

**PONTIFÍCIA UNIVERSIDADE CATÓLICA DO PARANA
ESCOLA DE MEDICINA
PÓS-GRADUAÇÃO EM CIÊNCIAS DA SAÚDE**

PRISCILA VERCHAI UASKA SARTORI

**MAPEAMENTO FINO DAS SEQUÊNCIAS EXÔNICAS DE
GENES DE SUSCEPTIBILIDADE DO HOSPEDEIRO À
HANSENÍASE**

**CURITIBA
2019**

PRISCILA VERCHAI UASKA SARTORI

**MAPEAMENTO FINO DAS SEQUÊNCIAS EXÔNICAS DE
GENES DE SUSCEPTIBILIDADE DO HOSPEDEIRO À
HANSENÍASE**

Tese apresentada ao Programa de Pós-Graduação em Ciências da Saúde, Escola de Medicina, Pontifícia Universidade Católica do Paraná, como requisito parcial à obtenção do título de Doutora em Ciências da Saúde.

Orientador: Prof. Marcelo Távora Mira, PhD

Co-orientador: Vinícius Medeiros Fava, PhD

**CURITIBA
2019**



Pontifícia Universidade Católica do Paraná
Escola de Medicina
Programa de Pós-Graduação em Ciências da Saúde

ATA DA SESSÃO PÚBLICA DE EXAME DE TESE DO PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS DA SAÚDE EM NÍVEL DE DOUTORADO DA PONTIFÍCIA UNIVERSIDADE CATÓLICA DO PARANÁ.

Aos 24 dias do mês de julho de 2019 às 08:30, realizou-se a sessão pública de Defesa de Tese “MAPEAMENTO FINO DAS SEQUÊNCIAS EXÔNICAS DE GENES DE SUSCEPTIBILIDADE DO HOSPEDEIRO À HANSENÍASE” apresentado por **Priscila Verchai Uaska Sartori** para obtenção do título de Doutor; Área de concentração: Medicina e áreas afins.

A banca examinadora foi composta pelos seguintes membros:

MEMBROS DA BANCA	ASSINATURA
Prof. Dr. Marcelo Távora Mira – Presidente - (PUCPR)	
Profa. Dra. Cristina Pellegrino Baena – (PUCPR)	
Prof. Dr. Cleber Machado de Souza - (PUCPR)	
Prof. Dr. Gerson Oliveira Penna (UnB)	
Profa. Dra. Angelica Beate Winter Boldt (UFPR)	

De acordo com as normas regimentais a Banca Examinadora deliberou sobre os conceitos a serem distribuídos e que foram os seguintes:

Prof. Dr. Marcelo Távora Mira – Presidente

Conceito Aprovada

Profa. Dra. Cristina Pellegrino Baena – (PUCPR)

Conceito Aprovada

Prof. Dr. Cleber Machado de Souza - (PUCPR)

Conceito Aprovada

Prof. Dr. Gerson Oliveira Penna (UnB)

Conceito Aprovada

Profa. Dra. Angelica Beate Winter Boldt (UFPR)

Conceito Aprovada

Parecer Final: Aprovada

Observações da Banca Examinadora:

Prof. Dr. Marcelo Távora Mira
Presidente da Banca Examinadora

Prof.ª Dra. Cristina Pellegrino Baena
Coordenadora do Programa de Pós-Graduação em Ciências da Saúde
Profa. Dra. Cristina Pellegrino Baena
Coordenadora do PPGCS-PUCPR

AGRADECIMENTOS

- Á Deus, pelo dom da vida e do conhecimento;
- Aos meus pais **Irene Uaska Sartori** e **João Sartori Junior** e irmã **Pietra Maria Uaska Sartori**, pelo apoio incondicional em todas as fases de minha vida;
- Ao meu marido, **Diego Verchai Nichetti**, pelo companheirismo e compreensão;
- Ao meu orientador, **Marcelo Távora Mira**, pelo exemplo de ética e excelência em pesquisa;
- Ao meu co-orientador **Vinícius Medeiros Fava**, por sempre se prontificar em me auxiliar nos momentos de dificuldade;
- Aos colaboradores desse estudo;
- Aos meus amigos do Laboratório Experimental Multiusuários, por serem muito mais do que colegas de bancada;
- Aos participantes desse estudo;
- Á CAPES e Fundação Araucária, pelo auxílio financeiro;
- Ao *Leprosy Research Initiative*, pelo financiamento desse estudo.

Tenho a impressão de ter sido uma criança brincando à beira-mar, divertindo-me em descobrir uma pedrinha mais lisa ou uma concha mais bonita que as outras, enquanto o imenso oceano da verdade continua misterioso diante de meus olhos.

Isaac Newton

Comece fazendo o que é necessário, depois o que é possível, e de repente você estará fazendo o impossível.

São Francisco de Assis

RESUMO

Introdução: A hanseníase é uma doença infecciosa crônica causada pelo *Mycobacterium leprae*, um patógeno intracelular obrigatório e de crescimento lento que possui tropismo por macrófagos da pele e células de Schwann do sistema nervoso periférico. O desenvolvimento da hanseníase é altamente dependente de fatores genéticos do hospedeiro; estudos moleculares de ligação e associação têm levado, respectivamente, à identificação de *loci* e genes candidatos ao controle de fenótipos da doença. Entretanto, esses desenhos de estudo têm baixo poder de detecção de variantes funcionais.

Objetivo: Identificar variantes genéticas potencialmente funcionais, localizadas nas sequências transcritas, traduzidas e não-traduzidas (UTR) de 39 genes previamente associados com a susceptibilidade do hospedeiro à hanseníase.

Métodos: Foram recrutados 177 indivíduos residentes na Vila Santo Antônio do Prata, Pará, Brasil, distribuídos em 40 *pedigrees*, que tiveram seu DNA obtido a partir de sangue periférico por *salting-out*. As regiões exônicas dos genes selecionados foram sequenciadas na plataforma *Illumina MiSeq*. O alinhamento foi realizado com o algoritmo *BWA-SW (Smith-Waterman)*, a identificação das variantes foi feita utilizando a *pipeline* do *GATK* e utilizou-se os programas *GATK* e *vcftool* para a filtragem das variantes. Análise de associação foi realizada pelo Teste de Desequilíbrio de Transmissão no programa *FBAT*; o programa *Haploview* foi usado para estimar desequilíbrio de ligação (*LD*). A previsão *in silico* do impacto das variantes encontradas foi realizada usando os programas *Polyphen*, *SIFT*, *LRT*, *MutationTaster* e *CADD*.

Resultados: Foi encontrada associação entre hanseníase e 19 variantes distribuídas em 13 genes. Das 19 variantes, 16 eram independentes, seis causaram mudança de aminoácido na proteína e duas foram consideradas danosas pela análise *in silico*. Associação haplotípica foi detectada para os genes *NOD2* e *IL1LR1*; análise envolvendo variantes localizadas no mesmo cromossomo identificou haplótipos associados à hanseníase nos cromossomos 1 e 6. As variantes associadas neste estudo explicam parcialmente associações anteriormente detectadas nos genes *CCDC88B*, *LTA*, *IL1RL1*, *IL12RB2* e *NOD2*.

Conclusão: Nossa estratégia revelou variantes genéticas em regiões exônicas que validam a associação entre hanseníase e 13 dos 39 genes analisados. Estes sinais

potencialmente explicam associações previamente descritas em regiões intrônicas/intergênicas de 5 dos genes sequenciados. Finalmente, esse estudo identificou sete variantes com possível impacto funcional/causal, 5 delas são raras/baixa frequência.

Palavras-chaves: Hanseníase; Genes candidatos; Susceptibilidade genética do hospedeiro; Sequenciamento de segunda geração.

ABSTRACT

Introduction: Leprosy is a chronic infectious disease caused by *Mycobacterium leprae*, a slow growth and obligate intracellular pathogen with tropism for the skin and peripheral nervous system. Leprosy occurrence is highly dependent on host genetic risk factors, and several linkage and association studies have successfully identified candidate *loci* and genes involved in the control of leprosy phenotypes. However, none of these studies are designed to detect the true functional variants underlying the observed association. Aim: To advance on the identification of functional variants contributing to leprosy susceptibility through sequencing of the coding and UTR regions of 39 genes previously associated with susceptibility to leprosy. Methods: A total of 177 individuals resident at the Vila Santo Antônio do Prata, Pará, Brazil and distributed in 40 pedigrees, were enrolled and had their DNA obtained from peripheral blood by salting-out. Genes associated with leprosy in at least two independent population samples were selected for sequencing of their exonic regions using the Illumina MiSeq platform. Variant calling was performed using an adaptation of GATK pipeline; variants were filtered using GATK and vcftools. Association was tested using the Transmission Disequilibrium Test as implemented in FBAT; linkage disequilibrium (LD) was estimated using Haploview. Functional impact was predicted *in-silico* using Polyphen, CADD, LRT, MutationTaster and SIFT. Results: A total of 19 variants distributed in 13 out of 39 genes selected for the study were found associated with leprosy in the Prata sample. Linkage disequilibrium analysis suggested independence of 16 out of the 19 association signals detected. Six variants led to amino acid changes and two were considered damaging by *in-silico* analysis. Haplotypic association was detected for *NOD2* and *IL1LR1* genes; additional analysis involving variants located in the same chromosome identified haplotypes associated with leprosy on chromosomes 1 and 6. The variants associated in this study partially explain the previously intronic associations in *CCDC88B*, *LTA*, *IL1RL1*, *IL12RB2* and *NOD2* genes. Conclusion: Our study was able to identify, through direct sequencing, genetic variants of exonic sequences that validate the association between

leprosy and 13 of the 39 analyzed genes. The signals identified in the present study potentially explain associations previously found in intronic/intergenic regions in 5 of the sequenced genes. In addition, this study was able to detect seven variants with possible functional/causal impact, 5 of which are rare/low frequency.

Keywords: Leprosy; Candidate genes; Host genetic susceptibility; Second generation sequencing.

LISTA DE ABREVIATURAS E SIGLAS

A	Adenina
Aa	Aminiócido
a.C.	Antes de Cristo
A260/230	<i>Ratio of the absorbance at 260 and 230nm</i>
A260/280	<i>Ratio of the absorbance at 260 and 280nm</i>
ACB	African Caribbeans in Barbados
ADO	<i>2-aminoethanethiol dioxygenase</i>
anti-PGL-1	Antígeno glicolípido fenólico-1
ASW	<i>Americans of African Ancestry in SW USA</i>
BAT1	<i>DEAD-box helicase 39B</i>
BATF3	<i>Basic leucine zipper ATF-like transcription factor 3</i>
BB	<i>Borderline-borderline</i>
BBS9	<i>Bardet-Biedl syndrome 9</i>
BCL10	<i>B-cell CLL/lymphoma 10</i>
BEB	<i>Bengali from Bangladesh</i>
BL	<i>Borderline-lepromatoso</i>
BT	<i>Borderline-tuberculoide</i>
BTNL2	<i>Butyrophilin like 2</i>
C	Citosina
C1orf52	<i>Chromosome 1 open reading frame 52</i>
C1orf141	<i>Chromosome 1 open reading frame 141</i>
CARD	<i>Caspase recruitment domain</i>
CCDC122	<i>Coiled-coil domain containing 122</i>
CCDC88B	<i>Coiled-coil domain containing 88B</i>
CDH18	<i>Cadherin 18</i>
CDX	<i>Chinese Dai in Xishuangbanna, China</i>
CEP	Comitê de Ética em Pesquisa <i>Utah Residents (CEPH) with Northern and Western</i>
CEU	<i>European Ancestry</i>
CFH	<i>Complement factor H</i>
CHB	<i>Han Chinese in Beijing, China</i>
CHS	<i>Southern Han Chinese</i>
chr	Cromossomo
CLM	<i>Colombians from Medellin, Colombia</i>
CN	Controle negativo
CONEP	Comissão Nacional de Ética em Pesquisa
COX4I1	<i>Cytochrome c oxidase subunit 4I1</i>
CUBN	<i>Cubilin</i>
CTSB	<i>Cathepsin B</i>
CYLD	<i>CYLD lysine 63 deubiquitinase</i>
d.C.	Depois de Cristo

<i>DEC1</i>	<i>Deleted in esophageal cancer 1</i>
DNA	Deoxyribonucleic acid, Ácido desoxirribonucleico
dNTP	Desoxirribonucleotídeos fosfatados
DZ	Dizigóticos
EDTA	Ácido etilenodiamino tetra-acético
<i>EGR2</i>	<i>Early growth response 2</i>
<i>eQTL</i>	<i>Expression quantitative trait loci</i>
ESN	<i>Esan in Nigeria</i>
<i>FastQC</i>	<i>Quality control toll for high throughput sequence data</i>
<i>FCN2</i>	<i>Ficolin 2</i>
FIN	<i>Finnish in Finland</i>
G	Guanina
<i>GATA3</i>	<i>GATA binding protein 3</i>
GBR	<i>British in England and Scotland</i>
GIH	<i>Gujarati Indian from Houston, Texas</i>
GWAS	<i>Genome-wide association study</i>
GWD	<i>Gambian in Western Division, The Gambia</i>
HLA	Human leukocyte antigen, Antígeno leucocitário humano
<i>HLA</i>	<i>Major histocompatibility complex</i>
<i>HLA-C</i>	<i>Major histocompatibility complex, class I, C</i>
<i>HLA-DQA1</i>	<i>Major histocompatibility complex, class II, DQ alpha 1</i>
<i>HLA-DQB1</i>	<i>Major histocompatibility complex, class II, DQ beta 1</i>
<i>HLA-DQB2</i>	<i>Major histocompatibility complex, class II, DQ beta 2</i>
<i>HLA-DQB3</i>	<i>Major histocompatibility complex, class II, DQ beta 3</i>
<i>HLA-DQB4</i>	<i>Major histocompatibility complex, class II, DQ beta 4</i>
<i>HLA-DQB5</i>	<i>Major histocompatibility complex, class II, DQ beta 5</i>
<i>HLA-DRB1/DQA1</i>	<i>Major histocompatibility complex, class II, DR beta 1/ class II, DQ alpha 1</i>
I	Indeterminada
<i>IFNG</i>	<i>Interferon gamma</i>
IFN- γ	Interferon-gama
<i>IL10</i>	<i>Interleukin 10</i>
IL-10	Interleucina 10
IL-12	Interleucina 12
<i>IL12B</i>	<i>Interleukin 12B</i>
IL-15	Interleucina 15
<i>IL17F</i>	<i>Interleukin 17F</i>
IL-18	Interleucina 18
<i>IL18R1</i>	<i>Interleukin 18 receptor 1</i>
<i>IL18RAP</i>	<i>Interleukin 18 receptor accessory protein</i>
IL-2	Interleucina 2
<i>IL23R</i>	<i>Interleukin 23 receptor</i>
IL-4	Interleucina 4
IL-5	Interleucina 5

IL-7	Interleucina 7
ITU	<i>Indian Telugu from the UK</i>
JPT	<i>Japanese in Tokyo, Japan</i>
KDa	Quilo daltons
KHV	<i>Kinh in Ho Chi Minh City, Vietnam</i>
Km	Quilômetro
LACC1	<i>Laccase (multicopper oxidoreductase) domain containing 1</i>
LD	<i>Linkage disequilibrium</i>
LL	Lepromatoso
LRI	<i>Leprosy research initiative</i>
LRRK2	<i>Leucine-rich repeat kinase 2</i>
LTA	<i>Lymphotoxin alpha</i>
LTA4H	<i>Leukotriene A4 hydrolase</i>
MAF	<i>Minor allele frequency</i>
MB	Multibacilar
MBL2	<i>Mannose binding lectin 2</i>
MED30	<i>Mediator complex subunit 30</i>
mg	Miligramas
MHC	<i>Major histocompatibility complex</i>
MICA	<i>MHC class I polypeptide-related sequence A</i>
MICB	<i>MHC class I polypeptide-related sequence B</i>
mL	Mililitros
MRC1	<i>Mannose receptor, C type 1</i>
MSL	<i>Mende in Sierra Leone</i>
MXL	<i>Mexican Ancestry from Los Angeles USA</i>
MZ	Monozigóticos
NEBL	<i>Nebulette</i>
NF-Kb	Fator nuclear kappa B
NFKBIL1	<i>NFKB inhibitor like 1</i>
ng	Nanogramas
NOD2	<i>Nucleotide-binding oligomerization domain containing 2</i>
NRAMP1	<i>Solute carrier family 11 member 1</i>
OMS	Organização Mundial da Saúde
OR	Odds ratio
PACRG	<i>PARK2 co-regulated gene</i>
PRKN	<i>Parkin RBR E3 ubiquitin protein ligase</i>
PB	Paucibacilar
pb	Pares de base
PCR	<i>Polymerase chain reaction</i>
PCR-RT	<i>Real-time Polymerase Chain Reaction</i>
PEL	<i>Peruvians from Lima, Peru</i>
pH	Potencial hidrogeniônico
PPGCS	Programa de Pós-Graduação em Ciências da Saúde
PQT	Poliquimioterapia

PUCPR	Pontifícia Universidade Católica do Paraná
PUR	<i>Puerto Ricans from Puerto Rico</i>
q.s.p	"Quantidade suficiente para"
<i>RAB32</i>	<i>RAB32, member RAS oncogene family</i>
<i>RIPK2</i>	<i>Receptor-interacting serinethreonine kinase 2</i>
<i>RMI2</i>	<i>Recq mediated genome instability 2</i>
<i>RPS6KA4</i>	<i>Ribosomal protein S6 kinase A4</i>
RT1	Reação do Tipo 1
RT2	Reação do Tipo 2
<i>SLC11A1</i>	<i>Solute carrier family 11 member 1</i>
SNP	<i>Single nucleotide polymorphism</i>
<i>SOD2</i>	<i>Superoxide dismutase 2, mitochondrial</i>
STU	<i>Sri Lankan Tamil from the UK</i>
<i>SYN2</i>	<i>synapsin II</i>
T	Timina
<i>Tag SNPs</i>	<i>Tagging single nucleotide polymorphism</i>
TCLE	Termo de Consentimento Livre e Esclarecido
TE	Tris-EDTA buffer
Th1	T helper1
Th2	T helper 2
<i>TLR1</i>	<i>Toll-like receptor 1</i>
<i>TLR-4</i>	<i>Toll-like receptor 4</i>
<i>TNFA</i>	<i>Tumor necrosis fator</i>
<i>TNFSF8</i>	<i>TNF superfamily member 8</i>
<i>TNFSF15</i>	<i>Tumor necrosis factor (ligand) superfamily, member 15</i>
TT	Tuberculoide
UCSC	<i>University of California Santa Cruz</i>
<i>UTR</i>	<i>Untranslated region</i>
v	Versão
<i>VDR</i>	<i>Vitamin D (1,25- dihydroxyvitamin D3) receptor</i>
YRI	<i>Yoruba in Ibadan, Nigeria</i>

LISTA DE ILUSTRAÇÕES

Figura 1. Crânio humano com lesões características da hanseníase.	1
Figura 2. Gerhard Henrick Armauer Hansen	2
Figura 3. Distribuição geográfica de novos casos (incidência) de hanseníase em 2017.	4
Figura 4. Microscopia de <i>Mycobacterium leprae</i>	5
Figura 5. Representação esquemática do espectro de classificação clínica da hanseníase.	8
Figura 6. Esquema recomendado pela OMS para o tratamento com a Poliquimioterapia.	9
Figura 7. Distribuição das cepas de <i>M. leprae</i> pelo mundo, de acordo com SNPs e as rotas migratórias humanas.	12
Figura 8. Genes associados no primeiro GWAS com hanseníase e suas validações.	22
Figura 9. <i>Illumina flow-cell</i>	29
Figura 10. Figura esquemática detalhando o preparo de template na plataforma Illumina.....	30
Figura 11. <i>Cluster</i> formado durante o preparo do <i>template</i> no equipamento <i>Illumina MySeq</i>	30
Figura 12. Figura esquemática do processo de sequenciamento em <i>Illumina</i>	32
Figura 13. Fluxograma das etapas de análise dos dados.....	41
Figura 14. Histograma do conteúdo GC por <i>read</i> obtido para todas as amostras em conjunto.	42
Figura 15. Scores de qualidade das bases obtidos para todas as amostras em conjunto, incluindo as reads dos dois extremos (r1 e r2).....	43
Figura 16. Distribuição de qualidade das <i>reads</i>	44
Figura 17. Média de cobertura da amostra populacional de alguns dos genes sequenciado.....	46
Figura 18. Bloco de desequilíbrio de ligação entre as variantes não sinônimas consideradas na análise gene-wide do gene <i>IL1RL1</i>	57
Figura 19. Bloco de desequilíbrio de ligação entre os marcadores associados dos genes <i>CUBN</i> e <i>ILRL1</i> , respectivamente.	59
Figura 20. Bloco de desequilíbrio de ligação entre os marcadores associados do gene <i>NOD2</i>	60
Figura 21. Bloco de desequilíbrio de ligação entre os marcadores associados nesse estudo e em estudo anterior no gene <i>LTA</i>	61
Figura 22. Bloco de desequilíbrio de ligação entre os marcadores associados nesse estudo e em estudo anterior nos genes <i>PRKN</i> e <i>PACRG</i>	62
Figura 23. Bloco de desequilíbrio de ligação entre os marcadores associados nesse estudo e em estudo anterior no gene <i>NOD2</i>	63

Figura 24. Bins das variantes já associadas com a susceptibilidade do hospedeiro à hanseníase no gene <i>LTA</i> utilizando dados do 1000G da população de YRI.	73
Figura 25. Bins das variantes já associadas com a susceptibilidade do hospedeiro à hanseníase no gene <i>NOD2</i> utilizando dados do 1000G da população de YRI.	74

LISTA DE TABELAS

Tabela 1: Caracterização da amostra populacional	36
Tabela 2. Dados de alinhamento do sequenciamento.	44
Tabela 3. Porcentagem de bases sequenciadas (em exons) com uma média de cobertura superior a 30 vezes, por gene.	45
Tabela 4. Número de variantes identificadas em cada um dos genes sequenciados	49
Tabela 5. Número de variantes identificadas com $MAF > 0,04$ em cada um dos genes sequenciados.	50
Tabela 6. SNPs associados no TDT	53
Tabela 7. Impacto das variantes não sinônimas encontradas associadas com a hanseníase <i>per se</i> no presente estudo.	54
Tabela 8. Número de variantes raras/baixa frequência identificadas em cada um dos genes sequenciados.....	55
Tabela 9. Resultado da análise <i>Gene-wide</i> para variantes raras e de baixa frequência	56
Tabela 10. Resultado da análise <i>Gene-wide</i> para variantes raras e comuns em conjunto	57
Tabela 11. Resultado da análise de regressão logística multivariada no SAS .	60
Tabela 12. Análise de <i>LD</i> entre variantes encontradas nesse estudo e em estudos anteriormente publicados.....	64
Tabela 13. Haplótipo das variantes associadas com a susceptibilidade do hospedeiro à hanseníase no gene <i>CUBN</i>	74
Tabela 14. Haplótipo das variantes associadas com a susceptibilidade do hospedeiro à hanseníase no gene <i>IL1RL1</i>	75
Tabela 15. Haplótipo das variantes associadas com a susceptibilidade do hospedeiro à hanseníase no gene <i>NOD2</i>	75
Tabela 16. Haplótipo das variantes associadas com a susceptibilidade do hospedeiro à hanseníase no cromossomo 1.	75
Tabela 17. Haplótipo das variantes associadas com a susceptibilidade do hospedeiro à hanseníase no cromossomo 6.	76
Tabela 18. Haplótipo das variantes associadas com a susceptibilidade do hospedeiro à hanseníase no cromossomo 10.	76

Sumário

AGRADECIMENTOS	IV
RESUMO	VI
ABSTRACT.....	VIII
LISTA DE ILUSTRAÇÕES	XIV
LISTA DE TABELAS.....	XVI
1. INTRODUÇÃO	1
1.1 ASPECTOS GERAIS E HISTÓRICOS DA HANSENÍASE	1
1.2 EPIDEMIOLOGIA.....	3
1.3 AGENTE ETIOLÓGICO	4
1.4 TRANSMISSÃO	5
1.5 DIAGNÓSTICO DA HANSENÍASE	6
1.6 CLASSIFICAÇÃO DA HANSENÍASE	6
1.7 TRATAMENTO DA HANSENÍASE.....	8
1.8 ESTADOS REACIONAIS DA HANSENÍASE	10
1.8.1 Tratamento dos estados reacionais.....	11
1.9 GENÉTICA DO <i>M.leprae</i>	11
1.10 SUSCEPTIBILIDADE GENÉTICA DO HOSPEDEIRO À HANSENÍASE	
13	
1.10.1 A região 6q25-q27 e os genes <i>PRKN</i> e <i>PACRG</i>	15
1.10.2 A região 10p13	17
1.10.3 Os genes associados a partir de <i>GWAS</i>	18
1.10.4 Os genes dos complexos <i>TLR</i> e <i>HLA</i>	22
1.10.5 Outros genes não- <i>HLA</i>	25
1.11 SEQUENCIAMENTO MULTIPARALELO.....	27
1.11.1 <i>MiSeq System – Illumina</i>	28
3. OBJETIVOS	34
3.1 OBJETIVO GERAL	34

3.2	OBJETIVOS ESPECÍFICOS	34
4.	MATERIAIS E MÉTODOS.....	35
4.1	ETAPAS PRÉ-SEQUENCIAMENTO.....	35
4.1.1	Considerações éticas	35
4.1.2	Suporte Financeiro	35
4.1.3	Amostra Populacional.....	36
4.1.4	Coleta do material biológico e extração de DNA.....	37
4.1.5	Seleção dos genes candidatos	37
4.2	SEQUENCIAMENTO	38
4.3	ANÁLISE DOS DADOS.....	38
4.3.1	Análise do sequenciamento.....	38
4.3.2	Análise de Associação.....	40
5.	RESULTADOS	42
5.1	ANÁLISE GERAL DO SEQUENCIAMENTO	42
5.2	MÉDIA DE COBERTURA DOS EXONS	45
5.3	IDENTIFICAÇÃO DE VARIANTES	49
5.4.	ANÁLISE DE ASSOCIAÇÃO	52
5.4.1	De variantes individualmente.....	52
5.4.2.	Variantes raras	54
5.4.3	Análise <i>gene-wide</i>	56
5.4.3.1	Análise <i>gene-wide</i> de variantes raras e de baixa frequência	56
5.4.3.2	Análise <i>gene-wide</i> concomitantemente de variantes raras/ baixa frequência e comuns.....	57
5.5	ANÁLISE DE DESEQUILÍBRIO DE LIGAÇÃO (<i>LD</i>).....	58
5.5.1	Entre marcadores associados nesse estudo	58
5.5.2	Entre marcadores deste e de outros estudos realizados no Prata 60	
5.5.3	Entre marcadores deste e de outros estudos realizados em amostras populacionais distintas.....	63
5.5.4	Entre todos os marcadores já associados em população de YRI da base de dados <i>1000G</i>	72
5.5.4.1	Gene <i>LTA</i>	73
5.5.4.2	Gene <i>NOD2</i>	73

5.6	ANÁLISE DE HAPLÓTIPOS	74
5.6.1	Por gene.....	74
5.6.2	Por cromossomo	75
6.	DISCUSSÃO	77
6.1	VARIANTES ASSOCIADAS - SINÔNIMAS OU EM REGIÃO <i>UTR</i>	79
6.2	ANÁLISE DOS HAPLÓTIPOS.....	84
6.3	ANÁLISE <i>GENE-WIDE</i>	85
7.	CONSIDERAÇÕES FINAIS	87
8	REFERÊNCIAS BIBLIOGRÁFICAS.....	90
	ANEXO I: TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO	110
	ANEXO II: APROVAÇÃO DO CONEP E DO CEP (2006 E 2007)	114
	ANEXO III: APROVAÇÃO PELO CEP (2014)	117
	ANEXO IV: TERMO DE ACEITAÇÃO DE APOIO FINANCEIRO	121
	ANEXO V: ARTIGO	123
	Human genetic susceptibility of leprosy recurrence	123
	ANEXO VI: ARTIGO	129
	The Prata Village: An Estimate of the Incubation Period and the Proportion of Naturally Susceptible Individuals to Leprosy in Humans	129

1. INTRODUÇÃO

1.1 ASPECTOS GERAIS E HISTÓRICOS DA HANSENÍASE

A hanseníase, outrora conhecida como “lepra”, é uma doença infecciosa crônica causada pelo *Mycobacterium leprae*, um bacilo álcool-ácido resistente, exclusivamente intracelular e de crescimento lento, que possui tropismo por macrófagos da pele e células de Schwann do Sistema Nervoso Periférico (1).

Uma das mais antigas evidências da doença trata-se de uma ossada datada de 2.000 a.C encontrada na Índia, que possui características do desenvolvimento da doença, como por exemplo lesões erosivas na região supraorbital e da glabella, necrose bilateral da maxila e remodelamento da abertura nasal, como mostrado na **Figura 1** (2).

Figura 1. Crânio humano com lesões características da hanseníase.

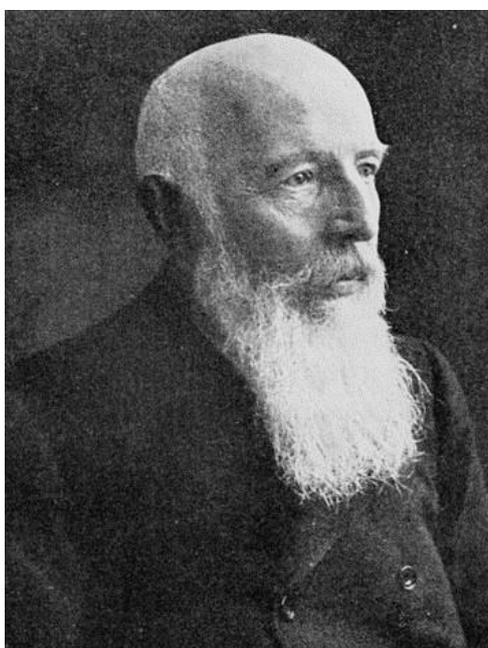


Fonte: Robbins, 2009 (2).

Na Bíblia, há cerca de 50 citações sobre a “lepra”, sendo 22 delas no capítulo 13 do livro do Levíticos, no qual é relatada a forma com a qual os sacerdotes diagnosticavam a doença (3). Entretanto, a “lepra” bíblica é diferente da doença que conhecemos hoje, já que não há qualquer menção sobre as deformidades, perdas sensoriais, cegueira ou paralisia, além das lesões poderem mostrar alterações em apenas sete dias, o que não ocorre com a hanseníase. Assim a doença foi confundida com outras enfermidades como vitiligo, porfirias, psoríase ou infecções fúngicas. Na época, os indivíduos acometidos pela doença eram considerados impuros e viviam marginalizados (4).

Apesar de a hanseníase ser uma doença milenar, a sua natureza infecciosa foi explorada apenas a partir de 1873 – até então, a doença era considerada hereditária – quando o médico norueguês Gerhard Henrick Armauer Hansen (**Figura 2**) analisou ao microscópio material de lesões de pacientes com hanseníase e observou a presença de estruturas com formato de bastonetes (5). A caracterização detalhada do bacilo foi realizada apenas em 1879 por Albert Ludwig Neisser, após utilizar técnicas de coloração por ele aprimoradas (6). O achado contribuiu, em longo prazo, para a redução do estigma religioso em torno dessa enfermidade e inaugurou uma nova era de estudos sobre a hanseníase e outras doenças infecciosas.

Figura 2. Gerhard Henrick Armauer Hansen



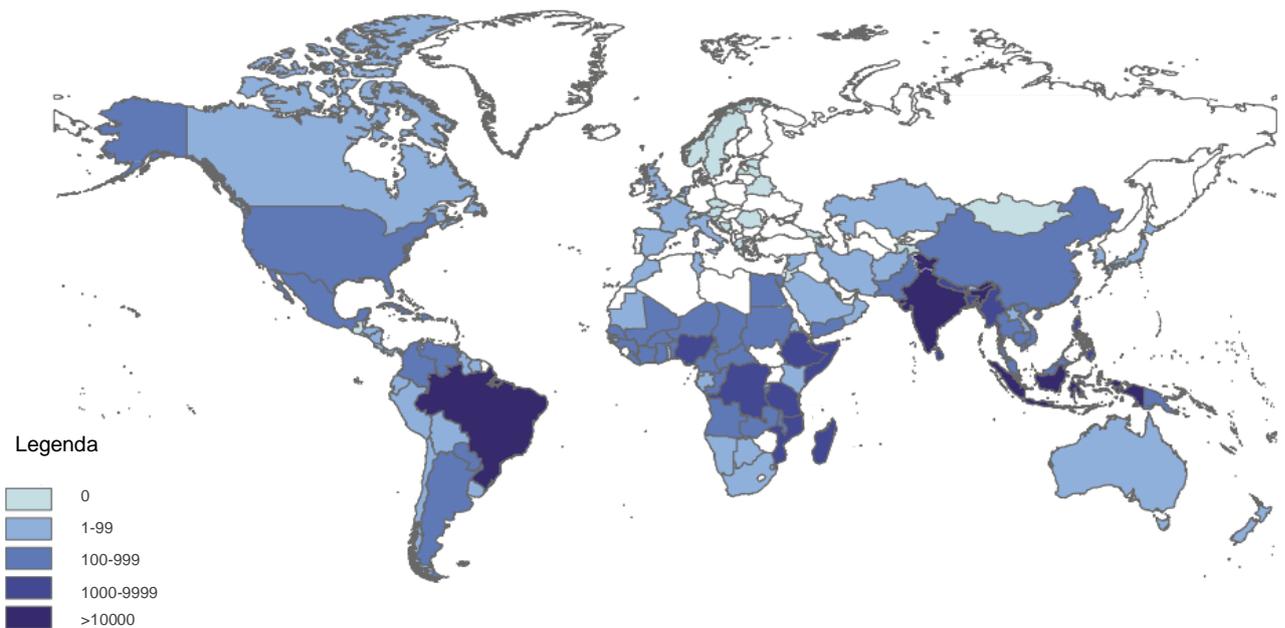
Fonte: Marmor, 2002 (7).

1.2 EPIDEMIOLOGIA

A prevalência da hanseníase em meados dos anos 80 era de 5.351.408 casos da doença no mundo, ou seja, cerca de 12 casos por 10.000 habitantes (8). O desenvolvimento de um esquema terapêutico poliquimioterápico de distribuição gratuita levou a Organização Mundial da Saúde (OMS) propor, em 1991, a eliminação da hanseníase em todo o mundo até o ano 2000, definida como a redução da sua prevalência para menos de um caso a cada 10.000 habitantes (9). Atualmente, a prevalência mundial da doença é de 192.713 casos, ou seja, 0,25 por 10.000 indivíduos. Contudo, apesar da redução da prevalência mundial, alguns países, incluindo o Brasil, ainda apresentam uma prevalência alarmante. Em 2017, foi registrada uma prevalência de 28.067 casos (1,33/10.000 habitantes) no Brasil e 90.709 (0,67/10.000 habitantes) na Índia (10). Neste mesmo ano, os estados brasileiros que apresentam a maior taxa de prevalência da hanseníase foram Mato Grosso, Tocantins, Maranhão, Rondônia e Pará (11). Dados do Sistema de Informação de Agravos de Notificação mostram que em 2017 foi registrada no Pará, estado do qual a amostra populacional deste estudo é proveniente, uma prevalência de 3,20 casos para cada 10.000 habitantes (11).

Segundo o último boletim informativo da OMS (10), no ano de 2017 foram reportados 210.671 novos casos (incidência) de hanseníase, 29.101 nas Américas, sendo que destes, 92,35% (26.875 casos) foram notificados no Brasil. Dentre os novos casos observados no país, 1.718 (6,39%) ocorreram em crianças com menos de 15 anos de idade, o que indica transmissão ativa da doença. Atualmente, o Brasil ocupa o segundo lugar entre os países com maior incidência de hanseníase, ficando atrás somente da Índia, que notificou 126.164 novos casos (**Figura 3**). Entretanto, se compararmos a taxa de incidência (novos casos/100.000 habitantes) o Brasil apresenta uma taxa mais elevada, de 12,74, enquanto a Índia apresenta uma taxa menor, de 9,31.

Figura 3. Distribuição geográfica de novos casos (incidência) de hanseníase em 2017.



Legenda: Os países em branco não reportaram casos, já os países em diferentes tons de azul reportaram 0, 1-99, 100-999, 1000-9999 e mais de 10000 casos de acordo com a tonalidade crescente (do mais claro ao mais escuro) de azul.

Fonte: Adaptado de: Organização Mundial da Saúde,(10).

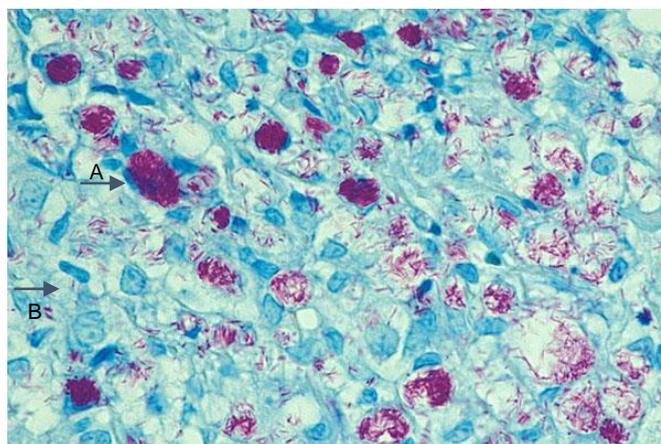
1.3 AGENTE ETIOLÓGICO

O *M. leprae* é um bacilo que pertence ao gênero *Mycobacterium*, da família *Mycobacteriaceae* e da ordem *Actinomycetales*. É uma bactéria gram positiva, álcool-ácido resistente que costuma ser corada pelos métodos de Ziehl-Neelsen (**Figura 4**) ou Fite-Faraco, podendo ser observadas isoladas ou em globias (visualmente são aglomerados de bacilos). A bactéria apresenta um formato bacilar reto ou ligeiramente curvado com um comprimento em torno de 1 – 8 μm e 0,3 – 0,6 μm de diâmetro. O tempo de duplicação do *M. leprae* é longo, de 12-14 dias, e sua temperatura ideal de crescimento é de 32°C; devido a isso, o bacilo possui tropismo pelas regiões mais externas (frias) do corpo humano, como a pele e os nervos periféricos (12). Além disso, a doença tem um longo período de incubação (+/- 15 anos), o que pode exercer grande impacto nos programas de saúde pública que visam a eliminação da doença (manuscrito em anexo – ANEXO V).

Além de ser um patógeno intracelular obrigatório, o *M. leprae* não pode ser cultivado *in vitro*, sendo possível apenas, mantê-lo viável em meios líquidos ou em cultura de macrófagos, por um curto período de tempo (12). Atualmente, o método mais utilizado para cultivar o *M. leprae* é *in vivo*, nos coxins plantares de camundongos atímicos (*nude*), técnica que foi desenvolvida em 1960 por Shepard (13). Alternativamente, é possível cultivar o bacilo em tatu de nove bandas (*Dasypus novemcinctus*), espécie que pode desenvolver a doença (14).

No ano de 2008, com o avanço de técnicas moleculares, foi descoberta uma nova espécie o *M. lepromatosis* (15). Essa foi apontada como agente etiológico específico da hanseníase lepromatosa difusa, ainda que esteja presente concomitantemente com o *M. leprae* em alguns casos de hanseníase multibacilar (16).

Figura 4. Microscopia de *Mycobacterium leprae*.



Legenda: Em A, *M. leprae* em globia; B, *M. leprae* isolado.

Fonte: Adaptado de: Pathology of Infectious Diseases. Lepromatous leprosy, (17).

1.4 TRANSMISSÃO

Que se tenha conhecimento, além do tatu nove-bandas (*Dasypus novemcinctus*) (18), os organismos capazes de contrair e desenvolver hanseníase são: o esquilo-vermelho (*Sciurus vulgaris*) (19), alguns primatas (20) e o ser humano, sendo este aceito como o principal reservatório natural conhecido do *M.*

leprae. O mecanismo exato de transmissão da hanseníase não se encontra completamente elucidado. A teoria mais aceita é de que a transmissão ocorra através das vias aéreas superiores de indivíduos afetados pela forma multibacilar da doença e não tratados, para indivíduos saudáveis susceptíveis, após um período prolongado de exposição (21). Estudos que demonstram a presença do *M. leprae* na mucosa nasal de pacientes hansenianos, de contatos ocupacionais e de controles de regiões endêmicas, corroboram com essa hipótese (22-24).

Outras vias de transmissão já foram sugeridas, dentre elas o contato com tatu nove-bandas infectados (18) e a infecção via insetos vetores (25), o que atribuiria à hanseníase uma característica de zoonose; além disso, há estudos envolvendo água (26) e solo contendo o *M. leprae* viável (27), porém, nenhuma dessas teorias foi comprovada até o momento.

1.5 DIAGNÓSTICO DA HANSENÍASE

O diagnóstico da hanseníase é essencialmente clínico. As principais características do desenvolvimento da doença são: (i) lesões de pele com perda de sensibilidade; (ii) espessamento de nervos periféricos, podendo levar à dor neuropática; (iii) baciloscopia positiva de lesão, sendo que, quando negativa, não indica a ausência da doença; (iv) detecção sérica de anticorpos anti-PGL-1 (antígeno glicolípido fenólico-1), entretanto sua ausência não significa a inexistência da doença. Além disso, contatos de pacientes podem possuir esses anticorpos mesmo não tendo desenvolvido a doença; e (v) PCR positivo para DNA de *M. leprae* em amostra de biópsia (28).

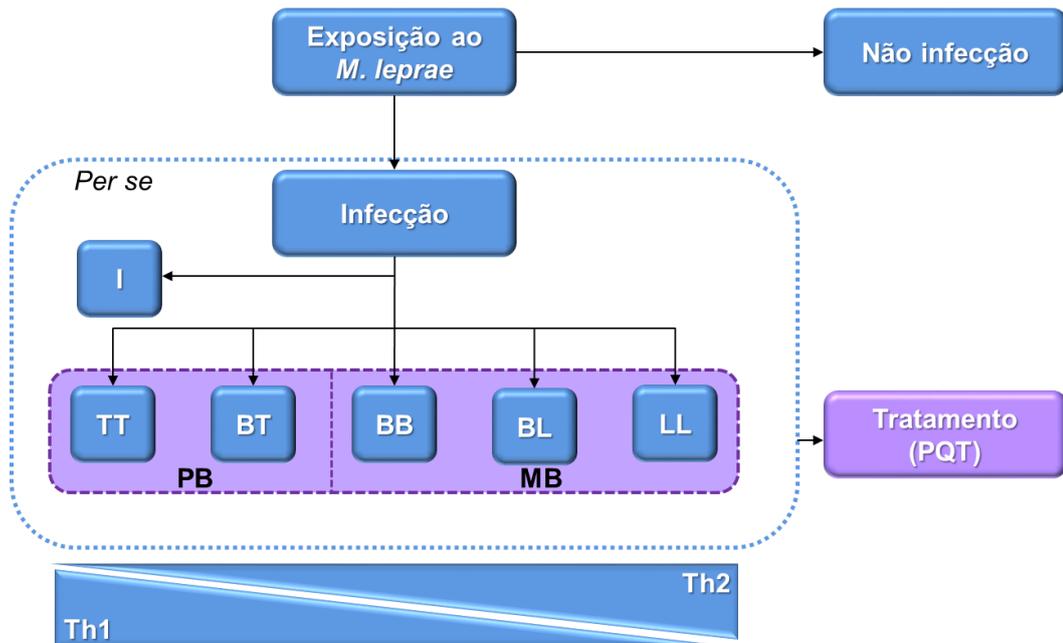
1.6 CLASSIFICAÇÃO DA HANSENÍASE

A hanseníase é uma doença espectral que possui formas clínicas distintas. Um sistema de classificação da hanseníase, postulado por Ridley e Jopling em 1966, adota cinco formas clínicas da doença, sendo duas polares e três

intermediárias. Um dos pólos, chamado de tuberculoide (TT) caracteriza-se por lesões bem definidas, ausência de bacilos na pele e nervos e resposta imunológica predominantemente Th1 (celular), com a predominância de citocinas pró-inflamatórias IFN- γ , IL-2, IL-7, IL-12, IL-15 e IL-18 (1). O pólo oposto, lepromatoso (LL), caracteriza-se por múltiplas lesões, presença de bacilos na pele e nervos e resposta imunológica predominantemente Th2 (humoral), caracterizada pelas citocinas IL-4, IL-5 e IL-10 (1). Entre estes pólos, há as formas intermediárias da doença, denominadas *borderline*, que se dividem em *borderline-tuberculoide* (BT), *borderline-borderline* (BB) e *borderline-lepromatosa* (BL), sendo que do polo TT para o LL há um aumento progressivo do número de bacilos em pele e nervos periféricos, no número de lesões na pele e de anticorpos específicos para o *M. leprae*. Quando a enfermidade não pode ser classificada em um desses cinco subgrupos, é chamada de hanseníase indeterminada (I) (29). A doença, independente de sua forma clínica, é chamada de hanseníase *per se* (**Figura 5**).

Para fins operacionais e terapêuticos, a OMS classifica a hanseníase em paucibacilar (PB) ou multibacilar (MB). A hanseníase PB é caracterizada pela presença de uma a cinco lesões na pele e ausência de bacilos na baciloscopia, enquanto que a forma MB é caracterizada por mais de cinco lesões na pele e/ou baciloscopia positiva (30). As formas clínicas TT e BT de Ridley e Jopling, bem como a Indeterminada geralmente possuem características da hanseníase PB, enquanto as demais contêm aspectos da forma clínica MB.

Figura 5. Representação esquemática do espectro de classificação clínica da hanseníase.



Legenda: O esquema apresenta a hanseníase *per se* e suas formas clínicas. TT-tuberculoide, BT-borderline-tuberculoide, BB-borderline-borderline, BL-borderline-lepromatosa, LL-lepromatosa, I-indeterminada, PB-paucibacilar, MB-multibacilar, Th1:T-helper 1, Th2:T-helper 2; PQT: Poliquimioterapia.

Fonte: Adaptado de Sauer, *et al.* 2016 (31).

1.7 TRATAMENTO DA HANSENÍASE

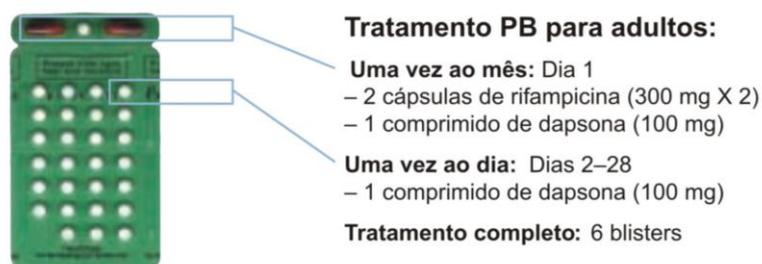
A primeira terapêutica que se tem relato do uso na hanseníase foi o óleo de Chaulmoogra, de eficácia questionável (32). Na década de 40, a antibioticoterapia com derivados sulfônicos começou a ser testada na hanseníase, sendo o promim (glicosulfona sódica) a primeira droga utilizada com sucesso. Por ser mais eficiente, barata e de fácil administração, a dapsona substituiu o promim e passou a ser amplamente empregada como monoterapia. Porém, em 1953 foi registrado o primeiro caso de resistência do *M. leprae* ao fármaco. Em 1970, a rifampicina, droga bactericida, começou a ser utilizada também como monoterapia no tratamento da hanseníase; no entanto seis anos mais tarde foi relatado o primeiro caso de resistência à rifampicina (33-35).

Principalmente em virtude dos casos de resistência às monoterapias dapsônica e rifampicínica, a OMS recomenda, desde 1981, a administração da

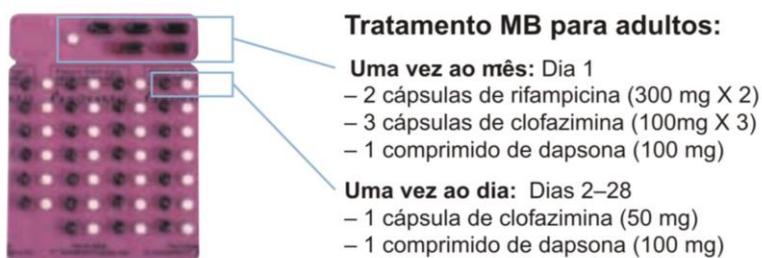
poliquimioterapia (PQT), que consiste na associação de rifampicina, dapsona e clofazimina, para o tratamento efetivo da hanseníase. Atualmente o tratamento deve ser realizado por seis meses (com rifampicina e dapsona) pelos portadores da forma PB da doença, e pelo período de um ano (com todos os fármacos da PQT) para os pacientes com hanseníase MB. Essas doses são ajustadas para uso pediátrico (**Figura 6**) (36). Ainda assim, após a introdução do tratamento poliquimioterápico, casos de *M. leprae* multirresistente às drogas da PQT começaram a ser registrados, e casos de resistência a componentes da PQT passaram a ser tratados com drogas alternativas, como a ofloxacina, miniciclina e claritromicina (37).

Nos últimos anos, têm sido proposto um tratamento unificado (U-MDT), para MB e PB, a ser realizado por um período de seis meses com todos os fármacos da PQT, nas mesmas dosagens que atualmente são utilizadas para tratar pacientes multibacilares (38, 39). Um estudo clínico randomizado mostrou uma maior redução no índice baciloscópico de pacientes tratados pela PQT convencional do que pela U-MDT, porém essa diferença não foi significativa (40).

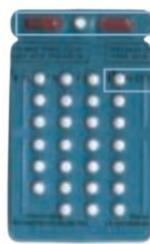
Figura 6. Esquema recomendado pela OMS para o tratamento com a Poliquimioterapia.



Blister para PB adultos



Blister para MB adultos



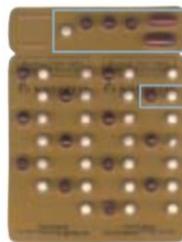
Blisters para PB crianças

Tratamento PB para crianças (10 a 14 anos)

- Uma vez ao mês:** Dia 1
– 2 cápsulas de rifampicina (300 mg+150 mg)
– 1 comprimido de dapsona (50 mg)
- Uma vez ao dia:** Dias 2–28
– 1 comprimido de dapsona (50 mg)

Tratamento completo: 6 blisters

Para crianças com menos de 10 anos as doses devem ser ajustadas de acordo com o peso corpóreo.



Blisters para MB crianças

Tratamento MB para crianças (10 a 14 anos)

- Uma vez ao mês:** Dia 1
– 2 cápsulas de rifampicina (300 mg+150 mg)
– 3 cápsulas de clofazimina (50 mg X 3)
– 1 comprimido de dapsona (50 mg)
- Uma vez ao dia:** Dias 2–28
– 1 cápsula de clofazimina dias intercalados (50 mg)
– 1 comprimido de dapsona (50 mg)

Tratamento completo: 12 blisters

Para crianças com menos de 10 anos as doses devem ser ajustadas de acordo com o peso corpóreo.

Legenda: PB: paucibacilar; MB: multibacilar; mg: miligramas

Fonte: Adaptado de Organização Mundial da Saúde, (41).

1.8 ESTADOS REACIONAIS DA HANSENÍASE

Os estados reacionais são processos inflamatórios agudos, repentinos e intensos que acometem de 30 a 50% dos indivíduos portadores de hanseníase em todos os estágios da doença, incluindo o momento do diagnóstico, durante o tratamento da hanseníase com a PQT e até mesmo após a cura terapêutica (42). Os estados reacionais são caracterizados por uma súbita e intensa ativação/reativação do sistema imune do hospedeiro, que frequentemente afeta os nervos periféricos, sendo hoje a principal causa de deformidades físicas e de deficiências sensoriais e motoras permanentes que atingem os pacientes hansenianos (42). Os dois principais tipos de estados reacionais são: a Reação Reversa, também chamada de Reação do Tipo 1 (RT1) e a Reação do Tipo 2 (RT2). A RT1 ocorre majoritariamente em pessoas acometidas pelas formas *borderline* (BT, BB e BL) da doença, e é caracterizada por uma resposta da imunidade celular tardia e abrupta (43). Os fatores de risco conhecidos para o desenvolvimento da RT1 são: a idade do indivíduo (no Brasil, pacientes com idade superior a 40 anos têm maior risco), o

índice de baciloscopia maior que quatro e mais que 5 lesões no momento do diagnóstico (42, 44). Já a RT2 ocorre em principalmente em pacientes acometidos pela hanseníase BL ou LL e relaciona-se com a deposição de imunocomplexos, semelhante à reação de hipersensibilidade do tipo III de Coombs e Gel (43). Pacientes que apresentam um índice baciloscópico maior que quatro são considerados de risco para desenvolver este tipo de reação (42).

1.8.1 Tratamento dos estados reacionais

Os pacientes acometidos pelos estados reacionais da hanseníase devem ser tratados o mais breve possível, a fim de evitar incapacidades permanentes. O tratamento recomendado para aqueles que desenvolveram RT1 é uma dose diária de prednisolona de 30-40 mg. Já para os pacientes com RT2, a dose diária de prednisolona é de 40-60 mg, podendo ser associada com 400 mg diários de talidomida. A talidomida deve ser usada com cautela em mulheres em idade fértil, devido ao seu potencial teratogênico (45).

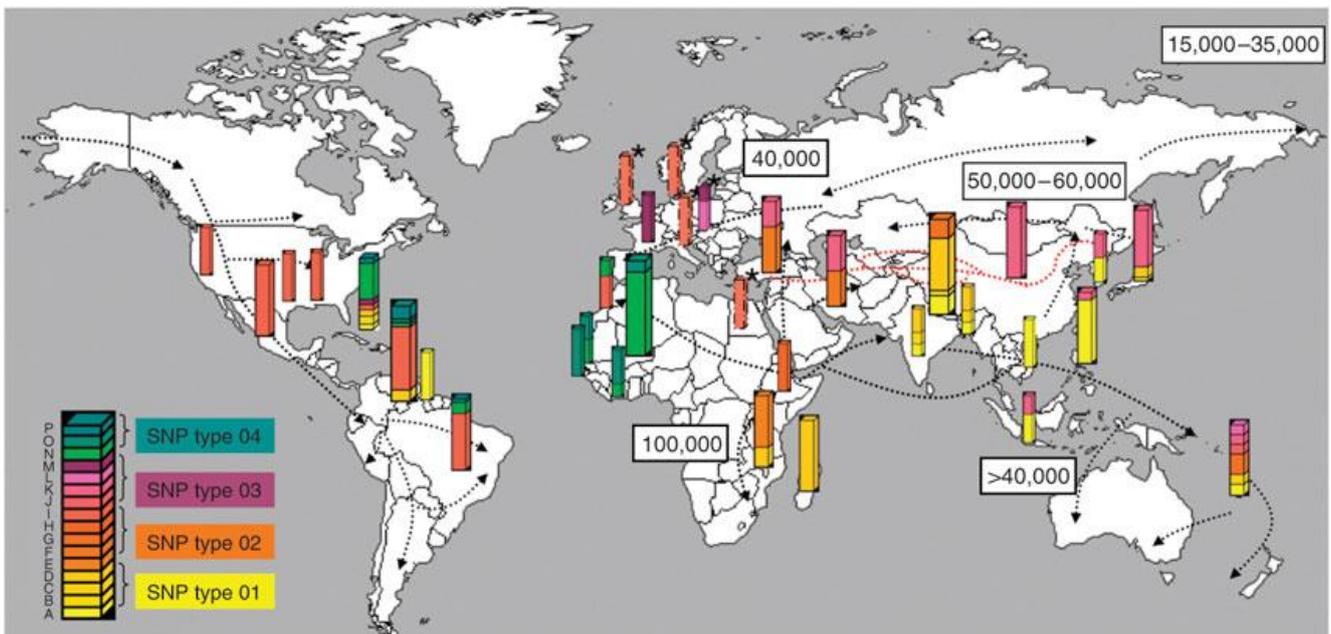
1.9 GENÉTICA DO *M.leprae*

O sequenciamento de genoma completo do *M. leprae* foi realizado pela primeira vez em 2001, por Stewart Cole e colaboradores. O estudo pioneiro revelou que este patógeno possui um genoma de 3.268.203 pb e um conteúdo médio G+C de 57,79% (46), semelhante ao que foi encontrado mais tarde para o *M. lepromatosis* – 3,206,741pb e G+C 57,89% (47) –, no entanto menor do que o de outras micobactérias, como o *M. tuberculosis* que possui um genoma de 4.411.532 pb e um conteúdo médio G+C de 65,61% (46). Além disso, em 2001 foi observado que apenas 49,5% do genoma do *M. leprae* contém genes codificantes de proteínas; 23,5% do genoma provavelmente corresponde a sequências regulatórias e 27% contém pseudogenes (sequências de leitura inativas com contrapartida funcional no *M. tuberculosis*) (46). Atualmente sabe-se que o *M. leprae* possui 1.115 pseudogenes e 2.770 genes, codificando um total de 1.605 proteínas (48).

Assumindo que o genoma do *M. leprae* já foi de tamanho semelhante aos das demais micobactérias, acredita-se que houve uma redução evolutiva e rearranjo de seu genoma durante a sua evolução. Este fato ajuda a explicar o parasitismo intracelular obrigatório, sua incapacidade de crescer *in vitro* e seu longo tempo de duplicação, quando comparado às demais espécies desse gênero (46).

Em estudo publicado em 2009, foi realizado o sequenciamento de quatro cepas de diferentes regiões do planeta (Índia, Brasil, Tailândia e Estados Unidos da América) que mostrou que essas cepas possuem 99,995% do genoma idêntico. Em seguida os pesquisadores realizaram genotipagem de um total de 175 cepas de 21 países dos cinco continentes; os resultados indicaram a clonalidade do *M. leprae* ao redor do mundo (49). O mesmo estudo também verificou 16 polimorfismos presentes em algumas cepas, que puderam ser utilizados para estimar a rota de distribuição da doença, que segue as rotas migratórias e comerciais humanas ao longo da história (Figura 7).

Figura 7. Distribuição das cepas de *M. leprae* pelo mundo, de acordo com SNPs e as rotas migratórias humanas.



Legenda: As rotas migratórias humanas estão indicadas pelas setas. As colunas estão localizados no país de origem da cepa de *M. leprae*, a cor é a codificação para cada um dos grupos de SNPs, a espessura da coluna corresponde ao número de amostras genotipadas e * indica os resultados obtido a partir de DNA de *M. leprae* da era medieval.

Fonte: Monot, 2009 (49).

Em 2013, foi realizada uma análise comparativa entre cepas medievais e modernas de *M. leprae*. Foram estudadas cinco cepas medievais, datadas de 1.010 a 1.383 d.C, utilizando DNA recuperado de antigos casos de hanseníase, obtido em cemitérios do Reino Unido, da Suécia e da Dinamarca, e 11 cepas modernas obtidas em diferentes regiões geográficas. Foi observado notável conservação do genoma do *M. leprae* ao longo do tempo, com exceção de quatro *SNPs* utilizados para a tipagem (50). Um estudo publicado recentemente analisou genomas de cepas de *M. leprae* obtidos de ossadas europeias do início da era medieval e identificou uma diversidade genética entre as cepas medievais que ainda não havia sido relatada. Ao analisá-las em conjunto com outras cepas medievais e modernas de estudos anteriores para a construção da árvore filogenética do *M. leprae*, as cepas medievais da Europa apareceram em quatro das principais ramificações, junto com cepas modernas de diferentes locais do globo. Assim, os autores propõem que i) a introdução de cepas de diferentes ramos, de diversos locais do mundo para a Europa aconteceu antes da era medieval ou ii) que o início da hanseníase se deu na Euroasia Ocidental ou até mesmo na Europa (51).

1.10 SUSCEPTIBILIDADE GENÉTICA DO HOSPEDEIRO À HANSENÍASE

O desenvolvimento e/ou a manifestação clínica de uma doença infecciosa após o contato patógeno-hospedeiro depende de dois fatores principais: a virulência do microrganismo e o grau de susceptibilidade inata do hospedeiro. Sabe-se que o genoma do *M. leprae* se mantém conservado ao longo dos anos (50) e que o bacilo é praticamente clonal ao redor do mundo (49). Portanto, é razoável atribuir grande parte do mecanismo que define o grau de suscetibilidade à hanseníase ao perfil genético do hospedeiro. Além disso, os subtipos clínicos de hanseníase são atrelados à resposta imune do hospedeiro o que sugere um papel importante do perfil genético do hospedeiro na definição dos diferentes estágios da doença. Estas hipóteses vem sendo reforçadas através de estudos observacionais com diferentes abordagens: estudos de gêmeos (52, 53), os estudos de agregação familiar (54) e

análises de segregação complexa (55-59) vêm reforçando que a genética do hospedeiro tem papel crítico nesses processos. Por exemplo, estudo de gêmeos realizado em 1973 por Vogel e Chakravarti em uma amostra populacional indiana composta por 102 pares de gêmeos, mostrou maior taxa de concordância para a hanseníase *per se* em gêmeos monozigóticos (MZ), quando comparado com gêmeos dizigóticos (DZ): dos 62 pares de gêmeos MZ analisados houve concordância em 37 (59,7%) para a hanseníase *per se*, enquanto que dos 40 pares de gêmeos DZ analisados apenas oito (20%) mostraram concordância no fenótipo (53). Em 1987, Shields *et al.* realizaram um estudo em 269 famílias contendo 552 indivíduos afetados, recrutados de uma população isolada de Papua – Nova Guiné. Nesse estudo, foi observada a passagem vertical da hanseníase (isto é, de pais para filhos através de múltiplas gerações) numa frequência de 89%, enquanto que a passagem horizontal da hanseníase foi observada numa taxa de 11% no grupo familiar (54). Em 2010, uma análise de segregação complexa realizada na Vila Santo Antônio do Prata, no estado do Pará – a mesma investigada no presente estudo –, envolvendo a população total da vila, de 2.005 indivíduos distribuídos em 112 famílias, com 225 casos, evidenciou, utilizando modelos matemáticos, uma forte dependência familiar com a presença de um componente majoritariamente genético, de herança codominante, influenciando a ocorrência da hanseníase (57).

No final do século XX, avanços no desenvolvimento de métodos e ferramentas de análise genética molecular possibilitaram o estudo do envolvimento de genes e regiões genômicas na susceptibilidade do hospedeiro à hanseníase, através de análises de ligação de regiões genômicas e de associação de genes candidatos. Mais recentemente, estudos pan-genômicos de associação (GWAS, em inglês *Genome-Wide Association Study*) têm sido realizados. Os GWAS, são estudos livres/geradores de hipótese no qual é feita a cobertura física completa de virtualmente todo o genoma em um único experimento de genotipagem. Esse desenho de estudo é possível, devido a um efeito muito importante na genética epidemiológica, denominado Desequilíbrio de Ligação (*LD* do inglês, *Linkage Disequilibrium*), um fenômeno de distribuição não aleatória de alelos em diferentes *loci* em uma população (60, 61). Entre as maneiras de medir estatisticamente o *LD*, o parâmetro mais usado é o r^2 , que pode variar entre 0 e 1. O r^2 indica a força de *LD* entre dois sítios polimórficos: quando igual a 1 indica perfeito *LD*, portanto os

marcadores capturam a mesma informação de associação (62). Dessa forma, em um estudo de associação, pode-se selecionar marcadores que capturam a informação de outros, denominados *tags*, diminuindo o custo e o tempo do experimento, bem como o número de testes estatísticos realizados (uma grande dificuldade de interpretação dos resultados dos GWAS é controlar para a grande inflação do erro tipo alfa (falso positivo) causada pelo grande número de testes de associação realizados sob uma única hipótese).

Esses estudos moleculares resultaram na identificação de variantes genéticas comuns relacionadas com a susceptibilidade do hospedeiro à hanseníase. Em seguida, parte dos genes previamente associados foram testados e validados em diferentes amostras populacionais, com *background* genético distinto. Serão discutidos nos tópicos a seguir os estudos que resultaram na associação dos genes relacionados com a susceptibilidade genética do hospedeiro à hanseníase, de interesse para o presente estudo.

1.10.1 A região 6q25-q27 e os genes *PRKN* e *PACRG*

Um *scan* genômico de ligação realizado em 2003, com a participação de nosso grupo de pesquisa, genotipou 388 marcadores microssatélites em 86 famílias sul-vietnamitas, contendo 205 irmãos afetados (63). Foi encontrada evidências de ligação entre a região cromossômica 6q25-q27 e a hanseníase *per se*. Em um estudo subsequentemente realizado pelo mesmo grupo, 64 *Single Nucleotide Polymorphism* (SNPs) localizados na região 6q25-q27 foram genotipados em uma amostra populacional composta de 197 famílias vietnamitas. Seis dos 64 SNPs foram associados com a susceptibilidade do hospedeiro à hanseníase, sendo que quatro desses estavam localizados na região promotora compartilhada pelos genes *PRKN* (*parkin RBR E3 ubiquitin protein ligase*) – anteriormente chamado de *PARK2* – e *PACRG* (*PARK2 co-regulated gene*). Para aumentar a cobertura deste *locus*, 81 marcadores genéticos distribuídos por toda a região dos genes *PRKN* e *PACRG* foram selecionados para serem genotipados. Destes, 17 marcadores localizados na região regulatória compartilhada por esses genes foram associados com a doença.

Finalmente, nove dessas variantes foram validadas em uma amostra populacional brasileira composta de 975 indivíduos (587 casos e 388 controles) (64). Curiosamente, as variantes genéticas *PRKN/PACRG* encontradas não explicam plenamente o sinal de ligação encontrado para a região 6q25-q27 em 2003; em 2016, variantes do gene *SOD2* (*superoxide dismutase 2, mitochondrial*) localizado nesta mesma região cromossômica, foram descritas associadas com hanseníase *per se* em duas amostras populacionais brasileiras independentes (65).

Desde a publicação dos achados originais, diversos estudos buscaram validar a associação dos genes *PRKN/PACRG* com a susceptibilidade do hospedeiro à hanseníase. Em 2006, Malhotra *et al.* realizaram um estudo caso-controle em uma amostra populacional composta por 636 indivíduos indianos (286 afetados), a fim de investigar a associação entre seis *SNPs* da região regulatória compartilhada pelos genes *PRKN* e *PACRG* e a susceptibilidade do hospedeiro à hanseníase. Evidência para associação foi novamente detectada, porém nenhum dos sinais resistiu à correção estatística de *Bonferroni* para múltiplos testes (66). Em 2012, um estudo caso-controle realizado por Li *et al.* em uma amostra populacional Han chinesa composta de 742 casos e 734 controles, analisou 15 *SNPs* dos genes *PRKN* e *PACRG*; nenhum deles mostrou associação significativa para o controle da susceptibilidade do hospedeiro à hanseníase nessa amostra populacional (67).

Em contraste com esses achados, Bakija-Konsuo *et al.* realizaram um estudo em uma amostra populacional composta por 88 indivíduos não hansenianos da ilha de Mljet (uma comunidade Croata geograficamente isolada, com um histórico bem documentado de hanseníase) e dois grupos controle populacionais sem histórico da doença: um grupo da ilha de Rab (93 controles) e outro da região do Split (160 controles). As frequências alélicas de dois *Tag SNPs* localizados no primeiro intron do gene *PACRG*, foram comparadas nesses dois grupos. Como resultado, observou-se que alelos de proteção para a hanseníase estão enriquecidos na população da ilha do Mljet, o que indica uma possível seleção positiva para esses alelos nessa população (68). Em 2013, um estudo caso-controle realizado por Chopra *et al.* envolvendo 2.685 indivíduos indianos (1.013 afetados), analisou 96 *SNPs* da região regulatória dos genes *PRKN* e *PACRG* e mostrou associação entre 11 desses *SNPs* e a hanseníase (69). Um estudo realizado por Alter *et al.* com 198 trios vietnamitas e 734 indivíduos indianos (364 afetados) mostrou que o *LD* e a

idade média de diagnóstico das amostras populacionais são cruciais para a validação dos achados originais para *PRKN/PACRG*. Com a construção de *bins* de marcadores em *LD* (utilizando um *cut-off* de $r^2 = 0,6$), os autores puderam observar diferenças na arquitetura de *LD* entre as duas amostras populacionais analisadas, indicando uma possível causa das diferenças de padrão de associação em diferentes amostras populacionais. No mesmo estudo, os autores realizaram uma estratificação da amostra populacional de acordo com a idade de diagnóstico, e observaram que a associação de marcadores da região promotora dos genes *PRKN/PACRG* com a hanseníase também são idade-dependente para dois dos três *bins* analisados (70). Em 2018, um estudo realizado em amostras populacionais chinesas, que buscou variantes funcionais através de estratégias de sequenciamento, identificou uma variante *missense* (rs1801582) no gene *PRKN*, sugestiva de associação com a susceptibilidade do hospedeiro à hanseníase (71).

1.10.2 A região 10p13

Além da identificação da região cromossômica 6q25-q27 ligada à hanseníase *per se*, o estudo na amostra populacional vietnamita também validou a ligação do *locus* 10p13 com o polo PB da doença (63), encontrada primeiramente por Siddiqui *et al.* em uma amostra populacional indiana (72). Assim, genes desta região cromossômica foram analisados em estudos subsequentes de associação. Em 2010, Alter *et al.* analisaram a associação do gene *MRC1* (*mannose receptor, C type 1*) com a susceptibilidade do hospedeiro à hanseníase e suas formas clínicas em duas amostras populacionais independentes, uma composta por famílias vietnamitas e a outra, por indivíduos brasileiros. A amostra vietnamita foi composta por 490 famílias *simplex* e 90 *multiplex* totalizando 704 pacientes, sendo 47% destes PB. Os indivíduos brasileiros foram recrutados no Rio de Janeiro e um total de 384 casos (51% PB) e 399 controles foram incluídos no estudo. Foi encontrada uma variante associada com a susceptibilidade do hospedeiro à hanseníase *per se* e à forma MB da doença nas duas populações estudadas (73). Na tentativa de validar esses achados, Wang *et al.* realizaram um estudo de associação em uma amostra populacional Chinesa composta por 527 casos (47% PB) e 583 controles. Foram

encontradas duas variantes no gene *MRC1* associadas com a hanseníase *per se* e a forma clínica PB e uma variante no gene *IFNG* (*interferon gamma*) associada com a forma MB da doença (74). Em 2014, Grant *et al.* realizaram um mapeamento de alta densidade na região cromossômica 10p13, cobrindo um total de 39 genes. Dos 1.335 *SNPs* genotipados na amostra populacional de descoberta – composta por 293 famílias vietnamita com 303 indivíduos afetados (37% PB) – 70 marcadores inicialmente associados ($P < 0,01$) foram genotipados em uma amostra de replicação. A amostra de replicação foi formada por 192 famílias *simplex* vietnamitas com 55% dos filhos afetados pela forma MB da doença. Foram identificados 3 marcadores associados com a susceptibilidade do hospedeiro à hanseníase *per se* e a forma MB da doença nos genes *CUBN* (*cubilin*) e *NEBL* (*nebulette*) (75). Em um estudo publicado em 2016 por Medeiros *et al.*, foi analisada a associação entre hanseníase e o gene *GATA3* (*GATA binding protein 3*), localizado na região 10p15. Dos sete *SNPs* genotipados em uma amostra populacional brasileira de 411 (25% PB) casos e 357 controles, um deles foi associado ($P < 0,01$) com a proteção do hospedeiro à hanseníase *per se*. Em seguida, este marcador foi genotipado em uma amostra brasileira de 511 casos (24% PB) e 380 controles e esta associação foi confirmada (76).

1.10.3 Os genes associados a partir de GWAS

O primeiro GWAS em hanseníase, foi realizado em 2009 por Zhang *et al.*, no qual 491.883 marcadores foram genotipados em uma amostra composta de 706 casos e 1.225 controles, formada predominantemente por chineses da etnia Han. Dos marcadores analisados, 93 mostraram forte evidência de associação e foram subsequentemente replicados em três amostras populacionais independentes, da mesma etnia, totalizando 3.254 casos e 5.955 controles. Como resultado, foi observada associação estatisticamente significativa entre *SNPs* de seis genes e a hanseníase: *CCDC122* (*coiled-coil domain containing 122*), *C13orf31* (atualmente conhecido como *LACC1* - *laccase (multicopper oxidoreductase) domain containing 1*), *NOD2* (*nucleotide-binding oligomerization domain containing 2*), *TNFSF15* (*tumor necrosis factor (ligand) superfamily, member 15*), *HLA-DR-DQ* (*major*

histocompatibility complex, class II DR-DQ) e *RIPK2* (receptor-interacting serine/threonine kinase 2). Além disso, uma tendência de associação foi observada entre o gene *LRRK2* (leucine-rich repeat kinase 2) e a doença (77).

Em 2011, foi realizada uma expansão do GWAS com a adição de 4.367 controles à primeira amostra. Um total de 1.701.673 SNPs em 706 casos e 5.581 controles foram analisados. Os SNPs com sinal significativo de associação que não estavam nas regiões identificadas anteriormente (*HLA-DR* e *HLA-DQA1*) foram genotipados em três amostras populacionais independentes, totalizando 3.301 casos e 5.299 controles de diferentes locais da China. Foram identificadas associações significativas entre *C1orf141* (chromosome 1 open reading frame 141), *IL23R* (interleukin 23 receptor), *RAB32* (*RAB32, member RAS oncogene Family*), *RPS6KA4* (ribosomal protein S6 kinase A4) e *CYLD* (*CYLD lysine 63 deubiquitinase*) e a susceptibilidade do hospedeiro à hanseníase *per se*. Na amostra combinada, dois destes, *IL23R* e *RAB32*, alcançaram valores de *P* significativo a nível de GWAS, de $2,67 \times 10^{-9}$ e $5,51 \times 10^{-12}$ respectivamente (78). Em um estudo subsequente, considerado como um terceiro estágio do GWAS, foram investigadas variantes em genes membros das famílias *Toll-like receptor (TLR)* e *CARD* (caspase recruitment domain), importantes reguladores da resposta imune inata. Após validações em amostras populacionais independentes, foi observada uma nova associação entre o gene *BCL10* (*B-cell CLL/lymphoma 10*) e a hanseníase *per se* (79).

Dando continuidade aos estudos utilizando os dados obtidos no GWAS, em 2015, Zhang *et al.* analisaram *loci* sugestivos de associação em 2.761 casos e 3.038 controles. Os SNPs com sinais significativos de associação foram replicados em cinco amostras populacionais distintas; nessas, além das associações já conhecidas em *IL23R*, *BCL10*, *IL12B* (interleukin 12B), *HLA-DR-DQ*, *RAB32*, *RIPK2*, *TNFSF15*, *LACC1* e *NOD2*, foram encontrados seis novos *loci* associados com a susceptibilidade do hospedeiro à hanseníase: *BATF3* (basic leucine zipper ATF-like transcription factor 3), *CDH18* (cadherin 18), *DEC1* (deleted in esophageal cancer 1), *ADO* (2-aminoethanethiol dioxygenase), *EGR2* (early growth response 2), *CCDC88B* (coiled-coil domain containing 88B) e *RMI2* (RecQ mediated genome instability 2). Também foram identificados sinais sugestivos de associação entre os genes *COX4I1* (cytochrome c oxidase subunit 4I1) e *LRRK2* e a hanseníase *per se* (80).

Em 2016, um novo GWAS foi publicado no qual 900.015 SNPs foram genotipados em 1.197 casos e 1.426 controles de hanseníase. Em seguida, foi feita imputação de variantes genéticas utilizando estes dados e os dados brutos dos GWAS anteriormente realizados pelo mesmo grupo separadamente (77, 80), e por fim uma meta-análise dessas populações. Os SNPs com resultados positivos de associação em *loci* que ainda não haviam sido associados com a hanseníase foram testados para validação utilizando uma estratégia *stepwise* em duas amostras independentes e uma nova meta-análise foi realizada, totalizando 8.156 casos e 15.610 controles com ancestralidade chinesa. Foram encontrados sinais significativos de associação com a susceptibilidade do hospedeiro à hanseníase *per se* para os genes *SYN2* (*synapsin II*), *BBS9* (*Bardet-Biedl syndrome 9*), *CTSB* (*cathepsin B*) e *MED30* (*mediator complex subunit 30*) (81). Em 2017, Liu *et al.* publicaram um GWAS de três estágios no qual foi feita uma busca por variantes em regiões codificantes. A amostra de descoberta foi composta por 1.670 casos e 2.321 controles chineses, na qual foram genotipados 273.028 variantes. Após filtragem das variantes, exclusão de variantes do complexo HLA e de variantes não raras de genes já associados com a hanseníase, 39 variantes foram genotipadas em 3.169 casos e 9.814 controles, e dessas 8 foram selecionadas para replicação em 2.231 casos e 2.266 controles. A meta-análise dos três estágios foi realizada e sete variantes codificantes – nos genes *NCKIPSD* (*NCK interacting protein with SH3 domain*), *CARD9* (*caspase recruitment domain family member 9*), *IL23R*, *FLG* (*filaggrin*), *TYK2* (*tyrosine kinase 2*), *SLC29A3* (*solute carrier family 29 member 3*) e *IL27* (*interleukin 27*) – mostraram associação significativa com a hanseníase (82).

Na tentativa de confirmar as associações encontradas no primeiro GWAS em hanseníase (77), foram realizados vários estudos independentes em diferentes amostras populacionais (**Figura 8**). No ano de 2010, um estudo de associação realizado por Wong *et al.* em dois grupos caso-controle indianos e uma amostra populacional do Mali, na África, encontrou associação entre a hanseníase e apenas dois desses genes, o *CCDC122* e o *LACC1* (83). No mesmo ano, um estudo caso-controle com 933 pacientes hansenianos, incluindo 240 com RT1 e 124 com RT2, e 101 controles mostrou associação entre o gene *NOD2* e a hanseníase *per se*, bem como com os estados reacionais da hanseníase (84). Em 2012, um estudo realizado por Grant *et al.* validou a associação dos genes *HLA-DR-DQ*, *RIPK2*,

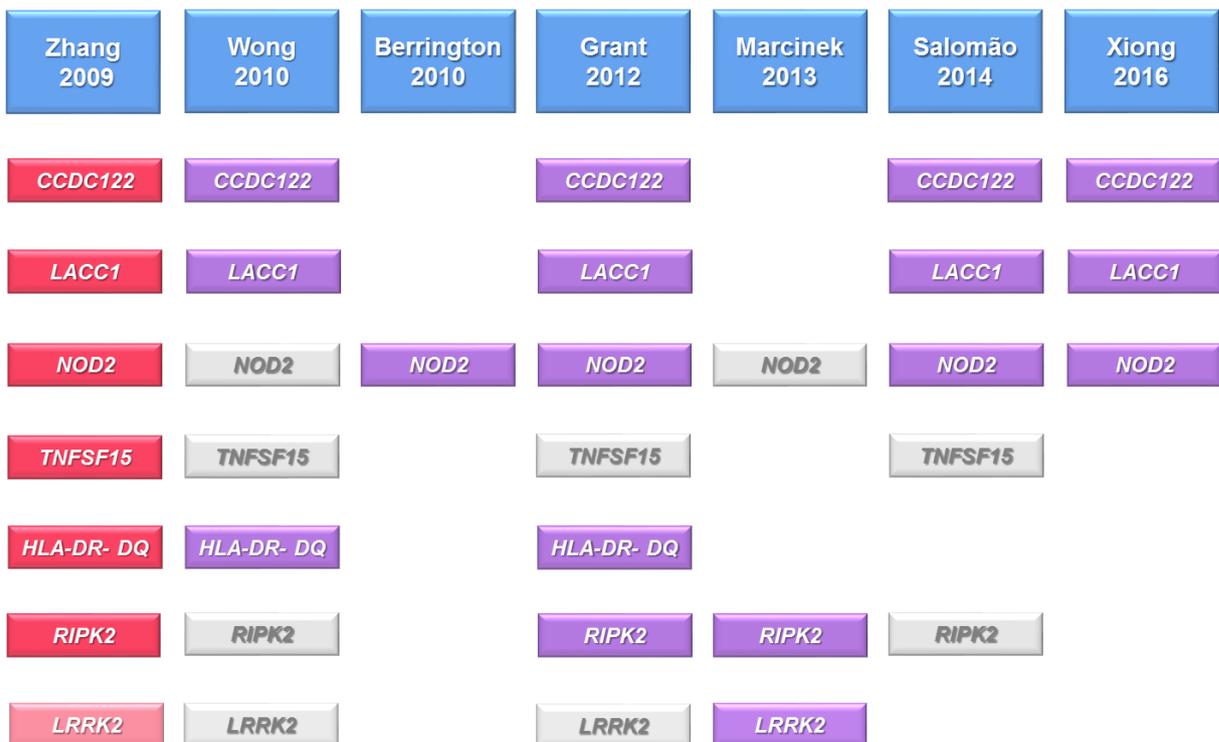
CCDC122/LACC1 e *NOD2* como fatores de susceptibilidade do hospedeiro à hanseníase em uma amostra populacional vietnamita (85). Um estudo realizado em 2013 em uma amostra populacional indiana mostrou associação dos genes *LRRK2* e *RIPK2* com o controle da susceptibilidade do hospedeiro à hanseníase (86). Em 2014, nosso grupo validou a associação entre o gene *NOD2* e o *locus CCDC122-LACC1* com a susceptibilidade do hospedeiro à hanseníase em amostras populacionais brasileiras (87).

Vários desses estudos (83, 85, 87) falharam em replicar a associação encontrada entre o gene *TNFSF15* e a hanseníase *per se* no GWAS de 2009 (77). Tendo isso em vista, Fava *et al.* notaram que mais de 85% da amostra populacional do estudo chinês possuíam incapacidades, o que é compatível com o endofenótipo da hanseníase de RT1, dessa forma o gene poderia estar associado com RT1 e não com hanseníase *per se* como foi anteriormente detectado. Neste sentido, em 2015 o grupo publicou um achado que mostra que – nas amostras analisadas, sendo a de descoberta composta por famílias vietnamitas e as amostras de replicação caso-controles brasileiras – na verdade o que ocorre é uma associação entre o gene *TNFSF8* (*TNF superfamily member 8*), gene vizinho ao *TNFSF15*, e RT1 (88). Em 2016, Xiong *et al.* publicaram um estudo no qual foi validada a associação entre a hanseníase *per se* e os genes *NOD2*, *LACC1* e *CCDC122* em uma população Yi chinesa (89). Neste mesmo ano, Fava *et al.* demonstraram, em uma amostra populacional composta por famílias vietnamitas, que a associação entre o gene *LRRK2* e a hanseníase ocorre preferencialmente na presença de RT1 (90). Por fim, em 2018, uma variante *missense* do gene *LACC1* foi associada com a susceptibilidade do hospedeiro à hanseníase, em amostra populacional chinesa através de estratégias *stepwise* de sequenciamento (71).

Dos genes associados com a susceptibilidade do hospedeiro à hanseníase *per se* no GWAS de 2009 (77), seis deles (*TNFSF15*, *NOD2*, *LRRK2*, *RIPK2*, *HLA-DR* e *LACC1*) foram associados anteriormente com duas doenças inflamatórias do intestino, a doença de Crohn e/ou colite ulcerativa. Provavelmente, essas doenças compartilham *loci* de susceptibilidade por possuírem mecanismos de resposta imunológicas/inflamatórias semelhantes. Além disso, há a hipótese de que a doença de Crohn também possa ser causada por uma micobactéria (91, 92). Em vista disso, Liu *et al.* analisaram 133 *SNPs* associados à susceptibilidade do hospedeiro às

doenças inflamatórias intestinais citadas acima, em amostras populacionais chinesas independentes totalizando 4.971 casos de hanseníase e 5.503 controles. Foram identificadas novas associações entre a susceptibilidade do hospedeiro à hanseníase *per se* e os genes *IL18RAP* (*interleukin 18 receptor accessory protein*), *IL18R1* (*interleukin 18 receptor 1*) e *IL12B* (93).

Figura 8 .Genes associados no primeiro GWAS com hanseníase e suas validações.



Legenda: Em vermelho, genes associados no primeiro GWAS em hanseníase; em vermelho claro tendência de associação; em roxo, genes validados e em cinza, genes testados e não validados com a susceptibilidade do hospedeiro à hanseníase.

Fonte: a autora,2019.

1.10.4 Os genes dos complexos *TLR* e HLA

Os membros da família de receptores *toll-like* têm sido estudados em doenças infecciosas, devido a sua importância no reconhecimento de patógenos e na resposta imune inata do hospedeiro (94). Um estudo publicado em 2009, por Bochud *et al*,

analisou a associação do gene *TLR-4* (*toll-like receptor 4*) com a susceptibilidade do hospedeiro à hanseníase *per se* e seus endofenótipos (forma clínica MB/PB, RT1 e RT2), em uma amostra populacional africana composta por 441 casos e 197 controles. Dos quatro *SNPs* analisados, dois mostraram sinais significativos de associação com proteção do hospedeiro à hanseníase *per se*. Para um dos marcadores analisados foi encontrado um sinal *borderline* de associação para proteção ao polo lepromatoso da doença (95). Em 2010, Wong *et al.* publicaram uma análise de associação *gene-centric*, cobrindo variantes em 2.092 genes em mais de 1.500 indivíduos de diferentes amostras populacionais, caso-controle e famílias. Os resultados sugeriram os genes *TLR1* (*toll-like receptor 1*) e *HLA-DRB1/DQA1* (*major histocompatibility complex, class II, DR beta 1/ class II, DQ alpha 1*) como importantes determinantes genéticos de susceptibilidade do hospedeiro à hanseníase (96). Em 2013, dois marcadores do gene *TLR1* foram testados em amostra populacionais de diferentes estados brasileiros; resultados significativos de associação com a hanseníase *per se* foram observados para a variante p.N248S nas duas amostras de descoberta, uma baseada em famílias de Minas Gerais e outra caso-controle de São Paulo, bem como nas amostras de replicação caso-controle do Rio de Janeiro e de Mato Grosso (97). Em um estudo publicado em 2018, esta mesma variante foi testada para validação em uma amostra populacional caso-controle, da mesma região de Minas Gerais que a amostra de famílias do estudo descrito acima (97). Foi obtido resultado estatisticamente significativo apenas após estratificação da amostra por gênero, que mostrou associação do genótipo heterozigoto com proteção à hanseníase em mulheres (98).

Além do pico de ligação observado no *locus* 6q25-q27 por Mira *et al.* (63), o mesmo estudo reportou um segundo pico de ligação na região cromossômica 6p21, que abriga o complexo HLA (antígeno leucocitário humano). Formado por genes altamente polimórficos com papel importante na regulação da resposta imune do hospedeiro (99). Com o objetivo de explorar essa região, Alcais *et al.* realizaram um *scan* de associação de 224 genes HLA em 194 famílias *simplex* vietnamitas. Os autores identificaram a associação entre um *SNP* funcional no gene *LTA* (*lymphotoxin alpha*) e a susceptibilidade do hospedeiro à hanseníase. Esse resultado foi validado em uma amostra populacional indiana, mas não na brasileira (100). Em 2011, um amplo estudo foi publicado, envolvendo quatro diferentes

amostras brasileiras, duas baseadas em famílias e duas caso-controles, compostas por i) 65 famílias nucleares de Duque de Caxias, Rio de Janeiro, ii) 54 famílias nucleares da Vila Santo Antônio do Prata, Pará, iii) 778 casos e 661 controles do Rio de Janeiro e por fim, iv) 368 casos e 375 controles de Bauru, São Paulo. Os autores testaram para associação de um *SNP* (rs1800629) do gene *TNF* (*tumor necrosis factor*) e genotiparam dois *SNPs* do gene *LTA* (nos controles), a fim de determinar se há *LD* entre variantes desses dois genes da região HLA classe III. Somente foi identificado resultado positivo de associação na amostra da Vila Santo Antônio do Prata. Em seguida foi realizada uma meta-análise, com dados publicados em estudos anteriores, que revelou evidência de associação deste *SNP* com a susceptibilidade do hospedeiro à hanseníase apenas quando resultados de amostras populacionais brasileiras estão incluídos, ou quando há na meta-análise apenas os dados de populações brasileiras. A análise de *LD* indicou uma independência dos marcadores dos genes *TNF* e *LTA* (101). Em 2012, Ali *et al.* publicaram um estudo no qual a região em questão foi explorada através de 111 *SNPs* em amostras populacionais indianas. Como resultado foram obtidos sinais positivos de associação entre o genes *BAT1* (*DEAD-box helicase 39B*), *NFKBIL1* (*NFκB inhibitor like 1*), *LTA*, *TNF*, *BTNL2* (*butyrophilin like 2*) e a susceptibilidade do hospedeiro à hanseníase *per se*, bem como à forma PB (102).

Um estudo de associação realizado em 2006 por Tosh *et al.* analisou o *locus* HLA 6p21 em uma amostra populacional de 196 famílias indianas com casos de hanseníase predominantemente PB. Foram encontradas associações significativas para os genes *HLA-DRB1*, *TNF*, *MICA* (*MHC class I polypeptide-related sequence A*) e *MICB* (*MHC class I polypeptide-related sequence B*) com a susceptibilidade do hospedeiro à forma PB da hanseníase (103). Em 2010, Alter *et al.* realizaram um estudo de associação de alta densidade, no qual 682 polimorfismos da região do complexo HLA foram genotipados em uma amostra populacional de 198 famílias vietnamitas com um único caso de hanseníase. Em seguida, os *SNPs* associados foram replicados em uma amostra contendo 292 famílias; foram observados 12 marcadores associados com um valor de *P* inferior a 0,01. Em seguida, foi realizada uma análise multivariada, que demonstrou que esses 12 marcadores podem ser representados por dois *SNPs* intragênicos da região HLA de classe I, próximos ao *HLA-C* (*major histocompatibility complex, class I, C*). Esses dois marcadores foram

então validados em uma amostra populacional indiana, envolvendo 364 casos e 371 controles (104). Um estudo conduzido em 2015 em uma amostra populacional brasileira – 202 casos *borderline* e 478 controles – levou à identificação da associação entre *HLA-C* e *HLA-DRB1* com a proteção do hospedeiro à hanseníase *borderline*. Além disso, foi observado efeito de proteção para um alelo do gene *HLA-DQB1* (*major histocompatibility complex, class II, DQ beta 1*) com a hanseníase BL (105). Em 2011, um estudo brasileiro realizado em uma amostra populacional composta por 223 casos (68,6% MB) e 201 controles paranaenses, analisou 17 variantes alélicas do gene *MICA*. Uma variante foi associada com a proteção do hospedeiro à hanseníase, essa mesma variante e uma outra foram associadas com a proteção para a forma MB da doença (106).

1.10.5 Outros genes não-HLA

No ano de 1998, Abel *et al.* encontraram ligação entre o gene *NRAMP1* (atualmente conhecido como *SLC11A1* - *solute carrier family 11 member 1*) e a susceptibilidade à hanseníase em famílias chinesas e vietnamitas (107). Roy *et al.* tentaram validar esses achados em uma amostra populacional indiana, contudo nenhuma associação entre o gene *NRAMP1* e a hanseníase foi encontrada. No mesmo estudo, foi identificada associação entre o gene *VDR* (*vitamin D (1,25-dihydroxyvitamin D3) receptor*) e as formas clínicas MB e PB da doença, com diferentes alelos de susceptibilidade (108). Em 2001, um estudo realizado com uma amostra populacional do Mali, na África, composta 273 pacientes e 201 controles, na qual três polimorfismos do gene *SLC11A1* foram estudados, não foi encontrada nenhuma associação com a hanseníase *per se*. No entanto, um desses polimorfismos mostrou resultado significativo de associação com a forma MB da doença (109). Em 2016, foi publicado um estudo envolvendo uma amostra populacional brasileira de 86 casos e 239 controles, na qual nove variantes do gene *SLC11A1* foram genotipadas; dessas, três foram associadas com a susceptibilidade do hospedeiro à hanseníase *per se*. Nenhuma variante foi associada com as formas clínicas da doença (110).

Um estudo realizado por Moraes *et al.* em uma amostra brasileira de 297 casos (44% PB) e 283 controles identificou *SNPs* do gene *IL-10* (*interleukin 10*) associados com a susceptibilidade à doença (111). Esta associação foi confirmada em diversos estudos de validação em amostras populacionais com *background* genético distinto (112). Anos mais tarde, o mesmo grupo conduziu um estudo em duas populações brasileiras, incluindo 2.125 indivíduos, que resultou na associação do gene *IFNG* com a proteção do hospedeiro à hanseníase (113).

Em 2007, um estudo realizado em uma amostra populacional do sul do Brasil, composta por 264 pacientes com hanseníase e 214 controles, identificou a associação entre haplótipos e variantes do gene *MBL2* (*mannose binding lectin 2*) com a susceptibilidade do hospedeiro à hanseníase *per se* e com o polo lepromatoso da doença (114). Um estudo publicado em 2009 em uma amostra populacional brasileira composta por 158 casos e 210 controles, apresentou evidências de que haplótipos funcionais do gene *FCN2* (*ficolin 2*) têm um efeito protetor na hanseníase (115).

Em 2010, um estudo revelou a associação entre polimorfismos do gene *LTA4H* (*leukotriene A4 hydrolase*) e a proteção à hanseníase MB em uma amostra populacional nepalesa (116). Em 2013, Zhang *et al.* realizaram um estudo de associação em uma amostra populacional Han Chinesa formada por 527 casos (48% PB) e 583 controles. Foram identificadas variantes associadas significativamente com a hanseníase e/ou suas formas clínicas (PB/MB) nos genes *FCN2*, *MBL2* e *CFH* (*complement factor H*) (117). Em 2018 um estudo publicado por Cardona-Pemberthy *et al* envolvendo 453 colombianos – sendo 114 pacientes, 111 contatos intradomiciliares não afetados e 228 familiares intradomiciliares não afetados – associou uma variante do gene *MBL2* com a susceptibilidade do hospedeiro à hanseníase PB (118).

Três variantes intrônicas do gene *CFH* foram associadas, após a genotipagem de 14 *SNPs* deste gene, em 527 casos e 583 controles chineses (119). Chaitanya *et al.* identificaram, em 2014, associação entre o gene *IL17F* (*interleukin 17F*) e a susceptibilidade do hospedeiro à hanseníase *per se* em uma amostra populacional da Índia de 140 casos e 84 controles (120).

Diversos estudos, em diferentes amostras populacionais, têm revelado associação entre genes distintos e a susceptibilidade do hospedeiro à hanseníase. Contudo, os estudos de associação tipicamente utilizam da estratégia de analisar marcadores *Tag SNPs* com frequência alélica mínima superior a 5%. Essa estratégia de baseia no *LD* desses marcadores com outros *SNPs*, a fim de ser realizada a cobertura física completa do gene com o menor número de marcadores genéticos possíveis. Dessa maneira, os estudos de associação são incapazes de prever se as variantes em questão, responsáveis por esses sinais significativos de associação, são causais ou estão em *LD* com a verdadeira variante causal. Dessa forma, as variantes causais ainda não foram identificadas para a maior parte desses genes.

1.11 SEQUENCIAMENTO MULTIPARALELO

O desenvolvimento das técnicas de sequenciamento multiparalelo (segunda geração, *NGS – Next generation sequencing*), que teve seu início em 2004, proporcionou a realização do sequenciamento de amplas frações do DNA e de diversas amostras em um único experimento, de forma mais eficiente, rápida e barata que o sequenciamento clássico pelo método de Sanger. As técnicas de sequenciamento são capazes de detectar variantes de DNA presentes em baixa frequência na amostra a ser sequenciada (121). A realização dessas técnicas consiste em três principais etapas, a saber:

- i) O preparo da biblioteca, que tem como principal objetivo a obtenção de fragmentos de DNA ligado a adaptadores;
- ii) O preparo de *template*, que consiste em uma PCR com a finalidade de amplificar o DNA alvo para o sequenciamento. Para que essa etapa ocorra, os adaptadores da biblioteca se anelam a oligonucleotídeos complementares ancorados em uma superfície fixa;
- iii) A reação de sequenciamento propriamente dita, que pode ou não utilizar nucleotídeos modificados.

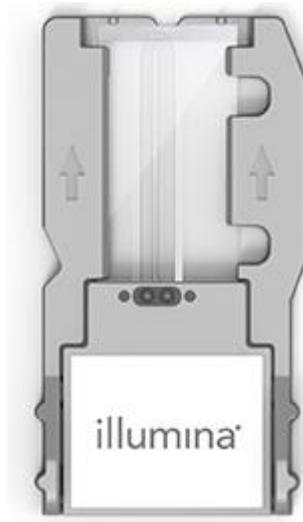
Neste estudo, utilizou-se técnicas de sequenciamento por síntese conforme implementados na plataforma *Illumina MiSeq* e descrito a seguir.

1.11.1 *MiSeq System – Illumina*

Utilizamos a tecnologia de sequenciamento na plataforma *MiSeq* da *Illumina*, com o intuito de realizar um re-sequenciamento de regiões alvo; dessa forma o preparo da biblioteca foi feito utilizando-se *primers* específicos de acordo com a região de interesse a ser sequenciada. Em um primeiro momento, esses *primers* hibridizam com o DNA genômico e ocorre a amplificação do fragmento de interesse. Em seguida, ocorre a ligação de diversos oligonucleotídeos necessários para a PCR, durante o preparo de *template*, e para o sequenciamento *pair-end*, sendo a estrutura da biblioteca composta por um grande *pool* de moléculas. Primeiramente são ligados os oligonucleotídeos necessários para dar início a leitura em ambas as extremidades do fragmento de DNA. Em seguida são ligados os indexadores, conhecidos como i5 e i7, usados em combinações distintas a fim de distinguir as *reads* de diferentes amostras sequenciadas simultaneamente. Por fim, são ligados, o P5 e o P7, oligonucleotídeos complementares àqueles fixados ao *flow cell*, necessários para dar o início ao preparo do *template*.

O preparo de *template* padronizado para a essa plataforma é realizado em uma superfície sólida conhecida como *flow-cell*, um chip contendo um canal (**Figura 9**).

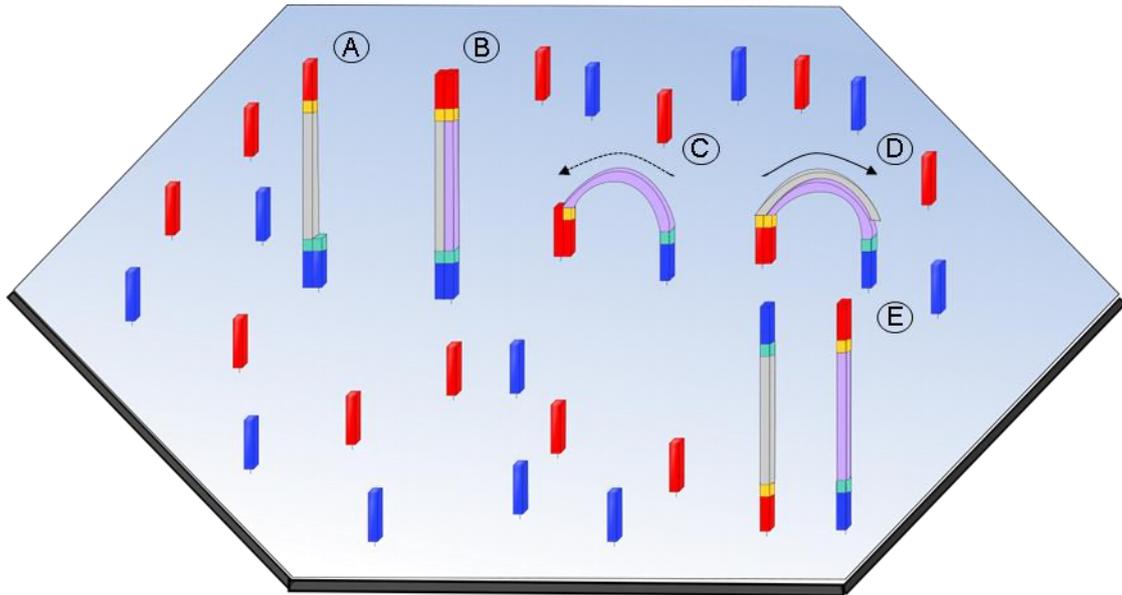
Figura 9. *Illumina flow-cell*



Fonte: *Illumina*,(122).

Em sua superfície, há dois tipos de oligonucleotídeos, *forward* (P5) e *reverse* (P7), ligados covalentemente e complementares a sequências ligadas ao DNA durante o preparo da biblioteca. Resumidamente, o DNA é desnaturado e hibridizado aos oligonucleotídeos da superfície (**Figura 10 A**), em seguida ocorre o alongamento da fita complementar (**Figura 10 B**) e a fita molde, por não estar anexada na superfície da *flow-cell*, se desliga. Posteriormente, é formada a ponte de amplificação com a hibridização do adaptador localizado na extremidade livre do fragmento de DNA a um oligonucleotídeo próximo anexado na *flow-cell* (**Figura 10 C**) e o alongamento da fita complementar (**Figura 10 D**). No final do processo obtêm-se a fita *forward* e a *reverse* da molécula de DNA (**Figura 10 E**). Essas etapas são repetidas inúmeras vezes até que, no final do processo são formados os *clusters* (**Figura 11**).

Figura 10. Figura esquemática detalhando o preparo de *template* na plataforma *Illumina*.

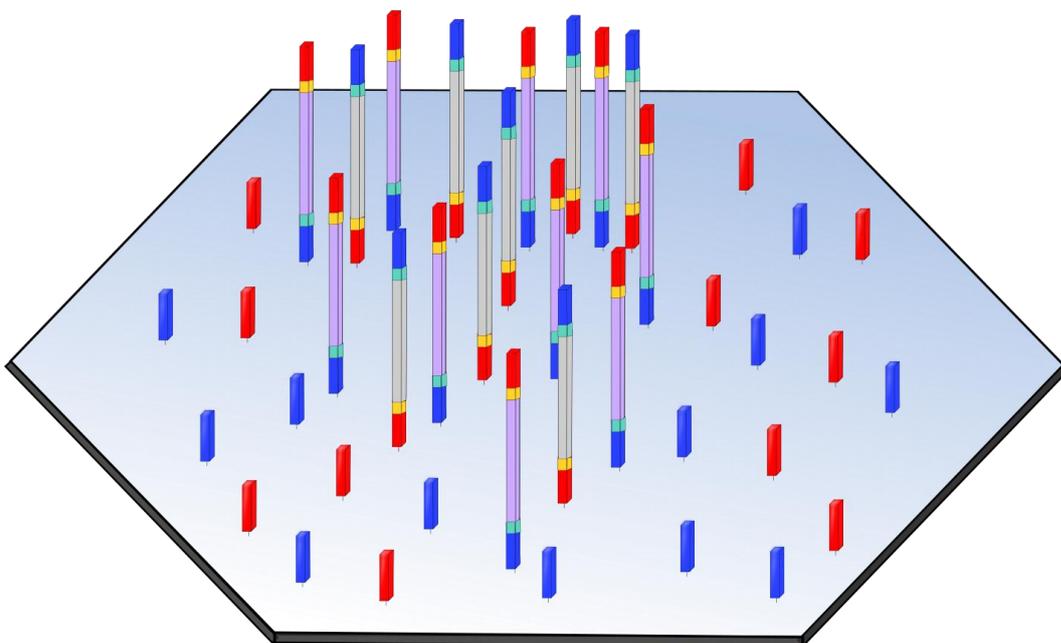


Legenda: Imagem esquematizando as etapas do preparo de *template* na plataforma *Illumina* MiSeq. Em A o anelamento do DNA, em B a extensão da fita complementar, em C a ponte para a amplificação, em D o alongamento da fita complementar, em ponte e em E as ambas fitas de DNA presentes. Em vermelho e azul temos os *primers forward (P5)* e *reverse (P7)*. A fita molde do DNA está em cinza e a complementar em roxo. Os indexadores estão em amarelo e verde.

A flecha tracejada indica a formação da ponte de amplificação e a flecha contínua o alongamento da molécula de DNA.

Fonte: A autora, 2016. Adaptado de Metzker, 2010 (123).

Figura 11. *Cluster* formado durante o preparo do *template* no equipamento *Illumina* MySeq.

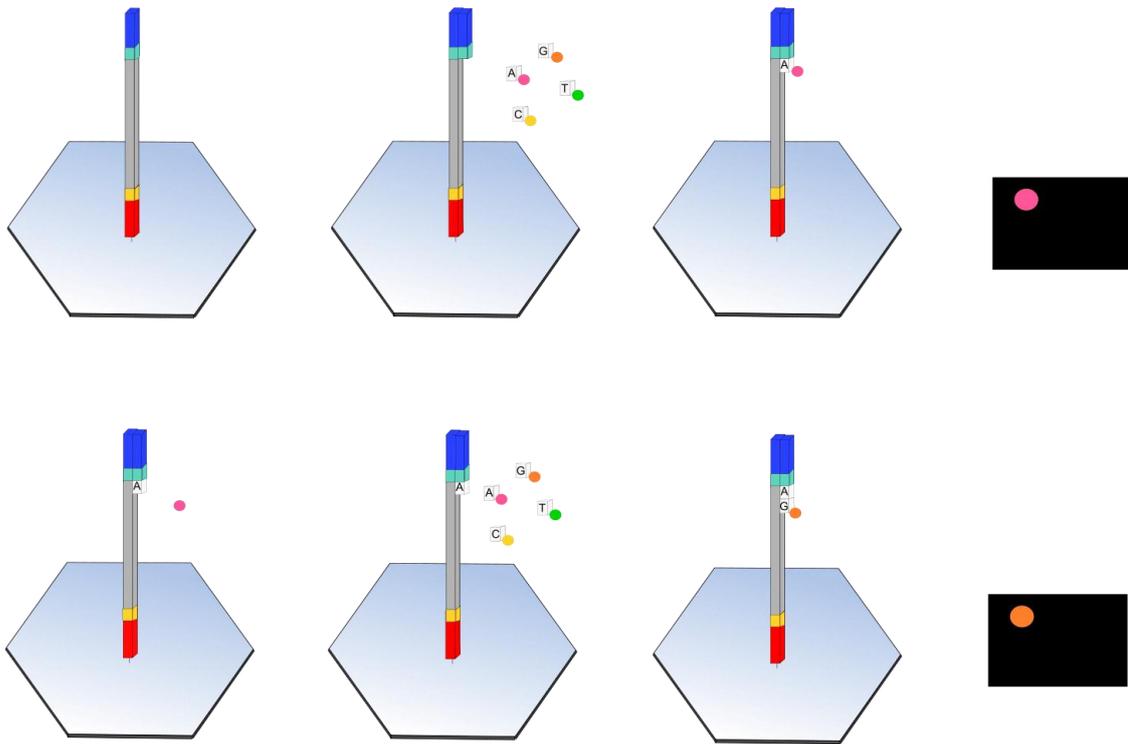


Legenda: Em vermelho e azul temos os *primers forward (P5)* e *reverse (P7)*. Os indexadores estão em amarelo e verde. A fita molde do DNA está em cinza e a complementar em roxo.

Fonte: A autora, 2016. Adaptado de Metzker, 2010 (123).

Após a formação do *cluster* ocorre o desligamento de uma das fitas de DNA da superfície da *flow-cell* e se inicia o sequenciamento. Durante esse processo, ocorre a hibridização do *primer*, para o início da leitura e o aporte de dNTPs marcados com fluorescência com um bloqueador na extremidade 3' terminal, para impedir a ligação de mais de um nucleotídeo. Ocorre então a inserção do nucleotídeo complementar na sequência pela ação da DNA polimerase. Posteriormente, os dNTPs não ligados são lavados e a fluorescência emitida pelos dNTPs ligados a cada cluster é detectada pelo sequenciador (**Figura 12**). Em seguida, ocorre a clivagem do fluoróforo e do bloqueador seguido de uma segunda lavagem, dando continuidade a um novo ciclo do sequenciamento. Após o sequenciamento da *read 1* ter sido concluído, ocorre a leitura dos indexadores i5 e i7. Em seguida, há uma nova PCR em ponte e o processo detalhado na **Figura 12** ocorre novamente para que possa ser feito sequenciamento da outra fita de DNA. Em seguida, é realizado o sequenciamento da *read 2* dá mesma forma que o da *read 1*, que é finalizado com a leitura dos indexadores. Dessa forma, cada *read* (sequência curta do DNA sequenciado) tem início em uma extremidade do fragmento, uma estratégia conhecida como sequenciamento *paired-end*.

Figura 12. Figura esquemática do processo de sequenciamento em *Illumina*.



Legenda: Em vermelho e azul temos os *primers forward (P5)* e *reverse (P7)*. Os indexadores estão em amarelo e verde. A fita do DNA está em cinza e os dNTPs em branco. O quadrado preto mostra a detecção da fluorescência para diferentes nucleotídeos (para A em rosa e G em laranja).

Fonte: A autora, 2016. Adaptado de Metzker, 2010 (123).

2. JUSTIFICATIVA E HIPÓTESE DO ESTUDO

Tipicamente, estudos de associações de genes candidatos e *GWAS* utilizam variantes comuns ($MAF > 0,05$), portanto informativas na população estudada (124, 125), e selecionam marcadores do tipo *Tag* de forma a capturar a informação de um conjunto de marcadores em *LD*. Desta forma, este modelo de estudo exclui um potencial papel de variantes raras (*MAF* menor que 1%) no controle da suscetibilidade à doença (124) e não são desenhados para detectar a variante causal (60) – quando se identifica associação, a causa é provavelmente *LD* entre o marcador associado e a(s) variante(s) funcional(is), que permanece(m) desconhecida(s). Seguindo este modelo, diversos estudos anteriores de associação, em diferentes amostras populacionais com *background* genético distintos identificaram genes associados com a susceptibilidade do hospedeiro à hanseníase. No entanto, devido à natureza desses estudos, descrita acima, não se pode afirmar se as variantes associadas são funcionais, portanto, com potencial efeito causal.

Nossa hipótese é de que os marcadores associados previamente à hanseníase estão em *LD* com variantes exônicas, raras e/ou desconhecidas, detectáveis apenas por sequenciamento direto, com potencial efeito funcional/causal. Esta hipótese foi testada através do sequenciamento paralelo massivo das regiões codificantes e *UTR* (*untranslated region*) de genes candidatos ao controle da susceptibilidade do hospedeiro à hanseníase em um grupo de famílias recrutadas na Vila Santo Antônio do Prata, uma população isolada, hiperendêmica para hanseníase e enriquecida em fatores de risco genéticos para a doença.

3. OBJETIVOS

3.1 OBJETIVO GERAL

O objetivo geral deste estudo foi identificar variantes genéticas raras ou comuns, localizadas nas regiões transcritas traduzidas e não-traduzidas (*UTR*) de 39 genes, candidatas a exercer relação causal com a reconhecida associação destes genes com susceptibilidade do hospedeiro à hanseníase.

3.2 OBJETIVOS ESPECÍFICOS

1. Obter as sequências exônicas dos 39 genes candidatos, na população de estudo, utilizando sequenciamento de segunda geração;
2. Processar as sequências (alinhamento e controle de qualidade) para identificar variantes genéticas presentes em regiões exônicas;
3. Analisar se os genes e/ou as variantes obtidas no sequenciamento estão associadas com a susceptibilidade do hospedeiro à hanseníase *per se*;
4. Verificar se as variantes associadas estão em desequilíbrio de ligação com variantes previamente associadas, no mesmo gene, em estudos anteriores;
5. Testar para associação haplotípica gênica ou cromossômica;
6. Realizar análises *in silico* do impacto funcional das variantes detectadas e filtradas.

4. MATERIAIS E MÉTODOS

4.1 ETAPAS PRÉ-SEQUENCIAMENTO

4.1.1 Considerações éticas

Os indivíduos envolvidos nesse estudo, ou seus representantes legais, foram anteriormente recrutados, assinaram um Termo de Consentimento Livre e Esclarecido (TCLE), responderam a um questionário e tiveram amostra de sangue periférico coletada. O presente estudo está incluído em um projeto maior, intitulado “Estudo de fatores de risco genéticos d susceptibilidade do hospedeiro à hanseníase”, aprovado pelo Comitê de Ética em Pesquisa (CEP) da PUCPR (Pontifícia Universidade Católica do Paraná) - Ofício nº 51/07/CEP-PUCPR de 29 de fevereiro de 2007 -, na Comissão Nacional de Ética em Pesquisa (CONEP) - Parecer nº 678/2006 de 30 de junho de 2006 - e no Comitê de Ética em pesquisa da Organização Mundial da Saúde (*The Ethics Research Commitee*). Por se tratar de uma tecnologia distinta aos estudos anteriores, mesmo que a amostra populacional já tenha sido recrutada anteriormente e amostras de DNA dos participantes da pesquisa já estejam armazenadas em nosso laboratório, optamos por submeter esse estudo como um novo projeto de pesquisa ao CEP da PUCPR; desta forma, o projeto foi submetido, analisado e aprovado, sob o parecer 807.698 de 26 de setembro de 2014.

4.1.2 Suporte Financeiro

O projeto foi contemplado com € 42.665,25 no edital lançado pelo *Leprosy Research Initiative*, chamada 704.16.31/LRI/DEF no ano de 2016.

4.1.3 Amostra Populacional

A amostra populacional desse estudo foi recrutada na Vila Santo Antônio do Prata, localizada no estado do Pará, norte do Brasil – especificamente, no município de Igarapé-Açú, a cerca de 110 Km a leste de Belém, capital do Pará. A Vila Santo Antônio do Prata foi fundada em 1898 por frades franciscanos. A partir de 1920, hansenianos foram enviados para lá com a finalidade de isolá-los da população em geral. O isolamento foi compulsório até o ano 1962; porém, a população permanece isolada e hiperendêmica para hanseníase até os dias atuais. Hoje, a população da Vila do Prata é composta por aproximadamente 3.500 habitantes, constituída predominantemente por indivíduos jovens, solteiros e pardos. Em 2010, cerca de 13% (257 casos em 2.005 habitantes) da população tinha histórico de hanseníase ou doença ativa, com predominância da forma lepromatosa (40,5%) (57).

Para este estudo, foi recrutada uma amostra baseada em famílias, composta por 177 indivíduos residentes na Vila Santo Antônio do Prata, distribuídos em 40 *pedigrees*. A **tabela 1** trás uma descrição das principais características demográficas da amostra populacional.

Tabela 1: Caracterização da amostra populacional

Sexo	n	%
Homens	87	49,15
Mulheres	90	50,85
Sexo filhos afetados (n=68)	n	%
Homens	33	48,53
Mulheres	35	51,47
Status	n	%
Afetado	116	65,54
Não afetado	59	33,33
Não informado	2	1,13
Etnia	n	%
Branca	18	10,17
Preta	29	16,39
Parda	127	71,75
Não informado	3	1,69
Etnia filhos afetados (n=68)	n	%
Branca	7	10,29
Preta	11	16,18

Parda	49	72,06
Não informado	1	1,47
Idade de diagnóstico em anos (n=105)		
Média (dp)	21,02 (13,58)	
Mínima	2	
Máxima	75	
Idade de diagnóstico dos filhos afetados, em anos (n=58)		
Média (dp)	15,12 (9,32)	
Mínima	2	
Máxima	45	

Legenda: n,número de indivíduos; dp, desvio padrão;

4.1.4 Coleta do material biológico e extração de DNA

Amostras de 5 mL de sangue periférico foram obtidas em expedições anteriores realizadas na Vila Santo Antônio do Prata, em tubos contendo anticoagulante ácido etilenodiamino tetra-acético (EDTA). A extração de DNA foi realizada pela técnica de *salting out* (126). O DNA recuperado foi quantificado por espectrofotometria usando o equipamento *NanoDrop 2000/2000c* e sua qualidade foi estimada através das razões de absorbância 260/280 e 260/230. Soluções de trabalho do DNA foram preparadas através da diluição das amostras originais em água MiliQ para a concentração de 20 ng/ μ L. Todas as amostras foram armazenadas em freezer a -20°C até o momento de seu uso.

4.1.5 Seleção dos genes candidatos

Os genes candidatos foram selecionados com base em revisão de literatura com foco em estudos publicados até o final de 2016 e realizados em diferentes amostras populacionais de *background* étnico distinto. Foram incluídos os genes não HLA (classe I e II) associados i) em ao menos duas amostras populacionais distintas ou ii) no GWAS de Zhang *et al.* (77) e suas fases seguintes (78-80). Esta estratégia resultou na seguinte lista de genes incluídos no estudo: *PACRG*, *PRKN*, *SOD2*, *MRC1*, *CUBN*, *NEBL*, *GATA3*, *LACC1*, *CCDC122*, *RIPK2*, *NOD2*, *IL23R*, *IL12RB2*, *BCL10*, *BATF3*, *ADO*, *CCDC88B*, *CDH18*, *DEC1*, *EGR2*, *RMI2*, *COX4I1*, *IL1RL1*,

IL18R1, IL18RAP, TLR1, TLR4, LTA, TNF, MICA, MICB, NRAMP1, VDR, IL10, LTA4H, MBL2, FCN2, IFNG e IL17.

Como sequências-alvo, foram incluídos os exons e região 3' e 5' UTR de todos os 39 genes candidatos.

4.2 SEQUENCIAMENTO

O sequenciamento, conforme implementado em plataforma *Illumina MiSeq*, foi realizado em parceria com grupo de pesquisa da *McGill University*, colaborador formal no estudo. O preparo da biblioteca foi executado utilizando *amplicons* desenhados especificamente para este estudo, conforme as recomendações descritas no protocolo *TruSeq Custom Amplicons v1.5* (127). O desenho total dos *amplicons* incluiu 1.344 regiões de interesse, com o objetivo de se obter fragmentos de DNA de em média 425 pares de base. Os preparos de *template* e as reações de sequenciamento foram realizadas conforme descrito no protocolo *Illumina TruSeq Custom Amplicom v1.5 Reference Guide v01* (128-130), utilizando 600 ciclos em um sequenciamento *pair-end*, gerando *reads* de 300pb.

4.3 ANÁLISE DOS DADOS

4.3.1 Análise do sequenciamento

A qualidade geral do sequenciamento foi verificada individualmente para cada amostra utilizando os programas *FastQC* versão 0.10.1 (131) e *PrinSeq* 0.20.4 (132). As *reads* obtidas foram alinhadas com as regiões de interesse do genoma humano de referência (UCSC Genome Browser hg19) utilizando o *BWA-SW* (Burrows-Wheeler Aligner, *Smith-Waterman algorithm*) que é recomendado quando se tem *reads* longas, com alta possibilidade de erros no sequenciamento (133). A partir das *reads* alinhadas, foi criado um arquivo de chamada de variantes para cada um dos indivíduos (gvcf), que possui informações de todas as bases da região

sequenciadas, com o *HaplotypeCaller* do *GATK* 3.8-1 (134). Em seguida, os arquivos de cada indivíduo foram unidos em um único arquivo contendo os dados de todos os indivíduos e foi feita a chamada das variantes a nível populacional com o *GenotypeGVCFs* do *GATK*. As variantes foram então filtradas, utilizando os seguintes *cut-offs* como critérios de qualidade:

i) Para *SNPs*: (razão qualidade por cobertura do alelo - *QualByDepth* (*QD*) < 2, qualidade do mapeamento – *MappingQuality* (*MQ*) < 40, *StrandOddsRatio* > 3, *MQRankSum* < -12,5 e *ReadPosRankSum* < -8,0)

ii) Para *InDels* (*QD* < 2, *StrandOddsRatio* > 10, *ReadPosRankSum* < -20,0 e *InbreedingCoeff* < -0,8).

Em seguida filtragem adicional, para garantir a qualidade das variantes e diminuir a possibilidade de falsos positivos, foi efetuada utilizando o *vcftools* (135) como se segue:

i) Genótipos com qualidade do genótipo <30 na escala *Phred* e/ou cobertura <10X foram considerados como *missing*;

ii) Indivíduos com mais de 10% do total de variantes não cobertas, foram considerados como *missing*; também foram excluídos aqueles com baixo fator de parentesco (*relatedness* <0,20) entre os filhos afetados e seus pais;

iii) Variantes com cobertura média na população <30X e *call rate* < 90% foram excluídas. Também foi feita a exclusão de variantes com o *P* do equilíbrio de Hardy-Weinberg menor de 0,05 nos indivíduos não afetados da amostra populacional.

A cobertura dos exons dos genes sequenciados foi estimada utilizando a ferramenta *DepthOfCoverage* do *GATK*, sendo que foram consideradas adequadas aquelas regiões que possuíam uma média de cobertura na amostra populacional superior a 30X.

A anotação das variantes encontradas foi feita através do *software ANNOVAR* (136).

4.3.2 Análise de Associação

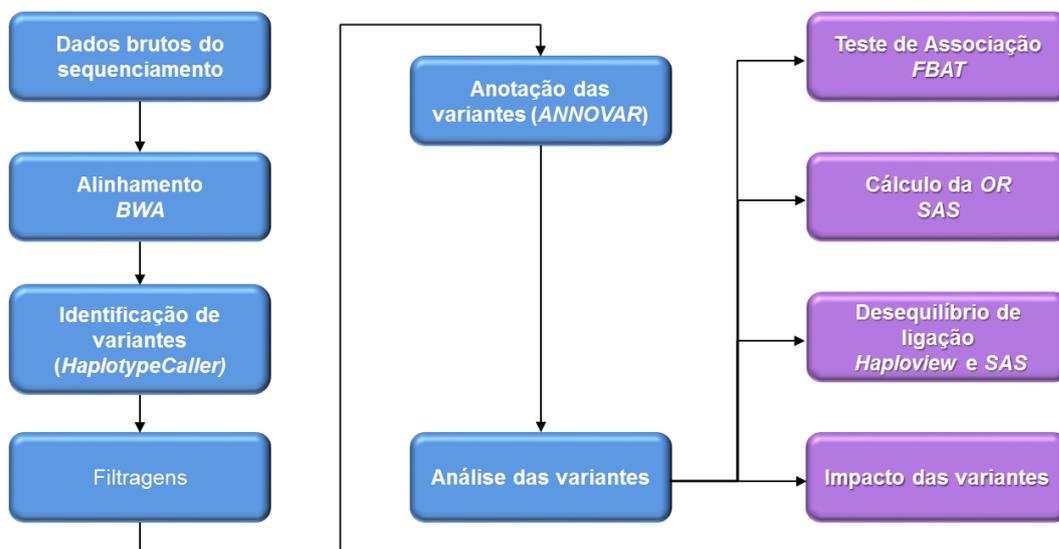
As análises de associação por gene (*gene-wise*) e por variante com $MAF > 0,04$, foram realizadas através do Teste de Desequilíbrio de Transmissão (TDT) no programa *Family-Based Association Test (FBAT)*, versão 2.0.4 (137). A análise *gene-wise* permite verificar a contribuição de variantes raras e de baixa frequência (frequência alélica $< 0,03$) em conjunto na associação entre o gene testado e o fenótipo em questão. São consideradas as variantes presentes em ao menos uma família informativa (138).

Análise por variante foi realizada sob os modelos aditivo e dominante. O programa *FBAT* também foi utilizado para análises haplotípicas de variantes associadas localizadas no mesmo gene ou cromossomo e que não estão em *LD*. Foram considerados haplótipos que possuem uma frequência mínima de 5% na população. A análise do desequilíbrio de ligação foi realizada em quatro cenários diferentes: i) entre marcadores associados nesse estudo e localizados no mesmo gene, utilizando-se o programa *Haploview* versão 4.2; para as variantes que tiveram um valor de r^2 superior a 0,5, o programa *SAS* versão 9.1 foi utilizado para realizar a análise de regressão logística multivariada, a fim de verificar a (in)dependência do sinal de associação; ii) entre marcadores deste estudo e variantes intrônicas associadas com a hanseníase em outros estudos de nosso grupo, realizados na Vila do Prata, quando localizadas no mesmo gene, utilizando o *Haploview* versão 4.2, a fim de verificar se a associação encontrada anteriormente é devido ao desequilíbrio de ligação entre essas e as variantes de regiões exônicas identificadas no presente estudo. Para isso, foi feita uma sobreposição entre os dados da amostra populacional desse estudo com as amostras dos estudos dos genes em questão; iii) entre marcadores associado nesse estudo e em outros já publicados, independentemente da população em que foi realizado. Essa análise foi feita utilizando-se a base de dados genômicos do projeto *Ensembl* (139), versão hg19 do genoma humano; e iv) entre todos os *SNPs* já associados em diferentes amostras populacionais, incluindo a desse estudo, utilizando a base de dados da amostra populacional africana loruba (YRI), disponível no projeto *1000 genomes*, a fim de

construir os *bins* (um grupo de *SNPs* fortemente correlacionados por *LD* e que não são necessariamente consecutivos (140)) entre os marcadores dos genes associados que possuem um coeficiente de correlação (r^2) maior ou igual à 0,6. A construção foi feita após análise de *LD* no *Haploview*; utilizando o programa *Inkscape*.

A técnica de *pseudosibs* (programa adaptado pelo nosso grupo) e o programa *SAS* foram utilizados para calcular a razão de chance (em inglês *Odds ratio*, *OR*) e seu intervalo de confiança, através da análise de regressão logística, bem como o valor de *P* para aquelas variantes em que foi necessário realizar a análise multivariada condicional entre marcadores. Por fim, a previsão *in silico* do impacto das variantes não sinônimas sobre a estrutura e função da proteína foi feita usando *Polyphen* (141), *SIFT* (142), *LRT*, *MutationTaster* e *CADD v.1.3* (143). O valor do *CADD* indica o dano baseado em uma escala na qual valores acima de 20 indicam as variantes que estão no *top* 1% de variantes mais deletérias do genoma humano e, acima de 30, no *top* 0,1%. Para variantes sinônimas ou em regiões *UTR*, foi estimado o impacto sobre a expressão dos genes utilizando-se o software *HaploReg v.4.1* (144), *GTEX v.8* (145), *Blood eQTL* (146) e *ImmunPOP* (147). Um fluxograma resumindo as etapas da análise dos dados está apresentado na **Figura 13**.

Figura 13. Fluxograma das etapas de análise dos dados



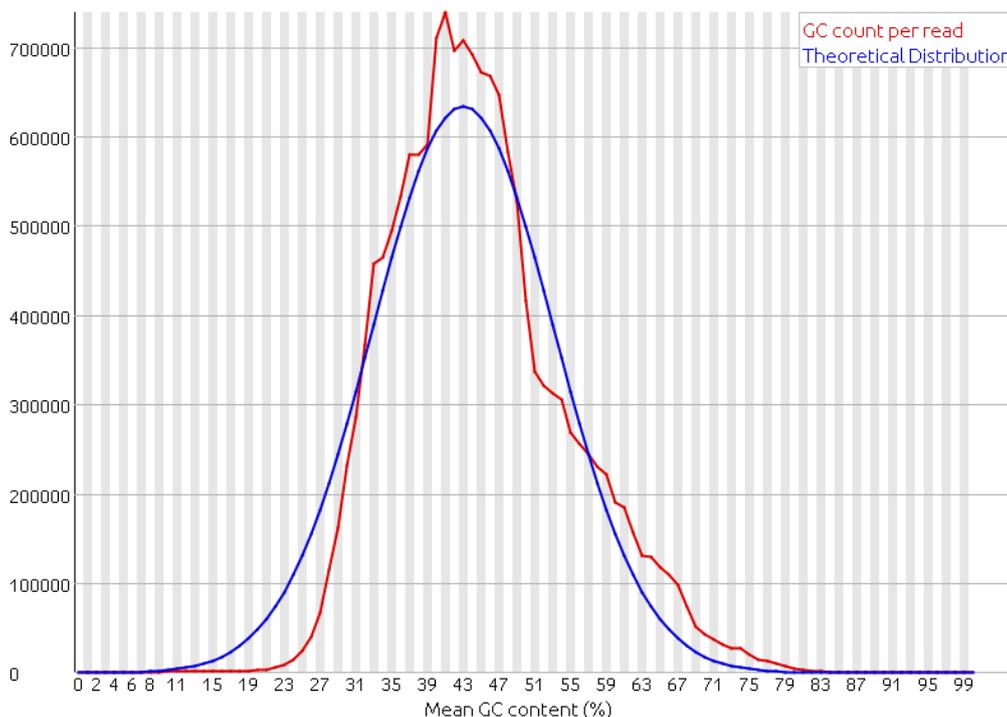
Fonte: A autora, 2018.

5. RESULTADOS

5.1 ANÁLISE GERAL DO SEQUENCIAMENTO

O sequenciamento dos genes foi realizado na plataforma *Illumina MiSeq*, em colaboração com a *McGill University*. Os resultados obtidos para cada indivíduo foram analisados individualmente, a fim de verificar a qualidade geral do sequenciamento. O conteúdo GC obtido está dentro do esperado de 44% para esses *amplicons* (**Figura 14**). A análise da qualidade das *reads* identificou uma queda da qualidade no final do fragmento. A queda de qualidade no final das *reads* é um fator conhecido do processo de sequenciamento de *Illumina*, em particular em *reads* superiores a 150 bp (**Figura 15**). No entanto, a maioria das *reads* possui uma média de qualidade maior que Q30 na escala *Phred*, ou seja, possuem uma acurácia superior a 99.9% (**Figura 16**).

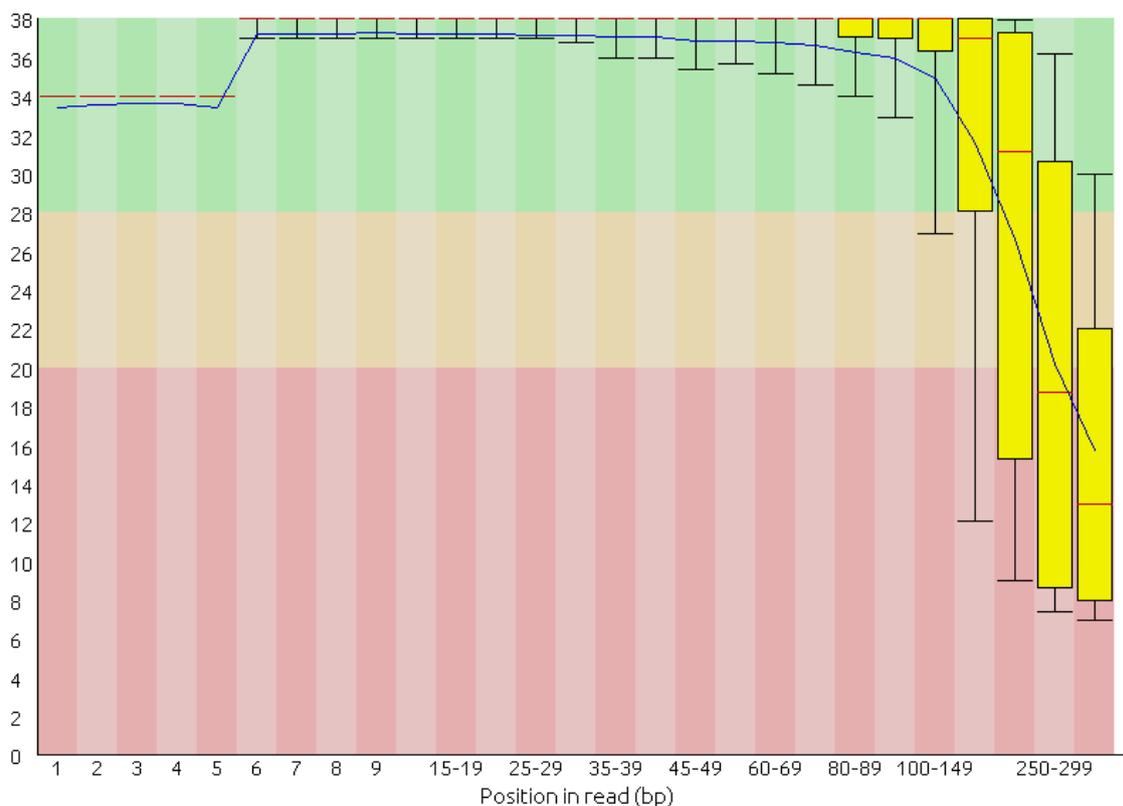
Figura 14. Histograma do conteúdo GC por *read* obtido para todas as amostras em conjunto.



Legenda: Histograma da média do conteúdo GC. O conteúdo obtido no sequenciamento foi semelhante à distribuição teórica esperada (azul). O eixo X mostra a média da porcentagem de conteúdo CG e o Y o número de *reads*. Em vermelho está representada a contagem CG obtida para todas as amostras em conjunto.

Fonte: A autora, 2016. Obtida no *software FastQC*.

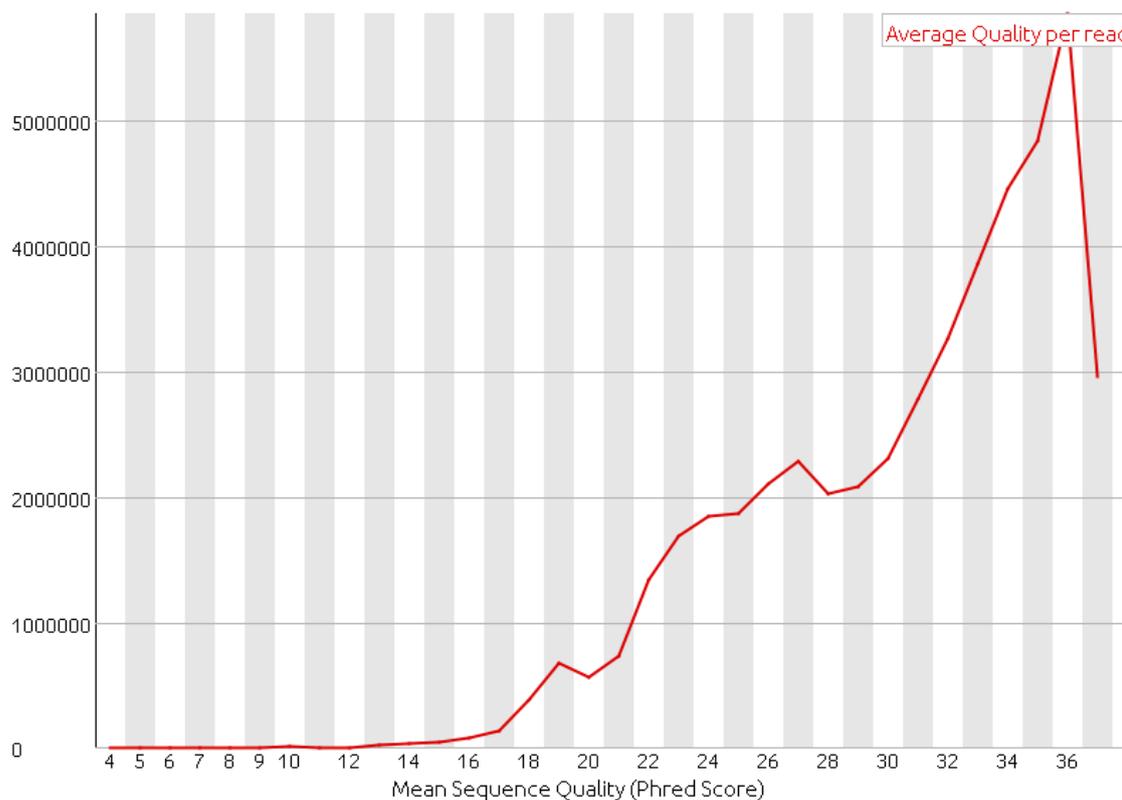
Figura 15. Scores de qualidade das bases obtidos para todas as amostras em conjunto, incluindo as *reads* dos dois extremos (r1 e r2).



Legenda: O eixo X mostra a posição em pb na *read* e o Y o *score* de qualidade. As linhas azuis e vermelhas representam a média e a mediana da qualidade. As caixas amarelas representam o intervalo inter-quartil de 25-75%. Os traços pretos marcam os limites de 10% e 90% desse índice. As cores de fundo gráfico dividem a qualidade em três grupos: boa qualidade (verde), qualidade aceitável (laranja) e baixa qualidade (vermelho).

Fonte: A autora, 2016. Obtida no *software FastQC*.

Figura 16. Distribuição de qualidade das *reads*.



Legenda: O eixo X mostra a escala *Phred* de qualidade e o eixo Y o número de *reads*. A linha vermelha representa a média de qualidade por *read*.

Fonte: A autora, 2016. Obtida no *software FastQC*.

Os dados do alinhamento das *reads* obtidas com o DNA de referência estão detalhados na **Tabela 2**. A cobertura média obtida por amostra, anteriormente às filtrações, foi de 211,96 para a *read1* e 169,71 para *read2*.

Tabela 2. Dados de alinhamento do sequenciamento.

Parâmetro	<i>Read1</i>	<i>Read2</i>
Cobertura média por amostra (pré filtração)	211,96 ±47,43	169,71 ±43,84
Alinhamento médio das <i>reads</i>	73,60%	58,49%

Total de bases alinhadas/amostra (\bar{x})	63.558.600	50.882.098
Bases com qualidade >Q30 (\bar{x})	79,65%	66,84%

Legenda: \bar{x} : média aritmética.

5.2 MÉDIA DE COBERTURA DOS EXONS

Para 18 dos 39 genes envolvidos nesse estudo, mais de 80% das bases sequenciadas em exons apresentaram média de cobertura na amostra populacional maior que 30X. Apenas para o gene *CCDC88B* não foi obtida uma média de cobertura maior que 30X em mais de 50% das bases exônicas sequenciadas (**Tabela 3**). A **Figura 17** mostra a média de cobertura por região para alguns exemplos de genes sequenciados; para construção da figura foram incluídos 150pb antes e depois de cada exon, a fim de facilitar a visualização dos mesmos.

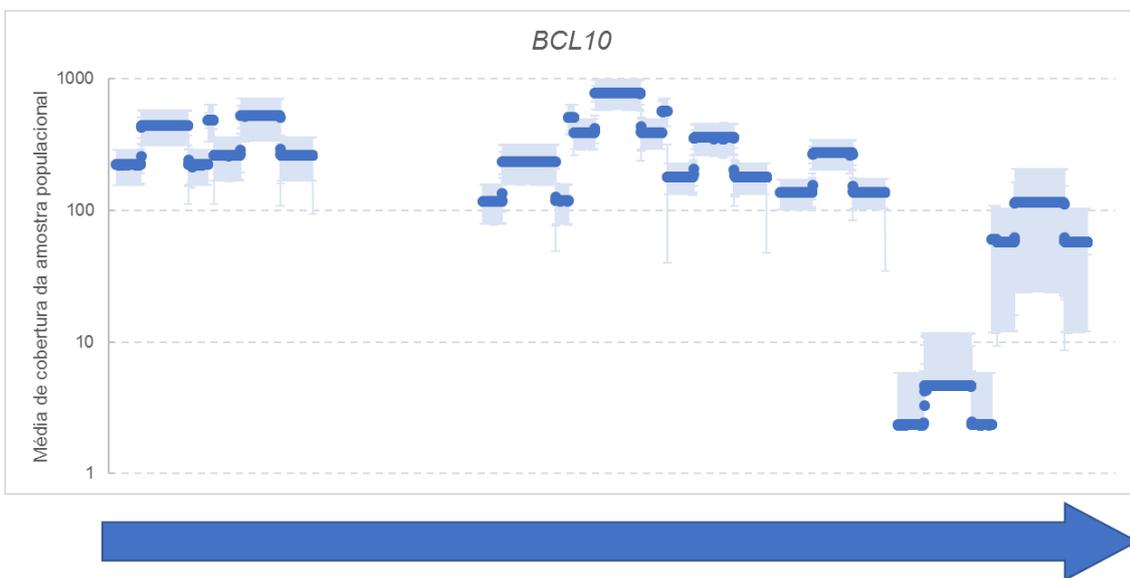
Tabela 3. Porcentagem de bases sequenciadas (em exons) com uma média de cobertura superior a 30 vezes, por gene.

Gene	% bases > 30X
<i>ADO</i>	64,25
<i>BATF3</i>	68,07
<i>BCL10</i>	61,72
<i>CCDC122</i>	65,33
<i>CCDC88B</i>	45,01
<i>CDH18</i>	79,30
<i>COX4I1</i>	78,00
<i>CUBN</i>	86,02
<i>DEC1</i>	100,00
<i>EGR2</i>	95,08
<i>FCN2</i>	53,48
<i>GATA3</i>	64,49
<i>IFNG</i>	100,00
<i>IL10</i>	60,86
<i>IL12RB2</i>	82,55
<i>IL17F</i>	100,00

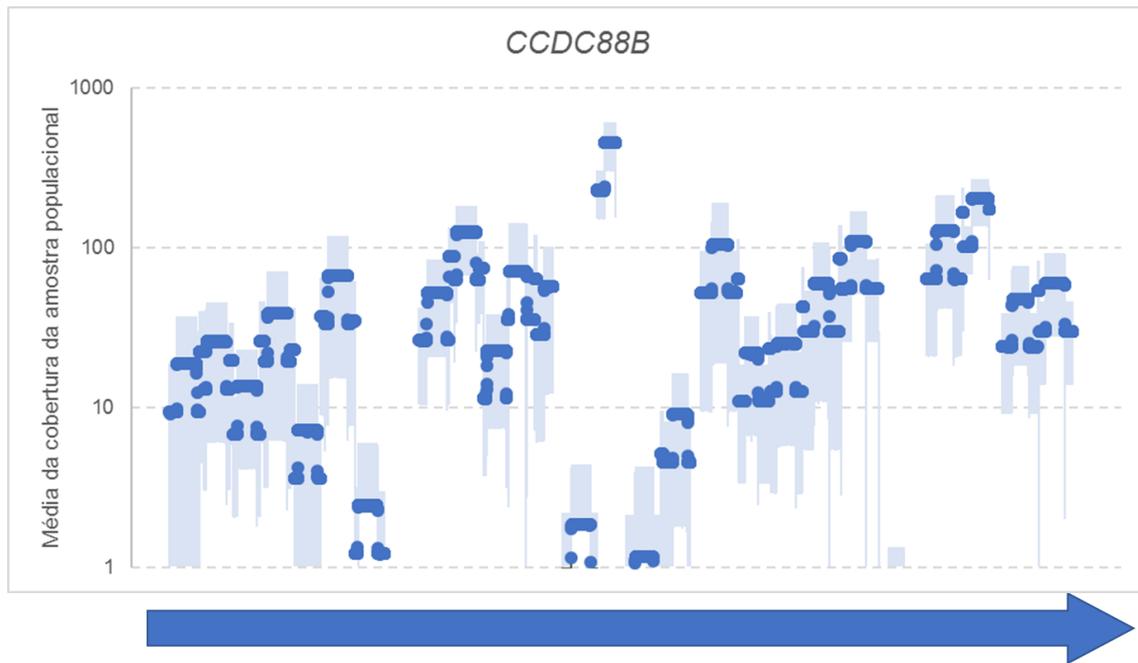
<i>IL18R1</i>	89,74
<i>IL18RAP</i>	79,09
<i>IL1RL1</i>	63,02
<i>IL23R</i>	78,66
<i>LACC1</i>	89,35
<i>LTA</i>	81,84
<i>LTA4H</i>	82,64
<i>MBL2</i>	100,00
<i>MICA</i>	82,97
<i>MICB</i>	68,89
<i>MRC1</i>	66,82
<i>NEBL</i>	94,76
<i>NOD2</i>	82,79
<i>SLC11A1</i>	57,88
<i>PACRG</i>	83,68
<i>PRKN</i>	72,24
<i>RIPK2</i>	88,92
<i>RMI2</i>	76,98
<i>SOD2</i>	64,81
<i>TLR1</i>	100,00
<i>TLR4</i>	95,02
<i>TNF</i>	72,49
<i>VDR</i>	62,98

Figura 17. Média de cobertura da amostra populacional de alguns dos genes sequenciado.

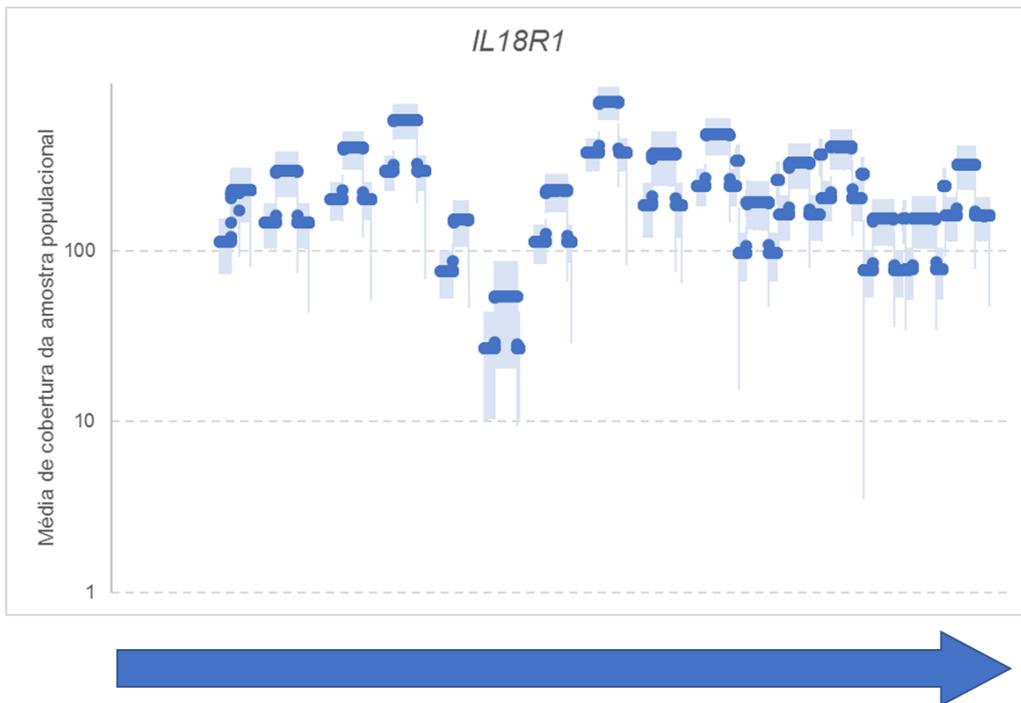
A) 3 exons do gene *BCL10*



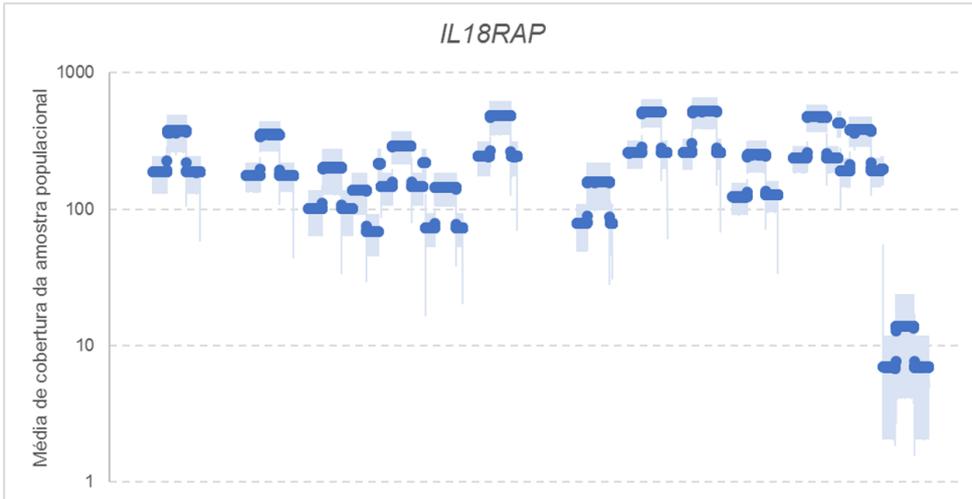
B) 27 exons do gene *CDC88B*



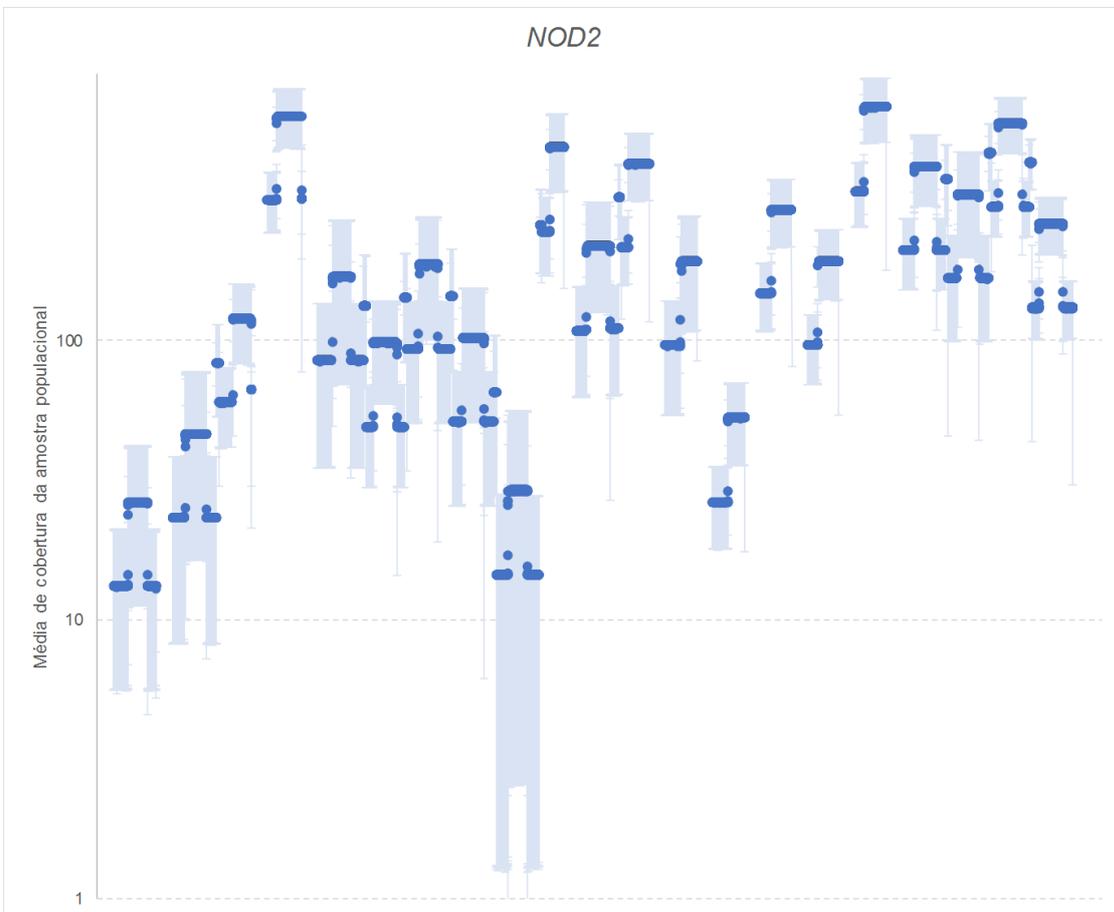
C) 12 exons do gene *IL18R1*



D) 12 exons do gene *IL18RAP*



E) 12 exons do gene *NOD2*



Legenda: O eixo Y indica a média de cobertura na amostra populacional, na escala logarítmica

na base 10; Cada ponto azul é a média de cobertura de uma base do gene, a barra azul clara mostra o desvio padrão da média e a flecha o sentido de transcrição do gene. Foram incluídos 150pb antes e depois de cada exon, que não são regiões alvos, a fim de facilitar a visualização dos mesmos.

5.3 IDENTIFICAÇÃO DE VARIANTES

Do total de variantes, 17 foram excluídas por estarem fora do equilíbrio de Hardy-Weinberg. Após as filtrações, foram identificadas 2003 variantes independentemente de sua frequência; dessas, 1066 são novas considerando a base de dados do *National Center for Biotechnology Information* dbSNP, Build 137. A distribuição do número de variantes por gene, encontra-se na **Tabela 4**. De interesse para este estudo, das 2003 variantes, 446 são em regiões UTR e 506 são em regiões codificantes; dentre as exônicas, 315 são não-sinônimas, 20 são *frameshift* e 21 são *stopgain*.

Tabela 4. Número de variantes identificadas em cada um dos genes sequenciados

Gene	UTR	Splicing	Exonica	Intronica	Upstream	Downstream
<i>ADO</i>	20	0	0	0	0	0
<i>BATF3</i>	2	0	4	6	0	2
<i>BCL10</i>	8	0	2	2	0	0
<i>CCDC122</i>	9	1	4	16	0	0
<i>CCDC88B</i>	2	0	10	19	0	3
<i>CDH18</i>	6	0	22	47	0	0
<i>COX4I1</i>	0	1	4	6	0	0
<i>CUBN</i>	0	0	109	154	1	0
<i>DEC1</i>	15	0	3	18	0	0
<i>EGR2</i>	11	0	6	5	0	1
<i>FCN2</i>	4	0	5	17	0	2
<i>GATA3</i>	9	0	1	3	0	2
<i>IFNG</i>	5	0	2	5	1	0
<i>IL10</i>	5	0	9	15	2	1
<i>IL12RB2</i>	11	2	24	45	0	0
<i>IL17F</i>	2	1	7	5	5	1
<i>IL18R1</i>	18	1	12	28	0	1
<i>IL18RAP</i>	8	0	9	42	1	0
<i>IL1RL1</i>	30	1	20	43	0	4
<i>IL23R</i>	5	0	8	26	7	0

<i>LACC1</i>	29	1	9	7	0	1
<i>LTA</i>	7	0	4	5	0	3
<i>LTA4H</i>	2	0	10	34	0	1
<i>MBL2</i>	42	0	8	13	5	1
<i>MICA</i>	1	0	24	26	0	1
<i>MICB</i>	11	1	15	15	0	0
<i>MRC1</i>	3	0	20	135	0	0
<i>NEBL</i>	31	0	25	73	0	2
<i>NOD2</i>	8	0	24	37	0	1
<i>SLC11A1</i>	14	1	10	17	0	1
<i>PACRG</i>	4	0	9	10	0	0
<i>PRKN</i>	18	0	25	36	0	0
<i>RIPK2</i>	8	0	14	29	0	1
<i>RMI2</i>	8	0	1	1	0	4
<i>SOD2</i>	22	0	1	3	0	0
<i>TLR1</i>	6	0	28	11	3	0
<i>TLR4</i>	32	0	12	8	1	1
<i>TNF</i>	7	0	1	2	0	1
<i>VDR</i>	23	0	5	16	0	0
Total	446	10	506	980	26	35

Seis indivíduos foram excluídos da análise por possuírem alto número de variantes não sequenciadas (*missing*), sendo, portanto, considerados com baixa qualidade de sequenciamento. Além disso, cinco indivíduos tiveram os genótipos de pai ou mãe desconsiderados por possuir baixa relação de parentesco identificada na análise de *relatedness*.

Após a aplicação dos filtros de qualidade, foram identificadas 193 variantes em regiões exônicas (n= 95), *UTR* (n=97) e de *splicing* (n=1) com frequência alélica mínima superior à 0,04 (do inglês *minor allele frequency, MAF*), ou seja, com ao menos 8 famílias informativas. Do total de variantes exônicas, 62 são não-sinônimas. A distribuição do número de variantes identificadas por genes, se encontra na **Tabela 5**.

Tabela 5. Número de variantes identificadas com $MAF > 0,04$ em cada um dos genes sequenciados.

Gene	UTR	Splicing	Exônica		
			Não sinônima	Sinônima	Stop Códon
<i>ADO</i>	3	0	0	0	0
<i>BATF3</i>	0	0	0	0	0

<i>BCL10</i>	2	0	0	0	0
<i>CCDC122</i>	2	0	0	1	0
<i>CCDC88B</i>	2	0	0	0	0
<i>CDH18</i>	1	0	0	2	0
<i>COX4I1</i>	0	0	1	0	0
<i>CUBN</i>	0	0	15	11	0
<i>DEC1</i>	3	0	1	0	0
<i>EGR2</i>	2	0	0	1	0
<i>FCN2</i>	1	0	2	0	0
<i>GATA3</i>	3	0	0	0	0
<i>IFNG</i>	0	0	0	0	0
<i>IL10</i>	0	0	0	0	0
<i>IL12RB2</i>	2	0	0	4	0
<i>IL17F</i>	0	0	2	0	0
<i>IL18R1</i>	7	0	0	0	0
<i>IL18RAP</i>	5	0	0	0	0
<i>IL1RL1</i>	9	0	4	1	0
<i>IL23R</i>	1	0	2	0	0
<i>LACC1</i>	5	1	1	1	0
<i>LTA</i>	0	0	3	0	0
<i>LTA4H</i>	0	0	0	0	0
<i>MBL2</i>	9	0	1	0	0
<i>MICA</i>	1	0	10	1	0
<i>MICB</i>	1	0	5	0	1
<i>MRC1</i>	0	0	2	2	0
<i>NEBL</i>	9	0	4	1	0
<i>NOD2</i>	2	0	1	3	0
<i>SLC11A1</i>	3	0	1	0	0
<i>PACRG</i>	0	0	0	0	0
<i>PRKN</i>	1	0	2	2	0
<i>RIPK2</i>	1	0	1	0	0
<i>RMI2</i>	5	0	0	1	0
<i>SOD2</i>	4	0	0	0	0
<i>TLR1</i>	3	0	2	1	0
<i>TLR4</i>	6	0	2	0	0
<i>TNF</i>	0	0	0	0	0
<i>VDR</i>	4	0	0	0	0
Total	97	1	62	32	1

5.4. ANÁLISE DE ASSOCIAÇÃO

5.4.1 De variantes individualmente

O teste TDT revelou 16 variantes associadas ($P < 0,05$) com hanseníase *per se* (**Tabela 6**), distribuídas em 12 genes: um *SNP* associado nos genes *ADO* (rs224082), *CCDC88B* (rs542907), *DEC1* (rs2285316), *IL1RL1* (rs13001714), *IL12RB2* (rs10489627), *IL23R* (rs10889677), *LTA* (rs1041981), *MBL2* (rs1800451), *PRKN* (rs1801582) e *SOD2* (rs5746140); dois *SNPs* em *CUBN* (rs1801225 e rs1801224); e quatro no *NOD2* (rs2066842, rs2066843, rs3135499 e rs3135500). Identificou-se ainda tendência de associação entre hanseníase *per se* e variantes dos genes *BCL10* (rs1060846), *IL1RL1* (rs1041973) e *NOD2* (rs1861759). Seis desses *SNPs*, localizados nos genes *CUBN*, *IL1RL1*, *LTA*, *MBL2*, *NOD2* e *PRKN*, levam à mudança de aminoácido na proteína.

Estimativa de risco de doença, através do cálculo das *Odds ratio*, revelou que as variantes exercendo maior impacto estavam localizadas nos genes *NOD2*, *SOD2*, *IL1RL1* e *ADO*; no entanto para todas elas há uma grande amplitude do intervalo de confiança (**Tabela 6**).

Tabela 6. SNPs associados no TDT

Gene	Posição	P HWE	dbSNP	Alelo	Freq.Alélica	n°FI	Z	P-value	Tipo de variante/ Mudança de Aa	OR (95% IC)
<i>ADO</i>	64566258	0,4884	rs224082	C	0,781	10	2,000	0,0455	UTR3 / NA	5,53 (1,32-23,23)
<i>BCL10</i>	85732065	0,1487	rs1060846	G	0,390	27	1,891	0,058682	UTR3 / NA	2,69 (1,03-6,99)
<i>CCDC88B</i>	64124980	0,3861	rs542907	G	0,643	35	2,101	0,035649	UTR3/NA	1,43 (1,05 – 1,94)
<i>CUBN</i>	17113563	0,787	rs1801225	T	0,386	25	2,132	0,032988	exonica / NA	NS
	17147521	0,6051	rs1801224	G	0,472	26	2,245	0,024796	exonica/ P389T	3,33 (1,20 -9,26)
<i>DEC1</i>	118054182	0,4	rs2285316	G	0,375	31	2,343	0,01915	UTR5 / NA	3,04 (1,15-8,00)
<i>IL1RL1</i>	102955468	0,2798	rs1041973	A	0,251	20	1,897	0,05778	exonica/ A78E	2,73 (1,00-7,52)
	102960485	1	rs13001714	A	0,688	13	2,335	0,019537	UTR3/ NA	9,86 (1,19-81,63)
<i>IL12RB2</i>	67779676	0,07082	rs10489627	G	0,497	31	2,126	0,033474	UTR5 / NA	2,59 (1,03 - 6,50)
<i>IL23R</i>	67725120	0,5184	rs10889677	C	0,723	12	2,214	0,026857	UTR3 / NA	NS
<i>LTA</i>	31540784	1	rs1041981	C	0,606	16	1,969	0,0489	exonica / T60N	4,04 (1,11 -14,95)
<i>MBL2</i>	54531226	0,1228	rs1800451	T	0,0489	9	-2,333	0,019631	exonica/ G57E	NS
<i>NOD2</i>	50744624	0,6645	rs2066842	T	0,174	22	3,128	0,001762	exonica / P268S	7,14 (1,96 - 26,31)
	50745199	1	rs2066843	T	0,185	23	3,591	0,00033	exonica / NA	12,19 (2,64 - 55,55)
	50745583	0,7497	rs1861759	T	0,721	12	1,938	0,052632	exonica / NA	9,14 (0,96 - 86,96)
	50766127	0,2744	rs3135499	A	0,615	17	2,390	0,016827	UTR3 / NA	6,46 (1,34 - 31,21)
	50766886	0,4147	rs3135500	G	0,610	17	2,390	0,016827	UTR3 / NA	6,46 (1,34 - 31,21)
<i>PRKN</i>	161807855	0,1837	rs1801582	G	0,243	27	2,846	0,004427	exonica / V380L	2,22 (1,25-3,92)
<i>SOD2</i>	160102770	1	rs5746140	T	0,043	8	2,333	0,019631	UTR3 / NA	10,64 (1,23 - 90,90)

Legenda: P HWE, Valor de P do equilíbrio de Hardy-Weinberg; n°FI, número de famílias informativas; UTR, *untranslated region*; Aa, aminoácido; P, Proline; T, Threonine; A, Alanine; E, Glutamate; N, Asparagine; G, Glycine; S, Serine; V, Valine; L, Leucine; NA, não se aplica, OR, *Odds Ratio*; IC, Intervalo de confiança; NS, não significativo.

Dentre as variantes encontradas associadas com a hanseníase *per se* no presente estudo, seis são não-sinônimas. Dessas, duas – rs1801224 no gene *CUBN* e rs1800451 no gene *MBL2* – foram identificadas como deletérias pelo programa *SIFT*, e possivelmente deletéria e provavelmente deletéria, respectivamente no programa *PolyPhen-2*. A variante em *MBL2* é considerada deletéria também na análise via software *LRT*. Estas variantes em *CUBN* e *MBL2* possuem um score de *CADD* superior a 20.

Tabela 7. Impacto das variantes não sinônimas encontradas associadas com a hanseníase *per se* no presente estudo.

Gene	dbSNP	Mudança de Aa	SIFT	PolyPhen 2	LRT	MutationTaster	CADD
<i>CUBN</i>	rs1801224	P389T	D	PD	Ne	P	23,6
<i>IL1RL1</i>	rs1041973	A78E	T	B	Ne	P	11,89
<i>LTA</i>	rs1041981	T60N	T	B	Ne	P	13,46
<i>MBL2</i>	rs1800451	G57E	D	pD	D	P	27,3
<i>NOD2</i>	rs2066842	P268S	T	B	Ne	P	1
<i>PRKN</i>	rs1801582	V380L	T	B	Ne	P	ND

Legenda: Aa, Aminoácido; P, Prolina; T, Treonina; A, Alanina; E, Ácido glutâmico; N, Asparagina; G, Glicina; S, Serina; V, Valina; L, Leucina; D, Deletéria; T, Tolerável; B, benigna; pD, Provavelmente danosa, PD; Possivelmente danosa; Ne, Neutra; P, Polimorfismo; ND, não disponível.

5.4.2. Variantes raras

Foram identificadas 1553 variantes raras e de baixa frequência (MAF<0,03) (138) nos genes sequenciados; dessas, 1058 não haviam sido descritas nos bancos de dados públicos do *National Center for Biotechnology Information dbSNP Build 137*. Do total de variantes raras e de baixa frequência, 336 são em regiões UTR e 399 são exônicas; destas, 177 são *singleton*, ou seja, ocorrem em apenas uma família informativa. Das variantes *singleton*, 97 são compartilhadas entre o filho afetado e um de seus pais, e para 80, o alelo menos frequente não foi transmitido ao filho afetado. Dentre as variantes em região codificante, 243 são não-sinônimas, 20 são *frameshift* e 20 *stopgain*. A distribuição do número de variantes raras e de baixa frequência nos genes sequenciados encontra-se resumida na **Tabela 8**.

Tabela 8. Número de variantes raras/baixa frequência identificadas em cada um dos genes sequenciados

Gene	Total	Novas	Tipo			Singleton*
			UTR	Sinônimas	Não sinônimas	
<i>ADO</i>	17	14	17	0	0	3
<i>BATF3</i>	11	9	2	1	1	2
<i>BCL10</i>	8	4	5	0	0	2
<i>CCDC122</i>	23	15	7	0	3	2
<i>CCDC88B</i>	29	24	0	3	6	4
<i>CDH18</i>	66	42	5	8	9	5
<i>COX4I1</i>	10	3	0	1	1	2
<i>CUBN</i>	199	132	0	24	54	24
<i>DEC1</i>	29	15	12	2	0	7
<i>EGR2</i>	20	14	9	0	4	1
<i>FCN2</i>	17	11	3	1	1	4
<i>GATA3</i>	12	10	6	0	1	3
<i>IFNG</i>	13	9	5	1	1	1
<i>IL10</i>	29	25	4	0	7	4
<i>IL12RB2</i>	66	50	9	4	11	6
<i>IL17F</i>	17	11	2	1	3	1
<i>IL18R1</i>	45	34	9	4	6	5
<i>IL18RAP</i>	41	34	3	2	5	3
<i>IL1RL1</i>	71	46	21	2	12	10
<i>IL23R</i>	34	22	4	0	3	2
<i>LACC1</i>	39	23	24	3	4	7
<i>LTA</i>	13	9	7	0	0	1
<i>LTA4H</i>	40	25	2	5	4	4
<i>MBL2</i>	50	33	33	2	3	4
<i>MICA</i>	22	10	0	2	8	3
<i>MICB</i>	31	18	10	3	5	0
<i>MRC1</i>	131	110	3	7	8	8
<i>NEBL</i>	101	69	20	4	15	12
<i>NOD2</i>	59	36	6	5	12	5
<i>SLC11A1</i>	32	22	10	1	7	5
<i>PACRG</i>	22	16	4	4	4	2
<i>PRKN</i>	64	37	16	9	10	10
<i>RIPK2</i>	47	35	7	4	6	2
<i>RMI2</i>	6	4	3	0	0	0
<i>SOD2</i>	19	14	17	0	1	0
<i>TLR1</i>	30	15	2	5	18	7
<i>TLR4</i>	46	33	26	2	7	7
<i>TNF</i>	10	7	7	0	1	4
<i>VDR</i>	34	18	16	2	2	5
Total	1553	1058	336	112	243	177

Singleton: Variantes que aparecem em apenas um trio; * Considerando variantes de regiões UTR e exônicas

5.4.3 Análise *gene-wide*

5.4.3.1 Análise *gene-wide* de variantes raras e de baixa frequência

Quando se incluiu na análise apenas variantes não-sinônimas, foi observada associação estatisticamente significativa entre a hanseníase *per se* e os genes *CCDC122*, *IL1RL1* e *MBL2*. Para a análise incluindo apenas variantes sinônimas, evidência de associação foi detectada para o gene *MICB* (**Tabela 9**).

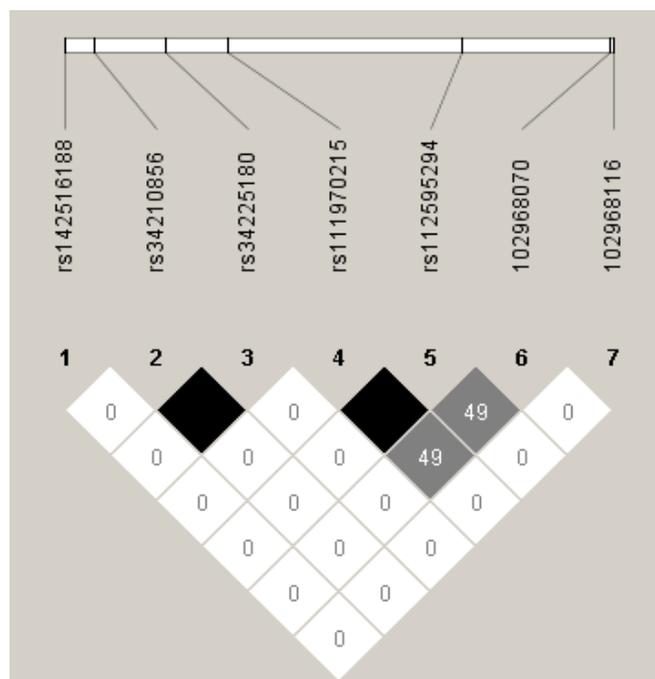
Tabela 9. Resultado da análise *Gene-wide* para variantes raras e de baixa frequência

Gene	Z	P	N	Tipo de variante incluída
<i>CCDC122</i>	-2,000	0,04	1	Não sinônimas
<i>IL1RL1</i>	-2,236	0,02	7	Não sinônimas
<i>MICB</i>	-2,000	0,04	2	Sinônimas
<i>MRC1</i>	-2,000	0,04	3	Não sinônimas

Legenda: N, Número de variantes consideradas na análise; Z, Teste Z, P, valor de P.

Dentre os sete marcadores envolvidos na análise *gene-wide* do gene *IL1RL1* (**Figura 18**) há perfeito *LD* entre os marcadores rs34210856 e rs4225180 e entre rs111970215 e rs112595249, além de um *LD* moderado ($r^2 = 0,49$) entre rs102968070 e rs111970215 e rs112595249.

Figura 18. Bloco de desequilíbrio de ligação entre as variantes não sinônimas consideradas na análise *gene-wide* do gene *IL1RL1*.



Legenda: os valores no interior dos losangos indicam os valores de r^2 em porcentagem e a intensidade de cores é maior quanto maior for esse valor. Sendo que, totalmente preto indica $r^2 = 1$.

Fonte: A autora, 2018. Obtido no programa *Haploview*.

5.4.3.2 Análise *gene-wide* concomitantemente de variantes raras/ baixa frequência e comuns

Quando a análise *gene-wide* foi realizada considerando tanto variantes raras/baixa frequência quanto variantes comuns (**Tabela 10**), associação – possivelmente causada por um acúmulo de variantes – foi novamente observada para *CCDC122* e *MBL2* quando se incluí na análise apenas variantes não-sinônimas; quando se inclui apenas as sinônimas, confirma-se a associação do gene *MICB*.

Tabela 10. Resultado da análise *Gene-wide* para variantes raras e comuns em conjunto

Gene	Z	P	N	Tipo de variante incluída
------	---	---	---	---------------------------

<i>CCDC122</i>	-2.000	0,04	1	Não sinônimas
<i>MBL2</i>	-1.971	0,05	4	Não sinônimas
<i>MICB</i>	-2.000	0,04	2	Sinônimas

Legenda: N, Número de variantes consideradas na análise; Z, Teste Z, P, valor de P.

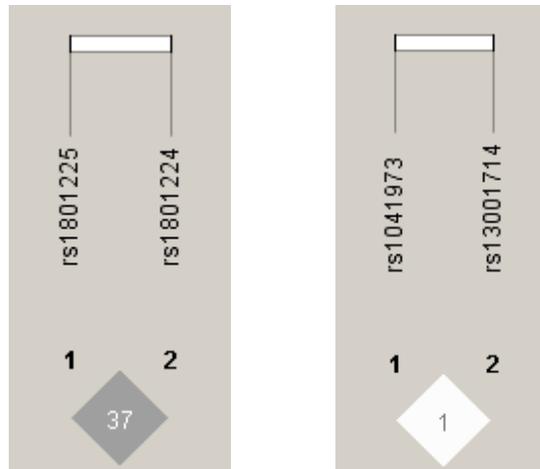
As variantes envolvidas na análise localizadas nos genes *CCDC122* e *MICB* possuem uma frequência alélica inferior a 0,03, portanto são as mesmas observadas na análise gene-wide para variantes raras/baixa frequência mostradas na **Tabela 9**. Já para o gene *MBL2*, há na análise duas variantes comuns e duas com frequência alélica inferior a 0,03. Todas as quatro variantes consideradas na análise desse gene são sinais independentes, ou seja, não estão em *LD* (dados não mostrados).

5.5 ANÁLISE DE DESEQUILÍBRIO DE LIGAÇÃO (*LD*)

5.5.1 Entre marcadores associados nesse estudo

A análise de desequilíbrio de ligação foi realizada quando mais de um marcador foi associado no mesmo gene. Para os genes *CUBN* e *IL1RL1*, não há *LD* entre os marcadores testados; portanto, os sinais de associação detectados são independentes (**Figura 19**).

Figura 19. Bloco de desequilíbrio de ligação entre os marcadores associados dos genes *CUBN* e *ILRL1*, respectivamente.

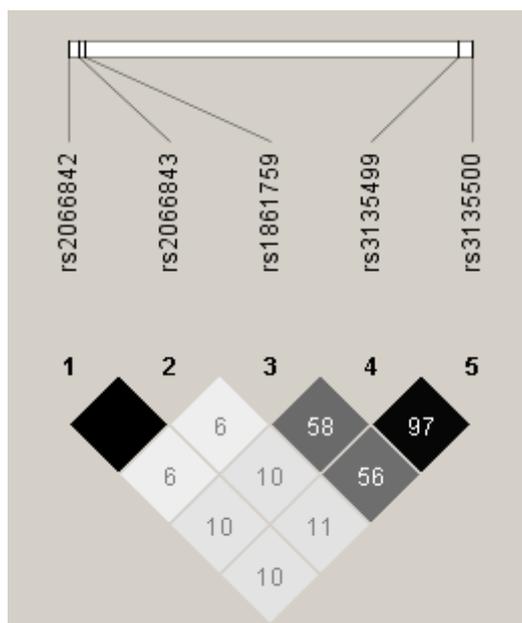


Legenda: os valores no interior dos losangos indicam os valores de r^2 em porcentagem e a intensidade de cores é maior quanto maior for esse valor.

Fonte: A autora, 2018. Obtido no programa *Haploview*.

Para o gene *NOD2* (**Figura 20**), dos cinco marcadores associados, dois – rs2066842 e rs2066843 – estão em *LD* perfeito ($r^2 = 1$) e outros dois – rs3135499 e rs3135500 – apresentam r^2 de 0,97. Ainda, o rs1861759 apresenta valores de r^2 de 0,58 e 0,56 com os marcadores rs3135499 e rs3135500, respectivamente. Análise de regressão logística multivariada revelou uma dependência de sinais entre os *SNPs* rs3135499 e rs1861759 (**Tabela 11**).

Figura 20. Bloco de desequilíbrio de ligação entre os marcadores associados do gene *NOD2*.



Legenda: os valores no interior dos losangos indicam os valores de r^2 em porcentagem e a intensidade de cores é maior quanto maior for esse valor. Sendo que, totalmente preto indica $r^2 = 1$.

Fonte: A autora, 2018. Obtido no programa *Haploview*.

Tabela 11. Resultado da análise de regressão logística multivariada no SAS

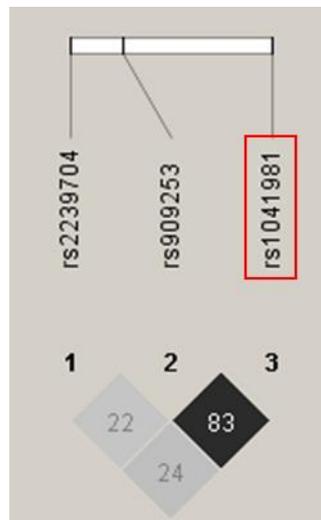
SNP	Valor de <i>P</i> - univariada	Valor de <i>P</i> – multivariada	
		rs3135499	rs1861759
rs2066843	0.0005	0,001	0,001
rs3135499	0.0126		0,069
rs1861759	0.0378	0,668	

5.5.2 Entre marcadores deste e de outros estudos realizados no Prata

Para os genes *LTA* e *PRKN* a análise foi realizada com 146 indivíduos distribuídos em 38 *pedigrees* e para o *NOD2* (87), a amostra foi composta de 152 indivíduos em 37 *pedigrees*. Os marcadores associados anteriormente no gene *LTA* (rs2239704 e rs909253; dados não publicados) possuem um valor de r^2 de 0,24 e 0,83 com o rs1041981, respectivamente (**Figura 21**). Dessa forma, há *LD* entre o marcador rs909253 associado anteriormente e o encontrado nesse estudo. No gene *PRKN*, não há *LD* entre os marcadores anteriormente associados com a hanseníase

nesse gene (rs10945859, rs2803073 e rs6930532; dados não publicados), nem em seu gene vizinho *PACRG* (rs1040079 e rs2276201) e o encontrado no presente estudo (rs1801582) (**Figura 22**).

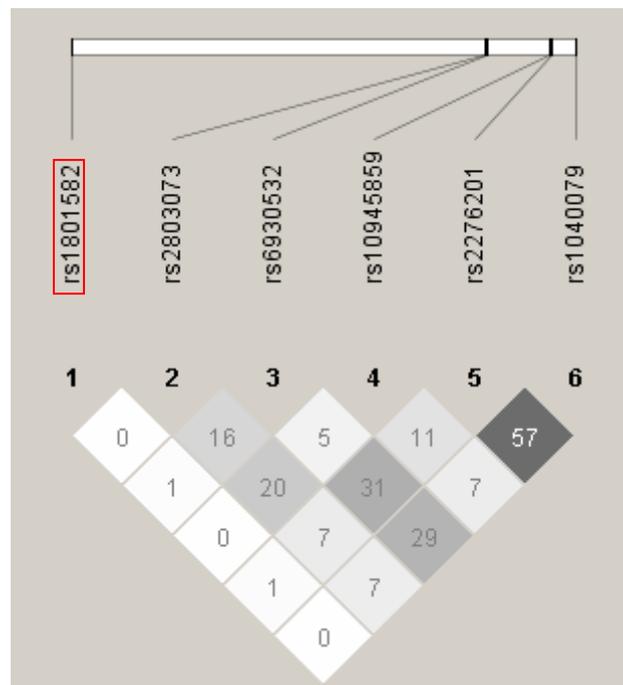
Figura 21. Bloco de desequilíbrio de ligação entre os marcadores associados nesse estudo e em estudo anterior no gene *LTA*.



Legenda: os valores no interior dos losangos indicam os valores de r^2 em porcentagem e a intensidade de cores é maior quanto maior for esse valor. Destacado em vermelho o marcador associado nesse estudo.

Fonte: A autora, 2018. Obtido no programa *Haploview*.

Figura 22. Bloco de desequilíbrio de ligação entre os marcadores associados nesse estudo e em estudo anterior nos genes *PRKN* e *PACRG*.

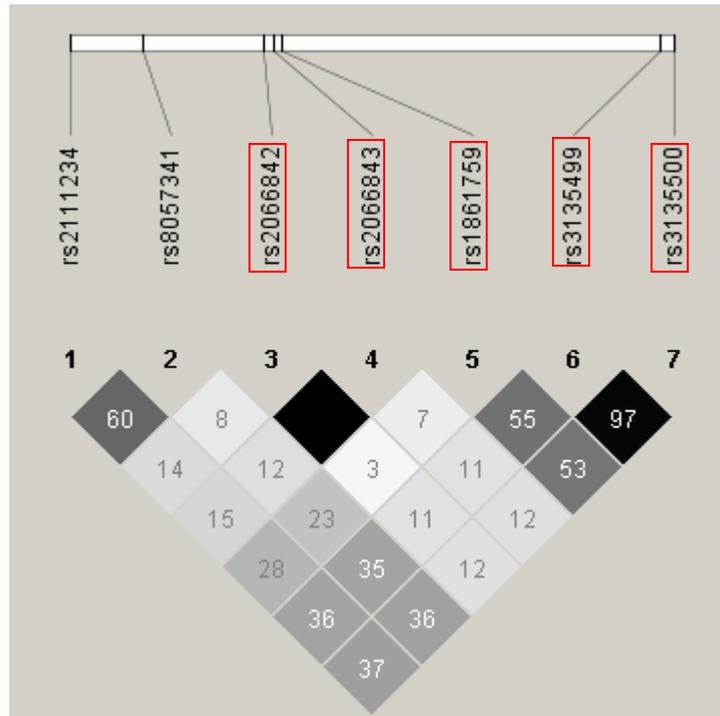


Legenda: os valores no interior dos losangos indicam os valores de r^2 em porcentagem e a intensidade de cores é maior quanto maior for esse valor. Destacado em vermelho o marcador associado neste estudo.

Fonte: A autora, 2018. Obtido no programa *Haploview*.

Para o gene *NOD2*, o marcador rs3135499 associado nesse estudo já havia sido descrito associado com hanseníase anteriormente (87). Este mesmo marcador está em fraco *LD* com outros dois, rs2111234 e rs8057341, associados com hanseníase em um estudo anterior de nosso grupo (87) ($r^2 = 0,36$ e $0,35$, respectivamente) (**Figura 23**)

Figura 23. Bloco de desequilíbrio de ligação entre os marcadores associados nesse estudo e em estudo anterior no gene *NOD2*.



Legenda: os valores no interior dos losangos indicam os valores de r^2 em porcentagem e a intensidade de cores é maior quanto maior for esse valor. Sendo que, totalmente preto indica $r^2 = 1$. Destacado em vermelho o marcador associado nesse estudo.

Fonte: A autora, 2018. Obtido no programa *Haploview*.

5.5.3 Entre marcadores deste e de outros estudos realizados em amostras populacionais distintas

A análise de *LD* entre marcadores associado nesse estudo e em outros já publicados está detalhada na **Tabela 12**. Um alto *LD* foi observado entre marcadores dos genes: *CCDC88B*, entre rs542907 e rs694739 ($r^2 = 1$ em africanos do Gâmbia e $r^2 = 0,91$ em caribenhos); *IL12RB2*, entre rs10489627 e rs3762315 ($r^2 = 1$ em chineses e $r^2 = 0,73$ em caucasianos de Utah) e entre rs10489627 e rs3762316 ($r^2 = 1$ em chineses e $r^2 = 0,73$ em caucasianos de Utah); *LTA*, entre rs1041981 e rs909253 ($r^2 = 1$ em caucasianos de Utah e africanos da Nigéria) e *NOD2* entre rs1861759 e rs7194886 ($r^2 = 0,87$ em ingleses e escoceses e $r^2 = 0,82$ em caucasianos de Utah), entre rs1861759 e rs3135499 ($r^2 = 0,97$ em ingleses e escoceses e $r^2 = 0,91$ em caucasianos de Utah), entre rs3135499 e rs7194886 ($r^2 =$

0,84 em ingleses e escoceses e $r^2 = 0,75$ em caucasianos de Utah), entre rs3135499 e rs13339578 ($r^2 = 1$ em japoneses e $r^2 = 0,83$ em africanos da Nigéria), entre rs3135499 e rs4785225 ($r^2 = 0,97$ em chineses e $r^2 = 0,81$ em africanos da Nigéria), entre rs3135499 e rs751271 ($r^2 = 1$ em japoneses e $r^2 = 0,81$ em africanos da Nigéria), entre rs3135499 e rs8057341 ($r^2 = 1$ em japoneses e $r^2 = 0,97$ em chineses), entre rs3135500 e rs7194886 ($r^2 = 0,80$ na população do Seri Lanca e escoceses e $r^2 = 0,74$ em caucasianos de Utah), entre rs3135500 e rs13339578 ($r^2 = 1$ em japoneses e $r^2 = 0,81$ em africanos da Nigéria), entre rs3135500 e rs4785225 ($r^2 = 0,97$ em chineses e $r^2 = 0,80$ em africanos da Nigéria), entre rs3135500 e rs751271 ($r^2 = 1$ em japoneses e $r^2 = 0,80$ em africanos da Nigéria), entre rs3135500 e rs8057341 ($r^2 = 1$ em japoneses e $r^2 = 0,97$ em chineses) e entre rs3135500 e rs3135499 ($r^2 = 1$ em chineses, $r^2 = 0,94$ em africanos da Nigéria e $r^2 = 0,91$ em caucasianos de Utah). O mesmo marcador encontrado no gene *MBL2* foi associado em um estudo anterior (118).

Tabela 12. Análise de LD entre variantes encontradas nesse estudo e em estudos anteriormente publicados.

Gene	SNPs desse estudo	SNPs estudos anteriores	Tipo de variantes do estudo anterior	Referência	r^2 (Ensemble hg19)
ADO	rs224082	rs58600253	Intergênica	Liu, J. 2015, Nat Gen	0,30 (KHV) 0,13 (CHB)
		rs233100	Intrônica	Liu, H. 2013 HMG	0,29 (FIN) 0,27 (CEU)
BCL10	rs1060846	rs2735591	Intrônica	Liu, H. 2013 HMG/ Liu, J. 2015, Nat Gen	0,61 (GIH) 0,33 (CHB) 0,37 (CEU) 0,18 (YRI)
		rs694739	Upstream	Liu, H. 2015, Nat Gen	1,0 (GWD) 0,91 (ACB) 0,74 (YRI) 0,52 (CEU)
CUBN	rs1801225	rs10904831 (MB)	Intrônica	Grant, A. 2014 Hum Gen	0,19 (FIN)
		rs1801241 (MB)	Sinônima	Grant, A. 2014 Hum Gen	0,19 (FIN)
	rs1801224	rs10904831 (MB)	Intrônica	Grant, A. 2014 Hum Gen	0,08 (PEL)
		rs1801241 (MB)	Sinônima	Grant, A. 2014 Hum Gen	0,12 (ASW) 0,06 (CEU)
DEC1	rs2285316	rs10817758	Intrônica	Liu, J. 2015, Nat Gen	0,09 (ITU)

<i>IL1RL1</i>	rs1041973	rs2058660 (<i>IL18RAP</i>)	Intrônica	Liu, H. 2012 AJHG	0,18 (PUR) 0,16 (CHB) 0,06 (CEU)
	rs13001714	rs2058660 (<i>IL18RAP</i>)	Intrônica	Liu, H. 2012 AJHG	0,71 (KHV) 0,56 (CDX) 0,17 (CEU)
<i>IL12RB2</i>	rs10489627	rs3762318	Intrônica	Zhang, F. 2011, Nat Gen.	0,13 (PEL) 0,05 (CHB)
		rs3762315 (LL)	<i>Upstream</i>	H Ohyama, 2005 J Clin Pathol	1 (CHS) 0,73 (CEU) 0,33 (YRI)
		rs3762316 (LL)	<i>Upstream</i>	H Ohyama, 2005 J Clin Pathol	1 (CHS) 0,73 (CEU) 0,33 (YRI)
		rs3762317 (LL)	<i>Upstream</i>	H Ohyama, 2005 J Clin Pathol	ND
		rs185002143	Não sinônima	Li GD.2016, J Dermatol Sci.	ND
		rs78198420	Não sinônima	Li GD.2016, J Dermatol Sci.	ND
<i>IL23R</i>	rs10889677	rs3762318	Intrônica	Zhang, F. 2011, Nat Gen./ Li GD.2016, J Dermatol Sci.	0,08 (FIN)
		rs2144658 (MB)	Intrônica	Li GD.2016, J Dermatol Sci.	0,12 (CLM) 0,057 (CEU)
		rs17129664 (PB)	Intrônica	Li GD.2016, J Dermatol Sci.	0,05 (ESN)
		rs6670134	Intrônica	Li GD.2016, J Dermatol Sci.	0,30 (MSL)
		rs539497366	Não sinônima	Li GD.2016, J Dermatol Sci.	ND
		rs199542433	Não sinônima	Li GD.2016, J Dermatol Sci.	ND
		rs201752419	Não sinônima	Li GD.2016, J Dermatol Sci.	ND
		rs76418789	Não sinônima	Liu H.2017, J Invest Dermatol.	0,05 (JPT)
<i>LTA</i>	rs1041981	rs2071590	<i>Upstream</i>	Alcaïs A.2007, Nat Genet	0,43 (ACB) 0,31 (KHV) 0,32 (CEU)
		rs2239704	UTR	Alcaïs A.2007, Nat Genet/ Angela	0,55 (JPT) 0,39 (KHV/YRI) 0,35 (CEU)
		rs746868	Intrônica	Alcaïs A.2007, Nat Genet	0,55 (JPT) 0,40 (KHV) 0,37(YRI) 0,35 (CEU) 0,28 (CDX)
		rs2844482	<i>Upstream</i>	Alcaïs A.2007, Nat Genet	0,16 (KHV/YRI) 0,09 (CEU)
		rs909253	Intrônica	Estudo do	1,00

			grupo	(CEU/YRI)	
MBL2	rs1800451	rs11003125 (MB)	Upstream	Zhang DF.2013, Hum Genet	0,06 (ESN)
		rs7100749	Upstream	Zhang DF.2013, Hum Genet	0,05 (PUR)
		rs11003124 (MB)	Upstream	Zhang DF.2013, Hum Genet	0,35 (ESN) 0,31 (YRI)
		rs7096206 (PB)	Upstream	Zhang DF.2013, Hum Genet	0,10 (MXL) 0,07 (YRI)
		rs1800451 (PB)	Não sinônima	Cardona- Pemberthy V.2018, Infect Genet Evol.	Mesmo marcador
NOD2	rs2066842	rs9302752	Upstream	Zhang FR.2009, N Engl J Med/ Grant AV.2012, J Infect Dis.	0,32 (ITU) 0,14 (CEU)
		rs7194886	Upstream	Zhang FR.2009, N Engl J Med	0,25 (GBR) 0,22 (CEU)
		rs12448797	Exônica não codificante	Berrington WR.2010, J Infect Dis	0,06 (ASW)
		rs2287195	Intrônica	Berrington WR.2010, J Infect Dis	0,28 (ITU) 0,14 (CEU)
		rs8044354	Upstream	Berrington WR.2010, J Infect Dis	0,21 (ITU) 0,14 (CEU)
		rs8043770	Upstream	Berrington WR.2010, J Infect Dis	0,33 (ITU) 0,14(CEU)
		rs13339578	Intrônica	Berrington WR.2010, J Infect Dis	0,29 (ITU) 0,19 (CEU)
		rs4785225	Splicing	Berrington WR.2010, J Infect Dis	0,31 (STU) 0,21 (CEU)
		rs751271	Intrônica	Berrington WR.2010, J Infect Dis	0,29 (STU) 0,21 (CEU)
		rs1477176	Intrônica	Berrington WR.2010, J Infect Dis	ND
		rs8057341	Intrônica	Sales- Marques C.2014, Hum Genet.	0,29 (ITU) 0,19 (CEU)
		rs2111234	Intrônica	Sales- Marques C.2014, Hum Genet.	0,24 (MXL) 0,18 (CEU)

	rs3135499	UTR	Sales- Marques C.2014, Hum Genet./ Xiong JH.2016,Int J Dermatol.	0,29 (CEU)
	rs9302752	<i>Upstream</i>	Zhang FR.2009, N Engl J Med/ Grant AV.2012, J Infect Dis.	0,32 (ITU) 0,14 (CEU)
	rs7194886	<i>Upstream</i>	Zhang FR.2009, N Engl J Med Berrington	0,25 (GBR) 0,22 (CEU)
	rs12448797	Exônica não codificante	WR.2010, J Infect Dis	0,06 (ASW)
	rs2287195	Intrônica	Berrington WR.2010, J Infect Dis	0,28 (ITU) 0,14 (CEU)
	rs8044354	<i>Upstream</i>	Berrington WR.2010, J Infect Dis	0,21 (ITU) 0,14 (CEU)
	rs8043770	<i>Upstream</i>	Berrington WR.2010, J Infect Dis	0,33 (ITU) 0,14(CEU)
rs2066843	rs13339578	Intrônica	Berrington WR.2010, J Infect Dis	0,29 (ITU) 0,19 (CEU)
	rs4785225	Splicing	Berrington WR.2010, J Infect Dis	0,31 (STU) 0,21 (CEU)
	rs751271	Intrônica	Berrington WR.2010, J Infect Dis	0,29 (STU) 0,21 (CEU)
	rs1477176	Intrônica	Berrington WR.2010, J Infect Dis	ND
	rs8057341	Intrônica	Sales- Marques C.2014, Hum Genet.	0,29 (ITU) 0,19 (CEU)
	rs2111234	Intrônica	Sales- Marques C.2014, Hum Genet.	0,24 (MXL) 0,18 (CEU)
	rs3135499	UTR	Sales- Marques C.2014, Hum Genet./ Xiong JH.2016,Int J Dermatol.	0,29 (CEU)
rs1861759	rs9302752	<i>Upstream</i>	Zhang FR.2009, N Engl J Med/ Grant AV.2012, J Infect Dis.	0,62 (PEL) 0,21 (CEU) 0,16 (CHB)

	rs7194886	<i>Upstream</i>	Zhang FR.2009, N Engl J Med	0,87 (GBR) 0,82 (CEU)
	rs12448797	Exônica não codificante	Berrington WR.2010, J Infect Dis	0,26 (JPT) 0,10 (CHS)
	rs2287195	Intrônica	Berrington WR.2010, J Infect Dis	0,60 (PEL) 0,21 (CEU)
	rs8044354	<i>Upstream</i>	Berrington WR.2010, J Infect Dis	0,58 (PEL) 0,21 (CEU)
	rs8043770	Upstream	Berrington WR.2010, J Infect Dis	0,51 (PEL) 0,21 (CEU)
	rs13339578	Intrônica	Berrington WR.2010, J Infect Dis	0,62 (JPT) 0,24 (CEU)
	rs4785225	Splicing	Berrington WR.2010, J Infect Dis	0,68 (PEL) 0,26 (CEU)
	rs751271	Intrônica	Berrington WR.2010, J Infect Dis	0,68 (PEL) 0,26 (CEU)
	rs1477176	Intrônica	Berrington WR.2010, J Infect Dis	ND
	rs8057341	Intrônica	Sales- Marques C.2014, Hum Genet.	0,62 (JPT) 0,24 (CEU)
	rs2111234	Intrônica	Sales- Marques C.2014, Hum Genet.	0,60 (PEL) 0,23 (CEU)
	rs3135499	UTR	Sales- Marques C.2014, Hum Genet./ Xiong JH.2016,Int J Dermatol.	0,97 (GBR) 0,91 (CEU) 0,14 (YRI)
	rs9302752	<i>Upstream</i>	Zhang FR.2009, N Engl J Med/ Grant AV.2012, J Infect Dis.	0,67 (PEL) 0,40 (CHB) 0,20(YRI)
rs3135499	rs7194886	<i>Upstream</i>	Zhang FR.2009, N Engl J Med	0,84 (GRB) 0,75 (CEU)
	rs12448797	Exônica não codificante	Berrington WR.2010, J Infect Dis	0,32 (JPT) 0,27 (CHB)
	rs2287195	Intrônica	Berrington WR.2010, J Infect Dis	0,65 (PEL) 0,14 (CEU)
	rs8044354	<i>Upstream</i>	Berrington WR.2010, J Infect Dis	0,63 (PEL) 0,13 (CEU)

	rs8043770	Upstream	Berrington WR.2010, J Infect Dis	0,50 (PEL) 0,14 (CEU)
	rs13339578	Intrônica	Berrington WR.2010, J Infect Dis	1,0 (JPT) 0,83 (YRI) 0,16 (CEU)
	rs4785225	Splicing	Berrington WR.2010, J Infect Dis	0,97 (CHB) 0,81 (YRI) 0,19 (CEU)
	rs751271	Intrônica	Berrington WR.2010, J Infect Dis	1,0 (JPT) 0,81 (YRI) 0,19 (CEU)
	rs1477176	Intrônica	Berrington WR.2010, J Infect Dis	ND
	rs8057341	Intrônica	Sales- Marques C.2014, Hum Genet.	1,0 (JPT) 0,97 (CHB) 0,28 (YRI) 0,16 (CEU)
	rs2111234	Intrônica	Sales- Marques C.2014, Hum Genet.	0,69 (CHB) 0,30 (YRI) 0,15 (CEU)
	rs3135499	UTR	Sales- Marques C.2014, Hum Genet./ Xiong JH.2016,Int J Dermatol.	Mesmo marcador
	rs9302752	Upstream	Zhang FR.2009, N Engl J Med/ Grant AV.2012, J Infect Dis.	0,64 (PEL) 0,40 (CHB) 0,23(YRI)
	rs7194886	Upstream	Zhang FR.2009, N Engl J Med	0,80 (STU) 0,74 (CEU)
	rs12448797	Exônica não codificante	Berrington WR.2010, J Infect Dis	0,32 (JPT) 0,27 (CHB)
rs3135500	rs2287195	Intrônica	Berrington WR.2010, J Infect Dis	0,67 (PEL) 0,11(CEU)
	rs8044354	Upstream	Berrington WR.2010, J Infect Dis	0,65 (PEL) 0,11 (CEU)
	rs8043770	Upstream	Berrington WR.2010, J Infect Dis	0,48 (PEL) 0,11 (CEU)
	rs13339578	Intrônica	Berrington WR.2010, J Infect Dis	1,0 (JPT) 0,81 (YRI) 0,14 (CEU)
	rs4785225	Splicing	Berrington WR.2010, J Infect Dis	0,97 (CHB) 0,80 (YRI) 0,16 (CEU)
	rs751271	Intrônica	Berrington WR.2010, J Infect Dis	1,0 (JPT) 0,80 (YRI) 0,16 (CEU)

	rs1477176	Intrônica	Berrington WR.2010, J Infect Dis	ND
	rs8057341	Intrônica	Sales- Marques C.2014, Hum Genet.	1,0 (JPT) 0,97 (CHB) 0,28 (YRI) 0,14 (CEU)
	rs2111234	Intrônica	Sales- Marques C.2014, Hum Genet.	0,69 (CHB) 0,31 (YRI) 0,13 (CEU)
	rs3135499	UTR	Sales- Marques C.2014, Hum Genet./ Xiong JH.2016,Int J Dermatol.	1,0 (CHB) 0,94 (YRI) 0,91 (CEU)
	rs10945859	Intrônica	Mira MT.2004, Nature/Chopra R.2013, PLoS Genet.	0,08 (MXL)
	rs2276201	Intrônica	Mira MT.2004, Nature	0,05 (PEL)
	rs9356058	Intrônica	Mira MT.2004, Nature./Angela	0,10 (GIH)
	rs9347684	Intrônica	Mira MT.2004, Nature/Chopra R.2013, PLoS Genet.	0,08 (MXL)
	rs1931223	Intrônica	Mira MT.2004, Nature	0,10 (GIH)
	rs1514343	Intrônica	Mira MT.2004, Nature	0,12 (PEL)
	rs1333955	Intrônica	Mira MT.2004, Nature/Alter A.2013,Hum Genet.	0,07 (MSL)
<i>PRKN</i>	rs1801582			
	rs1040079	Intrônica	Mira MT.2004, Nature/Angela	0,11 (PEL)
	rs4495257	Intrônica	Mira MT.2004, Nature	0,12 (PEL)
	rs9347683	UTR	Chopra R.2013, PLoS Genet.	0,08 (MXL)
	rs9346929	Intrônica	Chopra R.2013, PLoS Genet.	0,08 (MXL)
	rs4709648	Intrônica	Chopra R.2013, PLoS Genet.	0,10 (MXL)
	rs12215676	Intrônica	Chopra R.2013, PLoS Genet.	0,10 (MXL)
	rs10806765	Intrônica	Chopra R.2013, PLoS Genet.	0,08 (MXL)
	rs6936373	Intrônica	Chopra R.2013, PLoS Genet.	0,10 (MXL)

rs1333957	Intrônica	Chopra R.2013, PLoS Genet.	0,08 (MXL)
rs9365492	Intrônica	Chopra R.2013, PLoS Genet.	0,08 (MXL)
rs9355403	Intrônica	Chopra R.2013, PLoS Genet.	0,08 (MXL)
rs10806768	Intrônica	Alter A.2013, Hum Genet.	0,07 (MSL)
rs6915128	Intrônica	Alter A.2013, Hum Genet.	0,07 (MSL)
rs2849545	Intrônica	Alter A.2013, Hum Genet.	0,05 (KHV)
rs13219524	Intrônica	Alter A.2013, Hum Genet.	0,06 (MXL)
rs9347669	Intrônica	Alter A.2013, Hum Genet.	0,05 (YRI)
rs2846556	Intrônica	Alter A.2013, Hum Genet.	0,06 (ASW)
rs9458610	Intrônica	Alter A.2013, Hum Genet.	0,06 (MXL)
rs9365469	Intrônica	Alter A.2013, Hum Genet.	0,05 (ITU)
rs9456798	Intrônica	Alter A.2013, Hum Genet.	0,05 (ITU)
rs9456796	Intrônica	Alter A.2013, Hum Genet.	0,05 (ITU)
rs2155510	Intrônica	Alter A.2013, Hum Genet.	0,05 (ITU)
rs9458609	Intrônica	Alter A.2013, Hum Genet.	0,06 (MXL)
rs9356050	Intrônica	Alter A.2013, Hum Genet.	0,05 (ITU)
rs4709640	Intrônica	Alter A.2013, Hum Genet./Angela	0,06 (MXL)
rs10945846	Exônica não codificante	Alter A.2013, Hum Genet.	0,06 (MXL)
rs9295193	Intrônica	Alter A.2013, Hum Genet.	0,06 (BEB)
rs4636000	Intrônica	Alter A.2013, Hum	0,05 (ITU)

	rs9365460	Intrônica	Genet. Alter A.2013,Hum Genet.	0,05 (BEB)
	rs9356038	Intrônica	Alter A.2013,Hum Genet.	0,06 (MXL)
	rs7755248	Intrônica	Alter A.2013,Hum Genet.	0,06 (MXL)
	rs2803073	Intrônica	Alter A.2013,Hum Genet./Angela	0,05 (MXL)
	rs2023004	Intrônica	Alter A.2013,Hum Genet.	0,07 (ASW)
	rs768150	Intrônica	Alter A.2013,Hum Genet.	ND
	rs10945859	Intrônica	Estudo do grupo	0,08 (MXL)
	rs6930532	Intrônica	Estudo do grupo	0,08 (PUR)
SOD2	rs295340	<i>Upstream</i>	Ramos GB. 2016, J Infect Dis	0,06 (PEL)
	rs4880	Não sinônima	Ramos GB. 2016, J Infect Dis	0,17 (PEL)

Legenda: db, database; LL, hanseníase lepromatosa; MB, multibacilar; PB, paucibacilar; ACB, African Caribbeans in Barbados; ASW, Americans of African Ancestry in SW USA; BEB, Bengali from Bangladesh; CDX, Chinese Dai in Xishuangbanna, China; CEU, Utah Residents (CEPH) with Northern and Western European Ancestry; CHB, Han Chinese in Beijing, China; CHS, Southern Han Chinese; CLM, Colombians from Medellin, Colombia; ESN, Esan in Nigeria; FIN, Finnish in Finland; GBR, British in England and Scotland; GIH, Gujarati Indian from Houston, Texas; GWD, Gambian in Western Division, The Gambia; ITU, Indian Telugu from the UK; JPT, Japanese in Tokyo, Japan; KHV, Kinh in Ho Chi Minh City, Vietnam; MSL, Mende in Sierra Leone; MXL, Mexican Ancestry from Los Angeles USA; PEL, Peruvians from Lima, Peru; PUR, Puerto Ricans from Puerto Rico; STU, Sri Lankan Tamil from the UK; YRI, Yoruba in Ibadan, Nigeria; ND, não disponível.

Fonte: Ensembl (139).

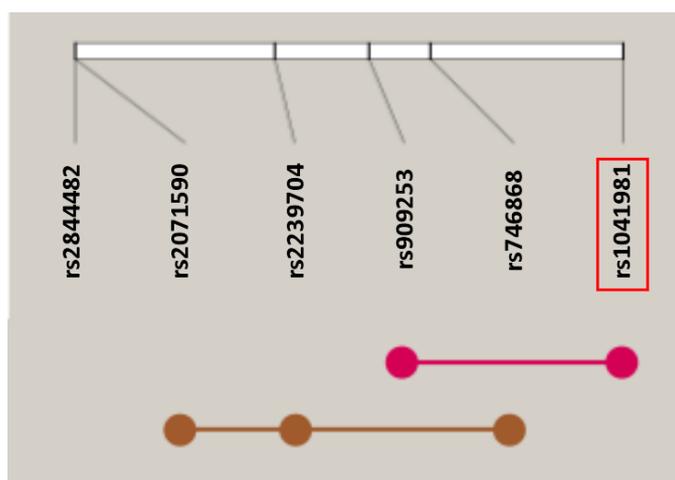
5.5.4 Entre todos os marcadores já associados em população de YRI da base de dados 1000G

Nos genes *ADO*, *BCL10*, *CUBN*, *DEC1*, *IL1RL1*, *IL23R*, *MBL2*, *PRKN* e *SOD2* não há *LD* – com um valor de $r^2 \geq 0,6$ – entre as variantes já associadas com a hanseníase, na população de YRI. No gene *IL12RB2* há um perfeito *LD* entre os marcadores rs3762315 e rs3762316 associados com o polo lepromatoso da doença por Ohyama e colaboradores (148). No gene *CCDC88B* há um elevado *LD* ($r^2 = 0,74$) entre as variantes rs542907, identificada nesse estudo, e rs694739 (80).

5.5.4.1 Gene *LTA*

A variante rs1041981 do gene *LTA* associada com a susceptibilidade do hospedeiro à hanseníase *per se* no presente estudo está em *LD* com a variante rs909253 previamente associada na amostra populacional do Prata (dados não publicados).

Figura 24. Bins das variantes já associadas com a susceptibilidade do hospedeiro à hanseníase no gene *LTA* utilizando dados do 1000G da população de YRI.



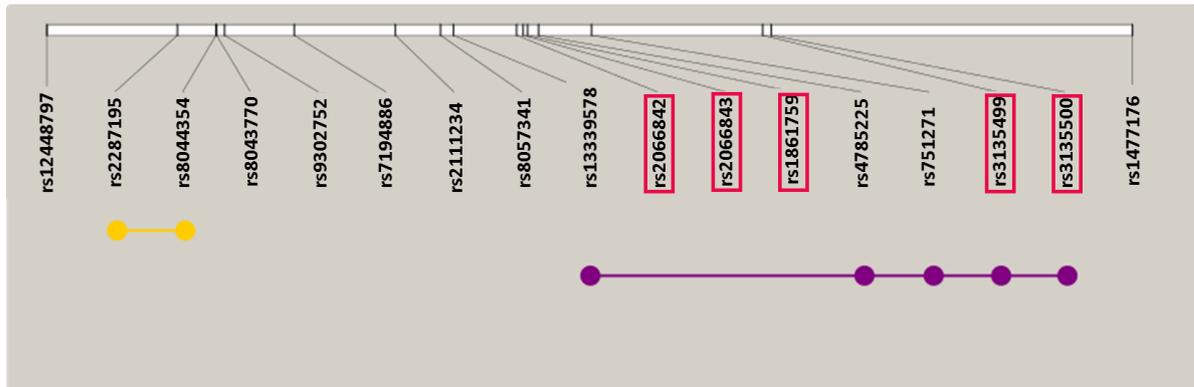
Legenda: Destacado em vermelho os marcadores associados neste estudo.

Fonte: A autora, 2018. Obtido nos programas *Ferret* e *Haploview*.

5.5.4.2 Gene *NOD2*

Foi verificado que os marcadores rs3135499 e rs3135500 do gene *NOD2* associados nesse estudo estão em *LD* com outros marcadores anteriormente associados com a susceptibilidade do hospedeiro à hanseníase por Berrington, 2010 (84) (**Figura 25**). Esse *bin* foi observado quando se considera a amostra de YRI disponível no projeto mil genomas.

Figura 25. Bins das variantes já associadas com a susceptibilidade do hospedeiro à hanseníase no gene *NOD2* utilizando dados do 1000G da população de YRI.



Legenda: Destacado em vermelho os marcadores associados neste estudo.

Fonte: A autora, 2018. Obtido nos programas *Ferret* e *Haploview*.

5.6 ANÁLISE DE HAPLÓTIPOS

5.6.1 Por gene

Não foi encontrada associação entre hanseníase e haplótipos do gene *CUBN* (**Tabela 13**). Já para o gene *IL1RL1* há uma associação significativa com proteção do hospedeiro à hanseníase *per se* e o haplótipo #2, que contém o alelo de proteção dos dois marcadores associados no presente estudo (**Tabela 14**). Para o gene *NOD2*, o haplótipo #3 está significativamente associado com a susceptibilidade do hospedeiro à hanseníase *per se* (**Tabela 15**), e contém o alelo de risco dos dois marcadores independentemente associados com a susceptibilidade à doença.

Tabela 13. Haplótipo das variantes associadas com a susceptibilidade do hospedeiro à hanseníase no gene *CUBN*.

Haplótipo	Frequencia	Alelos		n°FI	Z	P
		rs1801225	rs1801224			
#1	0,450	C	G	26	1,178	0,2387
#2	0,338	T	T	22	0,172	0,8635
#3	0,176	C	T	22	-1,830	0,0672

Legenda: n°FI, número de famílias informativas. Em vermelho o alelo de risco

Tabela 14. Haplótipo das variantes associadas com a susceptibilidade do hospedeiro à hanseníase no gene *IL1RL1*.

Haplótipo	Frequencia	Alelos		n°FI	Z	P
		rs1041973	rs13001714			
#1	0,522	C	A	25	0,959	0,3377
#2	0,241	C	G	21	-2,050	0,0404
#3	0,161	A	A	20	1,603	0,1088
#4	0,076	A	G	14	-0,435	0,66384

Legenda: n°FI, número de famílias informativas. Em vermelho o alelo de risco

Tabela 15. Haplótipo das variantes associadas com a susceptibilidade do hospedeiro à hanseníase no gene *NOD2*.

Haplótipo	Frequencia	Alelos		n°FI	Z	P
		rs2066842	rs3135499			
#1	0,434	C	A	37	-0,775	0,4386
#2	0,412	C	C	29	-0,885	0,3763
#3	0,154	T	A	20	2,205	0,0275

Legenda: n°FI, número de famílias informativas. Em vermelho o alelo de risco

5.6.2 Por cromossomo

Para o cromossomo 1 (**Tabela 16**), foram incluídos *SNPs* dos genes *BCL10* (rs1060846), *IL12RB2* (rs10489627) e *IL23R* (rs10889677). Apenas um dos haplótipos foi associado com proteção à doença, contendo o alelo de risco do marcador rs10889677.

Tabela 16. Haplótipo das variantes associadas com a susceptibilidade do hospedeiro à hanseníase no cromossomo 1.

Haplótipo	Frequência	Alelos			n°FI	Z	P
		<i>BCL10</i> rs1060846	<i>IL12RB2</i> rs10489627	<i>IL23R</i> rs10889677			
#1	0,233	A	G	C	26	0,276	0,7826
#2	0,207	G	G	C	22	1,661	0,0968
#3	0,197	A	A	C	24	-2,622	0,0087
#4	0,179	A	A	A	22	0,241	0,8099
#5	0,107	G	A	C	16	0,592	0,5538

Legenda: n°FI, número de famílias informativas; Em vermelho o alelo de risco.

Para o cromossomo 6 (**Tabela 17**), foram incluídas variantes dos genes *LTA* (rs1041981), *PRKN* (rs1801582) e *SOD2* (rs5746140). Todos os haplótipos com frequência maior que 5% possuem o alelo G para o rs5746140. O haplótipo associado com proteção do hospedeiro à hanseníase *per se* (#2) não inclui os alelos de risco dos marcadores rs1041981 e rs1801582, e os dois associados com risco (#3 e #4) possuem o alelo de risco do gene *PRKN*.

Tabela 17. Haplótipo das variantes associadas com a susceptibilidade do hospedeiro à hanseníase no cromossomo 6.

Haplótipo	Frequência	Alelos			n°FI	Z	P
		<i>LTA</i> rs1041981	<i>PRKN</i> rs1801582	<i>SOD2</i> rs5746140			
#1	0,459	C	C	G	29	-0,660	0,5093
#2	0,281	A	C	G	22	-3,603	0,0003
#3	0,112	A	G	G	13	2,608	0,0091
#4	0,105	C	G	G	15	2,862	0,0042

Legenda: n°FI: número de famílias informativas. Em vermelho o alelo de risco.

No cromossomo 10, foram analisados marcadores dos genes *ADO* (rs224082), *CUBN* (rs1801225 e rs1801224) e *MBL2* (rs1800451). Nenhum dos haplótipos foi associado significativamente com a doença (**Tabela 18**).

Tabela 18. Haplótipo das variantes associadas com a susceptibilidade do hospedeiro à hanseníase no cromossomo 10.

Haplótipo	Frequência	Alelos				n°FI	Z	P
		<i>ADO</i> rs224082	<i>CUBN</i> rs1801225	<i>CUBN</i> rs1801224	<i>MBL2</i> rs1800451			
#1	0,312	C	C	G	C	21	1,670	0,094844
#2	0,292	C	T	T	C	18	0,704	0,481728
#3	0,122	C	C	T	C	16	-1,667	0,095581
#4	0,118	T	C	G	C	10	-0,153	0,878791

Legenda: n°FI, número de famílias informativas. Em vermelho o alelo de risco.

6. DISCUSSÃO

Nas últimas décadas, diversos estudos com desenhos distintos têm sido realizados a fim de se avançar no entendimento do mecanismo de controle genético de suscetibilidade à hanseníase. Apesar do amplo corpo de evidências e de uma série de genes já terem sido associados com a susceptibilidade do hospedeiro à doença, possíveis variantes funcionais/causais raramente são propostas. Isso se deve ao fato de que tipicamente os estudos utilizam marcadores do tipo *Tag SNP* que, apesar de possibilitarem a cobertura completa do gene em estudo, são incapazes de definir se a variante genotipada é funcional/causal; a hipótese mais aceita é que os *Tag SNP* associados estão em desequilíbrio de ligação com a verdadeira variante responsável pelo sinal de associação observado. Esta hipótese é reforçada pelo fato de que com frequência as variantes associadas são comuns e não estão situadas em regiões codificantes. Com o objetivo de avançar na identificação de possíveis variantes causais/ funcionais, foi realizado neste estudo o sequenciamento e análise das regiões transcritas de 39 genes previamente associados com a susceptibilidade do hospedeiro à hanseníase, em uma amostra populacional composta por famílias da Vila Santo Antônio do Prata.

A amostra populacional da Vila Santo Antônio do Prata possui diversos fatores que a tornam uma oportunidade única para estudos genéticos: i) o enriquecimento de variantes de risco para a hanseníase, por se tratar de uma ex-colônia formada por indivíduos afetados por hanseníase e seus descendentes; ii) a homogeneidade dos fatores ambientais e iii) a provável alta exposição de seus habitantes ao *M. leprae*, por se tratar de uma população hiperendêmica e isolada (57, 149). Estudos de associação realizados com amostras populacionais da Vila do Prata corroboram esta hipótese (101). Ainda que seja uma população singular, uma limitação de estudos envolvendo a Vila do Prata é claramente o tamanho amostral pequeno, ainda mais se comparada com amostras populacionais comumente utilizadas em estudos genéticos de larga escala. Assim, deve-se sempre levar em conta o baixo poder de análise estatística em estudos envolvendo amostras do Prata, e a validação dos resultados observados em amostras populacionais independentes é sempre necessária.

Dos 39 genes analisados, foram encontrados sinais significativos de associação com a susceptibilidade do hospedeiro à hanseníase nos genes *ADO*, *BCL10*, *CCDC88B*, *CUBN*, *DEC1*, *IL1RL1*, *IL12RB2*, *I23R*, *LTA*, *MBL2*, *NOD2*, *PRKN* e *SOD2*. Para a maioria das associações observadas, há uma grande amplitude no intervalo de confiança da *odds ratio*, provavelmente devido ao reduzido número de famílias informativas envolvidas na análise. Dentre as variantes associadas, duas delas possuem evidências de impacto funcional, observadas nos genes *CUBN* e *MBL2*.

No gene *CUBN*, a variante não sinônima rs1801224 identificada nesse estudo leva a mudança do aminoácido hidrofóbico prolina por um polar, treonina, na posição 389 entre o quinto e o sexto domínio *EGF-like* da proteína. Além disso, essa variante foi encontrada como reguladora da expressão do gene *TRDMT1* (*tRNA aspartic acid methyltransferase 1*) em células dendríticas infectadas com *M. tuberculosis* (150). O gene *TRDMT1* codifica uma proteína com atividade de metiltransferase (151). A associação do gene *CUBN* com a hanseníase foi observada apenas entre pacientes com a forma MB da doença (75); no presente estudo, o gene *CUBN* foi associado com a hanseníase *per se*, sendo que, dos 116 indivíduos afetados nas famílias de nosso estudo, 104 se tem informação quanto a forma clínica da doença de acordo com a classificação da OMS, e desses 42 (40,38%) possuem a forma MB. No entanto, nenhuma das variantes identificadas anteriormente está em desequilíbrio de ligação com a associada nesse estudo.

O gene *MBL2* codifica uma proteína que é um importante elemento do sistema imune inato, responsável pelo reconhecimento de manose e N-acetilglucosamina de microrganismos e capaz de ativar a via das lectinas do complemento (152). A variante não-sinônima rs1800451 associada no presente estudo altera um aminoácido na posição 57 (de glicina, que é apolar, por um ácido glutâmico, polar ácido e de médio tamanho) considerada danosa, no domínio semelhante ao colágeno ("*Collagen-like*") da proteína; essa variante foi previamente associada com baixo nível da proteína no sangue (153) e com uma capacidade reduzida de ligação do ligante e portanto, falha na ativação do complemento (154). O *SNP* rs1800451 foi anteriormente associado com tuberculose em uma meta-análise envolvendo 22 estudos caso controle (155); e com a hanseníase PB em uma amostra populacional colombiana (118). Do total de hansenianos de que se tem

informação da forma clínica da doença em nosso estudo 59,61% são PB, o que corrobora com o achado anterior. Além disso a variante rs1800451 está moderadamente em *LD* ($r^2=0,35$) com o rs11003124 na população africana.

6.1 VARIANTES ASSOCIADAS - SINÔNIMAS OU EM REGIÃO *UTR*

Variantes associadas detectadas no presente estudo, ainda que sejam sinônimas ou de regiões *UTR*, mostraram possível impacto danoso na análise *in silico* ou exercem efeito sobre a expressão gênica (do inglês, *Expression quantitative trait loci*, *eQTL*). Esse efeito de *eQTL* também pôde ser visto em algumas variantes não sinônimas inicialmente consideradas benignas. Dessa forma, apesar de não serem *missense*, algumas das variantes associadas podem vir a apresentar impacto funcional.

A variante rs224082 do gene *ADO*, embora localizada região 3' *UTR*, apresentou um valor de CADD de 20,4; interessante, consta na base de dados do *HaploReg* que esta variante interfere na ligação da proteína *ADO* com a *GATA2*, importante fator regulatório da transcrição de genes envolvidos no desenvolvimento e proliferação de células hematopoiéticas. Além disso, interfere na expressão do gene *ADO* no sangue (156), tecido adiposo subcutâneo, músculo esquelético, testículo, pulmão e esôfago; e do gene *EGR2* em fibroblastos (157). Ainda, o *SNP* rs224082 está em *LD* ($r^2=0,88$ na população caucasiana) com o rs224090, associado com doença de Crohn (158). O *locus ADO-EGR2* foi também associado à Doença de Behçet (159), uma condição autoimune que parece poder ser desencadeada e/ou exacerbada por infecção (160). Assim, o gene *ADO* não tem sido associado apenas com a hanseníase, mas também com outras doenças com importante envolvimento do sistema imune e que podem ser causadas/desencadeadas pelo contato do hospedeiro com microrganismos. O rs58600253 anteriormente associado com a hanseníase por Liu *et al* (80) está localizado entre os genes *ADO* e *ALDH7A1P4* e não está em *LD* com a variante detectada nesse estudo.

O gene *BCL10* localizado na região 1p22.3 codifica uma proteína que contém um domínio de recrutamento de caspase (CARD) e demonstrou ativar o NF-kappaB (fator nuclear kappa B) (161), que por sua vez desempenha um importante papel na regulação da resposta imune. A variante rs1060846 associada com hanseníase neste estudo se encontra na região 3'UTR do gene, portanto não apresenta impacto funcional evidente. No entanto, a ferramenta de previsão *in silico* HaploReg v.4.1 mostrou que essa está em LD ($r^2=0,83$ na população caucasiana) com a variante rs11161570 do gene *C1orf52* (*chromosome 1 open reading frame 52*) que foi identificada como eQTL de *BCL10* no sangue (162) e em células dendríticas infectadas com *M. tuberculosis* (150). Estudos anteriores de associação revelaram duas variantes (rs233100 e rs2735591) associadas com hanseníase que estavam em LD entre si no estudo original mas que não apresentam relação com a variante descoberta nesse estudo (79). No entanto, não foi obtida uma boa cobertura média no sequenciamento (**Figura 17 A**) nos dois últimos exons desse gene, dessa maneira ele não pôde ser completamente explorado.

O marcador rs10489627 do gene *IL12RB2* aqui encontrado associado com a susceptibilidade à hanseníase está localizado na região 5'UTR do gene; e está correlacionado com o nível de *IL12RB2* no sangue (156, 157, 162) e em fibroblastos (157). As duas variantes associadas anteriormente com a forma LL da doença, rs3762315 e rs3762316 (148) estão em LD entre si e também com o SNP rs10489627. Dos 57 pacientes deste estudo classificado pelo critério de Ridley e Jopling, apenas 10 (17,54%) possuem a forma LL da doença; portanto, nossa amostra parece ser diferente para este endofenótipo, quando comparada com a do estudo anterior (148); fica, portanto, a ser explorada a possibilidade destas variantes estarem contribuindo para susceptibilidade à doença *per se* ou serem modificadoras para a forma LL. O gene *IL12RB2*, localizado no braço curto do cromossomo 1 (1p31.3), possui 89.537 pares de nucleotídeos, codifica uma proteína transmembrana do tipo 1 identificada como uma subunidade do complexo de receptores da interleucina 12 (IL-12) e participa da diferenciação de células Th1. Esse gene também já foi associado com doença de Crohn (163).

A variante rs542907 do gene *CCDC88B* associada com hanseníase nesse estudo possui efeito de eQTL sob esse gene no cérebro, esôfago, cólon e na pele (157). Ainda, está em perfeito desequilíbrio de ligação com o SNP rs519763 que

exerce efeito na expressão deste gene em células dendríticas tratadas com *M. tuberculosis* (150). O *SNP* aqui associado está em *LD* com o associado anteriormente por Liu *et al* (rs694739) (80) em algumas populações de bases de dados (de Gâmbia, Japão, Barbado, Los Angeles, China, entre outras). No entanto, é difícil afirmar se na amostra populacional do Prata haveria a associação entre esses dois marcadores, devido à alta miscigenação da população brasileira. Este gene foi também associado com doença de Crohn (164). Apesar de termos identificado uma variante associada, a qualidade geral do sequenciamento do gene *CCDC88B* foi baixa em muitas regiões exônicas (**Figura 17 B**). Dessa forma, é possível que haja outras variantes associadas à hanseníase nesse gene, que não foram testadas.

No presente estudo, duas variantes (rs1041973 e rs13001714) independentes foram significativamente associadas com susceptibilidade à hanseníase *per se* no gene *IL1RL1*. O *SNP* rs13001714 é uma variante 3'*UTR* que possui efeito na expressão desse gene no pulmão e cérebro (157). Além disso, interfere na expressão de *IL18RAP* na pele, tecido mamário, esôfago (157) e em macrófagos desafiados com *Salmonella typhimurium* (*P value*: $1,2 \times 10^{-2}$) (165). Ainda, se encontra em *LD* com outras correlacionadas com a expressão de *IL18R1* no sangue (146). Possui ainda um moderado *LD*, em amostras populacionais do Vietnã e China, com o *SNP* rs2058660 anteriormente associado com hanseníase, e localizado no gene *IL18RAP* (93). O *SNP* rs2058660 foi também anteriormente associado com doença de Crohn (166). Por sua vez, o *SNP* rs1041973 leva a mudança do aminoácido apolar alanina pelo ácido glutâmico na posição 78 da proteína e é considerada tolerável/benigna. Contudo, influencia na regulação da expressão dos genes *IL18R1* e *IL18RAP* no sangue (156, 162) e de *IL18RAP* em macrófagos desafiados com *Salmonella typhimurium* (*P value*: $5,4 \times 10^{-3}$) (165). Os genes *IL18R1* e *IL18RAP* foram sequenciados no presente estudo, porém, entre os *SNPs* com *MAF* >0,04 nenhuma associação com a hanseníase foi observada, apesar da alta qualidade geral do sequenciamento, ainda que com algumas regiões de baixa cobertura (**Figura 17 C e D**). É razoável propor que a associação anteriormente encontrada entre os genes *IL18R1* e *IL18RAP* com a susceptibilidade do hospedeiro à hanseníase (93) pode ocorrer devido a regulação da expressão desses genes.

Localizado na região MHC de classe III, o gene *LTA* (6p21.33) codifica uma citocina membro da família de fatores de necrose tumoral produzida em linfócitos (167). O rs1041981, associado com a hanseníase nesse estudo, é uma variante não sinônima, que causa uma troca de treonina por asparagina, ambos polares e de tamanho médio, na posição do aminoácido 60 da proteína. Identificada como benigna nos programas *SIFT*, *PolyPhen*, *LRT*, *MutationTaster* e *CADD*. Contudo, interfere na expressão de *LTA* no cérebro (157) e em macrófagos desafiados com *Salmonella typhimurium* (*P value*: $1,2 \times 10^{-5}$) e *Listeria monocytogenes* (*P value*: $1,5 \times 10^{-2}$) (165). Foi identificado, que o *SNP* rs1041981 influencia a expressão de outros genes HLA em células de linhagem de linfoblastos, como *HLA-DQA1*, *HLA-DRB1*, *HLA-DRB2*, *HLA-DRB3*, *HLA-DRB4* e *HLA-DRB5* (168), e no sangue periférico (*TNF*) (156, 162). Outros marcadores que estão em *LD* com rs1041981 já foram associados com infarto do miocárdio, artrite psoriática e miastenia grave (144). Ainda, esse marcador está em *LD* (**Figura 24**) com o rs909253 já associado com a hanseníase na Vila do Prata em um estudo anterior do nosso grupo (resultados não publicados).

Cinco polimorfismos do gene *NOD2* classicamente associado com a hanseníase foram encontrados associados com a doença na nossa amostra populacional. A proteína *NOD2* é expressa em leucócitos do sangue periférico e participa do reconhecimento de bactérias intracelulares através do dipeptídeo muramil presente nas suas paredes celulares, desencadeando a resposta imune (169, 170). Apesar do dipeptídeo muramil da parede celular do *M. leprae* ter uma estrutura distinta, se comparado com de outras bactérias patogênicas, estudos funcionais demonstram que a infecção de monócitos por *M. leprae* induz a produção de interleucina 32 dependente de *NOD2*, bem como a diferenciação de monócitos em células dendríticas (171). Das cinco variantes associadas no Prata, dois sinais de associação são independentes; do *bin* composto por rs2066842 e rs2066843, destaca-se o *SNP* rs2066843 que apesar de sinônimo está correlacionado com a expressão do gene no sangue, esôfago e pele (156, 157, 162, 172). Do segundo *bin*, formado pelos marcadores rs3135499, rs31354500 e rs1861759, o primeiro já foi associado anteriormente com hanseníase (87), está correlacionado com a expressão de *NOD2* no sangue periférico, esôfago, nervo tibial e pele (156, 157); e altera a expressão de *SNX20* (*sorting nexin 20*) em células dendríticas tratadas com

M. tuberculosis (150) e em macrófagos desafiados com *Salmonella typhimurium* (*P* value: $8,2 \times 10^{-3}$) (165). Finalmente, o SNP rs1861759 já foi associado com a expressão do gene *NOD2* na pele, nervo tibial (157), monócitos do sangue periférico (172) e em macrófagos desafiados com *Salmonella typhimurium* (*P* value: $8,5 \times 10^{-3}$) (165).

A análise de dependência dos sinais de associação de variantes do gene *NOD2* já associadas com a susceptibilidade do hospedeiro à hanseníase e as variantes identificadas nesse estudo, mostrou que alguns marcadores associados na população nepalesa (rs13339578, rs4785225 e rs751271) (84) estão em *LD* com marcadores do presente estudo (rs2066843 e rs2066842), quando se considera a população de YRI (**Figura 24**). A população YRI é antiga e por isso possui uma estrutura de fragmentação mais fina de *bins* de *LD*. Contudo, este precisaria ser confirmado na amostra em estudo. Ainda, podemos perceber que os SNPs associados nesse estudo estão na segunda metade do gene, e quando verificamos a média de cobertura da população no gene *NOD2*, percebe-se que a qualidade do sequenciamento não foi homogênea ao longo do gene, com sua primeira metade apresentando menor qualidade (**Figura 17 E**), o que pode explicar o fato de nenhum marcador identificado neste estudo explicar algumas das associações anteriormente encontradas, envolvendo marcadores localizados na porção inicial do gene *NOD2*.

Nosso estudo identificou variantes para as quais efeitos deletérios, nos genes *CUBN* e *MBL2*, e de expressão, nos genes *ADO*, *BCL10*, *IL12RB2*, *CCDC88B*, *IL1RL1* e *NOD2*, foram previstos *in silico*, por conseguinte, com possíveis relações causais sobre susceptibilidade do hospedeiro à hanseníase. Além de identificar variantes com predição de impacto deletério e de expressão, diversas associações detectadas neste estudo, são compatíveis com a hipótese de que *LD* pode ser a causa das associações observadas em estudos de genes candidatos e *GWAS*: as variantes de região transcrita, rs542907 do gene *CCDC88B*, rs13001714 do gene *IL1RL1*, rs10489627 do gene *IL12RB2*, rs1041981 do gene *LTA*, rs2066843 e rs2066842 do gene *NOD2* estão todas em forte *LD* com outras – com frequência, *tag* SNPs – descritas em estudos anteriores. Porém, outros sinais positivos de associação envolvendo variantes de regiões transcritas não explicam totalmente os anteriormente identificados, que podem ser devido a diversos fatores, incluindo: i) *background* genético distinto entre a amostra populacional de descoberta de

associação com a hanseníase e a do presente estudo; ii) a associação ser devido a algum endofenótipo da hanseníase; iii) cobertura insatisfatória de algumas regiões sequenciadas dos genes no presente estudo e iv) a associação de variantes *tag SNP* ser a causal de fato, e não devido ao *LD* com uma variante de região exônica.

Contudo, apesar desses resultados serem estimulantes e promissores esses polimorfismos devem ter seus sinais de associação confirmados em uma amostra populacional maior, preferencialmente com *background* étnico/genético distinto. Além disso, os efeitos de variantes não-sinônimas identificadas podem ser testados por modelagem molecular de proteínas, para verificar, mesmo que *in silico*, o impacto desta alteração nos domínios proteicos em que se encontram e conseqüentemente na proteína. Os impactos deletérios dessas variantes e de *eQTL* identificados *in silico* ou em estudo envolvendo outras condições devem ser testadas através de estudos funcionais, como por exemplo i) de expressão em um ou mais tipos celulares específicos e ii) edição genômica por CRISPR-Cas9, a fim de verificar o efeito molecular das mesma, individualmente e em conjunto, já que se trata de uma doença complexa. A partir do reconhecimento de variantes causais/funcionais, pode ser possível a previsão do risco de desenvolvimento da hanseníase em pessoas em áreas de risco, com potencial impacto sobre os programas de controle da doença.

6.2 ANÁLISE DOS HAPLÓTIPOS

A análise de haplótipos formados por alelos independentes significativamente associados com a hanseníase e estão no mesmo gene mostrou, para os genes *IL1RL1* e *NOD2*, associação com proteção e susceptibilidade, respectivamente, quando os alelos associados individualmente com proteção e risco são transmitidos em conjunto. Para o gene *CUBN*, nenhum haplótipo foi significativamente associado.

Na análise haplotípica envolvendo variantes de genes localizados no cromossomo 6, observou-se que o alelo G do rs5746140 do gene *SOD2* estava presente em todos os haplótipos com uma frequência superior a 5% (**Tabela 17**). Isso ocorre provavelmente devido à baixa frequência (4,3%) do outro alelo na amostra populacional investigada. Todos os haplótipos de susceptibilidade do

cromossomo 6 têm o alelo de risco do marcador rs1801582 do gene *PRKN*, que apesar de ser uma troca de aminoácido considerada benigna para a estrutura da proteína, a presença deste alelo define a associação com a susceptibilidade do hospedeiro à hanseníase observada para estes haplótipos. A parquína, proteína codificada pelo gene *PRKN* – localizado na região 6q26 e composto por 1.380.245 nucleotídeos – é uma E3 ubiquitina ligase responsável pela degradação proteossomal (173). Ademais, possui papel na mitofagia, no *clearance* de bactérias intracelulares e na apresentação de antígeno mitocondrial (170). Além de classicamente estar associado à doença de Parkinson (174), o gene *PRKN* já foi associado a infecções por patógenos intracelulares como *M. leprae* (64, 70) e *Salmonella enterica sorovar Typhi* (175). Em 2013, Manzanillo *et al.* realizaram um estudo funcional *in vitro* com o *Mycobacterium tuberculosis* e outros patógenos intracelulares que comprovam que a atividade ubiquitina ligase da parquína é essencial para o controle de patógenos bacterianos intracelulares (176).

No haplótipo associado com a proteção à hanseníase no cromossomo 1 (**Tabela 16**) está incluso o alelo de risco do *SNP* rs10889677 do gene *IL23R*. No entanto, apesar desse marcador ter sido associado no teste de desequilíbrio de transmissão no *FBAT*, quando testada sua associação no SAS o valor de *P* não foi significativo ($P = 0,0757$). Assim, novos experimentos devem ser realizados para confirmar se a associação desse *SNP* com a hanseníase é verdadeira ou se ocorreu de forma espúria.

6.3 ANÁLISE GENE-WIDE

A associação do gene *MRC1* (**Tabela 9**) com a susceptibilidade do hospedeiro à hanseníase *per se*, se dá no presente estudo devido a um acúmulo de variantes raras (todas elas com um *MAF* de 0,0041) não sinônimas, que não estão em *LD* entre si (Chr10:18141995, rs141126876 e Chr10:18170020). Essas três variantes, contribuem igualmente na análise, devido ao fato de possuírem a mesma frequência alélica, contudo o rs141126876 foi considerado causador da doença no *MutationTaster* e possui um CADD de 22.6. Sendo, portanto, uma variante rara com possível impacto causal na hanseníase. O gene *MRC1* codifica uma proteína

receptora de membrana que medeia a endocitose de glicoproteínas, de microrganismos com potencial patogênico, por macrófagos (177).

A análise *gene-wide*, deve ser olhada com cautela devido ao fato de que o número de variantes envolvidas em algumas das análises é reduzido, dessa maneira uma variante específica pode ser responsável pelo sinal de associação observado. Como no caso da análise para variantes não-sinônimas do gene *CCDC122*, tanto para variantes raras/baixa frequência individualmente quanto para variantes raras e comuns em conjunto, apenas uma variante (rs79782223) foi considerada pelo programa, sendo assim a associação ocorre devido a ela e não ao um acúmulo de polimorfismos. Esta variante não foi incluída na análise de associação por variantes devido ao fato de possuir frequência alélica mínima baixa, e conseqüentemente, um baixo número de famílias informativas (n=4). Trata-se de uma troca do aminoácido ácido glutâmico por uma lisina na posição 273 da proteína, essa troca é considerada deletéria no *SIFT* e possui um *score* no *CADD* de 23,5, sendo assim é possível que essa variante tenha um impacto causal na hanseníase. O gene *CCDC122* está localizado na região 13q14.11, possui 43.339 nucleotídeos e sua função ainda não é conhecida (178).

Na análise de variantes sinônimas do gene *MICB*, há apenas dois SNPs, rs45627734 com *MAF* de 0,011 e rs77208519 com *MAF* de 0,016, que não estão em desequilíbrio de ligação entre si ($r^2 = 0$), ambos presentes em duas famílias informativas. No entanto, pela característica das variantes, sinônimas e com *MAF* muito próximo, é difícil afirmar se há uma que contribui mais que a outra para a associação observada, sendo provável que ambas contribuam de forma praticamente igualitária para associação do *MICB* com a susceptibilidade do hospedeiro à hanseníase *per se*. Esse gene está localizado na região 6p21.33, possui 16.848pb e sua proteína codificada ativa a resposta citotóxica das células *natural killer* e células T (179).

Dos sete sinais considerados para análise de variantes raras não sinônimas do gene *IL1RL1*, pelo menos dois não são independentes (**Figura 18**). As frequências alélicas das sete variantes variam de 0,0041 a 0,0083, e em uma análise *gene-wise*, quanto menor a frequência alélica maior a contribuição da variante no resultado obtido (180). Dessa forma, é provável que determinadas variantes envolvidas na análise contribuam de forma majoritária para a associação

observada do gene *IL1RL1* com a susceptibilidade do hospedeiro à hanseníase *per se*. Essa hipótese é corroborada pelo fato de uma variante nova, localizada em Chr10:102968116, ter uma frequência alélica de 0,0041 e ser considerada danosa no *SIFT* e *Polyphen*, causador da doença no *MutationTaster* e possuir um CADD de 24,4; portanto, com um possível impacto causal na hanseníase. A proteína codificada por *IL1RL1* é um membro da família de receptores da interleucina 1 que pode ser induzido por estímulos pró-inflamatórios e estar envolvida na função das células T auxiliares (181).

Todas as variantes não sinônimas envolvidas na análise do gene *MBL2* são deletérias, independentemente da sua frequência, com valor de CADD superior a 20 e consideradas danosa no *Polyphen* e *SIFT*. Contudo, a variante nova localizada em chr10: 54528099 e o rs5030737 são consideradas causadoras da doença no *MutationTaster*. Ainda, essas variantes possuem uma frequência alélica de 0,005 e 0,003, respectivamente, portanto possuem um maior peso estatístico do que as demais (rs1800451 com frequência alélica de 0,048 e rs1800450 com frequência alélica de 0,18) no algoritmo utilizado pelo *FBAT* na análise *gene-wise* (180). Por conseguinte, é razoável propor que as variantes chr10: 54528099 e o rs5030737 contribuem predominantemente na associação observada entre o gene *MBL2* e a hanseníase *per se* nessa análise.

Apesar desses resultados serem instigantes, por se tratar de variantes raras e considerando que múltiplos testes foram realizados, mesmo que as associações sejam em genes conhecidamente associados à hanseníase, há a necessidade de checar a possibilidade de falsos positivos. Dessa maneira, assim como para as variantes comuns associadas, esses resultados devem ser testados em uma amostra populacional maior e independente.

7. CONSIDERAÇÕES FINAIS

O objetivo central desse estudo foi utilizar estratégia de sequenciamento de segunda geração para, primeiro, identificar variantes já conhecidas ou novas, independentemente de sua frequência alélica, localizadas em sequências exônicas

dos genes candidatos selecionados. Na sequência, essas variantes puderam ser testadas para associação, com o objetivo de avançar na identificação de variantes em regiões transcritas e prováveis funcionais/causais que possivelmente explicassem os achados anteriores – produzidos seguindo desenho clássico de estudo de associação utilizando marcadores comuns na população (de alta frequência).

Empregando essa estratégia, nosso estudo foi capaz de replicar/validar a associação entre a susceptibilidade genética do hospedeiro à hanseníase e os genes *ADO*, *BCL10*, *CCDC88B*, *CCDC122*, *CUBN*, *DEC1*, *IL1RL1*, *IL12RB2*, *I23R*, *LTA*, *MBL2*, *MICB*, *MRC1*, *NOD2*, *PRKN* e *SOD2*. Além disso, foram validadas as associações entre hanseníase *per se* e os SNPs rs1800451 do gene *MBL2* e rs1801582 do gene *PRKN*. A ausência de associação de genes que foram associados a susceptibilidade do hospedeiro à hanseníase em diferentes amostras populacionais, como por exemplo o *IL10*, *LACC1* e *TLR1*, pode ser em virtude de diversos fatores, entre eles destaca-se i) a cobertura inadequada de algumas regiões alvo; ii) o *background* genético/étnico distinto entre a amostra de descoberta e a do presente estudo; iii) a associação original pode envolver variantes de fato localizadas em regiões regulatórias não codificantes dos genes; e iv) a associação pode ser devido a algum endofenótipo da hanseníase.

Em seis dos genes associados no presente estudo, nós identificamos variantes não sinônimas: rs1801224 no *CUBN*, rs1041973 no *IL1RL1*, rs1041981 no *LTA*, rs1800451 no *MBL2*, rs2066842 no *NOD2* e rs1801582 no *PRKN*, além das variantes raras rs79782223 no gene *CCDC122*, rs141126876 no gene *MRC1*, Chr10:102968116 no gene *IL1RL1* e chr10: 54528099 e o rs5030737 no gene *MBL2*. Sete dos polimorfismos não-sinônimos – rs1801224 no gene *CUBN*, rs1800451, chr10: 54528099 e o rs5030737 no gene *MBL2*, rs79782223 no *CCDC122*, rs141126876 no gene *MRC1* e Chr10:102968116 no gene *IL1RL1* – foram classificados como danosos e são os melhores candidatos a serem variantes causais, a serem investigadas em estudos funcionais. Contudo, é necessária uma validação desses achados em uma amostra independente já que devido ao pequeno número amostral e a grande quantidade de variantes testadas, algumas das associações aqui detectadas podem ser aleatórias.

Além das validações/replicações encontradas nesse estudo que, *per se*, já são de grande importância em genética epidemiológica, nossa estratégia revelou variantes que apresentam um possível impacto causal/funcional na hanseníase. Apesar de nossa amostra populacional possuir um background genético distinto das utilizadas para os achados originais, as variantes em regiões transcritas encontradas nesse estudo podem explicar, através do desequilíbrio de ligação, as associações previamente detectadas entre a hanseníase e os genes *CCDC88B*, *IL1RL1*, *IL12RB2*, *LTA* e *NOD2*.

8 REFERÊNCIAS BIBLIOGRÁFICAS

1. Britton WJ, Lockwood DN. Leprosy. *Lancet*. 2004 Apr 10;363(9416):1209-19. PubMed PMID: 15081655.
2. Robbins G, Tripathy VM, Misra VN, Mohanty RK, Shinde VS, Gray KM, et al. Ancient skeletal evidence for leprosy in India (2000 B.C.). *PloS one*. 2009;4(5):e5669. PubMed PMID: 19479078. Pubmed Central PMCID: 2682583.
3. Bíblia Sagrada. 169 ed. São Paulo: Ave Maria; 2006.
4. Lewis G. A Lesson from Leviticus: Leprosy. *Man*. 1987;22(4):593-612.
5. Jay V. The legacy of Armauer Hansen. *Arch Pathol Lab Med*. 2000 Apr;124(4):496-7. PubMed PMID: 10747301.
6. Bechler RG. [Hansen versus Neisser: scientific controversies over the 'discovery' of the bacillus of leprosy]. *Historia, ciencias, saude--Manguinhos*. 2012 Sep;19(3):815-42. PubMed PMID: 23070374. Hansen versus Neisser: controversias científicas na 'descoberta' do bacilo da lepra.
7. Marmor MF. The ophthalmic trials of G. H. A. Hansen. *Survey of ophthalmology*. 2002 May-Jun;47(3):275-87. PubMed PMID: 12052414.
8. Leprosy. Global situation. *Releve epidemiologique hebdomadaire / Section d'hygiene du Secretariat de la Societe des Nations = Weekly epidemiological record / Health Section of the Secretariat of the League of Nations*. 2002 Jan 4;77(1):1-8. PubMed PMID: 11808001.
9. WHO. The final push strategy to eliminate leprosy as a public health problem : questions and answers. 2 ed: Geneva : World Health Organization; 2003.
10. WHO. Weekly epidemiological record - Global leprosy update, 2017: reducing the disease burden due to leprosy 2018. p. 445-56.
11. Sinan/SVS-MS B. Taxa de prevalência de hanseníase por 10.000 habitantes - Estados e regiões, Brasil, 1990 a 2017 Brasília/DF2018 [updated Segunda, 27 de Agosto de 2018; cited 2018 10/09/2018]. Available from: <http://portalarquivos2.saude.gov.br/images/pdf/2018/julho/13/Taxa-de-prevalencia-de-hanseníase-1990a2017.pdf>.
12. Clapasson AC, S. *Leprosy : A Practical Guide*: Springer-Verlag Mailand; 2012.

13. Shepard CC. The Experimental Disease That Follows the Injection of Human Leprosy Bacilli into Foot-Pads of Mice. *The Journal of experimental medicine*. 1960 Sep 1;112(3):445-54. PubMed PMID: 19867175. Pubmed Central PMCID: 2137235.
14. Balamayooran G, Pena M, Sharma R, Truman RW. The armadillo as an animal model and reservoir host for *Mycobacterium leprae*. *Clinics in dermatology*. 2015 Jan-Feb;33(1):108-15. PubMed PMID: 25432816.
15. 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 Dec;130(6):856-64. PubMed PMID: 19019760.
16. Kowalska M, Kowalik A. *Mycobacterium leprae*: pathogenic agent in leprosy. Discovery of new species *Mycobacterium lepromatosis*. *Perspectives in research and diagnosis of leprosy. International maritime health*. 2012;63(4):213-8. PubMed PMID: 24595978.
17. Case 165. Lepromatous leprosy: *Pathology of Infectious Diseases*; [cited 2019]. Available from: <https://pathos223.com/en/case/case165.htm>.
18. Truman R. Leprosy in wild armadillos. *Leprosy review*. 2005 Sep;76(3):198-208. PubMed PMID: 16248207.
19. 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 Nov 11;354(6313):744-7. PubMed PMID: 27846605.
20. 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 Jan;12(1):e0006190. PubMed PMID: 29381722. Pubmed Central PMCID: 5790234.
21. Brasil. Guia para o Controle da hanseníase. Brasilia: Ministerio da Saude; 2002.
22. de Wit MY, Douglas JT, McFadden J, Klatser PR. Polymerase chain reaction for detection of *Mycobacterium leprae* in nasal swab specimens. *J Clin Microbiol*. 1993 Mar;31(3):502-6. PubMed PMID: 8458942. Pubmed Central PMCID: 262809.
23. Klatser PR, van Beers S, Madjid B, Day R, de Wit MY. Detection of *Mycobacterium leprae* nasal carriers in populations for which leprosy is

endemic. *J Clin Microbiol.* 1993 Nov;31(11):2947-51. PubMed PMID: 8263180. Pubmed Central PMCID: 266167.

24. Job CK, Jayakumar J, Kearney M, Gillis TP. Transmission of leprosy: a study of skin and nasal secretions of household contacts of leprosy patients using PCR. *The American journal of tropical medicine and hygiene.* 2008 Mar;78(3):518-21. PubMed PMID: 18337353.

25. 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. PubMed PMID: 27203082. Pubmed Central PMCID: 4874629.

26. Matsuoka M, Izumi S, Budiawan T, Nakata N, Saeki K. *Mycobacterium leprae* DNA in daily using water as a possible source of leprosy infection. *Indian J Lepr.* 1999 Jan-Mar;71(1):61-7. PubMed PMID: 10439326. Epub 1999/08/10. eng.

27. 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 Sep;8(5):627-31. PubMed PMID: 18599381.

28. Jacobson RR, Krahenbuhl JL. Leprosy. *Lancet.* 1999 Feb 20;353(9153):655-60. PubMed PMID: 10030346.

29. Ridley DS, Jopling WH. Classification of leprosy according to immunity. A five-group system. *Int J Lepr Other Mycobact Dis.* 1966 Jul-Sep;34(3):255-73. PubMed PMID: 5950347.

30. Asia. WHOROfS-E. Guidelines for the diagnosis, treatment and prevention of leprosy. 2018. Available from: <http://www.who.int/iris/handle/10665/274127>.

31. Sauer ME, Salomao H, Ramos GB, D'Espindula HR, Rodrigues RS, Macedo WC, et al. Genetics of leprosy: Expected-and unexpected-developments and perspectives. *Clinics in dermatology.* 2016 Jan-Feb;34(1):96-104. PubMed PMID: 26773629.

32. Santos FSDd SL, Siani AC. . Chaulmoogra oil as scientific knowledge: the construction of a treatment for leprosy. *História, Ciências, Saúde-Manguinhos*. 2008;15:29-46.
33. Matsuoka M. Drug resistance in leprosy. *Japanese journal of infectious diseases*. 2010 Jan;63(1):1-7. PubMed PMID: 20093754.
34. Gelber RH, Grosset J. The chemotherapy of leprosy: an interpretive history. *Leprosy review*. 2012 Sep;83(3):221-40. PubMed PMID: 23356023.
35. Noordeen SK. History of chemotherapy of leprosy. *Clinics in dermatology*. 2016 Jan-Feb;34(1):32-6. PubMed PMID: 26773621.
36. group Ws. Chemotherapy of leprosy 1994.
37. Matsuoka M, Kashiwabara Y, Namisato M. A *Mycobacterium leprae* isolate resistant to dapsone, rifampin, ofloxacin and sparfloxacin. *International journal of leprosy and other mycobacterial diseases : official organ of the International Leprosy Association*. 2000 Dec;68(4):452-5. PubMed PMID: 11332288.
38. Penna GO, Buhner-Sekula S, Kerr LRS, Stefani MMA, Rodrigues LC, de Araujo MG, et al. Uniform multidrug therapy for leprosy patients in Brazil (U-MDT/CT-BR): Results of an open label, randomized and controlled clinical trial, among multibacillary patients. *PLoS neglected tropical diseases*. 2017 Jul;11(7):e0005725. PubMed PMID: 28704363. Pubmed Central PMCID: 5526599.
39. Hungria EM, Buhner-Sekula S, Oliveira RM, Aderaldo LC, Pontes MAA, Cruz R, et al. *Mycobacterium leprae*-Specific Antibodies in Multibacillary Leprosy Patients Decrease During and After Treatment With Either the Regular 12 Doses Multidrug Therapy (MDT) or the Uniform 6 Doses MDT. *Frontiers in immunology*. 2018;9:915. PubMed PMID: 29867930. Pubmed Central PMCID: 5960669.
40. Penna ML, Buhner-Sekula S, Pontes MA, Cruz R, Goncalves Hde S, Penna GO. Results from the clinical trial of uniform multidrug therapy for leprosy patients in Brazil (U-MDT/CT-BR): decrease in bacteriological index. *Leprosy review*. 2014 Dec;85(4):262-6. PubMed PMID: 25675650.
41. WHO. MDT Regiments.

42. Fava V, Orlova M, Cobat A, Alcais A, Mira M, Schurr E. Genetics of leprosy reactions: an overview. *Memorias do Instituto Oswaldo Cruz*. 2012 Dec;107 Suppl 1:132-42. PubMed PMID: 23283464.
43. Souza CS. Hanseníase: Formas Clínicas e Diagnóstico Diferencial. *Medicina*. 1997 jul./set;30:325-34.
44. Ranque B, Nguyen VT, Vu HT, Nguyen TH, Nguyen NB, Pham XK, et al. Age is an important risk factor for onset and sequelae of reversal reactions in Vietnamese patients with leprosy. *Clin Infect Dis*. 2007 Jan 1;44(1):33-40. PubMed PMID: 17143812.
45. Kahawita IP, Walker SL, Lockwood DNJ. Leprosy type 1 reactions and erythema nodosum leprosum. *Anais Brasileiros de Dermatologia*. 2008;83:75-82.
46. Cole ST, Eiglmeier K, Parkhill J, James KD, Thomson NR, Wheeler PR, et al. Massive gene decay in the leprosy bacillus. *Nature*. 2001 Feb 22;409(6823):1007-11. PubMed PMID: 11234002.
47. 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*. *Proceedings of the National Academy of Sciences of the United States of America*. 2015 Apr 7;112(14):4459-64. PubMed PMID: 25831531. Pubmed Central PMCID: 4394283.
- 48.
- Mycobacterium leprae* [Internet]. 2016. Available from: <http://www.ncbi.nlm.nih.gov/genome/?term=Mycobacterium+leprae>.
49. 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 Dec;41(12):1282-9. PubMed PMID: 19881526.
50. 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 Jul 12;341(6142):179-83. PubMed PMID: 23765279.
51. 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 May;14(5):e1006997. PubMed PMID: 29746563. Pubmed Central PMCID: 5944922.

52. MohamedAli. P, Ramanujam K. Leprosy in twins. International Journal of Leprosy. 1966;34(4):405-7.

53. Chakravarti MR, Vogel F. A twin study on leprosy: Georg Thieme; 1973.

54. Shields ED, Russell DA, Pericak-Vance MA. Genetic epidemiology of the susceptibility to leprosy. J Clin Invest. 1987 Apr;79(4):1139-43. PubMed PMID: 3549780. Pubmed Central PMCID: 424295.

55. Abel L, Demenais F. Detection of major genes for susceptibility to leprosy and its subtypes in a Caribbean island: Desirade island. American journal of human genetics. 1988 Feb;42(2):256-66. PubMed PMID: 3341381. Pubmed Central PMCID: 1715260.

56. Feitosa MF, Borecki I, Krieger H, Beiguelman B, Rao DC. The genetic epidemiology of leprosy in a Brazilian population. American journal of human genetics. 1995 May;56(5):1179-85. PubMed PMID: 7726175. Pubmed Central PMCID: 1801459.

57. Lazaro FP, Werneck RI, Mackert CC, Cobat A, Prevedello FC, Pimentel RP, et al. A major gene controls leprosy susceptibility in a hyperendemic isolated population from north of Brazil. The Journal of infectious diseases. 2010 May 15;201(10):1598-605. PubMed PMID: 20388034.

58. Abel L, Vu DL, Oberti J, Nguyen VT, Van VC, Guilloud-Bataille M, et al. Complex segregation analysis of leprosy in southern Vietnam. Genet Epidemiol. 1995;12(1):63-82. PubMed PMID: 7713401.

59. Wagener DK, Schauf V, Nelson KE, Scollard D, Brown A, Smith T. Segregation analysis of leprosy in families of northern Thailand. Genet Epidemiol. 1988;5(2):95-105. PubMed PMID: 3402727.

60. Pearson TA, Manolio TA. HOw to interpret a genome-wide association study. JAMA. 2008;299(11):1335-44.

61. Mueller JC. Linkage disequilibrium for different scales and applications. Briefings in bioinformatics. 2004 Dec;5(4):355-64. PubMed PMID: 15606972.

62. Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L, Nickerson DA. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. American journal of human

genetics. 2004 Jan;74(1):106-20. PubMed PMID: 14681826. Pubmed Central PMCID: 1181897.

63. Mira MT, Alcais 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 Mar;33(3):412-5. PubMed PMID: 12577057.

64. 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 Feb 12;427(6975):636-40. PubMed PMID: 14737177.

65. Ramos GB, Salomao H, Francio AS, Fava VM, Werneck RI, Mira MT. Association Analysis Suggests SOD2 as a Newly Identified Candidate Gene Associated With Leprosy Susceptibility. *The Journal of infectious diseases*. 2016 Aug 1;214(3):475-8. PubMed PMID: 27132285.

66. Malhotra D, Darvishi K, Lohra M, Kumar H, Grover C, Sood S, et al. Association study of major risk single nucleotide polymorphisms in the common regulatory region of PARK2 and PACRG genes with leprosy in an Indian population. *European journal of human genetics : EJHG*. 2006 Apr;14(4):438-42. PubMed PMID: 16391553.

67. Li J, Liu H, Liu J, Fu X, Yu Y, Yu G, et al. Association study of the single nucleotide polymorphisms of PARK2 and PACRG with leprosy susceptibility in Chinese population. *European journal of human genetics : EJHG*. 2012 May;20(5):488-9. PubMed PMID: 22009144. Pubmed Central PMCID: 3330206.

68. Bakija-Konsuo A, Mulic R, Boraska V, Pehlic M, Huffman JE, Hayward C, et al. Leprosy epidemics during history increased protective allele frequency of PARK2/PACRG genes in the population of the Mljet Island, Croatia. *Eur J Med Genet*. 2011 Nov-Dec;54(6):e548-52. PubMed PMID: 21816242.

69. Chopra R, Ali S, Srivastava AK, Aggarwal S, Kumar B, Manvati S, et al. Mapping of PARK2 and PACRG overlapping regulatory region reveals LD structure and functional variants in association with leprosy in unrelated indian population groups. *PLoS genetics*. 2013;9(7):e1003578. PubMed PMID: 23861666. Pubmed Central PMCID: 3701713.

70. Alter A, Fava VM, Huong NT, Singh M, Orlova M, Van Thuc N, et al. Linkage disequilibrium pattern and age-at-diagnosis are critical for replicating genetic associations across ethnic groups in leprosy. *Human genetics*. 2013 Jan;132(1):107-16. PubMed PMID: 23052943.

71. Wang D, Fan Y, Malhi M, Bi R, Wu Y, Xu M, et al. Missense Variants in HIF1A and LACC1 Contribute to Leprosy Risk in Han Chinese. *American journal of human genetics*. 2018 May 3;102(5):794-805. PubMed PMID: 29706348. Pubmed Central PMCID: 5986702.
72. 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 Apr;27(4):439-41. PubMed PMID: 11279529.
73. Alter A, de Leseleuc 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 Mar;127(3):337-48. PubMed PMID: 20035344. Pubmed Central PMCID: 2891106.
74. Wang D, Feng JQ, Li YY, Zhang DF, Li XA, Li QW, et al. Genetic variants of the MRC1 gene and the IFNG gene are associated with leprosy in Han Chinese from Southwest China. *Human genetics*. 2012 Jul;131(7):1251-60. PubMed PMID: 22392581.
75. 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 Jul;133(7):883-93. PubMed PMID: 24563210.
76. Medeiros P, da Silva WL, de Oliveira Gimenez BB, Vallezi KB, Moraes MO, de Souza VN, et al. The GATA3 gene is involved in leprosy susceptibility in Brazilian patients. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases*. 2016 Apr;39:194-200. PubMed PMID: 26807920.
77. Zhang FR, Huang W, Chen SM, Sun LD, Liu H, Li Y, et al. Genomewide association study of leprosy. *N Engl J Med*. 2009 Dec 31;361(27):2609-18. PubMed PMID: 20018961.
78. 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 Dec;43(12):1247-51. PubMed PMID: 22019778.
79. 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. *Human molecular genetics*. 2013 Nov 1;22(21):4430-7. PubMed PMID: 23784377. Pubmed Central PMCID: 3792694.

80. 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 Mar;47(3):267-71. PubMed PMID: 25642632.
81. Wang Z, Sun Y, Fu X, Yu G, Wang C, Bao F, et al. A large-scale genome-wide association and meta-analysis identified four novel susceptibility loci for leprosy. *Nature communications*. 2016 Dec 15;7:13760. PubMed PMID: 27976721. Pubmed Central PMCID: 5172377.
82. Liu H, Wang Z, Li Y, Yu G, Fu X, Wang C, et al. Genome-Wide Analysis of Protein-Coding Variants in Leprosy. *The Journal of investigative dermatology*. 2017 Dec;137(12):2544-51. PubMed PMID: 28842327.
83. Wong SH, Hill AV, Vannberg FO, India-Africa-United Kingdom Leprosy Genetics C. Genomewide association study of leprosy. *N Engl J Med*. 2010 Apr 15;362(15):1446-7; author reply 7-8. PubMed PMID: 20393182.
84. Berrington WR, Macdonald M, Khadge S, Sapkota BR, Janer M, Haggie DA, et al. Common polymorphisms in the NOD2 gene region are associated with leprosy and its reactive states. *The Journal of infectious diseases*. 2010 May 1;201(9):1422-35. PubMed PMID: 20350193. Pubmed Central PMCID: 2853728.
85. Grant AV, Alter A, Huong NT, Orlova M, Van Thuc N, Ba NN, et al. Crohn's Disease Susceptibility Genes are Associated With Leprosy in the Vietnamese Population. *Journal of Infectious Diseases*. 2012 December 1, 2012;206(11):1763-7.
86. Marcinek P, Jha AN, Shinde V, Sundaramoorthy A, Rajkumar R, Suryadevara NC, et al. LRRK2 and RIPK2 variants in the NOD 2-mediated signaling pathway are associated with susceptibility to *Mycobacterium leprae* in Indian populations. *PLoS one*. 2013;8(8):e73103. PubMed PMID: 24015287. Pubmed Central PMCID: 3756038.
87. Sales-Marques C, Salomao H, Fava VM, Alvarado-Arnez LE, Amaral EP, Cardoso CC, et al. NOD2 and CCDC122-LACC1 genes are associated with leprosy susceptibility in Brazilians. *Human genetics*. 2014 Dec;133(12):1525-32. PubMed PMID: 25367361.
88. Fava VM, Cobat A, Van Thuc N, Latini AC, Stefani MM, Belone AF, et al. Association of TNFSF8 regulatory variants with excessive inflammatory

responses but not leprosy per se. *The Journal of infectious diseases*. 2015 Mar 15;211(6):968-77. PubMed PMID: 25320285.

89. Xiong JH, Mao C, Sha XW, Jin Z, Wang H, Liu YY, et al. Association between genetic variants in NOD2, C13orf31, and CCDC122 genes and leprosy among the Chinese Yi population. *International journal of dermatology*. 2016 Jan;55(1):65-9. PubMed PMID: 26235265.

90. 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 Feb;10(2):e0004412. PubMed PMID: 26844546. Pubmed Central PMCID: 4742274.

91. Schurr E, Gros P. A common genetic fingerprint in leprosy and Crohn's disease? *The New England journal of medicine*. 2009 Dec 31;361(27):2666-8. PubMed PMID: 20018963.

92. Ferwerda G, Kullberg BJ, de Jong DJ, Girardin SE, Langenberg DM, van Crevel R, et al. Mycobacterium paratuberculosis is recognized by Toll-like receptors and NOD2. *Journal of leukocyte biology*. 2007 Oct;82(4):1011-8. PubMed PMID: 17652449.

93. Liu H, Irwanto A, Tian H, Fu X, Yu Y, Yu G, et al. Identification of IL18RAP/IL18R1 and IL12B as leprosy risk genes demonstrates shared pathogenesis between inflammation and infectious diseases. *American journal of human genetics*. 2012 Nov 2;91(5):935-41. PubMed PMID: 23103228. Pubmed Central PMCID: 3487119.

94. Jaeger M, Stappers MH, Joosten LA, Gyssens IC, Netea MG. Genetic variation in pattern recognition receptors: functional consequences and susceptibility to infectious disease. *Future microbiology*. 2015;10(6):989-1008. PubMed PMID: 26059622.

95. Bochud PY, Sinsimer D, Aderem A, Siddiqui MR, Saunderson P, Britton S, et al. Polymorphisms in Toll-like receptor 4 (TLR4) are associated with protection against leprosy. *European journal of clinical microbiology & infectious diseases* : official publication of the European Society of Clinical Microbiology. 2009 Sep;28(9):1055-65. PubMed PMID: 19430824. Pubmed Central PMCID: 3016895.

96. Wong SH, Gochhait S, Malhotra D, Pettersson FH, Teo YY, Khor CC, et al. Leprosy and the adaptation of human toll-like receptor 1. *PLoS pathogens*. 2010;6:e1000979. PubMed PMID: 20617178. Pubmed Central PMCID: 2895660.
97. Marques Cde S, Brito-de-Souza VN, Guerreiro LT, Martins JH, Amaral EP, Cardoso CC, et al. Toll-like receptor 1 N248S single-nucleotide polymorphism is associated with leprosy risk and regulates immune activation during mycobacterial infection. *The Journal of infectious diseases*. 2013 Jul;208(1):120-9. PubMed PMID: 23547143.
98. Alvarenga Niitsuma EN, Fernandes GDR, Lana FCF. The TLR1 gene is associated with higher protection from leprosy in women. *PloS one*. 2018;13(10):e0205234. PubMed PMID: 30289892.
99. Trowsdale J. The MHC, disease and selection. *Immunology letters*. 2011 Jun 30;137(1-2):1-8. PubMed PMID: 21262263.
100. Alcais A, Alter A, Antoni G, Orlova M, Van Thuc N, 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 04//print;39(4):517-22.
101. Cardoso CC, Pereira AC, Brito-de-Souza VN, Duraes SM, Ribeiro-Alves M, Nery JA, et al. TNF -308G>A single nucleotide polymorphism is associated with leprosy among Brazilians: a genetic epidemiology assessment, meta-analysis, and functional study. *The Journal of infectious diseases*. 2011 Oct 15;204(8):1256-63. PubMed PMID: 21917899.
102. 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. *Human genetics*. 2012 May;131(5):703-16. PubMed PMID: 22071774.
103. Tosh K, Ravikumar M, Bell JT, Meisner S, Hill AV, Pitchappan R. Variation in MICA and MICB genes and enhanced susceptibility to paucibacillary leprosy in South India. *Human molecular genetics*. 2006 Oct 1;15(19):2880-7. PubMed PMID: 16923796.
104. Alter A, Huong NT, Singh M, Orlova M, Van Thuc N, Katoch K, et al. Human leukocyte antigen class I region single-nucleotide polymorphisms are associated with leprosy susceptibility in Vietnam and India. *The Journal of*

- infectious diseases. 2011 May 1;203(9):1274-81. PubMed PMID: 21459816.
Pubmed Central PMCID: 3069725.
105. de Souza-Santana FC, Marcos EV, Nogueira ME, Ura S, Tomimori J. Human leukocyte antigen class I and class II alleles are associated with susceptibility and resistance in borderline leprosy patients from Southeast Brazil. *BMC infectious diseases*. 2015;15:22. PubMed PMID: 25605482.
Pubmed Central PMCID: 4307149.
106. do Sacramento WS, Mazini PS, Franceschi DA, de Melo FC, Braga MA, Sell AM, et al. Frequencies of MICA alleles in patients from southern Brazil with multibacillary and paucibacillary leprosy. *International journal of immunogenetics*. 2012 Jun;39(3):210-5. PubMed PMID: 22168250.
107. Abel L, Sanchez FO, Oberti J, Thuc NV, Hoa LV, Lap VD, et al. Susceptibility to leprosy is linked to the human NRAMP1 gene. *The Journal of infectious diseases*. 1998 Jan;177(1):133-45. PubMed PMID: 9419180.
108. Roy S, Frodsham A, Saha B, Hazra SK, Mascie-Taylor CG, Hill AV. Association of vitamin D receptor genotype with leprosy type. *The Journal of infectious diseases*. 1999 Jan;179(1):187-91. PubMed PMID: 9841838.
109. Meisner SJ, Mucklow S, Warner G, Sow SO, Lienhardt C, Hill AV. Association of NRAMP1 polymorphism with leprosy type but not susceptibility to leprosy per se in west Africans. *The American journal of tropical medicine and hygiene*. 2001 Dec;65(6):733-5. PubMed PMID: 11791966.
110. Brochado MJ, Gatti MF, Zago MA, Roselino AM. Association of the solute carrier family 11 member 1 gene polymorphisms with susceptibility to leprosy in a Brazilian sample. *Memorias do Instituto Oswaldo Cruz*. 2016 Feb;111(2):101-5. PubMed PMID: 26814595. Pubmed Central PMCID: 4750449.
111. Moraes MO, Pacheco AG, Schonkeren JJ, Vanderborght PR, Nery JA, Santos AR, et al. Interleukin-10 promoter single-nucleotide polymorphisms as markers for disease susceptibility and disease severity in leprosy. *Genes and immunity*. 2004 Nov;5(7):592-5. PubMed PMID: 15306847.
112. Alvarado-Arnez LE, Amaral EP, Sales-Marques C, Duraes SM, Cardoso CC, Nunes Sarno E, et al. Association of IL10 Polymorphisms and Leprosy: A Meta-Analysis. *PloS one*. 2015;10(9):e0136282. PubMed PMID: 26340474.
Pubmed Central PMCID: 4560376.

113. Cardoso CC, Pereira AC, Brito-de-Souza VN, Dias-Baptista IM, Maniero VC, Venturini J, et al. IFNG +874 T>A single nucleotide polymorphism is associated with leprosy among Brazilians. *Human genetics*. 2010;128(5):481-90.
114. de Messias-Reason IJ, Boldt AB, Moraes Braga AC, Von Rosen Seeling Stahlke E, Dornelles L, Pereira-Ferrari L, et al. The association between mannan-binding lectin gene polymorphism and clinical leprosy: new insight into an old paradigm. *The Journal of infectious diseases*. 2007 Nov 1;196(9):1379-85. PubMed PMID: 17922403.
115. de Messias-Reason I, Kremsner PG, Kun JF. Functional haplotypes that produce normal ficolin-2 levels protect against clinical leprosy. *The Journal of infectious diseases*. 2009 Mar 15;199(6):801-4. PubMed PMID: 19434912.
116. Tobin DM, Vary JC, Jr., Ray JP, Walsh GS, Dunstan SJ, Bang ND, et al. The *Ita4h* locus modulates susceptibility to mycobacterial infection in zebrafish and humans. *Cell*. 2010 Mar 5;140(5):717-30. PubMed PMID: 20211140. Pubmed Central PMCID: 2907082.
117. Zhang DF, Huang XQ, Wang D, Li YY, Yao YG. Genetic variants of complement genes ficolin-2, mannose-binding lectin and complement factor H are associated with leprosy in Han Chinese from Southwest China. *Human genetics*. 2013 Jun;132(6):629-40. PubMed PMID: 23423485.
118. Cardona-Pemberthy V, Rendon M, Beltran JC, Soto-Ospina A, Munoz-Gomez A, Araque-Marin P, et al. Genetic variants, structural, and functional changes of Myelin Protein Zero and Mannose-Binding Lectin 2 protein involved in immune response and its allelic transmission in families of patients with leprosy in Colombia. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases*. 2018 Jul;61:215-23. PubMed PMID: 29627640.
119. Zhang DF, Wang D, Li YY, Yao YG. Mapping genetic variants in the CFH gene for association with leprosy in Han Chinese. *Genes and immunity*. 2014 Oct;15(7):506-10. PubMed PMID: 25030427.
120. Chaitanya VS, Jadhav RS, Lavania M, Singh M, Valluri V, Sengupta U. Interleukin-17F single-nucleotide polymorphism (7488T>C) and its association with susceptibility to leprosy. *International journal of immunogenetics*. 2014 Apr;41(2):131-7. PubMed PMID: 24164796.

121. Morey M, Fernandez-Marmiesse A, Castineiras D, Fraga JM, Couce ML, Cocho JA. A glimpse into past, present, and future DNA sequencing. *Mol Genet Metab.* 2013 Sep-Oct;110(1-2):3-24. PubMed PMID: 23742747.
122. Illumina. Images for general use 2019. Available from: <https://www.illumina.com/company/news-center/multimedia-images.html>.
123. Metzker ML. Sequencing technologies - the next generation. *Nat Rev Genet.* 2010 Jan;11(1):31-46. PubMed PMID: 19997069.
124. Frazer KA, Murray SS, Schork NJ, Topol EJ. Human genetic variation and its contribution to complex traits. *Nat Rev Genet.* 2009 Apr;10(4):241-51. PubMed PMID: 19293820.
125. Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. *American journal of human genetics.* 2012 Jan 13;90(1):7-24. PubMed PMID: 22243964. Pubmed Central PMCID: 3257326.
126. John SW, Weitzner G, Rozen R, Scriver CR. A rapid procedure for extracting genomic DNA from leukocytes. *Nucleic acids research.* 1991 Jan 25;19(2):408. PubMed PMID: 2014181. Pubmed Central PMCID: 333617.
127. Inc I. TruSeq Custom Amplicon v1.5 A new and improved amplicon sequencing solution for interrogating custom regions of interest. U.S.2016.
128. Illumina I. TruSeq Custom Amplicom v1.5 Reference Guide. U.S.2016.
129. Illumina I. MiSeq® System Guide. U.S.2015.
130. Illumina I. MiSeq® Reagent Kit v3. Reagent Preparation Guide. U.S.2013.
131. Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010.
132. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics.* 2011 Mar 15;27(6):863-4. PubMed PMID: 21278185. Pubmed Central PMCID: 3051327.
133. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2010 Mar 1;26(5):589-95. PubMed PMID: 20080505. Pubmed Central PMCID: 2828108.
134. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature genetics.* 2011 May;43(5):491-8. PubMed PMID: 21478889. Pubmed Central PMCID: 3083463.

135. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics*. 2011 Aug 1;27(15):2156-8. PubMed PMID: 21653522. Pubmed Central PMCID: 3137218.
136. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic acids research*. 2010 Sep;38(16):e164. PubMed PMID: 20601685. Pubmed Central PMCID: 2938201.
137. Horvath S, Xu X, Laird NM. The family based association test method: strategies for studying general genotype--phenotype associations. *European journal of human genetics : EJHG*. 2001 Apr;9(4):301-6. PubMed PMID: 11313775.
138. Yip WK, De G, Raby BA, Laird N. Identifying causal rare variants of disease through family-based analysis of Genetics Analysis Workshop 17 data set. *BMC proceedings*. 2011 Nov 29;5 Suppl 9:S21. PubMed PMID: 22373204. Pubmed Central PMCID: 3287856.
139. Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, et al. Ensembl 2018. *Nucleic acids research*. 2018 Jan 4;46(D1):D754-D61. PubMed PMID: 29155950. Pubmed Central PMCID: 5753206.
140. Kim SA, Cho CS, Kim SR, Bull SB, Yoo YJ. A new haplotype block detection method for dense genome sequencing data based on interval graph modeling of clusters of highly correlated SNPs. *Bioinformatics*. 2018 Feb 1;34(3):388-97. PubMed PMID: 29028986. Pubmed Central PMCID: 5860363.
141. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. *Nature methods*. 2010 Apr;7(4):248-9. PubMed PMID: 20354512. Pubmed Central PMCID: 2855889.
142. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature protocols*. 2009;4(7):1073-81. PubMed PMID: 19561590.
143. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nature genetics*. 2014 Mar;46(3):310-5. PubMed PMID: 24487276. Pubmed Central PMCID: 3992975.

144. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic acids research*. 2012;40(D1):D930-D4.
145. Consortium GT. The Genotype-Tissue Expression (GTEx) project. *Nature genetics*. 2013 Jun;45(6):580-5. PubMed PMID: 23715323. Pubmed Central PMCID: 4010069.
146. Westra HJ, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nature genetics*. 2013 Oct;45(10):1238-43. PubMed PMID: 24013639. Pubmed Central PMCID: 3991562.
147. Nedelec Y, Sanz J, Baharian G, Szpiech ZA, Pacis A, Dumaine A, et al. Genetic Ancestry and Natural Selection Drive Population Differences in Immune Responses to Pathogens. *Cell*. 2016 Oct 20;167(3):657-69 e21. PubMed PMID: 27768889.
148. Ohya H, Ogata K, Takeuchi K, Namisato M, Fukutomi Y, Nishimura F, et al. Polymorphism of the 5' flanking region of the IL-12 receptor beta2 gene partially determines the clinical types of leprosy through impaired transcriptional activity. *Journal of clinical pathology*. 2005 Jul;58(7):740-3. PubMed PMID: 15976343. Pubmed Central PMCID: 1770718.
149. Werneck RI, Lazaro FP, Cobat A, Grant AV, Xavier MB, Abel L, et al. A major gene effect controls resistance to caries. *Journal of dental research*. 2011 Jun;90(6):735-9. PubMed PMID: 21364090. Pubmed Central PMCID: 3318028.
150. Barreiro LB, Talleux L, Pai AA, Gicquel B, Marioni JC, Gilad Y. Deciphering the genetic architecture of variation in the immune response to *Mycobacterium tuberculosis* infection. *Proceedings of the National Academy of Sciences of the United States of America*. 2012 Jan 24;109(4):1204-9. PubMed PMID: 22233810. Pubmed Central PMCID: 3268270.
151. TRDMT1 tRNA aspartic acid methyltransferase 1 [Internet]. NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION. 2019 [cited 29-Jan-2019]. Available from: <https://www.ncbi.nlm.nih.gov/gene/1787>.
152. MBL2 mannose binding lectin 2 [Homo sapiens (human)] [Internet]. NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION. 2018. Available from: <https://www.ncbi.nlm.nih.gov/gene/4153>.

153. Sun BB, Maranville JC, Peters JE, Stacey D, Staley JR, Blackshaw J, et al. Genomic atlas of the human plasma proteome. *Nature*. 2018 Jun;558(7708):73-9. PubMed PMID: 29875488.
154. Larsen F, Madsen HO, Sim RB, Koch C, Garred P. Disease-associated mutations in human mannose-binding lectin compromise oligomerization and activity of the final protein. *The Journal of biological chemistry*. 2004 May 14;279(20):21302-11. PubMed PMID: 14764589.
155. Cao Y, Wang X, Cao Z, Wu C, Wu D, Cheng X. Genetic polymorphisms of MBL2 and tuberculosis susceptibility: a meta-analysis of 22 case-control studies. *Archives of medical science : AMS*. 2018 Oct;14(6):1212-32. PubMed PMID: 30393476. Pubmed Central PMCID: 6209707.
156. Harm-Jan Westra MJP, Tõnu Esko, Hanieh Yaghootkar, Claudia Schurmann, Johannes Kettunen, Mark W Christiansen, Benjamin P Fairfax, Katharina Schramm, Joseph E Powell, Alexandra Zhernakova, Daria V Zhernakova, Jan H Veldink, Leonard H Van den Berg, Juha Karjalainen, Sebo Withoff, André G Uitterlinden, Albert Hofman, Fernando Rivadeneira, Peter A C 't Hoen, Eva Reinmaa, Krista Fischer, Mari Nelis, Lili Milani, David Melzer, Luigi Ferrucci, Andrew B Singleton, Dena G Hernandez, Michael A Nalls, Georg Homuth, Matthias Nauck, Dörte Radke, Uwe Völker, Markus Perola, Veikko Salomaa, Jennifer Brody, Astrid Suchy-Dicey, Sina A Gharib, Daniel A Enquobahrie, Thomas Lumley, Grant W Montgomery, Seiko Makino, Holger Prokisch, Christian Herder, Michael Roden, Harald Grallert, Thomas Meitinger, Konstantin Strauch, Yang Li, Ritsert C Jansen, Peter M Visscher, Julian C Knight, Bruce M Psaty, Samuli Ripatti, Alexander Teumer, Timothy M Frayling, Andres Metspalu, Joyce B J van Meurs and Lude Franke. Blood eQTL browser [cited 2019]. Available from: <https://genenetwork.nl/bloodeqtlbrowser/>.
157. Consortium GT, Laboratory DA, Coordinating Center -Analysis Working G, Statistical Methods groups-Analysis Working G, Enhancing Gg, Fund NIHC, et al. Genetic effects on gene expression across human tissues. *Nature*. 2017 Oct 11;550(7675):204-13. PubMed PMID: 29022597. Pubmed Central PMCID: 5776756.
158. Liu JZ, van Sommeren S, Huang H, Ng SC, Alberts R, Takahashi A, et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nature genetics*.

2015 Sep;47(9):979-86. PubMed PMID: 26192919. Pubmed Central PMCID: 4881818.

159. Takeuchi M, Mizuki N, Meguro A, Ombrello MJ, Kirino Y, Satorius C, et al. Dense genotyping of immune-related loci implicates host responses to microbial exposure in Behcet's disease susceptibility. *Nature genetics*. 2017 Mar;49(3):438-43. PubMed PMID: 28166214. Pubmed Central PMCID: 5369770.

160. Zierhut M, Mizuki N, Ohno S, Inoko H, Gul A, Onoe K, et al. Immunology and functional genomics of Behcet's disease. *Cellular and molecular life sciences : CMLS*. 2003 Sep;60(9):1903-22. PubMed PMID: 14523551.

161. NCBI. BCL10 B cell CLL/lymphoma 10 [Homo sapiens (human)]. 2018 18-Oct-2018:Gene ID: 8915, updated on 13-Oct-2018.

162. Fehrmann RS, Jansen RC, Veldink JH, Westra HJ, Arends D, Bonder MJ, et al. Trans-eQTLs reveal that independent genetic variants associated with a complex phenotype converge on intermediate genes, with a major role for the HLA. *PLoS genetics*. 2011 Aug;7(8):e1002197. PubMed PMID: 21829388. Pubmed Central PMCID: 3150446.

163. IL12RB2 interleukin 12 receptor subunit beta 2 [Homo sapiens (human)] [Internet]. NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION. 2018 [cited 17-Oct-2018]. Available from: <https://www.ncbi.nlm.nih.gov/gene/3595>.

164. de Lange KM, Moutsianas L, Lee JC, Lamb CA, Luo Y, Kennedy NA, et al. Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nature genetics*. 2017 Feb;49(2):256-61. PubMed PMID: 28067908. Pubmed Central PMCID: 5289481.

165. Yohann Nédélec JS, Golshid Baharian, Zachary A. Szpiech, Alain Pacis, Anne Dumaine, Jean-Christophe Grenier, Andrew Freiman, Aaron J. Sams, Steven Hebert, Ariane Pagé Sabourin, Francesca Luca, Ran Blekhman, Ryan D. Hernandez, Roger Pique-Regi, Jenny Tung, Vania Yotova, Luis B. Barreiro. ImmunPop browser [cited 2019]. Available from: <http://132.219.138.157/nedelec/eQTL/>.

166. Franke A, McGovern DP, Barrett JC, Wang K, Radford-Smith GL, Ahmad T, et al. Genome-wide meta-analysis increases to 71 the number of confirmed

- Crohn's disease susceptibility loci. *Nature genetics*. 2010 Dec;42(12):1118-25. PubMed PMID: 21102463. Pubmed Central PMCID: 3299551.
167. LTA lymphotoxin alpha [*Homo sapiens* (human)] [Internet]. NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION. 2018. Available from: <https://www.ncbi.nlm.nih.gov/gene/4049>.
168. Dixon AL, Liang L, Moffatt MF, Chen W, Heath S, Wong KC, et al. A genome-wide association study of global gene expression. *Nature genetics*. 2007 Oct;39(10):1202-7. PubMed PMID: 17873877.
169. NOD2 nucleotide binding oligomerization domain containing 2 [*Homo sapiens* (human)] [Internet]. NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION. 2018 [cited 17-Oct-2018]. Available from: <https://www.ncbi.nlm.nih.gov/gene/64127>.
170. 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. PubMed PMID: 30079069. Pubmed Central PMCID: 6062607.
171. Schenk M, Mahapatra S, Le P, Kim HJ, Choi AW, Brennan PJ, et al. Human NOD2 Recognizes Structurally Unique Muramyl Dipeptides from *Mycobacterium leprae*. *Infection and immunity*. 2016 Sep;84(9):2429-38. PubMed PMID: 27297389. Pubmed Central PMCID: 4995902.
172. Zeller T, Wild P, Szymczak S, Rotival M, Schillert A, Castagne R, et al. Genetics and beyond--the transcriptome of human monocytes and disease susceptibility. *PLoS one*. 2010 May 18;5(5):e10693. PubMed PMID: 20502693. Pubmed Central PMCID: 2872668.
173. PRKN parkin RBR E3 ubiquitin protein ligase [*Homo sapiens* (human)] [Internet]. NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION. 2018 [cited 17-Oct-2018]. Available from: <http://www.ncbi.nlm.nih.gov/gene/5071>.
174. Kitada T, Asakawa S, Hattori N, Matsumine H, Yamamura Y, Minoshima S, et al. Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. *Nature*. 1998 Apr 9;392(6676):605-8. PubMed PMID: 9560156.
175. Ali S, Vollaard AM, Widjaja S, Surjadi C, van de Vosse E, van Dissel JT. PARK2/PACRG polymorphisms and susceptibility to typhoid and paratyphoid fever. *Clinical and experimental immunology*. 2006 Jun;144(3):425-31. PubMed PMID: 16734611. Pubmed Central PMCID: 1941982.

176. 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 Sep 26;501(7468):512-6. PubMed PMID: 24005326. Pubmed Central PMCID: 3886920.
177. MRC1 mannose receptor C-type 1 [Homo sapiens (human)] [Internet]. NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION. 2018 [cited 17-Oct-2018]. Available from: <https://www.ncbi.nlm.nih.gov/gene/4360>.
178. CCDC122 coiled-coil domain containing 122 [Internet]. NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION. 2018 [cited 17-Oct-2018]. Available from: <http://www.ncbi.nlm.nih.gov/gene/160857>.
179. MICB MHC class I polypeptide-related sequence B [Homo sapiens (human)] [Internet]. NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION. 2018 [cited 17-Oct-2018]. Available from: <https://www.ncbi.nlm.nih.gov/gene/4277>.
180. Gourab De SH, Nan Laird, Steve Lake, Christoph Lange, Kristel Van Steen,, Lin Wang W-KY, Xin Xu, and Jin Zhou. FBAT USER'S MANUAL. 2013.
181. IL1RL1 interleukin 1 receptor like 1 [Homo sapiens (human)] [Internet]. NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION. 2018 [cited 17-Oct-2018]. Available from: <https://www.ncbi.nlm.nih.gov/gene/9173>.

ANEXO I: TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Estudo de Fatores de Riscos Genéticos de Susceptibilidade do Hospedeiro à Hanseníase

A Pontifícia Universidade Católica do Paraná (PUCPR) estará realizando um estudo que irá investigar o papel da genética na susceptibilidade à hanseníase (lepra) e na gravidade da doença. Se bem sucedido, o projeto poderá facilitar o tratamento da hanseníase e melhorar a qualidade de vida dos pacientes, seus familiares e pessoas próximas. Este estudo será coordenado por Dr. Marcelo Távora Mira, professor adjunto e pesquisador do programa de Pós-Graduação em Ciências da Saúde da PUCPR.

1) Métodos

Se você concordar em participar ou autorizar seu dependente a participar neste estudo, você responderá a uma entrevista com perguntas sobre dados pessoais, estilo de vida e saúde seus e/ou de seu dependente. Sob sua autorização, informações de prontuários clínicos também poderão ser lidas pelos cientistas e utilizadas no estudo.

2) Local do estudo

A unidade de origem deste estudo é o programa de Pós-Graduação em Ciências da Saúde da PUCPR. Colaboração entre a PUCPR e outros institutos de pesquisa deverão aumentar as chances de sucesso e o impacto do estudo. Portanto, está sendo requerida sua autorização para o uso de sua amostra em estudos da hanseníase envolvendo outras instituições. Sua autorização só valerá sob a condição de se manter o desenho, objetivos e metodologia do projeto original, além da avaliação e aprovação de eventuais alterações pelo Comitê de Ética da PUCPR.

3) Riscos Físicos para Saúde/Desconfortos

Não existe riscos físicos para sua saúde e/ou de seu dependente decorrente da participação neste estudo.

4) Alternativas

Se você e/ou seu dependente estiver afetado por hanseníase, acesso a procedimentos médicos para diagnóstico e tratamento da doença será providenciado mesmo que não concorde em participar ou não autorize a participação de seu dependente no estudo. Portanto, se você decidir não participar, ou cessar sua participação e/ou de seu

dependente no estudo a qualquer momento, todos os procedimentos para diagnósticos e tratamento médico serão mantidos pelo Hospital Santa Casa de Misericórdia da PUCPR e instituições colaboradoras.

Se você e/ou seu dependente não estiver afetado pela hanseníase, sua decisão de participar e/ou autorizar ou não a participação de seu dependente, ou de cessar sua participação a qualquer momento, não irá interferir de nenhuma forma nos procedimentos médicos para diagnóstico ou tratamento da hanseníase que você e/ou seu dependente possa necessitar no futuro. Da mesma forma, sua decisão não irá refletir no acesso a procedimentos médicos necessários à algum familiar ou contato afetado pela hanseníase.

5) Custos para os participantes

No caso de você decidir participar ou autorizar a participação do seu dependente no estudo, você ou seu dependente não terão nenhum custo.

6) Benefícios

A longo prazo, resultados deste projeto podem facilitar a detecção de resistência ao tratamento da hanseníase, tornando possível evitar tratamentos inadequados. Além disso, espera-se que conhecimentos científicos adicionais sejam alcançados, com consequente melhoria do tratamento de pessoas afetadas pela hanseníase.

7) Reembolso

Nem você nem seu dependente serão reembolsados por participar deste estudo.

8) Confidenciabilidade dos dados

A participação em projetos de pesquisa pode resultar em perda de privacidade. Além disso, a descoberta de fatores de risco genéticos para hanseníase podem expor susceptibilidades de certos grupos de pessoas, possivelmente levando a outros ou certas empresas a considerar estes grupos diferentes de uma forma negativa. Entretanto, procedimentos serão adotados pelos responsáveis por este estudo no intuito de proteger a confidencialidade das informações que você fornecer e as informações produzidas pelo projeto. Nenhuma informação genética individual será tornada pública. As informações serão codificadas e mantidas num local reservado o tempo todo. Somente os pesquisadores envolvidos neste estudo terão acesso às informações. Após o término deste estudo, as informações serão transcritas dos questionários para arquivos de computador, mantidos em local restrito com acesso permitido apenas aos mesmos pesquisadores. Os dados deste estudo poderão ser discutidos com pesquisadores de outras instituições, mas nenhuma identificação será fornecida.

CONSENTIMENTO

Estudo de Fatores de Riscos Genéticos de Susceptibilidade do Hospedeiro à Hanseníase

Você receberá uma cópia deste Termo de Consentimento para mantê-lo consigo. Se você tiver qualquer dúvida no futuro sobre a sua participação neste estudo, você pode e deve utilizar os seguintes meios de contato com o pesquisador responsável: Dr. Marcelo Távora Mira, telefone: (41) 3271-2618, celular: (41) 9164-4045, e-mail: m.mira@pucpr.br

A PARTICIPAÇÃO EM PESQUISA É VOLUNTARIA

Você tem o direito de não concordar a participação de seu dependente, ou mesmo de cessar a participação de seu dependente do estudo em qualquer momento que queira, sem riscos para o seu tratamento dele/dela. Se você autorizar a participação de seu dependente do estudo, permitirá desejar e concordar em participar, deve assinar ou fornecer sua impressão digital na linha apropriada abaixo.

Se você desejar participar do estudo, permitirá que seu endereço e telefone sejam anotados em uma folha separada, para facilitar contato quando necessário. Como já foi esclarecida anteriormente, toda informação, pessoal será mantida em sigilo.

Assinatura ou impressão digital do voluntário Nome completo e nº do prontuário

Assinatura do entrevistador Nome do entrevistador

Assinatura testemunha 1 Assinatura testemunha 2

_____/_____/_____

ANEXO II: APROVAÇÃO DO CONEP E DO CEP (2006 E 2007)



PONTIFICAL CATHOLIC UNIVERSITY OF PARANÁ

RESEARCH ETHICS COMMITTEE

Curitiba (Brazil), 2008-06-23

To whom it may concerns

We hereby, certify that the study entitled “**Estudo de fatores de risco genéticos de susceptibilidade do hospedeiro à hanseníase**” (brazilian version), was submitted to the Research Ethics Committee of the Pontifical Catholic University of Paraná. After analysis the project was approved for execution at the committee meeting of **December 12, 2005**, under the protocol number **988**. This research protocol was also submitted and approved by The Brazilian National Research Ethics Commission (CONEP), under registration number 12610, at June 21, 2006.

A handwritten signature in blue ink, appearing to read "S. Surugi de Siqueira", is written over a horizontal line.

Prof. Sergio Surugi de Siqueira, Ph.D.

Coordinator

Research Ethics Committee of the Pontifical Catholic University of Paraná





Pontifícia Universidade Católica do Paraná
Pró-Reitoria Acadêmica e de Pesquisa
Núcleo de Bioética

Curitiba, 29 de fevereiro de 2007
Of.nº 51/07/CEP-PUCPR

Projeto de Pesquisa: “Estudo de Fatores de Risco Moleculares de Susceptibilidade do Hospedeiro à Hanseníase”.

Pesquisador Responsável: Marcelo Távora Mira.

Instituição: Programa de Pós-Graduação em Ciências da Saúde da PUCPR

Comentários: O projeto de pesquisa: “Estudo de Fatores de Risco Moleculares de Susceptibilidade do Hospedeiro à Hanseníase”, a ser desenvolvido junto ao Programa de Pós-Graduação em Ciências da Saúde da PUCPR, sob coordenação do Prof. Marcelo Távora Mira é o resultado da combinação literal, sem qualquer modificação, de dois outros projetos, intitulados “Estudo de Fatores de Risco Genéticos de Susceptibilidade à Hanseníase” e “Estudo funcional de susceptibilidade celular ao *Mycobacterium leprae* em macrófagos primários humanos com mutação no gene *PARK2*”. O primeiro projeto foi avaliado e aprovado previamente pelo CEP-PUCPR, pelo CONEP e pelo Comitê de Ética da Organização Mundial da Saúde (OMS), baseado em Genebra, Suíça. O segundo foi aprovado pelo CEP-PUCPR.

Devido ao exposto, o Comitê de Ética em Pesquisa da PUCPR, de acordo com as exigências das Resoluções Nacionais 196/96, relacionadas a pesquisas envolvendo seres humanos, em reunião realizada no dia 28 de fevereiro de 2007, manifesta-se pela reiteração da aprovação conjunta dos projetos acima.

Situação: Aprovado.

Prof. Sergio Surugi de Siqueira

Coordenador do Comitê de Ética em Pesquisa - PUCPR

ANEXO III: APROVAÇÃO PELO CEP (2014)



Comitê de Ética
em Pesquisa da
PUCPR

ASSOCIAÇÃO PARANAENSE
DE CULTURA - PUCPR



PARECER CONSUBSTANCIADO DO CEP

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: Análise comparativa de sequências de genes associados com susceptibilidade do hospedeiro à hanseníase

Pesquisador: Marcelo Távora Mira

Área Temática: Genética Humana:

(Trata-se de pesquisa envolvendo Genética Humana que não necessita de análise ética por parte da CONEP;);

Versão: 1

CAAE: 35941614.7.0000.0020

Instituição Proponente: Pontifícia Universidade Católica do Paraná - PUCPR

Patrocinador Principal: Financiamento Próprio

DADOS DO PARECER

Número do Parecer: 807.698

Data da Relatoria: 17/09/2014

Apresentação do Projeto:

Análise comparativa de sequências de genes associados com susceptibilidade do hospedeiro à hanseníase

A população da Vila do Prata é composta por aproximadamente 4.000. Em torno de 13% da população é afetada por hanseníase.

Esse estudo incluirá 70 de cerca de 200 indivíduos que já aceitaram (e assinaram o TCLE) participar de estudos envolvendo análise genética molecular, realizados por nosso grupo. Posteriormente, a amostra populacional do presente projeto poderá ser ampliada, com vistas a acrescentar poder estatístico ao estudo. Amostras de 5mL de sangue periférico destes indivíduos foram obtidas em expedições anteriores realizadas na Vila do Prata, em tubos vacutainer® contendo anticoagulante EDTA.

O presente estudo está incluindo em um projeto maior, intitulado "Estudo de fatores de risco genéticos de susceptibilidade do hospedeiro à hanseníase", aprovado pelo Comitê de Ética em

Endereço: Rua Imaculada Conceição 1155

Bairro: Prado Velho

CEP: 80.215-901

UF: PR

Município: CURITIBA

Telefone: (41)3271-2292

Fax: (41)3271-2292

E-mail: nep@pucpr.br



Comitê de Ética
em Pesquisa da
PUCPR

ASSOCIAÇÃO PARANAENSE DE CULTURA - PUCPR



Continuação do Parecer: 807.698

Pesquisa (CEP) da PUCPR (Pontifícia Universidade Católica do Paraná) - Of. nº 51/07/CEPPUCPR de 29 de fevereiro de 2007 -, na Comissão Nacional de Ética em Pesquisa (CONEP) - Parecer nº 678/2006 de 21 de junho de 2006 - e no Comitê de ética em pesquisa da Organização Mundial da Saúde (The Ethics Research Committee). Os indivíduos anteriormente recrutados envolvidos nesse projeto, ou seus representantes legais, responderam e assinaram um termo de consentimento livre e esclarecido (TCLE), e tiveram amostra de sangue periférico coletada. Como informação adicional, o TCLE aprovado na pesquisa supracitada está sendo encaminhado em anexo.

Todas as informações contidas no TCLE serão mantidas. **IMPORTANTE:** embora esta proposta introduza uma nova técnica de sequenciamento, inexistente na época do recrutamento original, a proposta do estudo permanece essencialmente a mesma: a identificação de variantes genéticas associadas com a ocorrência de hanseníase. A presente submissão propõe apenas a substituição de um método por outro mais atual, de capacidade ampliada de forma exponencial, que aumenta na mesma proporção a chance de sucesso do estudo.

Objetivo da Pesquisa:

O objetivo primário deste estudo é a aplicação da técnica de sequenciamento de próxima geração para o sequenciamento da região codificadora (Cds), promotora (5000 pares de bases) e não transcrita (UTR) de cinco genes previamente associados com a hanseníase em estudos anteriormente realizados em uma amostra populacional baseada em famílias da Vila Santo Antônio do Prata, norte do Brasil, com a finalidade de

encontrar variantes raras, que auxiliem no entendimento dos componentes genéticos que controlam a susceptibilidade do hospedeiro à hanseníase.

Avaliação dos Riscos e Benefícios:

Riscos

Não existe risco decorrente da participação neste estudo, tendo em vista que o material biológico (DNA) a ser utilizado já se encontra disponível em nosso laboratório, como produto de coletas efetuadas em estudos anteriores. Não há previsão de necessidade de nova coleta de amostra biológica.

Benefícios

Não há benefícios imediatos.

Endereço: Rua Imaculada Conceição 1155

Bairro: Prado Velho

CEP: 80.215-901

UF: PR

Município: CURITIBA

Telefone: (41)3271-2292

Fax: (41)3271-2292

E-mail: nep@pucpr.br

Continuação do Parecer: 807.698

Comentários e Considerações sobre a Pesquisa:

sem comentários

Considerações sobre os Termos de apresentação obrigatória:

sem considerações

Recomendações:

sem recomendações

Conclusões ou Pendências e Lista de Inadequações:

-

Situação do Parecer:

Aprovado

Necessita Apreciação da CONEP:

Não

Considerações Finais a critério do CEP:

CURITIBA, 26 de Setembro de 2014

**Assinado por:
NAIM AKEL FILHO
(Coordenador)**

Endereço: Rua Imaculada Conceição 1155
Bairro: Prado Velho **CEP:** 80.215-901
UF: PR **Município:** CURITIBA
Telefone: (41)3271-2292 **Fax:** (41)3271-2292 **E-mail:** nep@pucpr.br

ANEXO IV: TERMO DE ACEITAÇÃO DE APOIO FINANCEIRO



Prof. Marcelo Távora Mira,
Pontifícia Universidade Católica do Paraná
Rua Imaculada Conceição, 1155
Prado Velho - CEP: 80215-901
Curitiba, Paraná - Brazil

PO Box 95005
1090 HA Amsterdam
The Netherlands
+ 31 20 595 05 34
info@leprosyresearch.org

Ref: 704.16.31

7 December 2015

Subject: Proposal – Comparative sequencing analysis of genes associated with susceptibility to leprosy and its reactive states

Dear Prof. Marcelo Távora Mira,

Thank you for submitting a proposal to the Leprosy Research Initiative (LRI). After careful consideration by the Scientific Review Committee and LRI Steering Committee, we are pleased to inform you that the LRI Executives Group approved your proposal for funding. The Turing Foundation will co-fund 50% of your project via LRI. The official LRI Grant Award Letter and Conditions will follow shortly.

Yours faithfully,

A handwritten signature in blue ink, appearing to read "Dinnissen", with a long horizontal flourish extending to the right.

Nicole Dinnissen
LRI Secretariat

ANEXO V: ARTIGO

Human genetic susceptibility of leprosy recurrence

OPEN

Human Genetic Susceptibility of Leprosy Recurrence

Priscila Verchai Uaska Sartori¹, Gerson O. Penna^{2,3,4}, Samira Bühner-Sékula⁵, Maria A. A. Pontes⁶, Heitor S. Gonçalves⁶, Rossilene Cruz⁷, Marcos C. L. Virmond⁸, Ida M. F. Dias-Baptista⁸, Patricia S. Rosa⁸, Maria L. F. Penna⁹, Vinicius Medeiros Fava¹⁰, Mariane M. A. Stefani^{4*} & Marcelo Távora Mira^{1*}

Host genetic susceptibility to leprosy has been intensively investigated over the last decades; however, there are no studies on the role of genetic variants in disease recurrence. A previous initiative identified three recurrent cases of leprosy for which none of the *M. leprae* strains, as obtained in the first and the second diagnosis, had any known genomic variants associated to resistance to Multidrug therapy; in addition, whole genome sequencing indicated that the same *M. leprae* was causing two out of the three recurrences. Thus, these individuals were suspected of being particularly susceptible to *M. leprae* infection, either as relapse or reinfection. To verify this hypothesis, 19 genetic markers distributed across 11 *loci* (14 genes) classically associated with leprosy were genotyped in the recurrent and in three matching non-recurrent leprosy cases. An enrichment of risk alleles was observed in the recurrent cases, suggesting the existence of a particularly high susceptibility genetic profile among leprosy patients predisposing to disease recurrence.

Leprosy is a chronic infectious disease caused by *Mycobacterium leprae*, an obligatory intracellular, slow growing, alcohol-acid resistant bacillus with tropism for skin macrophages and Schwann cells of the peripheral nervous system¹. The disease affects more than 200.000 people every year globally, with high incidence and prevalence concentrating in India and Brazil². Leprosy presents a large spectrum of clinical manifestations that varies depending on the host's immune response, ranging from the localized tuberculoid to the systemic lepromatous forms as defined in the classic Ridley & Jopling classification system³. Alternative, the World Health Organization (WHO) classification protocol distributes leprosy cases as paucibacillary (PB) or multibacillary (MB) for treatment purposes⁴.

Leprosy is treated by a multidrug therapy (MDT) consisting of rifampicin, dapsone and clofazimine carried out for six monthly doses for PB patients and for 12 monthly doses for MB patients⁵; treatment is distributed for free in endemic countries since 1995⁶. Currently, there is the proposal for a uniform multidrug therapy regimen for leprosy patients that consists of six doses of MDT prescribed regardless of any clinical classification criteria. A Clinical Trial for Uniform Multidrug Therapy Regimen for Leprosy Patients in Brazil, (U-MDT/CT-BR) was recently conducted to compare outcomes of regular 12 doses MDT (R-MDT) and the uniform six doses MDT (U-MDT). The study included results from laboratory tests (bacilloscopic index, serology and histopathology) and clinical evaluation during long term follow-up⁷. Among the recurrent cases of leprosy identified in the U-MDT/CT-BR trial; for three of them, *M. leprae* isolates obtained in the first and second diagnosis were submitted to whole genome sequencing (WGS); analyses did not show any mutations associated with drug resistance.

¹Graduate Program in Health Sciences, School of Medicine, Pontifícia Universidade Católica do Paraná, Rua Imaculada Conceição 1155 - Prado Velho, Curitiba, Paraná, Brazil. ²Tropical Medicine Centre, University of Brasília, UnB - Brasília, DF, Brazil. ³Campus Universitário Darcy Ribeiro, S/N, Asa Norte, Brasília - DF, CEP, 70.904.970, Brasil. ⁴Escola Fiocruz de Governo, Fiocruz Brasília, Avenida L3 Norte, s/n, Campus Universitário Darcy Ribeiro, Gleba A, Brasília, Brazil. ⁵Tropical Pathology and Public Health Institute, Federal University of Goiás, Rua 235 - s/n - Setor Universitário - Goiania, Goiania, Goiás, Brazil. ⁶Dona Libânia Dermatology Centre, R. Pedro I, 1033 - Centro, Fortaleza, Ceará, Brazil. ⁷Tropical Dermatology and Venerology, Fundação Alfredo da Matta, Av. Codajás, 24 - Cachoeirinha, Manaus, Amazonas, Brazil. ⁸Lauro Souza Lima Institute, Rodovia Comandante João Ribeiro de Barros, km 225/226, s/n - Distrito Industrial Marcus Vinícius Feliz Machado, Bauru, São Paulo, Brazil. ⁹Epidemiology and Biostatistics Department, Universidade Federal Fluminense, Rua Marquês do Paraná, 303, Centro, Niterói, Rio de Janeiro, Brazil. ¹⁰Infectious Diseases and Immunity in Global Health (IDIGH) Program at the Research Institute of the McGill University Health Centre (RI-MUHC), 1001 Boulevard Décarie, Bioinformatics suite ES1.5561, H4A 3J1., Montreal, Quebec, Canada. *email: mmastefani@gmail.com; m.mira@puccpr.br

Gene	Marker	Risk Allele	Frequency on recurrent cases	Frequency on controls	Frequency on public database ^a	Cases			Controls		
						3208	1126*	2188	9162	9638	9536
<i>IL10</i>	rs1800871	A	0.67	0.50	0.43	AG	AA	AG	AA	AG	GG
<i>PARK2/PACRG</i>	rs2803073	G	0.17	0.33	0.34	AA	AG	AA	AG	AG	AA
	rs1040079	G	0.67	0.0	0.32	GG	AA	GG	AA	AA	AA
	rs2276201	C	0.67	0.0	0.41	CC	TT	CC	TT	TT	TT
	rs9356058	T	0.67	0.17	0.75	TT	CC	TT	CC	CT	CC
<i>SOD2</i>	rs4880	G	0.33	0.50	0.41	AG	AA	AG	AG	AA	GG
<i>NEBL</i>	rs625903	G	0.33	0.67	0.59	AA	AG	AG	AG	AG	GG
<i>CCDC122/LACCI</i>	rs2275252	C	0.17	0.33	0.61	AA	AC	AA	AC	AC	AA
	rs4942254	T	0.17	0.33	0.57	CC	CT	CC	CT	CT	CC
<i>NOD2</i>	rs8057341	G	1.00	0.67	0.52	GG	GG	GG	GG	AG	AG
	rs3135499	A	0.17	0.50	0.64	CC	AC	CC	AC	AC	AC
	rs2111234	A	0.83	0.67	0.51	AA	AG	AA	AA	AG	AG
<i>HLA-DRB1/DQA1</i>	rs602875	A	0.83	0.67	0.73	AA	AG	AA	AG	AG	AA
	rs1071630	T	0.67	0.50	0.45	TT	CC	TT	CT	CT	CT
<i>LTA</i>	rs2239704	A	0.50	0.50	0.35	AA	AC	CC	CC	AA	AC
	rs909253	A	0.67	0.67	0.61	AA	AG	AG	GG	AA	AA
<i>GATA3</i>	rs10905284	C	0.67	0.67	0.56	AC	AC	CC	AC	CC	AC
<i>IFNG</i>	rs2069727	T	0.83	0.50	0.73	CT	TT	TT	TT	CC	CT
<i>TLR1</i>	rs4833095	C	0.83	0.33	0.57	CC	CC	CT	CT	CT	TT

Table 1. Genotypes for all markers on all patients. A: adenine, G: guanine, C: cytosine, T: thymine. *Patient suspected of reinfection ^a: Data from 1000 genomes, phase 3, all superpopulations combined, obtained on Ensembl Genome Browser.

Based on comparative analysis between *M. leprae* genomes obtained in the first and second infection of each patient, the authors concluded for two cases of relapses caused by same strain and one case of reinfection caused by a different *M. leprae* strain⁸.

Considering that (i) all patients were followed and completed the treatment correctly, (ii) U-MDT was efficient in treating the disease, and (iii) no mutation associated with drug resistance was detected in the genomes of any *M. leprae* strain recovered from the three reinvestigated recurrent cases, it is reasonable to assume that leprosy recurrence was not due to characteristics of the etiologic agent but yet, to a combination of continuous exposure to *M. leprae* associated with an increased host susceptibility genetic background.

Host genetics accounts for a great proportion of disease heritability in leprosy and many genes were already identified with substantial evidences in support of their association with leprosy⁹. Among them, genes *HLA-DR-DQ* (major histocompatibility complex, class II, DR – class II DQ)¹⁰, *LTA* (lymphotoxin alpha)¹¹, *IL10* (interleukin 10)¹², *PRKN/PACRG* (parkin RBR E3 ubiquitin protein ligase/parkin coregulated gene)¹³, *NOD2* (nucleotide binding oligomerization domain containing 2)¹⁴, *LACCI/CCDC122* (laccase domain containing 1/coiled-coil domain containing 122)¹⁴, *SOD2* (superoxide dismutase 2)¹⁵, *NEBL* (nebullette) (unpublished data), *GATA3* (GATA binding protein 3)¹⁶, *IFNG* (interferon gamma)¹⁷ and *TLR1* (toll like receptor 1)¹⁸ have been associated with leprosy in Brazilian population samples, the majority of them playing important roles related to the host immune system. However, to date, no attempt has been made to translate this genetic knowledge into the definition of an individual susceptibility profile that could help explain complex leprosy phenotypes such as recurrence/reinfection. Here we genotyped 19 Single Nucleotide Polymorphism (SNP) distributed across these 11 *loci* associated with leprosy *per se* on six leprosy cases from the U-MDT/CT-BR cohort: two cases previously reported as relapses and one as reinfection⁸, and three matching, successfully treated leprosy patients that did not relapse. Allele frequencies were compared across the two groups and with publicly available databases.

Results and Discussion

All 19 SNPs were successfully genotyped for all six samples of leprosy patients and results are summarized in Table 1. A global count including all genes/markers assessed revealed that risk alleles were more frequently observed in recurrent patients #3208 and #2188, in contrast to controls and recurrent patient #1126.

Comparison of the combined genotypic distribution of all variants tested between the two groups revealed an enrichment of homozygous genotypes for the risk alleles among the three recurrent patients (Table 2): 40% vs. 19% among controls ($P = 0.02$; $OR = 2.83$; 95% $CI 1.22-6.58$). The same comparison applied for the combined allelic distribution also revealed an enrichment of risk alleles among cases, with a suggestive difference that failed to reach statistical significance at the 0.05 level (57% vs. 45% among controls; $P = 0.08$). These results must be interpreted with caution, given the extremely low sample size and the moderate to high degree of linkage disequilibrium across markers located within gene *loci* *PRKN/PACRG* ($r^2 = 1$ between rs1040079 and rs2276201 and $r^2 = 0.7$ between rs1040079 or rs2276201 and rs9356058), *LTA*, ($r^2 = 0.5$ between rs2239704 and rs909253), *NOD2* ($r^2 = 0.6$ between rs2111234 and rs8057341, and $r^2 = 0.66$ between rs2111234 and rs3135499), *HLA-DRB1/DQA1* ($r^2 = 0.46$ between rs602875 and rs1071630) and *CCDC122/LACCI* ($r^2 = 1$ between rs2275252 and rs4942254). Importantly: it is possible that high LD is in fact reflecting the additive effect of two or more variants of distinct

	Cases	Controls	3208	1126*	2188	9162	9638	9536
Risk alleles	65 (0.57)	51 (0.45)	24	18	23	18	18	15
Non-risk allele	49 (0.43)	63 (0.55)	14	20	15	20	20	23
Homozygous for the risk allele	23 (0.40)	11 (0.19)	10	4	9	4	3	4
Non-homozygous for the risk allele	34 (0.60)	46 (0.81)	9	15	10	15	16	15

Table 2. Count of risk allele on cases and controls. *Patient suspected of reinfection.

Patients	Age at First Diagnosis	Ethnic Group	Leprosy Classification	Leprosy Reaction	Time to Relapse (Months)
Case/Relapse Patients					
1126*	32	Mixed	BL	T2R/Neuritis/T1R	46
3208	17	Mixed	LL	Neuritis/T1R/T2R	86
2188	20	Mixed	LL	T1R/T2R	79
Control/Patients Without Relapses					
9162	35	Mixed	BL	T1R	NA
9536	32	Mixed	LL	Neuritis/T2R	NA
9638	29	Mixed	LL	T2R	NA

Table 3. Demographic and diagnostic characteristics of leprosy patients with/without recurrence.

Abbreviations: BL: Borderline Leprosy, LL: Lepromatous Leprosy, T1R: Type 1 Reaction, T2R: Type 2 Reaction, ND: not declared, NA: not applicable. *Patient reported as reinfection.

functional impact upon the net risk for that gene. This effect would be similar to what has been observed previously for gene *TNFSF8*¹⁹; extensive LD has been detected between associated markers identified as independent expression quantitative trait loci (eQTL) for *TNFSF8*; the authors argue that this combined functional impact could reflect the high correlation observed between the SNP alleles.

It is interesting to note that recurrent patient #1126 presented allelic and genotypic global counts similar to the controls and in contrast with the two other recurrent individuals. Patient #1126 has been described previously as the only case of recurrence caused by distinct *M. leprae* strains isolated in the first and the second diagnosis⁸. Also, patient #1126 presented the shortest time to relapse and the latest age of diagnosis – 32 years – similar to the controls (Table 3). One cannot disregard the possibility that recurrent patients #3208 and #2188 have also suffered reinfection – and not relapse – caused by the same *M. leprae* strain. This hypothesis is reinforced by the fact that none of the *M. leprae* strains have the classically recognized resistance mutations⁸, thus relapse caused by treatment failure is unlikely. Thus, the most striking host-related, non-genetic difference between recurrent patients #1126 and #3208 and #2188 is the age of diagnosis. In leprosy, the presence of the age-dependent association has already been shown for *PRKN-PACRG*²⁰, *LTA*¹¹, *miR-146a*²¹ (leprosy *per se*) and *TNFSF15-TNFSF8*²² (leprosy reaction). The enrichment of risk alleles observed for recurrent patients #3208 and #2188 is compatible with the hypothesis that early onset leprosy is more heavily dependent on genetic mechanisms, whereas late onset disease is a balanced combination of genetic and environmental factors.

A higher frequency of the risk allele is observed for markers of several well-known leprosy susceptibility genes – *IL10*, *PACRG*, *NOD2*, *HLA-DRB1/DQA1*, *LTA*, *GATA3*, *IFNG* and *TLR1* – when relapse cases are compared with controls and the frequency available on public databases (Table 1). This observation, although interesting, has only qualitative value in the context of a description of a series of cases and should be interpreted with extreme caution.

Materials and Methods

Ethics approval. This study was performed following the Declaration of Helsinki and Brazilian research regulations and was approved by the National Ethics Commission of Research (CONEP) of the Ministry of Health, protocol number 12949/2007, as previously described⁷. Informed consent was obtained from all individual participants included in the study.

Subjects. Six male leprosy patients were selected from the previous U-MDT/CT-BR study; three did not relapse during an eight-year follow-up; two were diagnosed as relapse caused by the same *M. leprae* strain and one was considered a reinfection caused by a distinct *M. leprae* isolate⁸. The three non-recurrent patients were matched with the recurrent cases by leprosy clinical form, ethnicity (estimated based on phenotypical features and skin color and distributed as Caucasian, African, Asian, Native American or Mixed), bacilloscopic index, treatment protocol, occurrence of leprosy reaction and place of residence (all six patients were from a leprosy hyperendemic area of Fortaleza, Brazil). A summary of the demographic and diagnostic information is presented in Table 3.

Gene selection, genotyping and analysis. Genes were selected for genotypic analysis based on previous evidence for validated/replicated association with leprosy in Brazilian population samples. SNP markers have been selected based on previous published, association with leprosy in one or more Brazilian population samples and/or in previous genome-wide association studies (GWAS) (Supplementary Table S1). Genomic DNA was obtained from peripheral blood by salting-out²³; the DNA was quantified spectrophotometrically using NanoDrop 2000/2000c and the concentration of genomic DNA was normalized to 20 ng/μL. Genotyping was performed by fluorescence-based allelic discrimination using TaqMan, as implemented in the Applied Biosystems® 7500 Real-Time PCR platform. The reactions mix was composed by ultrapure water (1.9 μL per sample), TaqMan® SNP Genotyping Assays (40×, 0.1 μL per sample), TaqMan™ Genotyping Master Mix (2×, 3 μL per sample) and template DNA (3 μL). Reaction was conducted using a denaturing temperature of 95 °C for 15 seconds and an oligonucleotide hybridization and extension temperature of 60 °C for 2 minutes on each of a total of 50 cycles.

Linkage disequilibrium (LD) was estimated using Haploview v.4.2 using the combined dataset of all six patients. A descriptive analysis was performed using the Statistical Package for the Social Sciences (SPSS) v.22. Genetic profiles of recurrent and non-recurrent cases were described against the frequencies of risk/protection alleles obtained from previous studies as available on Ensembl Genome Browser with data of 1000 genomes, phase 3, combined population. The Fisher's exact test for small sample sizes was used to compare combined allelic and genotypic frequencies across the two groups.

Conclusion

Our study presents, for the first time, suggestive evidence that genes associated with susceptibility to leprosy may also play an important role on disease recurrence. Our data indicates that a combination of alleles in different genes may confer hyper-susceptibility to leprosy detectable even among leprosy susceptible individuals, increasing the chances of disease recurrence. In addition, one could speculate, based on the allelic profile of patient #1126 – similar to the non-recurrent and distinct of the other two recurrent individuals – on the possible existence of distinct genetic mechanisms controlling disease relapse and reinfection. We are fully aware of the limitations of our design, particularly related to the small sample size; thus, this mainly descriptive, qualitative report of a series of cases must be perceived as a hypothesis-generating initiative, opening the possibility for additional studies of larger sample sizes of non-recurrent and recurrent leprosy affected individuals, aiming to identify biomarkers of risk of reinfection that may be used to monitor treated patients continuously exposed to the bacilli in endemic areas, with potential positive effects over leprosy control programs.

Data availability

The authors state that all data produced in the study is available.

Received: 20 May 2019; Accepted: 30 December 2019;

Published online: 28 January 2020

References

- Britton, W. J. & Lockwood, D. N. Leprosy. *Lancet* **363**, 1209–1219, [https://doi.org/10.1016/S0140-6736\(04\)15952-7](https://doi.org/10.1016/S0140-6736(04)15952-7) (2004).
- WHO, W. H. O. Weekly epidemiological record, Global leprosy update, 2017: reducing the disease burden due to leprosy (2018).
- Ridley, D. S. & Jopling, W. H. Classification of leprosy according to immunity. A five-group system. *International journal of leprosy and other mycobacterial diseases: official organ of the International Leprosy Association* **34**, 255–273 (1966).
- World Health Organization. A guide for surveillance of antimicrobial resistance in leprosy: 2017 update. (2017).
- WHO/Department of Communicable Disease Prevention, C. a. E. 38 (2000).
- WHO. *Multidrug therapy (MDT)*, <http://www.who.int/lep/mdt/en/> (2018).
- Penna, G. O. *et al.* Uniform multidrug therapy for leprosy patients in Brazil (U-MDT/CT-BR): Results of an open label, randomized and controlled clinical trial, among multibacillary patients. *PLoS neglected tropical diseases* **11**, e0005725, <https://doi.org/10.1371/journal.pntd.0005725> (2017).
- Stefani, M. M. A. *et al.* Whole genome sequencing distinguishes between relapse and reinfection in recurrent leprosy cases. *PLoS neglected tropical diseases* **11**, doi:ARTN e000559810.1371/journal.pntd.0005598 (2017).
- Sauer, M. E. *et al.* Genetics of leprosy: expected and unexpected developments and perspectives. *Clinics in dermatology* **33**, 99–107, <https://doi.org/10.1016/j.clindermatol.2014.10.001> (2015).
- da Silva, S. A. *et al.* HLA-DR and HLA-DQ alleles in patients from the south of Brazil: markers for leprosy susceptibility and resistance. *BMC infectious diseases* **9**, 134, <https://doi.org/10.1186/1471-2334-9-134> (2009).
- Alcais, A. *et al.* Stepwise replication identifies a low-producing lymphotoxin-alpha allele as a major risk factor for early-onset leprosy. *Nature genetics* **39**, 517–522, <https://doi.org/10.1038/ng2000> (2007).
- Alvarado-Arnez, L. E. *et al.* Association of IL10 Polymorphisms and Leprosy: A Meta-Analysis. *PLoS one* **10**, e0136282, <https://doi.org/10.1371/journal.pone.0136282> (2015).
- Mira, M. T. *et al.* Susceptibility to leprosy is associated with PARK2 and PACRG. *Nature* **427**, 636–640, <https://doi.org/10.1038/nature02326> (2004).
- Sales-Marques, C. *et al.* NOD2 and CCDC122-LACC1 genes are associated with leprosy susceptibility in Brazilians. *Human genetics* **133**, 1525–1532, <https://doi.org/10.1007/s00439-014-1502-9> (2014).
- Ramos, G. B. *et al.* Association Analysis Suggests SOD2 as a Newly Identified Candidate Gene Associated With Leprosy Susceptibility. *The Journal of infectious diseases* **214**, 475–478, <https://doi.org/10.1093/infdis/jiw170> (2016).
- Medeiros, P. *et al.* The GATA3 gene is involved in leprosy susceptibility in Brazilian patients. *Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases* **39**, 194–200, <https://doi.org/10.1016/j.meegid.2016.01.015> (2016).
- Cardoso, C. C. *et al.* IFNG +874T>A single nucleotide polymorphism is associated with leprosy among Brazilians. *Human genetics* **128**, 481–490, <https://doi.org/10.1007/s00439-010-0872-x> (2010).
- Marques Cde, S. *et al.* Toll-like receptor 1 N248S single-nucleotide polymorphism is associated with leprosy risk and regulates immune activation during mycobacterial infection. *The Journal of infectious diseases* **208**, 120–129, <https://doi.org/10.1093/infdis/jit133> (2013).
- Fava, V. M. *et al.* Association of TNFSF8 regulatory variants with excessive inflammatory responses but not leprosy per se. *The Journal of infectious diseases* **211**, 968–977, <https://doi.org/10.1093/infdis/jiu566> (2015).

20. Alter, A. *et al.* Linkage disequilibrium pattern and age-at-diagnosis are critical for replicating genetic associations across ethnic groups in leprosy. *Human genetics* **132**, 107–116, <https://doi.org/10.1007/s00439-012-1227-6> (2013).
21. Cezar-de-Mello, P. F. *et al.* Pre-miR-146a (rs2910164 G > C) single nucleotide polymorphism is genetically and functionally associated with leprosy. *PLoS neglected tropical diseases* **8**, e3099, <https://doi.org/10.1371/journal.pntd.0003099> (2014).
22. Fava, V. M., Sales-Marques, C., Alcais, A., Moraes, M. O. & Schurr, E. Age-Dependent Association of TNFSF15/TNFSF8 Variants and Leprosy Type 1 Reaction. *Frontiers in immunology* **8**, 155, <https://doi.org/10.3389/fimmu.2017.00155> (2017).
23. John, S. W., Weitzner, G., Rozen, R. & Scriver, C. R. A rapid procedure for extracting genomic DNA from leukocytes. *Nucleic acids research* **19**, 408, <https://doi.org/10.1093/nar/19.2.408> (1991).

Acknowledgements

This study was funding from UMDT Clinical Trial supported by Department of Science and Technology and National Council for Scientific and Technological Development (CNPq Number 403293/2005-7). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions

P.V.U.S.: molecular analysis, statistical analysis, manuscript preparation; G.O.P.: clinical assessment, study design, financial support, manuscript preparation; S.B.S.: clinical assessment, study design; M.A.A.P.: clinical assessment, study design; H.S.G.: clinical assessment; R.C.: clinical assessment; M.C.L.V.: clinical assessment; I.M.F.D.B.: molecular analysis. P.S.R.: molecular analysis. M.L.F.P.: clinical assessment, statistical analysis; V.M.F.: statistical analysis, manuscript preparation; M.M.A.S.: study design, senior supervisor, manuscript preparation; M.T.M.: study design, senior supervisor, manuscript preparation.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-58079-3>.

Correspondence and requests for materials should be addressed to M.M.A.S. or M.T.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020

ANEXO VI: ARTIGO

The Prata Village: An Estimate of the Incubation Period and the Proportion of Naturally Susceptible Individuals to Leprosy in Humans

The Prata Village: An Estimate of the Incubation Period and the Proportion of Naturally Susceptible Individuals to Leprosy in Humans

Running title: Leprosy susceptibility and incubation period

Ciane Cristina de Oliveira Mackert^{a,1}, Fernando Panissa Lázaro^a, Márcia Olandowski^a, Helena Regina Salomé D’Espindula^a, Priscila Verchai Uaska Sartori^a, Rafael Saraiva de Andrade Rodrigues^a, Geison Cambri^a, Marília Brasil Xavier^b, Erwin Schurr^c, Alexandre Alcaïs^d, Marcelo Távora Mira^a

^a *Graduate Program in Health Sciences - Pontifícia Universidade Católica do Paraná – R. Imac. Conceição, 1155 - Prado Velho, Curitiba - PR, 80215-901;*

^b *Department of Tropical Medicine - Universidade do Estado do Pará – Travessa Angustura, 2219 - Pedreira, Belém - PA, 66087-310;*

^c *Centre for the Study of Host Resistance - McGill University - 21,111 Lakeshore Road - Ste. Anne de Bellevue - Quebec, Canada H9X 3V9*

^d *Laboratory of Human Genetics of Infectious Diseases - Institut National de la Santé et de la Recherche Médicale and University Paris René Descartes - 12 Rue de l'École de Médecine, 75006 Paris, França*

¹ In memorian

Reprints and Correspondence: Dr. Marcelo Távora Mira, Graduate Program in Health Sciences, School of Medicine, Pontifícia Universidade Católica do Paraná. Rua Imaculada Conceição, 1155, Prado Velho; CEP: 80215-901, Curitiba, Paraná, Brazil. Fone: +55 41 3271-2030. E-mail: m.mira@pucpr.br

Abstract

Introduction: It has been widely accepted that only a proportion of individuals exposed to an infectious agent develop clinical disease after a period of incubation that can range from days to years. However, the exact estimation of the incubation period and the proportion of individuals naturally susceptible to an infectious disease in a human population is a major epidemiological challenge. Here we use leprosy as a model to infer these elusive epidemiological parameters, as obtained at the Prata Village, a hyper-endemic, former leprosy colony located in the outskirts of the Amazon forest, north of Brazil. By the time of enrollment, approximately half of the Prata population has been born at the Village, therefore, have been under constant, heavy and lifelong exposure to the disease.

Methods: Our enrolment strategy was complete ascertainment of the entire Prata population. Clinical, socio-economical and demographical data were obtained. Clinical variables were cross-checked using three independent information sources; validated data was used for a population-based, descriptive epidemiological analysis.

Results: From a total of 2,005 inhabitants by the time of the enrollment, 1,084 (56.19%) were born at the Village. Among this population subset, 5.9% have been affected by leprosy, with a median age of diagnosis of 15 years old. Given the unique characteristics of exposure of the colony-born population, we propose the first as an estimate of the proportion of leprosy susceptible individuals and the second an estimate of the incubation period of the disease.

Keywords: leprosy; population; epidemiology; incubation period; innate susceptibility

Significance Statement:

It is well known that only a proportion of individuals exposed to an infectious agent develops clinical disease, usually after an incubation period that can range from days to years. However, the exact estimation of the incubation period and the proportion of individuals naturally susceptible to an infectious disease in human populations is a very difficult, if not impossible, given that controlled experiment involving human inoculation and lifelong follow-up imposes obvious ethical limitations. Here we present a study involving a very peculiar, rare human population with a long history of exposure to leprosy that allowed for a rare, accurate estimate of the proportion of naturally susceptible individuals and the incubation period of the disease.

INTRODUCTION

It has been classically accepted that only a proportion of individuals exposed to an infectious agent develops clinical disease, usually after an incubation period that can range from weeks to years. However, the exact estimation of the incubation period and the proportion of individuals naturally susceptible to an infectious disease in human populations is a very difficult – if not impossible – task; a controlled experiment involving human inoculation and lifelong follow-up imposes obvious ethical limitations. Thus, estimates of incubation period and natural resistance to human infection are often speculative, based on experiments with animal models, epidemics in isolated population or on the observation of rare natural occurrences, such as accidental inoculations – for example, the Lübeck accident in 1930 (1). If a disease is restricted to a few species, such as leprosy that naturally affects only humans and armadillos (2), these estimates are even more difficult to obtain.

Leprosy is a chronic infectious disease caused by *Mycobacterium leprae*, an obligatory, highly specialized intracellular parasite of the skin macrophages and Schwann cells of the peripheral nervous system (3). Clinical manifestation of leprosy distributes on a spectrum displaying two polar – tuberculoid and lepromatous – and three borderline forms (4). The disease is treated through multidrug-therapy (MDT) following World Health Organization (WHO) protocol applied accordingly to number of lesions and bacillary load (5). According to the last WHO report, in 2016, the global incidence of leprosy was 214,783 cases (6).

Transmission of leprosy, though still poorly understood, is assumed to occur from a bacillary patient to a contact through aerial nasal secretions, over a long period of exposure (7, 8). More recently, additional sources of infection have been proposed, including soil (9), water (10), hematophagous insects (11) and infected armadillos (2, 12). Defining exposure is challenging since serological and molecular tests underestimate the real proportion of exposed individuals mainly due to lack of sensitivity or specificity to identify low bacillary patients (13, 14). It is often assumed that around 5% of the population is susceptible to leprosy (15, 16) and the disease is often regarded as having an incubation period of 5 years (5). An early attempt (1936) to estimate the incubation period of leprosy reached 8.4 years, though little information

about how the study was conducted is reported and period calculations were made only based on patient's self-reports (17). Data from secondary cases of leprosy household contacts reveals a mean time of leprosy onset of 4.3 (± 3.3) years independently of clinical form (18); the authors considered incubation as the time between the appearance of an index leprosy patient in the household and the appearance of a leprosy secondary case in a household contact. Nevertheless, the study was performed in a hyper endemic region and incubation period is probably reduced by the assumption that exposure occurs only through the index case living in the same household. A study on 35 military veterans revealed a median time of onset of 2-5 years for tuberculoid and 8-12 for lepromatous leprosy; however, the finding lacks statistical significance (19). Lastly, incubation period estimative in animal models are based on chimpanzees, a non-consensus model for leprosy studies (20). Similar, previous attempts to define the proportion of individuals naturally susceptible to leprosy also led to inconsistent results, ranging from less than 1% to 12% (15, 21). An ideal scenario to assess both susceptibility and incubation period for leprosy would involve an isolated geographic site under high endemicity for the disease, as the Village of Santo Antônio do Prata (Prata Village).

The Prata Village is a former leprosy colony located in the outskirts of the Amazon forest in the Northern Brazilian state of Pará. Franciscan monks have founded the Village in 1898; in the 1920's the site became a venue for compulsory isolation of leprosy patients diagnosed across the Northern and Northeastern states of Brazil. Isolation was mandatory until 1962 (22); however, the strong social stigma associated with the disease has been preventing both emigration of affected and immigration of non-affected families to the Village. As a result, the Prata Village remains socially and geographically isolated up to date, and despite the efforts to control the disease, leprosy is still highly prevalent and homogeneously distributed across the entire community (23). A recent genetic epidemiology study performed using the Prata families revealed the presence of a strong genetic component controlling susceptibility to leprosy in the Prata population, a finding compatible with the hypothesis that, due to the history of the Village, genetic risk factors for leprosy are enriched within this population (23). By the time of recruitment, all inhabitants, the vast majority descendants of the original isolated leprosy affected individuals, shared the same limited environment; for example, the Village has only one church, two elementary schools, one social club and one large central square, used by the entire population for leisure activities such as soccer games and outdoor celebrations. In this context, it is

reasonable to assume that all individuals are constantly and heavily exposed to *M. leprae*. Of particular interest is the sub-population that has been born at the Village; these individuals have been living their entire lives in a homogenous environment of constant, high exposure to leprosy.

Here we present a population-based epidemiological study involving the entire population of the Prata Village, with the objective to explore the population's unique characteristics and obtain accurate estimations for incubation period and proportion of naturally susceptible individuals to leprosy.

RESULTS

Out of 2,005 individuals enrolled, 1,672 self-reported as non-affected, 319 self-reported as affected by leprosy and for 14 individuals it was not possible to obtain this information. Data checking led to the confirmation of 257 out of the 319 self-reported cases. A prevalence rate (active and historical cases) of 1,282/10,000 was estimated for the total population (257 cases/2,005 individuals). A total of 76 individuals with unknown leprosy status (14, above mentioned + 62, unavailable medical reports) were removed from subsequent analyzes (Tab. Suppl. 01). A complete demographical, socio-economic and clinical characteristics description of the population is provided as supplementary material (Tab. Suppl. 01 to 03).

Close to half of the Prata population – 1,084 individuals (56.19%) were born at the Village; among these, 64 cases of leprosy were confirmed, resulting in a cumulative leprosy prevalence of 5.9%. The same estimate reached 22.9% among the individuals not born at the Village (193 confirmed cases/844 individuals) (Figure 1 and Tab. Suppl. 02).

The parameter “age at diagnosis” also behaved differently across the two sub-populations (Tab. Suppl. 03): the median age of diagnosis among the population born at the Village was 15 years old (mean = 18.17, ranging from 5-75 years old and

standard deviation of ± 12.95) in contrast with 28 years old observed for the not born at the Village (mean = 30.19, ranging from 2-74 years old and standard deviation of ± 15.25) (Figure 2).

DISCUSSION

It has been well known that infectious agents are necessary but not sufficient to explain clinical infection; however, due to obvious ethical issues, attempts to estimate the proportion of susceptible individuals has been often performed under biased strategies. Assuming that half of the Prata population has been born in an environment highly favorable to infection, our finding that 5.9% of these Prata-born individuals in fact became affected provides a rare formal estimate of this critical epidemiological parameter. In addition, assuming that all individuals born at the Prata Village are exposed to the disease since their birth, the median age of diagnosis among this sub-population is likely to be, to our knowledge, the best estimate of the disease's incubation period. Interestingly, 15 years is much higher than what has been widely accepted among epidemiologists. This may have an impact over public health, leprosy control programs, since such a large incubation period may reflect on a much longer persistence of incidence rates, as indeed observed for leprosy worldwide (24).

One interesting outcome of this study is the finding that although several leprosy reviews and studies present incubation period and natural susceptibility data (often 5 years and 5%, respectively), authors commonly cite textbooks or reviews and rarely, the original studies. Incubation period is often based on a book chapter (25) describing it as being between 2-5 years; nevertheless, careful analysis of the references used in the book reveals that the information is based on studies of extremely different experimental designs (17) resulting in findings ranging from 2-30 years, as described in the introduction.

We are aware of the limitations of our study. For our estimates to be exact, one has to assume equal and permanent exposure of all newborn individuals to *M. leprae* throughout their entire lives, which is obviously impossible to prove. For example, our strategy does not take into account the exact nature of each leprosy case; therefore, well known differential risk factors for infection have not been considered, such as the presence or not of an index case within the household of a new case (26), as well as the clinical form of disease of the index case. However, stratification of all Prata patients by its clinical form revealed that 104 (40.5%) cases were lepromatous, 53 (20.6%) tuberculoid, 53 (20.6%) borderline and 47 (18.3%) of indeterminate clinical form (Suppl. table 3). Therefore, a high proportion of the Prata patients are of higher bacillary load; considering that the observed 12.82% cumulative prevalence of leprosy in the Village is, to our knowledge, the highest ever reported for this infectious disease, these observations favors our assumption that the Prata population is under high exposure. Thus, despite the limitations, we believe that the unique historical, geographical, socio-economic and epidemiological characteristics of the Prata Village conspired to create a setting that is the closest possible to an ideal, experimental-like situation; our estimates, if not exact, are likely to be the best possible approximation for a human population.

It seems reasonable to assume that the Prata population is particularly susceptible to leprosy, due to an enrichment of genetic risk factors (23). This is compatible with the young age of diagnosis among the sub-population of the Prata-born cases: it has been demonstrated that strong genetic effects controlling susceptibility to leprosy are more clearly detected among cases with early age of onset (27, 28). If this increased genetic effect is true, the proportion of naturally susceptible individuals in the Prata is likely to be overestimated. On the same note, one could speculate that the increased natural predisposition would result in a shorter incubation period of disease; intriguingly, though, the incubation period of leprosy in the Prata seems to be much higher than what has been usually reported (17-19). Therefore, it is possible that, in the general population, leprosy incubation period is even longer than 15 years, maximizing the impact over the expectations of the public leprosy control programs.

To conclude, here we use leprosy as a model to estimate the proportion of naturally susceptible individuals, as well as the incubation period of the disease. The finding may

have extended impact, beyond leprosy, and may be of value to the advance on the understanding of the epidemiology of human infectious diseases in general.

SUBJECTS AND METHODS

Complete enrollment of the entire population of the Prata Village has been performed between 2006 and 2007 for a population-based Complex Segregation Analysis, and a detailed description of the enrollment strategy has been published elsewhere (23). In brief, a member of the research team has visited every household at the Village and all adult members of the resident families have been interviewed. The parents have provided information about individuals younger than 18 years old. Socio-economic data has been collected through a questionnaire. All individuals self-reported their present and previous leprosy status as affected or non-affected. All self-reported leprosy cases, active and historical, have been checked using three independent sources available at the Village's health unit: (i) the original medical records, (ii) copies of the compulsory notification forms and (iii) a log book containing registries of all cases for treatment follow-up. Self-reported cases confirmed at all three sources were tagged as a leprosy case. Upon confirmation of a leprosy case, clinical information was obtained from the medical records. All subjects enrolled in the study agreed to participate and signed a written informed consent that have been approved along with the research project by the Research Ethics Committee of the *Pontifícia Universidade Católica do Paraná* (51/07/CEP-PUCPR) and the Brazilian National Board for Ethics in Research (222/2006; 25000.001992/2006-17).

Descriptive statistical analysis was used to calculate the proportion of affected individuals, given as a percentage of the total and of the sub-populations that has been born and not born at the Village. Median and mean age at diagnosis was estimated using the Kaplan-Meyer and box plot. Given that this is a population-based study that involved the recruitment of the entire Prata village, no inferential statistics have been applied.

CONTRIBUTORS

CCOM, FPL and MBX: collected the data.

MO: performed the statistical analysis.

HRSD, PVUS, RSAR and GC: did the statistical analysis and wrote the article.

ES, AA and MTM: designed the study, provided senior supervision, reviewed and wrote the article.

Potential Conflict of Interests: none reported.

Acknowledgments

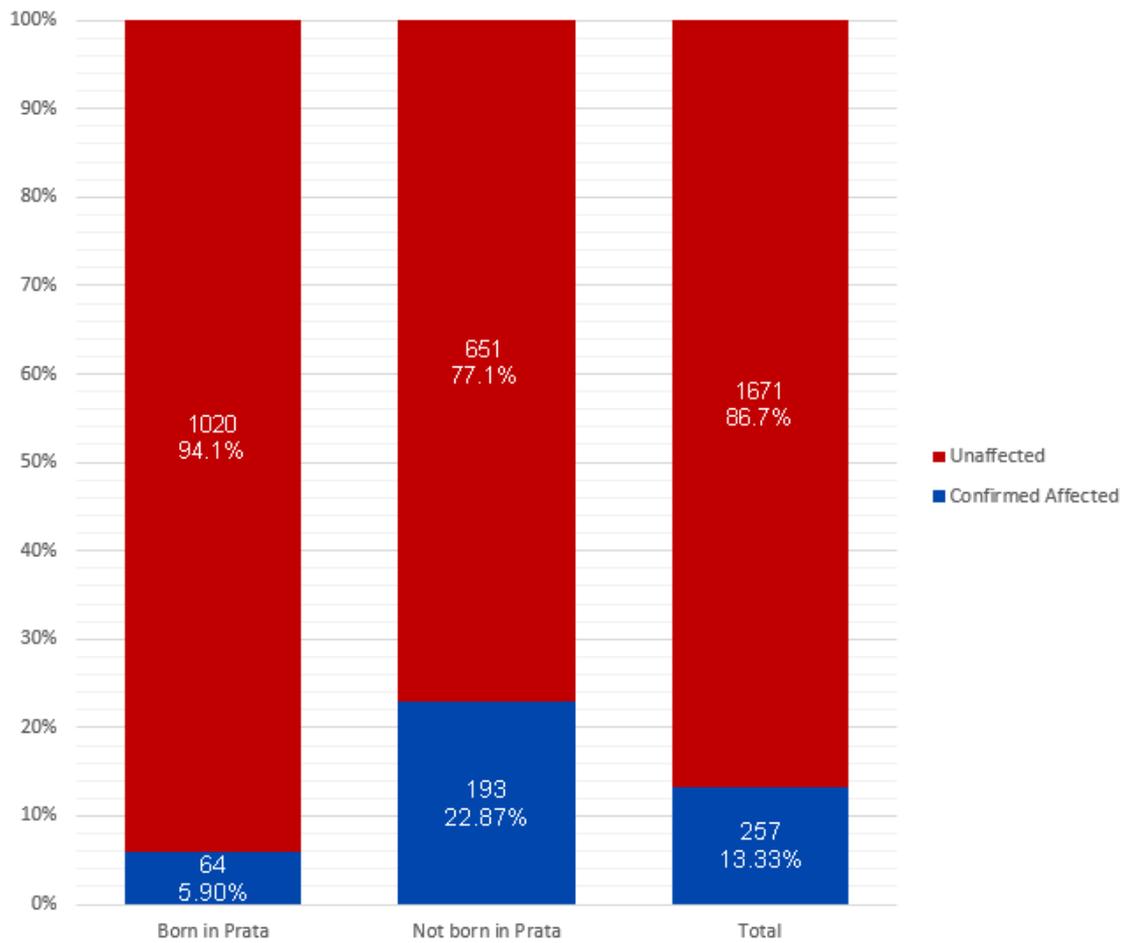
We thank all Prata Village population for acceptance in participating the study. This study was supported by grants of Unicef/UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases (TDR), MCT/CNPq/MS-SCTIE-DECIT 35/2005 – “Estudos da Hanseníase” and MCT/CNPq/MS-SCTIE-DECIT 25/2006 – “Estudo de Doenças Negligenciadas”.

References

1. Fox GJ, Orlova M, & Schurr E (2016) Tuberculosis in Newborns: The Lessons of the "Lubeck Disaster" (1929-1933). *PLoS pathogens* 12(1):e1005271.
2. Truman R (2005) Leprosy in wild armadillos. *Leprosy review* 76(3):198-208.
3. Scollard DM, *et al.* (2006) The continuing challenges of leprosy. *Clinical microbiology reviews* 19(2):338-381.
4. Ridley DS & Jopling WH (1966) Classification of leprosy according to immunity. A five-group system. *International journal of leprosy and other mycobacterial diseases : official organ of the International Leprosy Association* 34(3):255-273.
5. World Health Organization (Updated April 2016) Leprosy. Available from: <http://www.who.int/mediacentre/factsheets/fs101/en/>.
6. World Health Organization (2017) Global leprosy update, 2016: accelerating reduction of disease burden. *Releve epidemiologique hebdomadaire / Section d'hygiene du Secretariat de la Societe des Nations = Weekly epidemiological record / Health Section of the Secretariat of the League of Nations* 92(35):501-519.
7. Torres P, *et al.* (2003) Comparison of PCR mediated amplification of DNA and the classical methods for detection of Mycobacterium leprae in different types of clinical samples in leprosy patients and contacts. *Leprosy review* 74(1):18-30.
8. Davey TF & Rees RJ (1974) The nasal discharge in leprosy: clinical and bacteriological aspects. *Leprosy review* 45(2):121-134.
9. Lavania M, *et al.* (2008) 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* 8(5):627-631.
10. Matsuoka M, Izumi S, Budiawan T, Nakata N, & Saeki K (1999) Mycobacterium leprae DNA in daily using water as a possible source of leprosy infection. *Indian journal of leprosy* 71(1):61-67.
11. Neumann Ada S, *et al.* (2016) Experimental Infection of Rhodnius prolixus (Hemiptera, Triatominae) with Mycobacterium leprae Indicates Potential for Leprosy Transmission. *PLoS one* 11(5):e0156037.
12. Sharma R, *et al.* (2015) Zoonotic Leprosy in the Southeastern United States. *Emerging infectious diseases* 21(12):2127-2134.
13. Penna ML, Penna GO, Iglesias PC, Natal S, & Rodrigues LC (2016) 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* 10(5):e0004703.

14. Martinez AN, Talhari C, Moraes MO, & Talhari S (2014) PCR-based techniques for leprosy diagnosis: from the laboratory to the clinic. *PLoS neglected tropical diseases* 8(4):e2655.
15. Joyce MP (2012) Historic aspects of human susceptibility to leprosy and the risk of conjugal transmission. *Memorias do Instituto Oswaldo Cruz* 107 Suppl 1:17-21.
16. Newell KW (1966) An epidemiologist's view of leprosy. *Bull World Health Organ* 34(6):827-857.
17. Bechelli LM (1936) O tempo de incubação da lepra. *Rev Bras Leprol* 4:355-360.
18. Prasad KV & Ali PM (1967) Incubation period of leprosy. *The Indian journal of medical research* 55(1):29-42.
19. Fine PE (1982) Leprosy: the epidemiology of a slow bacterium. *Epidemiologic reviews* 4:161-188.
20. Suzuki K, Tanigawa K, Kawashima A, Miyamura T, & Ishii N (2011) Chimpanzees used for medical research shed light on the pathoetiology of leprosy. *Future microbiology* 6(10):1151-1157.
21. Sales AM, *et al.* (2011) Leprosy among patient contacts: a multilevel study of risk factors. *PLoS neglected tropical diseases* 5(3):e1013.
22. Rosa Maciel L, Oliveira ML, Gallo ME, & Damasco MS (2003) Memories and history of Hansen's disease in Brazil told by witnesses (1960-2000). *Historia, ciencias, saude--Manguinhos* 10(Suppl 1):308-335.
23. Lazaro FP, *et al.* (2010) A major gene controls leprosy susceptibility in a hyperendemic isolated population from north of Brazil. *The Journal of infectious diseases* 201(10):1598-1605.
24. World Health Organization (2016) Global Leprosy Update, 2015: Time For Action, Accountability and Inclusion. *Weekly Epidemiological Record* :405-420.
25. Noordeen SK (1994) Elimination of leprosy as a public health problem. *International journal of leprosy and other mycobacterial diseases : official organ of the International Leprosy Association* 62(2):278-283.
26. Fine PE, *et al.* (1997) Household and dwelling contact as risk factors for leprosy in northern Malawi. *American journal of epidemiology* 146(1):91-102.
27. Alter A, *et al.* (2013) Linkage disequilibrium pattern and age-at-diagnosis are critical for replicating genetic associations across ethnic groups in leprosy. *Human genetics* 132(1):107-116.
28. Alcais A, *et al.* (2007) Stepwise replication identifies a low-producing lymphotoxin-alpha allele as a major risk factor for early-onset leprosy. *Nature genetics* 39(4):517-522.

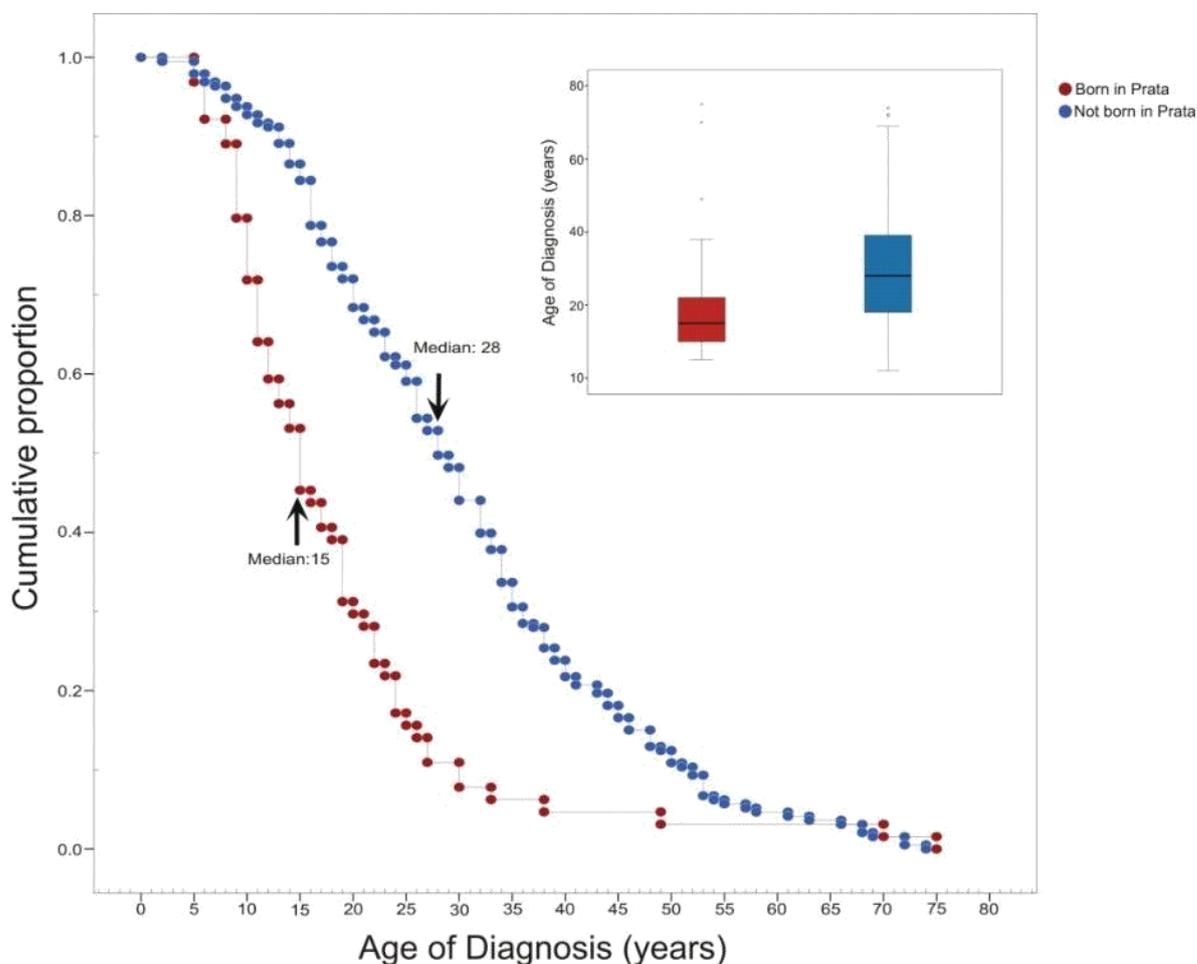
Figure 1 - Proportion of affected individuals in the total population and the sub-populations of individuals born and not born at the Prata Village.



Legend:

Numbers inside the bars: total number of individuals (proportion); Total: entire population of the Prata Village, as previously published on (23); Born in Prata: sub-population of individuals that has been born at the Village; Not born in Prata: sub-population of individuals that has not been born at the Village.

Figure 2 - Kaplan-Meier plot for age at diagnosis among affected individuals of the Prata Village, according to the place of birth (born at the Village; not born at the Village).



Legend:

Not born in Prata (blue): sub-population of individuals that has not been born at the Prata Village; Born in Prata (red): sub-population of individuals that has been born at the Prata Village. Kaplan-Meier plot : Y-axis: cumulative proportion of non-affected individuals; Median: median age at diagnosis. Box plot: the back band inside the boxes are the median; the bottom and top of the boxes represents the first and third quartiles respectively; the ends of the whiskers represents the 10th and the 90th percentile.

Supplementary Table 1 – Epidemiological characterization of the confirmed affected and unaffected subjects in Prata Village

	Confirmed Affected Subjects		Unaffected Subjects		Total Subjects	
	n		n		n	
Total Sample Size (n)	257		1672		1929	
Age (years) *	n		n		n	
Total individuals with information available	257		1668		1925	
Mean	47.88		20.37		24.04	
Median	49		16		19	
Minimum	10		<1		<1	
Maximum	89		82		89	
Standard deviation	19.99		15.99		19.03	
Sex	n	%	n	%	n	%
Male	153	59.5%	819	49.0%	972	50.4%
Female	104	40.5%	853	51.0%	957	49.6%
Ethnicity	n	%	n	%	n	%
White	27	10.5%	171	10.2%	198	10.3%
Black	50	19.5%	195	11.7%	245	12.7%
Mixed Race	180	70.0%	1301	77.8%	1481	76.8%
Others	0	0.0%	4	0.2%	4	0.2%
NA	0	0.0%	1	0.1%	1	0.1%
Marital Status	n	%	n	%	n	%
Single	84	48.6%	1147	68.6%	1231	63.8%
Stable Union / Married	139	80.3%	474	28.3%	613	31.8%
Divorced	2	1.2%	9	0.5%	11	0.6%

Widowed	31	17.9%	38	2.3%	69	3.6%
Others	1	0.6%	3	0.2%	4	0.2%
NA	0	0.0%	1	0.1%	1	0.1%
Birthplace	n	%	n	%	n	%
Prata	64	24.9%	1020	61.0%	1084	56.2%
Others	193	75.1%	651	38.9%	844	43.8%
NA	0	0.0%	1	0.1%	1	0.1%
Education	n	%	n	%	n	%
Illiterate	66	25.7%	86	5.1%	152	7.9%
Incomplete Elementary School	92	35.8%	590	35.3%	682	35.4%
Complete Elementary School	19	7.4%	71	4.2%	90	4.7%
Incomplete Middle School	39	15.2%	312	18.7%	351	18.2%
Complete Middle School	7	2.7%	71	4.2%	78	4.0%
Incomplete High School	13	5.1%	110	6.6%	123	6.4%
Complete High School	11	4.3%	111	6.6%	122	6.3%
More than High School	3	1.2%	276	16.5%	279	14.5%
NA	7	2.7%	45	2.7%	52	2.7%
Alcohol drink	n	%	n	%	n	%
Yes	89	34.6%	337	20.2%	426	22.1%
No	168	65.4%	1335	79.8%	1503	77.9%
Smoking	n	%	n	%	n	%
Yes	92	35.8%	292	17.5%	384	19.9%
No	165	64.2%	1380	82.5%	1545	80.1%
Drug User	n	%	n	%	n	%
Yes	5	1.9%	25	1.5%	30	1.6%

No	249	96.9%	1635	97.8%	1884	97.7%
NA	3	1.2%	12	0.7%	15	0.8%

Legend:

n = number.

Supplementary Table 2 – Epidemiological characterization by birthplace of the subjects in the study

	Born in Prata		Not born in Prata	
	n		n	
Total Sample Size (n)	1084		844	
Status	n	%	n	%
Confirmed Affected Subjects	64	5.9%	193	22.9%
Unaffected Subjects	1020	94.1%	651	77.1%
Age (years)	n		n	
Total individuals with information available	1084		840	
Mean	13.52		37.62	
Median	12		36	
Maximum	0		2	
Minimum	80		89	
Standard deviation	10.48		19.00	
Sex	n	%	n	%
Male	558	51.5%	413	48.9%
Female	526	48.5%	431	51.1%
Ethnicity				
White	110	10.1%	88	10.4%
Black	119	11.0%	126	14.9%
Mixed Race	854	78.8%	626	74.2%
Others	0	0.0%	4	0.5%
NA	1	0.1%	0	0.0%
Marital Status	n	%	n	%
Stable Union / Married	146	13.5%	466	55.2%
Single	936	86.3%	295	35.0%

Divorced	2	0.2%	9	1.1%
Widowed	0	0.0%	69	8.2%
Others	0	0.0%	4	0.5%
NA	0	0.0%	1	0.1%
Education	n	%	n	%
Illiterate	14	1.3%	138	16.4%
Incomplete Elementary School	375	34.6%	307	36.4%
Complete Elementary School	38	3.5%	52	6.2%
Incomplete Middle School	213	19.6%	138	16.4%
Complete Middle School	35	3.2%	43	5.1%
Incomplete High School	65	6.0%	58	6.9%
Complete High School	59	5.4%	63	7.5%
More than High School	253	23.3%	26	3.1%
NA	32	3.0%	19	2.3%
Alcohol drink	n	%	n	%
Yes	148	13.7%	278	32.9%
No	936	86.3%	566	67.1%
Smoking	n	%	n	%
Yes	112	10.3%	272	32.2%
No	972	89.7%	572	67.8%
Recreative Drug User	n	%	n	%
Yes	12	1.1%	18	2.1%
No	1067	98.4%	816	96.7%
NA	5	0.5%	10	1.2%

Legend:

n = number.

Supplementary Table 3 – Clinical characterization of the confirmed affected subjects in Prata Village

Confirmed Affected Subjects

Total Sample Size (n)	257
Age of Diagnosis (years)	
Total individuals with information available	257
Mean	27.20
Median	25
Minimum	2
Maximum	75
Standard deviation	15.58
Age of Diagnosis - Born in Prata (years)	
Total individuals with information available	64
Mean	18.17
Median	15
Minimum	5
Maximum	75
Standard deviation	12.95
Age of Diagnosis - Not born in Prata (years)	
Total individuals with information available	193
Mean	30.19
Median	28
Minimum	2
Maximum	74
Standard deviation	15.25
Clinical Classification	
Tuberculoid	53 (20.6%)

Borderline	53 (20.6%)
Lepromatous	104 (40.5%)
Indeterminate	47 (18.3%)

Legend:

n = number;