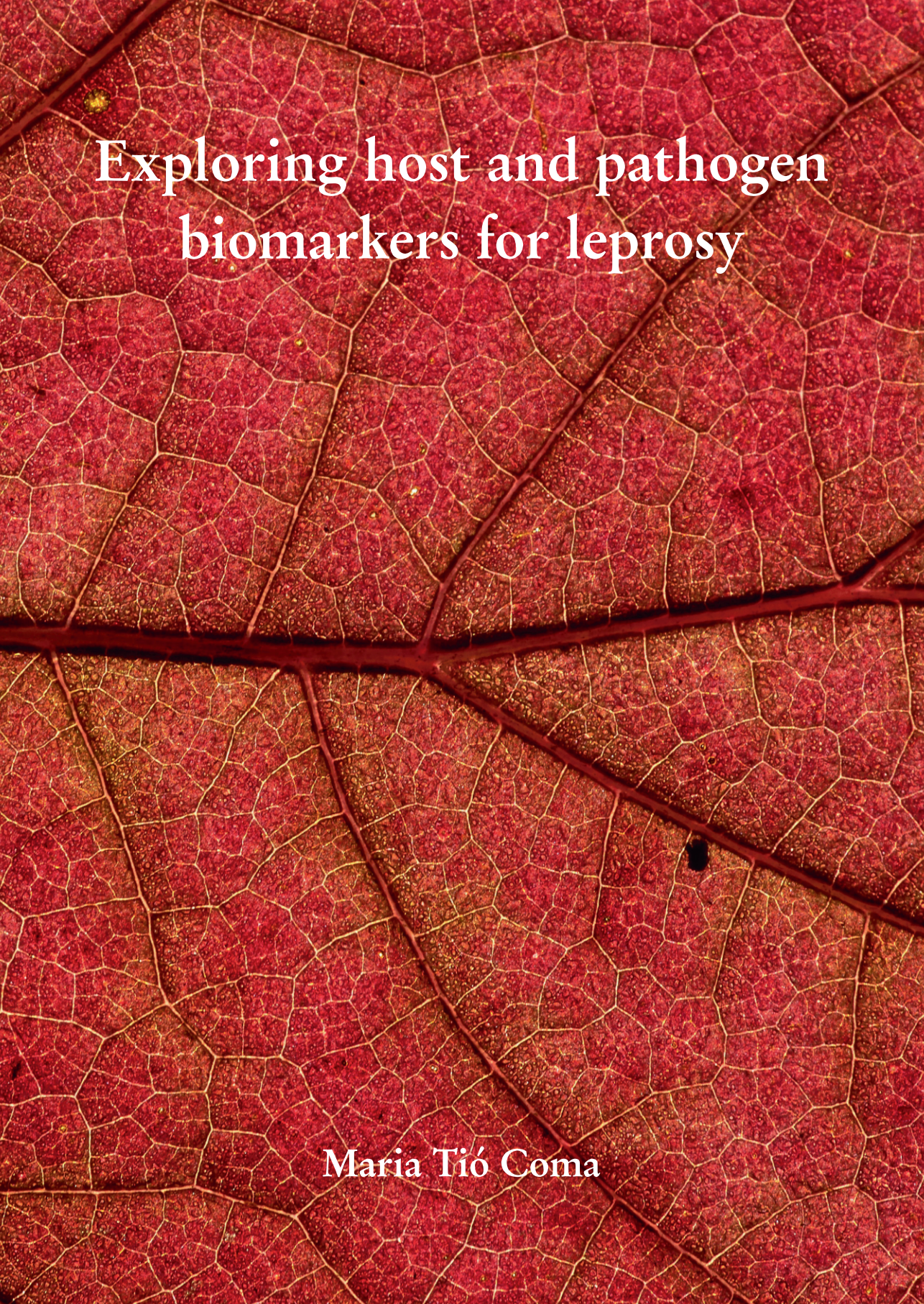
Exploring host and pathogen biomarkers for leprosy

Tio Coma, M.



Exploring host and pathogen biomarkers for leprosy

Maria Tió Coma

Exploring host and pathogen biomarkers for leprosy

Maria Tió Coma

Exploring host and pathogen biomarkers for leprosy
PhD thesis, Leiden University, The Netherlands

Author: Maria Tió Coma
Cover design: Maria Tió Coma
Lay-out: Sergio Arredondo Alonso and Maria Tió Coma
Printed by: Print Amsterdam BV

The research presented in this thesis was supported by an R2STOP Research Grant from effect:hope together with the Mission to End Leprosy, the Order of Malta-Grants-for-Leprosy-Research (MALTALEP), the Q.M. Gastmann-Wichers Foundation, the Netherlands Leprosy Relief Foundation (NLR, ILEP# 702.02.73 and the Leprosy Research Initiative (LRI, ILEP# 703.15.07) both together with the Turing Foundation.

Printing of this thesis was financially supported by the Q.M. Gastmann-Wichers Foundation.

Exploring host and pathogen biomarkers for leprosy

Proefschrift

ter verkrijging van
de graad van doctor aan de Universiteit Leiden,
op gezag van rector magnificus prof.dr.ir. H. Bijl,
volgens besluit van het college voor promoties
te verdedigen op donderdag 28 oktober 2021
klokke 13.45 uur

door

Maria Tió Coma

geboren te Gelida (Barcelona), Spanje
in 1993

Promotores:
Prof. dr. A. Geluk
Prof. dr. J.H. Richardus

Leden Promotiecommissie:
Prof. dr. J. J. Goeman
Prof. dr. P. Slagboom
Prof. dr. C.H. Hokke
Prof. dr. M.T. Mira (Pontifícia Universidade Católica do Paraná)
Dr. C. Avanzi (Colorado State University)

Table of Contents

1

General Introduction

Maria Tió-Coma

Highlights & Hallmarks of leprosy

Leprosy, a chronic disease affecting the skin and nerves, is caused by *Mycobacterium leprae* or *Mycobacterium lepromatosis* (1). Leprosy has been reported in humans for millennia with ancient reports in Roman, Egyptian and Greek manuscripts. In Europe it was endemic during the Middle Ages until the sixteenth century when it dissipated. In 1873, during his work in a leprosarium in Bergen (Norway), Gerhard Armauer Hansen discovered *M. leprae* in unstained material. *M. leprae* is unculturable in present media and this has hindered the study of the disease, bacterial transmission and pathogenesis. Nevertheless, leprosy is a curable disease currently treated with multi-drug therapy (MDT) (2). The first antimicrobial treatment for leprosy was introduced in 1943 and was based on sulphones. In 1950s dapsone monotherapy was established, in 1982 MDT was recommended to multibacillary (MB) patients and in 2002 this was extended to all patients (3).

Leprosy classification

Leprosy can be distributed according to the Ridley Jopling (4) or the World Health Organization (WHO) classifications (5) based on clinical features and the immunological response against the leprosy bacilli. The Ridley Jopling classification consist of five forms, at one end of the spectrum tuberculoid leprosy (TT) patients display a low bacterial load and a strong cell-mediated immune response, characterized by Th1 and Th17 cells (6, 7), granuloma formation and elimination of bacteria. At the other end, lepromatous leprosy (LL) is characterized by a high bacterial load and humoral immunity with Th2 cells but almost no protective cell mediated immunity, allowing accumulation of high numbers of bacilli around foamy macrophages (8-12). However, the majority of leprosy patients present unstable borderline forms, between LL and TT and classified as borderline lepromatous (BL), borderline borderline (BB) and borderline tuberculoid (BT) (9).

The WHO classifies leprosy in two groups, paucibacillary (PB) and multibacillary (MB) leprosy and is based on the number of skin lesions and presence of acid fast bacilli in slit skin smears (SSS) (5). Leprosy patients with up to five skin lesions and negative SSS are considered PB whereas patients with six or more skin lesions and a positive SSS are classified as MB.

Leprosy reactions

Leprosy reactions present a major challenge in the prevention of (permanent) disabilities (13) because reactions can cause peripheral nerve damage and are the major cause of irreversible neuropathy (14). Leprosy reactions are increased inflammation episode which may occur at any time before, during or after MDT (15, 16). Up to 30-50% of leprosy patients may experience leprosy reactions, however, these are most common in unstable BL patients presenting a high number of bacilli (17). There are two types of reactions: type

1 reactions (T1R) or reversal reactions (RR) and type 2 reactions (T2R) or erythema nodosum leprosum (ENL). While T1R are characterized by inflammation of nerves and/or skin lesions, T2R, which can be very painful, are characterized by erythematous skin lesions which may present in combination with fever and/or malaise (18). T1R are delayed hypersensitivity reactions caused by an increased cellular immune response against the bacteria mediated via Th1 cells resulting in a switch from borderline leprosy to TT leprosy (19, 20). Several cytokines and chemokines such as IFNγ, IL-12, CXCL10 and IL-6 show higher levels in lesions or plasma during T1R (18). Moreover, anti-helminth treatment (21-24), extensive anti-TNF-α therapy (16, 20), HIV highly active antiretroviral therapy (HAART) or BCG vaccination (25) produce a shift from Th2 to Th1 cells that may trigger T1R.

ENL are likely immune complex-mediated reactions, however, the mechanism of action is still unclear (18). Nevertheless, lower expression of Treg cells and higher CD4/CD8 ratios are observed (26) in patients presenting ENL.

Leprosy epidemiology

In 2000 leprosy elimination as a public health problem, defined as a prevalence of less than 1 case per 10,000 in the world population, was achieved (2). However, the number of new cases has been steady during the last decade, with 208,619 new leprosy cases worldwide in 2018 (27). The majority of new cases occur in low- and middle-income countries where leprosy is still endemic. India with 120,334 new cases in 2018 is the country with the highest number of new cases, followed by Brazil with 28,660 new cases and Indonesia with 17,017 new cases. Numerous studies of this thesis were performed in Bangladesh, a leprosy endemic country reporting up to 3,729 new leprosy cases in 2018 (27).

The incubation period of leprosy is significantly long being 5-10 years for MB cases and 2-5 for PB cases (28). Moreover, only 5% of people exposed to *M. leprae* become infected, and from those, barely 20% eventually develop the disease (29). The long incubation period and the low amount of leprosy progressors within individuals exposed to *M. leprae* in addition to the limited awareness of leprosy by the public and healthcare providers as well as the social stigma hinders the identification of new cases, particularly amongst household contacts (HC) of leprosy patients who are at highest risk of developing disease (30).

The incidence of leprosy in females is lower than in males. Women represent around 35-37% of the new cases, however, this lower rate may be influenced by an under-diagnosis due to limitations women face in endemic countries such as restricted access to health services, illiteracy and low status (28). Moreover, MB cases are more common in men than in women. From all new cases children represent 9% and since these are recent infections, it indicates that *M. leprae* transmission is still ongoing.

Leprosy: the pathogen side

Transmission and One Health

M. leprae transmission is still not completely understood, however, aerosol transmission via the respiratory route and skin-to-skin contact are assumed to be the most probable ways of bacterial dissemination (31, 32). Besides bacterial exposure, other risk factors have been shown to be associated with leprosy development, including host genetic polymorphisms (33-37), close contact with untreated, MB patients (38), infection with soil transmitted helminths (21), immunosuppression (38), nutritional factors (39), food shortage (40), living conditions (41, 42) and individual characteristics (43, 44). HC of leprosy patients present the highest risk to develop leprosy (45), due to the continuous contact with a person infected with the leprosy bacilli. Besides, they might not develop leprosy, but bear *M. leprae* serving as asymptomatic carriers who contribute to transmission.

M. leprae and *M. lepromatosis* have been identified in several animals as well as environmental samples representing a reservoir that could potentially become a source of infection (46, 47). Moreover, a leprosy-like disease caused by *Mycobacterium lepraemurium*, *Mycobacterium tarwinense* or *Mycobacterium lepraefelis* has been reported in cats (48).

In the British Isles, Scotland and Ireland *M. leprae* and *M. lepromatosis* were detected in red squirrels causing a leprosy-like disease (46, 49-52). However, molecular testing by PCR showed absence of the leprosy bacilli in squirrels from France, Germany, Switzerland, Italy and Mexico (53), indicating that leprosy diseases and reservoirs of *M. leprae* and *M. lepromatosis* in squirrels are only present in the British Isles.

M. leprae was first found in nine-banded armadillos from southern United States in 1977 (54) and was thereafter extensively studied (55-57). Probable zoonotic transmission of *M. leprae* from armadillos was also identified in the southeastern United States where wild armadillos and patients were infected with the same genotype (3I-2-v1) (58). This genotype was also identified both in armadillos and humans from Florida (59) and is closely related to *M. leprae* strains circulating in medieval Europe, suggesting that leprosy arrived in the United States from Europe before it disappeared in the sixteenth century. Armadillos bearing *M. leprae* have also been found in Brazil, Mexico, and Colombia (60-63). A study in Pará (Brazilian Amazon), Brazil found a higher rate of leprosy amongst armadillo hunters and people who eat armadillo meat more than 12 times per year (60). However, other studies from Brazil found no association with armadillo meat consumption and leprosy in Curitiba (Paraná) (64) and absence of *M. leprae* in armadillos from Coari (Amazonas state) (65).

Furthermore, *M. leprae* infections have been reported in wild and captive nonhuman primates (66-68). The *M. leprae* strain identified in a captive nonhuman primate (branch 0, 3K genotype) in The Philippines was phylogenetically close to the human strains, suggest-

ing possible transmission between humans and nonhuman primates (66). Whereas wild chimpanzees in West Africa were found to be infected with *M. leprae* strains belonging to different and rare genotypes: 2F and 4N/O. Genotype 4N/O was also identified in a captive nonhuman primate in West Africa (67).

In addition to animals, *M. leprae* DNA has also been detected in the environment, namely in soil (47, 69-74) and water (74-76) from India and Brazil. It has been described that viable bacilli could survive in the environment in free living amoebic cysts up to 8 months (77). Thus, these environmental reservoirs should be taken into consideration to investigate transmission chains.

Pathogen genomics

M. leprae genome was first sequenced in 2001 (78) leading to the identification of several repetitive elements such as RLEP which is currently used as the PCR target for *M. leprae* DNA detection (79-82). *M. leprae* underwent a massive gene reduction resulting in a genome of 3.27 Mb, whilst the closely related *Mycobacterium tuberculosis* possess a genome of 4.41 Mb (78). The majority of *M. leprae* genome is composed of pseudogenes which correspond to active genes in *M. tuberculosis*. Several of the genes absent in *M. leprae* are involved in vital metabolic activity, thus converting *M. leprae* into an obligate intracellular bacterium. *M. leprae* phylogeny is composed of four genotypes (1-4) (83) and 16 subtypes (A-P) (58, 84). This phylogenetic information has been recovered from *M. leprae* genomes obtained from contemporary leprosy patients (83-85), but also from ancient skeletons (86-90), red squirrels (46, 53), armadillos (58, 59) and non-human primates (66, 67). Extensive whole genome sequencing (WGS) from different sources has proven to be a relevant tool to study *M. leprae* transmission.

The genome of *M. leprae* is highly conserved, presenting a mutation rate of 18 ± 30 mutations per 1000 years (87), which facilitates reconstruction of historic *M. leprae* transmission and human migration patterns (91). The most ancestral linages of *M. leprae* are 3K-0 or branch 0 followed by 3K-1 or branch 5, which are predominant in modern East Asia, particularly in China, Japan and Korea (85, 87). In Medieval Europe different *M. leprae* genotypes were present reflecting ancient human migrations (91). Genotypes 2F and 3I were common in north Europe, whilst in Hungary, Byzantine Turkey and the Czech Republic subtypes 3M and the ancestral lineage 3K were present, likely introduced trough the Silk Road from central Asia (86, 92). The origin of *M. leprae* is still ambiguous and two possible locations have been suggested: Western Eurasia or East Asia and the Middle East. Genotype 3K-0 is found in modern East Asia whilst this ancestral lineage was also present in medieval East Europe indicating that either of these regions could be the origin of leprosy (87). Genotype 3K-0 possibly spread through Europe, the Middle East and East Asia giving rise to the different genotypes 3 whilst genotype 1 is predominant in south Asia.

Genotype 4 is thought to have evolved from European genotype 3 into West Africa and later arrived to America through the slave trade. In addition to genotype 4, subtype 3I has also been identified in America, suggesting a likely introduction of leprosy by European immigrants (83, 84).

In 2008 *M. lepromatosis* was identified in two patients presenting diffuse lepromatous leprosy and a set of genes also present in *M. leprae* were sequenced (1). Although *M. lepromatosis* is closely related to *M. leprae*, 2.1% of divergence can be observed in the highly conserved bacterial marker 16S rRNA (1). In addition, protein coding genes show 93% nucleotide sequence identity between the two species and pseudogenes 82% (93). It was suggested that both *M. leprae* and *M. lepromatosis* diverged from the most recent common ancestor around 13.9-20 Mya (93, 94). Since *M. lepromatosis* has retained all functions required to infect Schwann cells of the peripheral nervous system it can also cause leprosy.

The host side

Host transcriptomic biomarkers for leprosy diagnosis

Leprosy, particularly PB leprosy, is difficult to diagnose and current diagnostic tests are not sensitive and/or specific enough, thus diagnosis still strongly relies on clinical symptoms. Early detection of leprosy together with identification of *M. leprae* asymptomatic carriers is crucial for an effective intervention aimed at stopping *M. leprae* transmission. Besides, an early diagnosis can prevent the development of disabilities. Late diagnosis or misdiagnosis of leprosy, particularly in non-endemic areas, is common due to the infrequent encounters of health personnel with this disease. Leprosy diagnosis is usually assisted by detection of acid-fast bacilli in tissue smears, lymph or histological sections using a Ziehl-Neelsen staining (95). This technique is not sensitive enough for the diagnosis of PB cases and also during the early stages of leprosy. Currently, molecular techniques based on (quantitative) PCRs to detect *M. leprae* DNA are also in use to support the diagnosis (79, 96, 97). While the sensitivity of such molecular techniques is higher than microscopy it is still challenging to detect *M. leprae* DNA in PB cases. In addition, it has been observed that HC, the group with the highest risk to develop leprosy, may be asymptomatic carriers of *M. leprae* (31, 70, 98-104). Therefore, presence of *M. leprae* DNA without clinical symptoms in HC is not a useful predictive marker of leprosy (105). Nevertheless, detection of *M. leprae* carriage or infection can be applied for targeted prophylactic treatment to reduce transmission.

Several host markers have been proposed to diagnose leprosy based on the immunological response to *M. leprae* (106-111). The most commonly used is detection of serum levels of IgM anti-*M. leprae* phenolic glycolipid I (PGL-I) which has been implemented as a point-of-care (POC) test (97, 109, 112). However, the majority of the available host markers fail to diagnose PB leprosy cases (106, 113) and some markers, such as PGL-I, could also be

present in HC who have been exposed to *M. leprae* but do not develop leprosy.

Transcriptomic host profiles have been proven to be effective to identify correlates of risk for tuberculosis (114, 115). In leprosy, some studies employed transcriptomics to investigate the immunological response to *M. leprae* but others also aimed at identifying potential biomarkers for leprosy or leprosy reactions (116-125). However, the studies that focused on biomarkers to identify leprosy were employed after clinical symptoms, thus are not useful for early or predictive diagnosis. In addition, the biosamples used for these transcriptional studies were skin biopsies, nerve biopsies or cell culture, and although these are particularly useful to study the pathogenesis of the disease they are not practical for POC diagnostic tests. Instead, blood or urine are less invasive and easy-to collect samples that would be favored for diagnosis. For this reason, blood samples were employed in a study to identify longitudinal differential gene expression (DGE) during T1R (126). Using dual-color Reverse Transcriptase Multiplex Ligation-dependent Probe Amplification (dcRT-MLPA) an increased expression during T1R was observed in genes associated with cytotoxic T-cell response, IFN-induced genes and *VEGF* whilst a decrease was found in T-cell regulation genes. In line with this, a blood signature including type I IFN components, autophagy, parkins and Toll like receptors was also identified during T1R (127). In a different study, applying RNA-Seq a 44-gene signature in blood was established that could differentiate between leprosy patients suffering a T1R and no leprosy reaction (121). The signature was formed by pro-inflammatory regulators, arachidonic acid metabolism mediators and regulators of antiinflammation. However, to date none of these studies has resulted in application of host transcriptomics to diagnostic (POC) tests.

Influence of host genetics on leprosy

Leprosy is an infectious disease, however, host genetics strongly influences the outcome of the disease. This is observed from the low rate of disease development in people exposed to *M. leprae* (29) and the wide spectrum of leprosy disease whilst bacterial genome variation is very limited, suggesting that host genetics play a crucial role in this variation. In the last 20 years the genetics of leprosy has been extensively investigated through genome-wide association studies (GWASs) and candidate gene association studies (CGASs) (128, 129). In 2009 the first GWAS in leprosy was performed in a Chinese population (33) and since then many studies have used GWAS to characterize the association of genetics with leprosy (130-133).

Most of the genes that have been associated with leprosy or leprosy reactions are immune related and involved in the innate or adaptive immunity. Numerous studies have explored the association of genetics with leprosy per se, leprosy type and leprosy reactions.

Leprosy per se, referring to the presence or absence of disease, was the first subject to be researched by genetic studies. Several genes have been associated with leprosy per se,

such as *TLR1*, *SLC11A1*, *VDR*, *NOD2, LACC1* and *TYK* or variants in the promoter regions of *PRKN*, *IL10* and *LTA* (Figure 1) (129). The major histocompatibility complex (MHC) has been strongly linked to leprosy, however, the large variation of HLA alleles has hampered its study (129). Most of the variants associated with leprosy per se identified in GWASs are non-coding, such as eQTLs for *NOD*2 or *IL18RAP* (134).

The association of genetic variation to leprosy polarization has been less studied than leprosy per se. Besides, classification into a leprosy subtype, particularly using the Ridley Jopling classification, depends on the physician's expertise and definition used, affecting the resulting classification. Most of the genes related to leprosy type have been found in CGASs whereas GWASs have focused mostly on leprosy per se (129). Genes such as *IL10*, *MBL2*, *MRC1*, *TGFB1*, *TLR2*, *TNF*, *CUBN* and *NEBL* have been associated to the subtype of leprosy (Figure 1) (135-140).

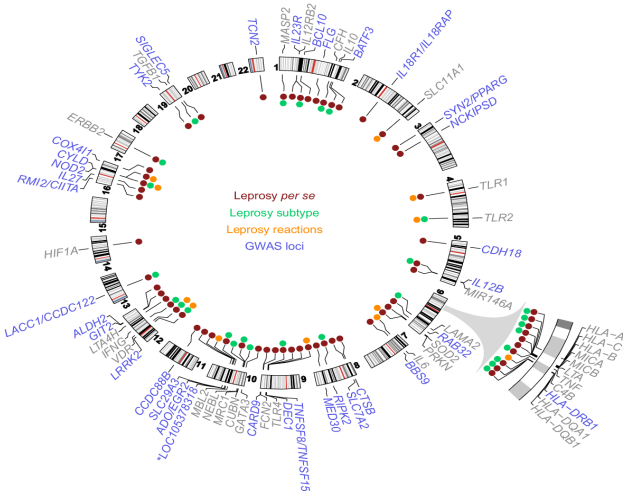


Figure 1. Genes and genome-wide association studies (GWAS) loci associated with leprosy per se (red), leprosy types (green) and leprosy reactions (orange). The human chromosomes 1–22 are presented in the circular plot. In blue, genes identified by GWAS and in gray genes identified using other approaches. Figure retrieved from Fava et al. (129).

CGASs have also been employed to study the relationship between host genetics and leprosy reactions. Since T1R are more common than ENL, association of genetics with T1R has been more widely investigated than ENL. Nevertheless, a study identified that allelic variance or absence of MHC class III protein C4B was associated with ENL occurrence (141).

Several *LRRK2* and *PRKN* variants as well as eQTLS for *LRRK2* were suggested to be linked with occurrence of T1R (142, 143). The mutations identified in leprosy patients presenting T1R have also been found in patients with Parkinson's disease, showing an overlapping inflammatory profile. In addition, eQTLs for lncRNA ENSG00000235140 (*LOC105378318*)

represent a risk for development of T1R (144). A selection of genes found to be associated with leprosy per se have also been linked to the occurrence of T1R: *PPARG*, *TNFSF8*, *TNFSF15*, *NOD2*, *LRRK2*, *TLR1*, and *TLR2* (Figure 1) (36, 135, 142, 144-148).

Outline of this thesis

The aim of the research described in this thesis was to combine the study on the pathogen *M. leprae* with the identification of transcriptomic and genetic host biomarkers associated with leprosy to aid the development of diagnostic tests as well as reduce transmission.

First, pathogen transmission was investigated through a One Health approach which incorporated analyses of *M. leprae* DNA derived from human, environmental and animal samples. In **chapter 2**, we identified leprosy patients and their asymptomatic HC carrying or infected with *M. leprae* in Bangladesh. We explored *M. leprae* genetic variation between individuals and intra-individually by WGS to identify transmission patterns. In **chapter 3**, we assessed whether *M. leprae* or *M. lepromatosis* were present in the environment. For this, we analyzed by PCR soil samples from the homes of leprosy patients in Bangladesh, the area where squirrels infected with *M. leprae* and *M. lepromatosis* were found in the British Isles and around holes of armadillos in Suriname. Since leprosy was previously observed in squirrels from the British Isles, in **chapter 4** we investigated whether Dutch and Belgian squirrels infected with *M. leprae* or *M. lepromatosis* could be identified.

Next, transcriptomic and genetic host biomarkers were identified to predict leprosy, leprosy reactions and genetic markers associated with susceptibility for leprosy. In **chapter 5**, we aimed to develop a transcriptomic signature that could predict leprosy development in HC of leprosy patients, 4 to 61 months before clinical symptoms. For this purpose, we analyzed gene expression variation in leprosy progressors using RNA-Seq: gene expression of leprosy progressors before clinical symptoms was compared with HC who remained without leprosy symptoms (cross-sectional analysis) and with the timepoint of clinical diagnosis, when symptoms were present (longitudinal analysis). In **chapter 6**, we studied gene expression differences in leprosy patients from Bangladesh, Brazil, Ethiopia and Nepal who developed leprosy reactions and identified a signature that predicted reversal reactions in leprosy patients before onset. In **chapter 7**, we investigated the association of 11 host genetic markers with leprosy in Bangladesh through a family-based study consisting of leprosy patients and both progenitors.

Finally, in **chapter 8** all findings and conclusions of this thesis are summarized and discussed.

References

1. Han XY, Seo Y-H, Sizer KC, Schoberle T, May GS, Spencer JS, et al. A New Mycobacterium Species Causing Diffuse Lepromatous Leprosy. American Journal of Clinical Pathology. 2008;130(6):856-64.
2. World Health Organization. Fact sheets: Leprosy Genova: World Health Organization; 2019

[Available from: https://www.who.int/news-room/fact-sheets/detail/leprosy.
3. Gelber RH, Grosset J. The chemotherapy of leprosy: an interpretive history. Leprosy review. 2012;83(3):221-40.
4. Ridley DS, Jopling WH. Classification of leprosy according to immunity. A five-group system. Int J Lepr Other Mycobact Dis. 1966;34(3):255-73.
5. Kumar B, Uprety S, Dogra S. Clinical diagnosis of leprosy. In: Scollard DM, Gills TP, editors. International textbook of leprosy. www.internationaltextbookofleprosy.org2017.
6. Barnes PF, Lu S, Abrams JS, Wang E, Yamamura M, Modlin RL. Cytokine production at the site of disease in human tuberculosis. Infect Immun. 1993;61(8):3482-9.
7. Saini C, Ramesh V, Nath I. CD4+ Th17 cells discriminate clinical types and constitute a third subset of non Th1, Non Th2 T cells in human leprosy. PLoS neglected tropical diseases. 2013;7(7):e2338.
8. Modlin RL, Melancon-Kaplan J, Young SM, Pirmez C, Kino H, Convit J, et al. Learning from lesions: patterns of tissue inflammation in leprosy. Proc Natl Acad Sci U S A. 1988;85(4):1213-7.
9. Scollard DM. Pathogenesis and Pathology of Leprosy. In: Scollard DM, Gillis TP, editors. International textbook of leprosy. www.internationaltextbookofleprosy.org2016.
10. Misch E, Berrington W, Vary J, Hawn T. Leprosy and the Human Genome. Microbiology and molecular biology reviews : MMBR. 2010;74:589-620.
11. Geluk A. Challenges in immunodiagnostic tests for leprosy. Expert Opin Med Diagn. 2013;7(3):265-74.
12. Geluk A. Biomarkers for leprosy: would you prefer T (cells)? Leprosy review. 2013;84(1): 3-12.
13. Assis BPN, Lyon S, Grossi MAF, Rocha M. Risk factors for physical disability upon release from multidrug therapy in new cases of leprosy at a referral center in Brazil. Rev Inst Med Trop Sao Paulo. 2019;61:e13.
14. Croft RP, Nicholls PG, Steyerberg EW, Richardus JH, Withington SG, Smith WC. A clinical prediction rule for nerve function impairment in leprosy patients-revisited after 5 years of follow-up. Leprosy review. 2003;74(1):35-41.
15. Lockwood DN, Saunderson P. Nerve damage in Leprosy: a continuing challenge for scientists, clinicians and service providers. Int Health. 2012;4(2012):77-85.
16. Scollard DM, Martelli CM, Stefani MM, Maroja MF, Villahermosa L, Pardillo F, et al. Risk factors for leprosy reactions in three endemic countries. Am J Trop Med Hyg. 2015;92(1):108-14.
17. Antunes DE, Ferreira GP, Nicchio MV, Araujo S, Cunha AC, Gomes RR, et al. Number of leprosy reactions during treatment: clinical correlations and laboratory diagnosis. Rev Soc Bras Med Trop. 2016;49(6):741-5.
18. Walker SL. Leprosy Reactions. In: Scollard DM, Gillis TP, editors. International textbook of leprosy. www.internationaltextbookofleprosy.org2020.
19. Ridley DS, Radia KB. The histological course of reactions in borderline leprosy and their outcome. International journal of leprosy and other mycobacterial diseases : official organ of the International Leprosy Association. 1981;49(4):383-92.
20. Geluk A. Correlates of immune exacerbations in leprosy. Semin Immunol. 2018.
21. Hagge DA, Parajuli P, Kunwar CB, Rana D, Thapa R, Neupane KD, et al. Opening a Can of Worms: Leprosy Reactions and Complicit Soil-Transmitted Helminths. EBioMedicine. 2017;23: 119-24.
22. Khadge S, Banu S, Bobosha K, JJ vdP-vS, Goulart IM, Thapa P, et al. Longitudinal immune profiles in type 1 leprosy reactions in Bangladesh, Brazil, Ethiopia and Nepal. BMC Infect Dis. 2015;15(1):477.
23. Scollard DM, Chaduvula MV, Martinez A, Fowlkes N, Nath I, Stryjewska BM, et al. Increased CXC ligand 10 levels and gene expression in type 1 leprosy reactions. Clin Vaccine Immunol. 2011;18(6):947-53.
24. Corstjens PL, van HA, Tjon Kon Fat EM, van den Eeden SJ, Wilson L, Geluk A. Field-Friendly Test for Monitoring Multiple Immune Response Markers during Onset and Treatment of Exacerbated Immunity in Leprosy. Clin Vaccine Immunol. 2016;23(6):515-9.
25. Richardus RA, Butlin CR, Alam K, Kundu K, Geluk A, Richardus JH. Clinical manifestations of leprosy after BCG vaccination: an observational study in Bangladesh. Vaccine. 2015;33(13):1562-7.
26. Lee DJ, Li H, Ochoa MT, Tanaka M, Carbone RJ, Damoiseaux R, et al. Integrated pathways for neutrophil recruitment and inflammation in leprosy. J Infect Dis. 2010;201(4):558-69.
27. World Health Organization. Global leprosy update, 2018: moving towards a leprosy-free world. Weekly Epidemiological Record. 2019;94(35/36):389-412.
28. Richardus JH, Ignotti E, Smith WCS. Epidemiology of Leprosy. In: Scollard DM, Gillis TP, editors. International textbook of leprosy. www.internationaltextbookofleprosy.org2016.
29. Nunzi E, Massone C. Leprosy: A Practical Guide. Milan: Springer; 2012.
30. Quilter EEV, Butlin CR, Singh S, Alam K, Lockwood DNJ. Patients with skin smear positive

leprosy in Bangladesh are the main risk factor for leprosy development: 21-year follow-up in the household contact study (COCOA). PLoS neglected tropical diseases. 2020;14(10):e0008687.
31. Araujo S, Freitas LO, Goulart LR, Goulart IM. Molecular evidence for the aerial route of infection of *Mycobacterium leprae* and the role of asymptomatic carriers in the persistence of leprosy. Clin Infect Dis. 2016;63(11):1412-20.
32. Bratschi MW, Steinmann P, Wickenden A, Gillis TP. Current knowledge on *Mycobacterium leprae* transmission: a systematic literature review. Leprosy review. 2015;86(2):142-55.
33. Zhang FR, Huang W, Chen SM, Sun LD, Liu H, Li Y, et al. Genomewide association study of leprosy. N Engl J Med. 2009;361(27):2609-18.
34. Mira MT, Alcais A, Nguyen VT, Moraes MO, Di Flumeri C, Vu HT, et al. Susceptibility to leprosy is associated with PARK2 and PACRG. Nature. 2004;427(6975):636-40.
35. Wang D, Xu L, Lv L, Su LY, Fan Y, Zhang DF, et al. Association of the LRRK2 genetic polymorphisms with leprosy in Han Chinese from Southwest China. Genes Immun. 2015;16(2):112-9.
36. Sales-Marques C, Cardoso CC, Alvarado-Arnez LE, Illaramendi X, Sales AM, Hacker MA, et al. Genetic polymorphisms of the IL6 and NOD2 genes are risk factors for inflammatory reactions in leprosy. PLoS neglected tropical diseases. 2017;11(7):e0005754.
37. Liu H, Irwanto A, Fu X, Yu G, Yu Y, Sun Y, et al. Discovery of six new susceptibility loci and analysis of pleiotropic effects in leprosy. Nature genetics. 2015;47(3):267-71.
38. Moet FJ, Meima A, Oskam L, Richardus JH. Risk factors for the development of clinical leprosy among contacts, and their relevance for targeted interventions. Leprosy review. 2004;75(4): 310-26.
39. Dwivedi VP, Banerjee A, Das I, Saha A, Dutta M, Bhardwaj B, et al. Diet and nutrition: An important risk factor in leprosy. Microbial pathogenesis. 2019;137:103714.
40. Feenstra SG, Nahar Q, Pahan D, Oskam L, Richardus JH. Recent food shortage is associated with leprosy disease in Bangladesh: a case-control study. PLoS neglected tropical diseases. 2011;5(5):e1029.
41. Serrano-Coll H, Mora HR, Beltrán JC, Duthie MS, Cardona-Castro N. Social and environmental conditions related to Mycobacterium leprae infection in children and adolescents from three leprosy endemic regions of Colombia. BMC infectious diseases. 2019;19(1):520.
42. Kerr-Pontes LRS, Montenegro ACD, Barreto ML, Werneck GL, Feldmeier H. Inequality and leprosy in Northeast Brazil: an ecological study. International Journal of Epidemiology. 2004;33(2):262-9.
43. Vieira MCA, Nery JS, Paixão ES, Freitas de Andrade KV, Oliveira Penna G, Teixeira MG. Leprosy in children under 15 years of age in Brazil: A systematic review of the literature. PLoS neglected tropical diseases. 2018;12(10):e0006788.
44. Sarkar R, Pradhan S. Leprosy and women. International Journal of Women's Dermatology. 2016;2(4):117-21.
45. Richardus R, Alam K, Kundu K, Chandra Roy J, Zafar T, Chowdhury AS, et al. Effectiveness of single-dose rifampicin after BCG vaccination to prevent leprosy in close contacts of patients with newly diagnosed leprosy: A cluster randomized controlled trial. International journal of infectious diseases : IJID : official publication of the International Society for Infectious Diseases. 2019;88:65-72.
46. Avanzi C, Del-Pozo J, Benjak A, Stevenson K, Simpson VR, Busso P, et al. Red squirrels in the British Isles are infected with leprosy bacilli. Science. 2016;354(6313):744-7.
47. Turankar RP, Lavania M, Singh M, Sengupta U, Siva Sai K, Jadhav RS. Presence of viable Mycobacterium leprae in environmental specimens around houses of leprosy patients. Indian J Med Microbiol. 2016;34(3):315-21.
48. O'Brien C, Malik R. History and mysteries of leprosy. Journal of feline medicine and surgery. 2017;19(5):496-7.
49. Schilling A-K, van Hooij A, Corstjens P, Lurz P, DelPozo J, Stevenson K, et al. Detection of humoral immunity to mycobacteria causing leprosy in Eurasian red squirrels (*Sciurus vulgaris*) using a quantitative rapid test2019.
50. Meredith A, Pozo JD, Smith S, Milne E, Stevenson K, McLuckie J. Leprosy in red squirrels in Scotland. Veterinary Record. 2014;175(11):285-6.
51. Simpson V, Hargreaves J, Butler H, Blackett T, Stevenson K, McLuckie J. Leprosy in red squirrels on the Isle of Wight and Brownsea Island. Veterinary Record. 2015;177(8):206-7.
52. Butler H, Stevenson K, McLuckie J, Simpson V. Further evidence of leprosy in Isle of Wight red squirrels. Vet Rec. 2017;180(16):407.
53. Schilling A-K, Avanzi C, Ulrich RG, Busso P, Pisanu B, Ferrari N, et al. British Red Squirrels Remain the Only Known Wild Rodent Host for Leprosy Bacilli. Frontiers in Veterinary Science. 2019;6(8).
54. Walsh GP, Storrs EE, Meyers W, Binford CH. Naturally acquired leprosy-like disease in the nine-banded armadillo (*Dasypus novemcinctus*): recent epizootiologic findings. J Reticuloendothel Soc. 1977;22(4):363-7.
55. Smith JH, File SK, Nagy BA, Folse DS, Buckner JA, Webb LJ, et al. Leprosy-like disease of wild

armadillos in French Acadiana, Louisiana. J Reticuloendothel Soc. 1978;24(6):705-19.
56. Stallknecht DE, Truman RW, Hugh-Jones ME, Job CK. Surveillance for naturally acquired leprosy in a nine-banded armadillo population. J Wildl Dis. 1987;23(2):308-10.
57. Loughry WJ, Truman RW, McDonough CM, Tilak MK, Garnier S, Delsuc F. Is leprosy spreading among nine-banded armadillos in the southeastern United States? J Wildl Dis. 2009;45(1): 144-52.
58. Truman RW, Singh P, Sharma R, Busso P, Rougemont J, Paniz-Mondolfi A, et al. Probable zoonotic leprosy in the Southern United States. The New England journal of medicine. 2011;364(17):1626-33.
59. Sharma R, Singh P, Loughry WJ, Lockhart JM, Inman WB, Duthie MS, et al. Zoonotic leprosy in the Southeastern United States. Emerg Infect Dis. 2015;21(12):2127-34.
60. da Silva MB, Portela JM, Li W, Jackson M, Gonzalez-Juarrero M, Hidalgo AS, et al. Evidence of zoonotic leprosy in Pará, Brazilian Amazon, and risks associated with human contact or consumption of armadillos. PLoS neglected tropical diseases. 2018;12(6):e0006532.
61. Amezcua ME, Escobar-Gutierrez A, Storrs EE, Dhople AM, Burchfield HP. Wild Mexican armadillo with leprosy-like infection. International journal of leprosy and other mycobacterial diseases : official organ of the International Leprosy Association. 1984;52(2):254-5.
62. da Silva Ferreira J, de Carvalho FM, Vidal Pessolani MC, de Paula Antunes JMA, de Medeiros Oliveira IVP, Ferreira Moura GH, et al. Serological and molecular detection of infection with *Mycobacterium leprae* in Brazilian six banded armadillos (*Euphractus sexcinctus*). Comp Immunol Microbiol Infect Dis. 2020;68:101397-.
63. Pedrini SC, Rosa PS, Medri IM, Mourão G, Bagagli E, Lopes CA. Search for *Mycobacterium leprae* in wild mammals. Braz J Infect Dis. 2010;14(1):47-53.
64. Schmitt JV, Dechandt IT, Dopke G, Ribas ML, Cerci FB, Viesi JMZ, et al. Armadillo meat intake was not associated with leprosy in a case control study, Curitiba (Brazil). Memórias do Instituto Oswaldo Cruz. 2010;105:857-62.
65. Stefani MMA, Rosa PS, Costa MB, Schetinni APM, Manhães I, Pontes MAA, et al. Leprosy survey among rural communities and wild armadillos from Amazonas state, Northern Brazil. PloS one. 2019;14(1):e0209491.
66. Honap TP, Pfister LA, Housman G, Mills S, Tarara RP, Suzuki K, et al. *Mycobacterium leprae* genomes from naturally infected nonhuman primates. PLoS neglected tropical diseases. 2018;12(1):e0006190.
67. Hockings KJ, Mubemba B, Avanzi C, Pleh K, Düx A, Bersacola E, et al. Leprosy in wild chimpanzees. bioRxiv. 2020:2020.11.10.374371.
68. Valverde CR, Canfield D, Tarara R, Esteves MI, Gormus BJ. Spontaneous leprosy in a wild-caught cynomolgus macaque. International journal of leprosy and other mycobacterial diseases : official organ of the International Leprosy Association. 1998;66(2):140-8.
69. Lavania M, Katoch K, Sachan P, Dubey A, Kapoor S, Kashyap M, et al. Detection of *Mycobacterium leprae* DNA from soil samples by PCR targeting RLEP sequences. J Commun Dis. 2006;38(3):269-73.
70. Turankar RP, Lavania M, Chaitanya VS, Sengupta U, Darlong J, Darlong F, et al. Single nucleotide polymorphism-based molecular typing of *M. leprae* from multicase families of leprosy patients and their surroundings to understand the transmission of leprosy. Clin Microbiol Infect. 2014;20(3):O142-9.
71. Turankar RP, Lavania M, Singh M, Siva Sai KS, Jadhav RS. Dynamics of *Mycobacterium leprae* transmission in environmental context: deciphering the role of environment as a potential reservoir. Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases. 2012;12(1):121-6.
72. Turankar RP, Pandey S, Lavania M, Singh I, Nigam A, Darlong J, et al. Comparative evaluation of PCR amplification of RLEP, 16S rRNA, rpoT and Sod A gene targets for detection of *M. leprae* DNA from clinical and environmental samples. Int J Mycobacteriol. 2015;4(1):54-9.
73. Lavania M, Katoch K, Katoch VM, Gupta AK, Chauhan DS, Sharma R, et al. Detection of viable *Mycobacterium leprae* in soil samples: insights into possible sources of transmission of leprosy. Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases. 2008;8(5):627-31.
74. Mohanty PS, Naaz F, Katara D, Misba L, Kumar D, Dwivedi DK, et al. Viability of *Mycobacterium leprae* in the environment and its role in leprosy dissemination. Indian J Dermatol Venereol Leprol. 2016;82(1):23-7.
75. Holanda MV, Marques LEC, Macedo MLB, Pontes MAA, Sabadia JAB, Kerr L, et al. Presence of *Mycobacterium leprae* genotype 4 in environmental waters in Northeast Brazil. Rev Soc Bras Med Trop. 2017;50(2):216-22.
76. Arraes M, Holanda MV, Lima L, Sabadia JAB, Duarte CR, Almeida RLF, et al. Natural environmental water sources in endemic regions of northeastern Brazil are potential reservoirs of viable

Mycobacterium leprae. Mem Inst Oswaldo Cruz. 2017;112(12):805-11.
77. Wheat WH, Casali AL, Thomas V, Spencer JS, Lahiri R, Williams DL, et al. Long-term survival and virulence of *Mycobacterium leprae* in amoebal cysts. PLoS neglected tropical diseases. 2014;8(12):e3405.
78. Cole ST, Eiglmeier K, Parkhill J, James KD, Thomson NR, Wheeler PR, et al. Massive gene decay in the leprosy bacillus. Nature. 2001;409(6823):1007-11.
79. Donoghue HD, Holton J, Spigelman M. PCR primers that can detect low levels of *Mycobacterium leprae* DNA. Journal of medical microbiology. 2001;50(2):177-82.
80. Truman RW, Andrews PK, Robbins NY, Adams LB, Krahenbuhl JL, Gillis TP. Enumeration of *Mycobacterium leprae* Using Real-Time PCR. PLoS neglected tropical diseases. 2008;2(11):e328.
81. Martinez AN, Ribeiro-Alves M, Sarno EN, Moraes MO. Evaluation of qPCR-based assays for leprosy diagnosis directly in clinical specimens. PLoS neglected tropical diseases. 2011;5(10):e1354.
82. Braet S, Vandelannoote K, Meehan CJ, Brum Fontes AN, Hasker E, Rosa PS, et al. The repetitive element RLEP is a highly specific target for detection of *Mycobacterium leprae*. Journal of clinical microbiology. 2018;56(3).
83. Monot M, Honore N, Garnier T, Araoz R, Coppee JY, Lacroix C, et al. On the origin of leprosy. Science. 2005;308(5724):1040-2.
84. Monot M, Honore N, Garnier T, Zidane N, Sherafi D, Paniz-Mondolfi A, et al. Comparative genomic and phylogeographic analysis of *Mycobacterium leprae*. Nature genetics. 2009;41(12): 1282-9.
85. Benjak A, Avanzi C, Singh P, Loiseau C, Girma S, Busso P, et al. Phylogenomics and antimicrobial resistance of the leprosy bacillus *Mycobacterium leprae*. Nat Commun. 2018;9(1):352.
86. Schuenemann VJ, Singh P, Mendum TA, Krause-Kyora B, Jager G, Bos KI, et al. Genome-wide comparison of medieval and modern *Mycobacterium leprae*. Science. 2013;341(6142):179-83.
87. Schuenemann VJ, Avanzi C, Krause-Kyora B, Seitz A, Herbig A, Inskip S, et al. Ancient genomes reveal a high diversity of *Mycobacterium leprae* in medieval Europe. PLoS pathogens. 2018;14(5):e1006997.
88. Mendum TA, Schuenemann VJ, Roffey S, Taylor GM, Wu H, Singh P, et al. *Mycobacterium leprae* genomes from a British medieval leprosy hospital: towards understanding an ancient epidemic. BMC genomics. 2014;15:270.
89. Suzuki K, Takigawa W, Tanigawa K, Nakamura K, Ishido Y, Kawashima A, et al. Detection of *Mycobacterium leprae* DNA from archaeological skeletal remains in Japan using whole genome amplification and polymerase chain reaction. PloS one. 2010;5(8):e12422.
90. Krause-Kyora B, Nutsua M, Boehme L, Pierini F, Pedersen DD, Kornell S-C, et al. Ancient DNA study reveals HLA susceptibility locus for leprosy in medieval Europeans. Nat Commun. 2018;9(1):1569-.
91. Donoghue HD. Tuberculosis and leprosy associated with historical human population movements in Europe and beyond - an overview based on mycobacterial ancient DNA. Annals of human biology. 2019;46(2):120-8.
92. Donoghue HD, Taylor GM, Mendum TA, Stewart GR, Rigouts L, Lee OY-C, et al. The Distribution and Origins of Ancient Leprosy. In: Ribòn W, editor. Hansen's Disease - The Forgotten and Neglected Disease. London: IntechOpen; 2018.
93. Singh P, Benjak A, Schuenemann VJ, Herbig A, Avanzi C, Busso P, et al. Insight into the evolution and origin of leprosy bacilli from the genome sequence of *Mycobacterium lepromatosis*. Proc Natl Acad Sci U S A. 2015;112(14):4459-64.
94. Han XY, Silva FJ. On the age of leprosy. PLoS neglected tropical diseases. 2014;8(2):e2544-e.
95. Fontes ANB, Lara FA, Santos AR, Suffys P. Pathogen Detection. In: Scollard DM, Gillis TP, editors. International textbook of leprosy. www.internationaltextbookofleprosy.org2017.
96. Martinez AN, Lahiri R, Pittman TL, Scollard D, Truman R, Moraes MO, et al. Molecular determination of *Mycobacterium leprae* viability by use of real-time PCR. Journal of clinical microbiology. 2009;47(7):2124-30.
97. Barbieri RR, Manta FSN, Moreira SJM, Sales AM, Nery JAC, Nascimento LPR, et al. Quantitative polymerase chain reaction in paucibacillary leprosy diagnosis: A follow-up study. PLoS neglected tropical diseases. 2019;13(3):e0007147.
98. Gama RS, Gomides TAR, Gama CFM, Moreira SJM, de Neves Manta FS, de Oliveira LBP, et al. High frequency of M. leprae DNA detection in asymptomatic household contacts. BMC infectious diseases. 2018;18(1):153.
99. Santos DFD, Mendonca MR, Antunes DE, Sabino EFP, Pereira RC, Goulart LR, et al. Molecular, immunological and neurophysiological evaluations for early diagnosis of neural impairment in seropositive leprosy household contacts. PLoS neglected tropical diseases. 2018;12(5):e0006494.
100. Brito e Cabral P, Junior JE, de Macedo AC, Alves AR, Goncalves TB, Brito e Cabral TC, et al. Anti-PGL1 salivary IgA/IgM, serum IgG/IgM, and nasal *Mycobacterium leprae* DNA in individuals with household contact with leprosy. International journal of infectious diseases : IJID : official publica-

tion of the International Society for Infectious Diseases. 2013;17(11):e1005-10.
101. Carvalho RS, Foschiani IM, Costa M, Marta SN, da Cunha Lopes Virmond M. Early detection of M. leprae by qPCR in untreated patients and their contacts: results for nasal swab and palate mucosa scraping. European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology. 2018;37(10):1863-7.
102. van Beers SM, Izumi S, Madjid B, Maeda Y, Day R, Klatser PR. An epidemiological study of leprosy infection by serology and polymerase chain reaction. International journal of leprosy and other mycobacterial diseases : official organ of the International Leprosy Association. 1994;62(1):1-9.
103. Romero-Montoya M, Beltran-Alzate JC, Cardona-Castro N. Evaluation and Monitoring of Mycobacterium leprae Transmission in Household Contacts of Patients with Hansen's Disease in Colombia. PLoS neglected tropical diseases. 2017;11(1):e0005325.
104. Cardona-Castro N, Beltran-Alzate JC, Manrique-Hernandez R. Survey to identify *Mycobacterium leprae*-infected household contacts of patients from prevalent regions of leprosy in Colombia. Mem Inst Oswaldo Cruz. 2008;103(4):332-6.
105. Manta FSN, Barbieri RR, Moreira SJM, Santos PTS, Nery JAC, Duppre NC, et al. Quantitative PCR for leprosy diagnosis and monitoring in household contacts: A follow-up study, 2011-2018. Scientific reports. 2019;9(1):16675.
106. van Hooij A, van den Eeden S, Richardus R, Tjon Kon Fat E, Wilson L, Franken K, et al. Application of new host biomarker profiles in quantitative point-of-care tests facilitates leprosy diagnosis in the field. EBioMedicine. 2019;47:301-8.
107. van Hooij A, Tjon Kon Fat EM, Batista da Silva M, Carvalho Bouth R, Cunha Messias AC, Gobbo AR, et al. Evaluation of Immunodiagnostic Tests for Leprosy in Brazil, China and Ethiopia. Scientific reports. 2018;8(1):17920.
108. van Hooij A, Tjon Kon Fat EM, Richardus R, van den Eeden SJ, Wilson L, de Dood CJ, et al. Quantitative lateral flow strip assays as User-Friendly Tools To Detect Biomarker Profiles For Leprosy. Scientific reports. 2016;6:34260.
109. van Hooij A, Tjon Kon Fat EM, van den Eeden SJF, Wilson L, Batista da Silva M, Salgado CG, et al. Field-friendly serological tests for determination of *M. leprae*-specific antibodies. Scientific reports. 2017;7(1):8868.
110. Chen X, You YG, Yuan YH, Yuan LC, Wen Y. Host immune responses induced by specific *Mycobacterium leprae* antigens in an overnight whole-blood assay correlate with the diagnosis of paucibacillary leprosy patients in China. PLoS neglected tropical diseases. 2019;13(4):e0007318.
111. Queiroz EA, Medeiros NI, Mattos RT, Carvalho APM, Rodrigues-Alves ML, Dutra WO, et al. Immunological biomarkers of subclinical infection in household contacts of leprosy patients. Immunobiology. 2019;224(4):518-25.
112. Penna ML, Penna GO, Iglesias PC, Natal S, Rodrigues LC. Anti-PGL-1 Positivity as a Risk Marker for the Development of Leprosy among Contacts of Leprosy Cases: Systematic Review and Meta-analysis. PLoS neglected tropical diseases. 2016;10(5):e0004703.
113. Spencer JS, Brennan PJ. The role of *Mycobacterium leprae* phenolic glycolipid I (PGL-I) in serodiagnosis and in the pathogenesis of leprosy. Leprosy review. 2011;82(4):344-57.
114. Zak DE, Penn-Nicholson A, Scriba TJ, Thompson E, Suliman S, Amon LM, et al. A blood RNA signature for tuberculosis disease risk: a prospective cohort study. Lancet. 2016;387(10035):2312-22.
115. Suliman S, Thompson E, Sutherland J, Weiner Rd J, Ota MOC, Shankar S, et al. Four-gene Pan-African Blood Signature Predicts Progression to Tuberculosis. American journal of respiratory and critical care medicine. 2018.
116. Jorge K, Souza RP, Assis MTA, Araujo MG, Locati M, Jesus AMR, et al. Characterization of MicroRNA Expression Profiles and Identification of Potential Biomarkers in Leprosy. Journal of clinical microbiology. 2017;55(5):1516-25.
117. Montoya DJ, Andrade P, Silva BJA, Teles RMB, Ma F, Bryson B, et al. Dual RNA-Seq of Human Leprosy Lesions Identifies Bacterial Determinants Linked to Host Immune Response. Cell reports. 2019;26(13):3574-85.e3.
118. Andrade PR, Mehta M, Lu J, Teles RMB, Montoya D, Scumpia PO, et al. The cell fate regulator NUPR1 is induced by *Mycobacterium leprae* via type I interferon in human leprosy. PLoS neglected tropical diseases. 2019;13(7):e0007589.
119. Inkeles MS, Teles RM, Pouldar D, Andrade PR, Madigan CA, Lopez D, et al. Cell-type deconvolution with immune pathways identifies gene networks of host defense and immunopathology in leprosy. JCI insight. 2016;1(15):e88843.
120. Realegeno S, Kelly-Scumpia KM, Dang AT, Lu J, Teles R, Liu PT, et al. S100A12 Is Part of the Antimicrobial Network against *Mycobacterium leprae* in Human Macrophages. PLoS pathogens. 2016;12(6):e1005705.
121. Orlova M, Cobat A, Huong NT, Ba NN, Van TN, Spencer J, et al. Gene set signature of reversal reaction type I in leprosy patients. PLoS Genet. 2013;9(7):e1003624.
122. Guerreiro LT, Robottom-Ferreira AB, Ribeiro-Alves M, Toledo-Pinto TG, Rosa BT, Rosa PS, et

al. Gene expression profiling specifies chemokine, mitochondrial and lipid metabolism signatures in leprosy. PloS one. 2013;8(6):e64748.
123. Bleharski JR, Li H, Meinken C, Graeber TG, Ochoa MT, Yamamura M, et al. Use of genetic profiling in leprosy to discriminate clinical forms of the disease. Science. 2003;301(5639):1527-30.
124. Salgado CG, Pinto P, Bouth RC, Gobbo AR, Messias ACC, Sandoval TV, et al. miRNome Expression Analysis Reveals New Players on Leprosy Immune Physiopathology. Frontiers in Immunology. 2018;9(463).
125. Guerreiro LT, Robottom-Ferreira AB, Ribeiro-Alves M, Toledo-Pinto TG, Rosa Brito T, Rosa PS, et al. Gene expression profiling specifies chemokine, mitochondrial and lipid metabolism signatures in leprosy. PloS one. 2013;8(6):e64748.
126. Geluk A, van Meijgaarden KE, Wilson L, Bobosha K, JJ vdP-vS, van den Eeden SJ, et al. Longitudinal immune responses and gene expression profiles in type 1 leprosy reactions. J Clin Immunol. 2014;34(2):245-55.
127. Rego JL, de Lima Santana N, Machado PRL, Ribeiro-Alves M, de Toledo-Pinto TG, Castellucci LC, et al. Whole blood profiling of leprosy type 1(reversal) reactions highlights prominence of innate immune response genes. BMC infectious diseases. 2018;18(1):422.
128. Mi Z, Liu H, Zhang F. Advances in the Immunology and Genetics of Leprosy. Front Immunol. 2020;11:567.
129. Fava VM, Dallmann-Sauer M, Schurr E. Genetics of leprosy: today and beyond. Hum Genet. 2019.
130. Zhang F, Liu H, Chen S, Low H, Sun L, Cui Y, et al. Identification of two new loci at IL23R and RAB32 that influence susceptibility to leprosy. Nature genetics. 2011;43(12):1247-51.
131. Wang Z, Sun Y, Fu Xa, Yu G, Wang C, Bao F, et al. A large-scale genome-wide association and meta-analysis identified four novel susceptibility loci for leprosy. Nat Commun. 2016;7(1):13760.
132. Wang Z, Mi Z, Wang H, Sun L, Yu G, Fu X, et al. Discovery of 4 exonic and 1 intergenic novel susceptibility loci for leprosy. Clinical Genetics. 2018;94(2):259-63.
133. Liu H, Wang Z, Li Y, Yu G, Fu Xa, Wang C, et al. Genome-Wide Analysis of Protein-Coding Variants in Leprosy. Journal of Investigative Dermatology. 2017;137(12):2544-51.
134. Andiappan AK, Melchiotti R, Poh TY, Nah M, Puan KJ, Vigano E, et al. Genome-wide analysis of the genetic regulation of gene expression in human neutrophils. Nat Commun. 2015;6(1):7971.
135. Bochud P-Y, Hawn TR, Siddiqui MR, Saunderson P, Britton S, Abraham I, et al. Toll-like receptor 2 (TLR2) polymorphisms are associated with reversal reaction in leprosy. J Infect Dis. 2008;197(2):253-61.
136. de Messias-Reason IJ, Boldt ABW, Braga ACM, Stahlke EVRS, Dornelles L, Pereira-Ferrari L, et al. The Association between Mannan-Binding Lectin Gene Polymorphism and Clinical Leprosy: New Insight into an Old Paradigm. J Infect Dis. 2007;196(9):1379-85.
137. Santos AR, Suffys PN, Vanderborght PR, Moraes MO, Vieira LMM, Cabello PH, et al. Role of Tumor Necrosis Factor–α and Interleukin-10 Promoter Gene Polymorphisms in Leprosy. J Infect Dis. 2002;186(11):1687-91.
138. Camargo RMd, Silva WLd, Medeiros P, Belone AdFF, Latini ACP. Polymorphisms in the TGFB1 and IL2RA genes are associated with clinical forms of leprosy in Brazilian population. Memórias do Instituto Oswaldo Cruz. 2018;113.
139. Alter A, de Léséleuc L, Van Thuc N, Thai VH, Huong NT, Ba NN, et al. Genetic and functional analysis of common MRC1 exon 7 polymorphisms in leprosy susceptibility. Human Genetics. 2010;127(3):337-48.
140. Grant AV, Cobat A, Van Thuc N, Orlova M, Huong NT, Gaschignard J, et al. CUBN and NEBL common variants in the chromosome 10p13 linkage region are associated with multibacillary leprosy in Vietnam. Human Genetics. 2014;133(7):883-93.
141. de Messias IJ, Santamaria J, Brenden M, Reis A, Mauff G. Association of C4B deficiency (C4B*Q0) with erythema nodosum in leprosy. Clin Exp Immunol. 1993;92(2):284-7.
142. Fava VM, Manry J, Cobat A, Orlova M, Van Thuc N, Ba NN, et al. A Missense LRRK2 Variant Is a Risk Factor for Excessive Inflammatory Responses in Leprosy. PLoS neglected tropical diseases. 2016;10(2):e0004412.
143. Fava VM, Xu YZ, Lettre G, Van Thuc N, Orlova M, Thai VH, et al. Pleiotropic effects for Parkin and LRRK2 in leprosy type-1 reactions and Parkinson's disease. Proceedings of the National Academy of Sciences. 2019;116(31):15616-24.
144. Fava VM, Manry J, Cobat A, Orlova M, Van Thuc N, Moraes MO, et al. A genome wide association study identifies a lncRna as risk factor for pathological inflammatory responses in leprosy. PLOS Genetics. 2017;13(2):e1006637.
145. Fava VM, Sales-Marques C, Alcais A, Moraes MO, Schurr E. Age-Dependent Association of TNFSF15/TNFSF8 Variants and Leprosy Type 1 Reaction. Front Immunol. 2017;8:155.
146. Fava VM, Cobat A, Van Thuc N, Latini ACP, Stefani MMA, Belone AF, et al. Association of TNFSF8 Regulatory Variants With Excessive Inflammatory Responses but not Leprosy Per Se. J Infect

Dis. 2014;211(6):968-77.
147. Misch EA, Macdonald M, Ranjit C, Sapkota BR, Wells RD, Siddiqui MR, et al. Human TLR1 Deficiency Is Associated with Impaired Mycobacterial Signaling and Protection from Leprosy Reversal Reaction. PLoS neglected tropical diseases. 2008;2(5):e231.
148. Berrington WR, Macdonald M, Khadge S, Sapkota BR, Janer M, Hagge DA, et al. Common Polymorphisms in the NOD2 Gene Region Are Associated with Leprosy and Its Reactive States. J Infect Dis. 2010;201(9):1422-35.

2

Genomic characterization of *Mycobacterium leprae* to explore transmission patterns identifies new subtype in Bangladesh

Maria Tió-Coma, Charlotte Avanzi, Els M. Verhard, Louise Pierneef, Anouk van Hooij, Andrej Benjak, Johan Chandra Roy, Marufa Khatun, Khorshed Alam, Paul Corstjens, Stewart T. Cole, Jan Hendrik Richardus and Annemieke Geluk

Front Microbiol. 2020 Jun 16;11:1220

Abstract

Mycobacterium leprae, the causative agent of leprosy, is an unculturable bacterium with a considerably reduced genome (3.27 Mb) compared to homologues mycobacteria from the same ancestry. In 2001, the genome of *M. leprae* was first described and subsequently four genotypes (1-4) and 16 subtypes (A-P) were identified providing means to study global transmission patterns for leprosy**.**

In order to understand the role of asymptomatic carriers we investigated *M. leprae* carriage as well as infection in leprosy patients (n=60) and healthy household contacts (HHC; n=250) from Bangladesh using molecular detection of the bacterial element RLEP in nasal swabs (NS) and slit skin smears (SSS). In parallel, to study *M. leprae* genotype distribution in Bangladesh we explored strain diversity by whole-genome sequencing (WGS) and Sanger sequencing.

In the studied cohort in Bangladesh, *M. leprae* DNA was detected in 33.3% of NS and 22.2% of SSS of patients with bacillary index of 0 whilst in HHC 18.0% of NS and 12.3% of SSS were positive.

The majority of the *M. leprae* strains detected in this study belonged to genotype 1D (55%), followed by 1A (31%). Importantly, WGS allowed the identification of a new *M. leprae* genotype, designated 1B-Bangladesh (14%), which clustered separately between the 1A and 1B strains. Moreover, we established that the genotype previously designated 1C, is not an independent subtype but clusters within the 1D genotype.

Intraindividual differences were present between the *M. leprae* strains obtained including mutations in hypermutated genes, suggesting mixed colonization/infection or in-host evolution.

In summary, we observed that *M. leprae* is present in asymptomatic contacts of leprosy patients fueling the concept that these individuals contribute to the current intensity of transmission. Our data therefore emphasize the importance of sensitive and specific tools allowing post-exposure prophylaxis targeted at *M. leprae*-infected or -colonized individuals.

Introduction

Mycobacterium leprae and the more recently discovered *Mycobacterium lepromatosis* (1) are the causative agents of leprosy in humans as well as animals (2-7). Leprosy is a complex infectious disease often resulting in severe, life-long disabilities and still poses a serious health threat in low- and middle income countries (8). Despite the very limited *M. leprae* genome variability (9), the disease presents with characteristically different clinico-pathological forms (10) due to genetically dependent differences in the immune response to the pathogen, resulting in the WHO classification from paucibacillary (PB) to multibacillary (MB) leprosy (11). Notwithstanding the efficacy of multidrug therapy (MDT), approximately 210,000 new cases are still annually diagnosed and this incidence rate has been stable over the last decade (8). Aerosol transmission via respiratory routes is generally assumed to be the most probable way of bacterial dissemination (12, 13). Besides bacterial exposure other risk factors have been shown to be associated with development of leprosy such as genetic polymorphisms (14-17), the clinical type of the leprosy index case within a household, immunosuppression (18), and nutritional factors (19).

M. leprae is closely related to *Mycobacterium tuberculosis*, however, its genome has undergone a reductive evolution resulting in a genome of only 3.27 Mb compared to the 4.41 Mb of *M. tuberculosis'* (20). Part of the genes lost in *M. leprae* included vital metabolic activity, causing it to be an obligate intracellular pathogen which cannot be cultured in axenic media that requires support of a host to survive. This poses major limitations to obtain sufficient bacterial DNA for research purposes including whole genome sequencing (WGS). Nevertheless, in 2001 the genome of *M. leprae* was first published (20) leading to the classification of *M. leprae* into four main genotypes (1-4) (21) and subsequently further allocation into 16 subtypes (A-P) (3, 22). The genome of *M. leprae* contains several repetitive elements such as RLEP which present 37 copies and has been widely applied in molecular diagnostics to specifically detect the presence of this mycobacterium (23-26).

Single-nucleotide polymorphism (SNP) genotyping and WGS are powerful approaches to investigate pathogen transmission as well as bacterial dissemination and evolution through genome characterization (21, 22, 27). The limited variation observed in the *M. leprae* genome permits the reconstruction of historic human migration patterns and the origin of *M. leprae* (28). Over the years, several studies have contributed to the detection and characterization of *M. leprae* genomes originating from patients all around the world (21, 22, 29) as well as from ancient skeletons (30-34), red squirrels (2, 7, 35), armadillos (3, 4), non-human primates (5) and soil (36-42). Moreover, skeleton remains have been successfully applied to retrospectively assess whether individuals who contributed to the care of leprosy patients such as the priest Petrus Donders, had developed leprosy (43). In the last few years, new tools were developed allowing direct sequencing of *M. leprae* from

various types of clinical isolates (2, 29, 31). However, these methods were never applied on challenging samples such as slit skin smears (SSS) and nasal swabs (NS) containing a low amount of bacterial DNA compared to skin lesions of patients.

Household contacts of leprosy patients are a high risk group for developing the disease (44), and might serve as asymptomatic carriers contributing to bacterial dissemination. PCR and quantitative PCR (qPCR) are reliable techniques to detect *M. leprae* DNA and have been proposed as tools for early diagnosis of leprosy, particularly among household contacts of newly diagnosed patients (45, 46). In Brazil, *M. leprae* DNA has been detected in 15.9% to 42.4% of healthy household contacts (HHC) in SSS, 9.7% to 35.2% in blood (45, 47) and 8.9 to 49.0% in NS (12, 48, 49). Other studies from India, Indonesia and Colombia reported 21% of *M. leprae* positivity in SSS of HHC (38), 7.8% (50) and 16.0% to 31.0% in NS (51, 52).

Detection of host markers, such as serum IgM levels of anti-*M. leprae* phenolic glycolipid I (PGL-I), represents an alternative approach to diagnose infected individuals (53-55). However, although detection of *M. leprae* DNA as well as antibodies against PGL-I indicate infection with *M. leprae,* this does not necessarily result in disease. Thus, these tests alone are not sufficient to identify the complete leprosy spectrum (56, 57).

Bangladesh is a leprosy endemic country reporting up to 3,729 new leprosy cases in 2018 (8). However, *M. leprae* whole genomes (n=4) from Bangladesh, have only been described in one study (22) in which genotypes 1A, 1C and 1D were identified. To gain more insight into *M. leprae* genome variation and transmission routes in endemic areas in Bangladesh as well as the potential role of asymptomatic carriers, we explored the diversity and transmission of *M. leprae* in four districts of the northwest of Bangladesh. We collected SSS and NS of 31 leprosy patients with a high bacterial load as well as 279 of their household contacts and characterized *M. leprae* DNA by WGS or Sanger sequencing. The resulting genotypes were correlated to the subjects' GIS location. Additionally, this is the first study to examine *M. leprae* DNA detection in comparison to anti-PGL-I IgM levels in plasma measured by up-converting phosphor lateral flow assays (UCP-LFAs).

Materials and methods

Study design and sample collection

Newly diagnosed leprosy patients (index case, n=31) with bacteriological index (BI) ≥ 2 and 3-15 household contacts of each index case (n=279) were recruited between July 2017 and May 2018 (Table S1, Supplementary Data 1) in four districts of Bangladesh (Nilphamari, Rangpur, Panchagar and Thakurgaon). Patients with five or fewer skin lesions and BI 0 were grouped as PB leprosy. Patients with more than five skin lesions were grouped as MB leprosy and BI was determined. The prevalence in the districts where this study

was performed was 0.9 per 10,000 and the new case detection rate 1.18 per 10,000 (Rural health program, the leprosy mission Bangladesh, yearly district activity report 2018).

For *M. leprae* detection and characterization, SSS from 2-3 sites of the earlobe and NS (tip wrapped with traditional fiber, CLASSIQSwabs, Copan, Brescia, Italy) were collected and stored in 1 ml 70% ethanol at -20 °C until further use. For immunological analysis, plasma was collected (53, 56, 58).

Subjects included in the study were followed up for surveillance of new case occurrence for ≥ 24 months after sample collection.

Ethics Statement

Subjects were recruited following the Helskinki Declaration (2008 revision). The National Research Ethics Committee approved the study (BMRC/NREC/2016-2019/214) and participants were informed about the study objectives, the samples and their right to refuse to take part or withdraw without consequences for their treatment. All subjects gave informed consent before enrollment and treatment was provided according to national guidelines.

DNA isolation from slit skin smears and nasal swabs

DNA was isolated using DNeasy Blood & Tissue Kit (Qiagen, Valencia, CA) as per manufacturer's instructions with minor modifications. Briefly, tubes containing 1 ml 70% ethanol and SSS were vortexed for 15 seconds. SSS were removed and tubes were centrifuged for 15 minutes at 14000 rpm. Supernatants were removed and buffer ATL (200 µl) and proteinase K (20 µl) added. NS were transferred to new microtubes and the microtubes containing the remaining ethanol were centrifuged at 14000 rpm for 15 minutes. Supernatants were removed and NS were inserted again in the tubes, prior addition of ATL buffer (400 µl) and proteinase K (20 µl). SSS and NS samples were incubated at 56 °C for 1 h at 1100 rpm. Next, AL buffer (200 µl) was added and incubated at 70 °C for 10 min at 1400 rpm. Column extraction was performed after absolute ethanol precipitation (200 µl) as per manufacturer's instructions. To avoid cross contamination tweezers were cleaned first with hydrogen peroxide and then with ethanol between samples.

RLEP PCR and qPCR

RLEP PCR (23) was performed as previously described (36). Briefly, the 129 bp RLEP sequence was amplified in 50 µl by addition of 10 µl 5x Gotaq® Flexi buffer (Promega, Madison, WI), 5 µl $MgCl_2$ (25 mM), 2 µl dNTP mix (5 mM), 0.25 µl Gotaq® G2 Flexi DNA Polymerase (5 u/µl), 5 µl (2 µM) forward and reverse primers (Table S2) and 5 µl template DNA, water (negative control) or *M. leprae* DNA (Br4923 or Thai-53 DNA, BEI Resources, Manassas, VA) as positive control. PCR mixes were subjected to 2 min at 95 °C followed by 40 cycles of 30 s at 95 °C, 30 s at 65°C and 30 s at 72 °C and a final extension of 10 min at 72 °C. PCR prod-

ucts (15µl) were used for electrophoresis in a 3.5% agarose gel at 130V. Amplified DNA was visualized by Midori Green Advance staining (Nippon Genetics Europe, Dueren, Germany) using iBright™ FL1000 Imaging System (Invitrogen, Carlsbad, CA).

Samples from index cases and a selection of contacts for sequencing were also evaluated by qPCR (59). The mix included 12.5 µl TaqMan Universal Master Mix II (Applied Biosystems, Foster City, CA), 0.5 µl (25 µM) forward and reverse primers (Table S2), 0.5 µl (10 µM) TaqMan probe (Table S2) and 5 µl template DNA were mixed in a final volume of 25 µl. DNA was amplified using the following profile: 2 min at 50ºC and 10 min at 95ºC followed by 40 cycles of 15 s at 95ºC and 1 min at 60ºC with a QuantStudio 6 Flex Real-Time PCR System (Applied Biosystems). Presence of *M. leprae* DNA was considered if a sample was positive for RLEP qPCR with a cycle threshold (Ct) lower than 37.5 or was positive for RLEP PCR at least in two out of three indecently performed PCRs to avoid false positives.

Library preparation and enrichment

A total of 60 DNA extracts were selected for WGS, including 30 from SSS and 30 from NS (Figure S1, Supplementary Data 1). At least one sample from each index case (MB leprosy patient) was selected as well as RLEP positive samples of HHC and MB or PB patients who were household contacts of the index case (selection based on Ct value and household overlap). For 12 subject both SSS and NS samples were selected for WGS. A maximum of 1µg of DNA in a final volume of 50µL was mechanically fragmented to 300 bp using the S220 Focused-ultrasonicator (Covaris) following the manufacturer's recommendations and cleaned-up using a 1.8x ratio of AMPure beads. Up to 1µg of fragmented DNA was used to prepare indexed libraries using the Kapa Hyperprep kit (Roche) and the Kapa dual-indexed adapter kit as previously described (29) followed by two rounds of amplification. All libraries were quantified using the Qubit fluorimeter (Thermo Fisher Scientific, Waltham, MA), and the fragment size distribution was assessed using a fragment analyzer.

Libraries were target enriched for the *M. leprae* genome using a custom MYbaits Whole Genome Enrichment kit (ArborBioscence) as previously described (5). Briefly, biotinylated RNA baits were prepared using DNA from *M. leprae* Br4923. A total of 1500 ng of each amplified library was used for enrichment. Each library was pooled prior to enrichment with another library with similar qPCR Ct value. Enrichment was conducted according to the MYbaits protocol with the hybridization being carried out at 65 °C for 24 hours. After elution, all pools were amplified using the Kapa amplification kit with universal P5 and P7 primers (Roche). All amplification reactions were cleaned up using the AMPure beads (1X ratio).

Illumina sequencing

Pools were multiplexed on one lane of a NextSeq instrument with a total amount of 20-30 million reads per pool. Some libraries were deep sequenced based on the mapping statis-

tics obtained in the first run.

Raw reads were processed and aligned to *M. leprae* TN reference genome (GenBank accession number AL450380.1) as previously described using an in-house pipeline (29). A minimum depth coverage of 5 was considered for further phylogenetic analysis.

Sequencing analysis

Genome comparison was based on analysis of SNPs (analyzed with VarScan v2.3.9(60)) and Indels (analyzed with Platypus v0.8.171(61)) as formerly reported (29). The newly sequenced *M. leprae* genomes were aligned with 232 genomes available in public databases (31, 62). Sites below 80 and above 20% alignment difference were also reported. A comparison to 259 *M. leprae* genomes (including 27 new genomes) allowed the identification of unique SNPs per index case. Each candidate SNP or Indel was checked manually on Integrative Genomics Viewer (63) .

Genotyping and antimicrobial resistance by Sanger sequencing

To further characterize the *M. leprae* strains for which the whole genome sequence was not obtained, specific primers were designed to perform Sanger sequencing based on unique SNPs (Table S3 and S4) of each index case strain. Additionally, Sanger sequencing was performed after amplifying several loci (Table S2) to subtype the genomes based on standard the *M. leprae* classification (3, 22) and to determine antimicrobial resistance to rifampicin (*rpoB*), dapsone (*folP1*) or ofloxacin (*gyrA*). Genotyping by Sanger sequencing was performed to all RLEP PCR positive samples (including samples obtained from leprosy patients and HHC) without a whole genome sequence with a depth coverage of ≥5. PCRs were performed with 5 µl of template DNA using the aforementioned PCR mixes. DNA was denatured for 2 minutes at 95ºC, followed by 45 cycles of 30 s at 95ºC, 30 s at 50-58 ºC and 30 s at 72 ºC and a final extension cycle of 10 min at 72ºC. PCR products were resolved by agarose gel electrophoresis as explained above. PCR products showing a band were purified prior to sequencing using the Wizard SV Gel and PCR Clean-Up System (Promega). Sequencing was performed on the ABI3730xl system (Applied Biosystems) using the BigDye Terminator Cycle Sequencing Kit (Thermo Fisher Scientific). Sequences were analyzed using Bioedit v7.0.5.3.

Anti-PGL-I UCP-LFA

UCP-LFAs were performed using the LUMC developed LFA based on luminescent up-converting reporter particles for quantitative detection of anti-*M. leprae* PGL-I IgM as previously described (53, 56, 58). Plasma samples (n=308, 2 samples excluded due to labeling mistake) were thawed and diluted (1:50) in assay buffer. Strips were placed in microtiter plate wells containing 50 µl diluted samples and target specific UCP conjugate (PGL-I, 400 ng). Immunochromatography continued for at least 30 min until dry. Scanning of the

LFA strips was performed by LFA strip readers adapted for measurement of the UCP label (UPCON; Labrox, Finland). Results are displayed as the Ratio (R) value between Test and Flow-Control signal based on relative fluorescence units (RFUs) measured at the respective lines. The threshold for positivity for the αPGL-I UCP-LFA was 0.10.

Results

M. leprae detection in patients and healthy household contacts

At diagnosis of the index cases and recruitment of contacts in this study out of 279 household contacts 250 presented no signs or symptoms of leprosy or other diseases (HHC), whereas 22 household contacts were diagnosed as PB and seven as MB patients (Table S1, Supplementary Data 1) and therefore were excluded from the HHC group.

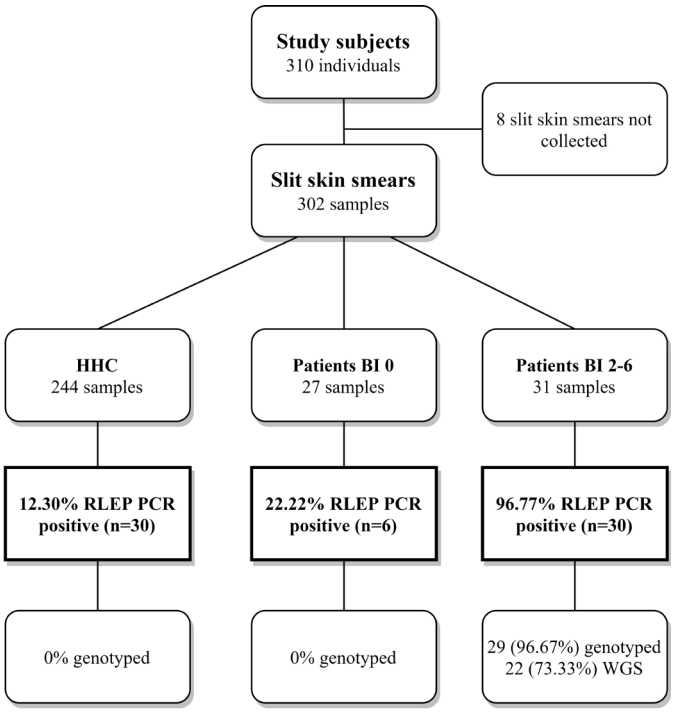
Presence of *M. leprae* DNA was determined by RLEP PCR or qPCR in SSS and NS of leprosy patients and HHC (Figure 1, Supplementary Data 1): as expected in MB patients with BI 2-6 *M. leprae* DNA was almost always detectable in both SSS (96.8%) and NS (90.9%). This was much lower in PB and MB patients with BI 0 ranging from 22.2% in SSS to 33.3% in NS. Positivity rates in HHC were not very different from those observed for PB and MB patients with BI 0, with 12.3% positive samples in SSS and 18.0% in NS. Showing a similar *M. leprae* carriage between HHC and patients with BI 0. Moreover, the overall Ct range was lower for SSS [16.3-37.1] compared to NS [20.1-39.4] showing that SSS contained more *M. leprae* DNA and is a preferred sample for its detection (Supplementary Data 1).

HHC (n=250) were followed up clinically for ≥ 24 months after sample collection and four of them developed leprosy within the first year. RLEP PCR performed on DNA isolated before disease occurrence showed a positive result from SSS in one patient (5 months before diagnosis) and a positive result from NS in another (8 months before diagnosis). All of the new cases developed PB leprosy with BI of 0 and three were genetically related to the index case (parent and child of index case H03 and second degree relative of index case H30) and one was the spouse (index case H10).

Genome typing and antimicrobial resistance

M. leprae genomes of SSS and NS were genotyped by WGS or Sanger sequencing. A total of 60 samples (30 SSS and 30 NS) from MB and PB leprosy patients as well as HHC were selected for WGS with an RLEP qPCR Ct ranging from 16.2 to 37.2 (Supplementary Data 1). A total of 27 samples from 21 subjects (21 SSS and 6 NS) passed the library quality check and were successfully sequenced with a coverage ≥ 5 (Figure S1, Table S5). The limiting Ct value was 26.2 for SSS and 24.2 for NS.

On applying the genotyping system described by (3, 22), the following genotypes were found for these 21 subjects: 1A (n=5), 1B (n=4), 1C (n=3) and 1D (n=9). Interestingly, the four newly sequenced 1B genotype strains do not cluster with the two previously de-



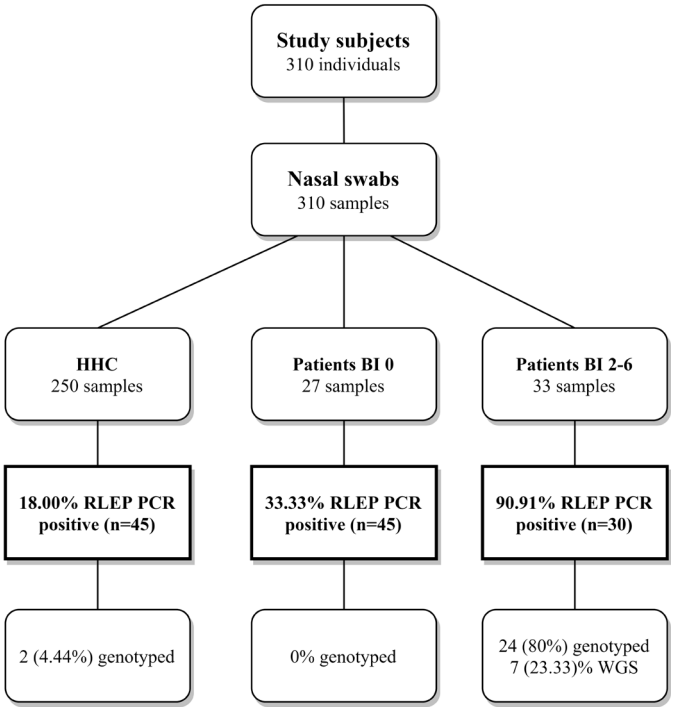


Figure 1. **Study design, RLEP positivity and genotyped samples.** Flow diagram providing an overview of the subjects recruited for this study. Slit skin smears (SSS) and nasal swabs (NS) collected per group; healthy household contacts (HHC), paucibacillary (PB) or multibacillary (MB) patients with BI 0, and MB patients with a bacteriological index (BI) 2-6. MB patients with BI 1 were not diagnosed within the course of this study. DNA was isolated from SSS and NS and screened for *M. leprae* DNA by RLEP PCR. Samples were genotyped by Sanger sequencing (3, 22) or Whole Genome Sequencing (29). Percentages of the samples positive for RLEP PCR and genotyped are shown.

scribed 1B strains from Yemen and Martinique (Figure 2). Instead, they form a new cluster in the phylogenetic tree located between genotypes the 1A and 1B, which we refer to as 1B-Bangladesh (Figure 2, blue, Supplementary Data 1). Using Sanger sequencing, the *M. leprae* strain for eight additional individuals were determined as 1A (n=4) or 1D (n=4). Three subjects, including 2 NS samples from HHC, carried genotype 1 but subtype could not be established due to lack of amplification of the subtyping loci (Supplementary Data 1).

The SNP used to differentiate genotype 1C (A61425G; Met90Thr, mutated in genotypes 1D and 2-4) is located at *esxA*. In contrast to previous observations (3, 22), we found that this position is not phylogenetically informative as it is also found unmutated (A; Met) in strains from the genotype 3I and 2E (Figure 2, green, Supplementary Data 2). Moreover, the 1C strains clustered in the middle of the 1D group suggesting that the previously described genotype 1C is part of the 1D genotype.

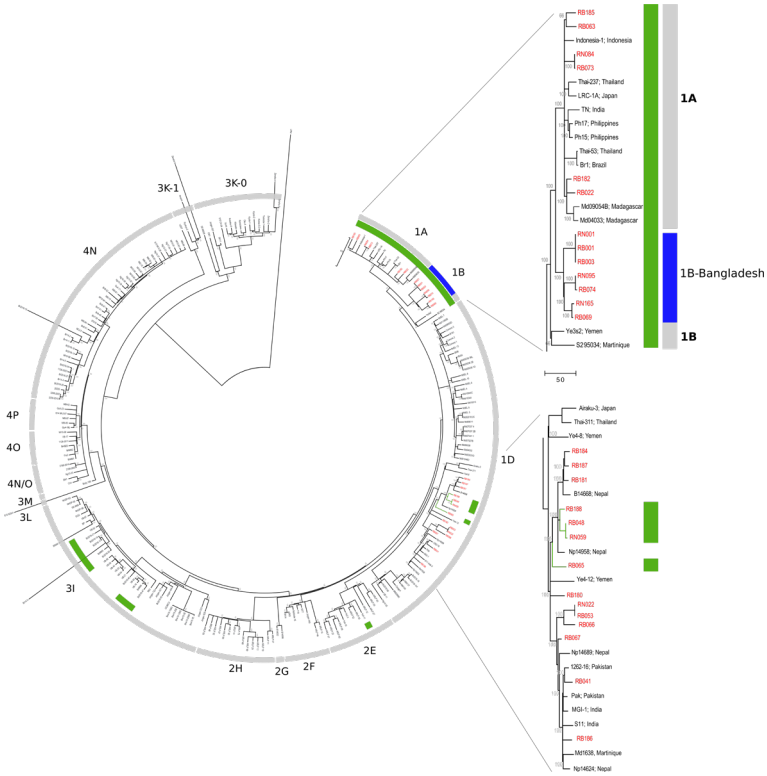


Figure 2. Phylogeography of *M. leprae* strains. Maximum parsimony tree of 259 genomes of *M. leprae* built in MEGA 7. Support values were obtained by bootstrapping 500 replicates. Branch lengths are proportional to nucleotide substitutions. The tree is rooted using *M. lepromatosis*. The strains from Bangladesh are shown in red and their exact organization in the tree is shown in the two zoomed sections of the genotypes 1A-B and 1D. Strains with an A at SNP61425 in the *esxA* gene are shown in green. The specific 1B-Bangladesh genotype/cluster of Bangladesh strains is shown in blue.

Finally, antimicrobial resistance was assessed in all genotyped strains either by WGS or Sanger sequencing. The latter was successful on 18 samples for *rpoB*, five sample for *folP1* and 15 samples for *gyrA* (Supplementary Data 1). None of the strains with a complete genome harbored drug-resistance mutations. One NS sample containing a missense mutation in the *rpoB* gene (Ser456Thr) in 50% of the sequences potentially leading to antimicrobial resistance (64) was identified by Sanger sequencing. Moreover, although not causing resistance, up to two silent mutations in three different positions of the *rpoB* gene relevant for antimicrobial resistance (432, 441 and 456) were also observed in several subjects.

Distribution and possible transmission of *M. leprae* genotypes

The most prevalent *M. leprae* genotype in the studied area of Bangladesh is 1D, found in 55% of the individuals (n=16, Table 1, Supplementary Data 1), followed by 1A in 31% (n=9), and 1B-Bangladesh in 14% (n=4). Genotype 1D is the most widely distributed throughout the whole area studied (Figure 3, blue and purple), whilst genotypes 1A and the here identified genotype 1B-Bangladesh are only observed in the eastern area (green and orange respectively). The latter genotype was found in 4 individuals: two from the same household and two unrelated subjects residing 56, 51 and 11 km from each other. However, due to privacy regulations on patient information to third parties it could not be established whether subjects in different households had had contact with any of the others.

Table 1. *M. leprae* genotypes identified in Bangladesh.

Genotype	Number of individuals	%
1A	9	31.0
1B-Bangladesh	4	13.8
1D	13	44.8
1D-*esxA*	3	10.4
1*	3	

M. leprae genotypes identified in patients and contacts from Bangladesh and the percentage of each subtype are shown. *M. leprae* DNA was isolated from slit skin smears (SSS) and/or nasal swabs (NS). Genotypes were determined by Whole Genome Sequencing (WGS) or Sanger sequencing according to (3, 22). The new subtype 1B-Bangladesh was identified by WGS and primers were then designed for use in Sanger sequencing (Table S2). 1D-*esxA* is 1D subtype containing an A at SNP61425 in the *esxA* gene, traditionally grouped as 1C (3, 22). This SNP is also found in strains from the genotype 3I and 2E (Figure 2, green). 1* are samples with genotype 1 for which the subtype could not be determined due to DNA concentration limit.

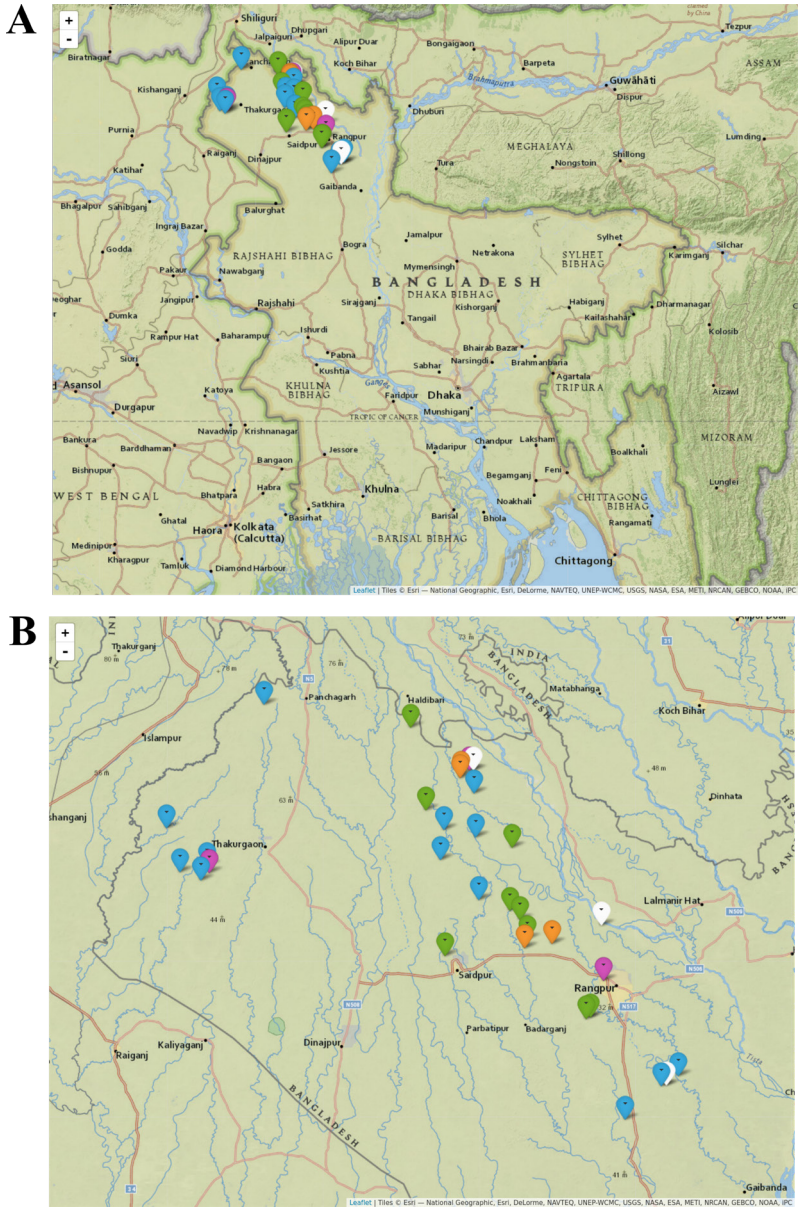


Figure 3. **Distribution of *M. leprae* genotypes in Bangladesh.** Map of Bangladesh including markers indicating the residence of every subject with at least one sample genotyped for *M. leprae* (A), and zoomed into the area of interest (B). Each marker indicates an individual for whom *M. leprae* genotype was determined, either from slit skin smear, nasal swab or both samples. Genotype 1A is shown in green, 1B-Bangladesh in orange, 1D in blue, 1D-*esxA* in purple and 1* in white. 1D-*esxA* is 1D subtype containing an A at SNP61425 in the *esxA* gene, formerly grouped as 1C (3, 22). 1* are samples with genotype 1 for which the subtype could not be determined. The figure was drawn in R (v3.4.3) with the package *leaflet* (v2.0.2) using maps from Esri – National Geographic with permission. Scale Not Given. "National Geographic World Map". December 13, 2011. http://www.arcgis.com/home/item.html?id=b9b1b422198944fbbd5250b3241691b6 (September 2, 2019).

n a total of four households the same *M. leprae* genotype was detected in two individuals (Supplementary Data 1). In the first household, both subjects were MB patients and WGS showed no genetic variation between both patients' genomes (RB001 and RB003, 1B-Bangladesh genotype, Supplementary Data 2). In the second household with two MB patients, the *M. leprae* whole genome was only obtained from the index case but the same genotype, 1A, and a strain-specific SNP of the index case (Table S3 and S4) was also identified by Sanger sequencing in the other patient (RB182 and RB266). In the last two households, the genotype of strains from both MB index cases' were determined by WGS (RB030, genotype 1D) and, by Sanger sequencing (RB065, genotype 1D-*esxA*), while the *M. leprae* genotype 1 was located in the NS of both HHC but no further subtyping was possible.

Comparison of *M. leprae* genomes from SSS and NS

M. leprae whole genomes of six patients were successfully recovered from both SSS and NS. The *M. leprae* genotypes obtained in each subject were in agreement between the two samples (Table 2). Genomic comparison showed no differences between DNA from SSS and NS for two patients: RB001-RN001 (genotype 1B-Bangladesh) and RB048-RN059 (genotype 1D-*esxA*, Supplementary Data 2, Figure 2).

In a third patient (RB073-RN084, genotype 1A), both strains were identical except that in the NS strain 17% of 24 reads in *ml1512* harbored a T1824441C (Gly56Asp) (Table 2). Interestingly, *ml1512* which encodes a ribonuclease J is one of the most mutated genes among all *M. leprae* strains (29) and mutations at this gene were also observed in two different patients: in the NS of RN022-RB053 (genotype 1D) 35% of 26 reads had a mutated allele (G1823127A; Ser494Leu) and 20% of 21 reads had an insertion of a C at position 1823613 probably leading to a deleterious frameshift; in the SSS of RB074-RN095 (genotype 1B-Bangladesh) 92% of 92 reads presented a missense mutation (G1823098A; Leu504Phe). Interestingly, RB074 harbored a G660474C mutation in *metK,* a probable methionine adenosyl-transferase, which was also found in 75% of 12 reads of the NS and is uniquely found in this subject's *M. leprae* genomes. Additionally, RN095 also displayed mutations at several positions in *ml1750* (a putative nucleotide cyclase): 48% of 21 reads had C2116695A (Pro100Thr), 20% of 20 reads had A2116670G (Gln108Arg) and 19% of 16 reads had C2116490T mutation (Arg168His). These positions were partially or totally mutated in other strains from different genotypes: SM1 (100% Pro100Ser; genotype 4), Ml9-81 (Mali, 30% Arg168His; genotype 4N) and Md05036 (Madagascar, 90% Gln108Arg, genotype 1D-Mada) (29, 62).

The patient with the *M. leprae* strains that were the most genetically different between the NS and SSS carried the genotype 1B-Bangladesh (RB069 and RN165). The NS strain had a mixed population in *glpQ* (29% of 76 reads C9231T, Leu34Phe) and *ml1752* (15% of 94

Table 2. Intraindividual *M. leprae* genomic differences.

Samples	Genotype	Mutation and amino-acid change	Genomic region	NS		SSS		Other genomes with similar position mutated (variant,% reads, genotype)
				% reads	Aligned reads	% reads	Aligned reads	
RB073-RN084	1A	T1824441C; Gly56Asp	*ml1512*	17%	24	-	-	-
RB053-RN022	1D	G1823127A; Ser494Leu	*ml1512*	35%	26	-	-	Ml10-98 (Ser494Ala, 95%, 4N)
		1823613_1823614insC;Asp332fs	*ml1512*	20%	21	-	-	ARLP-23 (Asp332fs, 80%, 2E)
RB074-RN095	1B-Bangladesh	G1823098A; Leu504Phe	*ml1512*	-	-	92%	92	-
		G660474C; Val252Leu	*metK*	75%	12	100%	92	-
		C2116695A; Pro100Thr	*ml1750*	48%	21	-	-	SM1, (Pro100Ser, 100%, 4N/O)
		A2116670G; Gln108Arg	*ml1750*	20%	20	-	-	LRC-1A (Gln108His, 100%,1A) Md0536 (Gln108Arg 90%, 1D-Mada)
		C2116490T; Arg168His	*ml1750*	19%	16	-	-	Br14-3 (5 Arg168Cys, 1%,3I) Arg168His: Br2016-17 (22%, 3I); Comore-3 (36%; 1D-Mada); Ml9-81 (29%; 4N)
RB069-RN165	1B-Bangladesh	C95231T; Leu34Phe	*glpQ*	29%	76	-	-	-
		C2121552T; Val226Ile	*ml1752*	15%	94	-	-	-

Genomic differences between *M. leprae* genomes obtained from slit skin smears (SSS; RB) and nasal swabs (NS; RN) of the same MB patient. The table shows the percentage of mutated **non-duplicated** reads and the total number of **non-duplicated** reads aligned at the position of the mutation for each sample type. Several positions were partially or totally mutated in other *M. leprae* strains from different genotypes. dup= duplication; *=stop; ins=insertion; del=deletion.

reads C2121552T, Val226Ile). These genes encode a glycerophosphoryl diester phosphodiesterase, a putative nucleotide cyclase, and a conserved hypothetical protein. Notably, *ml1752* is also one of the most hypermutated genes in *M. leprae* (29).

For 11 patients a whole genome sequence was recovered only from SSS but Sanger sequencing was successfully performed to identify the subtype in NS. The same subtype observed in SSS was also found in the NS of these 11 patients. Moreover, unique *M. leprae* SNPs identified in the genomes of the SSS (Table S3 and S4) were also detected in seven of the genomes of the NS of these patients (Supplementary Data 1).

Combining host and pathogen detection

Anti-PGL-I IgM levels were determined in plasma of 308 subjects. All MB patients with BI 2-6 (n=33) showed high levels for anti-PGL-I IgM (Table 3) in line with the general consensus (53, 65). Out of the patients (both MB and PB) with BI 0 (n=27), nine (33.3%) were positive for anti-PGL-I IgM. Similarly, 36.8% of HHC showed positivity (n=92). From these 92 positive individuals, 70 were neither positive for SSS nor NS RLEP PCR (Supplementary Data 1).

Table 3. Anti-PGL-I IgM positivity.

Genotype	Number of positive individuals	% of positivity
MB patients BI 2-6 (n=33)	33	100.0
Patients BI 0 (n=27)	9	33.3
Healthy household contacts (n=250)	92	36.8

Anti-PGL-I antibody levels were measured by up-converting phosphor lateral flow assay specific for *M. leprae* PGL-I IgM antibodies (UCP-LFA) using the Ratio (R) of the Test (T) and flow control (FC) lines as units. Ratios of ≥ 0.10 were considered positive.

Of the four contacts who developed leprosy within the first year after sample collection, two were positive for anti-PGL-I IgM whilst negative for RLEP PCRs 10 and 12 months before diagnosis. Since the two other subjects had a positive RLEP PCR in SSS or NS 5 or 8 months before diagnosis, it can be concluded that all of the new cases showed positivity either for host- or pathogen-associated diagnostics 5-12 months before developing disease.

Individual anti-PGL-I levels were compared to RLEP Ct values in SSS and NS samples (Figure 4), showing an expected negative correlation between anti-PGL-I ratio and Ct value since both values are associated with BI. A subtle difference can be observed in the correlation between anti-PGL-I IgM levels and RLEP Ct if the qPCR was performed on either SSS or NS DNA, with a coefficient of determination (R^2) 0.73 and 0.69 respectively.

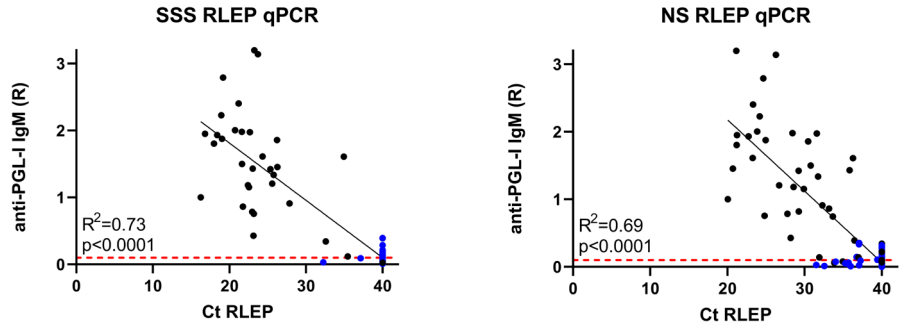


Figure 4. **Correlation of IgM antibodies against PGL-I to Ct of RLEP qPCR.** Quantified levels of pathogen DNA (qPCR) and host immunity were correlated for samples selected for qPCR analysis based on RLEP positivity in multiple individuals in one household. Each dot represents a sample from one individual; leprosy patients are indicated in black, and healthy household contact in blue. Anti-PGL-I antibody levels were measured by up-converting reporter particles lateral flow assay specific for *M. leprae* PGL-I IgM antibodies (αPGL-I UCP-LFA) using the Ratio (R) of the Test (T) and flow control (FC) lines as units. Ratios of ≥ 0.10 were considered positive as indicated by the red dashed line. RLEP cycle threshold (Ct) values are indicated on the x-axis and were measured by qPCR to detect *M. leprae* DNA in slit skin smears (SSS, left) and nasal swabs (NS, left). Undetermined Cts are depicted as Ct 40.

Discussion

In this study we investigated *M. leprae* transmission patterns in Bangladesh by detecting and sequencing *M. leprae* DNA derived from SSS and NS of patients and their household members. Our data represents the first report of *M. leprae* DNA detection in HHC from Bangladesh. We observed moderate positivity in HHC which was similar to positivity of leprosy patients with BI 0. A new genotype, 1B-Bangladesh, was sequenced and we showed that the previously described 1C genotype is part of the 1D group. Additionally, a negative correlation between RLEP Ct values indicating the amount of *M. leprae* DNA and anti-PGL-I IgM levels was observed.

M. leprae DNA detection frequency in HHC from Bangladesh (12.3% in SSS and 18.0% in NS) was in line with previous studies conducted in several hyperendemic areas of Brazil, Colombia and Indonesia (45, 48-52). In India higher positivity (21.0%) in SSS of HHC was reported (38) whereas in two Brazilian studies from Uberlandia, up to 42.4% positivity in SSS (47) and 49.0% in NS (12) were observed. Three factors may limit the translation of these high positive results from India and Brazil to our study: i) the sample sizes of the Indian (38) and one of the Brazilian studies (12) were smaller (n=28 and n=104, respectively versus n=250 HHC in this study); ii) we conducted a more stringent approach by testing the samples in three independent PCRs; and iii) the epidemiology and incidence of MB cases in India and Brazil differ from the studied area in Bangladesh where MB leprosy cas-

es occur less frequently than PB and also usually display a low BI (56, 66).

M. leprae DNA in the nose does not indicate disease but (transient) colonization whilst presence of *M. leprae* in SSS indicates infection. Thus, the higher RLEP PCR positivity in NS compared to SSS in patients with BI 0 and HHC likely represents the (virtual) absence of bacteria causing infection in these individuals despite colonization.

A longitudinal study conducted in Brazil (67), investigated SSS from 995 HHC by qPCR including follow-up for at least 3 years with occurrence of five new cases. The authors reported 20% qPCR positivity in HHC representing future new cases compared to 9% in HHC without disease. However, this difference was not significant. In line with that study, we found that *M. leprae* DNA detection was slightly higher (25% vs 18% in NS and 25% vs 12% in SSS) in contacts who developed disease compared to those who did not. Additionally, we determined anti-PGL-I IgM levels, which correlated well with Ct qPCR values. Notwithstanding this correlation, serology provided added value: when positivity in any of the three techniques was considered (NS PCR, SSS PCR or anti PGL-I), all of the contacts (n=4) who developed leprosy within the first year after sample collection, were identified. In agreement with this, a combination of host and pathogen markers was previously integrated in a machine learning model using qPCR and serological data (antibodies against LID-1 or ND-O-LID) (46) to identify prospective leprosy patients among contacts leading to an increased sensitivity in diagnosis, particularly in PB leprosy. It is of note that in our study, three of the four contacts who developed leprosy were genetically related to the index cases in their households, stressing the previously described role of genetic inheritance in the development of leprosy (14-17, 68). For this reason, the association between leprosy and the genetics of this Bangladeshi population is currently being studied.

Genotype 1 was identified in all the *M. leprae* genomes retrieved from Bangladesh, consistent with previous data from (22). In Bangladesh, leprosy was likely introduced through the southern Asian route (genotype 1) leading to the spread of *M. leprae* into the Indian subcontinent, Indonesia and the Philippines (22, 29). Subtype 1D was predominantly present in Bangladesh but in addition we detected 1A and identified a new 1B-Bangladesh genotype. The presence of multiple subtypes of *M. leprae* genotype 1 in Bangladesh is in line with previous studies in South Asian countries such as India, Nepal, Thailand, Indonesia and Pakistan (22, 29). The new 1B-Bangladesh genotype is thus far restricted to Bangladesh and two of the four individuals carrying this strain were part of the same household whilst the other two did not have any relationship with each other and were located in different areas with a distance of up to 56 km between them. This suggests that this genotype could be a common subtype in Bangladesh although additional studies are required to confirm this. Thus, it is of interest to include the 1B-Bangladesh SNP specific primers in future epidemiological studies, particularly in other (neighbouring) Asian

countries such as India where genotype 1 is widely established (22).

In contrast to the general belief (3, 22), we observed that subtype 1C does not form an independent subtype but actually belongs to subtype 1D. SNP61425 used to distinguish genotypes 1A-C is located at *esxA* encoding the virulence factor ESAT-6 (22). The Esx protein family also revealed high diversity in the more pathogenic mycobacterium, *M. tuberculosis* (69), and is involved in host-pathogen interaction. Of note is that ESAT-6 (ML0049) is a potent T-cell antigen (70, 71), thus mutations in *esxA* gene might indicate drift due to immune pressure potentially explaining the occurrence of mutations at SNP61425 in different genotypes.

In a recent survey in 19 countries during 2009-2015 (72), 8% of the cases presented mutations resulting in antimicrobial resistance and resistance to up to two different drugs was detected. In our study, which is the first investigating *M. leprae* drug resistance in Bangladesh, we detected no resistance by WGS, however, a partial missense mutation in the codon for Ser456 of the *rpoB* gene potentially leading to rifampicin resistance (n=1) was observed by Sanger sequencing. This could be the result of a mixed infection or an emerging mutation of the *M. leprae* strain occurring in the patient. Silent mutations in the *rpoB* gene were detected in several locations, which indicates that mutations do occur, and this may eventually lead to missense mutations conferring antimicrobial resistance. However, drug resistance is not only induced by genetic mutations in drug targets, efflux systems resulting in antimicrobial resistance have also been described for *M. leprae* (73). This mechanism of drug resistance is unnoticed in genomic tests and needs to be further investigated for leprosy especially in the light of the huge efforts recently initiated and WHO-endorsed for post-exposure prophylaxis (PEP) using antibiotic regimens (44, 74, 75).

Despite our finding that NS samples were more frequently positive for *M. leprae* DNA, recovery of *M. leprae* whole genomes from SSS has proven to be more successful than from NS. This is due to the higher number of bacteria in SSS of patients. However, the importance of genotyping NS as well as skin biopsies or SSS to better understand transmission has been previously discussed (76), as the nasal respiratory route remains one of the most plausible modes of infection (12, 13). In a recent study, skin biopsies and NS of patients were compared by VNTR typing and the authors found that out of 38 patients, differences between SSS and NS in seven loci were observed in 33 patients (77). Although the *M. leprae* genomes from SSS and NS analysed in our study were almost identical, we observed that genomes obtained from NS harboured more mutations, especially in previously reported (29) hypermutated genes. This could be an indication of in-host evolution in the nasal mucosa, mixed infection or mixed colonization. Thus, it may imply that colonization occurred with two different strains causing a co-infection or that one is present, likely from a later colonization, but does not cause the disease.

The presence of mixed infections emphasises once more the importance of monitoring asymptomatic carriers, who may contribute to the spread of the pathogen. Therefore, providing PEP only to the (close) contacts of leprosy patients might not be sufficient to stop transmission. Instead, an approach including the entire community but targeting only individuals testing positive for *M. leprae* DNA or host immune markers associated to *M. leprae* infection, would represent a preferred strategy for PEP.

Data availability

Sequence data are available from the NCBI Sequence Read Archive (SRA) under the bioprojects PRJNA605605 and PRJNA592722, biosamples SAMN14072760-775 and SAMN13438761-771.

https://www.ncbi.nlm.nih.gov/bioproject/PRJNA605605

https://www.ncbi.nlm.nih.gov/bioproject/PRJNA592722

Acknowledgements

The authors gratefully acknowledge all patients and control participants. LUMC. Erasmus MC and TLMI,B are part of the IDEAL (***I***nitiative for ***D***iagnostic and ***E***pidemiological ***A***ssays for ***L***eprosy) Consortium.

Funding statement

This study was supported by an R2STOP Research grant from effect:hope, Canada and The Mission to End Leprosy, Ireland; the Order of Malta-Grants-for-Leprosy-Research (MALTA-LEP, to AG); the Foundation Raoul Follereau (to STC); the Q.M. Gastmann-Wichers Foundation (to AG); the Leprosy Research Initiative (LRI) together with the Turing Foundation (ILEP#703.15.07).

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. Han XY, Seo YH, Sizer KC, Schoberle T, May GS, Spencer JS, et al. A new Mycobacterium species causing diffuse lepromatous leprosy. American journal of clinical pathology. 2008;130(6):856-64.
2. Avanzi C, Del-Pozo J, Benjak A, Stevenson K, Simpson VR, Busso P, et al. Red squirrels in the British Isles are infected with leprosy bacilli. Science. 2016;354(6313):744-7.
3. Truman RW, Singh P, Sharma R, Busso P, Rougemont J, Paniz-Mondolfi A, et al. Probable zoonotic leprosy in the Southern United States. The New England journal of medicine. 2011;364(17):1626-33.
4. Sharma R, Singh P, Loughry WJ, Lockhart JM, Inman WB, Duthie MS, et al. Zoonotic leprosy in the Southeastern United States. Emerg Infect Dis. 2015;21(12):2127-34.
5. Honap TP, Pfister LA, Housman G, Mills S, Tarara RP, Suzuki K, et al. *Mycobacterium leprae* genomes from naturally infected nonhuman primates. PLOS Neglected Tropical Diseases. 2018;12(1):e0006190.
6. Schilling A-K, van Hooij A, Corstjens P, Lurz P, DelPozo J, Stevenson K, et al. Detection of humoral immunity to mycobacteria causing leprosy in Eurasian red squirrels (Sciurus vulgaris) using a quantitative rapid test2019.

7. Tio-Coma M, Sprong H, Kik M, van Dissel JT, Han XY, Pieters T, et al. Lack of evidence for the presence of leprosy bacilli in red squirrels from North-West Europe. Transbound Emerg Dis. 2019.
8. World Health Organization. Global leprosy update, 2018: moving towards a leprosy-free world. Weekly Epidemiological Record. 2019;94(35/36):389-412.
9. Singh P, Cole ST. *Mycobacterium leprae*: genes, pseudogenes and genetic diversity. Future Microbiol. 2011;6(1):57-71.
10. Ridley DS, Jopling WH. Classification of leprosy according to immunity. A five-group system. Int J Lepr Other Mycobact Dis. 1966;34(3):255-73.
11. Kumar B, Uprety S, Dogra S. Clinical diagnosis of leprosy. In: Scollard DM, Gills TP, editors. International textbook of leprosy. www.internationaltextbookofleprosy.org2017.
12. Araujo S, Freitas LO, Goulart LR, Goulart IM. Molecular evidence for the aerial route of infection of *Mycobacterium leprae* and the role of asymptomatic carriers in the persistence of leprosy. Clin Infect Dis. 2016;63(11):1412-20.
13. Bratschi MW, Steinmann P, Wickenden A, Gillis TP. Current knowledge on *Mycobacterium leprae* transmission: a systematic literature review. Lepr Rev. 2015;86(2):142-55.
14. Zhang FR, Huang W, Chen SM, Sun LD, Liu H, Li Y, et al. Genomewide association study of leprosy. N Engl J Med. 2009;361(27):2609-18.
15. Mira MT, Alcais A, Nguyen VT, Moraes MO, Di Flumeri C, Vu HT, et al. Susceptibility to leprosy is associated with PARK2 and PACRG. Nature. 2004;427(6975):636-40.
16. Wang D, Xu L, Lv L, Su LY, Fan Y, Zhang DF, et al. Association of the LRRK2 genetic polymorphisms with leprosy in Han Chinese from Southwest China. Genes Immun. 2015;16(2):112-9.
17. Sales-Marques C, Cardoso CC, Alvarado-Arnez LE, Illaramendi X, Sales AM, Hacker MA, et al. Genetic polymorphisms of the IL6 and NOD2 genes are risk factors for inflammatory reactions in leprosy. PLoS Negl Trop Dis. 2017;11(7):e0005754.
18. Moet FJ, Meima A, Oskam L, Richardus JH. Risk factors for the development of clinical leprosy among contacts, and their relevance for targeted interventions. Lepr Rev. 2004;75(4):310-26.
19. Dwivedi VP, Banerjee A, Das I, Saha A, Dutta M, Bhardwaj B, et al. Diet and nutrition: An important risk factor in leprosy. Microb Pathog. 2019;137:103714.
20. Cole ST, Eiglmeier K, Parkhill J, James KD, Thomson NR, Wheeler PR, et al. Massive gene decay in the leprosy bacillus. Nature. 2001;409(6823):1007-11.
21. Monot M, Honore N, Garnier T, Araoz R, Coppee JY, Lacroix C, et al. On the origin of leprosy. Science. 2005;308(5724):1040-2.
22. Monot M, Honore N, Garnier T, Zidane N, Sherafi D, Paniz-Mondolfi A, et al. Comparative genomic and phylogeographic analysis of *Mycobacterium leprae*. Nat Genet. 2009;41(12):1282-9.
23. Donoghue HD, Holton J, Spigelman M. PCR primers that can detect low levels of *Mycobacterium leprae* DNA. J Med Microbiol. 2001;50(2):177-82.
24. Truman RW, Andrews PK, Robbins NY, Adams LB, Krahenbuhl JL, Gillis TP. Enumeration of *Mycobacterium leprae* Using Real-Time PCR. PLOS Neglected Tropical Diseases. 2008;2(11):e328.
25. Martinez AN, Ribeiro-Alves M, Sarno EN, Moraes MO. Evaluation of qPCR-based assays for leprosy diagnosis directly in clinical specimens. PLoS Negl Trop Dis. 2011;5(10):e1354.
26. Braet S, Vandelannoote K, Meehan CJ, Brum Fontes AN, Hasker E, Rosa PS, et al. The repetitive element RLEP is a highly specific target for detection of *Mycobacterium leprae*. J Clin Microbiol. 2018;56(3).
27. Han XY, Silva FJ. On the age of leprosy. PLoS neglected tropical diseases. 2014;8(2):e2544-e.
28. Donoghue HD. Tuberculosis and leprosy associated with historical human population movements in Europe and beyond - an overview based on mycobacterial ancient DNA. Ann Hum Biol. 2019;46(2):120-8.
29. Benjak A, Avanzi C, Singh P, Loiseau C, Girma S, Busso P, et al. Phylogenomics and antimicrobial resistance of the leprosy bacillus *Mycobacterium leprae*. Nat Commun. 2018;9(1):352.
30. Schuenemann VJ, Singh P, Mendum TA, Krause-Kyora B, Jager G, Bos KI, et al. Genome-wide comparison of medieval and modern *Mycobacterium leprae*. Science. 2013;341(6142):179-83.
31. Schuenemann VJ, Avanzi C, Krause-Kyora B, Seitz A, Herbig A, Inskip S, et al. Ancient genomes reveal a high diversity of *Mycobacterium leprae* in medieval Europe. PLoS Pathog. 2018;14(5):e1006997.
32. Mendum TA, Schuenemann VJ, Roffey S, Taylor GM, Wu H, Singh P, et al. *Mycobacterium leprae* genomes from a British medieval leprosy hospital: towards understanding an ancient epidemic. BMC Genomics. 2014;15:270.
33. Suzuki K, Takigawa W, Tanigawa K, Nakamura K, Ishido Y, Kawashima A, et al. Detection of *Mycobacterium leprae* DNA from archaeological skeletal remains in Japan using whole genome amplification and polymerase chain reaction. PLoS One. 2010;5(8):e12422.
34. Krause-Kyora B, Nutsua M, Boehme L, Pierini F, Pedersen DD, Kornell S-C, et al. Ancient DNA study reveals HLA susceptibility locus for leprosy in medieval Europeans. Nature communications. 2018;9(1):1569-.

35. Schilling A-K, Avanzi C, Ulrich RG, Busso P, Pisanu B, Ferrari N, et al. British Red Squirrels Remain the Only Known Wild Rodent Host for Leprosy Bacilli. Frontiers in Veterinary Science. 2019;6(8).
36. Tió-Coma M, Wijnands T, Pierneef L, Schilling AK, Alam K, Roy JC, et al. Detection of *Mycobacterium leprae* DNA in soil: multiple needles in the haystack. Scientific Reports. 2019;9(1):3165.
37. Lavania M, Katoch K, Sachan P, Dubey A, Kapoor S, Kashyap M, et al. Detection of *Mycobacterium leprae* DNA from soil samples by PCR targeting RLEP sequences. J Commun Dis. 2006;38(3):269-73.
38. Turankar RP, Lavania M, Chaitanya VS, Sengupta U, Darlong J, Darlong F, et al. Single nucleotide polymorphism-based molecular typing of *M. leprae* from multicase families of leprosy patients and their surroundings to understand the transmission of leprosy. Clin Microbiol Infect. 2014;20(3):O142-9.
39. Turankar RP, Lavania M, Singh M, Siva Sai KS, Jadhav RS. Dynamics of *Mycobacterium leprae* transmission in environmental context: deciphering the role of environment as a potential reservoir. Infect Genet Evol. 2012;12(1):121-6.
40. Turankar RP, Lavania M, Singh M, Sengupta U, Siva Sai K, Jadhav RS. Presence of viable *Mycobacterium leprae* in environmental specimens around houses of leprosy patients. Indian J Med Microbiol. 2016;34(3):315-21.
41. Lavania M, Katoch K, Katoch VM, Gupta AK, Chauhan DS, Sharma R, et al. Detection of viable *Mycobacterium leprae* in soil samples: insights into possible sources of transmission of leprosy. Infect Genet Evol. 2008;8(5):627-31.
42. Turankar RP, Lavania M, Darlong J, Siva Sai KSR, Sengupta U, Jadhav RS. Survival of *Mycobacterium leprae* and association with Acanthamoeba from environmental samples in the inhabitant areas of active leprosy cases: A cross sectional study from endemic pockets of Purulia, West Bengal. Infect Genet Evol. 2019;72:199-204.
43. Van Dissel JT, Pieters T, Geluk A, Maat G, Menke HE, Tio-Coma M, et al. Archival, paleopathological and aDNA-based techniques in leprosy research and the case of Father Petrus Donders at the Leprosarium 'Batavia', Suriname. Int J Paleopathol. 2019;27:1-8.
44. Richardus R, Alam K, Kundu K, Chandra Roy J, Zafar T, Chowdhury AS, et al. Effectiveness of single-dose rifampicin after BCG vaccination to prevent leprosy in close contacts of patients with newly diagnosed leprosy: A cluster randomized controlled trial. Int J Infect Dis. 2019;88:65-72.
45. Gama RS, Gomides TAR, Gama CFM, Moreira SJM, de Neves Manta FS, de Oliveira LBP, et al. High frequency of *M. leprae* DNA detection in asymptomatic household contacts. BMC Infect Dis. 2018;18(1):153.
46. Gama RS, Souza MLM, Sarno EN, Moraes MO, Goncalves A, Stefani MMA, et al. A novel integrated molecular and serological analysis method to predict new cases of leprosy amongst household contacts. PLoS Negl Trop Dis. 2019;13(6):e0007400.
47. Santos DFD, Mendonca MR, Antunes DE, Sabino EFP, Pereira RC, Goulart LR, et al. Molecular, immunological and neurophysiological evaluations for early diagnosis of neural impairment in seropositive leprosy household contacts. PLoS neglected tropical diseases. 2018;12(5):e0006494.
48. Brito e Cabral P, Junior JE, de Macedo AC, Alves AR, Goncalves TB, Brito e Cabral TC, et al. Anti-PGL1 salivary IgA/IgM, serum IgG/IgM, and nasal *Mycobacterium leprae* DNA in individuals with household contact with leprosy. Int J Infect Dis. 2013;17(11):e1005-10.
49. Carvalho RS, Foschiani IM, Costa M, Marta SN, da Cunha Lopes Virmond M. Early detection of M. leprae by qPCR in untreated patients and their contacts: results for nasal swab and palate mucosa scraping. Eur J Clin Microbiol Infect Dis. 2018;37(10):1863-7.
50. van Beers SM, Izumi S, Madjid B, Maeda Y, Day R, Klatser PR. An epidemiological study of leprosy infection by serology and polymerase chain reaction. Int J Lepr Other Mycobact Dis. 1994;62(1):1-9.
51. Romero-Montoya M, Beltran-Alzate JC, Cardona-Castro N. Evaluation and Monitoring of *Mycobacterium leprae* Transmission in Household Contacts of Patients with Hansen's Disease in Colombia. PLoS Negl Trop Dis. 2017;11(1):e0005325.
52. Cardona-Castro N, Beltran-Alzate JC, Manrique-Hernandez R. Survey to identify *Mycobacterium leprae*-infected household contacts of patients from prevalent regions of leprosy in Colombia. Mem Inst Oswaldo Cruz. 2008;103(4):332-6.
53. van Hooij A, Tjon Kon Fat EM, van den Eeden SJF, Wilson L, Batista da Silva M, Salgado CG, et al. Field-friendly serological tests for determination of *M. leprae*-specific antibodies. Sci Rep. 2017;7(1):8868.
54. Penna ML, Penna GO, Iglesias PC, Natal S, Rodrigues LC. Anti-PGL-1 Positivity as a Risk Marker for the Development of Leprosy among Contacts of Leprosy Cases: Systematic Review and Meta-analysis. PLoS Negl Trop Dis. 2016;10(5):e0004703.
55. Barbieri RR, Manta FSN, Moreira SJM, Sales AM, Nery JAC, Nascimento LPR, et al. Quantitative polymerase chain reaction in paucibacillary leprosy diagnosis: A follow-up study. PLoS Negl Trop Dis. 2019;13(3):e0007147.

56. van Hooij A, van den Eeden S, Richardus R, Tjon Kon Fat E, Wilson L, Franken K, et al. Application of new host biomarker profiles in quantitative point-of-care tests facilitates leprosy diagnosis in the field. EBioMedicine. 2019;47:301-8.
57. Spencer JS, Brennan PJ. The role of *Mycobacterium leprae* phenolic glycolipid I (PGL-I) in serodiagnosis and in the pathogenesis of leprosy. Lepr Rev. 2011;82(4):344-57.
58. van Hooij A, Tjon Kon Fat EM, Batista da Silva M, Carvalho Bouth R, Cunha Messias AC, Gobbo AR, et al. Evaluation of Immunodiagnostic Tests for Leprosy in Brazil, China and Ethiopia. Sci Rep. 2018;8(1):17920.
59. Martinez AN, Lahiri R, Pittman TL, Scollard D, Truman R, Moraes MO, et al. Molecular determination of *Mycobacterium leprae* viability by use of real-time PCR. J Clin Microbiol. 2009;47(7):2124-30.
60. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. 2012;22(3):568-76.
61. Rimmer A, Phan H, Mathieson I, Iqbal Z, Twigg SRF, Wilkie AOM, et al. Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. Nat Genet. 2014;46(8):912-8.
62. Avanzi C, Lécorché E, Rakotomalala FA, Benjak A, Rapelanoro Rabenja F, Ramarozatovo LS, et al. Population Genomics of *Mycobacterium leprae* Reveals a New Genotype in Madagascar and the Comoros. Frontiers in microbiology. 2020;11(711).
63. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nature biotechnology. 2011;29(1):24-6.
64. World Health Organization. A guide for surveillance of antimicrobial resistance in leprosy. New Delhi: World Health Organization, Region Office for South-East Asia; 2017.
65. Geluk A, Duthie MS, Spencer JS. Postgenomic *Mycobacterium leprae* antigens for cellular and serological diagnosis of M. leprae exposure, infection and leprosy disease. Leprosy review. 2011;82(4):402-21.
66. Richardus RA, van der Zwet K, van Hooij A, Wilson L, Oskam L, Faber R, et al. Longitudinal assessment of anti-PGL-I serology in contacts of leprosy patients in Bangladesh. PLOS Neglected Tropical Diseases. 2017;11(12):e0006083.
67. Manta FSN, Barbieri RR, Moreira SJM, Santos PTS, Nery JAC, Duppre NC, et al. Quantitative PCR for leprosy diagnosis and monitoring in household contacts: A follow-up study, 2011-2018. Sci Rep. 2019;9(1):16675.
68. Uaska Sartori PV, Penna GO, Bührer-Sékula S, Pontes MAA, Gonçalves HS, Cruz R, et al. Human Genetic Susceptibility of Leprosy Recurrence. Scientific reports. 2020;10(1):1284-.
69. Uplekar S, Heym B, Friocourt V, Rougemont J, Cole ST. Comparative genomics of Esx genes from clinical isolates of *Mycobacterium tuberculosis* provides evidence for gene conversion and epitope variation. Infection and immunity. 2011;79(10):4042-9.
70. Geluk A, van Meijgaarden KE, Franken KLMC, Subronto YW, Wieles B, Arend SM, et al. Identification and characterization of the ESAT-6 homologue of *Mycobacterium leprae* and T-cell cross-reactivity with Mycobacterium tuberculosis. Infection and immunity. 2002;70(5):2544-8.
71. Geluk A, van Meijgaarden KE, Franken KLMC, Wieles B, Arend SM, Faber WR, et al. Immunological crossreactivity of the *Mycobacterium leprae* CFP-10 with its homologue in *Mycobacterium tuberculosis*. Scandinavian journal of immunology. 2004;59(1):66-70.
72. Cambau E, Saunderson P, Matsuoka M, Cole ST, Kai M, Suffys P, et al. Antimicrobial resistance in leprosy: results of the first prospective open survey conducted by a WHO surveillance network for the period 2009-15. Clin Microbiol Infect. 2018;24(12):1305-10.
73. Machado D, Lecorche E, Mougari F, Cambau E, Viveiros M. Insights on *Mycobacterium leprae* Efflux Pumps and Their Implications in Drug Resistance and Virulence. Front Microbiol. 2018;9:3072.
74. Mieras LF, Taal AT, van Brakel WH, Cambau E, Saunderson PR, Smith WCS, et al. An enhanced regimen as post-exposure chemoprophylaxis for leprosy: PEP+. BMC Infect Dis. 2018;18(1):506.
75. Barth-Jaeggi T, Steinmann P, Mieras L, van Brakel W, Richardus JH, Tiwari A, et al. Leprosy Post-Exposure Prophylaxis (LPEP) programme: study protocol for evaluating the feasibility and impact on case detection rates of contact tracing and single dose rifampicin. BMJ Open. 2016;6(11):e013633.
76. Fontes ANB, Lima L, Mota RMS, Almeida RLF, Pontes MA, Goncalves HS, et al. Genotyping of *Mycobacterium leprae* for better understanding of leprosy transmission in Fortaleza, Northeastern Brazil. PLoS Negl Trop Dis. 2017;11(12):e0006117.
77. Lima L, Fontes ANB, Li W, Suffys PN, Vissa VD, Mota RMS, et al. Intrapatient comparison of *Mycobacterium leprae* by VNTR analysis in nasal secretions and skin biopsy in a Brazilian leprosy endemic region. Lepr Rev. 2016;87(4):486-500.

Supplementary Material

Table S1. Cohort characterization.

Group	Subjects	Gender	Age	RJ Classification	BI
MB patients BI 2-6*	33	18% Female 82% Male	34	17 LL 15 BL 1 BT	14 BI-6 3 BI-5 5 BI-6 6 BI-5 4 BI-4 1 BI-2
PB/MB patients BI 0	27	63% Female 37% Male	31	24 BT 2 TT 1 UD	24 BI-0 3 UD
HHC	250	52% Female 48% Male	30	-	-

Group, number of subjects, percentage of female and male, median of age, Ridley-Jopling classifications and BI of patients are shown. *31 were index cases of the study. UD: Undetermined.

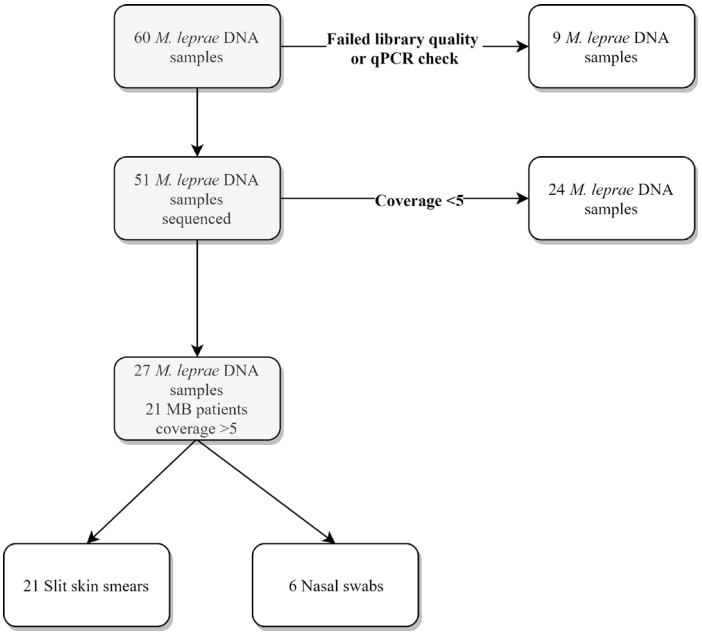


Figure S1. Samples analysed by whole genome sequencing. Number of DNA samples isolated from slit skin smears (SSS) or nasal swabs (NS) analysed by whole genome sequencing, samples that failed quality checks and samples with a query coverage higher or lower than 5 for *Mycobacterium leprae*. Origin of DNA samples sequenced with a coverage>5 is shown. All samples sequenced were collected from multibacillary (MB) patients. For all samples obtained from NS a sample of the same subject from SSS was also successfully sequenced.

3

Detection of *Mycobacterium leprae* DNA in soil: multiple needles in the haystack

Maria Tió-Coma, Thomas Wijnands, Louise Pierneef, Anna Katarina Schilling, Korshed Alam, Johan Chandra Roy, William R. Faber, Henk Menke, Toine Pieters, Karen Stevenson, Jan Hendrik Richardus and Annemieke Geluk

Sci Rep. 2019 Feb 28;9(1):3165

Abstract

Background: Leprosy is an infectious disease caused by *Mycobacterium leprae* affecting the skin and nerves. Despite decades of availability of adequate treatment, transmission is unabated and transmission routes are not completely understood. Despite the general assumption that untreated *M. leprae* infected humans represent the major source of transmission, scarce reports indicate that environmental sources could also play a role as a reservoir.

Methodology: We investigated whether *M. leprae* DNA is present in soil of regions where leprosy is endemic or areas with possible animal reservoirs (armadillos and red squirrels). Soil samples (n=73) were collected in Bangladesh, Suriname and the British Isles. Presence of *M. leprae* DNA was determined by RLEP PCR and genotypes were further identified by Sanger sequencing.

Results: *M. leprae* DNA was identified in 16.0% of soil from houses of leprosy patients (Bangladesh), in 10.7% from armadillos' holes (Suriname) and in 5% from the habitat of lepromatous red squirrels (British Isles). Genotype 1 was found in Bangladesh whilst in Suriname the genotype was 1 or 2.

Conclusions: *M. leprae* DNA can be detected in soil near human and animal sources, suggesting that environmental sources represent (temporary) reservoirs for *M. leprae*.

Introduction

Leprosy is a debilitating infectious disease caused by *Mycobacterium leprae* and *Mycobacterium lepromatosis* that is still considered a major threat in developing countries by WHO, remaining persistently endemic in regions in Africa, South America and Asia. Every year more than 200,000 new patients are still diagnosed and this new case detection rate has been virtually stable over the last decade (1). These facts indicate that multidrug therapy (MDT), although effective to treat leprosy, is insufficient to prevent transmission (2).

Granting *M. leprae* transmission is not completely understood, risk factors for development of leprosy have been identified including close contact with untreated, multibacillary patients (3), human susceptibility genes (4, 5), infection with soil transmitted helminths (6), as well as food shortage (7).

The mechanism by which bacteria are transmitted from one organism to another has not been unequivocally demonstrated (8). However, based on existing evidence, skin-to-skin contact, aerosols as well as shedding of bacteria into the environment subsequently followed by infection of other individuals remain the most obvious options for human leprosy (8, 9). Still these routes provide no explanation for the occurrence of leprosy in individuals without known contact to leprosy patients or in areas without any reported new cases (9, 10).

Through PCR amplification of *M. leprae* DNA, its presence has been detected in environmental samples such as soil (11-16) and water (17, 18) in areas inhabited by leprosy patients in Brazil and India. The viability of *M. leprae* was assessed by its multiplication in footpads of wild type mice and showed that *M. leprae* can remain alive in wet soil for 46 days (19). Moreover, viability of *M. leprae* bacilli in soil from India has been studied by 16S ribosomal RNA gene analysis (20). This study showed that 25% of the soil samples collected from patients' areas contained *M. leprae* 16S ribosomal RNA, suggesting the presence of viable *M. leprae* in the soil. Additionally, if environment–free living amoebic cysts cultured in the laboratory are artificially infected with *M. leprae* (bacilli:amoebae ratio of 5–10:1), the bacteria can survive up to 8 months (21).

Recently, *M. leprae* and *M. lepromatosis* were identified in red squirrels from the British Isles causing lepromatous disease in several animals (22, 23). Phylogenetic analyses determined that the *M. leprae* strain in squirrels (3I) was related to the lineage circulating in Medieval England, suggesting the red squirrels as a contemporary reservoir of the bacilli.

Zoonotic transmission of *M. leprae* from armadillos has been detected in the southeastern United States where wild armadillos and patients were infected with the same genotype (3I-2-v1) (24).

Furthermore, although the prevalence of leprosy in nonhuman primates (NHP) seems to

be quite low, *M. leprae* infections have also been reported in NHP (25) carrying *M. leprae* strains closely related to the human strains, suggesting that NHPs transmission can occur from human (or human sources like trash), but also among NHPs.

In this study, we aimed to explore whether besides humans and animals, environmental sources may function as a reservoir of *M. leprae*. For this purpose, we investigated the presence of *M. leprae* DNA in soil from regions with varying human leprosy endemicity in Bangladesh, Suriname, Brownsea Island and the Isle of Arran (22).

Materials and methods

DNA extraction from soil

Moist soil samples from 3 regions (Supplementary Table 1) were collected at a depth of 2 cm (Bangladesh and Suriname) or 8 cm (British Isles) in areas without sun light and stored in 50 ml tubes (Greiner Bio-One, Kremsmünster, Austria): i) in Bangladesh in front of the bedroom (right on the doorstep) in the houses of leprosy patients (n=25) and from areas without known leprosy patients (n=2); ii) in Suriname (Batavia and Groot Chatillon (former leprosy colonies), Pikin Slee and Gujaba) from areas known to be inhabited by nine-banded armadillos (n=28) (samples Suriname 2, 3 and 6 from Batavia and Groot Chatillon were previously described (van Dissel et al. submitted) and are presented here for reference purposes); iii) in the British Isles in the habitat of Eurasian red squirrels carrying *M. leprae* (Brownsea Island, n=10) and *M. lepromatosis* (Isle of Arran, n=10).

As a negative control soil was obtained from the surroundings of the Leiden University Medical Centre (The Netherlands) and spiked with 10^8 cells of *M. leprae* NHPD-63 as positive control.

DNA was extracted from 10 g of soil using DNeasy PowerMax Soil (Qiagen, Valencia, CA) as per manufacturer's instructions.

PCR amplification of RLEP and LPM244

To detect the presence of *M. leprae* DNA in soil, a PCR amplifying an *M. leprae*-specific repetitive sequence (RLEP) was performed. PCR amplification of a 129 bp sequence of RLEP (26) was carried by addition of 10 µl 5x Gotaq® Flexi buffer (Promega, Madison, WI), 5 µl $MgCl_2$ (25 mM), 2 µl dNTP mix (5 mM), 0.25 µl Gotaq® G2 Flexi DNA Polymerase (5 u/µl), 5 µl (2 µM) forward and reverse primers (Supplementary Table 2) and 5 µl template DNA in a final volume of 50 µl. DNA from *M. bovis* BCG P3 and *M. tuberculosis* H37Rv were used to assess PCR-specificity. As PCR positive controls DNA from *M. leprae* Br4923 and Thai-53 were used.

To detect inhibition of PCR due to remaining soil components, 1 µl of *M. leprae* DNA was added to the aforementioned PCR mixes together with 5 µl template DNA. In samples presenting PCR inhibition, 5 µl (2mM) Bovine Serum Albumin (BSA) Fraction V (Roche Di-

agnostics, Indianapolis, IN) were added to the PCR mixes.

PCR mixes were denatured for 2 min at 95 ºC followed by 40 cycles of 30 s at 95 ºC, 30 s at 65ºC and 30 s at 72 ºC and a final extension of 10 min at 72 ºC. PCR products (15µl) were used for electrophoresis in a 3.5% agarose gel at 130V. Amplified DNA was visualized by Midori Green Advance staining (Nippon Genetics Europe, Dueren, Germany) using a Gel Doc System (Bio-Rad Laboratories, Hercules, CA).

PCR to detect *M. lepromatosis* was performed for soil from the British Isles. The primers (LPM244) amplify a 244 bp region of the *hemN* gene not present in *M. leprae* or other mycobacteria (27). PCR was performed as explained above with LMP244 primers (Supplementary Table 2) and an annealing temperature of 53 ºC. *M. lepromatosis* DNA was used as a positive control.

Genotyping

To determine the genotype (1, 2, 3 or 4) of *M. leprae,* SNP-14676 (locus 1), SNP-1642875 (locus 2) and SNP-2935685 (locus 3) were amplified and sequenced as described (28) with minor modifications: PCRs were performed with 5 µl of template DNA using the aforementioned PCR mixes and forward and reverse primers for loci 1-3 (Supplementary Table 2) in a final volume of 50 µl. DNA was denatured for 2 minutes at 95ºC, following 45 cycles of 30 s at 95ºC, 30 s at 58 ºC and 30 s at 72 ºC and a final extension cycle of 10 min at 72ºC. PCR products were resolved by agarose gel electrophoresis as explained above. PCR products showing a band were purified prior to sequencing using the Wizard SV Gel and PCR Clean-Up System (Promega, Madison, WI). Sequencing was performed on the ABI3730xl system (Applied Biosystems, Foster City, CA) using the BigDye Terminator Cycle Sequencing Kit (Thermo Fisher Scientific, Waltham, MA).

Results

Detection of *M. leprae* DNA in soil

To determine whether *M. leprae* DNA is present in the environment surrounding leprosy patients, the habitat of armadillos and red squirrels with leprosy-like disease, soil was collected in each area. PCR amplification of a 129 bp sequence of the RLEP region from *M. leprae* was performed in a total of 75 soil samples from 3 different regions (Supplementary Table 1). Control soil samples did not show amplification of the fragment in RLEP PCR, whereas the same sample spiked with *M. leprae* bacilli presented a clear band confirming the applicability of the method to isolate, purify and detect *M. leprae* in soil. PCR amplification of 5 µl of *M. bovis* BCG P3 and *M. tuberculosis* H37Rv DNA did not show amplification of RLEP showing specificity of the PCR for *M. leprae* DNA.

In Bangladesh, 4 out of 25 collected samples were positive for RLEP PCR (Figure 1, Table 1; Supplementary Table 3), all of which were collected in houses of leprosy patients with

high bacillary load (BI 5-6, Figure 2). *M. leprae* DNA was not detected in the two soil samples from areas in Bangladesh without any reported leprosy cases (Supplementary Figure 1).

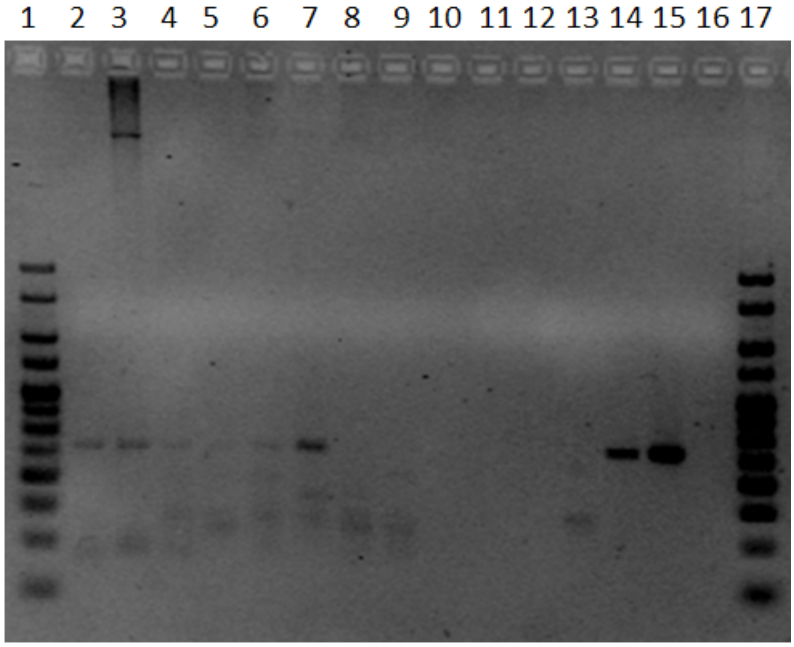


Figure 1. **Gel of PCR for RLEP region to detect presence of *M. leprae* in soil samples.** PCR products were electrophoresed in a 3.5% agarose gel. The size of the amplified RLEP sequence is 129 bp. Lanes 2 to 4 represent soil samples collected in Suriname (Suriname 2, 3, and 6), lanes 5 to 14 are soil samples collected in Bangladesh (01/65959/00, 01/65922/00, 01/65958/00, 02/65971/00, 02/22705/00, 01/65945/00, 01/65942/00, 01/65975/00, 01/22711/00 and 01/22723/00), lane 15 is DNA of *M. leprae* Thai-53 strain, lane 16 is a negative PCR control and lanes 1 and 17 are 25 bp HyperLadder (Bioline, Taunton, MA).

Table 1. RLEP PCR results for *M. leprae* DNA derived from soil samples.

	Positive		Negative	
Origin	**Number**	**%**	**Number**	**%**
Bangladesh	4	16.0	21	84.0
Suriname	3	10.7	25	89.3
Brownsea Island	1	10.0	9	90.0
Isle of Arran	0	0.0	10	100.0

RLEP PCR result to detect *M. leprae* DNA in soil samples from Bangladesh, Suriname, Brownsea Island and Isle of Arran. A positive result is determined by a visible band of 129 bp in an agarose gel.

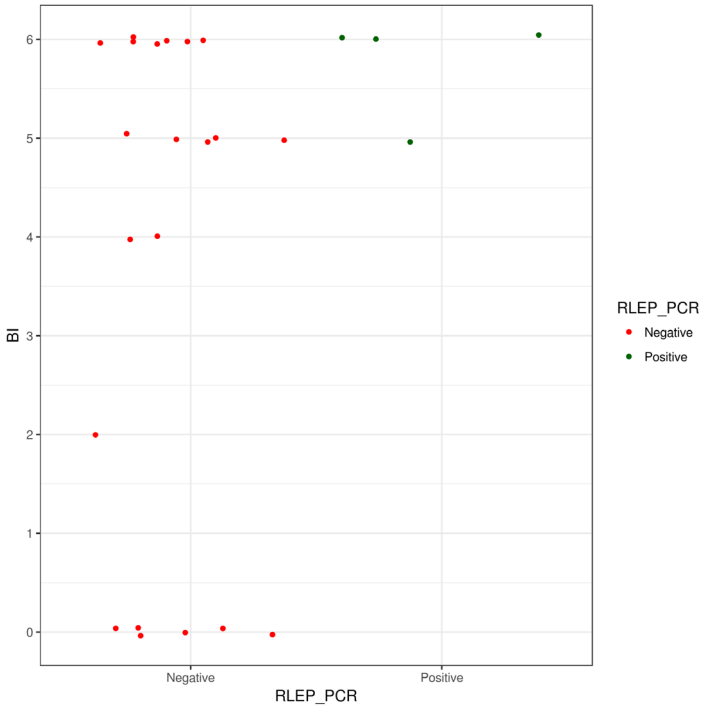


Figure 2. **RLEP PCR positivity in soil samples from Bangladesh and bacillary load (BI) of patient.** Soil samples collected in Bangladesh are represented in the graph by dots and sorted based on RLEP PCR results and bacillary load of the patient living in the household where the soil was collected.

In Suriname, samples (n=28) were taken in three different locations inhabited by armadillos and *M. leprae* DNA was detected in 3 samples obtained at former leprosy colonies in Batavia and Groot-Chatillon (Figure 1, Table 1; Supplementary Table 4).

Since all PCRs performed with UK samples were negative, we investigated whether PCRs were inhibited by compounds in the soil. DNA of *M. leprae* was added to the PCR mixes containing the DNA isolated from all soil samples and inhibition of PCR was determined by a negative PCR result. Inhibition was observed in 7 of the 10 soil samples from Brownsea Island, 8 out of the 10 from the Isle of Arran and 1 out of the 28 from Suriname. Since humic acid in soil can act as a PCR inhibitor (29, 30), 5 µl of 2 mM BSA was added to the PCRs with soil samples from the British Isles to overcome inhibition. Indeed, addition of BSA to soil-DNA spiked with *M. leprae* DNA (Br4923 or Thai-53), resulted in PCR-positivity for all spiked samples, indicating that BSA can prevent PCR inhibition due to undetermined soil compounds (data not shown).

Ten soil samples were collected in the surroundings of the infected red squirrels one of which was RLEP PCR positive (Table 1 and 2). To determine whether *M. lepromatosis* DNA

was also present in soil from the Isle of Arran with reported *M. lepromatosis* infection in red squirrels, PCRs were performed amplifying a 244 bp region of the *hemN* gene unique of *M. lepromatosis* (27). None of the 10 soil samples collected resulted in PCR-positivity using LPM244 primers.

Table 2. **SNP typing results**.

	Locus 1	Locus 2	Locus 3	Genotype
Tamil Nadu (reference strain)	C	G	A	1
Br4923 (reference strain)	T	T	C	4
Suriname 2	UD	UD	A	1 or 2
Suriname 3	UD	UD	A	1 or 2
Suriname 6	C	UD	A	1 or 2
Bangladesh 01/65922/00	UD	G	UD	1
Bangladesh 01/65958/00	UD	G	UD	1
Bangladesh 01/22723/00	C	G	A	1

Polymorphic sites in the genome of *M. leprae*: locus 1 (SNP-14676), locus 2 (SNP-1642875) and locus 3 (SNP-2935685) and the corresponding genotype. Nucleic acid corresponding to each polymorphic site of *M. leprae* reference strains Tamil Nadu and Br4923 and soil samples that were successfully sequenced. When PCR amplification or sequencing of the locus was not successful it is marked as undetermined (UD).

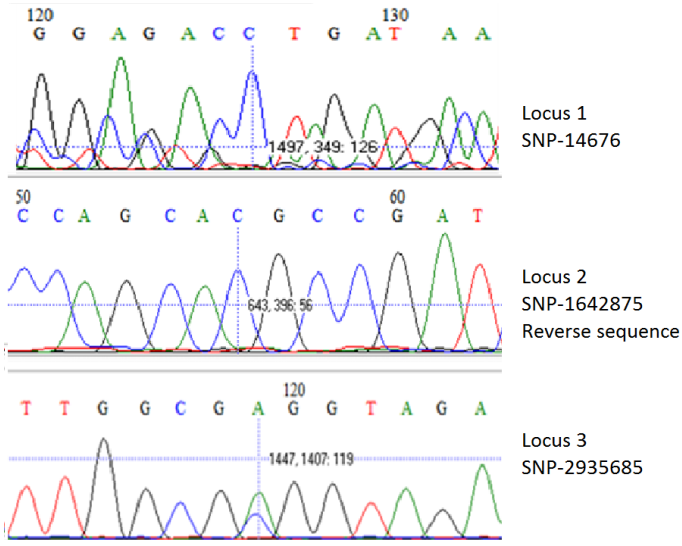


Figure 3. SNP analysis of loci 1, 2 and 3 from a representative *M. leprae* positive soil sample collected in Bangladesh. Sequencing results of locus 1 (SNP-146763) top, reverse sequence of locus 2 (SNP-1642875) middle and locus 3 (SNP-2935685) bottom, from soil sample Bangladesh 01/22723/00 used to determine the genotype of the *M. leprae* strain identified (genotype 1). SNP positions are based on the *M. leprae* TN strain. Vertical bars indicate the polymorphic base.

Next, for all RLEP PCR positive samples from Bangladesh (n=4), Suriname (n=3) and the British Isles (n=1) the PCR-amplified 129 bp RLEP region was sequenced. Sequence alignment with the RLEP region of *M. leprae* was found for all 8 samples, confirming that *M. leprae* specific DNA can be identified in soil using the above-described procedure.

Genotyping

Genotypes of the RLEP PCR positive soil samples (n=8) were investigated and determined according to the combination of SNPs in loci 1-3 as described by Monot *et al*. (28) RLEP-positive soil from Bangladesh were classified as genotype 1 (Table 2) according to the polymorphism in locus 2 or loci 1-3 (01/22723/00, Figure 3). For the soil from Suriname the genotype was narrowed down to either 1 or 2 since only sequencing of locus 3 (Suriname 2, 3 and 6) and locus 1 (Suriname 6) were identified. For the RLEP positive sample from Brownsea Island it was not possible to obtain sequence information for any of the polymorphic loci to assign a genotype. This was most likely due to the small amount of *M.leprae* DNA in the samples.

Discussion

Human leprosy still poses a considerable health threat in developing countries where transmission is generally assumed to take place via aerosol droplets from nasal cavities of untreated *M. leprae* infected individuals to their close contacts (8, 9). However, nonhuman animal and environmental sources have also been suggested to play a role in the pathogen's dissemination (9). As paleopathological evidence of leprosy in pre-Columbian America is lacking, leprosy was very likely introduced to the continent by European colonists or the African slave route (28) also resulting in transmission to armadillos. However, nowadays infected armadillos may even be responsible for new cases in human individuals who have never had contact with leprosy patients nor travelled to leprosy endemic areas (10, 31). In addition, another living host that could potentially represent an environmental reservoir for *M. leprae* are amoebae as it has been shown that *M. leprae* can survive in free living amoebae (21). Thus, amoebas or other protists might represent an intermediate host which would allow indirect infection with *M. leprae* through environmental samples.

In this study, *M. leprae* DNA was identified in soil collected in the houses of leprosy patients and the habitats of armadillos and red squirrels, suggesting that soil may represent a (temporary) reservoir. However, this study did not asses viability of the bacteria and since *M. leprae* is an obligate intracellular pathogen further investigation is needed to elucidate the role of the environment in *M. leprae* transmission.

Understanding how *M. leprae* is transmitted, and identifying sources of infection is crucial to prevent new cases and thus blocking transmission is essential to ultimately eradicate leprosy.

Although human leprosy was eradicated from the British Isles centuries ago, Eurasian red squirrels have remained a reservoir for *M. leprae*, containing a strain closely related to the strain present in Medieval England (31). This indicates that *M. leprae* may have persisted in the environment after the human reservoir disappeared. However, *M. leprae* DNA was not abundantly present in soil, suggesting that the risk of environmental contamination is low.

Because the genome of *M. lepromatosis* contains only one copy of the *hemN* gene (32) detected by LPM244 whereas 37 copies (33) are present in the RLEP region (34) of *M. leprae*, an equal amount of bacteria would be less well detectable by LPM244 PCR for *M. lepromatosis* than by RLEP PCR for *M. leprae*. Added to the fact that *M. lepromatosis* prevalence in the squirrel population is low, it is therefore possible that sensitivity was not sufficient to detect *M. lepromatosis*.

In Bangladesh, *M. leprae* was only found in soil collected in the houses of patients with high BI index (Figure 2). At those locations more bacteria are shed and thus the likelihood of encountering bacteria in the soil is higher. However, a high BI index of the patient where the soil sample was collected was not necessarily associated with a positive RLEP PCR result. The higher percentage of RLEP positive soil in Bangladesh is likely due to a more targeted selection of the sample location in the houses of leprosy patients as well as the higher leprosy prevalence.

In previous phylogeographic analysis genotype 1 was identified as the predominant strain type in South Asia (35, 36) and was likely introduced to South Asia from other parts of that continent (36). The genotype found in soil samples from Bangladesh (1) is therefore in accordance with previous phylogeographic data (35).

In summary, this study demonstrates the presence of *M. leprae* DNA in soil, contributing to a OneHealth view on transmission including humans, animals and the environment. Further research is needed, however, to confirm whether *M. leprae* DNA in soil is derived from viable bacteria that can survive in smaller hosts such as helminths or amoebas. Thus, strategies aimed at prevention of transmission by administration of post-exposure prophylaxis to infected individuals should, besides human reservoirs of *M. leprae,* also consider environmental sources of (re)infection.

Acknowledgements

The authors gratefully acknowledge all patients and control participants. LUMC and TL-MI,B are part of the IDEAL (***I***nitiative for ***D***iagnostic and ***E***pidemiological ***A***ssays for ***L***eprosy) Consortium.

We thank Dr. L. Adams (Louisiana State University, LA) for providing *M. leprae* cells and stimulating discussions and Prof. Xiang-Yang Han (MD Anderson Cancer Center, TX, USA)

for providing *M. lepromatosis* DNA.

Funding statement

This study was supported by an R2STOP Research grant from Effect hope/ The Leprosy Mission Canada, The Order of Malta-Grants-for-Leprosy-Research (MALTALEP), the Q.M. Gastmann-Wichers Foundation, the Leprosy Research Initiative (LRI; ILEP#703.15.07) together with the Turing Foundation and the Principal's Career Development PhD scholarship provided by the University of Edinburgh.

Competing interests

The authors declare no competing interests.

References

1. World Health Organization. Global leprosy update, 2016: accelerating reduction of disease burden. Weekly Epidemiological Record. 2017;92(35):501-20.
2. Blok DJ, de Vlas SJ, Geluk A, Richardus JH. Minimum requirements and optimal testing strategies of a diagnostic test for leprosy as a tool towards zero transmission: A modeling study. PLOS Neglected Tropical Diseases. 2018;12(5):e0006529.
3. Moet FJ, Meima A, Oskam L, Richardus JH. Risk factors for the development of clinical leprosy among contacts, and their relevance for targeted interventions. Lepr Rev. 2004;75(4):310-26.
4. Liu H, Irwanto A, Fu X, Yu G, Yu Y, Sun Y, et al. Discovery of six new susceptibility loci and analysis of pleiotropic effects in leprosy. Nat Genet. 2015;47(3):267-71.
5. Alter A, Grant A, Abel L, Alcais A, Schurr E. Leprosy as a genetic disease. Mamm Genome. 2011;22(1-2):19-31.
6. Hagge DA, Parajuli P, Kunwar CB, Rana D, Thapa R, Neupane KD, et al. Opening a Can of Worms: Leprosy Reactions and Complicit Soil-Transmitted Helminths. EBioMedicine. 2017;23:119-24.
7. Feenstra SG, Nahar Q, Pahan D, Oskam L, Richardus JH. Recent food shortage is associated with leprosy disease in Bangladesh: a case-control study. PLOS Neglected Tropical Diseases. 2011;5(5):e1029.
8. Bratschi MW, Steinmann P, Wickenden A, Gillis TP. Current knowledge on *Mycobacterium leprae* transmission: a systematic literature review. Lepr Rev. 2015;86(2):142-55.
9. Araujo S, Freitas LO, Goulart LR, Goulart IM. Molecular evidence for the aerial route of infection of *Mycobacterium leprae* and the role of asymptomatic carriers in the persistence of leprosy. Clin Infect Dis. 2016;63(11):1412-20.
10. Bonnar PE, Cunningham NP, Boggild AK, Walsh NM, Sharma R, Davis IRC. Leprosy in Nonimmigrant Canadian Man without Travel outside North America, 2014. Emerg Infect Dis. 2018;24(1):165-6.
11. Lavania M, Katoch K, Sachan P, Dubey A, Kapoor S, Kashyap M, et al. Detection of *Mycobacterium leprae* DNA from soil samples by PCR targeting RLEP sequences. J Commun Dis. 2006;38(3):269-73.
12. Turankar RP, Lavania M, Chaitanya VS, Sengupta U, Darlong J, Darlong F, et al. Single nucleotide polymorphism-based molecular typing of *M. leprae* from multicase families of leprosy patients and their surroundings to understand the transmission of leprosy. Clin Microbiol Infect. 2014;20(3):O142-9.
13. Turankar RP, Lavania M, Singh M, Siva Sai KS, Jadhav RS. Dynamics of *Mycobacterium leprae* transmission in environmental context: deciphering the role of environment as a potential reservoir. Infect Genet Evol. 2012;12(1):121-6.
14. Turankar RP, Lavania M, Singh M, Sengupta U, Siva Sai K, Jadhav RS. Presence of viable *Mycobacterium leprae* in environmental specimens around houses of leprosy patients. Indian J Med Microbiol. 2016;34(3):315-21.
15. Turankar RP, Pandey S, Lavania M, Singh I, Nigam A, Darlong J, et al. Comparative evaluation of PCR amplification of RLEP, 16S rRNA, rpoT and Sod A gene targets for detection of *M. leprae* DNA from clinical and environmental samples. Int J Mycobacteriol. 2015;4(1):54-9.
16. Lavania M, Katoch K, Katoch VM, Gupta AK, Chauhan DS, Sharma R, et al. Detection of viable *Mycobacterium leprae* in soil samples: insights into possible sources of transmission of leprosy.

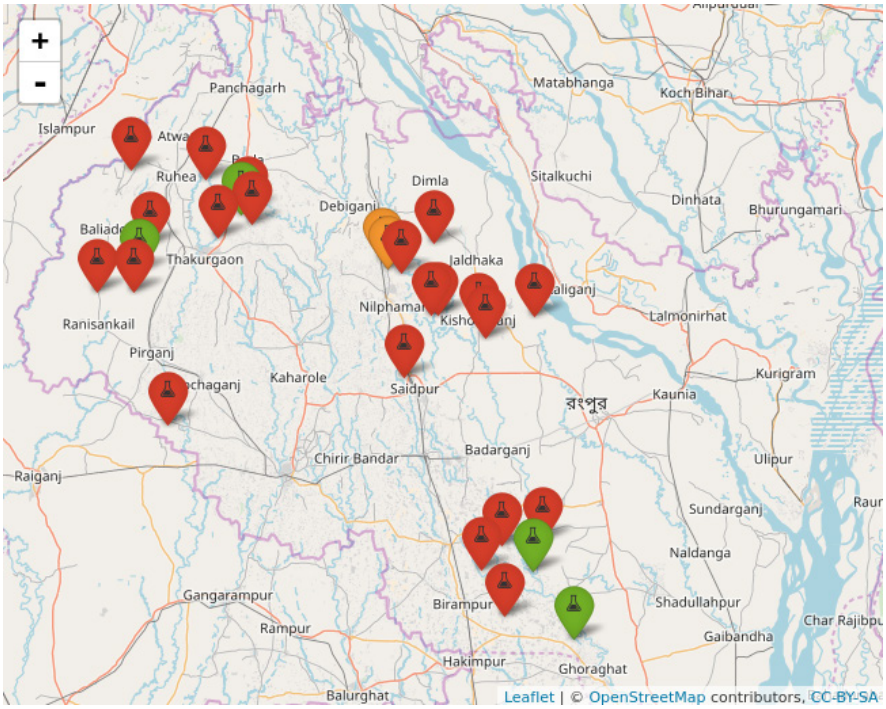
Infect Genet Evol. 2008;8(5):627-31.
17. Holanda MV, Marques LEC, Macedo MLB, Pontes MAA, Sabadia JAB, Kerr L, et al. Presence of *Mycobacterium leprae* genotype 4 in environmental waters in Northeast Brazil. Rev Soc Bras Med Trop. 2017;50(2):216-22.
18. Arraes M, Holanda MV, Lima L, Sabadia JAB, Duarte CR, Almeida RLF, et al. Natural environmental water sources in endemic regions of northeastern Brazil are potential reservoirs of viable *Mycobacterium leprae*. Mem Inst Oswaldo Cruz. 2017;112(12):805-11.
19. Desikan KV, Sreevatsa. Extended studies on the viability of *Mycobacterium leprae* outside the human body. Lepr Rev. 1995;66(4):287-95.
20. Mohanty PS, Naaz F, Katara D, Misba L, Kumar D, Dwivedi DK, et al. Viability of *Mycobacterium leprae* in the environment and its role in leprosy dissemination. Indian J Dermatol Venereol Leprol. 2016;82(1):23-7.
21. Wheat WH, Casali AL, Thomas V, Spencer JS, Lahiri R, Williams DL, et al. Long-term survival and virulence of *Mycobacterium leprae* in amoebal cysts. PLoS Neglected Tropical Diseases. 2014;8(12):e3405.
22. Avanzi C, Del-Pozo J, Benjak A, Stevenson K, Simpson VR, Busso P, et al. Red squirrels in the British Isles are infected with leprosy bacilli. Science. 2016;354(6313):744-7.
23. Simpson V, Hargreaves J, Butler H, Blackett T, Stevenson K, McLuckie J. Leprosy in red squirrels on the Isle of Wight and Brownsea Island. The Veterinary record. 2015;177(8):206-7.
24. Truman RW, Singh P, Sharma R, Busso P, Rougemont J, Paniz-Mondolfi A, et al. Probable zoonotic leprosy in the Southern United States. The New England journal of medicine. 2011;364(17):1626-33.
25. Honap TP, Pfister LA, Housman G, Mills S, Tarara RP, Suzuki K, et al. *Mycobacterium leprae* genomes from naturally infected nonhuman primates. PLOS Neglected Tropical Diseases. 2018;12(1):e0006190.
26. Donoghue HD, Holton J, Spigelman M. PCR primers that can detect low levels of *Mycobacterium leprae* DNA. J Med Microbiol. 2001;50(2):177-82.
27. Vera-Cabrera L, Escalante-Fuentes W, Ocampo-Garza SS, Ocampo-Candiani J, Molina-Torres CA, Avanzi C, et al. *Mycobacterium lepromatosis* infections in Nuevo León, Mexico. Journal of Clinical Microbiology. 2015;53(6):1945-6.
28. Monot M, Honore N, Garnier T, Araoz R, Coppee JY, Lacroix C, et al. On the origin of leprosy. Science. 2005;308(5724):1040-2.
29. Tsai YL, Olson BH. Rapid method for separation of bacterial DNA from humic substances in sediments for polymerase chain reaction. Applied and Environmental Microbiology. 1992;58(7):2292-5.
30. Watson RJ, Blackwell B. Purification and characterization of a common soil component which inhibits the polymerase chain reaction. Can J Microbiol. 2000;46(7):633-42.
31. da Silva MBP, Juliana M., Li W, Jackson MG-J, Mercedes , Belisle JT, Bouth RC, Gobbo ARB, Josafá G. , et al. Evidence of zoonotic leprosy in Pará, Brazilian Amazon, and increased anti-PGL-I titer in individuals who consume armadillos in their diet. PLOS Neglected Tropical Diseases. 2018;12(6):e0006532.
32. Singh P, Benjak A, Schuenemann VJ, Herbig A, Avanzi C, Busso P, et al. Insight into the evolution and origin of leprosy bacilli from the genome sequence of *Mycobacterium lepromatosis*. Proc Natl Acad Sci U S A. 2015;112(14):4459-64.
33. Cole ST, Supply P, Honoré N. Repetitive sequences in *Mycobacterium leprae* and their impact on genome plasticity. Leprosy review. 2001;72(4):449-61.
34. Braet S, Vandelannoote K, Meehan CJ, Brum Fontes AN, Hasker E, Rosa PS, et al. The repetitive element RLEP is a highly specific target for detection of *Mycobacterium leprae*. J Clin Microbiol. 2018;56(3).
35. Monot M, Honore N, Garnier T, Zidane N, Sherafi D, Paniz-Mondolfi A, et al. Comparative genomic and phylogeographic analysis of *Mycobacterium leprae*. Nat Genet. 2009;41(12):1282-9.
36. Benjak A, Avanzi C, Singh P, Loiseau C, Girma S, Busso P, et al. Phylogenomics and antimicrobial resistance of the leprosy bacillus *Mycobacterium leprae*. Nat Commun. 2018;9(1):352.

Supplementary Material

Supplementary Table 1. **Origin, number and location of soil samples.**

Origin	Number	Area of collection
Bangladesh	25	Houses of leprosy patients
Bangladesh	2	Area without any reported case of leprosy
Suriname	28	Surroundings of armadillos' habitats
British Isles	20	Areas frequented by red squirrels infected with *M. leprae* (Brownsea Island) or *M. lepromatosis* (Arran Isle)
The Netherlands	2	Control soil

Summary of soil collected and brief description of area.



Supplementary Figure 1. **Soil sampling sites in Bangladesh.** Red markers indicate a negative result for RLEP PCR. Green markers indicate presence of *M. leprae* determined by a positive result for RLEP PCR. Orange markers illustrate the soil collected in areas not known to be inhabited by leprosy patients. Both soil samples collected in areas not known to be inhabited by leprosy patients were negative for RLEP PCR. The figure was drawn in R (v3.4.3) with the package *leaflet* (v1.1.0.9000) using maps available under the Open Database License from © OpenStreetMap contributors (https://www.openstreetmap.org/copyright).

4

Lack of evidence for the presence of leprosy bacilli in red squirrels from North-West Europe

Maria Tió-Coma, Hein Sprong, Marja Kik, Jaap T. van Dissel, Xiang-Yang Han, Toine Pieters and Annemieke Geluk

Transbound Emerg Dis. 2020 Mar;67(2):1032-1034

Abstract

Leprosy is a human infectious disease caused by *Mycobacterium leprae* or *Mycobacterium lepromatosis* that can also occur in animals and even manifest as zoonosis. Recently, both mycobacteria were detected in red squirrels (*Sciurus vulgaris*) from the British Isles.

To further explore the presence of leprosy-bacilli in North-West Europe, we screened Belgian and Dutch squirrels. Tissue samples from 115 animals tested by qPCR were negative for both pathogens. No molecular or pathological evidence was found of the presence of these zoonotic pathogens in North-West Europe.

Introduction

Leprosy, caused by *Mycobacterium leprae* or *Mycobacterium lepromatosis*, is a debilitating disease occurring in several low- and middle-income countries. Transmission is unabated as shown by stable numbers of new cases worldwide (1). Close contact to multibacillary patients (2) and, to a lesser extent, infected animals and environmental sources are presumed to play a role in transmission (3, 4).

M. leprae and *M. lepromatosis* can cause leprosy-like disease in several animal hosts, including nine-banded armadillos (4), red squirrels (5) and nonhuman primates (6).

Previously, *M. leprae* and *M. lepromatosis* have been detected in Eurasian red squirrels (*Sciurus vulgaris*) from the British Isles (5, 7), where human leprosy has not occurred for centuries. Squirrels of other European countries and Mexico, however, were not positive for DNA of these pathogens when screened by PCR (8).

To further explore the presence of *M. leprae* or *M. lepromatosis* in continental squirrels from North-West Europe we screened squirrels from the Netherlands and Belgium.

Materials and methods

Sample collection

Sixty-one wild red squirrels (*Sciurus vulgaris*) and one Japanese squirrel (*Sciurus lis*) were found dead in the Netherlands and submitted to the Dutch Wildlife Health Centre (Figure S1). Animals died due to infection with *Toxoplasma gondii* (n=21), traumatic injuries (n=23), or other pathologies (n=18). Skin lesions consistent with leprosy were not detected. Red squirrels (n=53) victims of road traffic were collected between 2010 to 2014 in Flanders (Table S1), Belgium (9).

Necropsy included macroscopic examination, cytological analysis of liver, spleen, lungs, and intestinal contents stained with HemacolorR (Hemacolor quick stain, Merck, Darmstadt, Germany), and histological examination.

Small biopsies from liver and spleen were collected in 2 ml tubes and stored at -20 °C until further processing.

DNA extraction

Biopsies from spleen [Dutch squirrels (n=62); Belgian squirrels (n=53)] and liver [Belgian squirrels only (n=53)] were used for DNA extraction.

DNA extraction was performed on 20 mg tissue using the Qiagen DNeasy Blood and Tissue Kit (Qiagen, Venlo, the Netherlands), according to manufacturer's protocol.

RLEP and 202 qPCR

A qPCR amplifying an *M. leprae*-specific repetitive sequence (RLEP) was performed (10). For *M. lepromatosis* primers and probe (202-qPCR) were designed based on criteria for

TaqMan PCR reactions to amplify a 168bp fragment in contig 202 which has two copies in the genome and is specific of *M. lepromatosis*. Amplification was carried in a final volume of 25 µl by addition of GoTaq™ Probe qPCR Master Mix (Promega, Madison, WI), 22.5 µM primers (Table 1), 6.25 µM TaqMan probe (Table 1) and 5.0 µl sample using the following profile: 2 min at 95ºC, 40 cycles of 15 s at 95ºC and 1 min at 60ºC. Nuclease-free water was used as negative control and *M. leprae* or *M. lepromatosis* DNA as positive controls. DNA of *M. leprae* Br4923 and Thai53, *M. bovis* BCG P3 and *M. tuberculosis* H37DNA were used to assess 202qPCR specificity. A Ct of 38.5 was taken as the limit for positivity.

Statistical analysis was performed in R (version 3.4.1).

Table 1. ***Primers and probes for RLEP and 202 qPCR.***

Primer/probe	Sequence 5‘ - 3‘
RLEP qPCR F	GCAGCAGTATCGTGTTAGTGAA
RLEP qPCR R	CGCTAGAAGGTTGCCGTAT
RLEP qPCR Probe FAM	CGCCGACGGCCGGATCATCGA
qPCR 202 F	CTGATCGCACACCTTGATGAGAG
qPCR 202 R	GTTAGGTTGATCGACATCTTCGGTGC
qPCR 202 Probe VIC	CACCACTAGCGCACCACGTCAGACAGGC

Forward (F) primers, reverse (R) primers and probes with dyes used to detect presence of *M. leprae* (RLEP) and *M. lepromatosis* (202) by TaqMan qPCR.

Results

To study whether leprosy-bacilli are present in Dutch and Belgian squirrels, we performed qPCR analysis on 115 red squirrels none of which showed clinical signs of leprosy. DNA samples were negative for both RLEP-qPCR (*M. leprae*) and 202-qPCR (*M. lepromatosis*). Thus, *M. leprae*- and *M. lepromatosis*-specific DNA was not detected in the Dutch or Belgian red squirrel populations with a 95% confidence interval (CI) of 0.0 to 4.1%. Part of these samples have been tested successfully for the presence of tick-borne pathogens by qPCR (9), and 45% were positive for the presence of Bartonella DNA (11).

Discussion

Since the discovery of Eurasian red squirrels suffering from leprosy-like disease in the British Isles (5) there has been an increased interest to screen other squirrels for *M. leprae* and *M. lepromatosis*. We examined 114 animals from the red squirrel population in the Netherlands and Belgium.

In Brownsea Island *M. leprae* DNA was found in 25 out of 25 red squirrels (8 with leprosy-like lesions) (5). In Scotland and Ireland *M. leprae* was not detected, but *M. lepromatosis*

was identified. The authors calculated that 21% of the squirrel population without clinical signs and all squirrels with clinical signs from the British Isles carried either *M. leprae* or *M. lepromatosis*.

As recently described (8), in a squirrel population showing no disease manifestations and with the same prevalence of leprosy as the British Isles (21%), a sample size of 19 should suffice to identify minimally one case with a 95% CI. Since our sample size (53 Belgian squirrels, 62 Dutch squirrels) was 2-3-fold larger, we should have identified leprosy-bacilli in the Dutch and Belgian squirrel population, if present, in at least the same prevalence as in the British Isles. When taking into account also previously tested continental European Eurasian red squirrels (n=96, (8)), the prevalence in mainland Europe is less than 2.2% with a 95% CI.

Our findings are consistent with previous observation in other European countries (France, Italy, United Kingdom, Germany and Switzerland) and Mexico (8) where *M. leprae*- or *M. lepromatosis*-specific DNA were not detected.

Pinnae samples have been reported to be the optimal tissue for molecular screening. Spleen and liver have also been successfully used to detect *M. leprae* and *M. lepromatosis* DNA in squirrels (5), however, the sensitivity might differ from pinnae samples. The lack of positive PCR results could be due to a lower prevalence of the leprosy-bacilli in the Dutch and Belgian red squirrel population compared to the reference UK population. It is conceivable that, red squirrels in UK could be more susceptible due to reduced immunity as a consequence of squirrelpox transmitted by grey squirrels (12). Alternatively, the animals could carry a bacillary load below the detection limit, or our tissue sampling might have been suboptimal. However, animals with leprosy-related lesions have not been observed in these populations suggesting that the absence of the pathogen is a more-plausible scenario. As we did not find evidence of the presence of *M. leprae* or *M. lepromatosis*, our results endorse the recent hypothesis (8) that Eurasian red squirrels in the British Isles are the only known wild rodent up to date carrying the leprosy-bacilli.

Acknowledgements

We thank Ir. M.G.E. Montizaan (DWHC, Utrecht University, The Netherlands) for creating the map of the Dutch squirrel location.

Ethics Statement

The authors confirm that the ethical policies of the journal, as noted on the journal's author guidelines page, have been adhered to. No ethical approval was required as wild animals were found dead and offered for necropsy.

Funding Statement

This study was supported by an R2STOP Research grant from Effect hope/ The Leprosy Mission Canada, and the Q.M. Gastmann-Wichers Foundation (to AG).

Conflicts of Interest

The authors declare that they have no conflict of interest.

References

1. WHO. Global leprosy update, 2017: reducing the disease burden due to leprosy. Weekly epidemiological record. 2018;93:11.
2. Moet FJ, Meima A, Oskam L, Richardus JH. Risk factors for the development of clinical leprosy among contacts, and their relevance for targeted interventions. Lepr Rev. 2004;75(4):310-26.
3. Tió-Coma M, Wijnands T, Pierneef L, Schilling AK, Alam K, Roy JC, et al. Detection of Mycobacterium leprae DNA in soil: multiple needles in the haystack. Scientific Reports. 2019;9(1):3165.
4. Truman RW, Singh P, Sharma R, Busso P, Rougemont J, Paniz-Mondolfi A, et al. Probable zoonotic leprosy in the Southern United States. The New England journal of medicine. 2011;364(17):1626-33.
5. Avanzi C, Del-Pozo J, Benjak A, Stevenson K, Simpson VR, Busso P, et al. Red squirrels in the British Isles are infected with leprosy bacilli. Science. 2016;354(6313):744-7.
6. Honap TP, Pfister LA, Housman G, Mills S, Tarara RP, Suzuki K, et al. *Mycobacterium leprae* genomes from naturally infected nonhuman primates. PLOS Neglected Tropical Diseases. 2018;12(1):e0006190.
7. Schilling A-K, van Hooij A, Corstjens P, Lurz P, DelPozo J, Stevenson K, et al. Detection of humoral immunity to mycobacteria causing leprosy in Eurasian red squirrels (*Sciurus vulgaris*) using a quantitative rapid test2019.
8. Schilling A-K, Avanzi C, Ulrich RG, Busso P, Pisanu B, Ferrari N, et al. British Red Squirrels Remain the Only Known Wild Rodent Host for Leprosy Bacilli. Frontiers in Veterinary Science. 2019;6(8).
9. Ruyts S, Frazer-Mendelewska E, Van Den Berge K, Verheyen K, Sprong H. Molecular detection of tick-borne pathogens *Borrelia afzelii*, *Borrelia miyamotoi* and *Anaplasma phagocytophilum* in Eurasian red squirrels (*Sciurus vulgaris*). European journal of wildlife research. 2017;2017 v.63 no.3(-no. 3):pp. 1-43.
10. Martinez AN, Lahiri R, Pittman TL, Scollard D, Truman R, Moraes MO, et al. Molecular determination of *Mycobacterium leprae* viability by use of real-time PCR. J Clin Microbiol. 2009;47(7):2124-30.
11. von Loewenich FD, Seckert C, Dauber E, Kik MJL, de Vries A, Sprong H, et al. Prosthetic Valve Endocarditis with *Bartonella washoensis* in a Human European Patient and Its Detection in Red Squirrels (*Sciurus vulgaris*). 2019;58(1):e01404-19.
12. McGowan NE, Marks NJ, McInnes CJ, Deane D, Maule AG, Scantlebury M. Effects of parasitism and morphology on squirrelpox virus seroprevalence in grey squirrels (*Sciurus carolinensis*). PloS one. 2014;9(1):e83106-e.

Supplementary Material

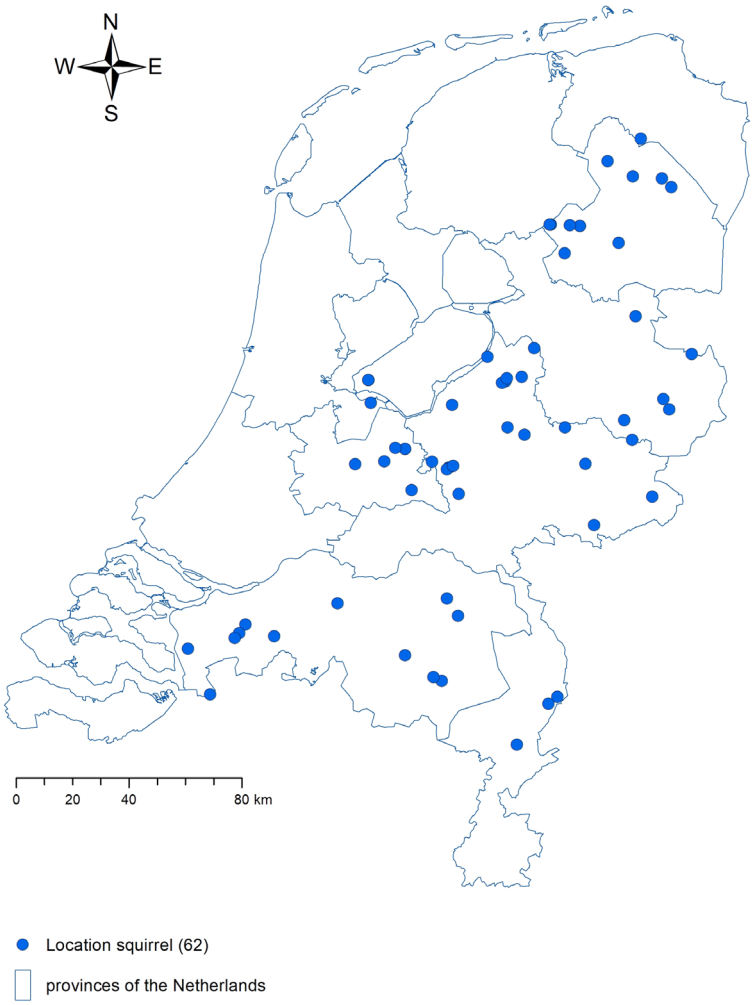


Figure S1. Distribution of necropsied and tested squirrels in the Netherlands (2015 – 2017). Locations of squirrel carcasses found in the Netherlands are indicated in blue circles. Figure designed by Ir. M.G.E. Montizaan.

5

Blood RNA signature RISK4LEP predicts leprosy years before clinical onset

Maria Tió-Coma, Szymon M. Kiełbasa, Susan J. F. van den Eeden, Hailiang Mei, Johan Chandra Roy, Jacco Wallinga, Marufa Khatun, Sontosh Soren, Abu Sufian Chowdhury, Khorshed Alam, Anouk van Hooij, Jan Hendrik Richardus and Annemieke Geluk

EbioMedicine. 2021 Jun;68:103379

Abstract

Background

Leprosy, a chronic infectious disease caused by *Mycobacterium leprae*, is often late- or misdiagnosed leading to irreversible disabilities. Blood transcriptomic biomarkers that prospectively predict those who progress to leprosy (progressors) would allow early diagnosis, better treatment outcomes and facilitate interventions aimed at stopping bacterial transmission. To identify potential risk signatures of leprosy, we collected whole blood of household contacts (HC, n=5,352) of leprosy patients, including individuals who were diagnosed with leprosy 4-61 months after sample collection.

Methods

We investigated differential gene expression (DGE) by RNA-Seq between progressors before presence of symptoms (n=40) and HC (n=40), as well as longitudinal DGE within each progressor. A prospective leprosy signature was identified using a machine learning approach (Random Forest) and validated using reverse transcription quantitative PCR (RT-qPCR).

Findings

Although no significant intra-individual longitudinal variation within leprosy progressors was identified, 1,613 genes were differentially expressed in progressors before diagnosis compared to HC. We identified a 13-gene prospective risk signature with an Area Under the Curve (AUC) of 95.2%. Validation of this RNA-Seq signature in an additional set of progressors (n=43) and HC (n=43) by RT-qPCR, resulted in a final 4-gene signature, designated RISK4LEP (*MT-ND2*, *REX1BD*, *TPGS1*, *UBC*) (AUC=86.4%).

Interpretation

This study identifies for the first time a prospective transcriptional risk signature in blood predicting development of leprosy 4 to 61 months before clinical diagnosis. Assessment of this signature in contacts of leprosy patients can function as an adjunct diagnostic tool to target implementation of interventions to restrain leprosy development.

Introduction

Leprosy, also known as Hansen's disease, is still a considerable health threat in pockets of several low- and middle-income countries worldwide. The annual number of new cases fluctuates around 200,000 people, reflecting a stable trend that has been observed during the last decade (1). Affecting the skin and peripheral nerves, leprosy presents as a spectrum including several clinical forms paralleling immunity against *Mycobacterium leprae*, the pathogen causing leprosy (2). On one pole of the immunopathological spectrum tuberculoid leprosy (TT) is situated, mainly characterized by low amount of bacteria and a cell-mediated immune response, and at the other pole lepromatous leprosy (LL) presenting high bacterial load, and a humoral response (3, 4). In between these polar forms patients present borderline leprosy (borderline tuberculoid [BT], borderline borderline [BB] and borderline lepromatous [BL]) (5).

Diagnosis still heavily relies on detection of clinical symptoms and early detection of leprosy represents a substantial hurdle in present-day leprosy health care. Besides, the reduced number of new cases has resulted in unfamiliarity of signs and symptoms of leprosy limiting suspicion and detection of leprosy. Only a small percentage (estimated 5%) of people exposed to *M. leprae* develop the disease (3). In addition, leprosy displays a long incubation period (2 to >10 years) (6, 7). These factors contribute to limited awareness of the disease among both the public and healthcare providers, hampering the early detection of new cases, and are reinforced by the strong social stigma of leprosy. Detection delay not only results in frequent delay of treatment leading to irreversible disabilities, but also contributes to perpetuating transmission.

Leprosy is a multi-factorial disease influenced by the infectious agent (dose and frequency of exposure) but also by genetics (8-14), nutritional factors (15, 16), living conditions (17, 18) and individual characteristics (age, sex) (19, 20). Household contacts (HC) of leprosy patients are at highest risk (21-24), and thus a recommendation for use of chemoprophylaxis as preventive treatment for contacts of leprosy patients was included in the WHO 2018 guidelines (25). Given the low proportions of individuals actually developing leprosy after *M. leprae* exposure, biomarkers identifying who will develop disease would be very useful to target prophylactic measures.

In the past years, several studies have searched for biomarkers to (early) detect leprosy either based on the host immune response (26-31), the pathogen (32-37), or a combination of both (38-45). Molecular detection by identification of the repetitive element RLEP by (quantitative) PCR (33, 46, 47) as well as detection of anti-*M. leprae* phenolic glycolipid I (PGL-I) IgM in blood (28, 29) are methods employed to assist leprosy diagnosis. Nevertheless, the sensitivity of these techniques to identify paucibacillary (PB) leprosy is not sufficient due to the low concentrations of bacilli in these patients (26, 44, 45). On the other

hand, PCR and anti-PGL-IgM, though useful to detect infection, are inadequate predictors of disease amongst HC of leprosy patients, as individuals remaining without disease may present positive PCR and/or PGL-I IgM (28, 32, 35, 44, 45). In addition, combinations of other host proteins (24) (27) have been shown to be useful to diagnose leprosy and detect *M. leprae* infection, but have not been studied prospectively yet.

Transcriptomic analysis of differential gene expression (DGE) represents an effective approach to identify new biomarkers for leprosy diagnosis (48). RNA-Seq, a high-throughput and unbiased technique which includes the whole transcriptome instead of a selection of genes, has been successfully used to prospectively identify correlates of risk for leprosy reversal reaction (49), as well as for tuberculosis caused by the closely related bacteria *Mycobacterium tuberculosis* (50-53).

The immune response during leprosy and leprosy reactions has also been investigated through transcriptomics (54-63). However, very few studies have employed transcriptomics to identify a biomarker risk signature for leprosy diagnosis: one study described that gene expression of *LDR* and *CCL4* in nerve biopsies identified up to 80% of pure neural (PN) leprosy patients (64). Likewise, a signature formed by four miRNA was identified using skin biopsies that could discriminate leprosy patients with 80% sensitivity and 91% specificity (54). Although these transcriptomic biomarkers show potential, both are based on samples that require invasive techniques (nerve and skin biopsies) and were applied when clinical symptoms were already visible.

In contrast to previous work, this study aimed to identify a prospective biomarker signature that can predict development of leprosy. For this purpose, whole blood samples were collected from HC who were followed up for several years and re-sampled in case they developed leprosy. Transcriptomic differences were investigated between progressors and HC who remained without leprosy. Variation in gene expression of those individuals who developed leprosy was assessed between the timepoint before leprosy diagnosis and at onset of disease. A risk signature for leprosy development can guide post-exposure prophylactic strategies to avoid disease progression, reduce disability and contribute to stop *M. leprae* transmission.

Methods

Sample collection and study design

HC (n=5,352) of newly diagnosed leprosy patients were recruited and a first blood sample was collected from April 2013 to April 2018 as part of a field trial (28, 65-67) in four districts in the northwest of Bangladesh (Nilphamari, Rangpur, Panchagarh and Thakurgaon). Patients and HC entered the study through the Rural Health Program of The Leprosy Mission International, Bangladesh, based at the Danish Bangladesh Leprosy Mission Hospital in

Nilphamari, a referral hospital specialized in the detection and treatment of leprosy. The population of the four districts, which was around 7,000,000 at the start of intake, is mainly rural, but includes six main towns. The new case detection rate and the prevalence in the study area were 1.18 and 0.9 per 10,000 correspondingly (68).

HC were defined as those living in the same house, in a house on the same compound and sharing the same kitchen, or direct neighbours (first neighbours). Exclusion criteria included previous leprosy, refused informed consent, pregnant women, tuberculosis, children younger than 5 years, liver disease or jaundice and temporary residency in the study area (67). Some HC in the study received BCG vaccination (n=657) after providing the first blood sample. Whole blood from HC was collected in PAXgene tubes at time of diagnosis of the index case (t=1) (Figure 1). All contacts were followed up annually and checked for the absence of clinical signs and symptoms of leprosy. All individuals were followed up for 36 months or longer. Follow up is still ongoing. Contacts who were clinically diagnosed with leprosy within 4-61 months after recruitment were considered progressors (n=85). A second blood sample was collected from progressors at the time of leprosy diagnosis, before start with multidrug therapy (t=2) and bacteriological index (BI) was determined. Leprosy was diagnosed by a medical officer following the Rural Health Program guidelines in accordance to the National Leprosy Control Program (69). Progressors who presented five or fewer skin lesions and BI 0 were classified as PB and those who presented more than five skin lesions were classified as MB (25).

An initial discovery set was drawn from the cohort including 40 HC and 40 progressors who were diagnosed with leprosy 4 to 60 months after recruitment. To replicate and validate the results from the discovery set, a validation set was drawn later from the same cohort which included 43 HC and 43 progressors who were diagnosed with leprosy 4 to 61 months after recruitment. Subjects who developed leprosy > 61 months after recruitment (available only during the validation analysis) were excluded (n=5). The control HC group were optimally matched to the progressors by age, sex, date of recruitment, follow up time and BCG vaccination within the study (Table 1).

Ethics Statement

This study was approved by the National Research Ethics Committee (BMRC/NREC/2016-2019/214) and followed the Helsinki Declaration (version Fortaleza, Brazil, October 2013). Participants were informed in the local language about the study objectives, the samples and their right to refuse to take part or withdraw without consequences for their treatment. All subjects gave written informed consent before enrolment and treatment was provided according to national guidelines (69).

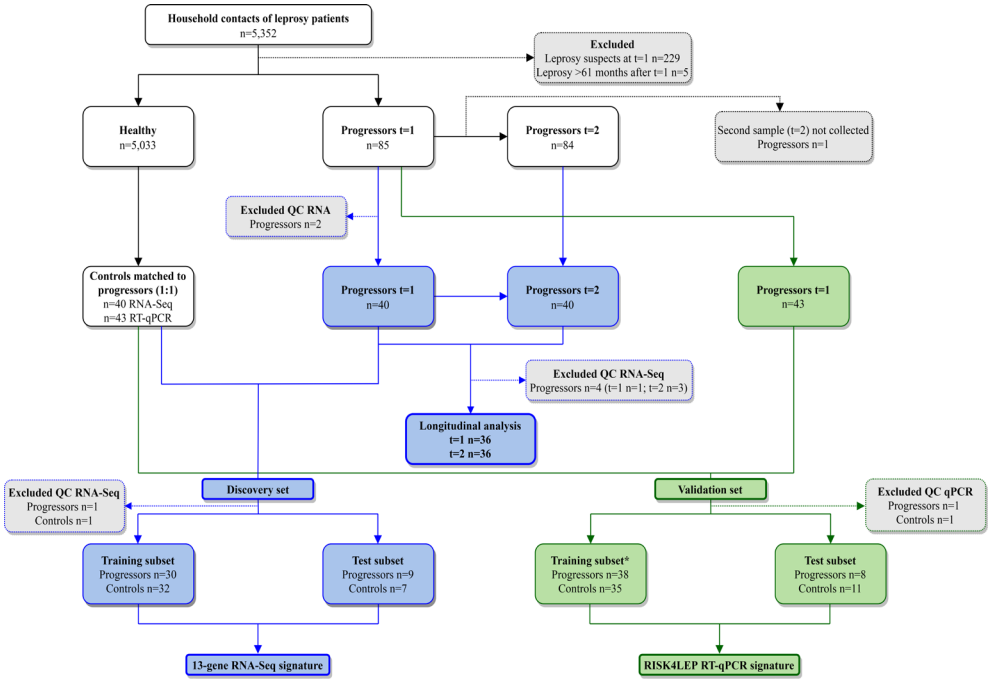


Figure 1. **Study design to identify a transcriptomic signature associated with leprosy risk.** In blue samples used in the discovery set (RNA-Seq) and in green samples used in the validation set (reverse transcription quantitative PCR (RT-qPCR)). Progressors are household contacts who developed leprosy within 4-61 months (Figure S1) after recruitment. t=1 is the timepoint before disease and t=2 is the timepoint of leprosy diagnosis. Excluded QC (quality check) RNA refers to samples that did not meet RNA quality check for RNA-Seq (RNA integrity number [RIN] ≤ 6) and were not used for RT-qPCR (validation set). Excluded QC RNA-Seq refers to samples for which RNA-Seq data did not meet the quality requirements with respect to number and distribution of reads (Figure S2). Excluded QC RT-qPCR were samples showing outlier Cycle threshold (Ct) values (>15) for the reference *GAPDH* gene (medians of two assays: 9.6 and 7.3). Training and test subsets were used in Random Forest to predict leprosy development. *RT-qPCR data of 8 samples (4 progressors and 4 HC controls) from the discovery set (RNA-Seq) were included in the training subset of the RT-qPCR Random Forest to improve the training of the model.

RNA isolation, library preparation and sequencing

Blood was collected in PAXgene tubes (BD Biosciences, Franklin Lakes, NJ) in Bangladesh and sent on dry ice to Leiden University Medical Centre (The Netherlands) for analysis. RNA isolation from PAXgene tubes was automated using a QIAcube (Qiagen, Hilden, Germany) and PAXgene Blood RNA kits (Qiagen) according to the manufacturers' protocol.

RNA concentrations were measured by Qubit RNA BR (Thermo Fisher Scientific, Waltham, MA) and integrity was determined by Fragment Analyzer (Agilent, Santa Clara, CA).

Table 1. Cohort characterization.

Discovery set, RNA-Seq (n=80)						
Group	**Subjects**	**Sex**	**Age range (n)**	**RJ Classification**	**BI**	**Time to diagnosis (n)**
Progressors	40	26 females 14 males	6-15 years (7) 16-30 years (9) 31-60 years (22) 61-70 years (2)	37 BT 1 TT 1 I 1 PN	34 BI-0 6 BI-und	4-12 months (6) 13-24 months (10) 25-36 months (10) 37-48 months (7) 49-61 months (7)
HC	40	27 females 13 males	6-15 years (7) 16-30 years (8) 31-60 years (24) 61-70 years (1)	-	-	
Validation set, RT-qPCR (n=86)						
Group	**Subjects**	**Sex**	**Age range (n)**	**RJ Classification**	**BI**	**Time to diagnosis (n)**
Progressors	43	23 females 20 males	6-15 years (12) 16-30 years (12) 31-60 years (16) 61-70 years (3)	40 BT 2 TT 1 I	35 BI-0 8 BI-und	4-12 months (7) 13-24 months (5) 25-36 months (9) 37-48 months (11) 49-61 months (11)
HC	43	23 females 20 males	6-15 years (12) 16-30 years (12) 31-60 years (16) 61-70 years (3)	-	-	

Group (leprosy progressors or household contact [HC] controls), number of individuals used for analyses, number of females and males, number of individuals in certain age range (at t=1), number of leprosy progressors according to Ridley-Jopling (RJ) classification (5), bacteriological index (BI) of progressors and time to diagnosis for progressors (time between the first sample before clinical diagnosis (t=1) and leprosy diagnosis (t=2)) are shown for the samples used in the RNA-Seq (discovery set) and the RT-qPCR (validation set) analyses. RT-qPCR: reverse transcription quantitative PCR. HC: Household contacts; BT: borderline tuberculoid leprosy; TT: tuberculoid leprosy; I: indeterminate leprosy; PN: pure neural leprosy; BI-und: bacteriological index undetermined as patient refused or was too young for skin slit smear and PB leprosy was diagnosed according to the number of lesions.

Samples that passed the quality check (RNA integrity number [RIN] ≥ 6) were considered for RNA-Seq. RNA-Seq was performed by GenomeScan (Leiden, The Netherlands): libraries were prepared using NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (New England Biolabs, Ipswich, MA) including poly(A) enrichment. Additionally, globin reduction was performed using GLOBINclear kit (Thermo Fisher Scientific). Briefly, mRNA was isolated from total RNA using the oligo-dT magnetic beads. After fragmentation of the mRNA cDNA was synthesized. This was used for ligation with the sequencing adapters and PCR amplification of the resulting product. The quality and yield after sample preparation was measured by Fragment Analyzer. The size of the resulting products was consistent with the expected size distribution.

Clustering and sequencing were performed in a NovaSeq6000 System (Illumina, San Diego, CA) with a 2*150bp paired-end protocol in one single batch to avoid a batch effect. A concentration of 1.1 nM of DNA was used.

Gene expression quantitative PCR

Reverse transcription quantitative PCR (RT-qPCR) was performed using Biomark HD system (Fluidigm, South San Francisco, CA). Reverse Transcription Master Mix (Fluidigm) was used to convert 40 ng of RNA into cDNA following manufacturer's instructions. Prior to real-time amplification with the 48.48 Dynamic Array™ integrated fluidic circuit (IFC), a pre-amplification of 14 cycles was performed using Preamp Master Mix and TaqMan Assays (Table S1) according to manufacturer's instructions. Data was analysed using the software Real-Time PCR Analysis (v 4.5.2, Fluidigm).

RNA sequencing analysis

RNA-Seq files were processed using the opensource BIOWDL RNA-Seq pipeline v2.0 (https://github.com/biowdl/RNA-seq/tree/v2.0.0) developed at Leiden University Medical Centre. This pipeline performs FASTQ pre-processing (including quality control, quality trimming, and adapter clipping), RNA-Seq alignment and read quantification. FastQC was used for checking raw read QC. Adapter clipping was performed using Cutadapt (v2.4) with default settings. RNA-Seq reads' alignment was performed using HISAT2 (v2.1.0) on GRCh38 reference genome analysis set. The gene read quantification was performed using HTSeq-count (v0.9.1) with setting "–stranded reverse". The gene annotation used for quantification was Ensembl version 94. DGE and read normalization is explained in the next section, statistics.

Functional analyses were performed using ClueGO plugin (70) to identify Gene Ontology (GO) terms and Ingenuity Pathway Analysis (IPA, Qiagen, Hildern, Germany) to establish canonical pathways.

Statistical analysis

Using RNA-Seq data, we performed DGE analysis to identify genes significantly differentially expressed between leprosy progressors at t=1 (n=40) and HC (n=40) and between progressors at t=1 and t=2 (n =40). We used an established R package, edgeR (71), executed according to their guidelines, and using raw counts normalized for library sizes with the Trimmed Mean of the M-values (TMM) method. The first comparison (t=1 vs HC) was evaluated in an unpaired design whereas the second comparison (t=1 vs t=2) was evaluated in a paired design because the samples were composed of 2 time points from the same individuals. We also modelled the difference of gene expression between the time points (t=1 vs t=2) as a linear function of the number of months which elapsed between the time points. Genes with false discovery rates below 0.05 (adjusted p-values <0.05) were classified as differentially expressed.

DGE of RT-qPCR data was measured using Mann-Whitney U test using the package stats (version 3.6.3) in RStudio (version 1.2.5033) and genes with p-value below 0.05 were considered significantly expressed.

Machine learning to predict leprosy progression

Random Forest, a machine learning approach (72), was applied to select gene features (chi-squared method) and design a model to predict leprosy progression using gene expression data from RNA-Seq and RT-qPCR. For this purpose, the package mlr (version 2.17.1) (73) was employed in RStudio (version 1.2.5033). Accuracy, sensitivity, specificity and Area Under the Curve (AUC) were also obtained using mlr.

The sample set for the RNA-Seq model (discovery set) included leprosy progressors at t=1 (n=40) and HC controls (n=40) (Figure 1). After RNA-Seq quality check (read count and MDS plot), two samples were excluded and the rest were divided into training (80%, n=62) and test (20%, n=16) subsets. An independent sample set (validation set) with 43 progressors and 43 controls was used for the RT-qPCR model. Samples were also placed 80% in the training (n=65) and 20% in the test (n=19) subsets (two samples excluded due to RT-qPCR quality check). Eight samples (four progressors and four controls) from the discovery set were added to the training subset (total training subset=73) for the RT-qPCR model to improve training input.

Training subsets were employed to train the models using a leave-one-out cross-validation (LOOCV) approach and subsequently evaluated in the test subsets. Parameters were set to ntree 50-1,000, mtry 1-10, nodsize 10-50, 72 iterations and 5 iterations of cross-validation. For the RT-qPCR model mtry and iterations were 1-4 and 1,000 respectively.

Using RNA-Seq data, an initial feature selection was performed limiting the model to 8-20 features. After feature selection lncRNA and pseudogenes were discarded from the se-

lection set and the model was retrained and re-evaluated using the final set of features (n=13). Gene expression in TMM-normalized counts per million mapped reads (CPM) of differentially expressed genes (n=1,613) were the input for the RNA-Seq Random Forest model (discovery set). ΔCts (Cycle threshold) of differentially expressed genes (Mann-Whitney U test, n=4) were used for the RT-qPCR model (validation set). ΔCts were calculated as the difference of Ct of target gene and Ct of the reference gene, where *GAPDH* (assay ID Hs99999905_m1) was the reference gene. Initially, two assays with different primers and probes of *GAPDH* (Hs99999905_m1 and Hs02786624_g1, Table S1) were included in the RT-qPCR. Mann-Whitney U test of Ct values showed that *GAPDH* from assay Hs02786624_g1 presented significant differences between groups, whilst Ct values obtained from assay ID Hs99999905_m1 did not differ significantly between groups (Table S2). Therefore, only Ct values from *GAPDH* assay ID Hs99999905_m1 were used to calculate the ΔCt values.

Role of the funding source

Funding sources had no role in the study design, data collection, data analyses, data interpretation, writing of the report and the decision to submit the manuscript for publication.

Results

Cohort characterization

Between 2013 and 2018 HC of leprosy patients (n=5,123) without any clinical signs and symptoms of leprosy were recruited in Bangladesh (Figure 1) and whole blood was collected for RNA isolation. HC who were suspected to have leprosy at recruitment (n=229) were excluded from the study. Leprosy progressors were defined as HC who were clinically diagnosed with leprosy 4-61 months after recruitment (n=85, Figure S1).

RNA quality of samples from progressors with two timepoints present (before disease [t=1] and at time of diagnosis [t=2]) was assessed and samples from progressors which passed the RNA-Seq quality check (RIN ≥ 6) of both timepoints (n=40) were further analysed by RNA-Seq (Figure 1). An equal number of HC (n=40) who did not develop disease (controls) were matched to progressors by age, sex, time of sample recruitment, follow up time and BCG vaccination (Table 1, discovery set). A separate sample set from the same area in Bangladesh, including samples from progressors (n=43) before disease (t=1) and matched controls (n=43) were used for RT-qPCR (Table 1, validation set).

In the discovery set BT leprosy was reported in 37 of the progressors, one presented TT leprosy, one indeterminate (I) and one PN (Table 1). Similarly, the progressors in the validation set included 40 BT, two TT and one I leprosy patients.

In the discovery set, two individuals in the progressors group and one in the control group (one paired control) received BCG vaccination after samples collection at t=1 as part of a field trial (65) in Bangladesh. None of the individuals in the validation set received BCG

vaccination after sample collection.

Gene expression differences in blood can be observed between leprosy progressors and contacts up to 5 years before leprosy diagnosis

RNA-Seq gene expression data from blood of leprosy progressors (n=40) 4-61 months before diagnosis (t=1) was compared to HC who did not develop disease (n=40) (Figure 2A). Initial quality analysis of the RNA-Seq revealed a low number of on-feature unique reads for two samples (one progressor at t=1 and one control, Figure S2) which were subsequently excluded and thus 39 samples per group were considered for further analysis (Figure 1). From the total of 17,435 genes, we identified 1,613 which were significantly differentially expressed (adjusted p-value < 0.05, Figure 2B) between progressors and HC using an unpaired analysis with edgeR.

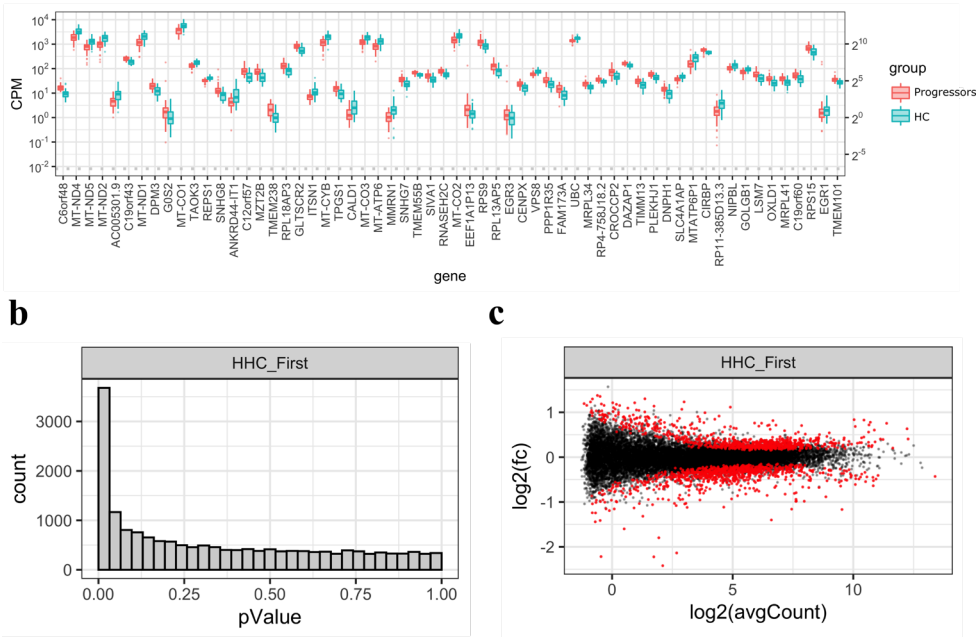


Figure 2. RNA-Seq differential gene expression analysis of leprosy progressors before clinical diagnosis and household contacts. RNA-Seq data of whole blood from leprosy progressors (n=39) 4-61 months before clinical diagnosis of leprosy (t=1 or First time point) was compared to control household contacts (HC/HHC, n=39), after exclusion of one sample per group due to low number of on-feature unique reads (Figure S2). A two-groups (unpaired samples) analysis was performed using edgeR (71) in R. a) Boxplot of Trimmed Mean of the M-values (TMM)-normalized counts per million mapped reads (CPM) per group of the most significantly differentially expressed genes. Y-axis shows CPM, expressed in power of 10 (left) or power of 2 (right). Progressors at t=1 are shown in red and HC controls in blue. b) Histogram of p-values. Number of genes (y-axis) with a determinate p-value (x-axis). c) MA plot showing log2 of fold change (FC) in gene expression (y-axis) and log2 of average CPM (x-axis) per gene. In red genes significantly differentially expressed (adjusted p-value < 0.05) and in black genes not differentially expressed. C6orf48 is also known as *SNHG32* and C19orf60 as *REX1BD*.

From these, 836 were upregulated and 777 were downregulated in leprosy progressors compared to HC (Figure 2C).

Enriched GO terms and pathways were identified in upregulated and downregulated genes. Upregulated GO terms and canonical pathways included "cotranslational protein targeting to membrane", "protein targeting to endoplasmic reticulum (ER)", "protein localization to endoplasmic reticulum", "eIF2 signalling", "mammalian target of rapamycin (mTOR) signalling", "regulation of eIF4 and p70S6K signalling" and "coronavirus pathogenesis pathway" (Table 2). Within the downregulated genes common GO terms and canonical pathways were "organelle organization", "cellular component organization", "clathrin-mediated endocytosis signalling", "integrin signalling", "Focal adhesion kinase (FAK) signalling" and "p70S6K signalling".

Gene expression in whole blood does not vary during leprosy development

To identify biomarkers indicative of disease development, we studied longitudinal variation of gene expression in leprosy progressors between 4-61 months before diagnosis and at time of diagnosis. Since the quality check for the RNA-Seq data of one sample at t=1 and three samples at t=2 failed due to low amount of aligned on-feature unique reads or the sample was an outlier (multidimensional scaling (MDS) plot) (Figure S2 and S3) these samples were excluded from the analysis with their paired sample (Figure 1). Thus, for a total of 36 progressors longitudinal comparison was feasible. Surprisingly, a paired DGE analysis showed no genes that were significantly differentially expressed (adjusted p-value < 0.05, edgeR) between timepoint of diagnosis compared to the timepoint before diagnosis (Figure S4), indicating that gene expression in blood does not vary intra-individually between the pre-clinical (no symptoms) and clinical (symptoms visible) phases of leprosy. Similarly, in a separate gene expression analysis we did not find any gene to display significant changes of expression level proportional to the number of months which elapsed between the two sample collection moments.

Machine learning identifies gene expression signature predicting leprosy

Next, a machine learning model was applied to select a subset of genes that optimize prediction of risk of leprosy development amongst HC. Random Forest was performed splitting the samples into training (80%, n=62) and test (20%, n=16) subsets, followed by a LOOCV approach and limiting the model to 8 to 20 features/genes. TMM-normalized CPM of genes differentially expressed (n=1,613) between progressors and HC in RNA-Seq of whole blood were used as input. The model (Table S3, 19-gene RNA-Seq) which included 19 genes (Table 3, Figure S5), showed a strong predictive potential for leprosy with an accuracy of 87.5% (sensitivity 100.0%, specificity 80.0%) and AUC of 96.7% (Figure 3A, Table S4). This set of genes contained protein coding genes but also long non-coding (lnc)RNA and pseudogenes.

Table 2. Functional analysis of differentially expressed genes in blood of leprosy progressors.

Upregulated genes			**Downregulated genes**		
GO terms	**adj p-value**	**% associated genes**	**GO terms**	**adj p-value**	**% associated genes**
SRP-dependent cotranslational protein targeting to membrane	1.50E-33	41.51	organelle organization	1.33E-20	5.92
cotranslational protein targeting to membrane	1.05E-32	40.00	cellular component organization	1.50E-17	5.02
protein targeting to ER	5.23E-32	37.50	regulation of cellular component organization	7.92E-15	6.31
establishment of protein localization to endoplasmic reticulum	2.87E-31	36.29	regulation of organelle organization	5.61E-14	7.49
protein localization to endoplasmic reticulum	2.22E-30	31.79	positive regulation of organelle organization	6.43E-13	9.28
Canonical pathway	**adj p-value**	**% associated genes**	**Canonical pathway**	**adj p-value**	**% associated genes**
eIF2 signalling	4.29E-28	21.90	clathrin-mediated endocytosis signalling	3.57E-08	11.40
mTOR signalling	3.04E-13	14.80	14-3-3-mediated signalling	6.72E-07	12.60
regulation of eIF4 and p70S6K signalling	1.70E-12	16.60	integrin signalling	8.44E-07	9.90
coronavirus pathogenesis pathway	1.63E-10	15.30	FAK signalling	2.92E-06	13.70
oxidative phosphorylation	1.04E-06	13.80	p70S6K signalling	4.13E-06	11.60

Top Gene Ontology (GO) terms identified by ClueGO (70) and canonical pathways identified by Ingenuity Pathway Analysis (Qiagen) from 836 upregulated and 777 downregulated genes in leprosy progressor before clinical diagnosis compared to household contacts who do not develop leprosy. P-values were adjusted for multiple testing with Bonferroni correction (adj p-value). Percentages of associated upregulated or downregulated genes from the pathway are shown.

Table 3. Gene selection using machine learning approach.

Gene name	Ensembl ID	Type of RNA
SNHG32* or *C6orf48	**ENSG00000204387**	**ncRNA, small nuclear RNA**
MT-ND4	**ENSG00000198886**	**protein coding**
MT-ND5	**ENSG00000198786**	**protein coding**
MT-ND2	**ENSG00000198763**	**protein coding**
lnc-IL17RA-36 or *AC005301.9*	ENSG00000283633	lncRNA
MT-CO1	**ENSG00000198804**	**protein coding**
TAOK3	**ENSG00000135090**	**protein coding**
REPS1	**ENSG00000135597**	**protein coding**
MT-CYB	**ENSG00000198727**	**protein coding**
TPGS1	**ENSG00000141933**	**protein coding**
MMRN1	**ENSG00000138722**	**protein coding**
UBC	**ENSG00000150991**	**protein coding**
MTATP6P1	ENSG00000248527	pseudogene
RP11-385D13.4	ENSG00000266538	lncRNA
REX1BD* or *C19orf60	**ENSG00000006015**	**protein coding**
CCDC85B	**ENSG00000175602**	**protein coding**
HCG4P12	ENSG00000225864	pseudogene
RNU6-238P	ENSG00000200183	pseudogene
AC009303.2	ENSG00000279227	lncRNA

Genes identified by Random Forest to predict leprosy progression amongst household contacts of leprosy patients. In bold genes that were included in the final RNA-Seq signature and tested by reverse transcription quantitative PCR (RT-qPCR). Underlined the genes present in the final RT-qPCR RISK4LEP signature.

To validate the signature in an independent sample set, we aimed at selecting a set of genes with commercially available probes for RT-qPCR. For this reason, the lncRNA and pseudogenes were excluded (n=6). A new model (Table S3, 13-gene RNA-Seq) was re-trained and re-evaluated in the reduced 13-gene signature (Table 3, *SNHG32/C6orf48, MT-ND4, MT-ND5, MT-ND2, MT-CO1, TAOK3, REPS1, MT-CYB, TPGS1, MMRN1, UBC, REX1BD/C19orf60, CCDC85B*) and showed an accuracy of 87.5% with a sensitivity of 88.9%, specificity of 85.7% and AUC of 95.2% (Figure 3B, Table S4). It is of note that five of these 13 genes (*MT-ND2*, *MT-ND4*, *MT-ND5*, *MT-CO1*, and *MT-CYB*) are mitochondrial genes involved in oxidative phosphorylation, and are all down-regulated in leprosy progressors.

In addition, we evaluated whether using genes from previously described tuberculosis risk signatures could also predict leprosy.

For this purpose a Random Forest was performed with genes from the Sweeney3 (*GBP5*, *DUSP3*, *KLF2*) (74), the Suliman2 (*ANKRD22*, *OSBPL10*) (51) or the RISK6 (*GBP2*, *FCGR1B*, *SERPING1*, *TUBGCP6*, *TRMT2A*, *SDR39U1*) (52) signatures as input. However, the tuberculosis risk signatures showed poor or moderate performance to predict leprosy with AUCs of 51.6%, 58.7% and 78.3% respectively (Table S4). Thus, the Sweeney3 and Suliman2 signatures resemble an algorithm that predicts leprosy randomly. The RISK6 signature, although presenting a reasonably good prediction of leprosy, showed lower performance compared to our novel 19-gene (AUC=96.7%) and 13-gene RNA-Seq (AUC=95.2%) signatures.

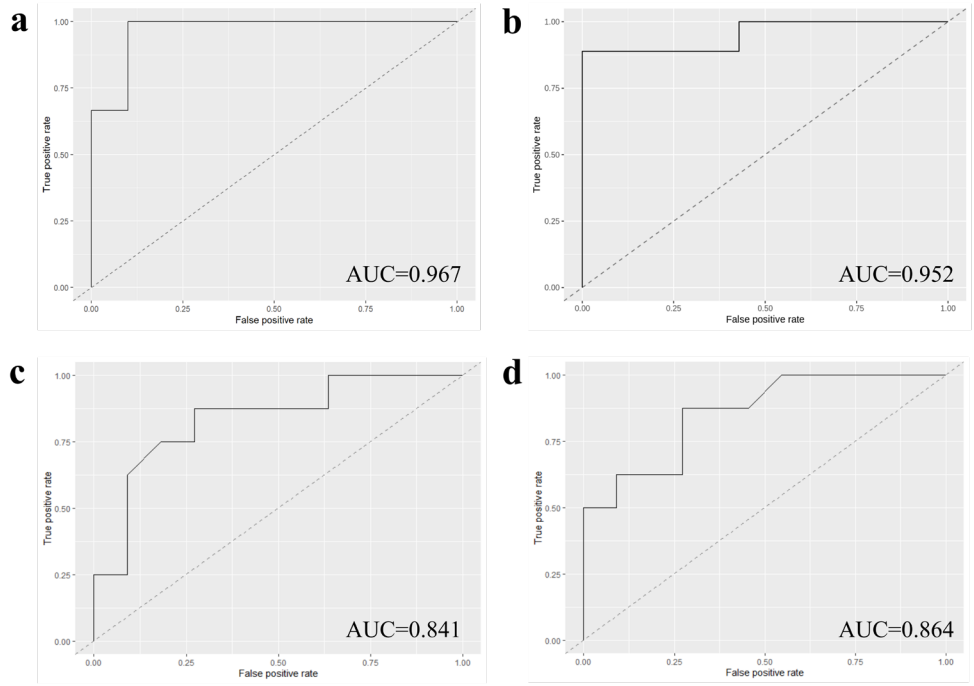


Figure 3. AUC of leprosy risk RNA-Seq and RT-qPCR signatures in blood. Area Under the Curve (AUC) of risk signatures in whole blood to prospectively predict leprosy progressors within household contacts (HC). The models were build using Random Forest, were trained with 80% of the sample sets and evaluated in 20%. a) AUC of RNA-Seq 19-gene signature where 8 to 20 features/genes were automatically selected by the model from a total of 1,613 features. b) AUC of RNA-Seq 13-gene signature based on the 19-gene signature but excluding pseudogenes and lncRNA (n=6). c) AUC of reverse transcription quantitative PCR (RT-qPCR) 13-gene signature selected in the RNA-Seq signature. d) AUC of RT-qPCR 4-gene signature RISK4LEP (final RT-qPCR signature) where only genes significantly differentially expressed in the RT-qPCR were selected.

Validation of a leprosy predictive biomarker signature

The 13-gene RNA-Seq signature was validated by RT-qPCR in an independent set of subjects. Gene expression of the 13 genes and a reference gene (*GAPDH_m1*) were tested using Biomark HD system (Fluidigm), a high-throughput RT-qPCR. Validation was performed on a separate set which included 43 leprosy progressors at t=1 and 43 HC controls as well as four progressors (at t=1) and four controls from the discovery set that were included to improve training of the model. Two outlier samples (one progressor and one control) presenting Cts of the reference gene >15 (median 7.3 *GAPDH*) were excluded from the analysis (Figure S6).

Significantly differentially expressed genes were determined using ΔCts (Ct of target gene – Ct of reference gene). Four genes, *MT-ND2*, *REX1BD*, *TPGS1* and *UBC* (Table 4), presented significant differential expression (p-values 0.0483, 0.0101, 0.0004 and 0.0060 respectively, Mann-Whitney U test) between leprosy progressors and controls (Figure 4).

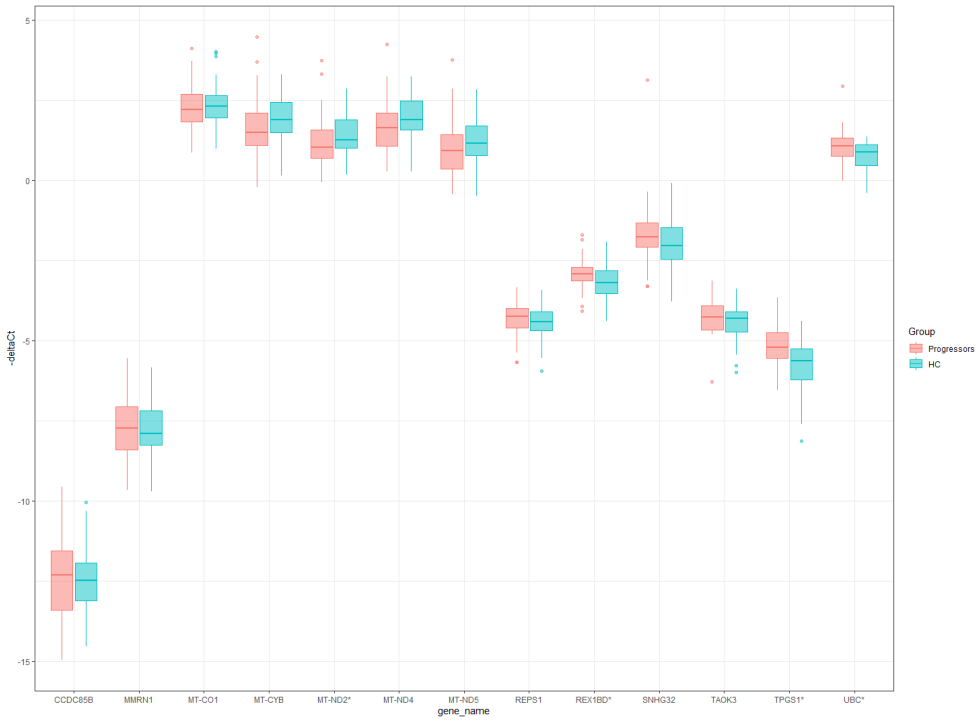


Figure 4. Boxplot showing -ΔCts of 13 genes. Boxplot of -ΔCts (-(Cycle threshold (Ct) target gene – Ct reference gene, *GAPDH*)) obtained by reverse transcription quantitative PCR (RT-qPCR) in whole blood. Genes identified in the RNA-Seq signature (n=13) are shown. Leprosy progressors before clinical diagnosis of leprosy are shown in red (t=1, n=47) and household contact (HC) controls in blue (n=47). *Genes significantly differentially expressed between the two groups using Mann-Whitney U test (*MT-ND2*, *REX1BD*, *TPGS1* and *UBC*).

Table 4. RT-qPCR ΔCts of the 13-gene signature in leprosy progressors and household contacts.

Gene	p-value	ΔCt progressors	ΔCt HC	ΔΔCt	FC	Log2FC
CCDC85B	0.840901	12.32	12.48	-0.16	1.12	0.16
MMRN1	0.654911	7.73	7.88	-0.15	1.11	0.15
MT-CO1	0.594361	-2.22	-2.31	0.09	0.94	-0.09
MT-CYB	0.054951	-1.50	-1.90	0.40	0.76	-0.40
MT-ND2	0.048303	-1.03	-1.26	0.23	0.85	-0.23
MT-ND4	0.062337	-1.64	-1.90	0.26	0.84	-0.26
MT-ND5	0.159386	-0.94	-1.17	0.24	0.85	-0.24
REPS1	0.298712	4.24	4.40	-0.16	1.12	0.16
REX1BD	0.010086	2.92	3.18	-0.27	1.20	0.27
SNHG32	0.238032	1.77	2.03	-0.26	1.20	0.26
TAOK3	0.178822	4.26	4.30	-0.04	1.03	0.04
TPGS1	0.000448	5.20	5.62	-0.42	1.34	0.42
UBC	0.005958	-1.07	-0.89	-0.18	1.13	0.18

P-values of Mann-Whitney U test of reverse transcription quantitative PCR (RT-qPCR) ΔCts (Cycle threshold (Ct) of target gene – Ct of reference gene) between leprosy progressors (n=47) and household contact (HC) controls (n=47). In bold genes significantly differentially expressed (p-value <0.05). Median of ΔCts per group, ΔΔCt (median ΔCt progressors – median ΔCt HC), Fold Change (FC, $2^{-\Delta\Delta Ct}$) for progressors and log2 of Fold Change (log2FC).

An RT-qPCR model to predict leprosy risk and validate the RNA-Seq signature was established by Random Forest using ΔCts as input. Samples from the discovery set (n=8) were only used in the training subset which included 80% of samples (n=73). Thus, the model was evaluated in a separate subset from the validation set consisting of 20% (n=19) of the sample set. A Random Forest model including the 13 genes (Table S3, 13-gene RT-qPCR) showed an AUC of 84.1% (Figure 3C, Table S4) and accuracy of 73.7% (sensitivity 87.5%, specificity 63.6%).

Slightly improved predictive potential for leprosy was observed if Random Forest was performed using only the four genes significantly differentially expressed (Table S3, 4-gene RT-qPCR RISK4LEP), showing an AUC of 86.4% (Figure 3D, Table S4), accuracy of 79.0%, sensitivity of 87.5% and specificity of 72.7%. From these four genes, *REX1BD*, *TPGS1* and *UBC* were upregulated, whilst *MT-ND2* was downregulated in leprosy progressors before clinical diagnosis compared to HC (Figure 4). This is in line with the RNA-Seq results in the discovery set, except for *UBC* (Figure 2). However, removing *UBC* from the signature (3-gene signature) and addition of the following gene with lowest p-value (*MT-CYB*) in the 3-gene signature led to decreased performances (Table S4). Addition of demographic

variables (sex, age and Ridley-Jopling (5) classification of the index leprosy contact) into the 4-gene or a reduced 2-gene (*TPGS1* and *UBC*) signature did not improve the performance either (Table S4). Therefore, the 4-gene RT-qPCR risk signature, which we named RISK4LEP, is preferred to predict leprosy development in HC due to the improved performance and the lower number of genes required (Table 3).

Discussion

Leprosy diagnosis is often ascertained after the occurrence of clinical symptoms, which may already coincide with the presence of irreversible tissue damage. Early diagnosis and prompt treatment are critical to reduce leprosy-associated disabilities and block *M. leprae* transmission. However, a sensitive diagnostic test with potential to predict the development of leprosy is not available.

To identify a transcriptomic risk signature for leprosy, this study investigated gene expression differences by RNA-Seq between HC of leprosy patients in Bangladesh who later developed leprosy and those who remained without clinical sign and symptoms. Initially a 13-gene signature that could predict leprosy development was identified using Random Forest, a machine learning approach. Subsequently, the signature was adapted and validated in a separate set by RT-qPCR. Validation of the signature in a new sample set (validation set) showed that reducing the signature to four genes improved prediction of leprosy in this sample set. The RISK4LEP signature allowed discrimination of leprosy progressors with a sensitivity of 87.5%, a specificity of 72.3% and an AUC of 86.4%. This 4-gene signature identified leprosy progressors amongst individuals exposed to leprosy bacilli from 4 to 61 months before clinical diagnosis, thus representing the first transcriptomic risk signature to prospectively predict leprosy progressors at an asymptomatic stage. This signature is unique for leprosy and does not overlap with known tuberculosis risk signatures. Since leprosy has a long incubation time and low disease prevalence, more than 5,000 samples had to be collected during 8 years to obtain samples of 85 individuals before and at disease onset. As such this is the first study of its kind in leprosy research.

The RISK4LEP predictive signature is composed by four genes: *MT-ND2*, *REX1BD*, *TPGS1* and *UBC*. *MT-ND2* encodes a subunit (core subunit 2) of the mitochondrial NADH:Ubiquinone Oxidoreductase (75). *MT-ND2* together with *MT-ND6* are the essential subunits forming the mitochondrial membrane respiratory chain NADH dehydrogenase which plays a critical role in oxidative phosphorylation. One of the functions of mitochondrial reactive oxygen species resulting from oxidative phosphorylation is to regulate immunity. *MT-ND2* is under-expressed in leprosy progressors, hence presenting a disadvantage to successful elimination of *M. leprae* (76). Little is known of *REX1BD* and *TPGS1*: *REX1BD* encodes the Required For Excision 1-B Domain Containing Protein and *TPGS1*, Tubulin Polyglutamylase Complex Subunit 1, is a gene related to microtubule binding and tubulin-glutamic acid

ligase activity that may act in the targeting of the tubulin polyglutamylase complex (75). *UBC* is one of the four genes encoding the human ubiquitin involved in several pathways such as protein degradations, DNA repair, cell cycle regulation, kinase modification, endocytosis and regulation of other cell signalling. It has been previously reported as having a high degree of connectivity in a protein-protein interaction network with differentially expressed genes in patients with active tuberculosis (77). The ubiquitin system is involved in the innate immune response in tuberculosis and has been suggested as a potential target for host-directed therapy, indicating that *UBC* might play a role in the innate response against *M. leprae* as well. Moreover, variants in the regulation regions of the *PRKN* gene (previously known as *PARK2*), which is part of the ubiquitin system, have been associated with susceptibility to leprosy (14). *PRKN* codes a ubiquitin ligase that is essential for autophagy of mycobacteria and damaged mitochondria (78, 79).

Our data show that differences in gene expression could be observed up to 61 months before the disease manifests (Figure S1). In contrast, intra-individual expression remains stable in individuals between the pre-symptomatic phase and time of diagnosis. This indicates that differences in expression of some genes in blood of leprosy progressors precede appearance of symptoms.

Further exploration of the pathways that could be responsible for the observed differences in gene expression between leprosy progressors and controls showed that genes overexpressed in leprosy progressors are involved in translation pathways and cotranslation of membrane and ER proteins. EIF2, eIF4 and p70S6K signalling pathways, overexpressed in leprosy progressors, are downstream pathways to the mTOR pathway which also displayed higher levels in progressors and regulates protein translation, gene expression, metabolic processes, immune receptor signalling and migratory activity (80, 81). Likewise, in several other diseases such as cancer, type 2 diabetes, rheumatoid arthritis and viral infections the mTOR pathway is deregulated (80, 82). In general, antigen recognition activates the mTOR signalling pathway as a result of which naïve CD4+ T-cells differentiate into Th1, Th2 and Th17 (81). This process may thus lead to a higher expression of mRNA related to Th1, Th2 and Th17 in leprosy patients compared to healthy individuals as previously observed (83). Interestingly, upregulation of the coronavirus pathogenesis pathway was also observed in leprosy progressors. This could be caused by activation of the inflammatory and autophagy regulation pathways in individuals infected with coronaviruses as well as BT leprosy patients (84, 85).

Downregulated gene expression in leprosy progressors was observed and occurred in organelle and cellular component organization pathways as well as integrin and FAK signalling pathways. FAK is a tyrosine kinase downstream of integrin growth factor. Nuclear FAK regulates transcription of inflammatory signalling, immune escape, angiogenesis and p53

(86). Moreover, overexpression of FAK has been linked to advanced cancer and metastasis (86, 87). Although FAK inhibitors are currently being tested for use in cancer treatment, the FAK signalling pathway has never been studied in leprosy. Hence the significance of under-expression in leprosy progressors before diagnosis as observed in this study requires further investigation.

Recently, Leal-Calvo and Moraes performed a comprehensive reanalysis of nine publicly available microarrays of leprosy patients from variable origin. The authors found DGE in skin development processes including genes such as *AQP3*, *AKR1C3*, *CYP27B1*, *LTB*, *VDR* and keratinocyte biology with *CSTA*, *DSG1*, *KRT14*, *KRT5*, *PKP1* and *IVL* (88). None of the genes identified by that study were, however, found in our analysis. This could be due to the fact that this reanalysis mainly investigated DGE between different leprosy types (BT vs LL), ENL reaction and LL or LL and healthy controls, whereas our study included mostly BT leprosy patients and HC. Moreover, RNA was obtained from skin biopsies or cell cultures instead of whole blood. Consequently, comparison of their results with the present work is limited and while possible biomarker genes were identified in the microarray reanalysis, application for diagnosis would be restricted to patients with visible symptoms as well as requiring more invasive samples (skin biopsies or cell culture). In contrast, the prospective 4-gene signature RISK4LEP identified in this study is measured in whole blood.

Similarly, other transcriptomic studies described leprosy biomarkers associated with leprosy but after occurrence of symptoms and using skin biopsies (54, 64, 89, 90). Serrano-Coll and colleagues showed that RT-qPCR of Oct-6 identified multibacillary (MB) patients (n=30) in Colombia with an AUC of 83.0% (89). However, S-100 immunohistochemistry alone showed a better AUC (96.0%). Pinto *et al.* investigated the expression of non-coding RNAs in leprosy patients (5 TT and 6 LL) from Brazil (90) and found five P-element-induced wimpy testis (PIWI)-interacting RNAs (piRNAs) that classified leprosy patients with an AUC of 90.0%. Jorge *et al*. established a non-coding RNA signature consisting of four miRNA, that discriminated leprosy patients (6 LL and 6 TT; AUC=87.3%) in Brazil (54). Furthermore, Guerreiro and colleagues using nerve biopsies of PN leprosy patients (n=28) identified a transcriptomic signature based on a classification tree including *LDR* and *CCL4* which could ascertain 80% of PN leprosy patients (64). Although *CCL4* and *LDR* were not significantly differentially expressed in our study, we also found lower expression levels of mitochondrial genes involved in the oxidative phosphorylation pathway in blood of leprosy progressors, in line with their findings in *M. leprae*-infected Schwann cells and nerve biopsies of Brazilian leprosy patients. This reduction may be caused by down-regulation of mitochondrial genes by mycobacteria during *M. leprae* infection to inhibit apoptosis and promote intracellular bacterial survival (91).

We found moderate prediction (AUC=78.3%) of leprosy when the RISK6 genes (*GBP2*, *FC-*

GR1B, *SERPING1*, *TUBGCP6*, *TRMT2A*, *SDR39U1*) were used as input in the Random Forest with RNA-Seq data. This is likely due to similarities in the immune response to mycobacteria in leprosy and tuberculosis patients (92). In line with this, *FCGR1A* and *GBP* genes were previously found to be upregulated in leprosy patients or during leprosy reactions in and outside Bangladesh (49, 55, 93).

It has been previously reported that RNA profiles in blood of leprosy patients are different from those derived from skin (63). Although transcriptomic analysis in skin of leprosy patients provide deeper insight into leprosy pathogenesis, the aim of this study was to identify leprosy predictive biomarkers, preferably measurable in rapid diagnostic tests. Thus, whole blood is a preferred biosample because it can be collected relatively easily and translated into field-friendly tests applying fingerstick blood (52).

In summary, the RISK4LEP signature described here, offers potential for the development of a point of care test allowing the identification of leprosy progression among HC in blood years before symptom development. Since the present study was performed in Bangladesh, additional, longitudinal studies are required to determine whether this signature predicts leprosy progression in endemic populations from different origins. Moreover, since the majority of the Bangladeshi patients who developed leprosy during this study developed BT leprosy, similar studies will also need to provide information on the performance of the signature to predict occurrence of LL types. It is tempting to speculate that this signature could identify early forms of BT leprosy, thus preventing the more severe LL types from developing.

Nevertheless, the novel RISK4LEP signature predicts development of (BT) leprosy up to 61 months before clinical diagnosis. Such signatures, when properly validated in other populations as well, can be applied for targeted preventive treatment and reduction of *M. leprae* transmission among HC.

Declaration of interests

The authors declare no conflicts of interests.

Acknowledgements

The authors gratefully acknowledge all patients and control participants in Bangladesh and would like to thank Suzanne van Veen and Mariëlle Haks for assistance and advice on the Biomark HD system (Fluidigm).

This study was supported by an R2STOP Research grant from effect:hope, Canada and the Mission to End Leprosy, Ireland; the Order of Malta-Grants-for-Leprosy-Research (MALTA-LEP); the Q.M. Gastmann-Wichers Foundation; the Leprosy Research Initiative (LRI) together with the Turing Foundation (ILEP; 702.02.73 and 703.15.07).

Data sharing statement

Sequence data have been submitted to NCBI Gene Expression Omnibus (GEO) under accession number GSE163498.

RNA-Seq files were processed using the opensource BIOWDL RNA-Seq pipeline v2.0 (https://github.com/biowdl/RNA-seq/tree/v2.0.0) developed at Leiden University Medical Centre (The Netherlands).

References

1. World Health Organization. Global leprosy update, 2018: moving towards a leprosy-free world. Weekly Epidemiological Record. 2019;94(35/36):389-412.
2. Scollard DM. Pathogenesis and Pathology of Leprosy. In: Scollard DM, Gillis TP, editors. International textbook of leprosy. www.internationaltextbookofleprosy.org 18 September 2016, posting date.
3. Scollard DM, Adams LB, Gillis TP, Krahenbuhl JL, Truman RW, Williams DL. The continuing challenges of leprosy. Clin Microbiol Rev. 2006;19(2):338-81.
4. Kumar B, Uprety S, Dogra S. Clinical diagnosis of leprosy. In: Scollard DM, Gills TP, editors. International textbook of leprosy. www.internationaltextbookofleprosy.org 9 February 2017, posting date.
5. Ridley DS, Jopling WH. Classification of leprosy according to immunity. A five-group system. Int J Lepr Other Mycobact Dis. 1966;34(3):255-73.
6. Richardus JH, Ignotti E, Smith WCS. Epidemiology of Leprosy. In: Scollard DM, Gillis TP, editors. International textbook of leprosy. www.internationaltextbookofleprosy.org 18 September 2016, posting date.
7. Walker SL, Withington SG, Lockwood DNJ. 41 - Leprosy. In: Farrar J, Hotez PJ, Junghanss T, Kang G, Lalloo D, White NJ, editors. Manson's Tropical Infectious Diseases (Twenty-third Edition). London: W.B. Saunders; 2014. p. 506-18.e1.
8. Uaska Sartori PV, Penna GO, Bührer-Sékula S, Pontes MAA, Gonçalves HS, Cruz R, et al. Human Genetic Susceptibility of Leprosy Recurrence. Scientific reports. 2020;10(1):1284-.
9. Zhang FR, Huang W, Chen SM, Sun LD, Liu H, Li Y, et al. Genomewide association study of leprosy. N Engl J Med. 2009;361(27):2609-18.
10. Cardoso CC, Pereira AC, de Sales Marques C, Moraes MO. Leprosy susceptibility: genetic variations regulate innate and adaptive immunity, and disease outcome. Future Microbiol. 2011;6(5):533-49.
11. Alter A, Grant A, Abel L, Alcais A, Schurr E. Leprosy as a genetic disease. Mamm Genome. 2011;22(1-2):19-31.
12. Wang D, Xu L, Lv L, Su LY, Fan Y, Zhang DF, et al. Association of the LRRK2 genetic polymorphisms with leprosy in Han Chinese from Southwest China. Genes Immun. 2015;16(2):112-9.
13. Sales-Marques C, Cardoso CC, Alvarado-Arnez LE, Illaramendi X, Sales AM, Hacker MA, et al. Genetic polymorphisms of the IL6 and NOD2 genes are risk factors for inflammatory reactions in leprosy. PLoS Negl Trop Dis. 2017;11(7):e0005754.
14. Mira MT, Alcais A, Nguyen VT, Moraes MO, Di Flumeri C, Vu HT, et al. Susceptibility to leprosy is associated with PARK2 and PACRG. Nature. 2004;427(6975):636-40.
15. Dwivedi VP, Banerjee A, Das I, Saha A, Dutta M, Bhardwaj B, et al. Diet and nutrition: An important risk factor in leprosy. Microb Pathog. 2019;137:103714.
16. Feenstra SG, Nahar Q, Pahan D, Oskam L, Richardus JH. Recent food shortage is associated with leprosy disease in Bangladesh: a case-control study. PLOS Neglected Tropical Diseases. 2011;5(5):e1029.
17. Serrano-Coll H, Mora HR, Beltrán JC, Duthie MS, Cardona-Castro N. Social and environmental conditions related to Mycobacterium leprae infection in children and adolescents from three leprosy endemic regions of Colombia. BMC infectious diseases. 2019;19(1):520.
18. Kerr-Pontes LRS, Montenegro ACD, Barreto ML, Werneck GL, Feldmeier H. Inequality and leprosy in Northeast Brazil: an ecological study. International Journal of Epidemiology. 2004;33(2):262-9.
19. Vieira MCA, Nery JS, Paixão ES, Freitas de Andrade KV, Oliveira Penna G, Teixeira MG. Leprosy in children under 15 years of age in Brazil: A systematic review of the literature. PLoS neglected tropical diseases. 2018;12(10):e0006788.
20. Sarkar R, Pradhan S. Leprosy and women. International Journal of Women's Dermatology. 2016;2(4):117-21.

21. Bakker MI, Hatta M, Kwenang A, Van Mosseveld P, Faber WR, Klatser PR, et al. Risk factors for developing leprosy--a population-based cohort study in Indonesia. Lepr Rev. 2006;77(1):48-61.
22. Goulart IM, Bernardes Souza DO, Marques CR, Pimenta VL, Goncalves MA, Goulart LR. Risk and protective factors for leprosy development determined by epidemiological surveillance of household contacts. Clin Vaccine Immunol. 2008;15(1):101-5.
23. Sales AM, Ponce de Leon A, Duppre NC, Hacker MA, Nery JA, Sarno EN, et al. Leprosy among patient contacts: a multilevel study of risk factors. PLoS Negl Trop Dis. 2011;5(3):e1013.
24. Quilter EEV, Butlin CR, Singh S, Alam K, Lockwood DNJ. Patients with skin smear positive leprosy in Bangladesh are the main risk factor for leprosy development: 21-year follow-up in the household contact study (COCOA). PLoS neglected tropical diseases. 2020;14(10):e0008687.
25. World Health Organization. Regional Office for South-East Asia. Guidelines for the diagnosis, treatment and prevention of leprosy. World Health Organization Regional Office for South-East Asia. 2018.
26. van Hooij A, van den Eeden S, Richardus R, Tjon Kon Fat E, Wilson L, Franken K, et al. Application of new host biomarker profiles in quantitative point-of-care tests facilitates leprosy diagnosis in the field. EBioMedicine. 2019;47:301-8.
27. van Hooij A, Tjon Kon Fat EM, Batista da Silva M, Carvalho Bouth R, Cunha Messias AC, Gobbo AR, et al. Evaluation of Immunodiagnostic Tests for Leprosy in Brazil, China and Ethiopia. Sci Rep. 2018;8(1):17920.
28. van Hooij A, Tjon Kon Fat EM, Richardus R, van den Eeden SJ, Wilson L, de Dood CJ, et al. Quantitative lateral flow strip assays as User-Friendly Tools To Detect Biomarker Profiles For Leprosy. Sci Rep. 2016;6:34260.
29. van Hooij A, Tjon Kon Fat EM, van den Eeden SJF, Wilson L, Batista da Silva M, Salgado CG, et al. Field-friendly serological tests for determination of M. leprae-specific antibodies. Sci Rep. 2017;7(1):8868.
30. Chen X, You YG, Yuan YH, Yuan LC, Wen Y. Host immune responses induced by specific Mycobacterium leprae antigens in an overnight whole-blood assay correlate with the diagnosis of paucibacillary leprosy patients in China. PLoS Negl Trop Dis. 2019;13(4):e0007318.
31. Queiroz EA, Medeiros NI, Mattos RT, Carvalho APM, Rodrigues-Alves ML, Dutra WO, et al. Immunological biomarkers of subclinical infection in household contacts of leprosy patients. Immunobiology. 2019;224(4):518-25.
32. Gama RS, Gomides TAR, Gama CFM, Moreira SJM, de Neves Manta FS, de Oliveira LBP, et al. High frequency of M. leprae DNA detection in asymptomatic household contacts. BMC Infect Dis. 2018;18(1):153.
33. Barbieri RR, Manta FSN, Moreira SJM, Sales AM, Nery JAC, Nascimento LPR, et al. Quantitative polymerase chain reaction in paucibacillary leprosy diagnosis: A follow-up study. PLoS Negl Trop Dis. 2019;13(3):e0007147.
34. Martinez AN, Ribeiro-Alves M, Sarno EN, Moraes MO. Evaluation of qPCR-based assays for leprosy diagnosis directly in clinical specimens. PLoS Negl Trop Dis. 2011;5(10):e1354.
35. Manta FSN, Barbieri RR, Moreira SJM, Santos PTS, Nery JAC, Duppre NC, et al. Quantitative PCR for leprosy diagnosis and monitoring in household contacts: A follow-up study, 2011-2018. Sci Rep. 2019;9(1):16675.
36. Carvalho RS, Foschiani IM, Costa M, Marta SN, da Cunha Lopes Virmond M. Early detection of M. leprae by qPCR in untreated patients and their contacts: results for nasal swab and palate mucosa scraping. Eur J Clin Microbiol Infect Dis. 2018;37(10):1863-7.
37. Pathak VK, Singh I, Turankar RP, Lavania M, Ahuja M, Singh V, et al. Utility of multiplex PCR for early diagnosis and household contact surveillance for leprosy. Diagnostic microbiology and infectious disease. 2019;95(3):114855-.
38. Gama RS, Souza MLM, Sarno EN, Moraes MO, Goncalves A, Stefani MMA, et al. A novel integrated molecular and serological analysis method to predict new cases of leprosy amongst household contacts. PLoS Negl Trop Dis. 2019;13(6):e0007400.
39. van Beers SM, Izumi S, Madjid B, Maeda Y, Day R, Klatser PR. An epidemiological study of leprosy infection by serology and polymerase chain reaction. Int J Lepr Other Mycobact Dis. 1994;62(1):1-9.
40. Bakker MI, Hatta M, Kwenang A, Van Mosseveld P, Faber WR, Klatser PR, et al. Risk factors for developing leprosy--a population-based cohort study in Indonesia. Leprosy review. 2006;77(1):48-61.
41. Martins ACdC, Miranda A, Oliveira MLW-d-Rd, Bührer-Sékula S, Martinez A. Nasal mucosa study of leprosy contacts with positive serology for the phenolic glycolipid 1 antigen. Brazilian journal of otorhinolaryngology. 2010;76(5):579-87.
42. Cardona-Castro N, Beltran-Alzate JC, Manrique-Hernandez R. Survey to identify Mycobacterium leprae-infected household contacts of patients from prevalent regions of leprosy in Colombia. Mem Inst Oswaldo Cruz. 2008;103(4):332-6.

43. Santos DFD, Mendonca MR, Antunes DE, Sabino EFP, Pereira RC, Goulart LR, et al. Molecular, immunological and neurophysiological evaluations for early diagnosis of neural impairment in seropositive leprosy household contacts. PLoS neglected tropical diseases. 2018;12(5):e0006494.
44. Tió-Coma M, Avanzi C, Verhard EM, Pierneef L, van Hooij A, Benjak A, et al. Genomic Characterization of Mycobacterium leprae to Explore Transmission Patterns Identifies New Subtype in Bangladesh. Frontiers in microbiology. 2020;11(1220).
45. van Hooij A, Tió-Coma M, Verhard EM, Khatun M, Alam K, Tjon Kon Fat E, et al. Household Contacts of Leprosy Patients in Endemic Areas Display a Specific Innate Immunity Profile. Frontiers in Immunology. 2020;11(1811).
46. Donoghue HD, Holton J, Spigelman M. PCR primers that can detect low levels of *Mycobacterium leprae* DNA. J Med Microbiol. 2001;50(2):177-82.
47. Martinez AN, Lahiri R, Pittman TL, Scollard D, Truman R, Moraes MO, et al. Molecular determination of *Mycobacterium leprae* viability by use of real-time PCR. J Clin Microbiol. 2009;47(7):2124-30.
48. Niemantsverdriet E, van den Akker EB, Boeters DM, van den Eeden SJF, Geluk A, van der Helm-van Mil AHM. Gene expression identifies patients who develop inflammatory arthritis in a clinically suspect arthralgia cohort. Arthritis Research & Therapy. 2020;22(1):266.
49. Teles RMB, Lu J, Tió-Coma M, Goulart IMB, Banu S, Hagge D, et al. Identification of a systemic interferon-γ inducible antimicrobial gene signature in leprosy patients undergoing reversal reaction. PLoS Negl Trop Dis. 2019;13(10):e0007764.
50. Zak DE, Penn-Nicholson A, Scriba TJ, Thompson E, Suliman S, Amon LM, et al. A blood RNA signature for tuberculosis disease risk: a prospective cohort study. Lancet. 2016;387(10035):2312-22.
51. Suliman S, Thompson E, Sutherland J, Weiner Rd J, Ota MOC, Shankar S, et al. Four-gene Pan-African Blood Signature Predicts Progression to Tuberculosis. Am J Respir Crit Care Med. 2018.
52. Penn-Nicholson A, Mbandi SK, Thompson E, Mendelsohn SC, Suliman S, Chegou NN, et al. RISK6, a 6-gene transcriptomic signature of TB disease risk, diagnosis and treatment response. Scientific reports. 2020;10(1):8629.
53. Turner CT, Gupta RK, Tsaliki E, Roe JK, Mondal P, Nyawo GR, et al. Blood transcriptional biomarkers for active pulmonary tuberculosis in a high-burden setting: a prospective, observational, diagnostic accuracy study. Lancet Respir Med. 2020;8(4):407-19.
54. Jorge K, Souza RP, Assis MTA, Araujo MG, Locati M, Jesus AMR, et al. Characterization of MicroRNA Expression Profiles and Identification of Potential Biomarkers in Leprosy. J Clin Microbiol. 2017;55(5):1516-25.
55. Tio-Coma M, van Hooij A, Bobosha K, van der Ploeg-van Schip JJ, Banu S, Khadge S, et al. Whole blood RNA signatures in leprosy patients identify reversal reactions before clinical onset: a prospective, multicenter study. Sci Rep. 2019;9(1):17931.
56. Montoya DJ, Andrade P, Silva BJA, Teles RMB, Ma F, Bryson B, et al. Dual RNA-Seq of Human Leprosy Lesions Identifies Bacterial Determinants Linked to Host Immune Response. Cell Rep. 2019;26(13):3574-85.e3.
57. Andrade PR, Mehta M, Lu J, Teles RMB, Montoya D, Scumpia PO, et al. The cell fate regulator NUPR1 is induced by Mycobacterium leprae via type I interferon in human leprosy. PLoS Negl Trop Dis. 2019;13(7):e0007589.
58. Inkeles MS, Teles RM, Pouldar D, Andrade PR, Madigan CA, Lopez D, et al. Cell-type deconvolution with immune pathways identifies gene networks of host defense and immunopathology in leprosy. JCI Insight. 2016;1(15):e88843.
59. Realegeno S, Kelly-Scumpia KM, Dang AT, Lu J, Teles R, Liu PT, et al. S100A12 Is Part of the Antimicrobial Network against Mycobacterium leprae in Human Macrophages. PLoS Pathog. 2016;12(6):e1005705.
60. Orlova M, Cobat A, Huong NT, Ba NN, Van TN, Spencer J, et al. Gene set signature of reversal reaction type I in leprosy patients. PLoS Genet. 2013;9(7):e1003624.
61. Guerreiro LT, Robottom-Ferreira AB, Ribeiro-Alves M, Toledo-Pinto TG, Rosa BT, Rosa PS, et al. Gene expression profiling specifies chemokine, mitochondrial and lipid metabolism signatures in leprosy. PLoS One. 2013;8(6):e64748.
62. Bleharski JR, Li H, Meinken C, Graeber TG, Ochoa MT, Yamamura M, et al. Use of genetic profiling in leprosy to discriminate clinical forms of the disease. Science. 2003;301(5639):1527-30.
63. Salgado CG, Pinto P, Bouth RC, Gobbo AR, Messias ACC, Sandoval TV, et al. miRNome Expression Analysis Reveals New Players on Leprosy Immune Physiopathology. Frontiers in Immunology. 2018;9(463).
64. Guerreiro LT, Robottom-Ferreira AB, Ribeiro-Alves M, Toledo-Pinto TG, Rosa Brito T, Rosa PS, et al. Gene expression profiling specifies chemokine, mitochondrial and lipid metabolism signatures in leprosy. PLoS One. 2013;8(6):e64748.
65. Richardus R, Alam K, Kundu K, Chandra Roy J, Zafar T, Chowdhury AS, et al. Effectiveness of single-dose rifampicin after BCG vaccination to prevent leprosy in close contacts of patients with

newly diagnosed leprosy: A cluster randomized controlled trial. Int J Infect Dis. 2019;88:65-72.
66. Richardus R, van Hooij A, van den Eeden SJF, Wilson L, Alam K, Richardus JH, et al. BCG and Adverse Events in the Context of Leprosy. Front Immunol. 2018;9:629.
67. Richardus RA, Alam K, Pahan D, Feenstra SG, Geluk A, Richardus JH. The combined effect of chemoprophylaxis with single dose rifampicin and immunoprophylaxis with BCG to prevent leprosy in contacts of newly diagnosed leprosy cases: a cluster randomized controlled trial (MALTALEP study). BMC infectious diseases. 2013;13:456.
68. Yearly district activity report 2018. Internal communication: Rural Health Program, the Leprosy Mission International Bangladesh.
69. National Leprosy Elimination Programme. National guidelines and technical manual on leprosy. 4th ed. Dhaka: Bangladesh National Leprosy Elimination Programme Directorate General of Health Services; 2014.
70. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics. 2009;25(8):1091-3.
71. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2009;26(1):139-40.
72. Breiman L. Random Forests. Machine Learning. 2001;45(1):5-32.
73. Lang M, Kotthaus H, Marwedel P, Weihs C, Rahnenführer J, Bischl B. Automatic model selection for high-dimensional survival analysis. Journal of Statistical Computation and Simulation. 2015;85(1):62-76.
74. Sweeney TE, Braviak L, Tato CM, Khatri P. Genome-wide expression for diagnosis of pulmonary tuberculosis: a multicohort analysis. Lancet Respir Med. 2016;4(3):213-24.
75. Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, et al. The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. Curr Protoc Bioinformatics. 2016;54:1.30.1-1..3.
76. Vanlalhruaii Tonsing M, Vanlalbiakdiki Sailo C, Zothansanga, Chhakchhuak L, Chhakchhuak Z, Pandit B, et al. Analysis of variants in mitochondrial genome and their putative pathogenicity in tuberculosis patients from Mizoram, North east India. Mitochondrion. 2020;54:21-5.
77. Wang Z, Arat S, Magid-Slav M, Brown JR. Meta-analysis of human gene expression in response to Mycobacterium tuberculosis infection reveals potential therapeutic targets. BMC Syst Biol. 2018;12(1):3-.
78. Manzanillo PS, Ayres JS, Watson RO, Collins AC, Souza G, Rae CS, et al. The ubiquitin ligase parkin mediates resistance to intracellular pathogens. Nature. 2013;501(7468):512-6.
79. Scarffe LA, Stevens DA, Dawson VL, Dawson TM. Parkin and PINK1: much more than mitophagy. Trends in Neurosciences. 2014;37(6):315-24.
80. Le Sage V, Cinti A, Amorim R, Mouland AJ. Adapting the Stress Response: Viral Subversion of the mTOR Signaling Pathway. Viruses. 2016;8(6):152.
81. Chi H. Regulation and function of mTOR signalling in T cell fate decisions. Nature Reviews Immunology. 2012;12(5):325-38.
82. Laplante M, Sabatini DM. mTOR signaling at a glance. Journal of Cell Science. 2009;122(20):3589-94.
83. Azevedo MCS, Marques H, Binelli LS, Malange MSV, Devides AC, Silva EA, et al. Simultaneous analysis of multiple T helper subsets in leprosy reveals distinct patterns of Th1, Th2, Th17 and Tregs markers expression in clinical forms and reactional events. Med Microbiol Immunol. 2017;206(6):429-39.
84. Vardhana SA, Wolchok JD. The many faces of the anti-COVID immune response. Journal of Experimental Medicine. 2020;217(6).
85. Sekine T, Perez-Potti A, Rivera-Ballesteros O, Strålin K, Gorin J-B, Olsson A, et al. Robust T Cell Immunity in Convalescent Individuals with Asymptomatic or Mild COVID-19. Cell. 2020;183(1):158-68.e14.
86. Zhou J, Yi Q, Tang L. The roles of nuclear focal adhesion kinase (FAK) on Cancer: a focused review. Journal of Experimental & Clinical Cancer Research. 2019;38(1):250.
87. Yoon H, Dehart JP, Murphy JM, Lim ST. Understanding the roles of FAK in cancer: inhibitors, genetic models, and new insights. J Histochem Cytochem. 2015;63(2):114-28.
88. Leal-Calvo T, Moraes MO. Reanalysis and integration of public microarray datasets reveals novel host genes modulated in leprosy. Mol Genet Genomics. 2020.
89. Serrano-Coll H, Salazar-Peláez LM, Mesa-Betancourt F, Cardona-Castro N. Oct-6 transcriptional factor a possible biomarker for leprosy diagnosis. Diagn Microbiol Infect Dis. 2020;99(2):115232.
90. Pinto P, da Silva MB, Moreira FC, Bouth RC, Gobbo AR, Sandoval TV, et al. Leprosy piRnome: exploring new possibilities for an old disease. Scientific reports. 2020;10(1):12648.
91. Dubey R. Assuming the role of mitochondria in mycobacterial infection. International Journal of Mycobacteriology. 2016;5(4):379-83.

92. Modlin RL, Bloom BR. TB or not TB: that is no longer the question. Sci Transl Med. 2013;5(213):213sr6.
93. Geluk A, van Meijgaarden KE, Wilson L, Bobosha K, JJ vdP-vS, van den Eeden SJ, et al. Longitudinal immune responses and gene expression profiles in type 1 leprosy reactions. J Clin Immunol. 2014;34(2):245-55.

Supplementary Material

Table S1. Primers and probe Assay IDs.

Gene name	Ensembl ID	Assay ID
SNHG32 or *C6orf48*	ENSG00000204387	Hs00382553_m1
MT-ND4	ENSG00000198886	Hs02596876_g1
MT-ND5	ENSG00000198786	Hs02596878_g1
MT-ND2	ENSG00000198763	Hs02596874_g1
MT-CO1	ENSG00000198804	Hs02596864_g1
TAOK3	ENSG00000135090	Hs00937694_m1
REPS1	ENSG00000135597	Hs01016191_m1
MT-CYB	ENSG00000198727	Hs02596867_s1
TPGS1	ENSG00000141933	Hs00293366_m1
MMRN1	ENSG00000138722	Hs01113299_m1
UBC	ENSG00000150991	Hs00824723_m1
REX1BD or *C19orf60*	ENSG00000006015	Hs00215835_m1
CCDC85B	ENSG00000175602	Hs00255227_s1
GAPDH	ENSG00000111640	Hs99999905_m1
GAPDH	ENSG00000111640	Hs02786624_g1

Assay IDs (Thermo Fisher Scientific) per gene for forward primers, reverse primers and probes with dye (FAM-MGB) used to measure gene expression with Biomark HD (Fluidigm).

Table S2. RT-qPCR median Ct values in leprosy progressors and HC controls at recruitment into the study (t=1).

Gene	p-value	Ct progressors	Ct HC
CCDC85B	0.520371	19.74	19.69
GAPDH_g1	0.017288	9.88	9.27
GAPDH_m1	0.407393	7.33	7.22
MMRN1	0.764781	15.09	15.04
MT-CO1	0.225810	4.91	4.83
MT-CYB	0.028590	5.61	5.17
MT-ND2	0.023773	6.09	5.57
MT-ND4	0.033548	5.48	5.17
MT-ND5	0.079548	6.18	6.09

REPS1	0.605180	11.60	11.47
REX1BD	0.567703	10.23	10.40
SNHG32	0.897977	8.99	9.01
TAOK3	0.758857	11.52	11.46
TPGS1	0.176301	12.49	12.68
UBC	0.616087	6.32	6.32

P-values of Mann-Whitney U test of reverse transcription quantitative PCR (RT-qPCR) Cycle threshold (Ct) values between household contacts (HC) who progressed to leprosy (progressors; n=47) and household contacts who remained without leprosy during follow-up (HC; n=47). Median of Ct values per group (Ct progressors, Ct HC). GAPDH_g1 corresponds to assay ID Hs02786624_g1 and GAPDH_m1 to assay ID Hs99999905_m1. t=1 represents timepoint of recruitment into the study, when no clinical signs or symptoms were present.

Table S3. Hyperparameters of the Random Forest models.

Random Forest Model	ntree	mtry	nodesize
19-gene RNA-Seq	527	4	22
13-gene RNA-Seq	380	1	30
13-gene RT-qPCR	200	2	41
4-gene RT-qPCR	202	1	38

Hyperparameters obtained for the Random Forest models. ntree: number of trees to grow; mtry: number of variables randomly sampled as candidates at each split; nodesize: minimum size of terminal nodes. RT-qPCR: quantitative reverse transcription PCR.

Table S4. Overall results of the signatures.

Signature	AUC	AUC CI	Accuracy	Sensitivity	Specificity
19-gene RNA-Seq	96.7%	88.9-100.0%	87.5%	100.0%	80.0%
13-gene RNA-Seq*	95.2%	84.9-100.0%	87.5%	88.9%	85.7%
13-gene RT-qPCR	84.1%	64.7-100.0%	73.7%	87.5%	63.6%
4-gene RT-qPCR RISK-4LEP	86.4%	69.9-100.0%	79.0%	87.5%	72.7%
3-gene RT-qPCR	84.7%	65.9-100.0%	68.4%	87.5%	54.6%
4-gene v2 RT qPCR	78.4%	55.9-100.0%	63.2%	75.0%	54.6%
4-gene RT-qPCR + sex + age + RJ	79.0%	56.7-100.0%	73.7%	75.0%	72.7%
2-gene RT-qPCR + sex + age + RJ	72.1%	48.0%-96.3%	63.2%	75.0%	54.6%
TB Sweeney3 RNA-Seq	51.6%	17.2-79.7%	56.3%	33.3%	85.7%
TB Suliman2 RNA-Seq	58.7%	25.9-91.5%	68.8%	77.8%	57.1%
TB RISK6 RNA-Seq	78.3%	55.0-100.0%	62.5%	100.0%	40.0%

Area Under the Curve (AUC), confidence interval (CI) of AUC, accuracy, sensitivity and specificity of the RNA-Seq, reserves transcription quantitative PCR (RT-qPCR) and tuberculosis (TB) risk signatures to predict leprosy development. The 4-gene signature refers to RISK4LEP which includes *MT-ND2*, *REX1BD*, *TPGS1* and *UBC* genes. The 3-gene signature includes *MT-ND2*, *REX1BD* and *TPGS* genes. The 4-gene signature v2 includes *MT-CYB*, *MT-ND2*, *REX1BD* and *TPGS1* genes. The 2-gene signature includes *TPGS1* and *UBC* genes. RJ: Ridley-Jopling (5) classification of the index leprosy contact. *Exclusion of 3 BCG vaccinated individuals (1 pair and 1 progressor) showed a similar performance of the model (AUC 92.6%).

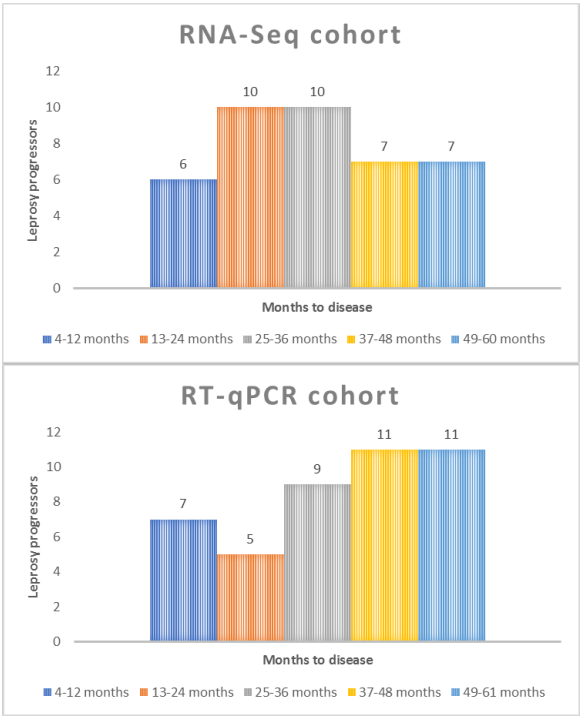


Figure S1. Months between samples at recruitment (t=1) and leprosy diagnosis (t=2) for leprosy progressors. Number of leprosy progressors (y-axis) and number of months elapsed between recruitment (t=1) and leprosy diagnosis (t=2) (x-axis).

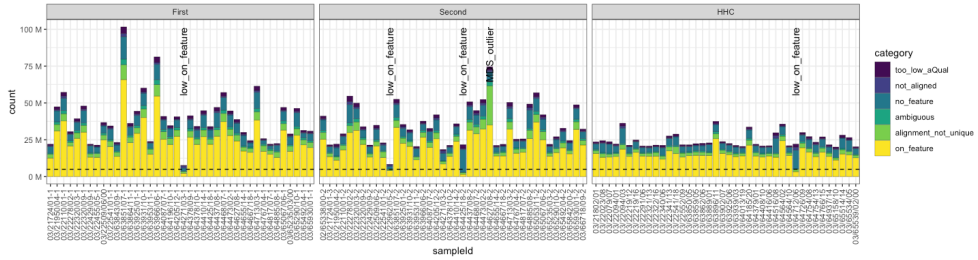


Figure S2. RNA-Seq read count. Read count of whole blood RNA samples (n=120) for quality control. The number of reads are shown per sample and color indicates different categories: reads with

low alignment mapping quality score (too_low_aQual), reads in the SAM file without alignment (not_aligned), reads which could not be assigned to any feature (no_feature), reads which could have been assigned to more than one feature and hence were not counted for any feature (ambiguous), reads with more than one reported alignment (alignment_not_unique), reads which could be assigned to a feature (on_feature). Samples with a low on-feature read count (yellow) were excluded from differential gene expression analyses.

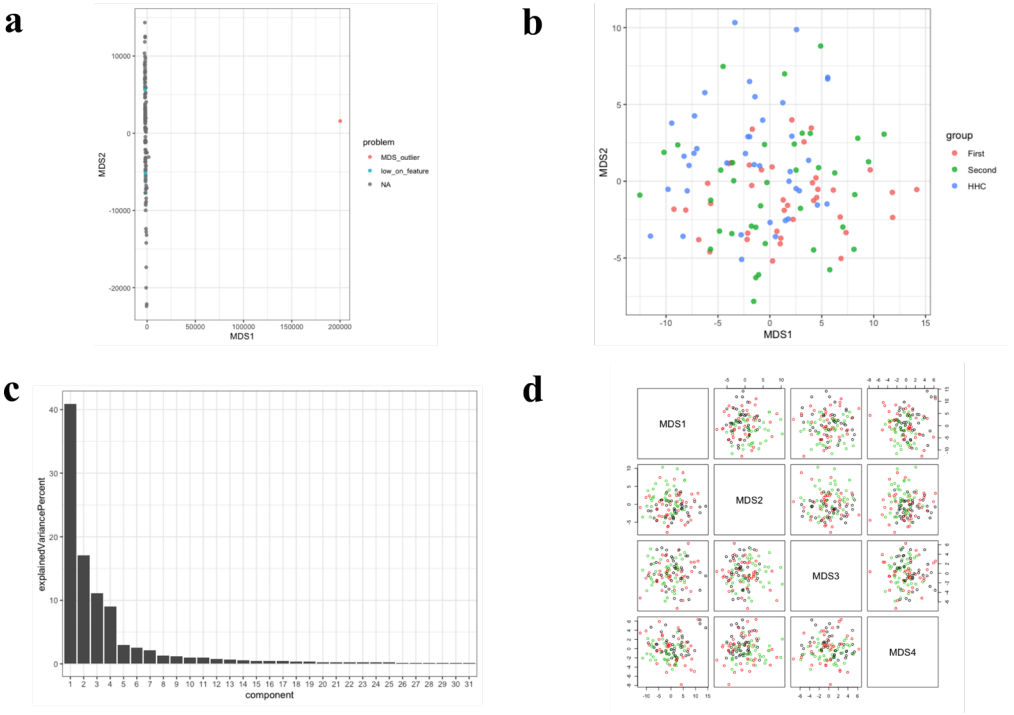


Figure S3. RNA-Seq Multidimensional scaling (MDS) plots. MDS plots of RNA-Seq read count of whole blood RNA samples (n=120) where each circle represents a sample. a) One sample (MDS_outlier, in pink) presented abnormal values showing as an outlier and was excluded from differential gene expression analyses. The rest of the samples (not applicable (NA) in grey and low on-feature in blue) clustered together. b) MDS plot after normalization and exclusion of one outlier in plot A. Samples from leprosy progressors 4-61 months before clinical diagnosis of leprosy (t=1, first) are shown in pink, progressors at timepoint of leprosy diagnosis (t=2, second) in green and household contact (HHC) controls in blue. c) Bar plot with the percentage of variance explained by each MDS component. d) Pairwise MDS plot of the first 4 dimensions (components) after normalization and exclusion of one outlier. Samples from leprosy progressors 4-61 months before clinical diagnosis of leprosy (t=1, first) are shown in black, progressors at timepoint of leprosy diagnosis (t=2, second) in red and controls in green.

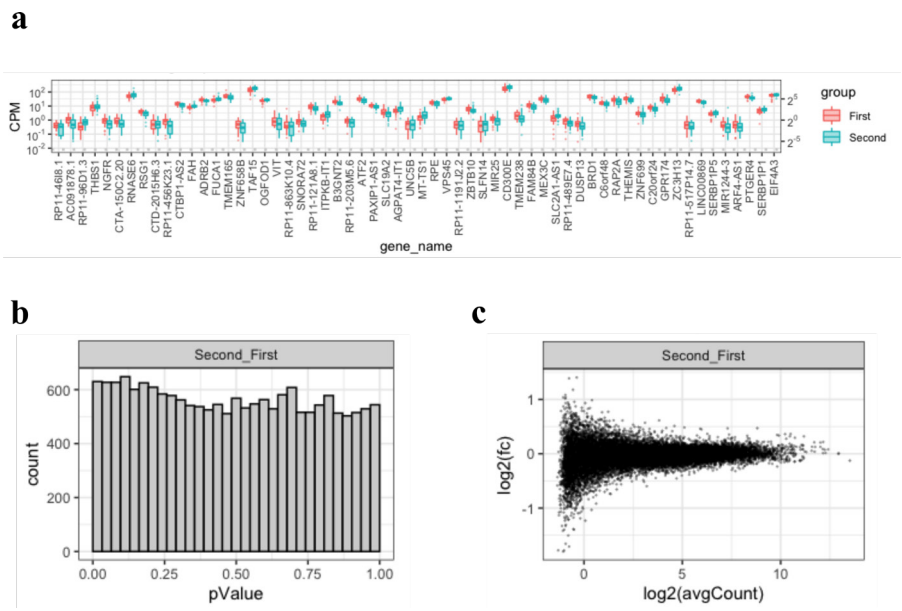


Figure S4. RNA-Seq differential gene expression of leprosy progressors before clinical diagnosis (t=1) and at leprosy diagnosis (t=2). RNA-Seq data of whole blood from leprosy progressors (n=36, after exclusion of samples with low quality control outcome, Figure S2 and S3) 4-61 months before clinical diagnosis of leprosy (t=1, first) was compared to timepoint of leprosy diagnosis (t=2, second). A two-groups (paired samples) analysis was performed using edgeR (71) in R. a) Boxplot of Trimmed Mean of the M-values (TMM)-normalized counts per million mapped reads (CPM) per group of the most significantly differentially expressed genes. Y-axis shows CPM, expressed in power of 10 (left) or power of 2 (right). First timepoint is shown in red and second timepoint in blue. b) Histogram of p-values. Number of genes (y-axis) with a determinate p-value (x-axis). c) MA plot showing log2 of fold change (FC) in gene expression (y-axis) and log2 of average CPM (x-axis) per gene. In red genes significantly differentially expressed (adjusted p-value < 0.05) and in black genes not differentially expressed.

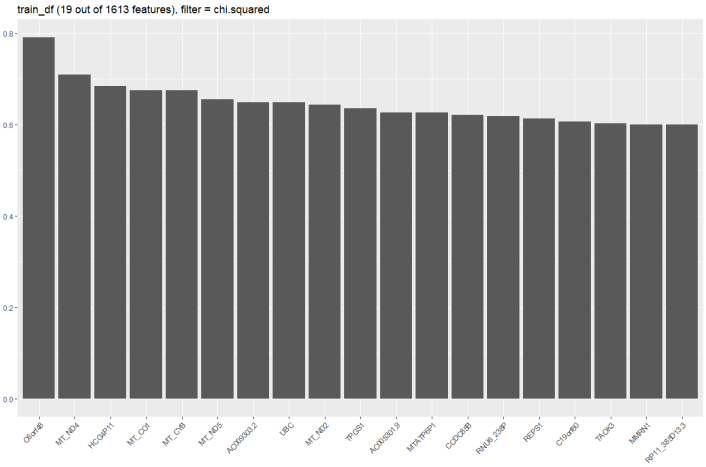


Figure S5. Importance of the RNA-Seq features. A Random Forest model with 1613 features was

fused with chi-squared feature selection strategy and the 19 most informative genes were selected for further analysis.

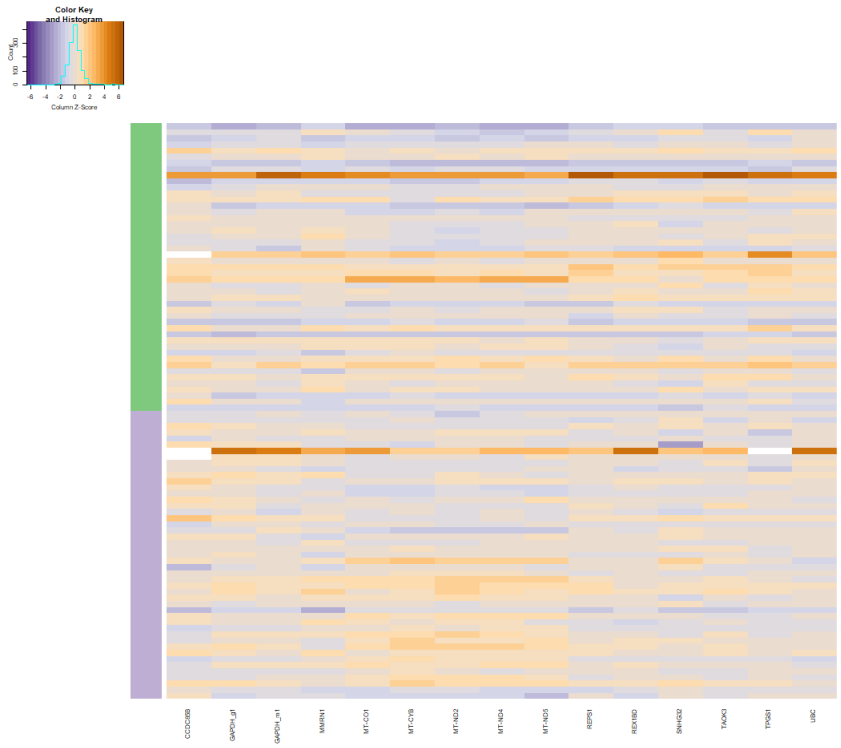


Figure S6. Heatmap of reverse RT-qPCR Ct. Heatmap depicting normalized reverse transcription quantitative PCR (RT-qPCR) Cycle threshold (Ct) values (z-scores) of different genes for leprosy progressors at t=1 (purple) and controls (green). Higher gene expression (lower Ct values) is shown in dark purple and lower gene expression (higher Ct values) in dark orange. No amplification is shown in white.

6

Whole blood RNA signatures in leprosy patients identify reversal reactions before clinical onset: a prospective, multicenter study

Maria Tió-Coma, Anouk van Hooij, Kidist Bobosha, Jolien J. van der Ploeg-van Schip, Sayera Banu, Saraswoti Khadge, Pratibha Thapa, Chhatra B Kunwar, Isabela M. Goulart, Yonas Bekele, Deanna A. Hagge, Milton O. Moraes, Rosane M.B. Teles, Roberta Olmo Pinheiro, Erik W. van Zwet, Jelle J. Goeman, Abraham Aseffa, Mariëlle C. Haks, Tom H.M. Ottenhoff, Robert L. Modlin and Annemieke Geluk

Sci Rep. 2019 Nov 29;9(1):17931

Abstract

Early diagnosis of leprosy is challenging, particularly its inflammatory reactions, the major cause of irreversible neuropathy in leprosy. Current diagnostics cannot identify which patients are at risk of developing reactions. This study assessed blood RNA-expression levels as potential biomarkers for leprosy.

Prospective cohorts of newly diagnosed leprosy patients, including reactions, and healthy controls were recruited in Bangladesh, Brazil, Ethiopia and Nepal. RNA-expression in 1,090 whole blood samples was determined for 103 target genes for innate- and adaptive immune profiling by dual color Reverse-Transcription Multiplex Ligation-dependent Probe Amplification (dcRT-MLPA) followed by cluster analysis.

We identified transcriptomic biomarkers associated with leprosy disease, different leprosy phenotypes as well as high exposure to *Mycobacterium leprae* which respectively allow improved diagnosis and classification of leprosy patients and detection of infection. Importantly, a transcriptomic signature of risk for reversal reactions consisting of five genes (*CCL2, CD8A, IL2, IL15* and *MARCO*) was identified based on cross-sectional comparison of RNA-expression. In addition, intra-individual longitudinal analyses of leprosy patients before, during and after treatment of reversal reactions, indicated that several IFN-induced genes increased significantly at onset of reaction whereas *IL15* decreased.

This multi-site study, situated in four leprosy endemic areas, demonstrates the potential of host transcriptomic biomarkers as correlates of risk for leprosy. Importantly, a prospective five-gene signature for reversal reactions could predict reversal reactions at least 2 weeks before onset. Thus, transcriptomic biomarkers provide promise for early detection of these acute inflammatory episodes and thereby help prevent permanent neuropathy and disability in leprosy patients.

Introduction

Leprosy is a chronic infectious disease caused by *Mycobacterium leprae*, a bacillus with tropism for skin and peripheral nerves. Despite decades of programs using multidrug therapy (MDT), leprosy remains persistently endemic or re-emerging in some regions where it predominantly affects poor and marginalized people. A featuring aspect regarding leprosy diagnosis is the plateauing annual new case detection rates of roughly 200,000 worldwide (1). The pathology of leprosy is complex as it presents as a spectral disease in which immunity against *M. leprae* matches the clinical manifestations after infection with the bacterium. At one pole of the spectrum, the disease manifests as tuberculoid leprosy (TT), characterized by strong pro-inflammatory cellular immunity including Th1 and Th17 cells (2, 3), granuloma formation and elimination of bacteria. At the other pole, lepromatous leprosy (LL) is characterized by humoral immunity against *M. leprae* along with Th2 cells but almost no protective cell mediated immunity, allowing accumulation of high numbers of bacilli around foamy macrophages (4-8). Nonetheless, the majority of individuals present unstable borderline phenotypes (BT, BB and BL) between the two poles (5).

A major challenge in leprosy control is the prevention of permanent disability due to nerve damage. Although leprosy is curable by MDT, nerve damage cannot always be avoided. Dynamic and unpredictable episodes of increased inflammation, leprosy reactions, can occur before, during and even after treatment, with a higher likelihood to occur in adults than in children (9, 10). These immunological complications are the principal cause of leprosy-associated irreversible neuropathy and are experienced by 30-50% of leprosy patients one or more times, mostly in the unstable borderline lepromatous patients with substantial bacterial loads (11). Two types of reactions are recognized: reversal reactions or type 1 (RR) and erythema nodosum leprosum (ENL). RRs are caused by changes in the host immune response against *M. leprae* which is upgrading from borderline to the TT pole characterized by an enhanced cell-mediated immunity, inflammation (12, 13). These reactions can occur spontaneously but are also linked to shifts from Th2 to Th1, e.g. occurring during anti-helminth treatment of co-infected leprosy patients (14-17), HIV highly active antiretroviral therapy (HAART) and at the end of extensive anti-TNF-α therapy (10, 13) and even BCG vaccination (18).

Prompt diagnosis and treatment of reactions significantly favors successful recovery (9, 19). Unfortunately, reactions are often late- or misdiagnosed, in part due to decreased expertise within integrated health services (19) which urges the need for new, diagnostic tools. Delays in diagnosis of reactions directly translate into negative clinical outcomes, as associated neuropathy not properly diagnosed or treated within the first 6 months of symptoms will likely become permanent (20) alongside the disabilities it may later initiate via recurrent ulcers and other related pathologies (21). Despite recent scientific progress

with respect to complement (22, 23) and serum-proteins, particularly CXCL10 (IP-10), as biomarkers associated with onset of reactions (15-17, 24-26), discovery of accurate, clinically useful prognostic biomarkers remains elusive, leaving early diagnosis of reactions a currently unmet need.

Since host transcriptomic biomarkers reflect early stages of or ongoing biological processes, they have been widely used to profile the host transcriptome for diagnostics for tuberculosis (TB) (27-30). Moreover, multicomponent host biomarker signatures have been described that predict development of disease in retro- and prospective cohorts (31, 32). In this respect dual color Reverse-Transcription Multiplex Ligation-dependent Probe Amplification (dcRT-MLPA) has proven to be a valuable tool for monitoring gene expression profiles in large cohorts (29, 33). Techniques such as RNA-Seq and microarray are costly, technically challenging and require high RNA concentrations which limits their application for large cohorts. Therefore, a selection of genes related to immune-mediated inflammatory pathways, which play a role in the immunopathology of leprosy can be assessed by dcRT-MLPA (29, 34).

Many reactions occur during MDT, with the highest rates reported within the first 6 months of treatment (11, 19, 35). To identify transcriptomic signatures for applications to surveillance of leprosy reactions, whole blood RNA of leprosy patients was monitored during MDT. To accommodate worldwide applicability, this study was executed in four prospective cohorts in Asia, Africa and South America. Improved knowledge on longitudinal fluctuations of RNA-expression associated with reactions will promote identification of patients with imminent reactions leading to timely interventions that can impact nerve damage in affected individuals.

Materials and methods

Participants

Patients and controls were recruited on a voluntary basis between February 2008 and March 2015 (Table 1) in four leprosy endemic populations: in Bangladesh (International Centre for Diarrhoeal Disease Research Bangladesh, Dhaka), Brazil (National Reference Centre for Sanitary Dermatology and Leprosy, Uberlandia and Leprosy Laboratory, Oswaldo Cruz Institute, Oswaldo Cruz Foundation (FIOCRUZ), Rio de Janeiro), Ethiopia (ALERT hospital and Health Centre, Addis Ababa) and Nepal (Mycobacterial Research Laboratories, Kathmandu). Leprosy was diagnosed based on clinical, bacteriological and histological observations and classified by skin biopsies according to Ridley-Jopling (36). Clinical monitoring for reactions was performed during monthly clinic visits. Participant information was collected with emphasis on standardizing data collection and definition of reactions between all cohorts (37, 38). Endemic controls (EC) were living in the same area without known contact with leprosy or TB patients and were assessed for the absence of

clinical signs and symptoms of leprosy and TB. Staff of leprosy or TB clinics and laboratory staff were excluded. Healthy household contacts (HHC) were defined as adults living in the same household as leprosy patients for at least the preceding six months.

Table 1. Samples of participants included in cross-sectional analysis.

Site	Category[1]	RR	Timepoint	number
Bangladesh	EC	na	na	61
	BB/BL/LL (MB) no LR t=0	no LR	t=0	62
	BB/BL/LL (MB) no LR t= end	no LR	t=end	26
	TT/BT (PB) no LR t=0	no LR	t=0	36
	TT/BT (PB) no LR t =end	no LR	t=end	9
	BB/BL/LL (MB) RR t=0	RR	t=0	5
	BB/BL/LL (MB) RR t= x	RR	t=x	30
	BB/BL/LL (MB) RR t= end	RR	t=end	42
	TT/BT (PB) RR	RR	t=0 or=x, or =end	9
	ENL/Neuritis	ENL/Neuritis	t=0 or=x, or =end	11/0
	HHC	na	na	38
Brazil	EC	na	na	46
	BB/BL/LL (MB) no LR t=0	no LR	t=0	26
	BB/BL/LL (MB) no LR t= end	no LR	t=end	20
	TT/BT (PB) no LR t=0	no LR	t=0	52
	TT/BT (PB) no LR t =end	no LR	t=end	38
	BB/BL/LL (MB) RR t=0	RR	t=0	17
	BB/BL/LL (MB) RR t= x	RR	t=x	20
	BB/BL/LL (MB) RR t= end	RR	t=end	20
	TT/BT (PB) RR	RR	t=0 or=x, or =end	28
	ENL or Neuritis	ENL/Neuritis	t=0 or=x, or =end	22/18
	HHC	na	na	14
Ethiopia	EC	na	na	51
	BB/BL/LL (MB) no LR t=0	no LR	t=0	83
	BB/BL/LL (MB) no LR t= end	no LR	t=end	9
	TT/BT (PB) no LR t=0	no LR	t=0	16
	TT/BT (PB) no LR t =end	no LR	t=end	1
	BB/BL/LL (MB) RR t=0	RR	t=0	2
	BB/BL/LL (MB) RR t= x	RR	t=x	36
	BB/BL/LL (MB) RR t= end	RR	t=end	12

	TT/BT (PB) RR	RR	t=0 or=x, or =end	6
	ENL or Neuritis	ENL/Neuritis	t=0 or=x, or =end	11/1
	HHC	na	na	33
Nepal	EC	na	na	42
	BB/BL/LL (MB) no LR t=0	no LR	t=0	14
	BB/BL/LL (MB) no LR t= end	no LR	t=end	5
	TT/BT (PB) no LR t=0	no LR	t=0	19
	TT/BT (PB) no LR t =end	no LR	t=end	8
	BB/BL/LL (MB) RR t=0	RR	t=0	6
	BB/BL/LL (MB) RR t= x	RR	t=x	12
	BB/BL/LL (MB) RR t= end	RR	t=end	7
	TT/BT (PB) RR	RR	t=0 or=x, or =end	29
	ENL or Neuritis	ENL/Neuritis	t=0 or=x, or =end	2/0
	HHC	na	na	6
Nether-lands	NEC first time point	na	na	19
	NEC second time points	na	na	10
			total	**1090**

[1]EC: endemic control; BB/BL/LL: borderline borderline/ borderline lepromatous/ lepromatous leprosy; TT/BT: tuberculoid leprosy/ borderline tuberculoid leprosy; LR: leprosy reaction; RR: reversal reaction; NEC: non-endemic control; na: not applicable; t=0: time point of enrolment before initiation of multidrug (MDT) therapy; t=x: time point of LR; t=end: time point of completion of MDT and/ or steroid therapy.

Recruitment

Newly diagnosed, untreated leprosy patients without clinical reactions were enrolled and blood was drawn before initiation of multidrug (MDT) therapy (t=0) as previously described (15). Patients with reactions within initiation of three months of therapy were excluded. Patients were often diagnosed with RR at first clinic visits, leading to a low frequency of untreated cases without RR at their first visits that subsequently developed RR during this study. If patients presented with reactions after more than three months of MDT, blood was drawn again before initiation of anti-reactional therapy (t=x). Patients diagnosed with RR at their first clinic visits were also recruited (t=x) but blood was collected after completion of MDT and/or after steroid therapy (t=end). For patients with RR this was done at least one month after completion of steroid therapy. Patients were assessed for absence of reactions one year after t=end. For patients showing clinical signs of reactions within three months after t=end, this time point was excluded from analyses. Thus, analyses included two samples of each patient without reactions [before (t=0) and after

treatment (t=end)] and three of each patient who developed RR [in the absence of clinical signs of reactions, ≥2 months before RR diagnosis (t=0); at RR diagnosis, before steroid--treatment (t=x); after RR, at least one month after ending steroid-treatment (t=end)]. Patients with leprosy relapse and pure neural leprosy were excluded from the analysis.

Ethics

This study was performed according to the Helsinki Declaration. Written informed consent was obtained before enrolment. Patients received treatment according to national guidelines. Ethical approval of the study-protocol was obtained through Ethical Review Committee of ICDDR,B (#PR-10032; #PR-2007-069); Brazilian National Council of Ethics in Research (CONEP) and the Fiocruz Ethical research Council CEP (# 555/10) or UFU Research Ethics Committee (#499/08); National Health Research Ethical Review committee Ethiopia (NERC # RDHE/127-83/08); Nepal Health Research Council (NHRC #751).

RNA isolation

RNA from PAXgene tubes was extracted using PAXgene Blood RNA kits (BD Biosciences, Franklin Lakes, NJ) according to the manufacturers' protocol. RNA yield was determined by a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE).

Dual color Reverse-Transcription Multiplex Ligation-dependent Probe Amplification (dcRT-MLPA) assays

dcRT-MLPA assay was performed as previously described (33, 39). In short, for each target--specific sequence, a specific RT primer was designed located downstream of the half-probe target sequences (Sigma-Aldrich, Saint Louis, MO). RNA (2.5 µl of a 50 ng/µl solution) was reverse transcribed with 1x MMLV reverse transcriptase buffer, dNTPs (0.4 mM of each nucleotide), and 80 nM of the target-specific RT primers in a final volume of 4.5 µl. After heating for 1 min to 80°C and incubation for 5 min at 45°C, 30U MMLV reverse transcriptase (Promega, Madison, WI) was added and incubated for 15 min at 37°C before heat inactivation of the enzyme for 2 min at 98°C. Subsequently, half-probes (6 nM) were added to the reaction, heat denatured for 1 min at 95°C followed by hybridization for 16 h at 60°C. Ligation of the annealed half-probes was performed for 15 min at 54°C by ligase-65 followed by heat inactivation for 5 min at 98°C. Ligation products were amplified by PCR. Thermal cycling conditions encompassed: 33 cycles of 30s/95°C, 30s/58°C, and 60s/72°C, followed by 1 cycle of 20min/72°C. PCR products were diluted 1:10 in HiDi formamide containing 400 HD ROX size standards and analyzed on an Applied Biosystems 3730 capillary sequencer in GeneScan mode (Applied Biosystems, Foster City, CA). MLPA reagents were from MRC-Holland (Amsterdam, The Netherlands).

Data were analyzed using GeneMapper software 5 (Applied Biosystems). The areas of each assigned peak (in arbitrary units) were exported for further analysis in R-Project, normali-

zed to *GAPDH* and log2 transformed. Signals below the threshold value for noise cut-off (peak area ≤ 7.644) were assigned the threshold value for noise cut-off. Pathway analysis was performed using Ingenuity Pathway Analysis (Qiagen, Hildern, Germany).

Statistical analyses dcRT-MLPA

Wilcoxon signed-rank test was applied for paired samples. For cross-sectional comparisons of different test groups at comparable time points, a Mann–Whitney test was used. P-values were corrected for multiple comparisons using the Benjamini-Hochberg method (40). To classify BL/LL patients according to their likelihood to develop RR a logistic regression model was fitted with RR (yes/no) as outcome (dependent variable), and gene expression values at time t=0 as covariates (risk factors). Genes were grouped based on correlation of their expression using a hierarchical clustering analysis (average linkage) based on absolute correlation difference. The global test (version 5.32.0) (41) in R (version 3.4.1) was performed on these groups. Genes that were significant differently expressed between the test groups (inheritance < 0.05) after multiple testing correction, constituted a biomarker signature for prediction of RR. To avoid overfitting, the biomarker signature was evaluated using a leave-one-out cross-validation (LOOCV): during the training of the model, one subsample (n-1) was reserved in each iteration for evaluation of the accuracy of the model on a sample excluded from the training set. The gene selection for the predictive signature was redone at every fold. The predictive biomarker signature was thus assessed in observations, which were not used to build the model. To evaluate the risk of developing RR, the cut-off for gene expression was determined based on the Youden Index. A score of 0 or 1*(weight of gene) was given per gene based on the association to RR (as indicated by the cut-off). The weight of gene was based on the results from the global test. All the scores from the significant genes were added and divided by the sum of weights to calculate the risk to develop RR. Area Under the Receiver Operating Characteristic curve (ROC-AUC) was calculated using GraphPad Prism (version 7.02).

Results

Prospective cohorts

To identify correlates of risk (CoR) for leprosy and RR, blood of 1,090 samples was obtained longitudinally in Bangladesh, Brazil, Ethiopia and Nepal (Table 1-2).

Leprosy-specific RNA-profiles

We first analyzed gene expression in *ex vivo* blood samples of newly diagnosed leprosy patients (irrespective of classification) without reactions (n=359) from Bangladesh, Brazil, Ethiopia and Nepal compared to EC from these regions (n=200). To this end, 103 target genes associated with innate and adaptive immunity (29, 33, 39) were analyzed by dcRT-MLPA (Table S1). A substantial variety (36 genes) was observed to significantly differ between

patients and EC at all sites. Expression of 13 genes was upregulated and 23 genes were downregulated in patients compared to EC (Table S2). When comparing leprosy patients to HHC (Table S3), 16 genes showed significantly different expression for leprosy patients (increased: *FCGR1A*, *IL6*, *IL15*, *LRKK2*, *MBP*, *MSR1*, *PACRGv1*, *TLR1*, *TLR4*; decreased: *CAMTA*, *CD3E*, *CTLA4*, *CXCL13*, *GATA3*, *LAG3*, *TFGB*). Importantly, whilst most of these genes were also differently expressed in leprosy patients compared to EC, *MBP*, *MSR1, TLR1, CAMTA, CXCL13* and *TFGB* were differentially expressed in leprosy patients exclusively when compared to HHC. Such genes are potential CoR for leprosy in contacts of leprosy patients who are highly exposed to *M. leprae*. Being part of innate/adaptive and macrophage signaling pathways (42), these genes, are not restricted to leprosy but are also relevant in rheumatoid arthritis and Crohn's disease (Table S4).

Table 2. Demographics and clinical characteristics of reactional patients recruited longitudinally[§].

#	area	RJ	LR	BI*	PGL-I*	sex	age*
1	Bangladesh	BL	RR	2+	0.95	male	36
2	Bangladesh	BL	RR	2+	0.27	female	39
3	Bangladesh	BL	RR	0	1.82	male	36
4	Bangladesh	BL	RR	2+	1.17	male	32
5	Brazil	BL	RR	3.2+	0,43	male	35
6	Brazil	BL	RR	2.42+	0.07	male	29
7	Brazil	BL	RR	4.28+	1.02	female	42
8	Nepal	BT	RR	0.25+	0.07	female	40
9	Ethiopia	BB	RR	0	0.15	male	33
10	Netherlands	BL	RR	5+	1.83	male	17

[§]*Blood samples were collected from 10 patients at 3 time points: at diagnosis of leprosy in the absence of any clinical signs of reactions, at diagnosis of reactions and after treatment; * at recruitment before treatment.*

RJ: Ridley Joplin classification; LR: Leprosy Reaction; RR: reversal reaction.

BI: bacterial index; PGL-I: OD at 450nm in anti-PGL-I antibody ELISA (threshold for positivity: $OD_{450} = 0.2$).

RNA-profiles for leprosy classification

To identify genes applicable for classification of different types of leprosy, the 103 target genes (Table S1) were assessed similarly by dcRT-MLPA in blood from newly diagnosed, untreated BL/LL (n=228) as well as TT/BT (n=131) leprosy patients from the four-different leprosy endemic populations. After correction for multiple comparisons (40) seven genes remained significantly different. Expression levels of *IL2* and *TLR6* were increased in TT/BT compared to BL/LL whereas *CD46, CXCL10, FCGR1A, HDAC2* and *TLR4* expression levels were increased in BL/LL (Figure 1; Table S5).

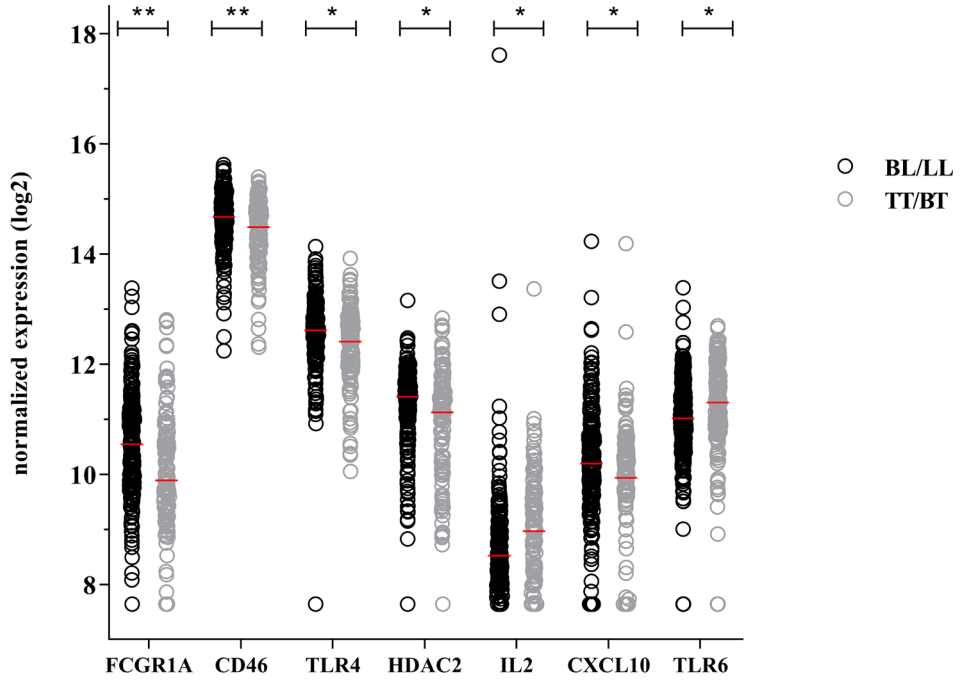


Figure 1. Transcriptional profiles of BL/LL and TT/BT patients without reactions.

Gene expression levels of 103 target genes were assessed by dual-color RT-MLPA performed on *ex vivo* RNA isolated from whole blood of newly diagnosed, untreated BL/LL (n=228; black circles) and TT/BT (n=131; grey circles) leprosy patients without reactions from Bangladesh, Brazil, Ethiopia and Nepal. Log2-transformations of peak areas (normalized to the housekeeping gene *GAPDH*) of genes that were significantly differentially expressed between BL/LL and TT/BT are shown on the *y*-axis. Raw p-values were calculated using the Mann–Whitney test and adjusted for multiple comparisons using the Benjamini-Hochberg correction (40).

*: adjusted p-values <0.05; **: adjusted p-values <0.01; (see Table S5).

RNA-profiles associated with exposure to *M. leprae*

Since HHC of leprosy patients have a higher risk to develop leprosy than the general population in an endemic area (43), biomarker profiles indicating this risk could help decision making on who needs preventive antibiotic treatment (44). These genes represent potential transcriptomic tools to identify individuals substantially exposed to *M. leprae*. Thus, RNA-expression profiles of HHC (n=83) were compared to EC (n=200) (Table S6). We identified 10 differentially expressed genes with either significantly higher expression in HHC (Figure 2; *FOXP3, TGFB* and *CCL3*) or in EC (*CCR6, GZMA, HDAC2, IL22RA1, PTPRCv2, TLR1* and *TLR7*).

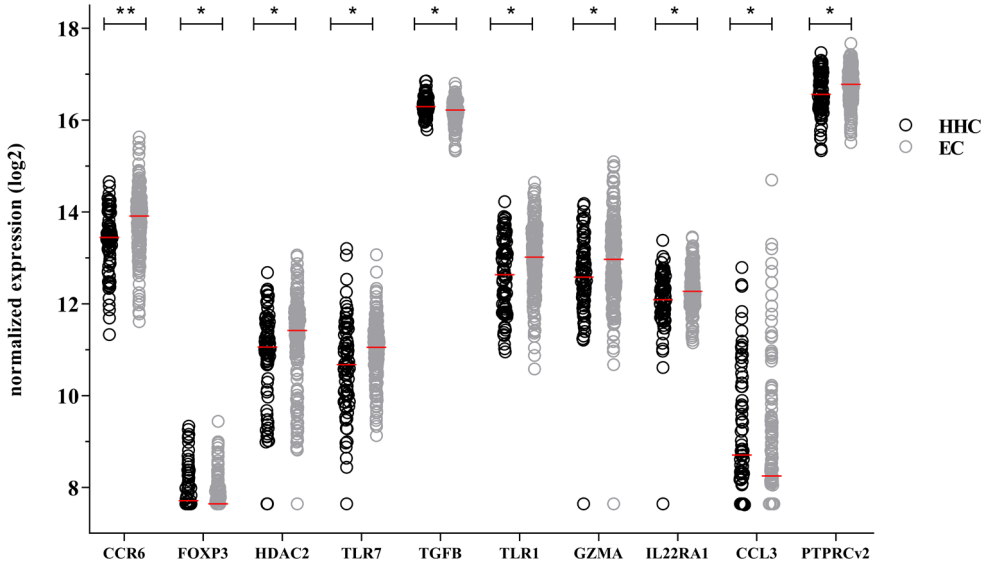


Figure 2. Transcriptional profiles of healthy household contacts and endemic controls.

Gene expression levels of 103 target genes were assessed by dual-color RT-MLPA performed on *ex vivo* RNA isolated from whole blood of healthy household contacts (HHC; n=83; black circles) and endemic controls (EC; n=200; grey circles) from Bangladesh, Brazil, Ethiopia and Nepal. Log2-transformations of peak areas (normalized to *GAPDH*) of genes significantly differentially expressed between HHC and EC are shown on the *y*-axis. Raw p-values were calculated using the Mann–Whitney test and adjusted for multiple comparisons using the Benjamini-Hochberg correction (40).

*: adjusted p-values <0.05; **: adjusted p-values <0.01; (see Table S6).

Transcriptomic risk factors for development of RR

To assess whether RNA-expression levels can be used as a predictive tool for reactions during MDT, cross-sectional comparison of gene expression levels was performed for samples at the time of leprosy diagnosis in the absence of clinical signs of reactions (t=0). Gene expression from BL/LL patients who developed RR ≥2 months later in the study (n=30) was compared to BL/LL patients who did not develop reactions at all (n=184) using the Mann–Whitney test. Transcriptomic profiles of the two groups of BL/LL patients resulted in a decreased expression of *CTLA4* and *GATA3* at diagnosis in patients who would later develop RR (Figure 3), whereas nine genes (*CCL2, IL2, IL15, IL18, MARCO, PHEX, TLR2, TLR6* and *TLR10*) were significantly increased (Figure 3; Table S7).

Next, we also determined the biomarker signature with the least number of genes and the highest discriminatory power to categorize BL/LL patients according to their likelihood to develop RR using the global test (41) (Figure 4; $p=1.28*10^{-5}$). Four genes (*CCL2, IL2, IL15* and *MARCO*) remained significantly (inheritance < 0.05) associated with occurrence

of RR at the time of recruitment, whereas *CD8A* was negatively associated with RR (Figure 3B). Although expression of CD8A was not statistically significant using Mann–Whitney, it contributed significantly to the transcriptomic global test-signature. From the five genes *CCL2* contributed most to the model (Figure 4B).

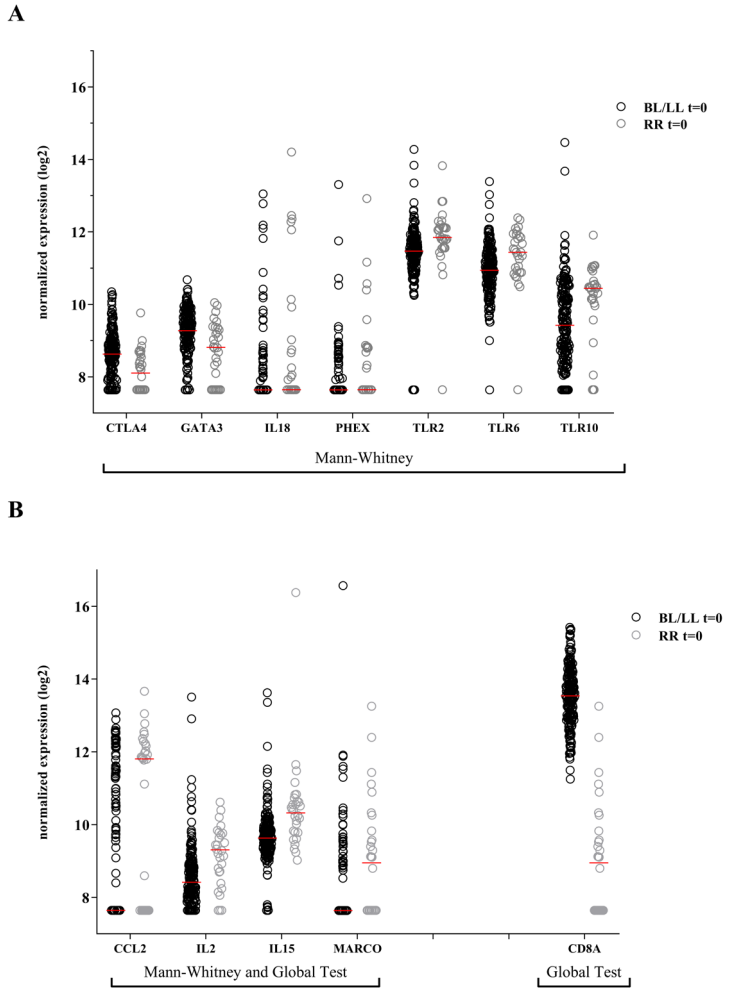
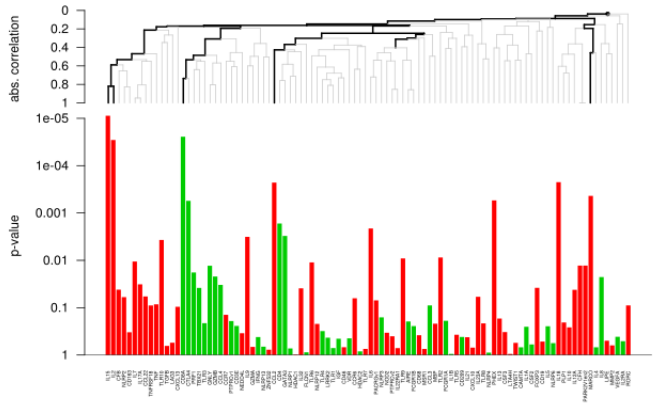


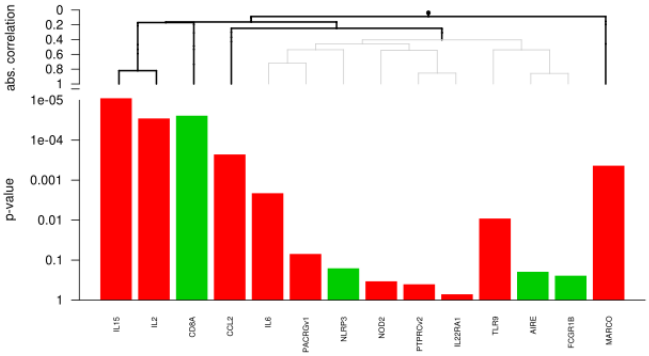
Figure 3. Identification of biomarker risk signature for developing reversal reactions.

Gene expression levels of 103 target genes were assessed by dual-color RT-MLPA performed on *ex vivo* RNA isolated from whole blood from BL/LL patients who developed reversal reactions (RR) at least two months later during the study (n=30; grey circles) and BL/LL who did not develop reactions (n=184; black circles) from Bangladesh, Brazil, Ethiopia and Nepal. Samples were analyzed at t=0: in the absence of clinical signs of reactions. Log2-transformations of peak areas (normalized for *GAPDH* expression) of genes with significantly different expression between both groups (at t=0) are shown on the *y*-axis. (**A**) Genes with a significant (p-value <0.05) different expression only using the Mann-Whitney test are show. P-values were adjusted for multiple comparisons using the Benjamini-Hochberg correction (40) (see Table S7). (**B**) Genes with a significant different expression in the global test and Mann-Whitney or the global test only are shown.

A



B



C

Global test

Gene	inheritance	p-value	weight
CCL2	0.00521	2.28e-04	0.666935
IL15	0.00783	8.82e-06	0.139147
IL2	0.00783	2.87e-05	0.188963
MARCO	0.02483	4.35e-04	0.253816
CD8A	0.02596	2.45e-05	0.170529

Figure 4. Identification of a minimal biomarker risk signature for developing reversal reactions.

Biomarker signature to assess the risk of BL/LL patients to develop reversal reactions (RR). Gene expression data obtained by dual-color RT-MLPA of RNA isolated from whole blood of BL/LL patients from Bangladesh, Brazil, Ethiopia and Nepal at t=0 were analyzed using the global test cluster analysis(41). The global test is a cluster analysis based on absolute correlation difference and average linkage developed for data sets in which many covariates (or features) have been measured for the same subjects, together with a response variable. Graphs (**A-B**) indicate genes that are higher expressed in future RR patients (red) or in non-reactional BL/LL patients (green). In (**A**) all genes analyzed are shown and in (**B**) figure only significant branches are shown. Table (**C**) shows values for the 5 genes that were statistically significant after correction for multiple testing (inheritance <0.05), representing the output signature of the global test shrinkage model.

To evaluate the prediction a LOOCV was performed in which in every iteration a subsample (n=1) was excluded in the global test to evaluate the model performance on unseen samples. The overall classifying ability of the biomarker-signature is indicated as a ROC--AUC=0.80 (Figure S1). Thus, transcriptomic profiles can prospectively differentiate, at the time of leprosy diagnosis in the absence of any clinical symptoms of reactions, patients who will develop RR.

Longitudinal transcriptomic changes: monitoring RR onset and treatment

Cross-sectional analysis of gene expression at different time points amongst patients with RR showed only a significant increase in IL10 at RR compared to before RR (Supplementary results; Tables S8-S9; Figures S2-S3).

Since RNA-expression levels of genes may vary over time, RNA-expression was also assessed longitudinally at three time points in whole blood of 10 leprosy patients (Table 2) developing RR during the study (Figures 5, S4-S5). This included besides 103 immune--associated genes, 38 IFN-induced genes (Table S1) previously identified as markers for mycobacterial disease (45-49). Expression of ten genes (*CXCL10, FCGR1A, IFI16, IFI44, IFI35, IFI44L, IFI6, IFIH1, IL15 and OAS1*) significantly differed when comparing time points before RR vs at RR as well as at RR vs after RR. All genes except *IL15* consistently increased with development of RR, and normalized after treatment, whereas *IL15* decreased at onset of RR (Figure 5). Most of the genes identified longitudinally are directly connected to the genes identified as predictive markers before clinical symptoms of RR (Figure S6).

Five genes (*IL10, PRF1, CCL2, CCL3* and *BMP6*) were significantly different only at onset of RR compared to before RR (Figure S4), while nine (*IFITM3, IFIT3, GBP5, GBP1, GBP2, OAS3, STAT1, STAT2* and *TAP1*) only displayed a significant effect upon treatment (Figure S5).

These data indicate that onset of RR can be monitored, based on differential expression of inflammatory genes in whole blood, allowing early detection and subsequent treatment of RR. Thus, helping to reduce irreversible nerve damage and associated disabilities.

Discussion

In the current state of leprosy elimination, reactions persist as a major problem since patients remain at risk due to *M. leprae* antigens that can persist for years post-MDT. RRs can occur at any time during, before or after MDT (12) and although several factors have been associated with reactions, the underlying mechanism is not completely known (10). Genetic susceptibility to reactions but also treatment for other diseases or co-morbidities have been suggested to play a role by causing an immunological shift from Th2 to Th1 (10, 13-17). In addition, Th17 cells have been proposed to play a role in RR pathogenesis as well (50). Still, no validated biomarker signature is available at the moment. The implementation of diagnostics tests for reactions in leprosy health care could make significant

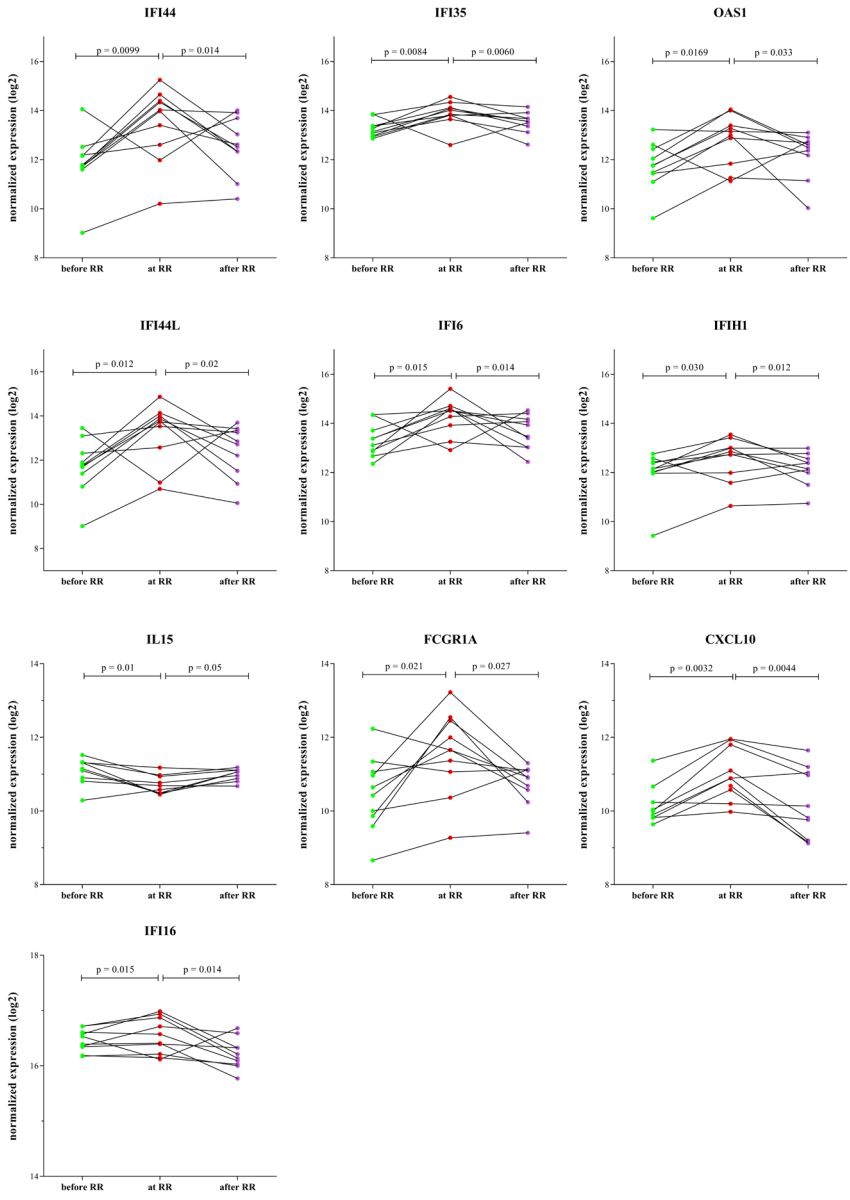


Figure 5. Intra-individual longitudinal expression of patients who developed RR.

Direct *ex vivo* RNA-expression values were assessed by dual-color RT-MLPA on whole blood of 10 leprosy patients who developed RR during this study. Blood was analyzed at three time points: in the absence of any clinical signs of reactions and at least two months before RR (t=0), at RR diagnosis before steroids (t=x) or after MDT and at least one month after end of steroids, in the absence of reactions (t=end). Log2-transformations of peak areas (normalized for *GAPDH* expression) are shown on the *y*-axis. Wilcoxon signed-rank test was performed. Genes with a significant difference (p-value <0.05) in expression between before RR-at RR and at RR-after RR are shown.

differences in clinical outcomes and help reduce nerve damage. Transcriptomic biomarker signatures provide indications as to how immune responses are oriented and instructed to develop reactions (31, 32). Due to the available samples set we have focused on the analysis of biomarkers for RR, biomarkers for CoR of leprosy, leprosy classification and *M. leprae* exposure have also been investigated.

This is the first multi-site, prospective study analyzing unstimulated whole blood-derived RNA-expression profiles of four endemic populations using dcRT-MLPA, a focused gene--expression profiling platform for monitoring gene expression in large cohorts (29, 34). We showed that host RNA-expression levels discriminate various stages of disease thereby offering potential for leprosy diagnostics. Since differentiating high exposure to *M. leprae* and early leprosy is difficult using serum proteins (51), the here identified transcriptomic biomarkers provide potential to identify cases among contacts in high endemic regions. A limitation of the current study is the lack of follow-up of HHC in order to validate the identified biomarkers of *M. leprae* infection as CoR for leprosy. However, this is currently addressed in ongoing studies. Moreover, validation in an independent cohort will provide more evidence on whether transcriptomic signatures may be applied to guide prophylactic strategies by discriminating contacts highly exposed to *M. leprae* and at risk of developing disease.

We have identified genes differentially expressed between leprosy patients and EC or HHC. Whilst most of the genes overlap between the two comparisons, some genes discriminate leprosy patients from HHC exclusively (*MBP*, *MSR1, TLR1, CAMTA, CXCL13* and *TFGB*). Thus, these genes could be correlated to development of disease. Additionally, we observed that a classification of leprosy patients is possible using transcriptomics since a set of genes were increased in BL/LL patients (*CD46, CXCL10, FCGR1A, HDAC2* and *TLR4*) and TT/BT showed a higher expression of *IL2* and *TLR6.*

Importantly, we identified a five-gene CoR signature (*CCL2, CD8A, IL2, IL15*, *MARCO*) for RR differentiating those developing RR ≥2 months prior to clinical symptoms. *CCL2*, was downregulated in leprosy patients without reactions compared to EC, in line with lower *CCL2* expression observed in nerves of leprosy patients compared to patients with non-leprous peripheral neuropathy (52). We observed a higher expression of *CCL2* in future RR patients compared to patients who did not develop RR, as well as a longitudinal increase during RR onset. Similarly, a Brazilian study analysing 90 immune related genes, showed that *CCL2* had the highest fold change in expression levels between leprosy patients with and without RR (53) and increased *CCL2* expression in RR patients has also been described in other studies (35, 54). *CCL2* is associated with excessive deposit of extracellular matrix and macrophage recruitment (55) which can be due to an increase of *M. leprae* antigens presented to the immune system after MDT. This leads to activation of pro-inflammatory

cytokines and attraction of $CD4^+$ T cells as confirmed by the similarly increased *IL2* expression in future RR patients. Upregulation of *IL2* and *IL15*, as well as downregulation of *CTLA4* and *GATA3* decreases regulation of Th2 and the lack of regulation leads to exacerbation of Th1, which is common in RR. Higher expression of *MARCO* in future RR patients is also in line with increased antigen presence. Differences in the expression of this scavenger receptor were also identified in several other diseases such as giant cell arteritis (56), enthesitis-related arthritis (57) and lupus (58) suggesting a general role for antigenic triggers in respective disease etiology. *IL15* which encodes a cytokine important for cytotoxic T-cell proliferation and increases *GZMB* expression (59), was higher expressed in future RR patients, but decreased during RR in the longitudinal analysis. Upregulation of *IL15* in patients who will develop RR may lead to an increase of cytotoxic T-cells and tissue destruction. *CD8A* is found on cytotoxic T lymphocytes, macrophages and dendritic cells leading to tissue damage in leprosy (60). Even though we did not find significant difference in the expression before and at RR, previous longitudinal analysis showed increased expression at RR onset (39). Thus, the decreased expression of *CD8A* as observed here in future RR patients when they still lack reactional symptoms (cross-sectional comparison) may indicate the number of $CD8^+$ cells like cytotoxic T cells are increasing during RR development.

To evaluate the role of reaction-associated biomarkers as etiological and early-disease--prediction targets, temporal associations are implicit to indicate the utility of novel biomarkers for application in diagnostic tests. In this multi-site study, RNA-expression was therefore also analyzed longitudinally: before, during reaction and after treatment. IFN-induced genes (*IFI44, IFI35, IFI44L, IFI6, IFIH1, CXCL10 and FCGR1A*) showed high expression at RR which decreased during reactional treatment; whereas *IL15* expression decreased at RR. These biomarkers corresponding with increased inflammation can be useful when monitoring patients during their monthly dose of rifampicin helping to timely detect incipient RR. However, larger longitudinal cohorts should be studied prospectively to confirm our findings.

Previously, using longitudinal analysis of blood and skin samples of one leprosy patient who developed RR, a candidate blood-derived CoR for RR was identified composed of genetic host factors associated with T-cell cytotoxicity, regulation, vasculogenogenesis and IFN-signaling (15). An interferon-dominant signature has also been identified for TB, first in 2010 (45) and in multiple subsequent studies (47, 49, 61). In line with our findings, higher levels of *CXCL10* have also been reported in association with episodes of RR (16, 54). *FCGR1A* can differentiate between active and latent tuberculosis, suggesting a major role in the immune response against mycobacterial diseases (34). Both *CXCL10* and *FCGR1A* showed higher RNA expression during reaction as well as in BL/LL patients compared to TT/BT patients. These markers of innate immunity (*CXCL10)* and infection

(*FCGR1A)* thus are increased in two different phases in leprosy associated either with *M. leprae*-specific T cell anergy and high bacterial load (BL/LL) or with a highly inflammatory state (RR). *CXCL10* is associated with Th1 responses occurring during RR (7, 13, 16). However, CXCL10-producing monocytes are induced during mycobacterial infections (62), in line with the observed increased expression in BL/LL patients with high bacillary load. Therefore, monitoring transcriptomic changes in an individual is relevant as one marker in various conditions may reflect a different disease process.

In this study 103 target genes for innate- and adaptive immune profiling were investigated in view of the immune mediated nature of the pathology of leprosy which strongly correlates with individuals' immune responses against the bacterium. This selection could limit the discovery of novel genes that could potentially be used as biomarkers for RR. However, in contrast to studies focused at identification of disease mechanisms, for diagnostic purposes only a limited amount of discriminating genes is sufficient, leaving out genes which strongly correlate and hence do not have added value to the signature. Further in- depth transcriptomic analysis of the samples described in this study have been performed by RNASeq in a separate study with the aim to identify additional genes providing increased insight into mechanisms and pathways involved in RR (63).

Although geographic differences were observed, our study showed that in prospective analysis of 4 different leprosy endemic areas, several genes are associated with onset of reactions at each site. Further analysis of the diagnostic signatures in extended cohorts worldwide need to be performed to validate the performance of the genes signature.

Transcriptomic analyses have shown that TB (31, 32) disease can be characterized by 16 (64) to as few as 2-4 genes (28). A gene set signature of reversal reaction consisting of 44 genes was previously described for the Vietnamese population using *M. leprae*-sonicate stimulated whole blood (35). Their set signature included pro-inflammatory regulator genes such as *CCL2* or *IL1A* in RR. We observed a decreased expression of IL1A after treatment for RR, however we did not find significant differences in patients who later developed RR compared to patients who did not have a reaction. We did not find significant differences either in other genes of that set such as *IL1B*, *IL6*, *IL23A*, *CCL3* or *CCL4*. Another study instead using PBMC of leprosy patients with and without reactions identified a role for complement-associated genes (54). They also found that *CCL2* and *MARCO*, which are part of our signature biomarker, were differentially expressed in RR patients, as well as an interferon γ significant upregulation during RR, which is in line with the increased expression of IFN-signaling genes we found longitudinally and our RNAseq data (63), which shows an enrichment at timepoint of reaction for both IFN- γ and IFN-β pathways.

The here identified biomarker signature is based on unstimulated whole blood and identifies development of RR ≥2 months prior to clinical symptoms using a five-gene signature

across four leprosy endemic populations on 3 different continents. Unstimulated whole blood and a signature with a small number of markers is more suitable for a point-of-care (POC) use as a field-friendly test. Early diagnosis and treatment of reactions are currently the primary research targets for reducing permanent neuropathy and disability development in future patients. The ability to predict reactions ≥2 months before development of clinical symptoms, using POC diagnostic tests detecting transcriptomic biomarkers for RR, would represent momentous advancements in global leprosy health care. Next steps include prospective evaluation and translation to clinically useful tools that can be implemented by clinicians. The challenge for the academic community and industry is to develop innovative methods to translate multi-transcript signatures into low-cost tests for leprosy diagnostics, suitable for use in health facilities in leprosy endemic areas.

Acknowledgements

The authors gratefully acknowledge all patients for their participation. AHRI, Fiocruz, ICDDR,B, LUMC and MRL Anandaban are part of the IDEAL (***I***nitiative for ***D***iagnostic and ***E***pidemiological ***A***ssays for ***L***eprosy) Consortium. We are indebted to Mr. Kapil Dev Neupane (Anandaban); Dr. J.A. Nery, A.M. Sales, Dr. Sarno (Fiocruz), Dr. Sheikh Abdul Hadi (Dhaka); S/r Genet Amare (Addis Ababa), S/r Haregewoin Yetesha, Mr. Alemayehu Kifle and Dr. Saba M. Lambert for recruitment of study participants and sample collection.

Funding Statement

This study was supported by the Order of Malta-Grants-for-Leprosy-Research (MALTALEP), the Heiser Program for Research in Leprosy in The New York Community Trust (P13-000392), the Q.M. Gastmann-Wichers Foundation, the UNICEF/UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases (TDR) and the Netherlands Leprosy Relief Foundation (NLR; ILEP#: 702.02.68, 7.01.02.48 and 701.02.49) together with the Turing Foundation.

Conflicts of Interest

The authors declare that they have no conflict of interest.

References

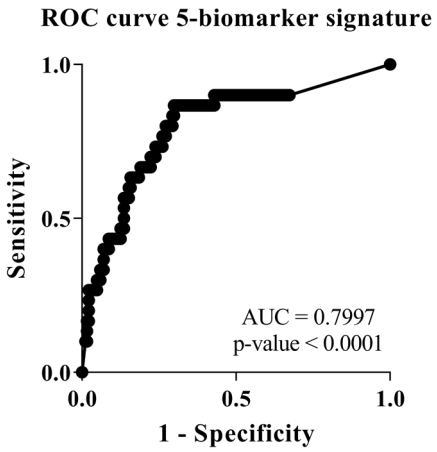
1. WHO. Global leprosy update, 2017: reducing the disease burden due to leprosy. Weekly epidemiological record. 2018;93:11.
2. Barnes PF, Lu S, Abrams JS, Wang E, Yamamura M, Modlin RL. Cytokine production at the site of disease in human tuberculosis. Infect Immun. 1993;61(8):3482-9.
3. Saini C, Ramesh V, Nath I. CD4+ Th17 cells discriminate clinical types and constitute a third subset of non Th1, Non Th2 T cells in human leprosy. PLoS Negl Trop Dis. 2013;7(7):e2338.
4. Modlin RL, Melancon-Kaplan J, Young SM, Pirmez C, Kino H, Convit J, et al. Learning from lesions: patterns of tissue inflammation in leprosy. Proc Natl Acad Sci U S A. 1988;85(4):1213-7.
5. Scollard DM. Pathogenesis and Pathology of Leprosy. In: Scollard DM, Gillis TP, editors. International textbook of leprosy. www.internationaltextbookofleprosy.org2016.
6. Misch E, Berrington W, Vary J, Hawn T. Leprosy and the Human Genome. Microbiology and

molecular biology reviews : MMBR. 2010;74:589-620.
7. Geluk A. Biomarkers for leprosy: would you prefer T (cells)? Leprosy review. 2013;84(1):3-12.
8. Geluk A. Challenges in immunodiagnostic tests for leprosy. Expert Opin Med Diagn. 2013;7(3):265-74.
9. Lockwood DN, Saunderson P. Nerve damage in Leprosy: a continuing challenge for scientists, clinicians and service providers. Int Health. 2012;4(2012):77-85.
10. Scollard DM, Martelli CM, Stefani MM, Maroja MF, Villahermosa L, Pardillo F, et al. Risk factors for leprosy reactions in three endemic countries. Am J Trop Med Hyg. 2015;92(1):108-14.
11. Antunes DE, Ferreira GP, Nicchio MV, Araujo S, Cunha AC, Gomes RR, et al. Number of leprosy reactions during treatment: clinical correlations and laboratory diagnosis. Rev Soc Bras Med Trop. 2016;49(6):741-5.
12. Ridley DS, Radia KB. The histological course of reactions in borderline leprosy and their outcome. International journal of leprosy and other mycobacterial diseases : official organ of the International Leprosy Association. 1981;49(4):383-92.
13. Geluk A. Correlates of immune exacerbations in leprosy. Semin Immunol. 2018;39:111-8.
14. Hagge DA, Parajuli P, Kunwar CB, Rana D, Thapa R, Neupane KD, et al. Opening a Can of Worms: Leprosy Reactions and Complicit Soil-Transmitted Helminths. EBioMedicine. 2017;23:119-24.
15. Khadge S, Banu S, Bobosha K, JJ vdP-vS, Goulart IM, Thapa P, et al. Longitudinal immune profiles in type 1 leprosy reactions in Bangladesh, Brazil, Ethiopia and Nepal. BMC Infect Dis. 2015;15(1):477.
16. Scollard DM, Chaduvula MV, Martinez A, Fowlkes N, Nath I, Stryjewska BM, et al. Increased CXC ligand 10 levels and gene expression in type 1 leprosy reactions. Clin Vaccine Immunol. 2011;18(6):947-53.
17. Corstjens PL, van HA, Tjon Kon Fat EM, van den Eeden SJ, Wilson L, Geluk A. Field-Friendly Test for Monitoring Multiple Immune Response Markers during Onset and Treatment of Exacerbated Immunity in Leprosy. Clin Vaccine Immunol. 2016;23(6):515-9.
18. Richardus RA, Butlin CR, Alam K, Kundu K, Geluk A, Richardus JH. Clinical manifestations of leprosy after BCG vaccination: an observational study in Bangladesh. Vaccine. 2015;33(13):1562-7.
19. Raffe SF, Thapa M, Khadge S, Tamang K, Hagge D, Lockwood DN. Diagnosis and treatment of leprosy reactions in integrated services - the patients' perspective in Nepal. PLoS Negl Trop Dis. 2013;7(3):e2089.
20. Croft RP, Nicholls PG, Steyerberg EW, Richardus JH, Withington SG, Smith WC. A clinical prediction rule for nerve function impairment in leprosy patients-revisited after 5 years of follow-up. Leprosy review. 2003;74(1):35-41.
21. Assis BPN, Lyon S, Grossi MAF, Rocha M. Risk factors for physical disability upon release from multidrug therapy in new cases of leprosy at a referral center in Brazil. Rev Inst Med Trop Sao Paulo. 2019;61:e13.
22. Bahia El Idrissi N, Hakobyan S, Ramaglia V, Geluk A, Morgan BP, Das PK, et al. Complement activation in leprosy: a retrospective study shows elevated circulating terminal complement complex in reactional leprosy. Clin Exp Immunol. 2016;184(3):338-46.
23. Negera E, Walker SL, Lema T, Aseffa A, Lockwood DN, Dockrell HM. Complement C1q expression in Erythema nodosum leprosum. PLoS Negl Trop Dis. 2018;12(3):e0006321.
24. Fava VM, Sales-Marques C, Alcais A, Moraes MO, Schurr E. Age-Dependent Association of TNFSF15/TNFSF8 Variants and Leprosy Type 1 Reaction. Frontiers in immunology. 2017;8:155.
25. Stefani MM, Guerra JG, Sousa AL, Costa MB, Oliveira ML, Martelli CT, et al. Potential plasma markers of Type 1 and Type 2 leprosy reactions: a preliminary report. BMC Infect Dis. 2009;9:75.
26. Iyer A, Hatta M, Usman R, Luiten S, Oskam L, Faber W, et al. Serum levels of interferon-gamma, tumour necrosis factor-alpha, soluble interleukin-6R and soluble cell activation markers for monitoring response to treatment of leprosy reactions. Clin Exp Immunol. 2007;150(2):210-6.
27. Maertzdorf J, McEwen G, Weiner J, III, Tian S, Lader E, Schriek U, et al. Concise gene signature for point-of-care classification of tuberculosis. EMBO Mol Med. 2015;8(2):86-95.
28. Sutherland JS, Loxton AG, Haks MC, Kassa D, Ambrose L, Lee JS, et al. Differential gene expression of activating Fcgamma receptor classifies active tuberculosis regardless of human immunodeficiency virus status or ethnicity. Clin Microbiol Infect. 2014;20(4):O230-O8.
29. Haks MC, Goeman JJ, Magis-Escurra C, Ottenhoff TH. Focused human gene expression profiling using dual-color reverse transcriptase multiplex ligation-dependent probe amplification. Vaccine. 2015;33(40):5282-8.
30. Anderson ST, Kaforou M, Brent AJ, Wright VJ, Banwell CM, Chagaluka G, et al. Diagnosis of childhood tuberculosis and host RNA expression in Africa. N Engl J Med. 2014;370(18):1712-23.
31. Sloot R, MF SvdL, van Zwet EW, Haks MC, Keizer ST, Scholing M, et al. Biomarkers Can Identify Pulmonary Tuberculosis in HIV-infected Drug Users Months Prior to Clinical Diagnosis. EBioMed-

icine. 2015;2(2):172-9.
32. Zak DE, Penn-Nicholson A, Scriba TJ, Thompson E, Suliman S, Amon LM, et al. A blood RNA signature for tuberculosis disease risk: a prospective cohort study. Lancet. 2016;387(10035):2312-22.
33. Joosten SA, Goeman JJ, Sutherland JS, Opmeer L, de Boer KG, Jacobsen M, et al. Identification of biomarkers for tuberculosis disease using a novel dual-color RT-MLPA assay. Genes Immun. 2012;13(1):71-82.
34. Gebremicael G, Kassa D, Quinten E, Alemayehu Y, Gebreegziaxier A, Belay Y, et al. Host Gene Expression Kinetics During Treatment of Tuberculosis in HIV-Coinfected Individuals Is Independent of Highly Active Antiretroviral Therapy. J Infect Dis. 2018;218(11):1833-46.
35. Orlova M, Cobat A, Huong NT, Ba NN, Van TN, Spencer J, et al. Gene set signature of reversal reaction type I in leprosy patients. PLoS Genet. 2013;9(7):e1003624.
36. Ridley DS, Jopling WH. Classification of leprosy according to immunity. A five-group system. Int J Lepr Other Mycobact Dis. 1966;34(3):255-73.
37. Roche PW, Le MJ, Butlin CR. Risk factors for type 1 reactions in leprosy. Int J Lepr Other Mycobact Dis. 1997;65(4):450-5.
38. Walker SL, Nicholls PG, Butlin CR, Nery JA, Roy HK, Rangel E, et al. Development and validation of a severity scale for leprosy type 1 reactions. PLoS Negl Trop Dis. 2008;2(12):e351.
39. Geluk A, van Meijgaarden KE, Wilson L, Bobosha K, JJ vdP-vS, van den Eeden SJ, et al. Longitudinal immune responses and gene expression profiles in type 1 leprosy reactions. J Clin Immunol. 2014;34(2):245-55.
40. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. J R Statist Soc B. 1995;57(1):289-300.
41. Goeman JJ, van de Geer SA, de KF, van Houwelingen HC. A global test for groups of genes: testing association with a clinical outcome. Bioinformatics. 2004;20(1):93-9.
42. Zhang FR, Huang W, Chen SM, Sun LD, Liu H, Li Y, et al. Genomewide association study of leprosy. N Engl J Med. 2009;361(27):2609-18.
43. Moet FJ, Meima A, Oskam L, Richardus JH. Risk factors for the development of clinical leprosy among contacts, and their relevance for targeted interventions. Lepr Rev. 2004;75(4):310-26.
44. Gillini L, Cooreman E, Wood T, Pemmaraju VR, Saunderson P. Global practices in regard to implementation of preventive measures for leprosy. PLoS Negl Trop Dis. 2017;11(5):e0005399.
45. Berry MP, Graham CM, McNab FW, Xu Z, Bloch SA, Oni T, et al. An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. Nature. 2010;466(7309):973-7.
46. Maertzdorf J, Repsilber D, Parida SK, Stanley K, Roberts T, Black G, et al. Human gene expression profiles of susceptibility and resistance in tuberculosis. Genes Immun. 2011;12(1):15-22.
47. Ottenhoff TH, Dass RH, Yang N, Zhang MM, Wong HE, Sahiratmadja E, et al. Genome-wide expression profiling identifies type 1 interferon response pathways in active tuberculosis. PLoS One. 2012;7(9):e45839.
48. Cliff JM, Lee JS, Constantinou N, Cho JE, Clark TG, Ronacher K, et al. Distinct phases of blood gene expression pattern through tuberculosis treatment reflect modulation of the humoral immune response. J Infect Dis. 2013;207(1):18-29.
49. Teles RM, Graeber TG, Krutzik SR, Montoya D, Schenk M, Lee DJ, et al. Type I interferon suppresses type II interferon-triggered human anti-mycobacterial responses. Science. 2013;339(6126):1448-53.
50. Chaitanya VS, Lavania M, Nigam A, Turankar RP, Singh I, Horo I, et al. Cortisol and proinflammatory cytokine profiles in type 1 (reversal) reactions of leprosy. Immunol Lett. 2013;156(1-2):159-67.
51. van Hooij A, Tjon Kon Fat EM, Richardus R, van den Eeden SJ, Wilson L, de Dood CJ, et al. Quantitative lateral flow strip assays as User-Friendly Tools To Detect Biomarker Profiles For Leprosy. Sci Rep. 2016;6:34260.
52. Guerreiro LT, Robottom-Ferreira AB, Ribeiro-Alves M, Toledo-Pinto TG, Rosa BT, Rosa PS, et al. Gene expression profiling specifies chemokine, mitochondrial and lipid metabolism signatures in leprosy. PLoS One. 2013;8(6):e64748.
53. Rego JL, de Lima Santana N, Machado PRL, Ribeiro-Alves M, de Toledo-Pinto TG, Castellucci LC, et al. Whole blood profiling of leprosy type 1(reversal) reactions highlights prominence of innate immune response genes. BMC Infect Dis. 2018;18(1):422.
54. Dupnik KM, Bair TB, Maia AO, Amorim FM, Costa MR, Keesen TSL, et al. Transcriptional changes that characterize the immune reactions of leprosy. J Infect Dis. 2015;211(10):1658-76.
55. Medeiros MF, Rodrigues MM, Vital RT, da Costa Nery JA, Sales AM, de Andrea Hacker M, et al. CXCL10, MCP-1, and other immunologic markers involved in neural leprosy. Appl Immunohistochem Mol Morphol. 2015;23(3):220-9.
56. Burja B, Feichtinger J, Lakota K, Thallinger GG, Sodin-Semrl S, Kuret T, et al. Utility of serological biomarkers for giant cell arteritis in a large cohort of treatment-naive patients. Clin Rheumatol. 2019;38(2):317-29.

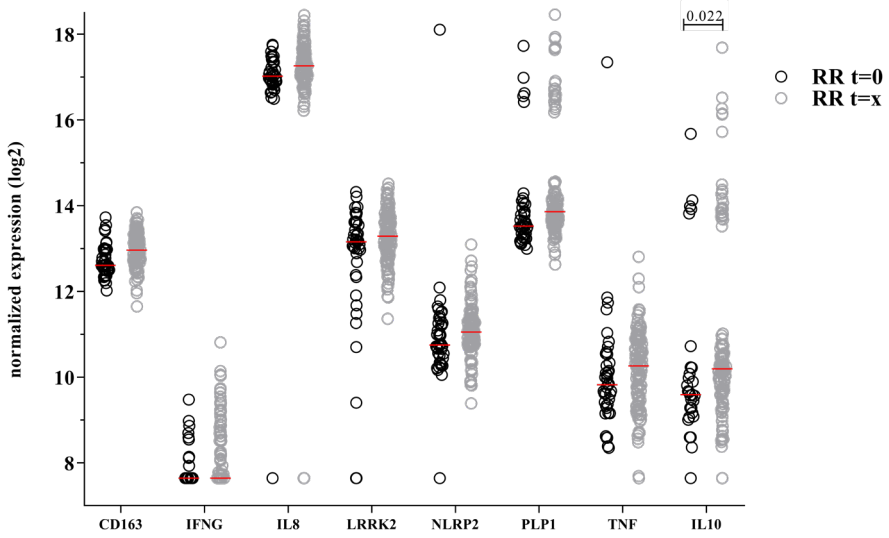
57. Gaur P, Myles A, Misra R, Aggarwal A. Intermediate monocytes are increased in enthesitis--related arthritis, a category of juvenile idiopathic arthritis. Clin Exp Immunol. 2017;187(2):234-41.
58. Han S, Zhuang H, Shumyak S, Wu J, Li H, Yang LJ, et al. A Novel Subset of Anti-Inflammatory CD138(+) Macrophages Is Deficient in Mice with Experimental Lupus. J Immunol. 2017;199(4):1261-74.
59. Balin SJ, Pellegrini M, Klechevsky E, Won ST, Weiss DI, Choi AW, et al. Human antimicrobial cytotoxic T lymphocytes, defined by NK receptors and antimicrobial proteins, kill intracellular bacteria. Sci Immunol. 2018;3(26).
60. Ottenhoff TH, Mutis T. Role of cytotoxic cells in the protective immunity against and immunopathology of intracellular infections. Eur J Clin Invest. 1995;25(6):371-7.
61. Maertzdorf J, Weiner J, Mollenkopf HJ, Bauer T, Prasse A, Muller-Quernheim J, et al. Common patterns and disease-related signatures in tuberculosis and sarcoidosis. Proc Natl Acad Sci U S A. 2012;109(20):7853-8.
62. Joosten SA, van Meijgaarden KE, Arend SM, Prins C, Oftung F, Korsvold GE, et al. Mycobacterial growth inhibition is associated with trained innate immunity. J Clin Invest. 2018;128(5):1837-51.
63. Teles RMB, Lu J, Tió-Coma M, Goulart IMB, Banu S, Hagge D, et al. Identification of a systemic interferon-γ inducible antimicrobial gene signature in leprosy patients undergoing reversal reaction. PLoS Negl Trop Dis. 2019;13(10):e0007764.
64. Levin M, Kaforou M. Predicting active tuberculosis progression by RNA analysis. Lancet. 2016;387(10035):2268-70.
65. Kramer A, Green J, Pollard J, Jr., Tugendreich S. Causal analysis approaches in Ingenuity Pathway Analysis. Bioinformatics. 2014;30(4):523-30.

Supplementary Material



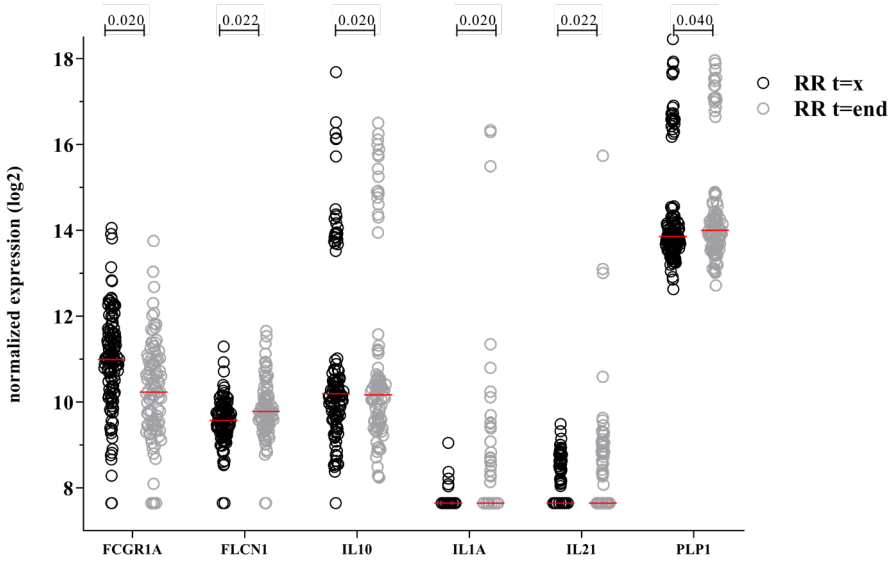
Supplementary Figure S1: AUC 5-biomarker signature to predict RR.

Performance of a 5-biomarker signature defined by global test and composed of gene expression data of *CCL2*, *CD8A*, *IL2*, *IL15* and *MARCO* to predict RR. And out-of-bag approach was used to assess the risk of developing RR in which a sample was excluded from the training set on each iteration. Receiver operator characteristics (ROC) curve is shown for the accuracy of the biomarker signature. The true positive rate (sensitivity) is plotted against the false positive rate (1-specificity). AUC: area under the curve.

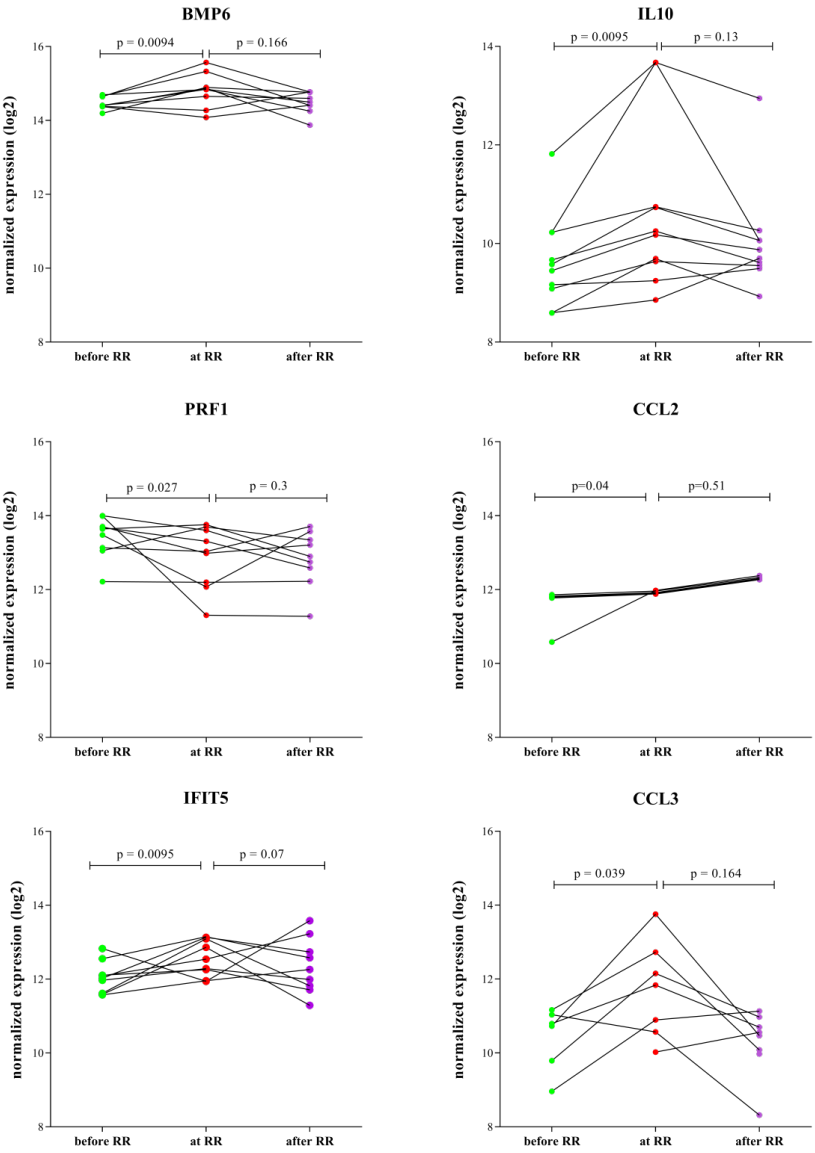


Supplementary Figure S2: Difference in gene expression in cross sectional samples between time points: before RR and at RR. Gene expression data obtained by dcRT-MLPA of RNA isolated from

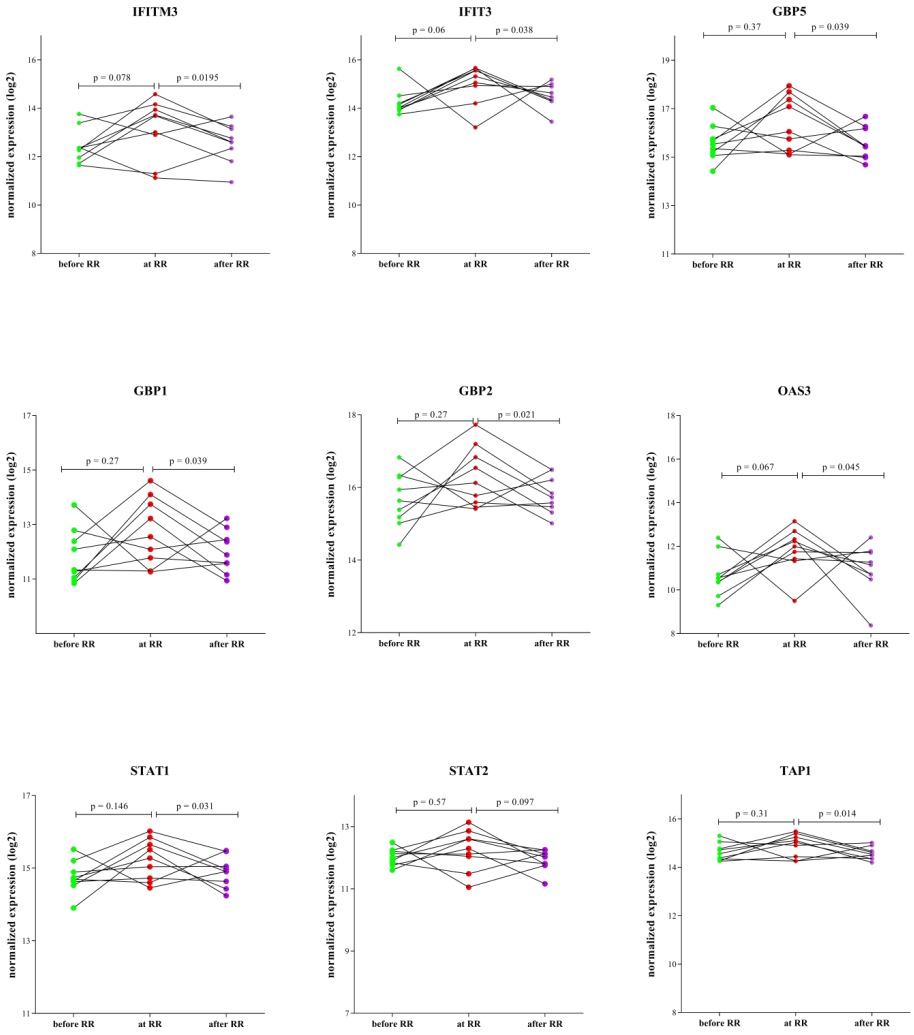
unstimulated whole blood of RR patients from Bangladesh, Brazil, Ethiopia and Nepal at t=0 (n= 41) and at clinical onset of RR (n=129, t=x). Log2-transformations of peak areas (normalized to the housekeeping gene *GAPDH*) of genes that were significantly different are shown on the *y*-axis. Raw p-values were calculated using the Mann–Whitney test and adjusted for multiple comparisons using the Benjamini-Hochberg correction (40).



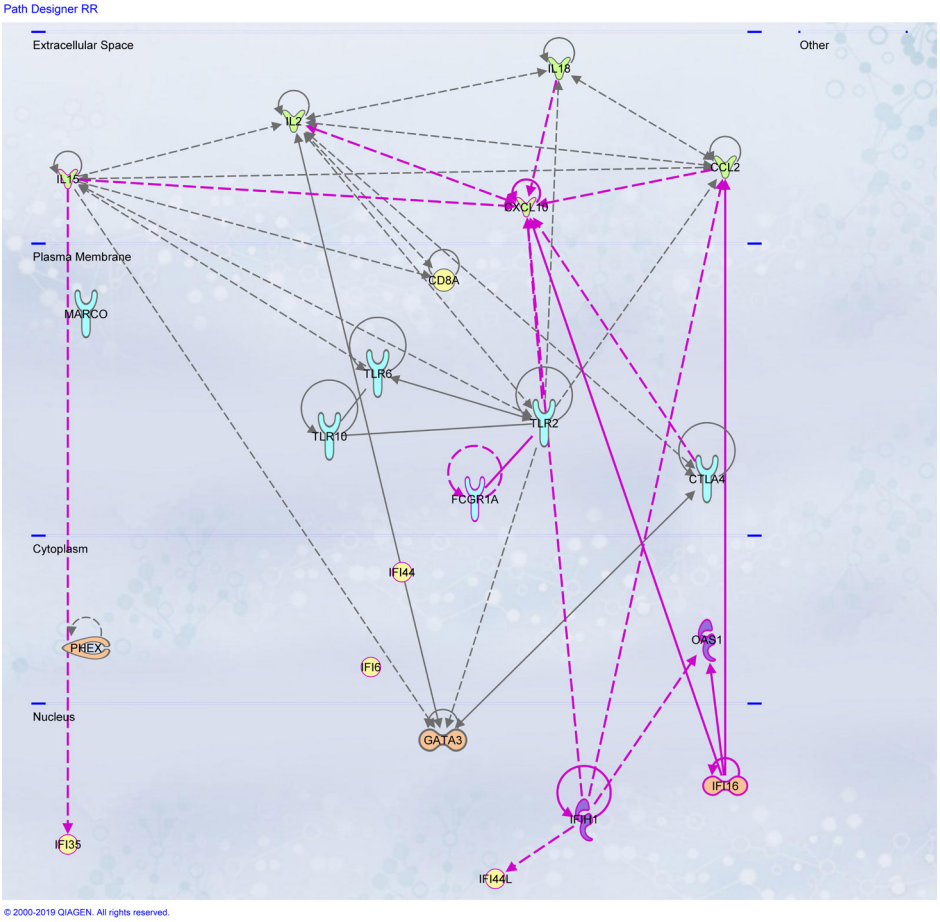
Supplementary Figure S3: Difference in gene expression in cross sectional samples between time points: at RR and after treatment. Gene expression data obtained by dcRT-MLPA of RNA isolated from unstimulated whole blood of RR patients from Bangladesh, Brazil, Ethiopia and Nepal at clinical onset of RR (n=129, t=x) and after treatment (n=110, t=end). Log2-transformations of peak areas (normalized to the housekeeping gene *GAPDH*) of genes that were significantly different are shown on the *y*-axis. Raw p-values were calculated using the Mann–Whitney test and adjusted for multiple comparisons using the Benjamini-Hochberg correction (40).



Supplementary Figure S4: Set of genes showing significant differences in expression levels during longitudinal follow-up before RR and at RR. Gene expression was assessed by dcRT-MLPA on *ex vivo* RNA of 10 leprosy patients who developed RR during this study. Blood was analyzed at 3 time points: in the absence of any clinical signs of reactions and at least two months before RR (t=0, in green), at RR diagnosis before steroids (t=x, in red) or after MDT and at least one month after end of steroids, in the absence of reactions (t=end, in purple). Log2-transformations of peak areas (normalized for *GAPDH* expression) are shown on the *y*-axis. Wilcoxon signed-rank test was performed. Significant differences (p-value <0.05) between gene expression before RR and at RR are indicated.



Supplementary Figure S5: Set of genes showing significant differences in expression levels during longitudinal follow-up between RR and after RR. Gene expression levels were assessed by dcRT-MLPA on *ex vivo* RNA of 10 leprosy patients who developed RR during this study. Blood was analyzed at 3 time points: in the absence of any clinical signs of reactions and at least two months before RR (t=0, in green), at RR diagnosis before steroids (t=x, in red) or after MDT and at least one month after end of steroids, in the absence of reactions (t=end, in purple). Log2-transformations of peak areas (normalized for *GAPDH* expression) are shown on the *y*-axis. Wilcoxon signed-rank test was performed. Significant differences (p-value <0.05) between gene expression before RR and at RR are indicated.



Supplementary Figure S6: Relationship of the genes identified as predictive CoR in RR and differently expressed longitudinally. Data were analyzed by Ingenuity Pathway Analysis (65) (QIAGEN Inc., https://www.qiagenbioinformatics.com/products/ingenuitypathway-analysis) showing the connections between the genes identified as predictive CoR for RR before clinical symptoms and genes with a significantly different expression in RR patients longitudinally. Genes identified in the longitudinal analysis are depicted with pink lines and genes identified as predictive CoR for RR in gray.

7

Family study identifies Single Nucleotide Polymorphisms associated with leprosy in Bangladesh

Maria Tió-Coma, Gal·la Garcia-Codina, Els M. Verhard, Priscila Verchai Uaska Sartori, Andressa Mayra dos Santos, Johan Chandra Roy, Marufa Khatun, Khorshed Alam, Jan Hendrik Richardus, Marcelo Távora Mira and Annemieke Geluk

Manuscript in preparation

Abstract

Introduction: The association between leprosy risk or protection and host genetics has been investigated in the last years. However, due to intrinsic genetic variations between different populations the impact of genetic variants in leprosy development might vary across populations.

Methods: In this study we investigated the effect in leprosy of 13 genetic markers located in 11 genes in the population of Bangladesh. These markers were found to be association with leprosy in Prata Village (Brazil). We performed a family-based analysis using 60 parent-affected child trios, followed by a case-control study consisting of 210 cases and 189 endemic controls. Genotypes for the 13 markers were determined by TaqMan SNP genotyping.

Results: Significant association was found in the family-based study between leprosy and three single nucleotide polymorphisms (SNPs), rs1801224, rs13001714, and rs1801582, located in *CUBN*, *IL1RL1*, and *PRKN* genes respectively. These findings did not replicate in the case-control sample.

Conclusion: Variants in the *CUBN*, *IL1RL1* and *PRKN* genes were associated with leprosy risk or protection in Bangladesh. Although replication in the case-control sample was not successful, the results presented in this study validated the association of these genes with leprosy as identified in Prata Village.

Introduction

Leprosy is a chronic infectious and neurological disease caused by *Mycobacterium leprae* that leads to an impairment of both the immune and peripheral nerve systems (1). The prevalence of this disease has globally decreased over the years as a result of treatment with multidrug therapy, which consists of a combination of rifampicin, clofazimine and dapsone. However, it still poses a significant threat to public health, especially in low- and middle-income countries, bearing more than 200.000 new cases annually (2). Leprosy presents different clinical manifestations depending on the immune response, ranging from few skin lesions with undetectable bacilli and strong cell-mediated immunity, corresponding to tuberculoid (TT) leprosy, to multiple lesions with high bacillary load and strong humoral immunity, corresponding to lepromatous (LL) leprosy. Between these two poles, borderline patients (BT, BB, BL) display variable degrees of TT and LL features that may develop into either of the two types (3, 4).

It is widely accepted that clinical leprosy, as an outcome of exposure to *M. leprae,* is likely to require a combination of pathogen burden as well as environmental and genetic susceptibility factors, meaning that the sole exposure to the pathogen is not enough to trigger the disease manifestations. Over the past century, more than 30 loci throughout the human genome have been associated with leprosy phenotypes, proving that the host genetic background strongly influences leprosy susceptibility and pathogenesis (3, 5).

The first positional candidate genes to harbor leprosy susceptibility were located at the 10p13, 6q25-27 and 6p21 chromosomal regions, described in Indian and Vietnamese populations, which included variants of the three classes of human leukocyte antigen (HLA) (6, 7). A subsequent high-resolution linkage disequilibrium (LD) mapping of the 6p21.3 region led to the identification of a functional single nucleotide polymorphism (SNP) in the HLA class III gene *LTA* as a risk factor for early-onset leprosy in the Vietnamese and North Indian populations (8, 9). Moreover, SNPs in *MRC1*, *CUBN* and *NEBL* genes, located in the 10p13 region, were found to be associated with MB leprosy also in the Vietnamese and Indian populations (6, 10, 11). Fine-mapping association studies performed to investigate further the 6q25-q27 region, showed that two SNPs located at the regulatory region shared by *PACRG* and *PRKN* (formerly known as *PARK2*) genes were strongly associated with leprosy (7, 12). The first genome-wide association study (GWAS) (13), which was performed on a Chinese population, identified *NOD2* polymorphisms, among others, to be associated with leprosy. Variants of *IL23R* were correlated with leprosy in Vietnamese and Chinese populations (14, 15) and an increased number of copies of this receptor were found in Indian leprosy patients (16).

Although the combination of different molecular strategies and study designs has led to the identification of several genes associated with leprosy, there are some incongruous

findings. These may be explained by diverse genetic backgrounds amongst populations or intrinsic differences between studies. Thus, it remains critical to further investigate and validate the association of genetic markers with leprosy in the different leprosy endemic areas. The aim of the present work was to test for validation, in a population sample from Bangladesh, of the association between leprosy and 13 genetic variants located in 11 genes (*ADO, BCL10*, CCDC88B, *CUBN*, *DEC1*, *IL1RL1*, *IL12RB2*, *IL23R*, *LTA*, *NOD2* and *PRKN*), all previously associated with an increased risk of leprosy development in different independent studies and populations (13, 15, 17, 18), including the Prata Village, an isolated former leprosy colony in Brazil (manuscript in preparation). The association between host genetics and leprosy has been understudied in Bangladesh, a country presenting areas with high leprosy endemicity. To date only one study has addressed host genomics of leprosy in Bangladesh, studying the N248S SNP in the Toll-like receptor 1 (*TLR1)* gene (19). Here, we further explored genetic associations with leprosy in Bangladesh through a family-based design formed by parent-affected offspring trios and a subsequent case-control study conducted with an independent sample.

Materials and methods

Study design and ethics statement

The family-based sample consisted of 60 family trios (Figure 1): leprosy patients (n=60) and both progenitors (n=118, including a multi-case family), recruited in Bangladesh (Table 1, family set). In addition, markers identified in the family-based sample were tested in a case-control set (Figure 1) which included leprosy patients (n=210) and unrelated endemic controls (EC) (n=189) from the same population (Table 1, case-control set).

Subjects from both sets were recruited in four districts of Bangladesh (Nilphamari, Rangpur, Panchagar and Thakurgaon) according to the Helsinki Declaration (2008 revision) and the study was approved by the National Research Ethics Committee (BMRC/NREC/2016-2019/214). Participants were informed about the study objectives and their right to refuse to take part or withdraw without consequences for their treatment. All subjects gave informed consent before enrollment and all patients received treatment according to national guidelines.

Blood samples were collected from all individuals in heparin tubes and were later transferred onto three different whole blood assays (WBA) containing different stimuli (*M. leprae* whole cell sonicate, *M. leprae* ML0840 recombinant proteins or no stimulus). Serum was removed and used for immunological analysis (data not shown) and the remaining was used for human genomic DNA isolation.

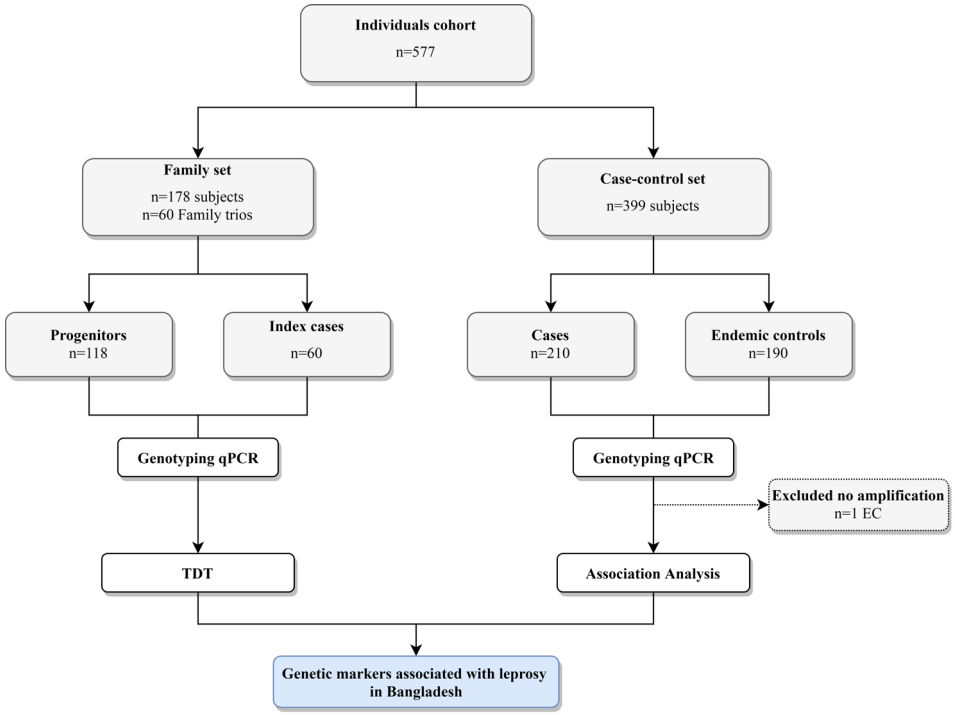


Figure 1. **Study design to identify genetic markers associated with leprosy in Bangladesh.** Schematic representation of the individuals recruited, experimental and analytical steps of the study. Samples excluded due to no amplification of any of the selected markers (family set markers tested =13, case-control set markers tested=3). TDT: Transmission disequilibrium test. ED: Endemic control.

DNA isolation

DNA isolation from WBA was performed using QIAamp DNA Blood Midi Kit (Qiagen, Hilden, Germany) or was automatized with Maxwell RSC Whole Blood DNA Kit (Maxwell, San Diego, CA) according to manufacturer's instructions. Starting sample volume for the QIAamp Kit was 2 ml and elution was performed using 200 µl of elution buffer (provided in the kit). For the Maxwell Kit starting volume was 300 µl and 60 µl of elution buffer (provided in the kit) was used.

SNP selection and genotyping

The genetic markers (n=13) selected for the present study were previously identified to be associated with leprosy susceptibility in an exon-sequencing-based study involving a Brazilian population from the Prata Village (manuscript in preparation). A list of 39 non-HLA class I and II candidate genes previously described to be associated with leprosy in independent populations of distinct ethnic backgrounds and candidate genes identified by GWAS (13, 15, 17, 18) were genotyped in leprosy patients from the Prata Village. Up to 19

genes were identified to be associated with leprosy and a selection of 13 genetic markers was made to be tested in the present study depending on whether the variation (i) caused an amino-acid change, (ii) was predicted to be deleterious *in silico*, (iii) was predicted as an eQTL, or (iv) explained (through LD) previous leprosy association findings. To analyze the selected SNPs, real-time quantitative PCR (qPCR) was performed using TaqMan Genotyping (Thermo Fisher Scientific, Waltham, MA). Mixes contained 20 ng of purified DNA, 0.5 µl of TaqMan SNP Genotyping Assay (Thermo Fisher, including forward and reverse primers and two TaqMan probes labeled differently, with VIC and FAM dyes, Table S1) and 5 µl of TaqMan Genotyping Master Mix in a final volume of 10 µl. The qPCR was performed by QuantStudio 6 Flex Real-Time PCR System (Applied Biosystems, Foster City, CA) starting with 10 minutes at 95ºC for the polymerase activation and then running 40 cycles of a 15 seconds at 95ºC, and one minute at 60ºC. Post-PCR reading was performed following manufacturer's instructions and the fluorescence measurements were collected. Genotypes were determined with Thermo Fisher Cloud Genotyping Application (Thermo Fisher Scientific) by analyzing the allelic discrimination plots.

Table 1. Characteristics of leprosy affected offspring in the family set and subjects in the case-control set.

	Family set, leprosy affected offspring (n=60)		**Case-control set, (n=399)**		
Group	**BT patients (n=50)**	**BL/LL patients (n=10)**	**BT patients (n=143)**	**BL/LL patients (n=67)**	**Endemic controls (n=189)**
Age (median)	21	26	35	35	36
Sex (n)	12 females 38 males	1 female 9 males	66 females 77 males	8 females 59 males	115 females 74 males
RJ classification (n)	50 BT	3 BL 7 LL	143 BT	37 BL 30 LL	-
BI (n)	43 BI-0 7 BI-und	2 BI-4 2 BI-5 6 BI-6	143 BI-0	1 BI-2 3 BI-3 21 BI-4 24 BI-5 18 BI-6	-

Characteristics of leprosy affected offspring in the family set and characteristics of leprosy patients and controls in the case-control set, including median age in years, number of females and males, Ridley-Jopling classification (borderline tuberculoid [BT], borderline lepromatous [BL] or lepromatous leprosy [LL]) and bacteriological index (BI). BI-und: bacteriological index undetermined.

Statistical analysis

Family-based association analyses were performed using FBAT v2.0.4Q (20) and PLINK v1.90b6.18 (21), by evaluating the transmission disequilibrium test (TDT). Hardy-Weinberg equilibrium (HWE) and LD were estimated using Haploview Software (version 4.2) (22). Comparative analysis between cases and controls were performed using a common logistic regression adjusted by sex. Analysis was carried out using R v4.0.4 (23) with the package SNPassoc (24).

Results

The genotyping success rate was >95 % for all tested markers except rs1801225 and rs224082. These two markers were excluded from the analysis due to MAF <0.05 for rs1801225 and to the presence of numerous no amplifications (79.2%) for rs224082 SNP, leading to insufficient data for analysis. The distributions of genotypes for the 11 genetic markers were in HWE and no LD has been detected between the markers.

Family-based association analysis included BT leprosy (n=50, 83%) and BL/LL leprosy (n=10, 17%) offspring (Table 1). TDT following the dominant model showed significant association of allele C of marker rs1801582 at *PRKN* gene (p-value=0.047) with leprosy *per se* (Table 2). Two markers indicated leprosy protection: allele G of marker rs1801224 at *CUBN* (p-value=0.045) and allele G of marker rs13001714 at *IL1RL1* (p-value=0.016). Another marker (rs1041973) at *IL1RL1* showed a tendency to leprosy protection for allele A (p-value=0.100), although not statistically significant.

The three markers showing evidence for association with leprosy in the family-based analysis (rs1801582, rs1801224 and rs13001714) were tested in a replication case-control sample set (Figure 1) including endemic controls (n=189) and leprosy patients (n=210) of which 143 (68%) presented BT leprosy and 67 (32%) BL/LL leprosy. Significant differences (p-value=$5.31 \cdot 10^{-7}$, Mann-Whitney U test) in the sex of individuals included in the case and control groups were observed, with a higher number of males in the case group. No difference was observed in the age of individuals in the case and control groups, however, a significant difference in the age was found between patients in the case group and index patients in the family set (p-value=$2.97 \cdot 10^{-13}$, Mann-Whitney U test). Association analysis, adjusted by sex, did not show significant association of any of the three markers with leprosy in the case-control sample set. However, rs13001714 marker at *IL1RL1* gene (Table 3) showed a tendency for association with a p-value of 0.133 in the dominant model. Association analysis separating the cases based on the type of leprosy (BT or BL/LL) did not reveal any further association signal with these phenotypes (data not shown).

Table 2. Association between leprosy *per se* and Single-Nucleotide Polymorphisms in the family set from Bangladesh.

Gene	Marker	Region	Allele	AF	IF	p-value[a]
BCL10	rs1060846	1p22.3	A	0.396	41	0.741
CCDC88B	rs542907	11q13.1	A	0.438	32	0.450
CUBN	rs1801224	10p13	G	0.429	33	**0.045**
DEC1	rs2285316	9q33.1	G	0.308	39	0.139
IL1RL1	rs1041973	2q12.1	A	0.104	21	0.100
	rs13001714		G	0.329	37	**0.016**
IL12RB2	rs10489627	1p31.3	G	0.292	40	0.167
IL23R	rs10889677	1p31.3	A	0.637	20	0.806
LTA	rs1041981	6p21.33	A	0.254	33	0.176
NOD2	rs3135499	16q12.1	C	0.204	27	0.768
PRKN	rs1801582	6q26	C	0.746	13	**0.047**

Results of gene markers (n=11) from 10 genes analyzed in 60 family trios formed by case-parents. Characteristics for each marker are shown: corresponding candidate gene, chromosomic region, allele, allele frequency (AF), number of informative families (IF), and p-value (in bold p-values <0.05).

[a] Results of the transmission disequilibrium test (p-value) are shown for the dominant model.

Table 3. Case-control replication analysis of markers associated with leprosy in the family-based analysis in Bangladesh.

Gene	Marker	Genotype	Cases	Controls	OR (95% CI)[a]	p-value[a]
CUBN	rs1801224	GG+GT	136 (0.67)	124 (0.67)	0.89 (0.57 – 1.39)	0.611
		TT	67 (0.33)	60 (0.33)		
IL1RL1	rs13001714	GG+AG	130 (0.62)	102 (0.54)	1.37 (0.91 – 2.07)	0.133
		AA	80 (0.38)	87 (0.46)		
PRKN	rs1801582	GG+GC	103 (0.50)	86 (0.55)	1.11 (0.74-1.67)	0.619
		CC	104 (0.50)	103 (0.45)		

Association analysis in leprosy cases (n=210) and controls (n=189) following a common logistic regression adjusted by sex. Number and frequencies of cases and controls presenting each genotype. OR: odds ratio. CI: confidence interval.

[a] Results of the association analysis (p-value and OR) are shown for the dominant model.

Discussion

The numerous studies that investigated the relationship between human genetics and leprosy over the past century have provided evidences that genetics play a significant role in susceptibility to leprosy phenotypes, including leprosy *per se*, clinical forms of disease and the occurrence of leprosy reactions (3, 12, 13, 25). Nevertheless, differences in the pattern of association of genetic variants and leprosy are found between distinct populations, emphasizing the importance to replicate and validate genetic associations in populations from diverse ethnic backgrounds. Thus far, only one study has been conducted in the Bangladeshi population, investigating the association with leprosy of genetic polymorphisms in one gene, *TLR1* (19). Analysis revealed that homozygous S248 was significantly associated with leprosy *per se*, whereas the heterozygous SN genotype was found to be protective against leprosy. In the present work, we further explored genetic associations with leprosy in Bangladesh by replicating the analysis of 13 SNPs previously associated with leprosy in the Prata population.

Using a family-based sample from Bangladesh we identified three markers (rs1801224, rs13001714 and rs1801582) associated with leprosy in the *CUBN*, *IL1RL1* and *PRKN* genes. A second marker in the *IL1RL1* gene (rs1041973) showed borderline evidence for association. Since no LD between genes was detected, this suggests two independent association signals between leprosy an *IL1RL1*. Replication of the results in the case-control sample was not successful, although suggestive evidence for association was observed for rs13001714 (p-value = 0.133). The lack of validation was possibly due to the observed difference in the age of leprosy onset between the two samples (cases in the family-based set were younger adults than those in the case-control), the slightly higher percentage of BL/LL leprosy cases in the case-control sample, or lack of power in the relatively small case-control sample. Family-based studies overcome population stratification since subjects present a common genetic composition (26). Thus, when the case-control sample is small and/or potentially harboring cryptic population stratification, family-based studies are advantageous.

In line with our results in the family-based sample, rs10904831, located in the same gene as rs1801224 (*CUBN*), shows a significant association with MB leprosy in the Vietnamese population (10). Interestingly, previous genome-wide scans conducted in India and Vietnam, identified a PB susceptibility locus in chromosome 10p13 (6, 7) where *CUBN* is located, suggesting that this region may be involved in leprosy polarization. *CUBN* encodes cubilin, a receptor for intrinsic factor-vitamin B12 complexes (27). It was shown that Rv1819c, an homolog of the *M. leprae* protein ML2084 in *Mycobacterium tuberculosis*, is involved in vitamin B12 uptake (28). Thus, the host's cubilin is possibly competing with mycobacterial proteins during infection (10). Association between leprosy *per se* and *IL-*

1RL1 was previously found in a Chinese population in the 2q12.1 region, where a cluster is formed by *IL1RL1* and *IL18* receptor genes (29). *IL1RL1* encodes the receptor of IL-33 which is present in Th2 cells and is responsible for the expression of Th2 cytokines (30). The connection of *IL1RL1* with its ligand leads to the activation of nuclear factor-kB (NF-kB), which is essential for mycobacterial immunity (13). Finally, in the Vietnamese and Brazilian populations, variants in the *PRKN* gene have been identified as major leprosy susceptibility locus (7, 12). *PRKN* codes Parkin, a protein part of the E3 ubiquitin ligase complex, which is involved in Parkinson's disease as well as the immune response to mycobacteria and other intracellular pathogens (31, 32). Parkin regulates the autophagy of damaged molecules, intracellular pathogens or organelles such as mitochondria (mitophagy) (33). Parkin-mediated xenophagy (bacterial autophagy) eliminates mycobacteria and is crucial to inhibit replication of *M. tuberculosis* in macrophages (32). In addition, Parkin is involved in other immunity pathways by regulating IL-6 and MCP-1/*CCL2* levels (31).

Thus, the three genes identified to be associated with leprosy in this study are either involved in controlling the immune response or the manifestation of different clinical forms of the disease.

In summary, our results in the family-based sample provide additional evidence of the association with leprosy of markers rs1801224, rs13001714 and rs1801582, located in the *CUBN, IL1RL1* and *PRKN* genes, previously identified in Prata Village (manuscript in preparation) as well as other ethnically distinct populations (12, 13, 15, 17, 18). Further ongoing studies in additional endemic populations will contribute to the validation of the 13 genetic markers investigated in this study. The association of the here described markers with leprosy in the Bangladeshi population aids the identification of a leprosy-risk genetic profile that could be applied in targeted preventive strategies in Bangladesh, potentially in combination with other biomarkers, to enhance early diagnosis of leprosy.

Acknowledgements

The authors gratefully acknowledge all patients and control participants.

References

1. Britton WJ, Lockwood DNJ. Leprosy. The Lancet. 2004;363(9416):1209-19.
2. World Health Organization. Global leprosy update, 2018: moving towards a leprosy-free world. Weekly Epidemiological Record. 2019;94(35/36):389-412.
3. Fava VM, Dallmann-Sauer M, Schurr E. Genetics of leprosy: today and beyond. Hum Genet. 2020;139(6-7):835-46.
4. Sauer ME, Salomão H, Ramos GB, D'Espindula HR, Rodrigues RS, Macedo WC, et al. Genetics of leprosy: expected and unexpected developments and perspectives. Clin Dermatol. 2015;33(1):99-107.
5. Cole ST, Eiglmeier K, Parkhill J, James KD, Thomson NR, Wheeler PR, et al. Massive gene decay in the leprosy bacillus. Nature. 2001;409(6823):1007-11.
6. Siddiqui MR, Meisner S, Tosh K, Balakrishnan K, Ghei S, Fisher SE, et al. A major susceptibility locus for leprosy in India maps to chromosome 10p13. Nature genetics. 2001;27(4):439-41.
7. Mira MT, Alcaïs A, Van Thuc N, Thai VH, Huong NT, Ba NN, et al. Chromosome 6q25 is linked to susceptibility to leprosy in a Vietnamese population. Nature genetics. 2003;33(3):412-5.
8. Alcaïs A, Alter A, Antoni G, Orlova M, Nguyen VT, Singh M, et al. Stepwise replication identifies a low-producing lymphotoxin-alpha allele as a major risk factor for early-onset leprosy. Nature genetics. 2007;39(4):517-22.
9. Ali S, Chopra R, Aggarwal S, Srivastava AK, Kalaiarasan P, Malhotra D, et al. Association of variants in BAT1-LTA-TNF-BTNL2 genes within 6p21.3 region show graded risk to leprosy in unrelated cohorts of Indian population. Hum Genet. 2012;131(5):703-16.
10. Grant AV, Cobat A, Van Thuc N, Orlova M, Huong NT, Gaschignard J, et al. CUBN and NEBL common variants in the chromosome 10p13 linkage region are associated with multibacillary leprosy in Vietnam. Human Genetics. 2014;133(7):883-93.
11. Alter A, de Léséleuc L, Van Thuc N, Thai VH, Huong NT, Ba NN, et al. Genetic and functional analysis of common MRC1 exon 7 polymorphisms in leprosy susceptibility. Human Genetics. 2010;127(3):337-48.
12. Mira MT, Alcais A, Nguyen VT, Moraes MO, Di Flumeri C, Vu HT, et al. Susceptibility to leprosy is associated with PARK2 and PACRG. Nature. 2004;427(6975):636-40.
13. Zhang FR, Huang W, Chen SM, Sun LD, Liu H, Li Y, et al. Genomewide association study of leprosy. N Engl J Med. 2009;361(27):2609-18.
14. Cobat A, Abel L, Alcaïs A, Schurr E. A general efficient and flexible approach for genome-wide association analyses of imputed genotypes in family-based designs. Genet Epidemiol. 2014;38(6):560-71.
15. Zhang F, Liu H, Chen S, Low H, Sun L, Cui Y, et al. Identification of two new loci at IL23R and RAB32 that influence susceptibility to leprosy. Nature genetics. 2011;43(12):1247-51.
16. Ali S, Srivastava AK, Chopra R, Aggarwal S, Garg VK, Bhattacharya SN, et al. IL12B SNPs and copy number variation in IL23R gene associated with susceptibility to leprosy. J Med Genet. 2013;50(1):34-42.
17. Liu H, Bao F, Irwanto A, Fu X, Lu N, Yu G, et al. An association study of TOLL and CARD with leprosy susceptibility in Chinese population. Hum Mol Genet. 2013;22(21):4430-7.
18. Liu H, Irwanto A, Fu X, Yu G, Yu Y, Sun Y, et al. Discovery of six new susceptibility loci and analysis of pleiotropic effects in leprosy. Nat Genet. 2015;47(3):267-71.
19. Schuring RP, Hamann L, Faber WR, Pahan D, Richardus JH, Schumann RR, et al. Polymorphism N248S in the human Toll-like receptor 1 gene is related to leprosy and leprosy reactions. J Infect Dis. 2009;199(12):1816-9.
20. Horvath S, Xu X, Laird NM. The family based association test method: strategies for studying general genotype--phenotype associations. Eur J Hum Genet. 2001;9(4):301-6.
21. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81(3):559-75.
22. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 2005;21(2):263-5.
23. R Development Core Team. R: A Language and Environment for Statistical Computing. http://www.R-project.org/. R Foundation for Statistical Computing; 2014.
24. González JR, Armengol L, Solé X, Guinó E, Mercader JM, Estivill X, et al. SNPassoc: an R package to perform whole genome association studies. Bioinformatics. 2007;23(5):654-5.
25. Fava VM, Mira MT. Leprosy: A Practical Guide. Nunzi E, Massone C, editors. Milan: Springer; 2012.
26. Suryadevara NCea. Genetic association of G896A polymorphism of TLR4 gene in leprosy through family-based and case-control study designs. Trop Med Hyg. 2013;107:777–82.
27. Seetharam B, Levine JS, Ramasamy M, Alpers DH. Purification, properties, and immu-

nochemical localization of a receptor for intrinsic factor-cobalamin complex in the rat kidney. J Biol Chem. 1988;263(9):4443-9.
28. Gopinath K, Venclovas C, Ioerger TR, Sacchettini JC, McKinney JD, Mizrahi V, et al. A vitamin B_{12} transporter in Mycobacterium tuberculosis. Open Biol. 2013;3(2):120175.
29. Liu Hea. Identification of IL18RAP/IL18R1 and IL12B as leprosy risk genes demonstrates shared pathogenesis between inflammation and infectious diseases. Am J Hum Genet. 2012;91:935–41.
30. Schmitz J, Owyang A, Oldham E, Song Y, Murphy E, McClanahan TK, et al. IL-33, an interleukin-1-like cytokine that signals via the IL-1 receptor-related protein ST2 and induces T helper type 2-associated cytokines. Immunity. 2005;23(5):479-90.
31. de Léséleuc L, Orlova M, Cobat A, Girard M, Huong NT, Ba NN, et al. PARK2 Mediates Interleukin 6 and Monocyte Chemoattractant Protein 1 Production by Human Macrophages. PLOS Neglected Tropical Diseases. 2013;7(1):e2015.
32. Manzanillo PS, Ayres JS, Watson RO, Collins AC, Souza G, Rae CS, et al. The ubiquitin ligase parkin mediates resistance to intracellular pathogens. Nature. 2013;501(7468):512-6.
33. Cambri G, Mira MT. Genetic Susceptibility to Leprosy-From Classic Immune-Related Candidate Genes to Hypothesis-Free, Whole Genome Approaches. Frontiers in immunology. 2018;9:1674.

Supplementary Material

Table S1. Predesigned TaqMan SNP Genotyping Assays from Thermo Fisher.

Gene name	Marker	Assay ID	Gene alias
ADO	rs224082	C____609393_20	*EGR2*
BCL10	rs1060846	C___2720591_10	*C1orf52, LOC646626*
CCDC88B	rs542907	C___2983453_10	*RPS6KA4*
CUBN	rs1801224	C___3135085_10	-
CUBN	rs1801225	C___3135052_20	**-**
DEC1	rs2285316	ANFVVYA	-
IL1RL1	rs1041973	C___1226160_1_	-
IL1RL1	rs13001714	ANGZPH7	-
IL12RB2	rs10489627	C__27853925_10	-
IL23R	rs10889677	C__11283764_10	-
LTA	rs1041981	C___7514870_20	*TNF, LTB, LOC100287329*
NOD2	rs3135499	C__31758802_10	**-**
PARK2	rs1801582	C___8701299_10	*PRKN*

Assay IDs (Thermo Fisher Scientific) per marker including primers and probes used for genotyping.

8

General discussion

Maria Tió Coma

Leprosy is a multifactorial disease that still affects more than 200.000 new cases per year. An early diagnosis is decisive to reduce and avoid leprosy-associated disabilities, however, leprosy is often late- or mis-diagnosed and clinical symptoms are visibly present at diagnosis. Furthermore, transmission of *Mycobacterium leprae* is still not completely understood, although it is clear that the major source of transmission is via human-to-human contact likely involving the respiratory route. To tackle transmission, One Health approaches could be advantageous, since not only humans but other animals and the environment may play a role as *M. leprae* reservoirs. In this thesis, pathogen transmission in leprosy patients as well as their household contacts (HC) was explored (Figure 1). In addition, possible animal and environmental reservoirs of *M. leprae* or *Mycobacterium lepromatosis* were investigated. Next, we searched for transcriptomic host biomarkers that could predict leprosy or leprosy reactions before occurrence of symptoms, and genomic biomarkers to identify individuals at higher risk of developing leprosy. Identification of predictive biomarkers will in due course lead to a prompt treatment, preventing leprosy-associated irreversible disabilities as well as reducing *M. leprae* transmission.

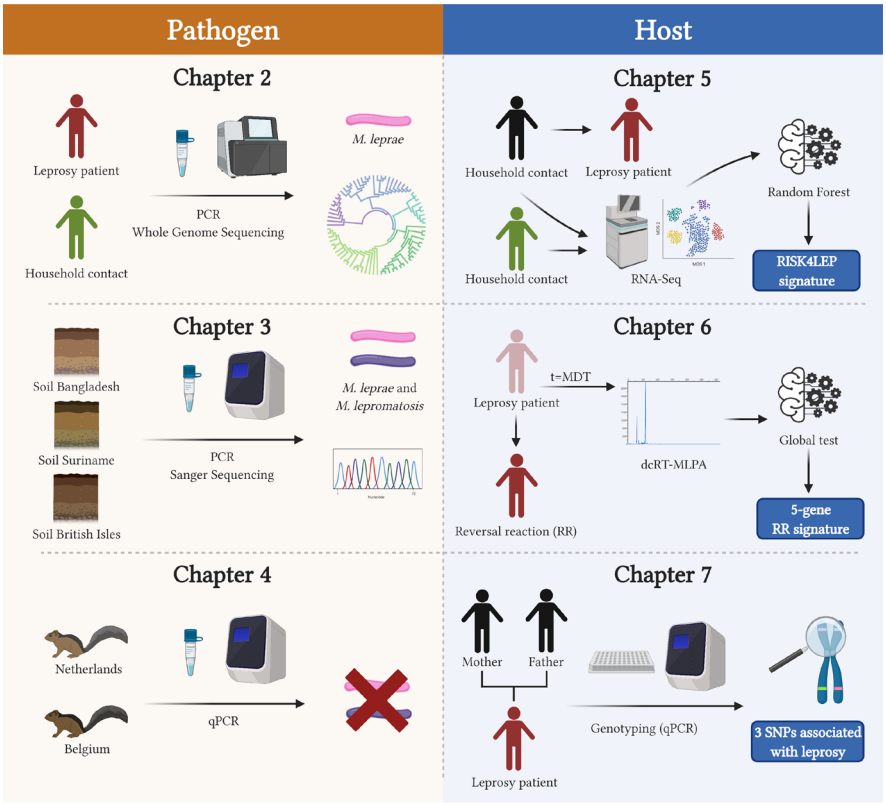


Figure 1. Summarizing overview of thesis.

Mycobacterium leprae genomics: what's new?

The low mutation rate of *M. leprae* (18 ± 30 mutations per 1,000 years (1)) enables the study of ancient transmission routes and ultimately facilitates the identification of the origin of leprosy (2). Although initially leprosy was described to have originated in Eastern Africa or the Middle East with *M. leprae* genotype 2 or 1 (3), recent studies identified genotype 3K-0, also known as branch 0, as the most ancestral *M. leprae* linage (1, 4, 5). This linage is found in modern East Asia and the Middle East (6, 7) and ancient Europe (8). Thus, currently two hypotheses of the origin of leprosy are considered: Western Eurasia or East Asia and the Middle East. Since only a limited number of *M. leprae* genomes from these locations have been sequenced, future studies will further validate or confront these hypotheses. Phylogenomic studies of *M. leprae* are valuable to identify the origin of leprosy, but they are also a tool to study transmission.

Over the years *M. leprae* whole genomes have been retrieved from modern leprosy patients (3, 4, 6) and animals (9-13), as well as ancient skeletons (1, 14-17) establishing a genotype classification consisting of four genotypes: 1-4 (3). These genotypes were later further divided into 16 subtypes: A-P (6, 11). *M. leprae* genotype 1 is predominant in Asia, however, recently it was also identified in the South of Africa (Madagascar and Comoros) (5). In Bangladesh, a leprosy endemic country, only one study had investigated *M. leprae* genomics including only four Bangladeshi patients by whole genome sequencing (WGS) and genotype 1 was identified (6).

In **chapter 2,** *M. leprae* genome diversity in Bangladesh was explored more extensively. Several subtypes of genotype 1 (1A, 1C and 1D) were identified, but most importantly, we characterized a new subtype that is thus far unique for Bangladesh, namely subtype 1B-Bangladesh. This new subtype clusters separately from the rest of genotype 1 subtypes and is phylogenetically positioned between subtypes 1B and 1A (Figure 2). In the four districts investigated in the study (Nilphamari, Rangpur, Panchagar and Thakurgaon) located in the North of Bangladesh, *M. leprae* subtype 1D was the most widely distributed throughout the entire territory while subtypes 1A and 1B-Bangladesh were only found in the eastern region. More studies covering other leprosy endemic areas, are needed to determine whether the new 1B-Bangladesh subtype is exclusive of Bangladesh or is also present in neighboring countries such as India. Thus, specific primers to detect genotype 1B-Bangladesh were designed in **chapter 2** which we suggest to include in future *M. leprae* genotyping studies.

In addition to identifying a new subtype, WGS phylogenetic analysis including the new *M. leprae* Bangladeshi strain as well as previously sequenced strains (1, 4, 18), demonstrated that the formerly described genotype 1C does not in fact constitute a separate subtype. The position previously used to differentiate genotype 1C (61425) located at *esxA* gene, is

not phylogenetically informative and the single nucleotide polymorphism (SNP) found in 1C can also be observed in genotypes 3I and 2E. Genotype 1C is actually distributed along the 1D genotype and does not form a separate cluster. Thus, in **chapter 2**, it is described how the *M. leprae* genotype classification is updated and concluded that *M. leprae* genotype 1 is composed by subtypes 1A, 1B, 1B-Bangladesh and 1D.

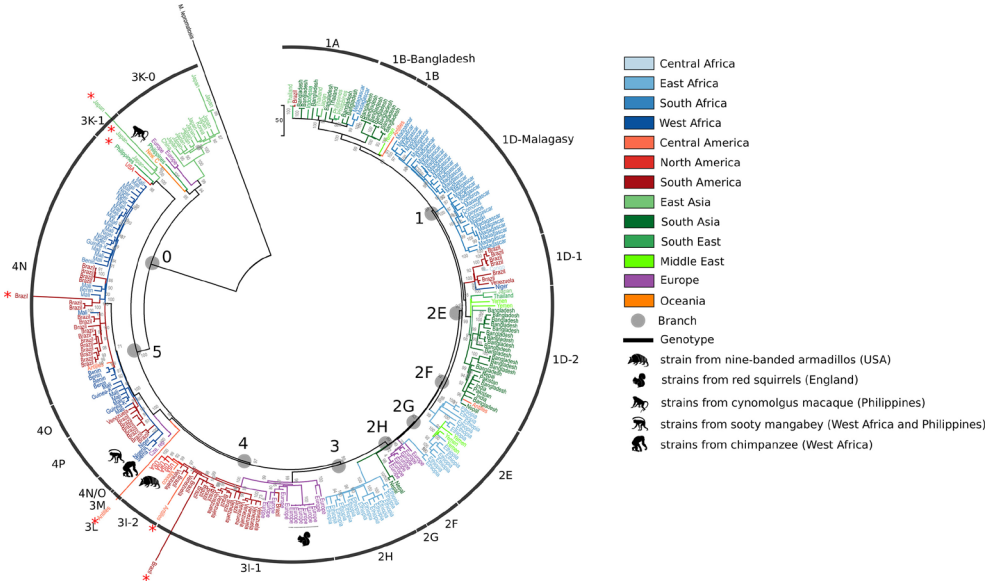


Figure 2. Updated *Mycobacterium leprae* phylogenomic tree. From Avanzi et al 2020 (5).

A separate study further investigated genotype 1D and two subgroups were identified: 1D-1 and 1D-2 (Figure 2) (18). Strains belonging to 1D-2 were mostly found in East Asia whereas 1D-1 strains were retrieved from West Africa and South America. The genotype previously classified as genotype 1C is actually part of the 1D-2 subgroup. Additionally, a new genotype was also described that is specific for Madagascar and Comoros, 1D-Malagasy.

The low mutation rate of *M. leprae*, the standard use of multidrug treatment, and the constraints to perform drug resistance profiling due to the fact that *M. leprae* can't be cultured in microbiological culture media, have limited the reports of antimicrobial resistant *M. leprae* (19). Currently, mutations at certain positions in *rpoB*, *folP1* and *gyrA* genes are surveilled for resistance to rifampicin, dapsone and ofloxacin respectively (20). In Bangladesh, antimicrobial resistant *M. leprae* has thus far not been reported (19), although this could be due to understudy of drug resistance profiles. We identified in a sample isolated from a nasal swab (NS), a mutation in the *rpoB* gene in 50% of the sequence that leads to an amino acid change in a position essential to determine antimicrobial resistance (20).

This is probably due to a mixed infection with a resistant strain or an in-patient *M. leprae* emerging mutation. Intra-host variation was found particularly in DNA retrieved from NS and mutations were often located in *ml1512* and *ml1750* genes, which were previously reported as hypermutated (4) and encode a ribonuclease J and a putative nucleotide cyclase.

In addition, several silent mutations were found in three different locations of the *rpoB* gene, indicating that mutations are definitely taking place in genes associated with antimicrobial resistance. We suggest increasing drug resistance surveillance in Bangladesh to identify potentially hidden antimicrobial resistant *M. leprae* strains, particularly in the light of the currently WHO endorsed PEP therapy for household contacts of newly diagnosed leprosy patients (21).

Leprosy bacilli transmission: One Health approach

Leprosy is not exclusive to humans, as red squirrels (9), armadillos (11, 12) and non-human primates (13, 22) can become naturally infected with *M. leprae* or *M. lepromatosis* as well. Moreover, other environmental sources such as soil (23-29) and water (29-31) have also been identified. A strategy to reduce transmission needs to incorporate not only human healthcare, but other potential sources as well. Thus, to understand the leprosy bacilli transmission patterns, a One Health approach may provide additional insight. This could be particularly of interest in areas where leprosy is not endemic (anymore) but cases are still found sporadically (11, 32). But also in areas where people are in frequent contact with proven animal sources of leprosy (33).

M. leprae transmission is still not completely understood and although skin-to-skin contact or aerosol transmission are the suggested transmission routes (34, 35), transmission is still subject of debate. In **chapter 2**, we observed that leprosy patients, but also asymptomatic individuals may carry *M. leprae*. A 33.3% *M. leprae* PCR positivity in NS of leprosy patient with bacillary index (BI) 0 was found, whereas in asymptomatic HC of leprosy patients with high BI (BI ≥ 2), PCR positivity was 18.0%. In addition, 36.8% of HC showed presence of anti-*M. leprae* phenolic glycolipid I (PGL-I) IgM in serum. This indicates that asymptomatic HC may play a parallel role contributing to *M. leprae* transmission, although nasal carriage is transient and dependent on the season (36). Future studies are needed to validate the viability of the bacteria in the nasal cavity and determine if the number of bacteria carried by asymptomatic contacts is comparable to that of leprosy patients with BI 0, paucibacillary. Using WGS, we observed direct transmission between two leprosy patients from the same household, who presented identical *M. leprae* genomes and between two leprosy patients with the same genotype including a strain-specific SNP. Unfortunately, *M. leprae* DNA recovery from slit skin smears (SSS) and NS of leprosy patients with BI 0 and HC is challenging and *M. leprae* genetic material is not sufficient to perform WGS or extensive

Sanger sequencing. This is a strong limitation to study *M. leprae* transmission, particularly in asymptomatic individuals from whom biopsies are not available. SSS and NS are, however, less invasive than biopsies and useful samples to characterize *M. leprae* from leprosy patients with BI ≥ 2.

Humans are not the only reservoir of the leprosy bacilli and since other source of infection could play a major role in transmission, we investigated other potential environmental and animal reservoirs. In **chapter 3**, we evaluated whether *M. leprae* or *M. lepromatosis* were present in soil from Bangladesh, Suriname and the British Isles. We identified *M. leprae* DNA in soil from the houses of leprosy patients in Bangladesh and from armadillos' holes at former leprosy colonies in Suriname. However, *M. leprae* DNA was only identified in a small number of samples and in low concentrations. In Bangladesh, the *M. leprae* genotype identified in the soil was the same that was found in the samples isolated from leprosy patients (genotype 1). This suggests that the *M. leprae* strains identified in the soil were originally shed from leprosy patients, but further subtyping would be necessary to confirm the association. *M. leprae* DNA in soil is very limited and in low concentration, therefore, extensive subtyping could not be performed. Additionally, *M. lepromatosis* was detected in soil obtained around an area where *M. lepromatosis* infected squirrels are located in the Isle of Arran.

Although we did not assess the viability of *M. leprae* in soil, it was previously described that viable bacteria can survive in wet soil for 46 days (37) and in free living amoebae in a laboratory up to 8 months (38). Moreover, small insects such as the kissing bug (*Rhodnius prolixu*) or ticks could be a vector for *M. leprae*, since viable *M. leprae* bacilli were found in the faeces of artificially-infected kissing bugs and induced transmission to rabbits was observed using *M. leprae*-infected tick larvae (39, 40). Thus, indirect transmission via environmental samples such as soil or water needs to be taken into account when the source of *M. leprae* is unknown and humans or animals infected with the leprosy bacilli are present. Nevertheless, the low concentration of *M. leprae* or *M. lepromatosis* DNA found in soil suggests that environmental contamination as a source of infection is not to be expected. More research is needed to further evaluate the viability of the leprosy bacilli in soil and determine the impact of environmental reservoirs as transmission sources. However, the scarce amount of DNA in such samples presents a challenge for epidemiological studies to investigate the role of these reservoirs.

Squirrels are among the small group of the animals presenting a leprosy-like disease caused by the leprosy bacilli. In the British Isles red squirrels infected with *M. leprae* or *M. lepromatosis* were identified (9, 41-43). The *M. leprae* strain identified in the squirrels was closely related to the strain circulating in humans in the Middle Ages, when leprosy disappeared in humans in the British Isles. Thus, possibly indicating that red squirrels or

an intermediary vector acquired *M. leprae* or *M. lepromatosis* during the Middle Ages and it persisted in the red squirrels as a reservoir. In **chapter 4**, we investigated whether Dutch and Belgian squirrels were carriers of the leprosy bacilli. We examined 114 squirrels by quantitative PCR (qPCR) and we did not detect *M. leprae* or *M. lepromatosis* DNA in any of the animals. This is in line with previous findings in France, Germany, Switzerland and Italy where PCR showed no presence of the leprosy bacilli either (10). Combining the results of this study with our data we concluded that if the leprosy bacillus is present in continental Europe, the prevalence in squirrels is 0 to 2.2% with a 95% confidence interval (CI). Thus, up to the present time, squirrels infected with *M. leprae* or *M. lepromatosis* have only been identified in the British Isles.

Transmission of the leprosy bacilli from red squirrels to humans has not been reported and due to limited squirrel-human interaction in the areas where squirrels with leprosy were found and the geographical limitations of Islands, zoonotic transmission and transmission to other areas should not be a cause of major concerns. However, in the south of United States, high prevalence of *M. leprae* infection in nine-banded armadillos infected with *M. leprae* was observed (11, 12, 44-47) and potential zoonotic transmission due to contact with armadillos was found (11, 12). Leprosy patients and wild armadillos were infected with *M. leprae* genotype 3I-2-v1, which is closely related to the variant circulating in Europe during the Middle Ages. In Brazil, a study observed that armadillo hunters and those who process or eat armadillo meat have a higher rate of leprosy (33), although other studies did not find an associations between armadillo meat consumption and leprosy (48, 49). The observed zoonotic transmissions between armadillos and humans highlights the necessity to not only investigate *M. leprae* and *M. lepromatosis* transmission between humans but incorporate animal and environmental reservoirs in strategies to effectively reduce transmission and ultimately eradicate leprosy.

Predicting leprosy and leprosy reactions: the potential of transcriptomic biomarkers

Early diagnosis of leprosy and leprosy reactions is crucial to avoid leprosy-associated disabilities which cause lifelong disabilities and stigma. Currently, diagnosis still heavily relies on appearance of clinical symptoms and prospective biomarkers are not available. Transcriptomic biomarkers aid the development of diagnostic tools by identifying differential gene expression in patients or individuals at risk of disease. In tuberculosis research, transcriptomic biomarkers have been successfully employed to predict active disease (50-52). Studies investigating potential biomarkers for leprosy diagnosis (53-59) are limited and do not predict the disease in individuals exposed to the leprosy bacilli. In addition, the starting materials used in previous studies are skin or nerve biopsies, which requires performing an invasive technique. In **chapter 5**, we aimed to develop a predictive tran-

scriptomic signature in blood that could assess whether an individual intensely exposed to *M. leprae* would develop leprosy. For this purpose, HC of leprosy patients (n=5,352) in Bangladesh have been followed up since 2013 and blood samples were collected at recruitment and if individuals developed leprosy (n=85). Gene expression differences between leprosy progressors and HC who remained without symptoms were studied using RNA-Seq. Minimal longitudinal intra-individual variation was found in gene expression of leprosy progressors between the pre-symptomatic phase and the time of clinical diagnosis of leprosy. This demonstrates that gene expression differences between healthy individuals and those who will develop leprosy can be observed months before clinical symptoms are visible.

A 4-gene transcriptomic signature, designated RISK4LEP, was identified and validated in Bangladesh. This RNA signature could identify HC of leprosy patients who developed leprosy before symptoms are visible and 4 to 61 months before clinical diagnosis. A machine learning algorithm, random forest, was used to identify the signature which was validated by reverse transcription quantitative PCR (RT-qPCR). RISK4LEP is formed by 4 genes: *MT-ND2*, *REX1BD*, *TPGS1* and *UBC*. Interestingly, *MT-ND2*, which encodes a mitochondrial NADH dehydrogenase, together with other mitochondrial genes involved in oxidative phosphorylation such as *MT-ND4*, *MT-ND5*, *MT-CO1*, and *MT-CYB* presented lower expression in individuals who would develop leprosy. Previous studies have also reported a lower expression of mitochondrial genes involved in oxidative phosphorylation during *M. leprae* infection (59, 60) as well as association of leprosy with genetic polymorphisms in mitochondrial genes (61, 62). This reduction caused by *M. leprae*, promotes intracellular bacterial survival through decreased oxidative stress and inflammasome activation as well as inhibition of apoptosis and xenophagy (60, 63, 64). Thus, individuals who present a lower expression of these genes face complications to efficiently eliminate *M. leprae* (65). Other pathways such as the mTOR signaling pathway and the coronavirus pathogenesis pathway were upregulated in leprosy progressors. The mTOR pathway which, among other functions, plays a role in immune receptor signaling and migratory activity, has also been found to be deregulated in rheumatoid arthritis, type 2 diabetes, cancer and other viral infections (66, 67). The upregulation of the coronavirus pathogenesis pathways is likely due to an activation of the inflammatory and autophagy regulation pathways which are observed both in coronaviruses infections and BT leprosy (68, 69).

The RISK4LEP signature, which was validated in an independent sample set from Bangladesh, presented a good predictive value with an Area Under the receiver operating characteristic Curve (AUC) in the validation set of 86.4%. This RNA signature was developed and validated using samples from mostly-BT Bangladeshi leprosy patients and their HC. Further research is needed to validate this signature in other populations, particularly in

South American and African countries, as well as to determine the efficiency of the signature to predict leprosy in the lepromatous (LL) pole. Additionally, future research should also include the use of the signature for monitoring efficient treatment of leprosy by follow up of patients during MDT.

Transcriptomic biomarkers to predict reversal reactions

Leprosy reactions are episodes of increased inflammation occurring unpredictably before, during or after multidrug treatment (MDT) (70, 71). These reactions are the main cause of leprosy-associated irreversible neuropathy and 30-50% of leprosy patients suffer an episode, mostly borderline lepromatous patients with substantial bacterial loads (72). Reactions are often late- or misdiagnosed (73), which may result into permanent neuropathy (74) or disabilities caused by ulcers and other recurrent pathologies (75). In **chapter 6**, we employed transcriptomics to identify a 5-gene signature which predicted leprosy reversal reactions (RR) ≥2 months before onset in leprosy patients from four endemic areas: Bangladesh, Brazil, Ethiopia and Nepal. In addition, a selection of genes differentially expressed between leprosy patients and HC and between different types of leprosy patients were identified. However, in comparison to **chapter 5**, where an unbiased selection was performed through RNA-Seq, in **chapter 6** gene expression was assessed for 103 genes using dual color Reverse-Transcription Multiplex Ligation-dependent Probe Amplification (dcRT-MLPA). Moreover, the RISK4LEP signature predicts leprosy in HC before symptoms have occurred whilst the leprosy reaction risk signatures in **chapter 6** were identified in patients at clinical diagnosis of leprosy.

The 5-gene RR signature is formed by *CCL2, CD8A, IL2, IL15* and *MARCO*. *CCL2* activates pro-inflammatory cytokines and attracts CD4+ T cells (76) whilst *IL2* and *IL15* decrease Th2 regulation and increase Th1. A previous study, had also identified a gene set signature for RR (77). This signature was identified in the Vietnamese population using *M. leprae*-sonicate stimulated whole blood, however, it included a high number of genes (n=44) which hinders its application in point of care (POC) tests. Applying our 5-gene RR signature in a POC test, would allow reduction of reaction-associated neuropathy. This 5-gene RR signature prospectively identified RR in four different leprosy endemic areas. Further validation in extended leprosy endemic populations will determine if the signature can be applied worldwide to prospectively detect RR and prevent irreversible nerve damage due to leprosy reactions.

Analysis of longitudinal (intraindividual) gene expression before, during and after RR identified an upregulation of IFN-induced genes during RR: *IFI44, IFI35, IFI44L, IFI6, IFIH1, CXCL10 and FCGR1A*. Expression of IFN-induced genes decreased after reaction treatment. RNA-Seq transcriptomic analysis of the longitudinal samples described in **chapter 6** was performed in a separate study to provide further information about the mechanisms and

pathways associated with RR (78). RNA-Seq data showed an increased expression of IFN-γ and IFN-β pathways due to the host antimicrobial response during RR. This is in line with the observed upregulation of IFN-signaling genes we found longitudinally using dcRT-MLPA. Moreover, an alternative RR signature consisting of 434 genes was identified in the RNA-Seq analysis, which in addition to the IFN-signaling genes included other genes involved in the antimicrobial response against mycobacteria such as the guanylate binding protein (GBP) family.

Potential application of transcriptomics in POC diagnostic tests

Early diagnosis and treatment of leprosy and leprosy reactions are the main topics that need to be addressed to reduce leprosy-associated disabilities. Implementation of POC diagnostic test, particularly predictive tests, for leprosy and leprosy reactions could substantially improve the clinical outcome of leprosy patients and would signify an advance in leprosy health care. In addition, biomarker signatures are of interest for targeting intervention strategies such as post-exposure prophylaxis with single dose rifampicin.

Transcriptomic biomarkers are often identified using unbiased approaches that include a large quantity of tested genes such as RNA-Seq or microarrays. These methods are useful for the identification of biomarkers, however, they are not field friendly techniques, entailing high costs, the use of specialized equipment and requiring significant expertise to interpret the results. Thus, after identification of differentially expressed genes, a small selection is employed in transcriptomic signatures for which expression levels are studied using quantitative RT-qPCR or similar targeted techniques.

Skin biopsies, nerve biopsies or cell culture are often used to study the pathogenesis of *M. leprae* since gene expression at the site of infection provides more thorough evidence. However, for a POC diagnostic test, easy-to collect samples such as urine or blood are preferred. Collecting capillary blood by finger stick blood is less burdensome and costly and requires less skills than collecting venous blood or SSS. Thus, the use of capillary blood facilitates the implementation of POC tests. Recently, a transcriptomic signature for tuberculosis, RISK6, was evaluated in venous and capillary blood (52). Small volumes of capillary blood showed similar performance (equivalent AUC) of the signature as when venous blood was used (79). In addition, proteomic POC test with lateral flow strips utilizing luminescent up-converting reporter particles have been developed for leprosy using capillary blood (80, 81). Other samples such as urine or saliva have also been employed in host transcriptomic diagnostic tests, particularly for the highly inflammatory leprosy reactions, and cancer and could be further considered in the future for leprosy diagnostics as well (82-84).

POC diagnostic tests might also be applied longitudinally to monitor transcriptomic changes. This is relevant to predict reactions in leprosy patients, since one marker may

reflect different disease mechanisms or stages at different time points. As an example, *IL15* was found to be higher expressed in leprosy patients before occurrence of RR with a successive longitudinal decrease leading to RR (**chapter 6**). Thus, longitudinal follow up of leprosy patients utilizing POC test may expose patterns useful to predict certain developments such as leprosy reactions or monitoring treatment efficiency.

The future challenge that transcriptomic biomarkers face is the evaluation and translation of transcriptomic signatures into POC tests that can be applied by clinicians in the field. For this purpose and to ensure implementation in endemic areas, novel, easy to interpret and economic methods need to be developed and implemented (85).

Leprosy host genetics in Bangladesh

Over the years it has been shown that host genetic variation plays an important role in leprosy outcome, including the clinical type of leprosy and occurrence of reactions (86-89). Results are not always replicated in different studies and conditions may vary, leading to incongruous findings. Thus, it remains important to validate genetic associations with disease susceptibility, particularly across populations from different ethnic backgrounds.

In **chapter 7**, we found three genetic markers (rs1801224, rs13001714 and rs1801582) that were associated with leprosy in a family study analysis in Bangladesh, using a targeted approach based on SNPs identified in the Prata Village (Brazil) (manuscript in preparation) as well as previous studies (87, 90-92). Although the results were not replicated in a separate control-case set from the same area in Bangladesh, our results are a validation of the association of these SNPs with leprosy in a distinct population.

The identified genetic markers are located at *CUBN*, *IL1RL1* and *PRKN* (formerly known as *PARK2*) genes. Cubilin, coded by *CUBN* is a crucial receptor for uptake of vitamin B12 (93). Obligate pathogens require vitamin B12 from their host to survive. *M. leprae* protein ML2084 is a homolog of Rv1819c, a *Mycobacterium tuberculosis* protein responsible for vitamin B12 uptake (94). It is to be expected that human cubilin is competing for the same substrate with mycobacteria, thus genetic variations in the *CUBN* gene likely impact this function (95). *IL1RL1* encodes the receptor of IL-33, IL-1 receptor ST 2 (96, 97). IL-33 receptor activates the nuclear factor-kB (NF-kB) and MAP kinases. In addition, *IL1RL1*, expressed in Th2 cells, drives production of Th2 associated cytokines IL-4, IL-5, and IL-13. These processes are crucial for the inflammatory response and mycobacterial immunity (87, 96). Parkin, encoded by *PRKN*, is a part of the multiprotein E3 ubiquitin ligase (98). Parking regulates autophagy of mycobacteria and damaged mitochondria and inhibits mycobacterial replication (99, 100). Interestingly, *PRKN* (presenting a SNP associated with leprosy in **chapter 7**) together with *UBC*, which was identified in **chapter 5** to be part of the leprosy predictive RISK4LEP signature, are part of the ubiquitin system. The ubiquitin system is involved in the innate immune response to mycobacteria as shown for tubercu-

losis (101). It appears that both, genetic variants of *PRKN* and differential gene expression of *UBC*, have a role in leprosy development, emphasizing the importance of the ubiquitin system in response to *M. leprae* infection.

Genetic variants associated with susceptibility to leprosy can be incorporated into a genetic profile to assess individuals exposed to *M. leprae* who present a higher risk of developing leprosy. Thus, the combination of demographic characteristics, pathogen detection, transcriptomic, and/or proteomic biomarkers and genetic markers can lead to a multifactorial leprosy signature applicable to early diagnosis of leprosy and/or guide intervention strategies.

Concluding remarks

In this thesis we combined the study of *M. leprae* with host transcriptomic and genetic biomarkers as a tool to understand and reduce transmission as well as early detect leprosy. We detected *M. leprae* in HC of leprosy patients as well as soil from several locations, indicating that both sources may play a role in transmission. In Bangladesh, we identified a new *M. leprae* genotype and determined the distribution of *M. leprae* genotypes in four districts. We proposed a new classification of genotype 1C and observed differences between *M. leprae* strains isolated from NS or SSS. We showed that Dutch and Belgian squirrels do not carry the leprosy bacilli, therefore, limiting the disease in squirrels to the British Isles (9). We presented evidence that a One Health approach bears value to identify new reservoirs, although further studies are necessary to determine the extent of the role of these reservoirs in transmission.

Moreover, we identified the first host transcriptomic signature in leprosy, RISK4LEP, that predicts leprosy among HC before clinical symptoms are visible. We described a 5-gene transcriptomic signature that could predict occurrence of leprosy reactions in leprosy patients. Validation of these signatures in different populations will confirm their potential to predict leprosy (reactions) in general. In addition, genetic markers associated with susceptibility to leprosy were also validated in Bangladesh. These SNPs could be implemented in combination with other biomarkers into a leprosy signature to assess individuals at higher risk of developing leprosy. The next challenges in leprosy diagnostics will be to implement these signatures in easy to interpret and field-friendly POC tests.

References

1. Schuenemann VJ, Avanzi C, Krause-Kyora B, Seitz A, Herbig A, Inskip S, et al. Ancient genomes reveal a high diversity of *Mycobacterium leprae* in medieval Europe. PLoS pathogens. 2018;14(5):e1006997.
2. Donoghue HD. Tuberculosis and leprosy associated with historical human population movements in Europe and beyond - an overview based on mycobacterial ancient DNA. Annals of human biology. 2019;46(2):120-8.
3. Monot M, Honore N, Garnier T, Araoz R, Coppee JY, Lacroix C, et al. On the origin of leprosy. Science. 2005;308(5724):1040-2.
4. Benjak A, Avanzi C, Singh P, Loiseau C, Girma S, Busso P, et al. Phylogenomics and antimi-

crobial resistance of the leprosy bacillus *Mycobacterium leprae*. Nat Commun. 2018;9(1):352.
5. Avanzi C, Singh P, Truman RW, Suffys PN. Molecular epidemiology of leprosy: An update. Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases. 2020;86:104581.
6. Monot M, Honore N, Garnier T, Zidane N, Sherafi D, Paniz-Mondolfi A, et al. Comparative genomic and phylogeographic analysis of *Mycobacterium leprae*. Nature genetics. 2009;41(12):1282-9.
7. Weng X, Xing Y, Liu J, Wang Y, Ning Y, Li M, et al. Molecular, ethno-spatial epidemiology of leprosy in China: Novel insights for tracing leprosy in endemic and non endemic provinces. Infection, Genetics and Evolution. 2013;14:361-8.
8. Donoghue HD, Michael Taylor G, Marcsik A, Molnár E, Pálfi G, Pap I, et al. A migration-driven model for the historical spread of leprosy in medieval Eastern and Central Europe. Infection, Genetics and Evolution. 2015;31:250-6.
9. Avanzi C, Del-Pozo J, Benjak A, Stevenson K, Simpson VR, Busso P, et al. Red squirrels in the British Isles are infected with leprosy bacilli. Science. 2016;354(6313):744-7.
10. Schilling A-K, Avanzi C, Ulrich RG, Busso P, Pisanu B, Ferrari N, et al. British Red Squirrels Remain the Only Known Wild Rodent Host for Leprosy Bacilli. Frontiers in Veterinary Science. 2019;6(8).
11. Truman RW, Singh P, Sharma R, Busso P, Rougemont J, Paniz-Mondolfi A, et al. Probable zoonotic leprosy in the Southern United States. The New England journal of medicine. 2011;364(17):1626-33.
12. Sharma R, Singh P, Loughry WJ, Lockhart JM, Inman WB, Duthie MS, et al. Zoonotic leprosy in the Southeastern United States. Emerg Infect Dis. 2015;21(12):2127-34.
13. Honap TP, Pfister LA, Housman G, Mills S, Tarara RP, Suzuki K, et al. *Mycobacterium leprae* genomes from naturally infected nonhuman primates. PLoS neglected tropical diseases. 2018;12(1):e0006190.
14. Schuenemann VJ, Singh P, Mendum TA, Krause-Kyora B, Jager G, Bos KI, et al. Genome-wide comparison of medieval and modern *Mycobacterium leprae*. Science. 2013;341(6142):179-83.
15. Mendum TA, Schuenemann VJ, Roffey S, Taylor GM, Wu H, Singh P, et al. *Mycobacterium leprae* genomes from a British medieval leprosy hospital: towards understanding an ancient epidemic. BMC genomics. 2014;15:270.
16. Suzuki K, Takigawa W, Tanigawa K, Nakamura K, Ishido Y, Kawashima A, et al. Detection of *Mycobacterium leprae* DNA from archaeological skeletal remains in Japan using whole genome amplification and polymerase chain reaction. PloS one. 2010;5(8):e12422.
17. Krause-Kyora B, Nutsua M, Boehme L, Pierini F, Pedersen DD, Kornell S-C, et al. Ancient DNA study reveals HLA susceptibility locus for leprosy in medieval Europeans. Nat Commun. 2018;9(1):1569-.
18. Avanzi C, Lécorché E, Rakotomalala FA, Benjak A, Rapelanoro Rabenja F, Ramarozatovo LS, et al. Population Genomics of *Mycobacterium leprae* Reveals a New Genotype in Madagascar and the Comoros. Frontiers in microbiology. 2020;11(711).
19. Cambau E, Saunderson P, Matsuoka M, Cole ST, Kai M, Suffys P, et al. Antimicrobial resistance in leprosy: results of the first prospective open survey conducted by a WHO surveillance network for the period 2009-15. Clin Microbiol Infect. 2018;24(12):1305-10.
20. World Health Organization. A guide for surveillance of antimicrobial resistance in leprosy. New Delhi: World Health Organization, Region Office for South-East Asia; 2017.
21. Richardus JH, Tiwari A, Barth-Jaeggi T, Arif MA, Banstola NL, Baskota R, et al. Leprosy post-exposure prophylaxis with single-dose rifampicin (LPEP): an international feasibility programme. Lancet Glob Health. 2021;9(1):e81-e90.
22. Hockings KJ, Mubemba B, Avanzi C, Pleh K, Düx A, Bersacola E, et al. Leprosy in wild chimpanzees. bioRxiv. 2020:2020.11.10.374371.
23. Lavania M, Katoch K, Sachan P, Dubey A, Kapoor S, Kashyap M, et al. Detection of *Mycobacterium leprae* DNA from soil samples by PCR targeting RLEP sequences. J Commun Dis. 2006;38(3):269-73.
24. Turankar RP, Lavania M, Chaitanya VS, Sengupta U, Darlong J, Darlong F, et al. Single nucleotide polymorphism-based molecular typing of *M. leprae* from multicase families of leprosy patients and their surroundings to understand the transmission of leprosy. Clin Microbiol Infect. 2014;20(3):O142-9.
25. Turankar RP, Lavania M, Singh M, Siva Sai KS, Jadhav RS. Dynamics of *Mycobacterium leprae* transmission in environmental context: deciphering the role of environment as a potential reservoir. Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases. 2012;12(1):121-6.
26. Turankar RP, Lavania M, Singh M, Sengupta U, Siva Sai K, Jadhav RS. Presence of viable *Mycobacterium leprae* in environmental specimens around houses of leprosy patients. Indian J Med Microbiol. 2016;34(3):315-21.

27. Turankar RP, Pandey S, Lavania M, Singh I, Nigam A, Darlong J, et al. Comparative evaluation of PCR amplification of RLEP, 16S rRNA, rpoT and Sod A gene targets for detection of *M. leprae* DNA from clinical and environmental samples. Int J Mycobacteriol. 2015;4(1):54-9.
28. Lavania M, Katoch K, Katoch VM, Gupta AK, Chauhan DS, Sharma R, et al. Detection of viable *Mycobacterium leprae* in soil samples: insights into possible sources of transmission of leprosy. Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases. 2008;8(5):627-31.
29. Mohanty PS, Naaz F, Katara D, Misba L, Kumar D, Dwivedi DK, et al. Viability of *Mycobacterium leprae* in the environment and its role in leprosy dissemination. Indian J Dermatol Venereol Leprol. 2016;82(1):23-7.
30. Holanda MV, Marques LEC, Macedo MLB, Pontes MAA, Sabadia JAB, Kerr L, et al. Presence of *Mycobacterium leprae* genotype 4 in environmental waters in Northeast Brazil. Rev Soc Bras Med Trop. 2017;50(2):216-22.
31. Arraes M, Holanda MV, Lima L, Sabadia JAB, Duarte CR, Almeida RLF, et al. Natural environmental water sources in endemic regions of northeastern Brazil are potential reservoirs of viable *Mycobacterium leprae*. Mem Inst Oswaldo Cruz. 2017;112(12):805-11.
32. Bonnar PE, Cunningham NP, Boggild AK, Walsh NM, Sharma R, Davis IRC. Leprosy in Nonimmigrant Canadian Man without Travel outside North America, 2014. Emerg Infect Dis. 2018;24(1):165-6.
33. da Silva MB, Portela JM, Li W, Jackson M, Gonzalez-Juarrero M, Hidalgo AS, et al. Evidence of zoonotic leprosy in Pará, Brazilian Amazon, and risks associated with human contact or consumption of armadillos. PLoS neglected tropical diseases. 2018;12(6):e0006532.
34. Araujo S, Freitas LO, Goulart LR, Goulart IM. Molecular evidence for the aerial route of infection of *Mycobacterium leprae* and the role of asymptomatic carriers in the persistence of leprosy. Clin Infect Dis. 2016;63(11):1412-20.
35. Bratschi MW, Steinmann P, Wickenden A, Gillis TP. Current knowledge on *Mycobacterium leprae* transmission: a systematic literature review. Leprosy review. 2015;86(2):142-55.
36. Smith WC, Smith CM, Cree IA, Jadhav RS, Macdonald M, Edward VK, et al. An approach to understanding the transmission of Mycobacterium leprae using molecular and immunological methods: results from the MILEP2 study. International journal of leprosy and other mycobacterial diseases : official organ of the International Leprosy Association. 2004;72(3):269-77.
37. Desikan KV, Sreevatsa. Extended studies on the viability of *Mycobacterium leprae* outside the human body. Leprosy review. 1995;66(4):287-95.
38. Wheat WH, Casali AL, Thomas V, Spencer JS, Lahiri R, Williams DL, et al. Long-term survival and virulence of *Mycobacterium leprae* in amoebal cysts. PLoS neglected tropical diseases. 2014;8(12):e3405.
39. Neumann Ada S, Dias Fde A, Ferreira Jda S, Fontes AN, Rosa PS, Macedo RE, et al. Experimental Infection of Rhodnius prolixus (Hemiptera, Triatominae) with *Mycobacterium leprae* Indicates Potential for Leprosy Transmission. PloS one. 2016;11(5):e0156037.
40. Ferreira JDS, Souza Oliveira DA, Santos JP, Ribeiro C, Baêta BA, Teixeira RC, et al. Ticks as potential vectors of *Mycobacterium leprae*: Use of tick cell lines to culture the bacilli and generate transgenic strains. PLoS neglected tropical diseases. 2018;12(12):e0007001.
41. Meredith A, Pozo JD, Smith S, Milne E, Stevenson K, McLuckie J. Leprosy in red squirrels in Scotland. Veterinary Record. 2014;175(11):285-6.
42. Simpson V, Hargreaves J, Butler H, Blackett T, Stevenson K, McLuckie J. Leprosy in red squirrels on the Isle of Wight and Brownsea Island. Veterinary Record. 2015;177(8):206-7.
43. Butler H, Stevenson K, McLuckie J, Simpson V. Further evidence of leprosy in Isle of Wight red squirrels. Vet Rec. 2017;180(16):407.
44. Walsh GP, Storrs EE, Meyers W, Binford CH. Naturally acquired leprosy-like disease in the nine-banded armadillo (*Dasypus novemcinctus*): recent epizootiologic findings. J Reticuloendothel Soc. 1977;22(4):363-7.
45. Smith JH, File SK, Nagy BA, Folse DS, Buckner JA, Webb LJ, et al. Leprosy-like disease of wild armadillos in French Acadiana, Louisiana. J Reticuloendothel Soc. 1978;24(6):705-19.
46. Stallknecht DE, Truman RW, Hugh-Jones ME, Job CK. Surveillance for naturally acquired leprosy in a nine-banded armadillo population. J Wildl Dis. 1987;23(2):308-10.
47. Loughry WJ, Truman RW, McDonough CM, Tilak MK, Garnier S, Delsuc F. Is leprosy spreading among nine-banded armadillos in the southeastern United States? J Wildl Dis. 2009;45(1):144-52.
48. Schmitt JV, Dechandt IT, Dopke G, Ribas ML, Cerci FB, Viesi JMZ, et al. Armadillo meat intake was not associated with leprosy in a case control study, Curitiba (Brazil). Memórias do Instituto Oswaldo Cruz. 2010;105:857-62.
49. Stefani MMA, Rosa PS, Costa MB, Schetinni APM, Manhães I, Pontes MAA, et al. Leprosy survey among rural communities and wild armadillos from Amazonas state, Northern Brazil. PloS

one. 2019;14(1):e0209491.
50. Zak DE, Penn-Nicholson A, Scriba TJ, Thompson E, Suliman S, Amon LM, et al. A blood RNA signature for tuberculosis disease risk: a prospective cohort study. Lancet. 2016.
51. Suliman S, Thompson E, Sutherland J, Weiner Rd J, Ota MOC, Shankar S, et al. Four-gene Pan-African Blood Signature Predicts Progression to Tuberculosis. American journal of respiratory and critical care medicine. 2018.
52. Penn-Nicholson A, Mbandi SK, Thompson E, Mendelsohn SC, Suliman S, Chegou NN, et al. RISK6, a 6-gene transcriptomic signature of TB disease risk, diagnosis and treatment response. Scientific reports. 2020;10(1):8629.
53. Jorge K, Souza RP, Assis MTA, Araujo MG, Locati M, Jesus AMR, et al. Characterization of MicroRNA Expression Profiles and Identification of Potential Biomarkers in Leprosy. Journal of clinical microbiology. 2017;55(5):1516-25.
54. Montoya DJ, Andrade P, Silva BJA, Teles RMB, Ma F, Bryson B, et al. Dual RNA-Seq of Human Leprosy Lesions Identifies Bacterial Determinants Linked to Host Immune Response. Cell reports. 2019;26(13):3574-85.e3.
55. Inkeles MS, Teles RM, Pouldar D, Andrade PR, Madigan CA, Lopez D, et al. Cell-type deconvolution with immune pathways identifies gene networks of host defense and immunopathology in leprosy. JCI insight. 2016;1(15):e88843.
56. Realegeno S, Kelly-Scumpia KM, Dang AT, Lu J, Teles R, Liu PT, et al. S100A12 Is Part of the Antimicrobial Network against *Mycobacterium leprae* in Human Macrophages. PLoS pathogens. 2016;12(6):e1005705.
57. Bleharski JR, Li H, Meinken C, Graeber TG, Ochoa MT, Yamamura M, et al. Use of genetic profiling in leprosy to discriminate clinical forms of the disease. Science. 2003;301(5639):1527-30.
58. Salgado CG, Pinto P, Bouth RC, Gobbo AR, Messias ACC, Sandoval TV, et al. miRNome Expression Analysis Reveals New Players on Leprosy Immune Physiopathology. Frontiers in Immunology. 2018;9(463).
59. Guerreiro LT, Robottom-Ferreira AB, Ribeiro-Alves M, Toledo-Pinto TG, Rosa Brito T, Rosa PS, et al. Gene expression profiling specifies chemokine, mitochondrial and lipid metabolism signatures in leprosy. PloS one. 2013;8(6):e64748.
60. Dubey R. Assuming the role of mitochondria in mycobacterial infection. International Journal of Mycobacteriology. 2016;5(4):379-83.
61. Xing Y, He J, Wen Y, Liu J, You Y, Weng X, et al. Polymorphisms in mitochondrial ribosomal protein S5 (MRPS5) are associated with leprosy risk in Chinese. PLoS neglected tropical diseases. 2020;14(12):e0008883.
62. Ramos GB, Salomão H, Francio AS, Fava VM, Werneck RI, Mira MT. Association Analysis Suggests SOD2 as a Newly Identified Candidate Gene Associated With Leprosy Susceptibility. J Infect Dis. 2016;214(3):475-8.
63. Thomas LW, Ashcroft M. Exploring the molecular interface between hypoxia-inducible factor signalling and mitochondria. Cellular and Molecular Life Sciences. 2019;76(9):1759-77.
64. Dröse S, Brandt U. Molecular mechanisms of superoxide production by the mitochondrial respiratory chain. Adv Exp Med Biol. 2012;748:145-69.
65. Vanlalhruaii Tonsing M, Vanlalbiakdiki Sailo C, Zothansanga, Chhakchhuak L, Chhakchhuak Z, Pandit B, et al. Analysis of variants in mitochondrial genome and their putative pathogenicity in tuberculosis patients from Mizoram, North east India. Mitochondrion. 2020;54:21-5.
66. Laplante M, Sabatini DM. mTOR signaling at a glance. Journal of Cell Science. 2009;122(20):3589-94.
67. Le Sage V, Cinti A, Amorim R, Mouland AJ. Adapting the Stress Response: Viral Subversion of the mTOR Signaling Pathway. Viruses. 2016;8(6):152.
68. Sekine T, Perez-Potti A, Rivera-Ballesteros O, Strålin K, Gorin J-B, Olsson A, et al. Robust T cell immunity in convalescent individuals with asymptomatic or mild COVID-19. bioRxiv. 2020:2020.06.29.174888.
69. Vardhana SA, Wolchok JD. The many faces of the anti-COVID immune response. Journal of Experimental Medicine. 2020;217(6).
70. Lockwood DN, Saunderson P. Nerve damage in Leprosy: a continuing challenge for scientists, clinicians and service providers. Int Health. 2012;4(2012):77-85.
71. Scollard DM, Martelli CM, Stefani MM, Maroja MF, Villahermosa L, Pardillo F, et al. Risk factors for leprosy reactions in three endemic countries. Am J Trop Med Hyg. 2015;92(1):108-14.
72. Antunes DE, Ferreira GP, Nicchio MV, Araujo S, Cunha AC, Gomes RR, et al. Number of leprosy reactions during treatment: clinical correlations and laboratory diagnosis. Rev Soc Bras Med Trop. 2016;49(6):741-5.
73. Raffe SF, Thapa M, Khadge S, Tamang K, Hagge D, Lockwood DN. Diagnosis and treatment of leprosy reactions in integrated services - the patients' perspective in Nepal. PLoS neglected tropical diseases. 2013;7(3):e2089.

74. Croft RP, Nicholls PG, Steyerberg EW, Richardus JH, Withington SG, Smith WC. A clinical prediction rule for nerve function impairment in leprosy patients-revisited after 5 years of follow-up. Leprosy review. 2003;74(1):35-41.
75. Assis BPN, Lyon S, Grossi MAF, Rocha M. Risk factors for physical disability upon release from multidrug therapy in new cases of leprosy at a referral center in Brazil. Rev Inst Med Trop Sao Paulo. 2019;61:e13.
76. Medeiros MF, Rodrigues MM, Vital RT, da Costa Nery JA, Sales AM, de Andrea Hacker M, et al. CXCL10, MCP-1, and other immunologic markers involved in neural leprosy. Appl Immunohistochem Mol Morphol. 2015;23(3):220-9.
77. Orlova M, Cobat A, Huong NT, Ba NN, Van TN, Spencer J, et al. Gene set signature of reversal reaction type I in leprosy patients. PLoS Genet. 2013;9(7):e1003624.
78. Teles RMB, Lu J, Tió-Coma M, Goulart IMB, Banu S, Hagge D, et al. Identification of a systemic interferon-γ inducible antimicrobial gene signature in leprosy patients undergoing reversal reaction. PLoS Negl Trop Dis. 2019;13(10):e0007764.
79. van Hooij A, Tjon Kon Fat EM, Richardus R, van den Eeden SJ, Wilson L, de Dood CJ, et al. Quantitative lateral flow strip assays as User-Friendly Tools To Detect Biomarker Profiles For Leprosy. Scientific reports. 2016;6:34260.
80. van Hooij A, Tjon Kon Fat EM, de Jong D, Khatun M, Soren S, Chowdhury AS, et al. Prototype multi-biomarker test for point-of-care leprosy diagnostics. iScience. 2021;24(1):102006.
81. Corstjens P, van Hooij A, Tjon Kon Fat EM, Alam K, Vrolijk LB, Dlamini S, et al. Fingerstick test quantifying humoral and cellular biomarkers indicative for M. leprae infection. Clin Biochem. 2019;66:76-82.
82. Mayboroda OA, van Hooij A, Derks R, van den Eeden SJ, Dijkman K, Khadge S, et al. Exploratory urinary metabolomics of type 1 leprosy reactions. International journal of infectious diseases : IJID : official publication of the International Society for Infectious Diseases. 2016;45:46-52.
83. Lee B, Mahmud I, Marchica J, Dereziński P, Qi F, Wang F, et al. Integrated RNA and metabolite profiling of urine liquid biopsies for prostate cancer biomarker discovery. Scientific reports. 2020;10(1):3716.
84. Khurshid Z, Zafar MS, Khan RS, Najeeb S, Slowey PD, Rehman IU. Role of Salivary Biomarkers in Oral Cancer Detection. Adv Clin Chem. 2018;86:23-70.
85. Pennisi I, Rodriguez-Manzano J, Moniri A, Kaforou M, Herberg JA, Levin M, et al. Translation of a Host Blood RNA Signature Distinguishing Bacterial From Viral Infection Into a Platform Suitable for Development as a Point-of-Care Test. JAMA Pediatr. 2021;175(4):417-9.
86. Fava VM, Mira MT. Leprosy: A Practical Guide. Nunzi E, Massone C, editors. Milan: Springer; 2012.
87. Zhang FR, Huang W, Chen SM, Sun LD, Liu H, Li Y, et al. Genomewide association study of leprosy. N Engl J Med. 2009;361(27):2609-18.
88. Mira MT, Alcais A, Nguyen VT, Moraes MO, Di Flumeri C, Vu HT, et al. Susceptibility to leprosy is associated with PARK2 and PACRG. Nature. 2004;427(6975):636-40.
89. Fava VM, Dallmann-Sauer M, Schurr E. Genetics of leprosy: today and beyond. Hum Genet. 2020;139(6-7):835-46.
90. Zhang F, Liu H, Chen S, Low H, Sun L, Cui Y, et al. Identification of two new loci at IL23R and RAB32 that influence susceptibility to leprosy. Nature genetics. 2011;43(12):1247-51.
91. Liu H, Bao F, Irwanto A, Fu X, Lu N, Yu G, et al. An association study of TOLL and CARD with leprosy susceptibility in Chinese population. Hum Mol Genet. 2013;22(21):4430-7.
92. Liu H, Irwanto A, Fu X, Yu G, Yu Y, Sun Y, et al. Discovery of six new susceptibility loci and analysis of pleiotropic effects in leprosy. Nature genetics. 2015;47(3):267-71.
93. Seetharam B, Levine JS, Ramasamy M, Alpers DH. Purification, properties, and immunochemical localization of a receptor for intrinsic factor-cobalamin complex in the rat kidney. J Biol Chem. 1988;263(9):4443-9.
94. Gopinath K, Venclovas C, Ioerger TR, Sacchettini JC, McKinney JD, Mizrahi V, et al. A vitamin B_{12} transporter in *Mycobacterium tuberculosis*. Open Biol. 2013;3(2):120175.
95. Grant AV, Cobat A, Van Thuc N, Orlova M, Huong NT, Gaschignard J, et al. CUBN and NEBL common variants in the chromosome 10p13 linkage region are associated with multibacillary leprosy in Vietnam. Human Genetics. 2014;133(7):883-93.
96. Liu Hea. Identification of IL18RAP/IL18R1 and IL12B as leprosy risk genes demonstrates shared pathogenesis between inflammation and infectious diseases. Am J Hum Genet. 2012;91:935–41.
97. Schmitz J, Owyang A, Oldham E, Song Y, Murphy E, McClanahan TK, et al. IL-33, an interleukin-1-like cytokine that signals via the IL-1 receptor-related protein ST2 and induces T helper type 2-associated cytokines. Immunity. 2005;23(5):479-90.
98. Manzanillo PS, Ayres JS, Watson RO, Collins AC, Souza G, Rae CS, et al. The ubiquitin ligase parkin mediates resistance to intracellular pathogens. Nature. 2013;501(7468):512-6.

99. Cambri G, Mira MT. Genetic Susceptibility to Leprosy-From Classic Immune-Related Candidate Genes to Hypothesis-Free, Whole Genome Approaches. Frontiers in immunology. 2018;9:1674.
100. Manzanillo PS, Ayres JS, Watson RO, Collins AC, Souza G, Rae CS, et al. The ubiquitin ligase parkin mediates resistance to intracellular pathogens. Nature. 2013;501(7468):512-6.
101. Wang Z, Arat S, Magid-Slav M, Brown JR. Meta-analysis of human gene expression in response to *Mycobacterium tuberculosis* infection reveals potential therapeutic targets. BMC Syst Biol. 2018;12(1):3-.

English summary
Nederlandse samenvatting
Acknowledgements
Curriculum vitae
List of publications

English summary

Leprosy is a multifactorial chronic disease caused by *Mycobacterium leprae* or *Mycobacterium lepromatosis* that affects the skin and nerves. More than 200.000 new cases are diagnosed per year; thus, transmission is still ongoing. The most likely way of transmission is the respiratory route form human-to-human; however, transmission is still not clearly understood. Early diagnosis of leprosy is crucial to reduce and avoid transmission as well as leprosy-associated disabilities, which are also a cause of stigma. Currently, diagnosis is performed based on clinical signs and symptoms and late- or mis-diagnosis are not uncommon.

In this thesis, we combined the study of pathogen transmission with host transcriptomic and genomic biomarkers. To explore *M. leprae* transmission a One Health approach was followed, where human, animal and environmental samples were studied.

In **chapter 2**, *M. leprae* transmission in multibacillary leprosy patients as well as their household contacts (HC) was studied in Bangladesh. We observed that *M. leprae* was not only present in leprosy patients, but also in asymptomatic individuals. Particularly in the nasal cavities, with up to 18% of asymptomatic HC showing presence of *M. leprae* DNA and 36.8% with phenolic glycolipid I (PGL-I) IgM in serum.

M. leprae whole genomes have been retrieved over the years and a genotype classification was established consisting of four genotypes (1-4) and 16 subtypes (A-P). In **chapter 2**, *M. leprae* genome diversity in Bangladesh was explored and several subtypes of genotype 1 (1A, 1C and 1D) were identified with subtype 1D being the most prevalent. Importantly, we describe a new subtype, 1B-Bangladesh, only found in Bangladesh up until now. Moreover, we demonstrated that the subtype 1C does not constitute a separate subtype and is part of genotype 1D.

In **chapter 3**, we investigated whether soil could be a potential reservoir of the leprosy bacilli. We identified *M. leprae* DNA in soil from the houses of leprosy patients in Bangladesh and from armadillos' holes at former leprosy colonies in Suriname. Additionally, *M. lepromatosis* was detected in soil obtained around an area where *M. lepromatosis* infected squirrels are located in the Isle of Arran. Nevertheless, the low concentration of *M. leprae* or *M. lepromatosis* DNA found in soil suggests that environmental contamination as a source of infection is not very likely.

Leprosy is not exclusive to humans, as red squirrels, armadillos and non-human primates can become naturally infected with *M. leprae* or *M. lepromatosis*. Squirrels infected with *M. leprae* or *M. lepromatosis* were identified in the British Isles. In **chapter 4,** we investigated whether Dutch and Belgian squirrels were carriers of the leprosy bacilli. We examined 114 squirrels by quantitative PCR (qPCR) and we did not detect *M. leprae* or *M. lepromatosis* DNA in any of the animals. This is in line with previous findings in France, Germany, Swit-

zerland and Italy where PCR showed no presence of the leprosy bacilli either. Thus, up to the present time, squirrels infected with *M. leprae* or *M. lepromatosis* have only been identified in the British Isles. Transmission of the leprosy bacilli from red squirrels to humans has not been reported and due to limited squirrel-human interaction in the areas where squirrels with leprosy were found and the geographical limitations of islands, zoonotic transmission and transmission to other areas should not be a cause of major concerns. Nevertheless, it remains necessary to keep vigilant and include the study of animal and environmental reservoirs in strategies to effectively stop transmission.

Next, we searched for transcriptomic host biomarkers that could predict leprosy or leprosy reactions before occurrence of symptoms. In **chapter 5**, we aimed to develop a predictive transcriptomic signature in blood that could assess whether an individual intensely exposed to *M. leprae* would develop leprosy. Gene expression differences between leprosy progressors and HC who remained without symptoms were studied using RNA-Seq. Minimal longitudinal intra-individual variation was found in gene expression of leprosy progressors between the pre-symptomatic phase and the time of clinical diagnosis of leprosy. This indicates that gene expression differences between healthy individuals and those who will develop leprosy can be observed months before clinical symptoms are visible. A 4-gene transcriptomic signature, designated RISK4LEP, was identified and validated in Bangladesh. This RNA signature could identify HC of leprosy patients who developed leprosy before symptoms were visible, 4 to 61 months before clinical diagnosis. A machine learning algorithm, random forest, was used to identify the signature which was validated by reverse transcription quantitative PCR (RT-qPCR). RISK4LEP is based on the expression of 4 genes: *MT-ND2*, *REX1BD*, *TPGS1* and *UBC*.

Leprosy reactions are episodes of increased inflammation occurring unpredictably before, during or after multidrug treatment (MDT). Reactions are often late- or misdiagnosed, which may result into permanent neuropathy or disabilities caused by ulcers and other recurrent pathologies. In **chapter 6**, we employed dual color Reverse-Transcription Multiplex Ligation-dependent Probe Amplification (dcRT-MLPA) to identify a 5-gene signature which predicted leprosy reversal reactions (RR) ≥ 2 months before onset in leprosy patients from four endemic areas: Bangladesh, Brazil, Ethiopia and Nepal. The 5-gene RR signature is formed by *CCL2, CD8A, IL2, IL15* and *MARCO*. Applying our 5-gene RR signature in a POC test, would allow reduction of reaction-associated neuropathy.

Finally, in **chapter 7**, we investigated genomic biomarkers to identify individuals at higher risk of developing leprosy. Three genetic markers (rs1801224, rs13001714 and rs1801582) were associated with leprosy in a family study analysis in Bangladesh. These markers were previously described to be associated with leprosy in the Prata Village. Although the results were not replicated in a separate control-case set from the same area in Bangladesh,

our results are a validation of the association of these SNPs with leprosy in a distinct population. The described genetic markers are located at *CUBN*, *IL1RL1* and *PRKN* (formerly known as *PARK2*) genes. Genetic variants associated with susceptibility to leprosy can be incorporated into a genetic profile to identify individuals exposed to *M. leprae* who present a higher risk of developing leprosy.

The combination of demographic characteristics, pathogen detection, genetic and/or transcriptomic biomarkers can be applied in a multifactorial leprosy signature applicable for early diagnosis of leprosy and/or to guide intervention strategies. Identification of predictive biomarkers will in due course lead to prompt treatment, preventing leprosy-associated irreversible disabilities as well as reducing *M. leprae* transmission.

Nederlandse samenvatting

Lepra is een multifactoriële chronische ziekte, veroorzaakt door *Mycobacterium leprae* of *Mycobacterium lepromatosis,* die de huid en zenuwen aantast. Meer dan 200.000 nieuwe patiënten worden gediagnostiseerd per jaar; transmissie vindt dus nog steeds plaats. De meest waarschijnlijke transmissie route is de respiratoire route van mens op mens; Echter, de manier waarop transmissie plaats vindt is nog steeds niet helemaal bekend. Vroege lepra diagnose is cruciaal om transmissie en handicaps geassocieerd met lepra te verminderen en vermijden, deze handicaps zijn ook vaak de oorzaak van stigma. Op dit moment is de diagnose gebaseerd op klinische symptomen, waarbij late of misdiagnose regelmatig voorkomt.

In deze thesis hebben we de studie van pathogeen transmissie gecombineerd met host transcriptomic en genomische biomarkers. Om *M. leprae* transmissie te onderzoeken is een One Health benadering toegepast, waarin menselijke, dierlijke en omgevingssamples zijn bestudeerd.

In **hoofdstuk 2** is *M. leprae* transmissie in multibacillaire lepra patiënten en hun huishoud contacten (HC) bestudeerd in Bangladesh. We observeerde dat *M. leprae* niet alleen aanwezig was in lepra patiënten, maar ook in asymptomatische individuen. Vooral in de neusholten was *M. leprae* DNA aanwezig in de asymptomatische HC (tot 18%) en in serum werd in 36.8% anti-phenolic glycolipid I (PGL-I) IgM gedetecteerd.

M. leprae whole genomes zijn verkregen gedurende de tijd en een aan de hand daarvan is een genotype classificatie bepaald, bestaande uit 4 genotypes (1-4) en 16 subtypes (A-P). In **hoofdstuk 2** is de diversiteit van het *M. leprae* genoom in Bangladesh onderzocht en meerdere subtypes van genotype 1 (1A, 1C and 1D) werden geïdentificeerd, waarbij subtype 1D het meest prevalent was. Van belang is de beschrijving van een nieuw subtype, 1B-Bangladesh, dat tot nu toe alleen nog maar in Bangladesh werd gevonden. Bovendien hebben we laten zien dat subtype 1C geen afzonderlijk subtype vormt maar onderdeel is van genotype 1D.

In **hoofdstuk 3** hebben we onderzocht of aarde een potentieel reservoir voor de lepra bacterie kan zijn. We identificeerde *M. leprae* DNA in de aarde van huizen van lepra patienten in Bangladesh en van armadillo holen in voormalige lepra koloniën in Suriname. Daarnaast werd *M. lepromatosis* gedetecteerd in aarde verkregen rond een gebied waar *M. lepromatosis* geïnfecteerde eekhoorns voorkomen in the Isle of Arran. Desalniettemin, de lage concentratie *M. leprae* of *M. lepromatosis* DNA die werd gevonden in aarde suggereert dat milieu contaminatie als een bron van infectie niet heel waarschijnlijk is.

Lepra komt niet exclusief voor in mensen, aangezien rode eekhoorns, armadillos and niet-humane primaten ook natuurlijk geïnfecteerd kunnen raken met *M. leprae* or *M. lepromatosis*. Eekhoorns geïnfecteerd met *M. leprae* or *M. lepromatosis* zijn geïdentificeerd

op de Britse eilanden. In **hoofdstuk 4** hebben we onderzocht of Nederlandse en Belgische eekhoorns dragers waren van de lepra bacterie. We onderzochten 114 eekhoorns met quantitative PCR (qPCR) en detecteerden in geen enkel dier *M. leprae* or *M. lepromatosis* DNA. Dit is in de lijn der verwachting met eerdere bevindingen in Frankrijk, Duitsland, Zwitserland en Italië, waar PCR ook niet de aanwezigheid van de lepra bacterie liet zien. Dus, tot nu toe, zijn eekhoorns geïnfecteerd met *M. leprae* or *M. lepromatosis* alleen nog maar geïdentificeerd op de Britse eilanden. Transmissie van de lepra bacterie van rode eekhoorns op mensen is nog niet beschreven en vanwege gelimiteerde interactie tussen de mens en de eekhoorn in de gebieden waar eekhoorns met lepra zijn gevonden en de geografische limitatie van de eilanden, is er geen reden tot grote zorg voor zoönose en transmissie naar andere gebieden. Desalniettemin blijft het noodzakelijk om waakzaam te blijven en om de studie naar dierlijke en omgevingsreservoirs te includeren in strategieën om de transmissie effectief te stoppen.

Vervolgens zochten we naar transcriptomic host biomarkers die lepra of lepra reacties zouden kunnen voorspellen voordat de symptomen verschijnen. In **hoofdstuk 5** hadden we als doel om een voorspellend transcriptomic signature in bloed te ontwikkelen dat zou kunnen aangeven of een individu, die intensief werd blootgesteld aan *M. leprae,* lepra zou gaan ontwikkelen. Gen expressie verschillen tussen lepra progressors en HC die geen symptomen ontwikkelden werden bestudeerd, gebruik makend van RNA-Seq. Minimale longitudinale intra-individuele variatie werd gevonden in gen expressie van lepra progressors tussen de pre-symptomatische fase en het tijdspunt van klinische diagnose. Dit geeft aan dat gen expressie verschillen tussen gezonde individuen en degene die lepra gaan ontwikkelen al geobserveerd kunnen worden maanden voordat de klinische symptomen zichtbaar zijn. Een 4-genen transcriptomic signature, RISK4LEP genoemd, werd geïdentificeerd en gevalideerd in Bangladesh. Dit RNA signature kon HC (van lepra patiënten) die lepra ontwikkelden identificeren voordat de symptomen zichtbaar waren, 4 tot 61 maanden voor de klinische diagnose. Een machine learning algoritme, random forest, werd gebruikt om het signature te identificeren dat vervolgens werd gevalideerd met reverse transcription quantitative PCR (RT-qPCR). RISK4LEP bestaat uit de expressie van 4 genen: *MT-ND2*, *REX1BD*, *TPGS1* and *UBC*.

Lepra reacties zijn episodes van verhoogde inflammatie die onvoorspelbaar voorkomen voor, tijdens of na multi-drug treatment (MDT). Reacties worden vaak te laat of fout gediagnosticeerd, wat zou kunnen resulteren in permanente neuropathie of handicaps veroorzaakt door zweren en andere terugkerende pathologie. In **hoofdstuk 6**, hebben we dual color Reverse-Transcription Multiplex Ligation-dependent Probe Amplification (dcRT-MLPA) toegepast om een 5-genen signature te identificeren dat lepra reversal reactions (RR) kon voorspellen ≥ 2 maanden voor de start in lepra patiënten uit vier en-

demische gebieden: Bangladesh, Brazil, Ethiopia and Nepal. Het 5-genen RR signature wordt gevormd door *CCL2, CD8A, IL2, IL15* en *MARCO*. Het toepassen van ons 5-genen RR signature in een POC test zou kunnen leiden tot een reductie van reactie-geassocieerde neuropathie.

Tenslotte hebben we in **hoofdstuk 7** genomische biomarkers onderzocht om individuen te kunnen identificeren die een hoger risico hebben om lepra te ontwikkelen. Drie genetische markers (rs1801224, rs13001714 and rs1801582) waren geassocieerd met lepra in een familie studie analyse in Bangladesh. De associatie van deze markers met lepra was eerder beschreven in de Prata Village. Hoewel de resultaten niet gerepliceerd werden in een control-case set uit dezelfde regio in Bangladesh, zijn onze resultaten wel een validatie van de associatie van deze SNPs met lepra in een andere populatie. De beschreven genetische markers zijn gelokaliseerd in de genen *CUBN*, *IL1RL1* en *PRKN* (voormalig bekend als *PARK2*). Genetische varianten geassocieerd met susceptibiliteit voor lepra kunnen worden geïncorporeerd in een genetisch profiel om individuen te identificeren die een hoger risico hebben om lepra te ontwikkelen na blootstelling aan *M. leprae*.

De combinatie van demografische kenmerken, genetische en/of transcriptomic biomarkers, en pathogeen detectie kan worden gebruikt in een multifactorieel lepra signature toepaspaar voor vroege diagnostiek van lepra en/of om interventie strategieën te begeleiden. Identificatie van voorspellende biomarkers zal te zijner tijd leiden tot een snellere behandeling, waarbij lepra-geassocieerde handicaps kunnen worden voorkomen en ook de transmissie van *M. leprae* kan worden verminderd.

Acknowledgements

In this section I would like to acknowledge all the people who have helped and supported me during the development of this thesis.

Annemieke, I am very grateful that you gave me the opportunity to perform my PhD in your group. I have learnt a lot during my PhD and I have been able to meet and learn from many international collaborators. Thank you for encouraging me since the first day to start writing. Even if I did not realize it at the time, looking back I see how crucial that was to finish all the projects in 4 years. Thank you for giving me enough independence to carry on the different projects, especially towards the end of my PhD, and for encouraging and giving me the space to develop other tasks that I found interesting during my PhD such as the database management.

Anouk, I am so happy to have had such a wonderful friend sharing PhD experiences. We have been working together on many projects and I could always count on your help for anything. We were very proud to finish the sorting of all the WBA together with *Els* as well as many different students who helped with this. It was very nice to travel together to the International Leprosy Congress in the Philippines and I will always remember fondly our dinners by the pool and many other good times in the Netherlands! I am grateful that you have been there during my PhD and I am sure my PhD would not have been as fun without you. Thank you for everything!

Susan, thank you for guiding me and showing me so much when I started my PhD. Even if we did not work on the same projects at the end of my PhD, I could always count on you and I was happy to see you in the early hours of the morning. You create a great working environment in the department and I am also happy to have been able to hang out with you outside of work.

Els, it has been great working with you! I am really happy that you joined the team and that we could work together in the diagnostics, DNA isolations and other molecular work. You have been a great support during my PhD. I had a lot of fun talking with you about our trips and other life experiences. Thank you for been such a nice colleague!

Louise, you were a great first student and I am happy that you joined the team again after you finished your studies. Thank you for all your hard work!

Marloes, *Jamie* thank you for your contributions during your internships. You were both involved in the DNA isolations and other side projects that have been very valuable for this thesis.

Gal·la, moltes gràcies pel teu entusiasme i tota la feina que has fet per fer possible l'últim capítol d'aquesta tesis. Ha estat un plaer tenir-te com a estudiant!

Calinda, *Suus*, *Anne* thank you for the good times at the LUMC. I have enjoyed having all of

you as colleagues, you all brought a very positive energy to the group. *Mariateresa*, *Mikolaj*, *Cassandra*, *Matthias* and *Frank*, thank you for creating a welcoming environment when I started my PhD. I am happy to have shared many experiences as PhD students together. Thank you, *Annelies* and *Krista* for making sure I had everything I needed for my experiments. *Thomas* and *Mayara*, thank you for your initial work setting up some techniques that I continued during my PhD.

I would also like to thank all my INZI colleagues for the support and good times during these years: *Louis, Kimberley, Eleonora, Paula, Arthur, Carina, Josimar, Zije, Gül, Amy, Linda, Miriam, Roel, Michella, Tannie, Adriëtte, Bep, Tom, Simone, Peter,* and *Mariëlle***.**

To *Jan Hendrik* and colleagues from Erasmus MC, thank you for your help during these years, particularly with the database. To our collaborators in Bangladesh, *Khorshed, Johan, Marufa* and many more, thank you all for your hard work to collect samples from patients and contacts, solving all my questions about the database and taking care of all the organization in Bangladesh. To our collaborators in Brazil, *Marcelo, Priscilla* and *Andressa*, thank you for your support and advice on the genetic study. And to other collaborators for their support, advice and help to collect samples, thank you: *Charlotte, Toine* and *Anna*.

Mama, papa i *Uri*, gràcies per donar-me suport en tots els projectes que he decidit emprendre i per haver sigut sempre un exemple a seguir.

Sergi, gràcies per estar el meu costat i ajudar-me en tots els moments que he necessitat, tant a casa com professionalment. Sense el teu suport aquest camí hagués estat molt més difícil. Gràcies per haver emprès aquesta aventura amb mi, al teu costat qualsevol cosa és possible!

Curriculum vitae

Maria Tió Coma was born on the 4th of June 1993, in Gelida. In 2011, she started her Bachelor studies in Microbiology at the Autonomous University of Barcelona. During her Bachelor's degree, she obtained a personal collaboration grant from the Spanish Ministry of Education to join the laboratory of Dr. Artur Xavier Roig and Dr. Ramón Gervilla at the Department of Animal and Food Sciences, where she worked on the evaluation of Ultraviolet C light (UVC) and Ultra High-Pressure Homogenization (UHPH) to inactivate *Alicyclobacillus acidoterrestris* spores in apple juice. Due to her strong interest in Molecular Biology, in 2016 she started a Master in Molecular Biotechnology at the University of Barcelona. During her Master studies, she obtained an Erasmus+ grant to perform an internship at the Department of Experimental Virology at Amsterdam Medical Center, under the supervision of Dr. Lia van der Hoek. The topic of the Master's thesis was to develop a protocol to identify and characterize new viruses transmitted by insect vectors using Next Generation Sequencing to aid in the identification of the cause of Nodding Disease. In May 2017, she started her PhD studies at the Leiden University Medical Center in the group of Prof. Annemieke Geluk with the aim to identify pathogen and host biomarkers for leprosy in a leprosy endemic area in Bangladesh. The results of her studies have been published in international scientific journals and are described in this thesis. In 2021, she decided to use her scientific background to advice and help other scientists and technicians as a Technical Support Specialist.

List of publications

1. **Tió-Coma M**, Wijnands T, Pierneef L, Schilling AK, Alam K, Roy JC, Faber WR, Menke H, Pieters T, Stevenson K, Richardus JH, Geluk A. Detection of Mycobacterium leprae DNA in soil: multiple needles in the haystack. Sci Rep. 2019 Feb 28;9(1):3165.

2. Teles RMB, Lu J, **Tió-Coma M**, Goulart IMB, Banu S, Hagge D, Bobosha K, Ottenhoff THM, Pellegrini M, Geluk A, Modlin RL. Identification of a systemic interferon-γ inducible antimicrobial gene signature in leprosy patients undergoing reversal reaction. PLoS Negl Trop Dis. 2019 Oct 10;13(10):e0007764.

3. **Tió-Coma M**, van Hooij A, Bobosha K, van der Ploeg-van Schip JJ, Banu S, Khadge S, Thapa P, Kunwar CB, Goulart IM, Bekele Y, Hagge DA, Moraes MO, Teles RMB, Pinheiro RO, van Zwet EW, Goeman JJ, Aseffa A, Haks MC, Ottenhoff THM, Modlin RL, Geluk A. Whole blood RNA signatures in leprosy patients identify reversal reactions before clinical onset: a prospective, multicenter study. Sci Rep. 2019 Nov 29;9(1):17931.

4. Van Dissel JT, Pieters T, Geluk A, Maat G, Menke HE, **Tió-Coma M**, Altena E, Laros JFJ, Adhin MR. Archival, paleopathological and aDNA-based techniques in leprosy research and the case of Father Petrus Donders at the Leprosarium 'Batavia', Suriname. Int J Paleopathol. 2019 Dec;27:1-8.

5. **Tió-Coma M**, Sprong H, Kik M, van Dissel JT, Han XY, Pieters T, Geluk A. Lack of evidence for the presence of leprosy bacilli in red squirrels from North-West Europe. Transbound Emerg Dis. 2020 Mar;67(2):1032-1034.

6. Avanzi C, Lécorché E, Rakotomalala FA, Benjak A, Rapelanoro Rabenja F, Ramarozatovo LS, Cauchoix B, Rakoto-Andrianarivelo M, **Tió-Coma M**, Leal-Calvo T, Busso P, Boy-Röttger S, Chauffour A, Rasamoelina T, Andrianarison A, Sendrasoa F, Spencer JS, Singh P, Dashatwar DR, Narang R, Berland JL, Jarlier V, Salgado CG, Moraes MO, Geluk A, Randrianantoandro A, Cambau E, Cole ST. Population Genomics of *Mycobacterium leprae* Reveals a New Genotype in Madagascar and the Comoros. Front Microbiol. 2020 May 11;11:711.

7. **Tió-Coma M**, Avanzi C, Verhard EM, Pierneef L, van Hooij A, Benjak A, Roy JC, Khatun M, Alam K, Corstjens P, Cole ST, Richardus JH, Geluk A. Genomic Characterization of *Mycobacterium leprae* to Explore Transmission Patterns Identifies New Subtype in Bangladesh. Front Microbiol. 2020 Jun 16;11:1220.

8. van Hooij A*, **Tió-Coma M***, Verhard EM, Khatun M, Alam K, Tjon Kon Fat E, de Jong D, Sufian Chowdhury A, Corstjens P, Richardus JH, Geluk A. Household Contacts of Leprosy Patients in Endemic Areas Display a Specific Innate Immunity Profile. Front Immunol. 2020 Aug 11;11:1811.

9. **Tió-Coma M**, Kiełbasa SM, van den Eeden SJF, Mei H, Roy JC, Wallinga J, Khatun M, Soren S, Chowdhury AS, Alam K, van Hooij A, Richardus JH, Geluk A. Blood RNA signature RISK4LEP predicts leprosy years before clinical onset. EBioMedicine. 2021 Jun;68:103379.

*contributed equally