculosis proteomes for candidate vaccine epitopes. *Tuberculosis* 2005;85:95-105.
4. Mazurek GH, Jereb J, Lobue P, Iademarco M.F, Metchock B, Vernon A. Guidelines for using the QuantiFERON-TB Gold test for detecting *Mycobacterium tuberculosis* infection. United States. *MMWR Recomm Rep* 2005;54:49-55.
5. Dewan PK, Grinsdale J, Liska S, Wong E, Robert Fallstad R, Kawamura LM. Feasibility, acceptability, and cost of tuberculosis testing by whole-blood interferongamma assay. *BMC Infectious Diseases* 2006;6:47-54.
6. Pai M, Riley LW, Colford JM. Interferon-gamma assays in the immunodiagnosis of tuberculosis:a systematic review. *Lancet Infect Dis* 2004;4:761-76.
7. Walker BD, Korber BT. Immune control of HIV: the obstacles of HLA and viral diversity. *Nature Immunol* 2001;2:473-475.
8. Seghrouchni F, Berretta F, Amicosante M. Selection of Microbial T-Cell Epitopes for Immune-Diagnosis, Immune-Therapy and Vaccine Design. *Current Pharmacogenomics* 2005;3:247-257.
9. Amicosante M, Gioia C, Montesano C, Casetti R, Topino S, D'Offizi G, Cappelli G, Ippolito G, Colizzi V, Poccia F, Pucillo LP. Computer-based design of an HLA-haplotype and HIV-clade independent cytotoxic T-lymphocyte assay for monitoring HIV-specific immunity. *Mol Med* 2002; 12:798-807.
10. Singh H, Raghava GPS. ProPred: Prediction of HLA-DR binding sites. *Bioinformatics* 2001;17:1236-37.
11. De Groot AS, Bosma A, Chinai N, Frost J, Jesdal BM, Gonzales MA, Martin W, Saint-Aubin CS. From genome to vaccine: in silico predictions, ex vivo verification. *Vaccine* 2001;19:4385-4395.
12. McMurry J, Sbai H, Gennaro ML, Carter EJ, Martin W, De Groot AS. Analyzing Mycobacterium tuberculosis proteomes for candidate vaccine epitopes. *Tuberculosis* 2005;85:95-105.
13. Mahairas GG, Sabo PJ, Hickey MJ, Singh DC, Stover CK. Molecular analysis of genetic differences between *Mycobacterium bovis* BCG and virulent *M. bovis*. *J Bacteriol* 1996;178:1274–82.
14. Berretta F, Butler RH, Diaz G, Sanarico N, Arroyo J, Fraziano M, Aichinger G, Wucherpfennig KW, Colizzi V, Saltini C, Amicosante M. Detailed analysis of the effects of Glu/Lys beta69 human leukocyte antigen-DP polymorphism on peptide-binding specificity. *Tissue Antigens* 2003;62:459-71.
15. Desjardin LE, Perkins MD, Wolski K, Haun S, Teixeira L, Chen Y, Johnson JL, Ellner JJ, Dietze R, Bates J, Cave MD, Eisenach KD. Measurement of sputum Mycobacterium tuberculosis messenger RNA as a surrogate for response to chemotherapy. *Am J Respir Crit Care Med* 1999;160:203-210.
16. Addo MM, Yu XG, Rathod A, Cohen D, Eldridge RL, Strick D, Johnston MN, Corcoran C, Wurcel AG, Fitzpatrick CA, Feeney ME, Rodriguez WR, Basgoz N, Draenert R, Stone DR, Brander C, Goulder PJ, Rosenberg ES, Altfeld M, Walker BD. Comprehensive epitope

analysis of human immunodeficiency virus type 1 (HIV-1)- specific T-cell responses directed against the entire expressed HIV-1 genome demonstrate broadly directed responses, but no correlation to viral load. *J Virol* 2003;77:2081-2092.
17. Rachman H, Strong M, Schaible U, Schuchhardt J, Hagens K, Mollenkopf H, Eisenberg D, Kaufmann SH. Mycobacterium tuberculosis gene expression profiling within the context of protein networks. *Microbes Infect* 2006;8:747-57.
18. Carrara S, Vincenti D, Petrosillo N, Amicosante M, Girardi E, Goletti D. Use of a T cell-based assay for monitoring efficacy of antituberculosis therapy. Clin. Infect. Dis. . 2006, 38:754-6.
19. Richeldi L. An Update on the Diagnosis of Tuberculosis Infection. *Am J Respir Crit Care Med* 2006; 174:736-42.
20. Ravn P, Demissie A, Eguale T, Wondwosson H, Lein D, Amoudy HA, Mustafa AS, Jensen AK, Holm A, Rosenkrands I, Oftung F, Olobo J, von Reyn F, Andersen P. Human T cell responses to the ESAT-6 antigen from Mycobacterium tuberculosis. *J Infect Dis* 1999;179:637-645.
21. Arend SM, van Meijgaarden KE, de Boer K, de Palou EC, van Soolingen D, Ottenhoff TH, van Dissel JT. Tuberculin skin testing and in vitro T cell responses to ESAT-6 and culture filtrate protein 10 after infection with Mycobacterium marinum or M. kansasii. *J Infect Dis* 2002;186: 1797–807.
22. Arend SM, Geluk A, van Meijgaarden KE, van Dissel JT, Theisen M, Andersen P, Ottenhoff TH. Antigenic equivalence of human T-cell responses to Mycobacterium tuberculosis-specific RD1-encoded protein antigens ESAT-6 and culture filtrate protein 10 and to mixtures of synthetic peptides. *Infect Immun* 2000;68:3314-21.
23. Vordermeier HM, Whelan A, Cockle PJ, Farrant L, Palmer N, Hewinson RG. Use of synthetic peptides derived from the antigens ESAT-6 and CFP-10 for differential diagnosis of bovine tuberculosis in cattle. Clin Diagn Lab Immunol 2001;8:571-8.
24. Mustafa AS. HLA-restricted immune response to mycobacterial antigens: relevance to vaccine design. *Hum Immunol* 2000;61:166-71.
25. Panigada M, Sturniolo T, Besozzi G, et al. Identification of a promiscuous T-cell epitope in Mycobacterium tuberculosis Mce proteins. *Infect Immun* 2002;70:79-85.
26. Agger EM, Brock I, Okkels LM, Arend SM, Aagaard CS, Weldingh KN, Andersen P.Human T-cell responses to the RD1-encoded protein TB27.4 (Rv3878) fromMycobacterium tuberculosis, Immunology. 2003;110: 507-12.
27. Mustafa AS, Shaban FA, Al-Attiyah R, Abal AT, El-Shamy AM, Andersen P, OftungF. Human Th1 cell lines recognize the Mycobacterium tuberculosis ESAT-6 antigen andits peptides in association with frequently expressed HLA class II molecules. Scand JImmunol. 2003;57:125-34.
28. Brock I, Weldingh K, Leyten EM, Arend SM, Ravn P, Andersen P. Specific T-cellepitopes for immunoassay-based diagnosis of Mycobacterium tuberculosis infection. J ClinMicrobiol. 2004;42:2379-87.
29. Mustafa AS, Al-Attiyah R, Sumaila N. Hanif M., Shaban FA. Efficient Testing of LargePools of *Mycobacterium tuberculosis* RD1 Peptides and Identification of Major Antigens and Immunodominant Peptides Recognized by Human Th1 Cells. Clin. Vaccine Immunol.2008;15: 916-924.
30. Gao Q, Kripke KE, Saldanha AJ, Yan W, Holmes S, Small P.M. Gene expression diversity among Mycobacterium tuberculosis clinical isolates. *Microbiology* 2005;151:5- 14.
31. Haile Y, Bjune G, Wiker HG. Expression of the mceA, esat-6 and hspX genes in Mycobacterium tuberculosis and their responses to aerobic conditions and to restricted oxygen supply. *Microbiology* 2002;148:3881-3886.

32. Rogerson BJ, Jung YJ, La Course R, Ryan L, Enright N, North RJ. Expression levels of Mycobacterium tuberculosis antigen-encoding genes versus production levels of antigen-specific T cells during stationary level lung infection in mice. *Immunology* 2006;118:195-201.
33. Shi L, North R, Gennaro ML. Effect of growth state on transcription levels of genes encoding major secreted antigens of Mycobacterium tuberculosis in the mouse lung. *Infect Immun* 2004;72:2420-2424.
34. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry III CE, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, HolroydS, Hornsby T, Jagels K, Krogh A, McLean J, Moule S, Murphy L, Oliver S, Osborne J, Quail MA, Rajandream MA, Rogers J, Rutter S, Seeger K, Skelton S, Squares S, Sqares R, Sulston JE, Taylor K, Whitehead S, Barrell BG. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature* 1998;393:537-544.
35. Dave JA, Gey van Pittius NC, Beyers AD, Ehlers MR, Brown GD. Mycosin-1, a subtilisin-like serine protease of Mycobacterium tuberculosis, is cell wall-associated and expressed during infection of macrophages. *BMC Microbiol* 2002;2:30.
36. Al-Attiyah R, Mustafa AS, Abal AT, El-Shamy AS, Dalemans W, Skeiky YA. In vitro cellular immune responses to complex and newly defined recombinant antigens of Mycobacterium tuberculosis. *Clin Exp Immunol* 2004;138:139-144.
37. Skeiky YA, Lodes MJ, Guderian JA, Mohamath R, Bement T, Alderson MR, Reed SG. Cloning, expression, and immunological evaluation of two putative secreted serine protease antigens of Mycobacterium tuberculosis. *Infect Immun* 1999;67:3998-4007.
38. Cockle PJ, Gordon SV, Lalvani A, Buddle BM, Hewinson RG, Vordermeier HM. Identification of novel *Mycobacterium tuberculosis a*ntigens with potential as diagnostic reagents or subunit vaccine. Candidates by comparative genomics. *Infect Immun* 2002;70:6996-7003.
39. Chaitra MG, Hariharaputran S, Chandra NR, Shaila MS, Nayak R. Defining putative T cell epitopes from PE and PPE families of proteins of Mycobacterium tuberculosis with vaccine potential. *Vaccine* 2005;23:1265-1272.
40. Choudhary RK, Mukhopadhyay S, Chakhaiyar P, Sharma N, Murthy KJ, Katoch VM, Hasnain SE. PPE antigen Rv2430c of Mycobacterium tuberculosis induces a strong B-cell response. *Infect Immun* 2003;71:6338-6343.
41. Okkels LM, Brock I, Follmann F, Agger EM, Arend SM, Ottenhoff TH, Oftung F, Rosenkrands I, Andersen P. PPE protein (Rv3873) from DNA segment RD1 of Mycobacterium tuberculosis: strong recognition of both specific T-cell epitopes and epitopes conserved within the PPE family. *Infect. Immun* 2003;71:6116-6123.

Figure 1.

IFN-γ response to the 75 multiepitopic HLA promiscuous MTB peptides obtained from 24 MTB proteins. Each column represents the enumeration of IFN-γ producing cells induced by one selected peptide. The measurement was performed by a ELISpot assay in active TST-negative (panel A), TST-positive (panel B) and TB patients (panel C). Data are reported as the mean of the SFCs per million PBMCs for each peptide. Peptides are grouped as belonging to genes expressed in activated human macrophages (MA), expressed during growth in non-activated human macrophages (MN-A), housekeeping genes (HKG) and RD1 region (RD1). Black bars represent peptides showing responses with significant differences between the three study groups ($p<0.05$) while open bars represent those did not. Peptides drived from CFP-10 (*) and ESAT-6 (§) are represented in the last four positions.

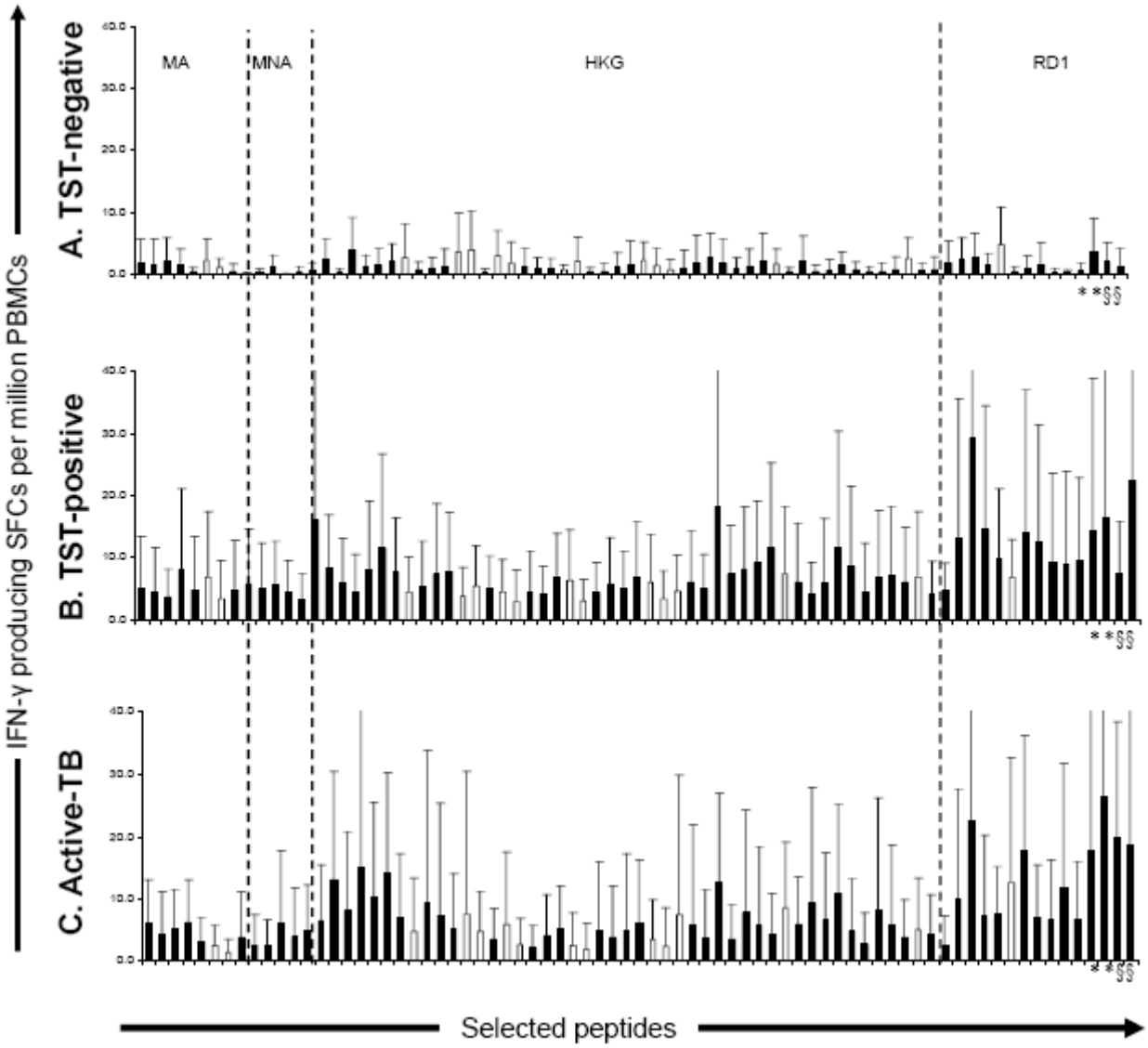


Figure 2.

Evaluation of the *in vitro* response to RD1 derived peptides (panel A) and to peptides designed from genes expressed in activated human macrophages (MA, panel B) at the time of diagnosis of active-TB and after 2 months of standard anti-tuberculous therapy in a subgroup of active-TB patients. Each study subject is represented by one line and the ordinate represents the sum of the number of IFN- γ producing spot forming cells (SFC) per million PBMCs in response to the peptides of each group.

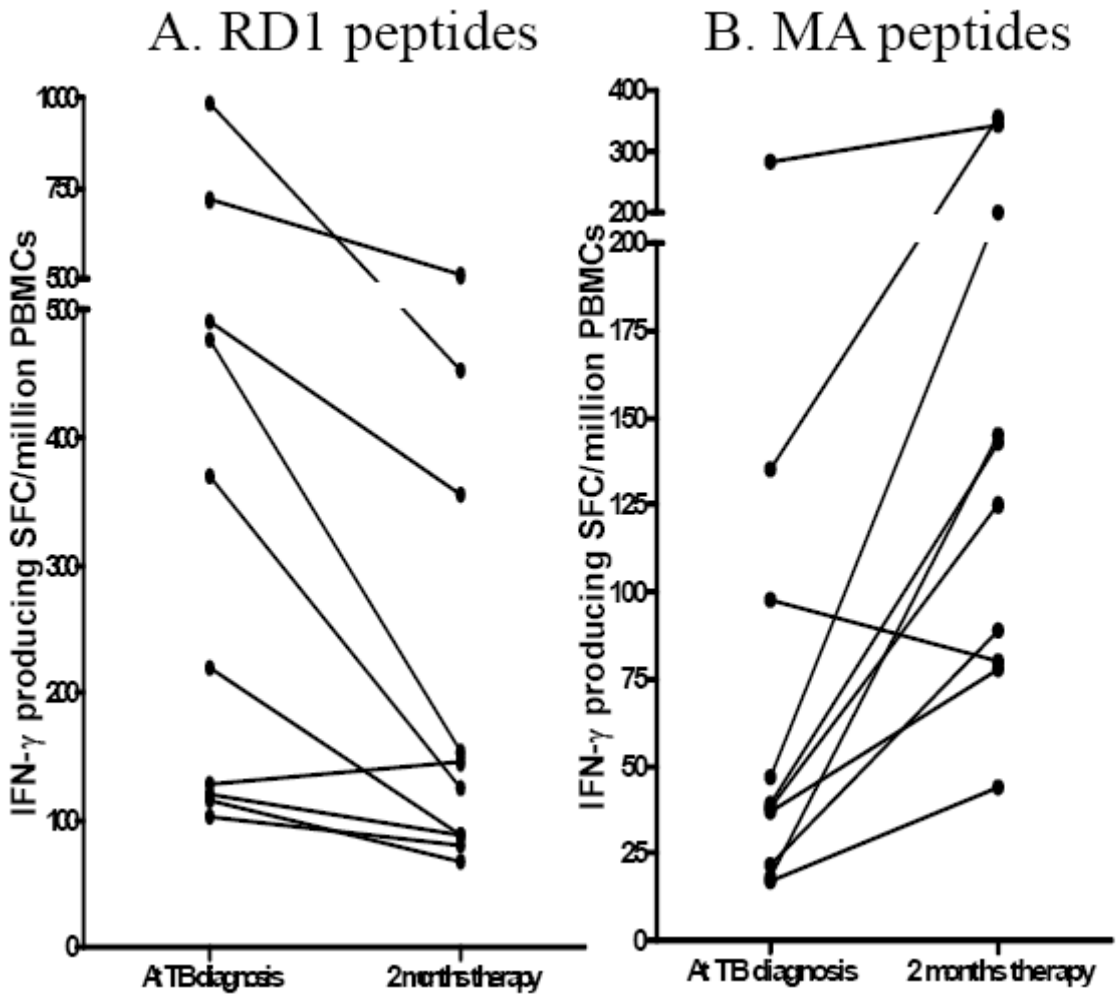


Table1. Study group

STUDY GROUP	n	Mean Age (years)	M/F	Nationality	Vaccination
MTB-Not Exposed TST-negative controls	9	30	5/4	Italian	0/9
MTB-Exposed TST-positive controls	25	46	13/12	Italian	18/25
Untreated active-TB patients	22	39	13/9	3 Bulgarian 9 Italian 7 Rumenian 1 Brasilian 1 Moldavian 1 Polish	12/21 (1 not know)
Treated active-TB patients	23	38	20/3	10 Bulgarian 2 Italian 2 Rumenian 1 Albanese 1 Macedone 1 Egyptian 1 Indian 1 Ucraino 1 Somala 1 Colombiana 1 Cinese 1 Filippino	15/18 (5 not known)

Table2. List of the selected peptides

#	MTB protein	Gi accession number	AA starting position in protein sequence	PEPTIDE SEQUENCE§	HLA-DR GROUP VARIANTS RECOGNISED BY THE EPITOPES PRESENTED IN THE PEPTIDE°	HOMOLOGY BY BLAST^	GROUP‖
1	Rv3883c	**15611019**	1*	AAMHRIFLITVALALLTASPASAIT	1, 3, 4, 7, 8, 11, 13, 15	M. leprae 77%	MA
2	Rv3883c	**15611019**	176	GSIRSLARAVVHAANLGVGV	1, 3, 4, 7, 8, 11, 13, 15	M. leprae 80%	MA
3	Rv3883c	**15611019**	219	IDYAVNVKGVVVVVAAGNTGGDC	1, 3, 4, 7, 8, 11, 13	M. leprae and M. marinum 86%	MA
4	Rv3883c	**15611019**	337	DDLVGAGVIDAVA	1, 3, 4, 8, 11, 13, 15	M. leprae and M. marinum 76%	MA
5	Rv3883c	**15611019**	372	APYNVRRLPPPVVEP	1, 3, 8, 11, 13, 15	Oryza sativa 66%	MA
6	Rv3042c	**15610179**	137	LELRVSVPPGCVGPLQIALTKVAAE	1, 3, 4, 8, 11, 13	M. marinum 82%	MA
7	Rv1114	**15608254**	46	RDVVLDLLANLYQFPVV	3, 4, 8, 11, 13, 15	Roseobacter sp. 64%	MA
8	Rv1114	**15608254**	60	FPVVTHDEVLRLVGRRRLWGRG	1, 3, 7, 8, 11, 13, 15	M. paratuberculosis and Janibacter sp. 68%	MA
9	Rv0125	**15607267**	216	GQVVGMNTAASDNF	1, 3, 4, 8, 11, 13	M. leprae 85%	MN-A
10	Rv0198c	**15607339**	56	TQVRDLIIQASQAGAA	1, 4, 11, 13	M. marinum 80%	MN-A
11	Rv1382	**15608522**	9	GSLIFAAVLVMLIAVLARLMMRGWR	1, 3, 4, 7, 8, 11, 13, 15	M. avium 79%	MN-A
12	Rv1382	**15608522**	119	RGVAGKVVAGIGILAIRWRLPSGT	1, 7, 8, 11, 13, 15	M. marinum and M. ulcerans 79%	MN-A
13	Rv0755c	**57116777**	107	VVHPAVVQANRVRTWLLAVSNVFGQ	1, 3, 4, 7, 8, 11, 13	M. marinum and M. ulcerans 59%	MN-A
14	Rv1582c	**15608720**	143	LDLHTLKLRPHAPADRIT	3, 4, 8, 13	Burkholderia ambifaria 50%	HKG
15	Rv1582c	**15608720**	254	GEMDLRGVRWVAVSESEK	1, 3, 4, 7, 8, 11, 13, 15	Janibacter sp. 55%	HKG
16	Rv1769	**15608907**	131	APIVMVDSVEHLDL	1, 3, 4, 7, 8, 11, 15	M. marinum 100%, M. ulcerans 92%, Francisella philomiragia71%	HKG
17	Rv1769	**15608907**	215	DNVAGKRVHNAIIRRMQRMSFEE	1, 3, 4, 7, 8, 11, 13, 15	M. marinum, M. ulcerans 65%	HKG
18	Rv1769	**15608907**	365	RRLKLGDKVYFRHTKAGELCE	1, 3, 7, 8, 13	M. marinum, M. ulcerans 95%, Saccharopolyspora erythraea 76%	HKG
19	Rv1769	**15608907**	391	HLVRGAEVVDTVP	1, 3, 4, 11, 13	M. marinum, M. ulcerans 83%, Dinoroseobacter shibae 69%	HKG
20	Rv1979c	**15609116**	138	LTFLGFIGVLLAINLFGNRAIKWAN	1, 3, 4, 7, 8, 11, 13, 15	M. leprae 76%	HKG
21	Rv1979c	15609116	187	QHVNNYATAWSAYS	4, 8, 11, 13, 15	M. leprae 71%	HKG
22	Rv1979c	15609116	297	ATFRTIIVVGALISMFGINVAASFGA	1, 3, 4, 7, 8, 11, 13, 15	M. leprae 92%, Fusobacterium nucleatum 65%	HKG
23	Rv1979c	15609116	413	DKVLPLVAIVVSVGLAVSYD	1, 3, 4, 7, 8, 11, 13, 15	M. leprae 70%	HKG
24	Rv1979c	15609116	448	IALIVITFVVVPAMAYLHYY	1, 3, 4, 7, 8, 11, 13, 15	M. leprae 70%	HKG
25	Rv1979c	15609116	465	HYYRIIRRVGDRPSTR	1, 3, 4, 7, 8, 11, 13, 15	Sclerotinia sclerotiorum 50%	HKG
26	Rv2994	15610131	13	IMIVSLGVTASSFLFINGVAFLIPR	1, 3, 4, 7, 8, 11, 13	M. marinum 92%, M. avium 80%, M. abscessus 56%	HKG
27	Rv2994	15610131	56	PSWGLVVTMFAWGYLLDHVGERMV	1, 3, 4, 7, 8, 11, 13, 15	M. marinum, M. avium 75%	HKG
28	Rv2994	15610131	98	HSLLWIGVFLFLGGMAAGGCNS	1, 3, 4, 7, 8, 11, 13, 15	M. marinum,68%	HKG
29	Rv2994	15610131	204	SPYRGSSILWRIHAASALLMMP	1, 3, 4, 7, 8, 11, 13, 15	M. marinum 86%, M. avium, M. smegmatis 73%	HKG
30	Rv2994	15610131	228	VTVTFMLVWLINHHGWSVAQ	1, 4, 7, 8, 11, 13, 15	M. marinum 85%, M. smegmatis 55%	HKG
31	Rv0418	15607559	1*	AAMVNKSRMMPAV	3, 8, 11, 13, 15	M. marinum 67%	HKG
32	Rv0418	15607559	8	PAVLAVAVVVAFLTTGCIRWST	3, 4, 8, 11, 13, 15	M. marinum, M. ulcerans 72%	HKG
33	Rv0418	15607559	85	VDYVVNTLRNSGFDVQTP	1, 3, 4, 8, 11, 13	M. marinum, M. ulcerans 94% M. avium 83%	HKG
34	Rv0418	15607559	157	SDYDRLPVSGAVVLVDRGVCPFAQ	3, 7, 8, 11, 13, 15	M. marinum, M. ulcerans 83% M. avium 73%	HKG
35	Rv0418	15607559	216	IPVVSVTKSVGFQLRGQSGPTTVK	1, 3, 4, 7, 8	M. marinum, M. ulcerans, M. avium 83% M. smegmatis 65%	HKG
36	Rv0418	15607559	377	AGIERTFVAYLKMAGKTAQDT	1, 3, 7, 8, 11, 13, 15	M. marinum, M. ulcerans, M. vanbaalenii 71%, M. abscessus 57%	HKG
37	Rv1983	57116933	331	GGLYYIFATYTTTVDF	1, 4, 7, 8, 11, 13, 15	M. marinum 81%	HKG
38	Rv1983	57116933	348	NGIVTAPTAVNVVLLSIPTSPFAI	1, 3, 4, 8, 11, 13, 15	M. marinum 73%	HKG
39	Rv1983	57116933	533	LPFRFQPVYIDYSPSGIGT	1, 3, 4, 8, 11, 13	M. marinum 78%	HKG
40	Rv1986	15609123	2	SPLVVGFLACFTLIAAIGAQN	1, 3, 4, 7, 8, 11, 13, 15	M. smegmatis 76%	HKG
41	Rv1986	15609123	36	LPVVALCTVSDIVLIAAGIAGFGA	1, 3, 4, 8, 11, 13, 15	M. gilvum 83%	HKG
42	Rv1986	15609123	102	TPVRLAEVLVTCA	1, 3, 4, 7, 8, 11, 13, 15	M. Smegmatis and Leishmania infantum 72%	HKG
43	Rv1986	15609123	165	GRLRGLFTNPGSW	1, 3, 4, 8, 11, 13	Burkholderia ubonensis 76%	HKG
44	Rv1986	15609123	176	SWRILDGLIAVMMVALGISLTVT	1, 3, 4, 7, 8, 11, 13, 15	M. vanbaalenii 71%	HKG
45	Rv0747	15608885	1*	AAMSWVMVSPELVVAAAADLAG	1, 3, 4, 7, 8, 11, 13, 15	M. marinum 68%	HKG
46	Rv0747	15608885	69	AAFYAQFVQALSAGGGA	1, 4, 7, 11, 13	M. marinum 76%	HKG
47	Rv0747	15608885	99	AQFVAATGRPLIGNGANGAPG	1, 4, 8, 11	M. marinum 64%	HKG
48	Rv0747	15608885	387	HTLQQDVINMVNDPFQTLTGRPLIG	1, 3, 4, 11	M. marinum 66%	HKG
49	Rv1434	15608572	1*	AAMRASPAERVDGAYAGAGPHTQSV	1, 3, 7, 11, 13	Xanthobacter autotrophicus 58%	HKG
50	Rv1866	15609003	91	DRYRHLVALSITDFGAAGPRSSWRA	1, 3, 4, 7, 8, 11, 13, 15	M. avium 83%	HKG

51	Rv1866	15609003	149	AAVQAAWAVLVAYFNRLRCGTGDY	1, 7, 8, 11, 13, 15	M. avium 82%	HKG
52	Rv1866	15609003	222	DGYVRFCVMAPRQWRGLRRWLGEP	1, 4, 7, 8, 11, 13, 15	M. marinum 91% M. avium 83%	HKG
53	Rv1866	15609003	372	YPFEGLRILDLGIIVAGGELS	1, 3, 4, 7, 8, 11, 13, 15	M. marinum 95% M. avium 82%	HKG
54	Rv1866	15609003	477	DVLHAFNPRIVLAGSSAFGN	1, 3, 4, 7, 8, 11, 13, 15	M. marinum 81% M. avium 77%	HKG
55	Rv1866	15609003	502	TRMGYGPLVRAATGVTRVWTSDEAQ	1, 3, 4, 8, 11	M. marinum and M. avium 84%	HKG
56	Rv1866	15609003	538	TTIFPDHVVGRVGALLALAALIHRD	1, 3, 4, 8, 11, 13, 15	M. marinum and M. avium 84%	HKG
57	Rv1866	15609003	568	AHVHISQAEVVVVNQLDTMFVAE	1, 3, 4, 7, 13, 15	M. marinum and M. avium 81%	HKG
58	Rv3219	15610355	68	RALKRRNARTKAR	4, 8, 11, 13	M. marinum, M. ulcerans, M. smegmatis and M. avium 100%, Saccharopolyspora erythraea 84%	HKG
59	Rv0956	15608096	12	ARLVVLASGTGSLLR	1, 3, 4, 7, 8, 11, 13, 15	M. marinum and M. ulcerans, 100%, M. smegmatis and M. avium 86%, Nocardia farcinica 80%	HKG
60	Rv0956	15608096	176	ETLHERIKVTERRLLVAAVAALATH	1, 4, 7, 8, 11, 13	M. smegmatis 83%	HKG
61	Rv3872	57117163	56	EGIQLLASNASQ	1, 3, 4, 8, 11, 13, 15	Caenorhabditis elegans 67%	RD1
62	Rv3873	57117164	68	TPMVVWLQTASTQAKTR	1, 4, 8, 11, 13	M. marinum 76%	RD1
63	Rv3873	57117164	121	TNFFGINTIPIALT	1, 4, 7, 11, 13	M. abscessus, M. avium, M. leprae, M. smegmatis, M. ulcerans, 86%	RD1
64	Rv3873	57117164	131	IALTEMDYFIRMWNQAALAMEVY	1, 4, 7, 8, 11, 13, 15	M. marinum, M. smegmatis 82%	RD1
65	Rv3878	15611014	138	PRVVATVPQLVQLAPHAVQMSQN	1, 4, 7, 8, 11, 13, 15	M. marinum 74%	RD1
66	Rv3878	15611014	216	GDVQPAEVVAAAR	1, 3, 4, 11, 13	Magnetospirillum magneticum 76%	RD1
67	Rv3879c	15611015	135	TAINSLVTATHGA	4, 8, 11, 13	M. marinum 77%	RD1
68	Rv3879c	15611015	325	PHVKPAALAEQPG	4, 8, 11, 13	Verminephrobacter eiseniа 77%	RD1
69	Rv3879c	15611015	484	DALRLARRIAAAL	1, 3, 7, 8, 11, 13, 15	M. marinum 92%, Burkholderia graminis 77%	RD1
70	Rv3879c	15611015	656	LWFELMKPMTSTATGR	1, 3, 4, 8, 11, 13	M. marinum 81%, Rhodobacter sphaeroides 56%	RD1
71	Rv3879c	15611015	674	AHLRAFRAYAAHSQEI	1, 4, 8, 11, 13, 15	M. marinum 81%, M. smegmatis 68%	RD1
72	Rv3874	15611010	53	AVVRFQEAANKQK	1, 3, 4, 11, 13	M. marinum and M. ulcerans 100%, Anaeba variabilis 69%	RD1
73	Rv3874	15611010	65	TNIRQAGVQYSRA	1, 3, 4, 8, 11, 13	M. marinum and M. ulcerans 100%, Strongylocentrotus purpuratus 61%	RD1
74	Rv3875	57117165	2	QQWNFAGIEAAASAIQGNVTSIHSL	1, 3, 4, 11, 13	M. marinum 96%, M. ulcerans 84%, M. smegmatis 68%	RD1
75	Rv3875	57117165	65	NALQNLARTISEA	3, 8, 11, 13	M. marinum, M. ulcerans M. riyadhense 100%, M. smegmatis 83%, Bacillus subtilis 69%	RD1

Notes:

|| Peptides are grouped according to the capability of the belonging protein to be selectively over expressed during growth in activated human macrophages (MA), expressed during growth in non-activated human macrophages (MN-A), constitutively expressed housekeeping genes (HKG) and belonging to the RD1 region proteins (RD1).

*Peptide epitopes starting from aminoacid position 1 in the protein sequence where two alanines were added to the N-terminus of the peptide in order to increase binding affinity (see also Method section).

§Peptide sequence reported as single letter code. The underlined letter(s) in the peptide sequence indicate the aminoacid acting as P1 of the epitope(s) presented along the peptide sequence as for quantitative implemented peptide binding motif analysis.

°HLA-DR group variants capable to bind *in silico* the peptide sequence with an affinity equivalent or higher than the 3% of the best binder peptides for the alleles present in the variant.

^Homology analysis of the peptide sequence performed by BLAST reporting reported the first microorganisms not beloning to the M. tuberculosiscomplex with the higher percentage of identity (expressed as percentage) to the selected protein sequence.

IMMUNOGENETICS OF SARCOIDOSIS

M. AVIUM BINDING TO HLA-DR EXPRESSED ALLELES *IN SILICO*: A MODEL OF PHENOTYPIC SUSCEPTIBILITY TO SARCOIDOSIS

Abstract

Sarcoidosis is a systemic granulomatosis disease of unknown origin where a number of microbes, in particular *M. tuberculosis* and non-tuberculous mycobacteria, have been hypotesized to play a role in disease pathogenesis, possibly through bacterial antigen-driven hypersensitivity. To test this concept, we used bioinformatic tools allowing the identification of antigenic peptides in whole microbial genomes to analyze the interaction between the expressed HLA-DR gene allelic variants and the HLA-DR immunome of all pathogenic bacteria in a population of 149 sarcoidosis affected subjects and 447 controls, all HLA-typed at high resolution.

We show here that patients with the Loefgren's syndrome, express HLA-DR alleles that recognize *in silico* a significantly higher number of bacterial antigen epitopes compared to the control population (18,496+9,114 vs 17,954+8,742; $p<0.00001$), and the chronic sarcoidosis affected popuation (17,954+8,742; $p<0.00001$ vs Loefgren's and controls). Further, the analysis of the ability of the HLA-DR allele combinations expressed by the Loefgren's and the chronic sarcoidosis affected subjects to recognize *M. avium* epitopes demonstrates that a significantly larger number of Loefgren's are capable of top affinity recognition, compared to chronic sarcoidosis (45% vs 17%, $p<0.0037$). Finally, both Loefgren'a and chronic sarcoidosis subjects expressed HLA-DR allele combinations capable of *M. tuberculosis* and *M. avium* epitope recognition at higher affinity than tuberculosis affected subjects ($p<0.01$ all comparisons).

In conclusion, we propose that -- at least in a subgroup of affected subjects-- sarcoidosis might be part of a spectrum of granulomatous responses to several agents where the Loefgren's syndrome represents the hyper-reactive end of the spectrum while pulmonary tuberculosis and atypical mycobacterial infections might represent the opposite end.

Introduction

Sarcoidosis is a chronic disorder characterized by the accumulation of increased numbers of T-lymphocytes and activated mononuclear phagocytes in the lower respiratory tract and other sites of disease, leading to the formation of noncaseating granulomas in involved tissues such as the lung, skin and eyes. Several lines of evidence suggest that the accumulation of activated T-cells and macrophages in the lower respiratory tract is the result of chronic stimulation by some (yet unknown) antigen(s) (1). Firstly, in sarcoidosis activated alveolar macrophages possess increased ability of presenting antigens to T-cells (2) and express HLA-DR molecules at higher level at the surface of the cell (3). Secondly, T-cells accumulate in the lower respiratory tract where they release interleukin 2 (4, 5) and interferon-γ at exaggerated rates (2). Thirdly, sarcoid lung T-lymphocytes display activated memory T-cell phenotypes (6, 7) and express oligoclonal Vβ T-cell receptor genes as they were locally expanded in response to a classical antigenic stimulus rather than a more broad-based stimulus such as from a superantigen (8, 9).

A number of potential causation agents have been implicated in the pathogenesis of sarcoidosis (10) and some sarcodosis-like disease and reactions could be identified as hypersensitivities to compounds such as beryllium (11), titanium (12) , zirconium (13) and aluminum (14). Recently, it has been shown that, in addition to chemicals, microbial agents otherwise capable of causing infection and disease may be implicated in the pathogenesis of granulomatous hypersensitivity pneumonias such as the hot tub lung, where *M. avium* has been identified as a causative agent and (15) and metal working fluid pneumonitis, where a variety of bacteria have been implicated (16). Furthermore, in a chronic granulomatous disease such as Chron's disease *M. avium* paratuberculosis has been implicated (17).

In sarcoidosis, since the application of the the polymerase chain reaction to the detection of mycobacterial genomes in clinical samples (18), the presence of nucleic acids of M. *tuberculosis*, non tuberculous mycobacteria and of *Propionibacterium acnes* has been observed both in bronchoalveolar lavage and tissue samples in different sarcoidosis patient populations (19-23). Collectively, these studies indicate that mycobacteria can be found in tissue samples in 23 to 30% of sarcoid patients, with a ten-fold higher probability compared to controls, and suggest an association between mycobacteria and sarcoidosis (24-26).

In this context and with the knowledge that certain microbial agents could cause hypersensitivity reactions leading to granuloma formation in the absence of clinical infection, the observations that antibody and T-cell mediated immune responses against mycobacterial antigens, such as catalase-peroxidase (mKatG) (27), heat shock protein (Mtb-hsp) (28), and Antigen 85A (29)

could be demonstrated in sarcoidosis and in the Kweim reaction, led to hypothesize that exposure to certain microbes might be causative of sarcoidosis in hyperreactive individuals (30-32).

Antigen presentation to CD4 T-cells is mediated by a class of peptide cell surface receptors, the HLA class II proteins (HLA-DP, -DQ and -DR), which are capable of binding peptides derived from the digestion of microbes and other agents in the phago-lysosome of antigen presenting cells and to carry them to the cell surface and to present them to the CD4 T-cell expressing the appropriate T-cell antigen receptor. Importantly, the selection of antigenic peptides presented to, hence recognized by, T-cells is driven by the chemico-physical interaction between (i) the amino acid side chains which are lining the receptor-like pockets on the floor of the HLA antigen binding groove and (ii) the aminoacid side chains of the antigenic peptides (33, 34).

The HLA class II genes have been implicated in susceptibility to sarcoidosis (35, 36) and it has been shown that the expression of different HLA allelic variants is associated with different clinical presentations and disease outcomes. The Loefgren's syndrome, an acute form of sarcoidosis characterized by fever, erythema nodosum, bilateral hilar lymphoadenopaty and/or ankle arthritis is associated with the expression of the HLA-DRB1*0301 allele and has a favourable clinical outcome (37). On the other hand, chronic forms of sarcoidosis with less favourable outcomes have been associated with the expression of the DRB1*1501 allele (37).

Key to the understanding of susceptibility to immune disorders is the knowledge that HLA alleles can select peptide antigens for antigen presentation with widely varying affinities, thus leading to different immune responses against the same agent. In this regard, it is well known that since the HLA polymorphisms characterizing allelic variants code for aminoacid changes in the peptide binding groove's, each HLA allele binds a unique set of aminoacid side chains. Each HLA class II allelic variant is therefore capable of selecting a discrete set of antigenic peptides for antigen presentation (38, 39) and to allow for the formation of unique HLA molecule/T-cell receptor/antigenic peptide complexes activating specific cytokine producing CD4 T-cells (40). Thus, since each individual coexpresses at least two HLA-DR molecules on the cell surface of the antigen presenting cell, it is reasonable to think that individual ability to generate immune responses to a given antigen shall be determined by the capability of each subject's two HLA-DR molecules combined to recognize antigen epitopes (41-43).

By using bioinformatics tools allowing the identification of antigenic peptides in whole microbial genomes (44), we have recently demonstrated that the expression of HLA-DR allele combinations endowed with a lower affinity for the *M. tubercuosis* genome and for its derived peptide antigens is associated with susceptibility to active tuberculosis (45). Interestingly, the HLA-DR alleles implicated in this reduced ability to bind and select for mycobacterial antigenic peptides

are the same which have also been associated with chronic sarcoidosis. Even more interestingly, the alleles endowed with higher affinity for *M. tuberculosis* antigenic peptides, and with "protection" against tuberculosis, are the same which have been previously associated with the acute forms of sarcoidosis (45).

With the background that the affinity for peptide antigens of the HLA-DR expressed allelic variants plays a key role in determining the ability of the individual to recognize and react to bacterial antigens, hence in generating the ineffective immune response leading to bacterial invasion and active disease, we hypothesized that it may also play a role in determining suceptibility to hypersensitivity reactions to the same bacteria. To test this hypothesis, we set out to to analyze the interaction between the expressed HLA-DR gene allelic variants and the genomes of all pathogenic bacterial in the Loefgren's syndrome and in chronic sarcoidosis, two immunologically contrasting forms of sarcoidosis, in order to determine whether hypersensitivity to mycobacterial, propionibacterial or other bacterial antigens may play a potential role in the pathogenesis of sarcoidosis.

Methods

Patients' characteristics:

The study population was composed by the patients with sarcoidosis and matched controls of an already described population evaluated for the genetic susceptibility factors in sarcoidosis. HLA-DR high resolution typing was available for all study subjects (35).

They were 149 patients with sarcoidosis, of which 39 with Loefgren's syndrome and 110 with chronic sarcoidosis, and 447 matched controls (see table 1). The quantitative implemented peptide binding motifs i.e., the algorithms with which to assess the ability of a given HLA-DR alleleic variant, are available only for 49 out of the more than 300 HLA-DRB1 alleles (46, 47) and cover, with at least one allele, about 90% of the HLA-DR variability of different human populations (48). For the purpose of this study, only the subjects carrying both HLA-DR alleles with an available HLA-DR binding motif were included (see table 1). Specifically, 349 out of 447 (78.1%) control subjects, 31 out of 39 (79.5%) patients with Loefgren syndrome and 93 out of 110 (84.5%) patients with chronic sarcoidosis, could be subjected to the analysis.

The selected subgroups did not differ for demographical characteristics from the subgroups of subjects excluded from the study for having one or both HLA-DR alleles without a known binding motif (data not shown).

Finally, a group of patients with active tuberculosis and matched controls already evaluated previously (45) were used as comparison. They include: (i) 303 patients with active tuberculosis of which 92 carrying both HLA-DR alleles with an available HLA-DR binding motif and (ii) 345 controls of which 111 carrying both HLA-DR alleles with an available HLA-DR binding motif (Table 1).

Genomes.

One hundred and twenty four, non redundant, bacterial complete genomes, classified as "relevant human pathogens" in the Genome OnLine Database v2.0 (GOLD database update October 30th 2008, http://www.genomesonline.org/), were used to analyse the number of potential epitopes in each genome which were recognised by the study population. Bacterial genome characteristics (accession code, species, family, human disease and number of open reading frames) are listed in table 2.

Software for the identification and enumeration of epitopes in whole genomes

To enumerate the T-cell epitopes present in the protein data sets of each microbial genome examined, we used a software specifically designed for the identification and enumeration of peptide binding epitopes to HLA-DR molecules in large set of protein databases, as already described (45).

For the purpose of this study, the bacterial genomes were analysed for enumerating the epitopes in silico recognised by the combination of the HLA-DR alleles carried by single subjects at an affinity level equivalent or higher than the 1% of the best binder peptides for the allele in analysis (49). The Threshold of affinity is a preselected numerical value used to differentiate between binders and non binders, any peptide frame scoring higher than this value is predicted as binder or vice versa; it correlates with the peptide score (39, 49) and therefore with HLA-ligand interaction, therefore it is an indicator for the likelihood that predicted peptide is capable of binding to a given HLA-molecule.

Statistical analysis.

All the data are expressed as mean $\pm$ standard deviation of the mean (SD), and frequency expressed as percentage. Interquartile distribution of epitope number recognised was used to classify allele combination. Comparisons between means are made by Student's t test. Comparison between frequencies is made by Chi-square, with Yates and Bonferroni's correction, Chi-square for trend and Fisher's exact test when appropriate. GraphPad Prism version 4.0 was used for all statistical analyses and graphs.

Results

Differential capability of chronic sarcoidosis and Loefgren's syndrome affected subjects to recognize genomes of human pathogen.

The capability of the HLA-DR expressed allelic variants of the study subjects to bind bacterial peptide epitopes was determined as the number of epitopes bound *in silico* at the top 1% affinity. At an affinity level equivalent to that of the top 1% of the binder peptides for the HLA-DR molecules of the study subjects, Loefgren syndrome affected subjects recognized a higher number of epitopes (mean epitope number for all the genomes: 18,496+9,114) than chronic sarcoidosis affected subjects (17,515+8,448; $p<0.00001$) and controls (17,954+8,742; $p<0.00001$). In addition, the comparison between chronic sarcoidosis and healthy controls showed that the capability of affected subjects to recognize bacterial pathogen genomes was significantly reduced compared to controls ($p<0.00001$).

Interestingly though, the capability of the HLA-DR allelic variants expressed by Loefgren's syndrome affected subjects to recognize a higher number of epitopes than chronic sarcoidosis affected subjects was not generalized to all the genomes examined, but it was more prominent for certain bacterial pathogens. In particular, statistically significant differences in the mean number of epitopes recognized by Loefgren syndrome affected subjects with respect to chronic sarcoidosis affected subjects was observed for the genomes of bacteria belonging to the genders Bordetella, Burkholderia, Granulibacter, Mycobacterium, Nocardia and Propionibactierium (table 2 and figure 1). The largest difference was observed for *Mycobacterium avium*, where there was 12% difference between the average HLA-DR immunomes, i.e. the number of putative epitopes deduced from the whole bacterial genome that were capable of being bound by the HLA-DR expressed allele of each individual subject, of the two disease groups (figure 1).

Contribution of the different HLA-DRB1 alleles combination in the hyper-recognition of genomes by patients with Loefgren's syndrome.

In order to analyze the nature of the exaggerated epitope recognition of bacterial genomes seen in the subjects with Loefgren's syndrome compared to those with chronic sarcoidosis, the data obtained in the study population for the HLA-DR immunome of *M. avium* i.e., the best binder of all tested genomes, were used.

The distribution of HLA-DR allelic variants in the study populations is shown in table 3. As previously reported (37), alleles HLA-DRB*0301 and *0407 were significantly more frequent in

the Loefgren's syndrome affected while the allele *1501 was significantly over-represented in the chronic sarcoidosis affected group (table 3).

Table 4 shows the distribution of *M. avium* genome recognition levels in the study groups, according to the percentile distribution from bottom to top recognition levels. Considering the affinity level equivalent to 1% of the higher affinity peptides, the 1,225 different combinations generated by the 49 HLA-DRB1 alleles for which the quantitative implemented peptide binding motifs is available, can recognize 6,149 to 43,781 epitopes in the genome of *M. avium*. Subjects with Loefgren's syndrome showed significantly higher levels of *M. avium* epitope recognition than subjects with chronic sarcoidosis as 45% of the Loefgren's affected were in the top 76-100° percentile compared to 17% of the chronic sarcoidosis affected (p=0.0037) and 26% of the controls (p=0.0386). Conversely, subjects with chronic sarcoidosis showed a significantly higher level of *M. avium* epitope recognition in the lower 26°-50° percentiles distribution with 39% compared to the 16% of the Loefgren's affected (p=0.0363) and the 24% of the controls (p=0.0059; table 4).

Furthermore, in order to determine whether the exaggerated recognition of microbial genomes such as those of mycobacteria, shown by Loefgren's syndrome affected subjects, was associated with the expression of specific HLA-DR allelic variants rather than with the expression of combinations of HLA-DR alleles endowed with higher binding capabilities, we analyzed the distribution of HLA-DR alleles in the *M. avium* recognition percentile subgroups.

This analysis showed that there were no biases in the expression of HLA-DR variants in the Loefgren's syndrome affected population, neither in the 26-50° nor in the 76-100° percentile groups (table 5). Only the association of the HLA-DRB1*1501 allele with chronic sarcoidosis remained statistically significant both in the higher and the lower affinity percentile groups (table 5).

Assessment of HLA-DR M. avium and M. tuberculosis epitope affinity in sarcoidosis and tuberculosis.

Altogether the data suggests that Loefgren's syndrome immune responsiveness might be characterized by the expression of HLA-DR genomes endowed with an exaggerated ability to recognize bacterial genomes while, in contrast, chronic sarcoidosis appears to be characterized by the expression of HLA-DR alleles with significantly lower ability to bind and present bacterial antigens. As already reported in tuberculosis affected patients a biased expression of the same HLA-DR alleles expressed in sarcoidosis could be observed (45).

Strikingly, the comparison of epitope binding affinity of the HLA-DR allelic combinations expressed by sarcoidosis affected subjects with those of tuberculosis patients showed a statistically significant lower binding affinity, both toward the *M. avium* and the *M. tuberculosis* immunomes

than that of sarcoidosis affected subjects. In addition, when the patient groups were ranked by their binding affinities, the Loefgren's affected ranked the highest, with the tuberculosis affected the lowest. Chronic sarcoidosis, although ranking lower than controls ranked higher than tuberculosis (table 6), thereby indicating that in tuberculosis the susceptibility to infection and active disease is likely associated with the expression of a defective epitope recognition repertoire .

Discussion

We show here that patients affected by the Loefgren's syndrome express HLA-DR alleleic combinations that recognize a significantly higher number of bacterial antigen epitopes compared to healthy individuals, as they do in comparison to chronic sarcoidosis.

With regard to sarcoidosis, the data are consistent with the observation of Grosser et al. (50) who recently reported the association of chronic sarcoidosis, the presence of mycobacterial DNA in affected tissue and the expression of DRB1*15, whereas acute sarcoidosis was associated with the absence of mycobacterial DNA and with the expression of HLA-DRB1*03. These data add further support to the concept that the manifestations and outcomes of sarcoidosis are driven by a complex interaction between antigen presentating molecules, antigens(s) and T.-cell receptors leading to antigen clearance or persistence (51). In this hypothesis, HLA Class II molecules with the higher affinity for mycobacterial antigenic epitopes might present to T cells wider antigen repertoires with more elevated efficiency. This would initiate a hypersensitivity response leading to acute onset of disease, antigen clearance and spontaneous disease remission. To the contrary, HLA Class II molecules, such as DRB1*1501, due to their lower affinity for bacterial epitopes, would express a less efficient antigen presentation process, thereby dictating a sluggish immune response resulting in insidious clinical presentation, antigen persistence, continuing granuloma formation: in summary, to chronic sarcoidosis.

Following the same line of interpretation, also for the *in silico* data presented here on tuberculosis patients, our observation would suggests that patients bearing even less recognizing HLA-DR allele combinations could be unable to mount an effective immune response, would fail to achieve bacterial containment and would, consequentially, develop active infectiuos disease.

Interestingly, it has been proposed that in tuberculosis a spectrum of immune reactions may be recognized, where poor cellular mediated immunity, associated with exuberant antibody production, represents one end of the spectrum while good cellular mediated immunity, with little or no antibody formation, would be at the other end (52). Similarly, in leprosy the immune spectrum is represented by non-reactive disseminated lepromatous leprosy, associated with HLA-DR 15 alleles (53, 54), at one end, while tuberculoid leprosy --the disease presentation characterized by stronger T-cell mediated immunity and a lesser invasiveness—is associated with HLA-DR3 alleles (55) and represents the opposite end. In this context, it is enticing to hypothesize that sarcoidosis may be part of a spectrum of granulomatous response to one or more closely related bacteria where the Loefgren's syndrome represents hypersensitivity, similarly to the hot tub lung

(15) and the metalworking fluid associated pneumonitis (16), while pulmonary atypical mycobacterial infection and tuberculosis might represent the other end of the spectrum.

In conclusion, even though these *in silico* immunogenetic data still need to be confirmed in other sarcoid populations with different ethnical backgrounds and to be validated by *in vitro* and *ex vivo* confirmatory analysis, our findings support a pathogenetic role of specific groups of bacteria in the hypersensitivity granulomatous reaction(s) of sarcoidosis.

References

1. Crystal RG, Roberts WC, Hunninghake GW, Gadek JE, Fulmer JD, and Line BR: Pulmonary sarcoidosis: a disease characterized and perpetuated by activated lung T-lymphocytes. *Ann. Intern. Med* 1981. 94:73-94.
2. Venet A, Hance AJ, Saltini C, Robinson BWS, Crystal RJ: Enhanced Alveolar Macrophage-mediated Antigen-induced T Lymphocyte Proliferation in Sarcoidosis. *J. Clin. Invest* 1985; 75: 293-301.
3. Spurzem JR, Saltini C, Kirby M, Konishi K, Crystal RG: Expression of HLA class II genes in alveolar macrophages of patients with sarcoidosis. *Am Rev Respir Dis* 1989;140:89-94
4. PinkstonP, Bitterman PB, and Crystal RG: Spontaneous release of interleukin-2 by lung T-lymphocytes in active pulmonary sarcoidosis. *N. Engl. J. Med* 1983; 308:793-800.
5. Saltini C, Spurzem JR, Lee JJ, Pinkston P, Crystal RG: Spontaneous Release of Interleukin 2 by Lung T Lymphocytes in Active Pulmonary Sarcoidosis Is Primarily from the Leu3+DR+ T Cell Subset. *J Clin Invest*. 1986; 77:1962-1970
6. du Bois RM, Kirby M, Balbi B, Saltini C and Crystal RG: T-lymphocytes that accumulate in the lung in sarcoidosis have evidence of recent stimulation of the T-cell antigen receptor. *Am Rev Respir Dis* 1992; 154: 1205-1211.
7. Saltini C, Kirby M, Trapnell BC, Tamura N, Crystal RG: Biased accumulation of T lymphocytes with "memory"-type CD45 leukocyte common antigen gene expression on the epithelial surface of the human lung. *J Exp Med*. 1990;171:1123-40
8. Silver RF, Crystal RG, Moller DR: Limited heterogeneity of biased T-cell receptor V□ gene usage in lung but not blood T cells in active pulmonary sarcoidosis. *Immunology* 1996; 88:516-523.
9. Moller DR: T-cell receptor genes in sarcoidosis. *Sarcoidosis Vasc Diffuse Lung Dis*. 1998;15:158-64.
10. Thomeer M, Demedts M, Wuyts W: Epidemiology of sarcoidosis. In: Dent M, Costabel U, eds. Sarcoidosis. *Eur Respir Mon* 2005, 32: 13–22.
11. Saltini C., Winestock K., Kirby M., Pinkston P., Crystal R.G. Maintenance of alveolitis in patients with chronic beryllium disease by beryllium-specific helper T-cells. *N. Engl. J. Med*. 1989; 320: 1103-1109.
12. Redline S., Barna B., Tomashefski J.F., Abraham J.L. Granulomatous disease associated with pulmonary deposition of titanium. *Br. J. Indust. Med*. 1986; 43: 652-6.
13. U. Werfel, J. Schneider, K. Rödelsperger, J. Kotter, W. Popp, H.J. Woitowitz, G. Zieger: Sarcoid granulomatosis after zirconium exposure with multiple organ involvement. *Eur Respir J* 1998; 12: 750.
14. De Vuyst P., Dumortier P., Schandené L., Estenne M., Verhest A., Yernault J.C. Sarcoidlike lung granulomatosis induced by aluminum dusts. Am. Rev. Respir. Dis.1987; 135: 493-97
15. Viktor Hanak, Sanjay Kalra, Timothy R. Aksamit, Thomas E. Hartman, Henry D. Tazelaar, Jay H. Ryu: Hot tub lung: Presenting features and clinical course of 21 patients. *Respiratory Medicine* 2006; 100: 610–615.
16. Hodgson MJ, Bracker A, Yang C, Storey E, Jarvis BJ, Milton D, Lummus Z, Bernstein D, Cole S: Hypersensitivity Pneumonitis In a Metal-Working Environment. *American J Industrial Medicine* 2001; 39: 616-628.
17. Sechi LA, Scanu AM, Molicotti P, Cannas S, Mura M, Dettori G, Fadda G, Zanetti S: Detection and Isolation of Mycobacterium avium subspecies paratuberculosis from intestinal mucosal biopsies of patients with and without Crohn's disease in Sardinia. *Am J Gastroenterol 2005*; 100: 1529-1536.

18. Brisson-Noël A, Gicquel B, Lecossier D, Lévy-Frébault V, Nassif X, Hance AJ: Rapid diagnosis of tuberculosis by amplification of mycobacterial DNA in clinical samples. *Lancet* 1989; 2: 1069-71.
19. Bocart D, Lecossier D, De Lassence A, Valeyre D, Battesti JP, Hance AJ: A search for mycobacterial DNA in granulomatous tissues from patients with sarcoidosis using the polymerase chain reaction. *Am Rev Respir Dis*. 1992;145: 1142-8.
20. Saboor SA, Johnson NM, McFadden J: Detection of mycobacterial DNA in sarcoidosis and tuberculosis with polymerase chain reaction. *Lancet* 1992; 339: 1012-5.
21. Eishi Y, Suga M, Ishige I, Kobayashi D, Yamada T, Takemura T, Takizawa T, Koike M, Kudoh S, Costabel U, Guzman J, Rizzato G, Gambacorta M, du Bois R, Nicholson AG, Sharma OP, Ando M: Quantitative analysis of mycobacterial and propionibacterial DNA in lymph nodes of Japanese and European patients with sarcoidosis. *J Clin Microbiol.* 2002; 40: 198-204.
22. Ishige I, Eishi Y, Takemura T, Kobayashi I, Nakata K, Tanaka I, Nagaoka S, Iwai K, Watanabe K, Takizawa T, Koike M: Propionibacterium acnes is the most common bacterium commensal in peripheral lung tissue and mediastinal lymph nodes from subjects without sarcoidosis. *Sarcoidosis Vasc Diffuse Lung Dis* 2005; 22: 33–42.
23. Ichikawa H, Kataoka M, Hiramatsu J, Ohmori M, Tanimoto Y, Kanehiro A, Nakata Y, Tanimoto M: Quantitative analysis of propionibacterial DNA in bronchoalveolar lavage cells from patients with sarcoidosis. *Sarcoidosis Vasc Diffuse Lung Dis* 2008; 25; 15-20.
24. Fidler HM: Mycobacteria and sarcoidosis: recent advances. *Sarcoidosis* 1994 Mar; 11: 66-8.
25. Mangiapan G, Hance AJ. Mycobacteria and sarcoidosis: an overview and summary of recent molecular biological data. *Sarcoidosis.* 1995;12: 20-37.
26. Gupta D, Agarwal R, Aggarwal AN, Jindal SK: Molecular evidence for the role of mycobacteria in sarcoidosis: a meta-analysis. *Eur Respir J*. 2007, 30:508-16.
27. Song Z, Marzilli L, Greenlee BM, Chen ES, Silver RF, Askin FB, Teirstein AS, Zhang Y, Cotter RJ, Moller DR: Mycobacterial catalase-peroxidase is a tissue antigen and target of the adaptive immune response in systemic sarcoidosis. *J Exp Med.* 2005, 201:755-67.
28. Dubaniewicz A, Trzonkowski P, Dubaniewicz-Wybieralska M, Dubaniewicz A, Singh M, Myśliwski A: Mycobacterial heat shock protein-induced blood T lymphocytes subsets and cytokine pattern: comparison of sarcoidosis with tuberculosis and healthy controls. *Respirology* 2007, 12:346-54.
29. Hajizadeh R, Sato H, Carlisle J, Nadaf MT, Evans W, Shepherd BE, Miller RF, Kalams SA, Drake WP: Mycobacterium tuberculosis Antigen 85A induces Th-1 immune responses in systemic sarcoidosis. *J Clin Immunol.* 2007, 27:445-54.
30. Drake WP: When a commensal becomes a pathogen. *Sarcoidosis Vasc Diffuse Lung Dis* 2008; 25: 10-11.
31. Moller DR: State of the Art. Potential Etiologic Agents in Sarcoidosis. *The Proceedings of the American Thoracic Society* 2007; 4: 465-468.
32. Tercelj M, Salobir B, Rylander R: Microbial antigen treatment in sarcoidosis a new paradigm? *Med Hypotheses.* 2008; 70: 831-4.
33. Davenport MP, Quinn CL, Valsasnini P, Sinigaglia F, Hill AV, Bell JI: Analysis of peptide-binding motifs for two disease associated HLA-DR13 alleles using an M13 phage display library. *Immunology* 1996, 88: 482-486.
34. Mallios RR: Predicting class II MHC/peptide multi-level binding with an iterative stepwise discriminant analysis meta-algorithm. *Bioinformatics* 2001,17: 942-8.
35. Voorter CE, Amicosante M, Berretta F, Groeneveld L, Drent M, van den Berg-Loonen EM: HLA class II amino acid epitopes as susceptibility markers of sarcoidosis. *Tissue Antigens* 2007, 70:18-27

36. Rossman MD, Thompson B, Frederick M, Maliarik M, Iannuzzi MC, Rybicki BA, Pandey JP, Newman LS, Magira E, Beznik-Cizman B, Monos D; ACCESS Group: HLA-DRB1*1101: a significant risk factor for sarcoidosis in blacks and whites. *Am J Hum Genet.* 2003,73:720-35
37. Berlin M, Fogdell-Hahn A, Olerup O, Eklund A, Grunewald J: HLA-DR predicts the prognosis in scandinavian patients with pulmonary sarcoidosis. *Am J Respir Crit Care Med* 1997; 156: 1601–1605.
38. Germain RN: MHC-dependent antigen processing and peptide presentation: providing ligands for T lymphocyte activation. *Cell* 1994, 76: 287-99.
39. Sturniolo T, Bono E, Ding J, Raddrizzani L, Tuereci O, Sahin U, Braxenthaler M, Gallazzi F, Protti MP, Sinigaglia F, Hammer J: Generation of tissue-specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices. *Nat Biotechnol* 1999, 17: 555-61.
40. Scherer A, Frater J, Oxenius A, Agudelo J, Price DA, Gunthard HF, Barnardo M, Perrin L, Hirschel B, Phillips RE, McLean AR; Swiss HIV: Quantifiable cytotoxic T lymphocyte responses and HLA-related risk of progression to AIDS. *Proc Natl Acad Sci U S A* 2004, 101: 12266-70.
41. Stunz L, Karr RW, Anderson R: HLA-DRB1 and -DRB4 genes are differentially regulated at the transcriptional level. *J Immunol* 1989, 143: 3081– 86.
42. Berdoz J, Gorski J,Termijtelen A, Dayer J, Irle C, Schendel D, Mach B: Constitutive and induced expression of the individual HLA-DR h and a chain loci in different cell types. *J Immunol* 1987, 139: 1336–41.
43. Cotner T, Charbonneau H, Mellins E, Pious D: mRNA abundance, rather than differences in subunit assembly, determine expression of HLA-DRh1 and DRh3 molecules. *J Biol Chem* 1996, 264: 11107 –11.
44. De Groot AS, Bosma A, Chinai N, Frost J, Jesdale BM, Gonzalez MA, Martin W, Saint-Aubin C: From genome to vaccine: in silico predictions, ex vivo verification. *Vaccine* 2001, 19: 4385-95.
45. Contini S, Pallante M, Vejbaesya S, Park M H, Chierakul N, Kim HS, Saltini C, Amicosante M: A model of phenotypic susceptibility to Tuberculosis: deficient in silico selection of Mycobacterium tuberculosis epitopes by HLA alleles. Sarcoidosis Vasc Diffuse Lung Dis 2008, 25:21-28
46. Marsh SGE: Nomenclature for factors of the HLA system, update January 1998. *Tissue Antigens* 1998, 51: 582-3.
47. Marsh SG: WHO Nomenclature Committee for Factors of the HLA System. Nomenclature for factors of the HLA system, update September 2006. *Tissue Antigens* 2006, 68: 540-2.
48. Kubo K, Yamazaki Y, Hanaoka M, Nomura H, Fujimoto K, Honda T, Ota M, Kamijou Y. Analysis of HLA antigens in Mycobacterium avium-intracellulare pulmonary infection. *Am J Respir Crit Care Med* 2000, 161: 1368-71.
49. Singh H, Raghava GP: ProPred: prediction of HLA-DR binding sites. *Bioinformatics* 2001, 17: 1236-7.
50. Grosser M, Luther T, Fuessel M, Bickhardt J, Magdolen V, Baretton G: Clinical course of sarcoidosis in dependence on HLA-DRB1 allele frequencies, inflammatory markers, and the presence of M. tuberculosis DNA fragments. *Sarcoidosis Vasc Diffuse Lung Dis* 2005, 22:66-74.
51. Moller DR, Chen ES: Genetic basis of remitting sarcoidosis. Triumph of the trimolecular complex? *Am J Respir Cell Mol Biol* 2002; 27:391–5.
52. Lenzini L, Rottoli P, Rottoli L: The spectrum of human tuberculosis. *Clin exp Immunol* 1977; 27: 230-237.

53. Soebono H, Giphart MJ, Schreuder GM, Klatser PR, de Vries RR: Associations between HLA-DRB1 alleles and leprosy in an Indonesian population. *Int J Lepr Other Mycobact Dis.* 1997; 65: 190-6.
54. Rani R, Fernandez-Vina MA, Zaheer SA, Beena KR, Stastny P: Study of HLA class II alleles by PCR oligotyping in leprosy patients from north India. *Tissue Antigens.* 1993; 42: 133-7.
55. De Vries RPP: An immunogenetic view of delayed type hypersensitivity. *Tubercle* 1991*;* 72: 161-167.

Figure 1. Relative recognition of all the Loefgren's syndrome affected subjects with respect to chronic sarcoidosis patients of the HLA-DR immunomes deduced from the genomes of pathogenic bacteria. The ratio of the total number of the recognized epitopes (on the ordinate) in the two study groups is reported for 124 non redundant genomes of bacterial pathogens analysed (shown in alphabetic order on the abscissa). The bacterial groups showing statistically significant differences in HLA-DR immunome recognition are indicated on the graph.

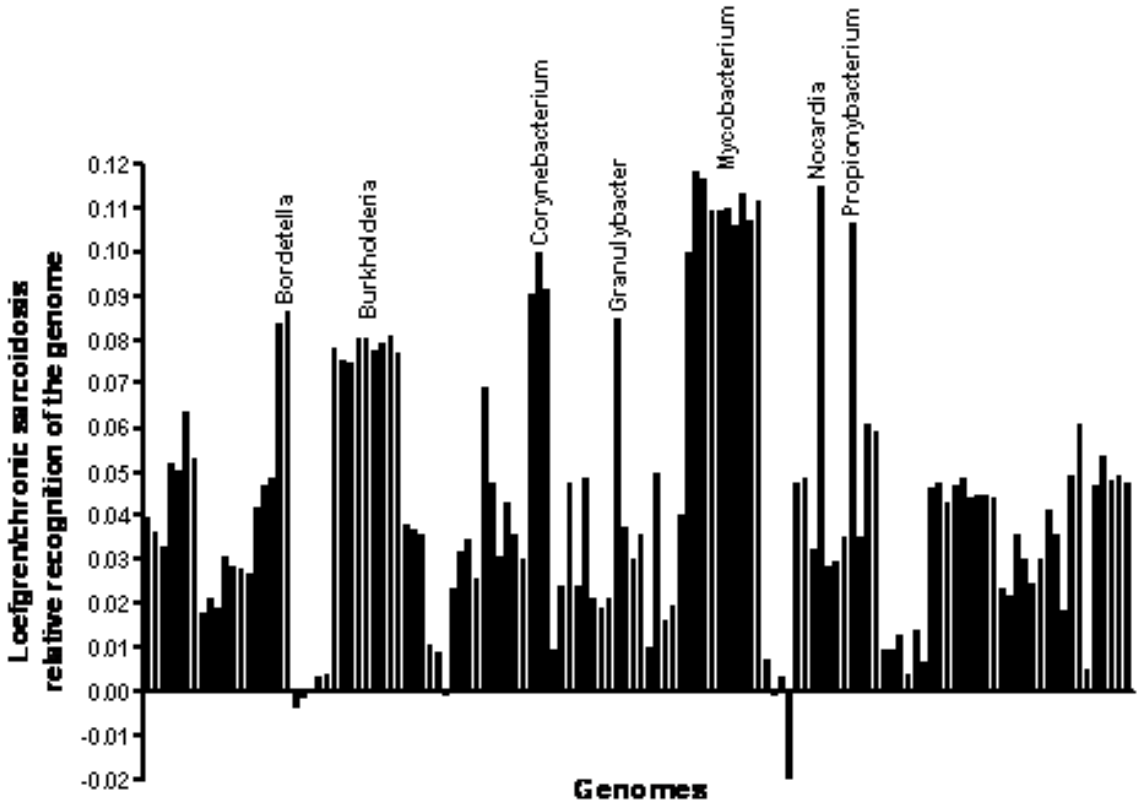


Table 1. Study population.

Study population	Subject number	Subjects with both HLA-DR alleles with known peptide binding motif[1]
Controls	447	349 (78.1%)
Loefgren	39	31 (79.5%)
Chronic Sarcoidosis	110	93 (84.5)

1. Numbers of subjects for each study population used for the epitope prediction analysis.

Table 2. List of non redundant, complete bacterial genomes classified as "relevant human pathogens" in the Genome OnLine Database and analysis of the number of potential epitopes in each genome recognized by the study populations.

ORGANISM[1]	DISEASE	GOLD DATABASE Entry	Genome (Kb)[2]	Number of ORFs[3]	Mean Number of potential epitopes (1% affinity) in the study population[4]			t-test (p value)[5]		
					Controls	Chronic sarcoidosis	Loefgren	Chronic Sarc vs CTR	Loefgren vs CTR	Loefgren vs Chronic Sarc
Acinetobacter baumannii	Meningitis, Pneumonia, Septicemia	Gc00522	3976	3352	17871±3963	17460±3881	18156±4008	0.1868	0.3504	0.1964
Acinetobacter calcoaceticus	Nosocomial infection	Gc00201	3598	3325	21074±4645	20626±4573	21373±4622	0.2036	0.3657	0.2168
Actinobacillus pleuropneumoniae	Necrotizing pleuropneumonia	Gc00708	2242	2036	13015±2854	12798±2808	13221±2833	0.2565	0.3499	0.2346
Aeromonas hydrophila hydrophila	Gastroenteritis, Septicemia, Food poisoning	Gc00456	4744	4122	24703±5644	24118±5315	25373±5553	0.1843	0.2633	0.1312
Aeromonas salmonicida salmonicida	Furunculosis	Gc00536	4702	4086	23482±5284	22944±4991	24103±5246	0.1890	0.2653	0.1355
Anaplasma marginale	Anaplasmosis	Gc00239	1197	949	6919±1536	6723±1466	7152±1639	0.1350	0.2108	0.0867
Anaplasma phagocytophilum	Anaplasmosis	Gc00351	1471	1264	7080±1571	6916±1544	7282±1617	0.1842	0.2470	0.1301
Arcobacter butzleri	Bacteremia, Gastroenteritis	Gc00657	2341	2259	16597±3902	16502±4006	16797±3699	0.4185	0.3915	0.3591
Bacillus anthracis	Anthrax	Gc00189	5227	5309	31589±7108	31303±7157	31971±6753	0.3655	0.3868	0.3245
Bacillus cereus	Food poisoning	Gc00173	5224	5603	33351±7428	33086±7480	33715±7080	0.3804	0.3964	0.3410
Bacillus licheniformis	Food poisoning	Gc00213	4222	4152	23972±5311	23715±5265	24447±5125	0.3388	0.3165	0.2507
Bacteroides fragilis	Diarrhea, Abscesses	Gc00260	5205	4184	30522±6601	30276±6557	31142±6555	0.3746	0.3083	0.2628
Bacteroides thetaiotaomicron	Peritonitis	Gc00126	6260	4778	35884±7660	35563±7597	36516±7751	0.3674	0.3154	0.2658
Bacteroides vulgatus	Opportunistic peritoneal disease	Gc00584	5163	4065	28893±6119	28651±6052	29445±6250	0.3597	0.3302	0.2742
Bartonella henselae	Bacillary angiomatosis	Gc00192	1931	1488	9206±1997	9048±1958	9427±2013	0.2483	0.2778	0.1780
Bartonella quintana	Bacillary angiomatosis, Endocarditis, Trench fever	Gc00191	1581	1142	7710±1699	7544±1666	7896±1705	0.1998	0.2801	0.1564
Bdellovibrio bacteriovorus	Cell lysis	Gc00168	3782	3587	21121±4611	20600±4493	21603±4929	0.1654	0.2899	0.1478
Bordetella parapertussis	Respiratory infection	Gc00147	4773	4185	22589±5550	21833±5067	23661±5670	0.1177	0.1520	**0.0469**
Bordetella pertussis	Respiratory infection	Gc00146	4086	3436	18390±4527	17731±4121	19260±4681	0.1024	0.1534	**0.0432**
Borrelia afzelii	Lyme disease, Acrodermatitis chronica atrophicans	Gc00408	905	856	7545±1708	7616±1778	7590±1624	0.3624	0.4436	0.4718
Borrelia burgdorferi	Lyme disease	Gc00012	910	851	7635±1733	7699±1803	7690±1652	0.3767	0.4331	0.4896
Borrelia garinii	Lyme disease	Gc00211	904	832	7395±1673	7446±1734	7447±1603	0.3963	0.4338	0.4996
Borrelia hermsii	Tick-borne relapsing fever	Gc00792	922	819	7605±1724	7668±1791	7695±1634	0.3786	0.3905	0.4707

Borrelia turicatae	Tick-borne relapsing fever	Gc00791	9173	818	7763±1784	7828±1859	7860±1684	0.3778	0.3851	0.4663
Brucella abortus	Brucellosis	Gc00263	3287	3085	15773±3805	15314±3590	16508±3809	0.1484	0.1516	0.0584
Brucella melitensis	Brucellosis	Gc00074	3294	3198	16505±3949	16052±3726	17265±3953	0.1605	0.1524	0.0623
Brucella suis	Brucellosis, Fever, Infectious abortions	Gc00699	3323	3241	16270±3886	15819±3661	17000±3886	0.1574	0.1585	0.0641
Burkholderia cenocepacia	Chronic infection, Necrotizing Pneumonia	Gc00744	7900	7008	34542±8237	33395±7622	36080±8521	0.1131	0.1604	0.0508
Burkholderia mallei	Glanders, Pneumonia	Gc00518	5900	5852	24519±5794	23776±5247	25689±6100	0.1316	0.1420	**0.0471**
Burkholderia multivorans	Cepacia syndrome	Gc00689	6820	6121	32560±7692	31521±7102	33960±8004	0.1202	0.1668	0.0557
Burkholderia pseudomallei	Pneumonia, Bacteremia, Melioidosis	Gc00520	7100	7183	32157±7571	31144±6874	33613±7981	0.1216	0.1537	**0.0495**
Burkholderia vietnamiensis	Pneumonia	Gc00531	7430	7617	33383±7916	32235±7273	34852±8335	0.1036	0.1624	**0.0486**
Burkholderia xenovorans (fungorum)	Opportunistic infection	Gc00365	9279	8702	46606±10914	45002±10138	48476±11425	0.1010	0.1815	0.0561
Campylobacter concisus	Gingivitis, Periodontitis, Periodontosis, Gastroenteritis	Gc00644	2052	1929	11849±2753	11758±2783	12204±2635	0.3881	0.2453	0.2174
Campylobacter curvus	Gastroenteritis, Periodontal infection	Gc00628	1971	1931	12086±2771	11987±2783	12425±2642	0.3799	0.2564	0.2219
Campylobacter fetus fetus	Bacteremia, Infertility, Septicemia, Meningitis	Gc00466	1773	1719	11959±2849	11858±2909	12283±2689	0.3809	0.2716	0.2374
Campylobacter jejuni jejuni	Diarrhea, Gastroenteritis, Food poisoning	Gc00491	1616	1653	11426±2562	11457±2621	11577±2383	0.4583	0.3761	0.4113
Campylobacter lari	Bacteremia, Diarrhea, Food poisoning, Gastroenteritis	Gc00851	1562	1599	10676±2320	10703±2362	10803±2214	0.4603	0.3854	0.4187
Candidatus Protochlamydia amoebophila	Pneumonia	Gc00185	2414	2031	13565±2812	13476±2807	13470±2836	0.3928	0.4286	0.4961
Chlamydia muridarum	Pharyngitis, Bronchitis, Pneumonia, Respiratory infection	Gc00028	1072	904	6170±1313	6049±1293	6241±1360	0.2159	0.3859	0.2407
Chlamydia trachomatis	Bronchitis, Heart disease, Pharyngitis, Pneumonia, Respiratory infection, Trachoma, Venereal disease	Gc00017	1042	895	6007±1291	5883±1267	6084±1331	0.2044	0.3743	0.2248
Chlamydophila abortus	Pharyngitis, Bronchitis, Pneumonia, Respiratory infection	Gc00262	1144	932	6630±1421	6521±1416	6687±1430	0.2550	0.4157	0.2869
Chlamydophila pneumoniae	Pharyngitis, Bronchitis, Pneumonia, Heart disease, Respiratory infection	Gc00143	1225	1113	7387±1560	7268±1543	7438±1582	0.2572	0.4305	0.2995
Chromobacterium violaceum	Diarrhea, Septicemia	Gc00157	4751	4407	22494±5283	21847±4846	23367±5441	0.1431	0.1898	0.0725
Citrobacter koseri	Bacteremia, Brain abscesses, Meningoencephalitis, Neonatal meningitis	Gc00642	4720	5003	28306±6188	27612±5922	28928±6208	0.1664	0.2961	0.1460

Clostridium botulinum	Botulism	Gc00741	3992	3655	23971±5592	23730±5681	24467±5397	0.3563	0.3178	0.2639
Clostridium difficile	Diarrhea, Colitis, Peritonitis	Gc00420	4290	3742	26564±6412	26124±6506	27255±6245	0.2791	0.2826	0.1996
Clostridium perfringens	Dysenteria, Enterocolitis, Enterotoxemia, Food poisoning, Gas gangrene	Gc00406	3256	2876	18878±4506	18730±4584	19400±4242	0.3898	0.2672	0.2370
Clostridium tetani	Tetanus	Gc00120	2799	2373	18240±4199	18081±4260	18631±4105	0.3730	0.3096	0.2656
Corynebacterium diphtheriae gravis	Diphtheria, Respiratory infection	Gc00163	2488	2272	12322±3052	11767±2926	12830±3196	0.0584	0.1882	**0.0447**
Corynebacterium jeikeium	Endocarditis, Septicemia, Meningitis, Nosocomial infection	Gc00271	2462	2104	11183±2838	10672±2686	11739±3007	0.0595	0.1494	**0.0327**
Corynebacterium urealyticum	Cystitis, Pyelitis, Urinary tract infection	Gc00755	2369	2024	10616±2640	10185±2486	11118±2738	0.0790	0.1560	**0.0401**
Coxiella burnetii	Food poisoning	Gc00692	2016	1930	12824±2643	12776±2602	12901±2601	0.4375	0.4386	0.4086
Ehrlichia chaffeensis	Ehrlichiosis, Human monocytic ehrlichiosis	Gc00353	1176	1105	7907±1729	7804±1763	7993±1720	0.3058	0.3958	0.3023
Enterobacter sakazakii	Meningitis, Septicemia, Necrotizing enterocolitis	Gc00638	4368	4277	24007±5218	23480±4948	24594±5261	0.1912	0.2746	0.1438
Enterococcus faecalis	Bacteremia, Endocarditis, Urinary infection	Gc00831	2739	2701	18533±4090	18323±4088	18760±3981	0.3306	0.3833	0.3026
Escherichia coli	Diarrhea	Gc00716	4686	4126	29392±6442	28654±6199	30050±6555	0.1618	0.2931	0.1433
Francisella philomiragia	Bacteremia, Pneumonia, Septicemia	Gc00722	2045	1911	14775±3443	14622±3505	14931±3225	0.3523	0.4040	0.3326
Francisella tularensis	Tularemia	Gc00241	1892	1603	11937±2715	11810±2749	12039±2605	0.3443	0.4206	0.3424
Fusobacterium nucleatum nucleatum	Periodontal infection	Gc00085	2170	2067	14465±3377	14476±3456	14784±3182	0.4891	0.3064	0.3309
Granulibacter bethesdensis	Chronic granulomatous	Gc00423	2708	2437	13697±3305	13213±3077	14332±3446	0.1017	0.1540	**0.0457**
Haemophilus ducreyi	Genital ulcer, Chancroid	Gc00142	1698	1717	9470±2049	9291±2014	9569±2051	0.2260	0.3986	0.2543
Haemophilus influenzae	Otitis, Meningitis, Septicemia, Sinusitis, Bronchitis	Gc00001	1830	1657	10640±2330	10462±2291	10852±2324	0.2560	0.3138	0.2077
Haemophilus somnus	Arthritis, Myocarditis, Pneumonia	Gc00743	2263	1980	11524±2533	11307±2499	11710±2538	0.2313	0.3476	0.2201
Helicobacter pylori	Gastric inflammation, Ulcer	Gc00879	1673	1567	10030±2103	9998±2111	10098±2064	0.4475	0.4316	0.4091
Klebsiella pneumoniae	Bacteremia, Pneumonia, Urinary tract infection	Gc00841	5641	5425	29171±6561	28491±6262	29909±6474	0.1855	0.2740	0.1406
Legionella pneumophila	Legionellosis	Gc00226	3345	2878	21074±4478	20803±4455	21141±4445	0.3014	0.4681	0.3571
Leptospira borgpetersenii hardjobovis	Leptospirosis	Gc00433	3920	2945	19500±4144	19288±4116	19667±4101	0.3306	0.4150	0.3290
Listeria monocytogenes	Food poisoning, Listeriosis	Gc00186	2905	2821	17937±4140	17600±4133	18305±3994	0.2431	0.3173	0.2044
Mycobacterium abscessus	Broncho-pulmonary infection, Respiratory infection, Wound infection	Gc00729	5067	4920	24543±6186	23410±5792	25751±6546	0.0562	0.1502	**0.0309**

Mycobacterium avium	Respiratory infection, Tuberculosis type pulmonary infection	Gc00462	5475	5120	23353±6182	22194±5736	24817±6661	0.0519	0.1049	**0.0182**
Mycobacterium avium paratuberculosis	Johne's disease, Paratuberculosis, Enteritis	Gc00169	4829	4350	21675±5713	20621±5284	23029±6117	0.0545	0.1047	**0.0184**
Mycobacterium bovis	Tuberculosis	Gc00138	4345	3920	19755±5050	18796±4688	20854±5432	**0.0497**	0.1247	**0.0221**
Mycobacterium bovis BCG	Bovine tuberculosis	Gc00489	4374	3952	19934±5100	18969±4732	21050±5486	0.0503	0.1234	**0.0220**
Mycobacterium leprae	Leprosy, Hanson's disease	Gc00045	3268	1605	9120±2335	8658±2202	9609±2524	**0.0434**	0.1341	**0.0235**
Mycobacterium marinum	Tuberculosis	Gc00784	6636	5423	28943±7417	27506±6949	30423±7938	**0.0466**	0.1452	**0.0266**
Mycobacterium smegmatis	Soft tissue lesions	Gc00461	6988	6716	32660±8582	31102±8048	34619±9073	0.0579	0.1131	**0.0217**
Mycobacterium tuberculosis	Tuberculosis	Gc00578	4424	3950	20040±5081	19073±4713	21124±5483	**0.0492**	0.1294	**0.0232**
Mycobacterium ulcerans	Buruli ulcer	Gc00469	5631	4160	20218±5218	19169±4838	21314±5711	**0.0405**	0.1334	**0.0217**
Mycoplasma arthritidis	Arthritis	Gc00825	820	631	4728±998	4682±1013	4715±1041	0.3462	0.4724	0.4377
Mycoplasma genitalium	Urogenital infection, Respiratory infection, Non-gonococcal urethritis	Gc00002	580	477	4034±836	4017±849	4015±860	0.4314	0.4542	0.4974
Mycoplasma penetrans	Urogenital infection, Respiratory infection	Gc00110	1358	1037	8755±1969	8662±1991	8690±1990	0.3445	0.4304	0.4736
Mycoplasma pneumoniae	Pneumonia, Tracheobronchitis, Respiratory infection	Gc00005	816	689	6227±1330	6293±1364	6172±1331	0.3367	0.4118	0.3333
Neisseria gonorrhoeae	Gonorrhea	Gc00258	2153	2002	10711±2295	10488±2158	10987±2423	0.2008	0.2617	0.1413
Neisseria meningitidis	Meningitis, Septicemia	Gc00026	2272	2063	10104±2187	9898±2081	10378±2282	0.2079	0.2529	0.1401
Neorickettsia sennetsu	Sennetsu fever	Gc00352	859	932	5766±1265	5712±1268	5895±1223	0.3561	0.2934	0.2419
Nocardia farcinica	Nocardiosis, Mastitis	Gc00224	6021	5683	26116±6921	24917±6314	27781±7320	0.0656	0.1010	**0.0189**
Parabacteroides distasonis	Opportunistic peritoneal disease	Gc00583	4811	3850	28852±6264	28651±6238	29468±6162	0.3916	0.2999	0.2639
Pasteurella multocida	Pasteurellosis, Cholera, Septicemia	Gc00048	2257	2015	13526±2959	13315±2905	13708±2895	0.2700	0.3716	0.2578
Porphyromonas gingivalis	Dental plaque, Periodontal infection	Gc00809	2354	2090	13339±2871	13275±2830	13739±2836	0.4251	0.2282	0.2155
Propionibacterium acnes	Acne	Gc00204	2560	2297	12996±3347	12326±3179	13641±3565	**0.0418**	0.1536	**0.0277**
Proteus mirabilis	Encephalitis, Pneumonia, Pyelonephritis, Septicemia, Surgical wound infection, Ulcer, Urinary tract infection, Urolithiasis	Gc00758	4063	3693	22802±5068	22364±4999	23152±4996	0.2293	0.3560	0.2243
Pseudomonas aeruginosa	Opportunistic infection, Nocosomial infection	Gc00432	6537	5892	32928±7633	32280±7041	34239±7495	0.2301	0.1797	0.0946
Pseudomonas mendocina	Endocarditis, Spondylodiscitis	Gc00544	5072	4594	25846±6022	25287±5593	26784±5865	0.2099	0.2026	0.1022
Rickettsia bellii	Epidemic typhus	Gc00661	1528	1476	9556±2129	9492±2186	9582±2043	0.3992	0.4735	0.4200
Rickettsia canadensis	Epidemic typhus	Gc00660	1159	1093	7112±1594	7078±1638	7144±1513	0.4272	0.4576	0.4217
Rickettsia conorii	Rocky Mountain Spotted Fever	Gc00062	1268	1374	8045±1775	7988±1816	8088±1709	0.3922	0.4489	0.3944

Rickettsia prowazekii	Typhus, Rocky Mountain Spotted Fever	Gc00018	1111	835	7260±1621	7249±1677	7277±1536	0.4777	0.4784	0.4684
Rickettsia rickettsii	Rocky Mountain Spotted Fever	Gc00736	1257	1345	7436±1648	7374±1687	7479±1583	0.3747	0.4447	0.3810
Rickettsia typhi	Typhus	Gc00208	1111	838	7186±1602	7164±1653	7212±1528	0.4544	0.4645	0.4432
Salmonella enterica enterica	Food poisoning, Gastroenteritis, Salmonellosis	Gc00833	4685	4318	25874±5635	25253±5414	26418±5671	0.1723	0.2980	0.1502
Salmonella enterica sv Paratyphi A	Food poisoning, Gastroenteritis, Salmonellosis, Typhoid fever	Gc00853	4581	4078	24585±5427	23997±5208	25131±5404	0.1708	0.3036	0.1537
Salmonella enterica sv Paratyphi A	Food poisoning, Gastroenteritis, Salmonellosis, Typhoid fever	Gc00238	4585	4093	28451±6197	27834±5959	29034±6155	0.1746	0.2958	0.1501
Salmonella enterica sv Typhi	Typhoid fever, Salmonellosis, Food poisoning	Gc00066	4809	4395	25456±5596	24843±5356	26012±5603	0.1951	0.3081	0.1688
Serratia proteamaculans	Pneumonia	Gc00647	5448	4891	30454±6854	29710±6536	31154±6733	0.1740	0.2929	0.1462
Shigella boydii	Dysenteria, Food poisoning	Gc00779	4615	4246	22229±4780	21699±4539	22648±4871	0.1686	0.3205	0.1621
Shigella dysenteriae	Dysenteria, Food poisoning	Gc00324	4369	4274	21088±4422	20522±4135	21444±4595	0.1331	0.3344	0.1488
Shigella flexneri	Dysenteria, Gastroenteritis, Food poisoning	Gc00130	4599	4068	23543±5078	22989±4839	24010±5130	0.1729	0.3121	0.1592
Shigella sonnei	Dysenteria, Food poisoning	Gc00323	5039	4461	24441±5249	23862±4983	24912±5377	0.1703	0.3163	0.1606
Staphylococcus aureus	Mastitis, Nosocomial infection	Gc00327	2515	2665	18106±4153	17948±4241	18367±3890	0.3727	0.3684	0.3140
Staphylococcus haemolyticus	Opportunistic infection	Gc00274	2685	2676	17340±3945	17207±4023	17582±3713	0.3869	0.3711	0.3238
Streptococcus agalactiae	Meningitis	Gc00100	2211	2094	12858±2930	12638±2948	13092±2877	0.2601	0.3353	0.2285
Streptococcus gordonii	Dental plaque, Endocarditis, Periodontal infection	Gc00643	2196	2051	13014±2945	12895±2958	13286±2829	0.3647	0.3106	0.2602
Streptococcus mutans	Dental caries	Gc00109	2030	1960	12205±2735	12118±2770	12415±2651	0.3939	0.3406	0.3015
Streptococcus pneumoniae	Meningitis, Pneumonia, Otitis media	Gc00437	2000	1914	12092±2702	11967±2722	12331±2619	0.3458	0.3182	0.2580
Streptococcus pyogenes	Bone infection, Endocarditis, Mastoiditis, Meningitis, Myositis, Necrotizing fasciitis, Otitis, Pharyngitis, Pneumonia, Sinusitis, Tonsillitis	Gc00285	1838	1865	10563±2399	10376±2396	10801±2403	0.2530	0.2985	0.1976
Streptococcus suis	Meningitis, Endocarditis, Septicemia, Arthritis	Gc00546	2096	2186	12022±2696	11867±2694	12293±2660	0.3108	0.2955	0.2225
Treponema pallidum pallidum	Syphilis	Gc00789	1139	1028	6849±1410	6676±1310	7002±1533	0.1435	0.2829	0.1264
Tropheryma whipplei	Whipple's disease	Gc00123	925	783	5486±1270	5331±1245	5654±1276	0.1467	0.2404	0.1079
Ureaplasma parvum	Respiratory infection, Urinary tract infection	Gc00742	751	609	5510±1196	5469±1221	5498±1212	0.3847	0.4797	0.4533
Vibrio cholerae	Cholera, Food poisoning	Gc00557	4148	3875	22495±4906	21872±4736	22891±5091	0.1368	0.3337	0.1552
Vibrio parahaemolyticus	Gastroenteritis	Gc00124	5165	4832	27525±6192	26701±6047	28124±6419	0.1263	0.3035	0.1330

Yersinia enterocolitica	Gastroenteritis, Food poisoning	Gc00481	4615	3979	25630±5744	24958±5547	26182±5722	0.1696	0.2958	0.1485
Yersinia pestis	Bubonic and Pneumonic plague	Gc00538	4517	3850	23982±5310	23393±5121	24516±5316	0.1566	0.3041	0.1466
Yersinia pseudotuberculosis	Gastroenteritis	Gc00776	4695	4150	25461±5702	24847±5530	26020±5664	0.1766	0.3005	0.1556

1. Non redundant, complete bacterial genomes classified as “relevant human pathogens” in the Genome OnLine Database v2.0 (GOLD database update October 30th 2008, http://www.genomesonline.org/) under analysis
2. Genome size in Kilobases
3. Number of Open Reading Frame in the genome
5. Mean Number of potential epitopes of the genome under analysis, equivalent to the affinity of 1% of the best binder peptides, recognized by the different study populations
5. p value for the comparison between groups indicated. Students' t-test. Statistically significant comparisons ($p<0.05$) are indicated in bold.

Table 3. HLA-DRB1 allelic frequency analysis of the study population.

	Controls		**Chronic Sarcoidosis**		**Loefgren**		**CTR vs Chr Sarc**	**CTR vs Loef**	**Chr Sarc vs Loef**
HLA-DRB1 Allele	N	Freq	N	Freq	N	Freq	χ-square p	χ-square p	χ-square p
0101	57	0.064	12	0.055	6	0.077	0.7250	0.8312	0.6627
0102	4	0.004	0	0.000	0	0.000	0.7152	0.7413	NA
0103	3	0.003	0	0.000	0	0.000	0.8932	0.5811	NA
0301	103	0.115	21	0.095	16	0.205	0.4746	**0.0321***	**0.0201***
0302	2	0.002	0	0.000	0	0.000	0.8519	0.3764	NA
0401	57	0.064	13	0.059	3	0.038	0.9199	0.5189	0.6876
0402	6	0.007	1	0.005	0	0.000	0.9108	0.9777	0.5872
0403	10	0.011	0	0.000	1	0.013	0.2393	0.6693	0.5872
0404	41	0.046	4	0.018	2	0.026	0.0936	0.5852	0.9473
0405	9	0.010	1	0.005	0	0.000	0.7048	0.7841	0.5872
0406	1	0.001	0	0.000	0	0.000	0.4471	0.1221	NA
0407	9	0.010	1	0.005	4	0.051	0.7048	**0.0116***	**0.0246***
0408	8	0.009	0	0.000	0	0.000	0.3358	0.8528	NA
0701	74	0.083	15	0.068	6	0.077	0.5644	0.9725	0.9986
0801	18	0.020	8	0.036	1	0.013	0.2384	0.9832	0.5100
0803	6	0.007	1	0.005	0	0.000	0.9108	0.9777	0.5872
0804	2	0.002	2	0.009	0	0.000	0.3716	0.3764	0.9698
0812	1	0.001	0	0.000	0	0.000	0.4471	0.1221	NA
0901	10	0.011	4	0.018	1	0.013	0.6194	0.6693	0.8444
1001	6	0.007	3	0.014	1	0.013	0.5435	0.9313	0.6039
1101	78	0.087	29	0.132	4	0.051	0.0598	0.3769	0.0823
1102	5	0.006	0	0.000	1	0.013	0.5831	0.9777	0.5872
1103	7	0.008	2	0.009	0	0.000	0.8156	0.9313	0.9698
1104	51	0.057	5	0.023	2	0.026	0.0555	0.3620	0.7725
1201	10	0.011	2	0.009	0	0.000	0.9244	0.7234	0.9698
1202	4	0.004	1	0.005	0	0.000	0.5831	0.7413	0.5872
1301	94	0.105	16	0.073	9	0.115	0.1876	0.9283	0.3524
1302	52	0.058	18	0.082	3	0.038	0.2543	0.6406	0.3039
1303	13	0.015	0	0.000	1	0.013	0.1474	0.7091	0.5872
1305	1	0.001	1	0.005	0	0.000	0.8519	0.1221	0.5872
1326	1	0.001	0	0.000	0	0.000	0.4471	0.1221	NA
1401	33	0.037	13	0.059	3	0.038	0.1964	0.8079	0.6876
1404	1	0.001	0	0.000	0	0.000	0.4471	0.1221	NA
1407	2	0.002	0	0.000	0	0.000	0.8519	0.3764	NA
1501	86	0.096	44	0.200	13	0.167	**0.00002^**	0.0754	0.6344
1502	9	0.010	0	0.000	0	0.000	0.2829	0.7841	NA
1601	19	0.021	2	0.009	1	0.013	0.3620	0.9305	0.7065
1602	1	0.001	1	0.005	0	0.000	0.8519	0.1221	0.5872

*Bonferroni corrected p>0.05

^Bonferroni corrected p=0.0013

Table 5. Analysis of the HLA-DRB1 alle frequency in the study population subgroups ranked by the percentile recognition of *M. avium* immunome.

A. 76°-100° percentile of M. avium immunome

	Loefgren		Chronic Sarcoidosis		Controls		CTR vs Chr Sarc	CTR vs Loef	Chr Sarc vs Loef
HLA-DRB1 Allele	N	Freq	N	Freq	N	Freq	χ-square p	χ-square p	χ-square p
0101	4	0,142857	2	0,0625	12	0,065934	0,752637	0,29569	0,545972
0102	0	0	0	0	3	0,016484	0,933209	0,86417	NA
0301	11	0,392857	12	0,375	71	0,39011	0,972137	0,856908	0,901165
0401	1	0,035714	1	0,03125	5	0,027473	0,644641	0,714702	0,532173
0404	2	0,071429	1	0,03125	6	0,032967	0,62522	0,645852	0,905486
0405	0	0	0	0	1	0,005495	0,324589	0,279601	NA
0408	0	0	0	0	2	0,010989	0,688938	0,625769	NA
0701	1	0,035714	3	0,09375	14	0,076923	0,976218	0,693497	0,703662
0801	1	0,035714	0	0	4	0,021978	0,889536	0,824374	0,94628
1101	2	0,071429	2	0,0625	11	0,06044	0,721656	0,844175	0,703662
1102	1	0,035714	0	0	0	0	NA	0,279601	0,94628
1104	1	0,035714	0	0	10	0,054945	0,366004	0,975771	0,94628
1301	3	0,107143	0	0	17	0,093407	0,147756	0,908241	0,191528
1305	0	0	0	0	1	0,005495	0,324589	0,279601	NA
1401	1	0,035714	4	0,125	12	0,065934	0,419583	0,844175	0,435253
1501	0	0	7	0,21875	13	0,071429	**0,020827**	0,298833	**0,025735**

B. 26°-50° percentile of M. avium immunome

	Loefgren		Chronic Sarcoidosis		Controls		CTR vs Chr Sarc	CTR vs Loef	Chr Sarc vs Loef
HLA-DRB1 Allele	N	Freq	N	Freq	N	Freq	χ-square p	χ-square p	χ-square p
0401	1	0,1	6	0,083333	31	0,186747	0,067584	0,788224	0,669285
0402	0	0	1	0,013889	2	0,012048	0,60619	0,235259	0,245069
0404	0	0	1	0,013889	11	0,066265	0,169496	0,866468	0,245069
0405	0	0	1	0,013889	3	0,018072	0,750289	0,407113	0,245069
0408	0	0	0	0	3	0,018072	0,60619	0,407113	NA
0701	0	0	3	0,041667	18	0,108434	0,155789	0,574302	0,809451
0801	0	0	6	0,083333	6	0,036145	0,227886	0,775284	0,763971
0804	0	0	0	0	1	0,006024	0,666602	0,054871	NA
1101	2	0,2	13	0,180556	22	0,13253	0,446225	0,897053	0,773789
1102	0	0	0	0	3	0,018072	0,60619	0,407113	NA!
1104	0	0	1	0,013889	3	0,018072	0,750289	0,407113	0,245069
1301	3	0,3	11	0,152778	18	0,108434	0,456291	0,189295	0,477119
1302	0	0	6	0,083333	15	0,090361	0,941676	0,681216	0,763971
1501	4	0,4	23	0,319444	28	0,168675	**0,01502**	0,155656	0,881651
1502	0	0	0	0	2	0,012048	0,871005	0,235259	NA

Table 6. Comparison of the recognition of the HLA-DR immunomes of *M. avium* and *M. tuberculosis* in the study populations in comparison to tuberculosis patients.

Genome	Mean Number of potential epitopes (1% affinity)[1]				Student's t-test (p-value)2		
	Controls	Chronic Sarcoidosis	Loefgren	Tuberculosis[3]	TB vs Control	TB vs Loefgren	TB vs Chronic Sarcoidosis
M. avium	23353±6182	22194±5736	24817±6661	20709±6894	**0.0002**	**>0.00001**	**0.003**
M. tuberculosis	20040±5081	19073±4713	21124±5483	17925±5732	**0.0008**	**>0.00001**	**0.01**

1. Mean Number of potential epitopes in the different study population of the *M. avium* and *M. tuberculosis*, equivalent or higher than the affinity of 1% of the best binder peptides for the subject's HLA-DR molecules
2. p value for the comparison between groups indicated. Students' t-test. Statistically significant comparisons ($p<0.05$) are indicated in bold
3. Matched controls of the tuberculosis patient populations did not differ from the control population of this study in terms of the number of potential epitopes recognized (*M. avium*: 23353±6182 vs 22591±7612; *M. tuberculosis*: 20040±5081 vs 19472±4972; $p>0.05$ all comparisons).

THANKS TO

My family, my dear husband Simone, and my beloved son Davide

My friends and coworkers Alessia Comandini, and Floriana Berretta, for it has been a pleasure to work with them.

The head of laboratory Massimo Amicosante for the enthusiasm he transmitted to me.

And a special thanks to my professor and my tutor prof. Cesare Saltini for his constant support and education for my career.