



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ
Pós-Graduação em Biologia Celular e Molecular

Ohanna Cavalcanti de Lima Bezerra

**Polimorfismos selecionados por patógenos em
genes do metabolismo energético e associação
com doenças micobacterianas**

Orientador: Dr. Milton Ozório Moraes

Co-orientador: Dr. Antonio Guilherme Pacheco

RIO DE JANEIRO

Dezembro, 2019



Ministério da Saúde

FIOCRUZ
Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ
Pós-Graduação em Biologia Celular e Molecular

Ohanna Cavalcanti de Lima Bezerra

Polimorfismos selecionados por patógenos em genes do metabolismo energético e associação com doenças micobacterianas

Tese apresentada ao Instituto Oswaldo Cruz como parte dos requisitos para obtenção do título de Doutor em Biologia Celular e Molecular, na área de Genética Epidemiológica e Evolutiva.

Orientador: Dr. Milton Ozório Moraes

Co-orientador: Dr. Antonio Guilherme Pacheco

RIO DE JANEIRO
Dezembro, 2019

Cavalcanti de Lima Bezerra, Ohanna.

Polimorfismos selecionados por patógenos em genes do metabolismo energético e associação com doenças micobacterianas / Ohanna Cavalcanti de Lima Bezerra. - Rio de Janeiro, 2019.

223 f.

Tese (Doutorado) - Instituto Oswaldo Cruz, Pós-Graduação em Biologia Celular e Molecular, 2019.

Orientador: Milton Ozório Moraes.

Co-orientador: Antonio Guilherme Pacheco.

Bibliografia: f. 154-178

1. associação genética. 2. polimorfismos. 3. PKLR, G6PD. 4. ancestralidade. 5. micobactérias. I. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da Biblioteca de Mangueiras/ICICT com os dados fornecidos pelo(a) autor(a).



Ministério da Saúde

FIOCRUZ
Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ
Pós-Graduação em Biologia Celular e Molecular

Ohanna Cavalcanti de Lima Bezerra

Polimorfismos selecionados por patógenos em genes do metabolismo energético e associação com doenças micobacterianas

Orientador: Dr. Milton Ozório Moraes

Co-orientador: Dr. Antonio Guilherme Pacheco

EXAMINADORES:

Dr. Fernando Regla Vargas – Fundação Oswaldo Cruz (IOC/FIOCRUZ) – Presidente

Dra. Cynthia Chester Cardoso – Universidade Federal do Rio de Janeiro (UFRJ) – Titular/Revisora

Dr. Diogo Meyer – Universidade de São Paulo (USP) – Titular

Dra. Cristiana Santos de Macedo – Fundação Oswaldo Cruz (CDTS/FIOCRUZ) – Suplente

Dra. Sayonara Maria de Carvalho Gonzalez – Fundação Oswaldo Cruz (IFF/FIOCRUZ) – Suplente

RIO DE JANEIRO

13 de dezembro de 2019

“Seja aquilo que você quer ver no mundo”

Mahatma Gandhi.

À minha família, pelo incentivo e amor incondicional.

Agradecimentos

A Deus, por me fortalecer e guiar na longa jornada até aqui.

À minha mãe, meu exemplo e inspiração, pelo amor incondicional e dedicação à minha formação, e ao meu pai, irmão e “paidraastro”, Amilton, Hallan e Thomas, pelo amor, segurança, por encorajar minha profissão, minhas decisões e me transmitir força e apoio mesmo à distância.

Ao meu avô, Manoel, fonte da minha força, que está sempre ao meu lado e acompanha o final dessa tese no céu. À minha avó, Maria, meu exemplo, minha alegria semanal, pelo amor que não mede esforços. Vocês são meu lar e minha inspiração! Agradeço à minha família, que sempre estiveram perto, me dando forças e acreditando no meu sucesso.

Ao meu amor e pesquisador favorito, Valter, por me trazer os melhores sorrisos, pelo amor diário, pelo apoio ao desenvolvimento desse trabalho, pela alegria, paz e o sentido que traz em minha vida. Aos meus sogros, Marlene e Valter Fernando, que trazem a simplicidade do meu bem-estar, a saudade e o amor que transcende mundos.

Ao meu orientador, Milton Ozório Moraes, pelo ensinamento desde a iniciação científica, pelas ricas discussões e confiança. Agradeço a oportunidade da orientação que me fez amadurecer como profissional ao longo desses anos. Meus eterno reconhecimento e sinceros agradecimentos.

Ao meu co-orientador, Antonio Guilherme Pacheco, pela disponibilidade, confiança, desenvolvimento das análises, sugestões valiosas e revisão cuidadosa dos resultados deste trabalho.

Às minhas amigas e mentoras, Lucia Elena e Carolinne Marques, por me acompanharem desde o início, me darem suporte e participarem diretamente da minha formação, além da revisão crítica do artigo e do debate dos resultados, e à Fernanda Kehdy e Fernanda Manta, pela tamanha contribuição nessa tese, desde o delineamento da hipótese formulada, aos experimentos ensinados, desenvolvimento das análises e enriquecimento das discussões. Tenho um pouco de cada uma de vocês na minha formação!

À Cynthia Chester, pela nobre contribuição nas revisões dos trabalhos e pela disponibilidade em auxiliar o grupo sempre.

Ao Dr. Diogo Meyer, pelos ensinamentos, por me orientar no delineamento das análises de seleção e me receber de braços abertos no Laboratório de Genética Evolutiva (Instituto de Biociências/USP).

À Kelly Nunes, pela grande ajuda no desenvolvimento das análises de seleção, pelas discussões, skypes, pelas caronas, boa companhia e disposição em se aprofundar no meu estudo. Agradeço também ao Vitor Aguiar, pelas sugestões e companheirismo, e aos meninos da iniciação científica do Laboratório de Genética Evolutiva.

Aos colaboradores, Jamile Leão Rêgo, André Leturiondo, Ariani Batista, Fabíola Rodrigues, Rafaela Mota, Ana Carla Pereira Latini, Nédio Mabunda, Amina de Sousa, Graça Salomé e Ricardo Pereira, pelo envio das amostras utilizadas nesse estudo e pelo auxílio nos trâmites e informações gerais das populações.

Ao Dr. Eduardo Tarazona Santos (Departamento de Biologia Geral/ Instituto de Ciências Biológicas da UFMG) que permitiu o acesso ao servidor e aos dados do projeto EPIGEN.

Aos colaboradores Mateus Gouveia (Center for Research on Genomics & Global Health/NIH, EUA) e Vinicius Fava (McGill University Health Centre, Canadá), por compartilharem dados genômicos (alguns ainda não publicados) que enriqueceram a discussão desse trabalho.

Ao Dr. Elizeu Fagundes e à Dra. Leonor Gusmão por abrirem as portas do Laboratório de Diagnóstico por DNA (UERJ) para as análises de ancestralidade.

À Dra. Roberta Olmo, por auxiliar na colaboração das dosagens hematológicas, e ao Dr. Adalberto Rezende, por compartilhar os reagentes e o sistema computacional para análise de fragmentos.

Aos meus alunos de iniciação científica, Laís Pereira Ferreira e Vinicius Maia da Cruz, por me trazerem o desafio do ensino e auxiliarem nos experimentos, em especial a Laís, que acompanhou, sempre disposta, o estudo até o final. Parcelas consideráveis desse trabalho foram desenvolvidas porque você esteve para me ajudar! Agradeço também à minha aluna “não oficial”, Ariani Batista, pelo auxílio, carinho e preocupação (mesmo de Manaus).

Aos amigos do laboratório, Suelen, Rafaela, Isabela, Brunna, Rychelle, Mayara, Thyago, PT, Meydson, Thadeu, Valcemir, Tamiris, Andrea, Paula, Thiago, Leonardo, Isabelle, Priscilla, Nath, Alexandre, Yerly e Ricardo, Vivian e à amiga que se foi, mas mora em meu coração, Renuka, pelas boas lembranças, pelas conversas, momentos de risadas, discussões, ajudas indispensáveis, pelo convívio saudável e pela boa companhia no dia-a-dia. Agradeço em especial a Rychelle, por ter discutido e compartilhado parte dos resultados dessa tese que envolve uma hipótese conjunta; Thyago, por ter se debruçado em auxiliar nos ensaios de expressão gênica e nas análises gerais, e Isabela, por auxiliar na bancada/análises e nas discussões sobre genética. Vocês sempre me estenderam a mão, sou eternamente grata!

Aos demais colegas que passaram pelo LAHAN, aos integrantes do corpo do Pavilhão da Hanseníase e do Ambulatório Souza Araújo pela adorável convivência e momentos de descontração. Agradeço às amigas do Laboratório de Genética Humana (IOC/FIOCRUZ) por compartilharem sabedoria e uma amizade especial.

Aos amigos que a FIOCRUZ, UNIRIO, Rio de Janeiro e São Paulo me deram, muitos deles vindo de cada cantinho do país, pelos bons momentos, pela compreensão nos momentos de ausência e pela torcida para que eu chegasse até aqui. Em especial, às minhas amigas Lu e Jaque, que mesmo em meio ao conturbado momento político do país, abriram suas portas para me hospedar durante os experimentos na USP.

Aos voluntários que participaram do estudo, permitindo a realização do mesmo.

Ao Laboratório Sérgio Franco, pelo auxílio nas dosagens sorológicas.

À Plataforma de Genômica de Sequenciamento de DNA – RPT01A (Rede de Plataformas Tecnológicas) – da FIOCRUZ.

Ao Programa de Biologia Celular e Molecular e aos professores das disciplinas e dos cursos nacionais e internacionais que requintaram a minha formação acadêmica.

Às agências de fomento CAPES, CNPq e IOC pelo suporte financeiro que possibilitaram o desenvolvimento da presente tese.

SUMÁRIO

| | |
|--|-------|
| Lista de abreviaturas, siglas e símbolos | xii |
| Lista de figuras | xv |
| Lista de tabelas | xviii |
| RESUMO | xx |
| ABSTRACT | xxii |
| CAPÍTULO I – Introdução, Justificativa, Objetivos e Metodologia | 1 |
| 1. INTRODUÇÃO | 2 |
| 1.1 Evolução humana e doenças infecciosas | 2 |
| 1.1.1 Resistência genética humana à malária | 3 |
| 1.1.2 Mutações de proteção à malária | 5 |
| 1.1.3 Assinatura de seleção natural no genoma | 7 |
| 1.2 Metabolismo energético | 12 |
| 1.2.1 O gene <i>PKLR</i> e a glicólise | 13 |
| 1.2.2 Polimorfismos no gene <i>PKLR</i> e associação com a malária | 15 |
| 1.2.3 O gene <i>G6PD</i> e a via das pentoses | 17 |
| 1.2.4 Mutações no gene <i>G6PD</i> e associação com a malária | 18 |
| 1.3 Hanseníase | 24 |
| 1.3.1 Breve histórico e epidemiologia | 24 |
| 1.3.2 <i>Mycobacterium leprae</i> | 26 |
| 1.3.3 Classificação clínica | 29 |
| 1.3.4 Transmissão, diagnóstico e tratamento | 31 |
| 1.3.5 Papel do ferro na infecção pelo <i>M. leprae</i> | 34 |
| 1.3.6 Suscetibilidade genética à hanseníase | 38 |
| 1.4 Heterogeneidade da população brasileira | 45 |
| 1.4.1 Estudo de associação genética em populações miscigenadas | 47 |
| 2. JUSTIFICATIVA | 52 |
| 3. OBJETIVOS | 54 |
| 3.1 Objetivo geral | 54 |
| 3.2 Objetivos específicos | 54 |
| 4. METODOLOGIA | 55 |
| 4.1 Desenho de estudo | 55 |
| 4.2 Análise de ancestralidade e estruturação populacional | 56 |
| 4.2.1 Populações do estudo de associação e ancestralidade genômica | 56 |
| 4.2.2 Considerações éticas | 62 |
| 4.2.3 Extração de DNA | 62 |
| 4.2.4 Genotipagem dos marcadores de ancestralidade (46-AIMS-Indels) | 63 |

| | |
|---|-----------|
| 4.1.5 – Análise de ancestralidade e estrutura populacional | 66 |
| 4.3 Análise de associação genética | 67 |
| 4.3.1 Seleção de SNPs candidatos | 67 |
| 4.3.2 Genotipagem por PCR em tempo real | 71 |
| 4.3.3 Análises do estudo de associação | 72 |
| 4.3.4 Busca de eQTLs e correlação dos SNPs com a literatura | 72 |
| 4.4 Análises de seleção natural | 73 |
| 4.4.1 Dados genômicos | 73 |
| 4.4.2 Métodos de detecção de seleção estudados | 73 |
| 4.5 Análises funcionais | 74 |
| 4.5.1 População do estudo funcional | 74 |
| 4.5.2 Quantificações sorológicas | 75 |
| 4.5.3 Extração de RNA de sangue total e síntese de cDNA | 75 |
| 4.5.4 Ensaios de infecção com <i>M. leprae</i> em monócitos primários | 76 |
| 4.5.5 Ensaios de atividade enzimática | 77 |
| 4.5.6 PCR em tempo real para análise da expressão gênica (qRT-PCR) | 77 |
| CAPÍTULO II – Ancestralidade genômica em grupos populacionais brasileiros utilizados em estudos de associação com a hanseníase | 79 |
| 1 – Caracterização da ancestralidade global e individual de casos e controles | 80 |
| 2 – Ancestralidade da população brasileira comparada aos estudos prévios | 85 |
| CAPÍTULO III – Associação genética, caracterização evolutiva e análise funcional do gene <i>PKLR</i> com a hanseníase | 87 |
| 1 – Análise de associação | 88 |
| 1.1 – Associação dos SNPs do gene <i>PKLR</i> com a suscetibilidade à hanseníase após o ajuste | 88 |
| 1.2 – Replicação do estudo de associação do <i>PKLR</i> com a hanseníase em outras populações brasileiras | 94 |
| 1.3 – Associação dos SNPs do gene <i>PKLR</i> com a suscetibilidade à tuberculose em Moçambique | 95 |
| 1.4 – Análise de desequilíbrio de ligação (LD) dos polimorfismos do gene <i>PKLR</i> | 97 |
| 1.5 – O haplótipo T/G/G e a associação com a suscetibilidade a hanseníase e tuberculose | 99 |
| 2 – Análises evolutivas | 102 |
| 2.1 – Frequências alélicas das variantes do gene <i>PKLR</i> diferem entre Europeus e Africanos | 102 |
| 2.2 – Análise de “ <i>F_{ST} outlier</i> ” entre europeus e africanos | 103 |
| 2.3 – Cálculo de Tajima D em europeus e africanos | 105 |
| 2.4 – Análise da homozigosidade do haplótipo estendido (iHS e xpEHH) | 108 |
| 2.5 – <i>F_{ST}</i> e xpEHH em um <i>dataset</i> de populações africanas | 111 |
| 3 – Análises funcionais | 113 |
| 3.1 – Comparação entre os carreadores dos genótipos de risco e as dosagens séricas . | 113 |

| | |
|--|------------|
| 3.2 – Busca de eQTLs do gene PKLR (GTEx) | 120 |
| CAPÍTULO IV – Associação de polimorfismos do gene <i>G6PD</i> com a hanseníase e o impacto funcional dessas variantes na infecção | 121 |
| 1 – Análise de associação | 122 |
| 1.1 – Associação dos polimorfismos do gene <i>G6PD</i> com a hanseníase na população do Rio de Janeiro | 122 |
| 1.2 – Análise de LD entre os polimorfismos do gene <i>G6PD</i> | 125 |
| 2 – Análise funcional | 127 |
| 2.1 – Comparação entre os genótipos das mutações em <i>G6PD</i> e as dosagens séricas | 127 |
| 2.2 – Impacto funcional das variantes de <i>G6PD</i> na infecção pelo <i>M. leprae</i> | 128 |
| CAPÍTULO V – Discussão e Conclusões | 132 |
| 1. DISCUSSÃO | 132 |
| 1.1 – Suscetibilidade do gene <i>PKLR</i> e doenças infecciosas | 134 |
| 1.2 – Pegadas de seleção nos genes <i>PKLR</i> e <i>G6PD</i> | 141 |
| 1.3 – Aspectos funcionais das variantes do gene <i>G6PD</i> e a patogênese da hanseníase . | 146 |
| 2. CONSIDERAÇÕES FINAIS | 151 |
| 3. CONCLUSÕES | 153 |
| 4. REFERÊNCIAS BIBLIOGRÁFICA | 154 |
| ANEXOS I | 186 |
| 1 – Lista dos pares de <i>primers</i> referente aos 46-AIMs-Indels e as características dos marcadores | 186 |
| ANEXOS II | 182 |
| 1 – Resultados complementares | 182 |
| ANEXOS III | 188 |
| 1 – Artigos em preparação para submissão | 188 |
| 2 – Artigos submetidos como colaboradora ao longo do período de doutoramento | 192 |
| 3 – Artigos publicados como colaboradora ao longo do período de doutoramento | 196 |

Lista de abreviaturas, siglas e símbolos

| | |
|-------------------------------|---|
| °C | Graus Celsius |
| r^2 | Coefficiente de correlação |
| χ^2 | Qui quadrado |
| 3'UTR | Região 3' não traduzida de um gene |
| ADP | Adenosina difosfato |
| ATP | Adenosina trifosfato |
| BAAR | Bacilo álcool-ácido resistente |
| BCG | Bacilo de Calmette e Guérin |
| BB | <i>Boderline-boderline</i> |
| BL | <i>Boderline-lepromatoso</i> |
| Brf | Bacterioferritina |
| BT | <i>Boderline-tuberculóide</i> |
| <i>CCDC122</i> | Gene cuja sigla vem do inglês, <i>coiled-coil domain containing 122</i> |
| CD | Do inglês, <i>cluster of differentiation</i> |
| CMb | Carboximicobactina |
| CNV | Número variável de cópias (do inglês, <i>copy number variantios</i>) |
| CO | Monóxido de carbono |
| <i>CUBN</i> | Gene que codifica a cubulina |
| <i>DAF</i> | Frequência do alelo derivado |
| DARC | Receptor do antígeno Duffy |
| DbGaP | Banco de dados de genótipos e fenótipos (do inglês, <i>database of Genotypes and Phenotypes</i>) |
| dbSNP | Do inglês, <i>SNPs database</i> |
| dNTP | Desoxirribonucleotídeos trifosfatados |
| DMT-1 | Proteína transportadora de metal divalente (do inglês, <i>divalent metal transporter 1</i>) |
| DNA | Ácido desoxiribonucléico |
| EDTA | Ácido etilenodiamino tetra-acético |
| EHH | Haplótipo estendido em homozigose |
| ELISA | Ensaio imunoenzimático (do inglês, <i>Enzyme-linked immunosorbent assay</i>) |
| ENH | Eritema nodoso hansênico |
| <i>EPAS1</i> | Gene envolvido com a concentração de oxigênio e resposta à hipóxia |
| eQTLs | Do inglês, <i>expression Quantitative Traits Loci</i> |
| EUA | Estados Unidos da América |
| Fe | Ferro |
| FPT | Ferroportina |
| F_{ST} | Índice de diferenciação populacional |
| G6DP | Glicose-fosfato |
| GSH | Glutationa na forma reduzida |
| GSSG | Glutationa na forma oxidada |
| GWAS | Estudo de associação do genoma completo |
| H ₂ O ₂ | Peróxido de hidrogênio |

| | |
|-------------------|---|
| Hb | Hemoglobina |
| Hp | Haptoglobina |
| HCl | Ácido Clorídrico |
| HLA | Antígeno leucocitário humano |
| HIV | Vírus da imunodeficiência humana |
| HO-1 | Heme-oxigenase-1 |
| HWE | Equilíbrio de Hardy-Weimberg |
| I | Forma indeterminada da hanseníase |
| IBGE | Instituto Brasileiro de Geografia e Estatística |
| ideR | Gene regulador de ferro (do inglês, <i>iron-dependent</i>) |
| IFNG | Gene que codifica o Interferon-gama |
| IFN γ | Interferon gama |
| IL | Interleucina |
| INCA | Instituto Nacional do Câncer |
| IOC | Instituto Oswaldo Cruz |
| IREPs | Proteínas alternativas reguladoras de ferro |
| IrtAB | Do inglês, <i>Iron transporter</i> |
| K ⁺ | Íon potássio |
| KCl | Cloreto de potássio |
| LAM | Lipoarabinomanana |
| LCT | Gene que codifica a lactase |
| LD | Desequilíbrio de ligação (do inglês, <i>Linkage Disequilibrium</i>) |
| LDH | Lactato desidrogenase |
| LID-1 | Antígeno recombinante (do inglês, <i>Leprosy IDRI Diagnostic</i>) |
| LL | Lepromatose |
| LTA | Gene que codifica linfotóxina |
| M | Molar |
| M. | Mycobacterium |
| Mb | Megabase / micobactinas |
| MB | Multibacilar |
| mg | Miligrama |
| Mg ²⁺ | Íon magnésio |
| MgCl ₂ | Cloreto de magnésio |
| miRNA | MicroRNA |
| mL | Mililitro |
| MR | Receptor de manose |
| MRC1 | Gene que codifica o receptor de manose (do inglês, <i>Mannose receptor C type 1</i>) |
| mRNA | Ácido ribonucleico mensageiro |
| NaCl | Cloreto de sódio |
| NCBI | Centro Nacional de Informações sobre Biotecnologia |
| NEBL | Gene que codifica nebulina |
| NF-kB | Fator nuclear kappa B |
| ng | Nanograma |
| nM | Nanomolar |

| | |
|--------------|---|
| <i>NOD</i> | gene que codifica o receptor NOD (do inglês, <i>Nucleotide-binding Oligomerization Domain Receptor</i>) |
| NP | Forma neural pura |
| OMS | Organização Mundial da Saúde |
| OR | <i>Odds ratio</i> (razão de chances) |
| pb | Pares de base |
| <i>PACRG</i> | Gene que compartilha o promotor do gene <i>PARK2</i> (do inglês, <i>Parkin co-regulated gene</i>) |
| PAMP | Padrões Moleculares Associados a Patógenos (do inglês, <i>Pathogen-associated molecular patterns</i>) |
| <i>PARK2</i> | Gene que codifica a proteína parkina |
| PB | Paucibacilar |
| PBMC | Célula Mononuclear do Sangue Periférico (do inglês, <i>Peripheral Blood Mononucleated Cell</i>) |
| PCR | Reação em cadeia da polimerase |
| PGL-1 | Glicolipídeo fenólico-1 |
| pH | Potencial hidrogeniônico |
| PK | Piruvato cinase |
| <i>PKM</i> | Gene que codifica a piruvato quinase do músculo (do inglês, <i>Muscle PK</i>) |
| PKL | Piruvato quinase do fígado (do inglês, <i>Liver PK</i>) |
| <i>PKLR</i> | Gene que codifica a piruvato quinase do fígado e eritrócitos (do inglês, <i>Pyruvate Kinase Liver and Red blood cells</i>) |
| PKR | Piruvato quinase dos eritrócitos (do inglês, <i>Red blood cell PK</i>) |
| PPP | Via das pentoses-fostato |
| rs | Do inglês, <i>reference SNP</i> |
| SC | Célula de Schwann (do inglês, <i>Schwann cell</i>) |
| tag | Etiqueta |
| Taq | <i>Thermus aquaticus</i> |
| Th | Célula T auxiliar |
| TB | Tuberculose |
| TDT | Teste de Desequilíbrio de Transmissão |
| TE | Tampão Tris EDTA |
| Trf | Transferrina |
| TNF | Fator de necrose tumoral |
| TLR | Receptor do tipo Toll (do inglês, <i>Toll-Like Receptor</i>) |
| TT | Pólo tuberculóide |
| UTR | Região não codificante (do inglês, <i>Untranslated region</i>) |
| <i>VDR</i> | Gene que codifica o receptor de vitamina D (do inglês, <i>Vitamin D receptor</i>) |
| VNTR | Repetição em tandem de número variável (do inglês, <i>Variable number of tandem repeats</i>) |

Lista de figuras

| | |
|---|----|
| Figura 1.1 – Distribuição da malária e risco de infecção a nível mundial (OMS, 2018) .. | 4 |
| Figura 1.2 – Impacto das varreduras seletivas (<i>sweeps</i>) nos padrões de variabilidade genética e desequilíbrio de ligação (LD) antes e depois da seleção | 8 |
| Figura 1.3 – Via glicolítica e via das pentoses | 14 |
| Figura 1.4 – Distribuição da frequência alélica das variantes de <i>G6PD</i> a nível global | 19 |
| Figura 1.5 – Distribuição das mutações ao longo do gene <i>G6PD</i> | 21 |
| Figura 1.6 – Mecanismos de hemólise na deficiência da PK e <i>G6PD</i> | 24 |
| Figura 1.7 – Novos casos de hanseníase distribuídos no mundo em 2018 reportados pela OMS | 25 |
| Figura 1.8 – Número total de novos casos no país de 2002 a 2018 (Ministério da Saúde) | 26 |
| | |
| Figura 1.9 – Distribuição dos genes de <i>M. leprae</i> e <i>M. tuberculosis</i> de acordo com a categoria funcional | 28 |
| Figura 1.10 – Classificação das formas clínicas da hanseníase baseado em Ridley e Jopling (1966) | 31 |
| Figura 1.11 – Vias de aquisição de ferro pelo <i>M. tuberculosis</i> (Mtb) no nicho intracelular | 36 |
| Figura 1.12 – Linha do tempo indicando os genes associados com a hanseníase nos últimos 20 anos | 44 |
| Figura 1.13 – Tráfego de africanos para o Brasil entre os séculos XVI e XIX | 47 |
| Figura 1.14 – Estruturação populacional e o viés de associação genética entre casos e controles | 49 |
| Figura 4.1 – Estratégia de estudo apresentada nesse trabalho | 55 |
| Figura 4.2 – Genotipagem dos marcadores do tipo Indels | 66 |
| Figura 4.3 – Fluxograma de seleção dos SNPs funcionais na região do gene <i>PKLR</i> e <i>G6PD</i> | 69 |
| Figura 4.4 – Análise de Componentes Principais (PCA) na população do 1000Genomes para a busca de SNPs funcionais | 70 |

CAPÍTULO II

| | |
|--|----|
| Figura 1.1 – Proporções de ancestralidade dos grupos populacionais estudados baseadas em 46 indels-AIMs | 83 |
| Figura 1.2 – Proporções (%) de ancestralidade genética entre casos e controles das populações endêmicas para a hanseníase entre as sub-regiões brasileiras | 84 |

CAPÍTULO III

| | |
|---|-----|
| Figura 1.1 – Regressão logística dos SNPs do gene <i>PKLR</i> na população do Rio de Janeiro sem ajuste (A), ajustando para sexo e etnia (B) e para sexo e ancestralidade (C) | 88 |
| Figura 1.2 – Valores de Odds ratios (ORs) e intervalos de confiança (95%) para a associação entre os polimorfismos do gene <i>PKLR</i> com a hanseníase e tuberculose | 96 |
| Figura 1.3 – Desequilíbrio de ligação entre os SNPs da região do <i>PKLR</i> entre as populações do estudo | 98 |
| Figura 1.4 – Densidade das variantes genéticas do cromossomo 1 em europeus (EUR) e africanos (AFR) | 103 |
| Figura 1.5 – Distribuição dos valores de F_{ST} para as variantes da região do gene <i>PKLR</i> .. | 104 |
| Figura 1.6 – Valores de Tajima D por janelas em africanos (linha azul) e europeus (linha verde) | 106 |
| Figura 1.7 – <i>Sweep</i> de seleção no gene <i>HCN3</i> detectado pela análise de xpEHH (1:155247308-155290457) entre africanos e europeus | 108 |
| Figura 1.8 – LD entre as variantes dos genes <i>HCN3</i> , <i>PKLR</i> e <i>FDPS</i> | 109 |
| Figura 1.9 – Decaimento de homozigosidade do haplótipo a partir do SNP “core” em europeus e africanos | 111 |
| Figura 1.10 – Níveis de ferro sérico, ferritina e haptoglobina entre os indivíduos do estudo | 115 |
| Figura 1.11 – Níveis de ferro sérico, ferritina e haptoglobina entre os pacientes do Rio de Janeiro e Rondonópolis | 116 |
| Figura 1.12 – Níveis de ferro sérico, ferritina e haptoglobina comparado com os genéticos de risco | 117 |
| Figura 1.13 – Níveis de ferro sérico, ferritina e haptoglobina comparado com a combinação dos genéticos | 118 |

| | |
|--|-----|
| Figura 1.14 – Níveis de ferro sérico, ferritina e haptoglobina entre o sexo feminino (F) e masculino (M) comparado com os genótipos do SNP rs1052176 | 119 |
| Figura 1.15 – Expressão do gene <i>PKLR</i> comparado aos genótipos dos SNPs eQTLs | 120 |
| CAPÍTULO IV | |
| Figura 1.1 – Valores de <i>Odds ratios</i> (ORs) e intervalos de confiança (95%) para a associação entre os polimorfismos do gene <i>G6PD</i> e a hanseníase | 125 |
| Figura 1.2 – Desequilíbrio de ligação entre os SNPs da região do <i>G6PD</i> nas populações do 1000Genomes (A), Rio de Janeiro (B) e Epigen (C) | 126 |
| Figura 1.3 – Níveis de ferro sérico, ferritina e haptoglobina por SNP do gene <i>G6PD</i> em indivíduos saudáveis | 128 |
| Figura 1.4 – Expressão do gene <i>G6PD</i> e atividade enzimática em monócitos de infectados com <i>M. leprae</i> | 130 |
| CAPÍTULO V | |
| Figura 1.1 – Modelo proposto para o mecanismo de sobrevivência micobacteriana em indivíduos carreadores dos diferentes haplótipos do gene <i>PKLR</i> | 139 |
| Figura 1.2 – Modulação do metabolismo de células de Schwann para a sobrevivência micobacteriana | 147 |

Lista de tabelas

| | |
|---|-----|
| Tabela 1.1 – Classificação das variantes de <i>G6PD</i> de acordo com os níveis de atividade enzimática (OMS) | 22 |
| Tabela 1.2 – Comparação entre os genomas do <i>M. leprae</i> , <i>M. lepromatosis</i> e <i>M. tuberculosis</i> | 28 |
| Tabela 4.1 – Taxa média de detecção de novos casos de hanseníase (por 100 mil habitantes) por região Brasileira estudada, registrada pelo Ministério da Saúde entre 2012 e 2016 | 60 |
| Tabela 4.2 – Características das populações do estudo | 61 |
| Tabela 4.3 – 46-AIM-Indels utilizados no estudo, incluindo a localização de cada marcador no genoma e os alelos relatados | 65 |
| Tabela 4.4 – Caracterização das populações do painel HGDP-CEPH utilizadas como referência | 67 |
| Tabela 4.5 – SNPs dos genes <i>PKLR</i> e <i>G6PD</i> candidatos ao estudo de associação | 70 |
| Tabela 4.6 – Sequência dos oligonucleotídeos utilizados nos ensaios de qRT-PCR | 78 |
| CAPÍTULO II | |
| Tabela 1.1 – Comparação das estimativas de ancestralidade entre os casos de hanseníase do Rio de Janeiro | 81 |
| Tabela 1.2 – Comparação da ancestralidade na população brasileira por diferentes estudos | 86 |
| CAPÍTULO III | |
| Tabela 1.1 – Frequência dos genótipos, alelos e carreadores dos SNPs do gene <i>PKLR</i> em casos e controles e regressão logística para associação com a hanseníase ou tuberculose | 90 |
| Tabela 1.2 – Frequência de haplótipos do gene <i>PKLR</i> nas populações do estudo e associação com a hanseníase e TB | 101 |
| Tabela 1.3 – Frequência alélica dos SNPs do gene <i>PKLR</i> em europeus e africanos do 1000Genomes | 102 |

| | |
|--|-----|
| Tabela 1.4 – Valores de F_{ST} para os SNPs individuais e por janelas em europeus e africanos | 105 |
| Tabela 1.5 – Valores de Tajima D calculado por janelas de 1000pb na região do gene <i>PKLR</i> | 107 |
| Tabela 1.6 – LD entre os 25 SNPs com sinal de seleção pelo teste de xpEHH entre europeus e africanos e os SNPs associados à hanseníase | 110 |
| Tabela 1.7 – Análise de F_{ST} , iHS e xpEHH dos SNPs do gene <i>PKLR</i> nas populações africanas de Gouveia e colaboradores (2019) | 112 |
| Tabela 1.8 – Frequência alélica dos SNPs do PKLR na população de Gouveia e colaboradores (2019) e 1000Genomes | 113 |
| Tabela 1.9 – Média das dosagens séricas entre homens e mulheres do estudo | 114 |
| CAPÍTULO IV | |
| Tabela 1.1 – Frequência dos polimorfismos do gene G6PD nas populações do 1000Genomes no Rio de Janeiro e associação com a hanseníase | 123 |
| CAPÍTULO V | |
| Tabela 1.1 – Associação do gene PKLR com a hanseníase na população de chineses Han de Zhang e colaboradores (2009) | 137 |



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

RESUMO

POLIMORFISMOS SELECIONADOS POR PATÓGENOS EM GENES DO METABOLISMO ENERGÉTICO E ASSOCIAÇÃO COM DOENÇAS MICOBACTERIANAS

TESE DE DOUTORADO

Ohanna Cavalcanti de Lima Bezerra

Patógenos atuam como poderosos agentes de seleção no genoma humano. Populações africanas de regiões endêmicas para a malária apresentam mutações associadas com a resistência ao *Plasmodium* em genes de enzimas cruciais ao metabolismo das células vermelhas, como a PK (piruvato cinase) e a G6PD (glicose-6-fosfato desidrogenase). A PK e a G6PD participam na produção de ATP e NADPH reduzido na via glicolítica e via das pentoses, respectivamente. A deficiência dessas enzimas leva ao quadro de anemia hemolítica, que tem como consequência funcional a liberação de hemoglobina, heme e ferro. Devido à dependência de ferro para a sobrevivência de bactérias intracelulares, a deficiência da PK também foi associada com a suscetibilidade a esses microrganismos. A hipótese defendida nesse trabalho é a de que mutações selecionadas por patógenos nos genes *PKLR* e *G6PD*, que codificam a enzima PK e a glicose-6-fosfato desidrogenase, estão associadas com a suscetibilidade a micobacterioses, utilizando a hanseníase como modelo de estudo. Previamente, identificamos a associação do gene *PKLR* com suscetibilidade a hanseníase e tuberculose (TB) na população do Rio de Janeiro e Moçambique, respectivamente. Entretanto, considerando que a população brasileira é geneticamente heterogênea, na tentativa de evitar falsos-positivos provocados por estruturação populacional, foi necessário utilizar a ancestralidade genômica para o ajuste da associação. Para isso, estimamos a ancestralidade genética de indivíduos que compõem os grupos populacionais estudados a partir de um painel de 46 indels marcadores informativos de ancestralidade (AIMs). Em seguida, a fim de confirmar a associação do gene *PKLR* com a hanseníase, realizamos a replicação do estudo em outras populações brasileiras, totalizando um número de 5.293 indivíduos analisados. O haplótipo T/G/G (rs1052176/rs4971072/rs11264359) do gene *PKLR* foi significativamente associado com a hanseníase na população do Rio de Janeiro (OR=1,26; $p=0,02$), seguido da associação sugestiva em Salvador (OR=1,35; $p=0,07$) e em Moçambique (OR=1,52; $p=0,07$), em que o desfecho é a TB. O genótipo TT do rs1052176 foi o SNP associado de forma significativa entre essas populações, confirmando o potencial do gene *PKLR* como um preditor de risco a doenças micobacterianas. Suposições acerca da seleção no gene *PKLR* têm pouca evidência em populações humanas. Embora os testes para assinaturas de seleção no nosso estudo não foram decisivos quanto à ocorrência de um regime seletivo na região, algumas sugestões foram apontadas. Uma varredura seletiva foi observada em 25 variantes no gene *HCN3* que se

encontram em alto desequilíbrio de ligação (LD) com os polimorfismos associados com a hanseníase, porém o sinal ocorreu em europeus. Não observamos uma correlação direta entre genótipo-fenótipo entre as variantes de risco e as dosagens de ferro, porém o SNP rs1052176 parece modular a expressão do gene *PKLR*. O gene *G6PD*, por sua vez, apresenta-se indispensável para a sobrevivência do *Mycobacterium leprae* em células de Schwann, como demonstrado anteriormente. Na tentativa de identificar uma correlação entre polimorfismos (incluindo as mutações A376G/G202A que formam a variante A-) no *G6PD* e a hanseníase, nenhuma associação foi observada em homens e mulheres do Rio de Janeiro após a correção para ancestralidade. Porém, o aumento significativo da expressão do *G6PD* bem como da atividade enzimática após a infecção em monócitos primários, pouco explorado até o momento, demonstra a relevância do gene na infecção. Assim, nossos dados contribuem para o melhor entendimento de como a relação patógeno-hospedeiro molda a suscetibilidade genética às doenças micobacterianas, incluindo o gene *PKLR* como um preditor de risco da hanseníase e TB. Além disso, a identificação de vias metabólicas cruciais ao *M. leprae* pode indicar formas alternativas para o tratamento da hanseníase.



Ministério da Saúde

FIOCRUZ
Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

ABSTRACT

POLYMORPHISMS SELECTED BY PATHOGENS IN THE ENERGETIC METABOLISM GENES AND ASSOCIATION WITH MYCOBACTERIAL DISEASES

TESE DE DOUTORADO

Ohanna Cavalcanti de Lima Bezerra

Pathogens are strong selective agents in the human genome. African populations from malaria-endemic regions present mutations associated with *Plasmodium* resistance in genes of crucial enzymes of red blood cells metabolisms, as PK (pyruvate kinase) and G6PD (glucose-6-phosphate dehydrogenase). PK and G6PD act on ATP and NADPH production in the glycolytic and pentose pathway, respectively. Hemolysis is one of the functional consequences of these enzyme's deficiency, which in turn leads to hemoglobin and iron releases. Therefore, PK deficiency has also been associated with susceptibility to intracellular bacteria. Our hypothesis is that pathogen-selected mutations in the *PKLR* and *G6PD* genes, which encode the enzyme PK and glucose-6-phosphate dehydrogenase, are associated with susceptibility to mycobacterious diseases, having leprosy as our model of study disease. Previously, we identified the association of the *PKLR* gene with susceptibility to leprosy and tuberculosis (TB) in the population of Rio de Janeiro and Mozambique, respectively. However, considering that the Brazilian population is admixed, attempting to avoid false-positives associations due to population substructure, it was necessary to use the genomic ancestry to adjust for the association. Then, we estimated the genetic ancestry of the individuals enrolled in the study from a panel of 46 Indels ancestry informative markers (AIMs). Then, in order to confirm the association of the *PKLR* gene with leprosy, we performed a replication study in other Brazilian populations, accounting for 5,293 individuals analyzed. The T/G/G haplotype (rs1052176/rs4971072/rs11264359) of the *PKLR* gene was significantly associated with leprosy in the population of Rio de Janeiro (OR = 1.26; p = 0.02), followed by a suggestive association in Salvador (OR = 1.35; p = 0.07) and in Mozambique (OR = 1.52; p = 0.07), where the outcome is TB. The TT genotype of rs1052176 was the unic significantly associated SNP among these populations, confirming the potential of the *PKLR* gene as a risk predictor for mycobacterial diseases. The hypothesis of selection in the *PKLR* gene have little evidence in human populations. Although the tests for selection signatures in our study were not decisive regarding the occurrence of a selective regime in the region, some suggestions were pointed out. A selective sweep was observed in 25 variants of the *HCN3* gene where the variants are in high linkage disequilibrium (LD) with the polymorphisms associated with leprosy, but the signal

occurred in Europeans. Moreover, we did not observe a direct correlation between genotype-phenotype among risk variants and iron dosages, but SNP rs1052176 seems to modulate *PKLR* gene expression. The *G6PD* gene, in turn, is indispensable for the survival of *Mycobacterium leprae* inside Schwann cells, as previously shown. In order to identify a correlation between polymorphisms (including the A376G/G202A mutations for the A- variant) in *G6PD* with leprosy, no association was observed for men or women from Rio de Janeiro after ancestry adjustments. However, the increased expression of *G6PD* as well as the enzymatic activity 48 hours post *M. leprae* infection in monocytes, not much explored to date, suggests the relevance of the gene along the infection. Thus, our data contribute to a better understanding of how host-pathogen-coevolution shapes the genetic susceptibility to mycobacterial diseases, highlighting the *PKLR* gene as a risk predictor for leprosy and TB. In addition, the identification of vital metabolic pathways for the bacillus survival, such as the *G6PD* pathway, may allow to alternative ways for leprosy treatment.

APRESENTAÇÃO À ESTRUTURA DA TESE

Este estudo aborda como as relações evolutivas entre patógeno-hospedeiro moldam a ampla variabilidade genética observada nas populações humanas e influenciam na suscetibilidade a doenças infecciosas. A hipótese defendida tange ao estudo de polimorfismos de alta frequência em populações africanas, devido à seleção pela malária, com a associação a doenças micobacterianas. A tese apresenta-se em capítulos, em que o capítulo I é referente a introdução, justificativa, objetivos e metodologia. Os capítulos II, III e IV descrevem os principais resultados obtidos durante o doutoramento e que originaram dois manuscritos que se encontram em fase final para a submissão (em ANEXO). O capítulo V, por fim, descreve a discussão, as principais conclusões e as referências bibliográficas.

Capítulo I

Introdução, Justificativa, Objetivos e Metodologia

1 – INTRODUÇÃO

1.1 – Evolução humana e doenças infecciosas

A recente história de evolução da população humana é caracterizada por bruscas mudanças ambientais e pela emergência de diversos agentes seletivos. A domesticação de plantas e animais no início do período Neolítico, cerca de 10.000 anos atrás, acompanhou o aumento da densidade da população humana, que por sua vez foi confrontada com a propagação de infecções, novos recursos alimentícios e novos ambientes culturais. Frente à essas modificações ambientais, fez-se necessário o surgimento de adaptações fisiológicas ao “novo”. Assim, pode-se considerar que os últimos 10.000 anos da história evolutiva humana tem sido o período de maior interesse biológico, já que importantes adaptações genéticas e diferentes mecanismos de resistência a doenças ocorreram (1).

A variedade de fenótipos observados atualmente nas populações humanas reflete os distintos processos de adaptação genética a diferentes condições ambientais. Numerosos agentes infecciosos atuaram em gerações promovendo o aparecimento dos alelos de resistência que se espalharam ao longo do tempo e gerações (2,3). Os alelos de resistência surgem quando alterações do DNA humano aumentam a proteção contra um determinado patógeno e reflete no aumento da capacidade reprodutiva dos indivíduos. A existência dessas variantes genéticas de resistência se deve a pressões seletivas exercidas por patógenos de doenças milenares que ocorrem de forma distinta entre diferentes populações. Dessa forma, o padrão de resposta imune e a suscetibilidade a doenças infecciosas varia consideravelmente entre diferentes populações (4). Como exemplos de doenças históricas de caráter seletivo na população europeia, cita-se a cólera, tuberculose e febre bubônica (“peste negra”). Na África, algumas diarreias crônicas, febre amarela e malária têm sido endêmicas até os dias atuais (5).

No início do século XX, estudiosos da evolução humana já defendiam a ideia de que as doenças são potentes agentes de seleção natural (6). Na década de 40, levantava-se a observação de que as distúrbios das células vermelhas, como a anemia falciforme e as talassemias, ocorriam em regiões tropicais onde a malária é endêmica e que, de alguma forma, protegiam os indivíduos contra essa doença. Considerando que a teoria da seleção natural havia sido proposta um século

atrás, a hipótese da malária, a qual iremos abordar adiante, destacou, pela primeira vez, um mecanismo de adaptação genética humana frente a uma doença (7,8).

Hoje, o acesso ao conjunto de dados genômicos para o estudo da variação genética humana permite explorar as assinaturas genéticas de seleção natural. Através de diferentes abordagens, é possível identificar variantes genéticas que conferem resistência a doenças infecciosas que se espalharam pelas populações humanas ao longo do tempo. A investigação de assinaturas de seleção natural pode ajudar a elucidar as adaptações evolutivas que permitiram aos seres humanos suportar alguns de nossos agentes seletivos mais complexos e desafiadores. Com isso, é possível identificar variantes que podem ser prontamente candidatas em estudos de associação genética com doenças, verificar os distintos padrões de variação genética no genoma humano e rastrear pistas sobre como os próprios patógenos evoluem tão rapidamente (7,9).

1.1.1 – Resistência genética humana à malária

A malária é considerada uma das maiores forças seletivas atuantes sobre o genoma humano desde a origem da agricultura há cerca de 10.000 anos (1). É uma doença potencialmente fatal causada pelo parasita *Plasmodium*, principalmente das espécies *P. falciparum* e *P. vivax*, e transmitida através da picada do mosquito fêmea do gênero *Anopheles* (10). Segundo o último levantamento do *World Malaria Report* (2018), a malária é responsável por cerca de 430.000 mortes anualmente, sendo 61% de crianças com menos de 5 anos de idade (11). A região africana carrega uma parcela desproporcional da carga mundial de malária, como mostrado na Figura 1.1, comportando 92% do número total de casos da doença e 93% das mortes a nível global. Os países que contêm aproximadamente metade dos casos mundiais são: Nigéria (25%), República Democrática do Congo (11%), Moçambique (5%), Índia (4%) e Uganda (4%) (12).

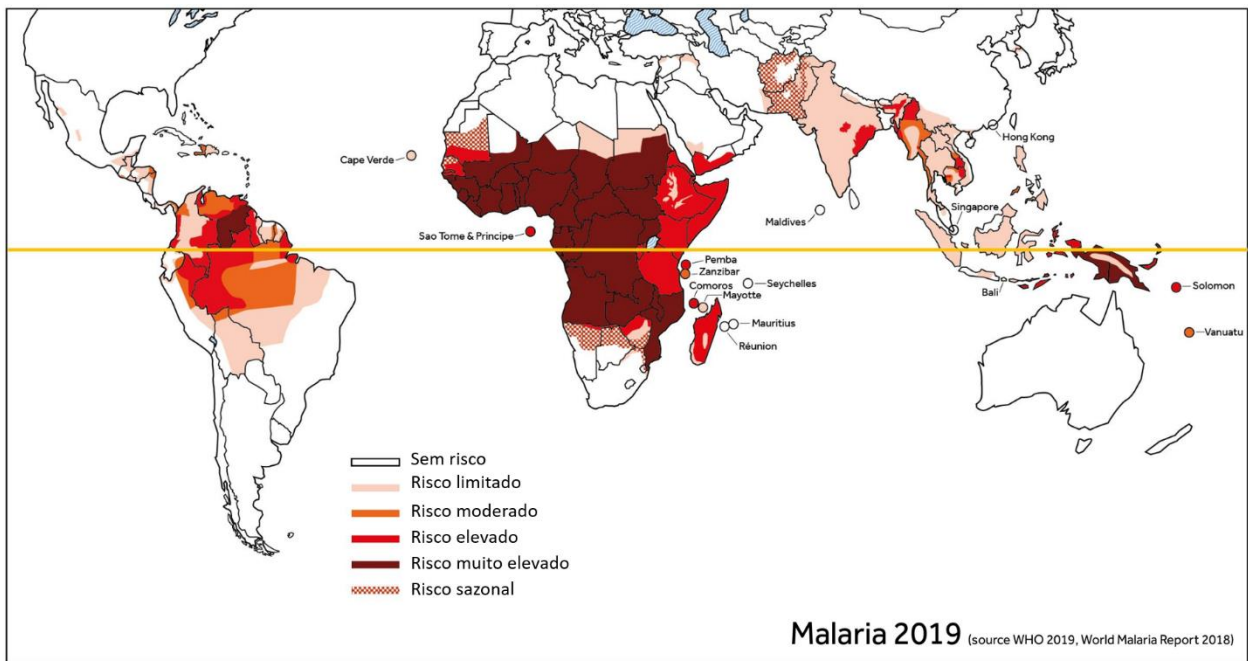


Figura 1.1 – Distribuição da malária e risco de infecção a nível mundial (OMS, 2018). Mapa adaptado de <https://www.itg.be/E/travel-health/malaria>.

O grande impacto de seleção natural da malária advém das severas manifestações clínicas que levam à morte de crianças que ainda não chegaram em idade reprodutiva. Especula-se que em vilas rurais do Oeste Africano, onde o acesso aos medicamentos adequados é escasso, aproximadamente 20% das crianças morrem até os cinco anos (13). Assim, uma vantagem seletiva tremenda ocorre em crianças passíveis de tolerar os episódios de infecção pelo *P. falciparum* e não progredir para a malária severa. Nesse sentido, devido a elevada morbidade e mortalidade milenares dessa doença, indivíduos tiveram que desenvolver mecanismos para sobreviver em face a essa doença devastadora (14).

O pesquisador JBS Haldane foi um dos articuladores iniciais a relacionar a malária com a resistência genética humana. Suas ideias foram discutidas de forma pioneira em 1948 no Congresso Internacional de Genética em Estocolmo (6). Segundo Haldane, os heterozigotos para talassemia são mais resistentes à malária severa do que os indivíduos que não apresentam a mutação. Em heterozigose, os indivíduos não apresentam a anemia grave e a malária não é capaz de reduzir o *fitness* desses indivíduos, portanto, eles são seletivamente vantajosos em populações expostas a malária (7). A primeira evidência reconhecida, entretanto, ocorreu no trabalho de Allison em 1954, em que foi visto que o *P. falciparum* foi significativamente reduzido em células

vermelhas heterozigotas (AS) para a hemoglobina S (que confere anemia falciforme) do que em células de indivíduos normais (AA) (15). Hoje, a “hipótese da malária” de Haldane se provou correta e é muito bem caracterizada em diversos trabalhos, apesar dos mecanismos de proteção para as diferentes variantes genéticas não serem simples de explicar.

1.1.2 – Mutações de proteção à malária

Muitas das variantes que conferem resistência à malária são mutações de perda de função e parecem ser polimorfismos recentes dos últimos 5.000 a 10.000 anos ou menos (5). Dentre elas, as mais comuns e bem caracterizadas envolvem proteínas estruturais específicas, tais como as hemoglobinas S, as talassemias, o antígeno Duffy e o grupo sanguíneo O, e algumas enzimas eritrocitárias, como a deficiência de glucose-6-fosfato desidrogenase, e, mais recentemente descrita, a deficiência de piruvato cinase (16). Considerando que os efeitos patogênicos do *P. falciparum* são exercidos em seu ciclo intraeritrocítico, os polimorfismos em genes das células vermelhas são os primeiros candidatos a se relacionar com a suscetibilidade à malária (10). Abordaremos brevemente as variantes genéticas estruturais que são mais estudadas.

A hemoglobina normal de um adulto é composta por duas cadeias de α -globina e duas de β -globina. A anemia falciforme (em inglês, *sickle cell disease*) é causada por uma mutação na posição 6 do gene da β -globina (*HbB*), que leva à substituição do ácido glutâmico (Glu) por uma valina (Val). A hemoglobina anormal produzida é denominada hemoglobina S e provoca rigidez dos glóbulos vermelhos, tornando-os em forma de foice, o que impede a passagem pelos vasos sanguíneos para oxigenar os tecidos (17). Esse distúrbio genético de herança autossômica recessiva em homozigotos SS pode diminuir a sobrevivência dos pacientes. Porém, heterozigotos AS são excepcionalmente vantajosos em regiões de malária, o que permite a persistência da mutação, principalmente na África subsaariana (18). De fato, inúmeros estudos evidenciam que a contagem de parasitas, a ocorrência da malária severa e as mortes por malária são consideravelmente menores em heterozigotos AS do que em homozigotos AA (selvagens) (19). Em geral, as razões para a resistência dessas mutações frente à malária são: i) crescimento prejudicado do parasito dentro das células vermelhas com a mutação (devido ao menor aporte

de oxigênio e presença de hipóxia) e ii) aumento do *clearance* (do inglês, “remoção”) do parasito pelo sistema imune do hospedeiro (7).

Diferente da anemia falciforme, em que há alteração da estrutura da hemoglobina, a talassemia é uma condição em que há diminuição na produção das formas estruturais normais das α ou β -globinas. As α e β -talassemias são provocadas por deleções e mutações pontuais em diferentes regiões dos genes das α e β -globinas, causando uma síntese insuficiente destas cadeias (18). Os sintomas também variam de acordo com a presença da mutação em um único cromossomo, em que há anemia pouco acentuada e os indivíduos são designados portadores do traço talassêmico, ou a mutação presente em ambos os cromossomos, em que se produz uma quantidade muito reduzida de globina, levando a anemias graves que podem ser fatais no período neonatal (20). Apesar disso, em algumas regiões, como no Sul do Nepal onde a malária é holoendêmica (alto grau de intensidade), a frequência de portadores chega a ser de 80% a 90% da população, ou seja, atinge quase a fixação (21).

O receptor do antígeno Duffy ou DARC (do inglês, *Duffy Antigen Receptor for Chemokines*) é um receptor de eritrócitos que se liga a uma séria de quimiocinas pró-inflamatórias (22). Ele é utilizado pelo *P. vivax*, a segunda espécie de *Plasmodium* mais prevalente no mundo, para invadir o eritrócito (23). Uma única mutação pontual (-33T>C) na região promotora de um dos alelos que codificam o receptor, o alelo FY*B, resulta no fenótipo Duffy negativo, Fy(a⁻b⁻), em que não há expressão do receptor na membrana da hemácia. A ausência do antígeno Duffy impede a invasão do eritrócito por *P. vivax* e leva a resistência à malária (24,25). O *P. vivax* é amplamente distribuído nos países tropicais, e praticamente inexistente na África Central e Ocidental (26). Esta ausência é justificava porque a maior parte da população nessas regiões africanas não contém o antígeno Duffy (27).

Além das mutações em genes relacionados à estrutura das hemácias, alguns polimorfismos importantes também ocorrem em genes que codificam as enzimas eritrocitárias. Devido à perda do núcleo, mitocôndrias e ribossomos, os eritrócitos maduros não têm a capacidade de realizar a fosforilação oxidativa e a síntese proteica. Porém, estas células necessitam de um metabolismo ativo para manter a integridade e flexibilidade da membrana plasmática, assim como para a manutenção da hemoglobina na sua forma funcional. Tal metabolismo é assegurado por enzimas que participam na glicólise (piruvato cinase, ou PK), via

das pentoses-fosfato (glicose-6-fosfato desidrogenas, ou G6PD) e metabolismo de nucleotídeos. Logo, mutações que conferem deficiência nessas enzimas irão comprometer as vias de produção de ATP e/ou de NADPH, provocando alterações na membrana e consequente lise dos eritrócitos (16,28). Ao contrário da anemia falciforme, as deficiências de G6PD ou PK não são traços tão graves. Na verdade, elas são, em sua maioria, assintomáticas. A deficiência de G6PD, por exemplo, tem relevância caso haja exposição a agentes exógenos como infecções ou drogas. A relação da deficiência dessas enzimas e o controle da infecção pelo *Plasmodium* será abordado com mais detalhes nesse trabalho.

1.1.3 – Assinaturas de seleção natural no genoma

Os principais mecanismos que determinam o destino de uma mutação inserida no *pool* genético de uma população são a deriva, processo que altera de forma randômica a frequência alélica na população, e a seleção natural, que leva à seleção de alelos adaptativos e eliminação de alelos desfavoráveis (29). Como a seleção natural é crucial para a manutenção da variabilidade genética, afetando funções biológicas e os valores adaptativos humanos frente ao ambiente, vale discutir como esses processos de seleção ocorrem e entender como eles podem modular a suscetibilidade a doenças complexas (30). De maneira geral, três regimes seletivos são considerados entre os geneticistas de populações. A seleção positiva ocorre quando um alelo que promove uma vantagem evolutiva tem a sua frequência aumentada na população ao longo das gerações, podendo alcançar a fixação. Com isso, há uma redução da diversidade genética ao redor do sítio sob seleção (31). O aumento brusco de um genótipo na população por efeito da seleção natural, além de diminuir a diversidade local, irá levar ao aumento do desequilíbrio de ligação (LD, *Linkage disequilibrium*) na região e a presença de variantes neutras por conta do “efeito carona”, como demonstrado na Figura 1.2. (32).

Sobre a variabilidade local e o padrão de desequilíbrio de ligação, um alelo mutante que surgiu numa população e está em completo LD com as variantes adjacentes pode ter três destinos, caso ele seja neutro: pode desaparecer por deriva; pode persistir, mas continuar raro ou pode persistir e aumentar de frequência. Se o alelo persistir, o LD poderá ser quebrado devido à recombinação gênica. Assim, em uma população que evolui por mutação e deriva, espera-se

encontrar alelos raros em alto LD (por serem recentes), alelos raros e com baixo LD (por serem antigos) e alelos comuns com baixo LD (por serem antigos) (33). Em relação à seleção positiva, quanto mais alta a frequência do alelo e mais distante ocorrer o LD, mais forte deve ter sido a seleção.

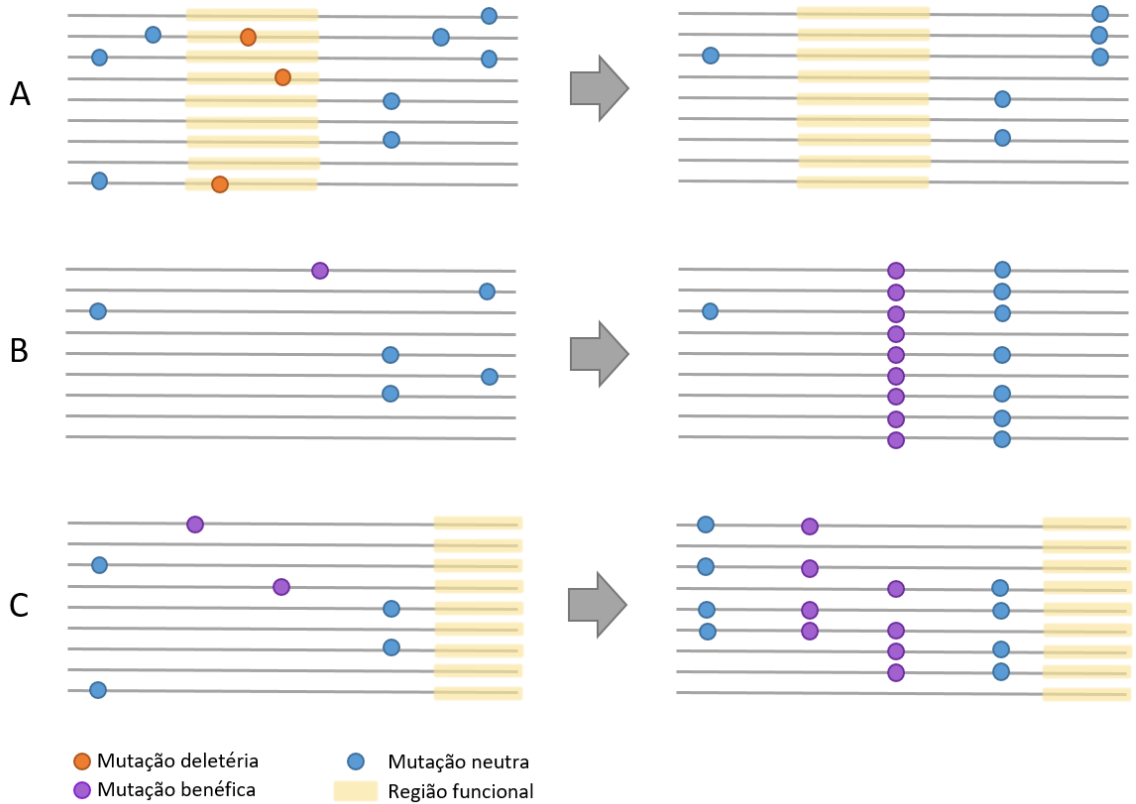


Figura 1.2 – Impacto das varreduras seletivas (*sweeps*) nos padrões de variabilidade genética e desequilíbrio de ligação (LD) antes e depois da seleção. Cada linha representa um haplótipo individual. **A** – Uma mutação deletéria não-sinônima é desvantajosa, portanto ela é eliminada através da seleção purificadora. **B** – A varredura total (*hard sweep*) ocorre quando uma mutação benéfica surge *de novo* e alcança a fixação de forma tão rápida que carrega por efeito carona as mutações neutras ao redor, aumentando o LD entre elas. **C** – A varredura parcial (*soft sweep*) envolve a seleção mutações permanentes ou *de novo*, que podem ou não ocorrer na mesma posição genômica, de modo que alelos adaptativos estejam presentes em vários haplótipos distintos. Todos os haplótipos portadores de um alelo adaptativo varrem simultaneamente para uma frequência mais alta, levando apenas a reduções modestas nos níveis de diversidade genética e LD (Adaptado de McCoy, 2017).

Ao contrário da seleção positiva, a seleção balanceadora leva à manutenção da diversidade genética em um nível acima do esperado sob neutralidade (34). Uma característica clássica desse processo evolutivo é a maior aptidão do genótipo heterozigoto (“vantagem do heterozigoto”), mas também é possível observar que o alelo vantajoso é o de menor frequência (35). Além disso, observa-se um excesso de polimorfismos segregando em frequências intermediárias (36). Nesse cenário, ao invés da seleção levar à fixação do alelo, ela mantém a frequência durante uma larga escala de tempo (37). Os genes HLA, do grupo sanguíneo ABO e gene da β -hemoglobina (anemia falciforme) são os exemplos bem caracterizados que sofrem esse tipo de seleção (38,39). A seleção purificadora é o processo que leva à eliminação do alelo mutante devido ao dano funcional que esse alelo pode gerar.

Scans para detectar seleção natural recente no genoma humano podem desvendar pistas sobre o histórico adaptativo de populações humanas ao ambiente (nutricional, climático ou patogênico) e apontar genes associados com variações fenotípicas, sendo elas neutras ou associadas a doenças, o que, portanto, tem relevância biológica. Até os anos 2000, o estudo da seleção natural em humanos restringia-se a comparação entre genes candidatos individuais e expectativas teóricas. A partir dos dados de sequenciamento e genotipagem em larga escala, foi possível identificar novas regiões candidatas a seleção natural bem como reavaliar os resultados anteriores através da comparação das distribuições empíricas das variações no genoma humano e também entre diferentes populações (40).

Assim, existem duas formas de se estudar indícios de seleção natural ao longo do genoma: i) a partir de métodos baseados em genes candidatos, quando já se tem uma hipótese *a priori* da ocorrência de seleção na região de interesse, ii) ou *scans* genômicos a fim de investigar os sinais sugestivos de seleção, e, posteriormente, avaliar o gene alvo. De ambas as formas, é necessário entender as características esperadas para cada regime evolutivo (34). Em geral, os testes para detecção de assinaturas de seleção se baseiam na observação de desvios (frequências alélicas, desequilíbrio de ligação e a diferenciação geográfica) em regiões ou sítios específicos do genoma que não são esperados sob neutralidade (31). No modelo neutro de seleção, as frequências alélicas flutuam de forma aleatória e não influenciam no *fitness* do indivíduo. Assim, por usarem o modelo neutro como modelo nulo, os testes para detectar assinaturas de seleção são também chamados de “testes de neutralidade” (41).

Uma das formas clássicas de detectar essas assinaturas é a análise da diferenciação genética entre duas populações para as variantes ou região de interesse (F_{ST}) (42,43). Nessa abordagem, é assumido que a frequência do alelo vantajoso aumenta somente na população que sofre seleção, aumentando a diferenciação entre essas populações (31). Entretanto, não é definido qual o grau de diferenciação que supera o esperado em decorrência dos eventos demográficos, como a deriva genética (43). A abordagem de F_{ST} *outlier* pode contornar o efeito da demografia: o alelo é considerado como selecionado caso ele apresente um nível de diferenciação extremo em relação ao esperado para o resto do genoma (44). Esse método muitas vezes é mais sensível para detectar varreduras de seleção do que outros testes poderosos, como o EHH (do inglês, *Extended Haplotype Homozygosity*), abordado adiante. Detectando SNPs altamente diferenciados, uma forma alternativa de verificar em qual população a seleção pode estar ocorrendo é analisar a frequência do alelo derivado – DAF (do inglês, *Derived Allele Frequency*).

Ainda no contexto do método por F_{ST} , outra forma de avaliar se um SNP tem uma diferenciação muito mais alta do que a esperada sob deriva é realizar simulações neutras, isto é, sem ocorrência de seleção. Nesse caso, ao invés de testar se os SNPs do gene são "os mais extremos" em relação aos SNPs fora da região, pergunta-se se eles são mais extremos do que o esperado para dados simulados sem seleção (45). Entretanto, a abordagem por simulação apresenta dificuldades em relação aos parâmetros demográficos populacionais que melhor se adequam ao modelo e que devem ser cuidadosamente estabelecidos, além do maior tempo de execução dessa análise (46).

A metodologia de homozigidade do haplótipo estendido (EHH) consiste na relação entre as frequências dos alelos e a extensão do LD. Nessa análise, define-se um alelo "core" e observa-se o decaimento da homozigidade ao redor desse alelo. O EHH, também conhecido por teste do haplótipo de longo alcance (do inglês, *Long-Range Haplotype – LRH*), é definido como a probabilidade de que dois cromossomos escolhidos aleatoriamente e que carreguem haplótipos de interesse sejam idênticos por descendência para todo o intervalo da região do "core" a uma distância "x". Assim, quanto mais alto o valor de EHH, maior a homozigidade do haplótipo estendido, e maior o LD entre as variantes adjacentes, sugerindo seleção. Esse teste foi inicialmente postulado por Sabeti e colaboradores (2002) para identificar seleção positiva (47). A

desvantagem é que os haplótipos estendidos persistem por pouco tempo (<30.000 anos) devido à recombinação (31,47). O teste iHS (do inglês, *integrated haplotype score*) foi formulado para observar o decaimento em homoziguidade considerando o alelo derivado e o alelo ancestral. Quando a razão de decaimento do alelo derivado e ancestral é similar, então $iHS \approx 0$. Valores de iHS negativos indicam que existe um haplótipo longo carregando o alelo derivado, e os valores de iHS positivos pontuam um haplótipo longo carregando o alelo ancestral (44). Ainda existe o xpEHH, que compara o decaimento da homoziguidade do haplótipo em relação ao alelo ancestral e derivado entre duas populações. No teste iHS, as informações genômicas são diretamente comparadas com a informação de um ancestral, de forma que este teste tem aplicação na detecção de assinaturas de seleção dentro da população. Já o XP-EHH possibilita a detecção de regiões candidatas à seleção entre populações pela comparação dos perfis de EHH de uma população referência e uma população observada (40). Em ambos iHS e xpEHH, valores acima de 2 são considerados indicativos de seleção.

O teste de Tajima D, amplamente utilizado, avalia desvio de neutralidade a partir da interpretação de padrões de variação de nucleotídeos. O teste de Tajima D compara a média das diferenças par a par sobre o número de sítios segregantes em uma amostra (48). O teste de Tajima D compara a média das diferenças par a par sobre o número de sítios segregantes em uma amostra. O valor de D de Tajima igual a 0 indica ausência de seleção. Caso não haja alterações demográficas (expansão ou contração da população, migração, etc.) em uma população que esteja submetida à seleção positiva, os valores de Tajima D serão negativos. Isso ocorre porque um haplótipo conservado estará presente na maioria da população, ou seja, observaremos poucas diferenças par a par. No caso da seleção balanceadora, alelos em frequência intermediária serão mantidos, e nesse caso, o valor de Tajima D será positivo pois o número de diferenças par a par será maior do que o número de sítios segregantes. Outro teste de neutralidade, como o McDonald Kreitman (1991), tem poder para detectar desvios de seleção já que eles consideram além das variações intraespecíficas, as divergências interespecíficas (49).

Uma outra forma de avaliar região sob seleção é calcular a relação entre a taxa de mutações não-sinônimas, que alteram o aminoácido da proteína (podendo ser uma mutação deletéria ou vantajosa), e sinônimas (que normalmente são mutações neutras). O modelo de seleção adaptativa (ou positiva) pode aumentar a taxa de substituição não-sinônimas (dN),

tornando-as mais altas do que as mutações sinônimas (dS). Então, caso a razão dN/dS seja maior que 1, observa-se um sinal de seleção positiva, mas também é possível predizer seleção balanceadora (31,34). Caso as mutações não-sinônimas sejam deletérias, serão alvo de seleção purificadora no sentido de eliminar essas mutações do *pool* gênico e, nesse caso, a razão dN/dS será menor que 1. Uma problemática desse método é que quando se avalia genes por inteiro, alguns códons podem estar sob seleção balanceadora, outros, positiva, mas a maioria das mutações não-sinônimas são alvo de seleção purificadora. Assim, recentemente há sugestões de analisar a razão dN/dS por grupos de códons ao longo do seguimento genômico (50).

Outros testes estão disponíveis e podem ser acessados em Sabeti e colaboradores (2006), Vitti e colaboradores (2013) entre outros (31,51). A detecção de marcas de seleção no genoma não é simples e envolve tanto a seleção do método mais adequado quanto dos dados genômicos onde os testes serão aplicados. Cabe destacar que devido as propriedades únicas de cada teste, os resultados gerados para assinaturas de seleção devem ser utilizados de forma complementar. Assinaturas foram evidenciadas em poucos genes até hoje, como o gene da β -hemoglobina, o locus HLA, gene da lactase (*LCT*), *EPAS1*, *G6PD*, etc (1,33,45,52). Todavia, especula-se que no genoma, em vez de se questionar se uma região é alvo de seleção, deve-se questionar o quão forte é a seleção atuante na região, pois sugere-se que a maior parte do genoma está sob seleção natural (53). Nesse sentido, novos estudos em genômica evolutiva devem ser formulados a fim de desvendar possíveis pegadas de seleção e melhor compreender como as frequências alélicas regulam os padrões de variabilidade genética humana e a resposta aos diferentes patógenos.

1.2 – Metabolismo da glicose

O metabolismo energético se refere ao complexo de reações metabólicas intracelulares envolvidas com o processamento dos alimentos provenientes da ingestão e a produção e armazenamento de energia necessária à vida. O transporte de glicose é fundamental para o metabolismo energético celular, disponibilizando glicose e intermediários para rotas metabólicas. Diversas reações anabólicas e catabólicas pertencentes a diferentes vias caracterizam esse sistema altamente regulado, em que a glicólise e a via das pentoses desempenham papel central.

1.2.1 – O gene *PKLR* e a glicólise

O gene *PKLR* (do inglês, *Pyruvate Kinase Liver and Red Blood Cells*) codifica a enzima piruvato cinase (PK), que atua na última etapa da via glicolítica em uma reação irreversível de transfosforilação do fosfoenolpiruvato em piruvato e ATP (Figura 1.3). De forma simplificada, após a ingestão e a metabolização de carboidratos pelo trato digestivo, a glicose liberada na corrente sanguínea pode ser captada por qualquer célula que demande energia através dos transportadores GLUT (do inglês, *Glucose transporter*). A velocidade desse transporte é acentuada pela ação da insulina. A quebra da glicose, um açúcar de 6 carbonos, em açúcares de 3 carbonos (piruvato) com a consequente produção de energia (ATP) é o processo ao qual se denomina glicólise, que ocorre em 2 fases. Na primeira fase, ou fase preparatória, 2 moléculas de ATP são utilizadas para converter a glicose em 2 moléculas de gliceraldeído-3-fosfato. Na segunda fase, ou fase produtora de energia, o gliceraldeído-3-fosfato é degradado em piruvato, com a produção de 4 equivalentes de ATP e 2 equivalentes de nicotinamida adenosina dinucleotídeo reduzida (NADH) (54). A PK, assim como a fosfofrutoquinase (PFK) e a hexoquinase (HK), atuam nas etapas irreversíveis da glicólise, e por isso são as principais enzimas reguladoras da via através de modificações covalentes ou dos níveis de substratos (55).

A glicólise ocorre no citoplasma de todos os seres vivos (exceto vírus) e, além da importância no fornecimento de energia para a célula, vários de seus intermediários fornecem ramificações para outras vias metabólicas, como a 6-fosfato-desidrogenase, que participa na via das pentoses, descrita adiante. Nas células nucleadas, a glicólise prossegue com as demais vias produtoras de ATP (ciclo de Krebs e cadeia respiratória) que ocorrem no interior das mitocôndrias. Entretanto, na ausência de mitocôndrias, como ocorre nos eritrócitos maduros, a glicólise se torna a via exclusiva de produção de energia e é a grande responsável pelo funcionamento e integridade dessas células (56). Nessas células, o ATP é necessário para a manutenção do formato da célula, fosforilação dos fosfolípidos e proteínas de membrana, transporte ativo de várias moléculas, síntese parcial de nucleotídeos de purina e pirimidina e síntese de glutatona reduzida (GSH) (57).

Nos mamíferos existem quatro isoformas da piruvato cinase (PK-M1, PK-M2, PK-L e PK-R) codificadas por dois genes distintos, *PKM* e *PKLR*, expressos em tecidos diferentes. O gene PK-M

codifica a piruvato cinase dos músculos. O *PKLR* codifica as proteínas PK-L, expressa no fígado, intestino delgado e córtex renal, e a PK-R, exclusivamente expressa nos eritrócitos. A isoforma PKM2 é produzido durante a maturação do eritrócito, porém é substituída por PKR na célula madura. O gene *PKLR* localiza-se no cromossomo 1 (1q21) e apresenta 12 éxons, dos quais dez são compartilhados pelas duas isoformas, enquanto os éxons 1 e 2 são específicos para as enzimas eritrocitária e hepática, respectivamente (58). As variantes genéticas nesse gene resultam em alterações tanto na enzima do eritrócito quando na enzima do fígado, porém os sintomas clínicos se devem às alterações nos glóbulos vermelhos, uma vez que a deficiência hepática é geralmente compensada pela síntese contínua nos hepatócitos (59). Assim, as manifestações clínicas da deficiência da PK compreendem diferentes graus de anemia (28,55).

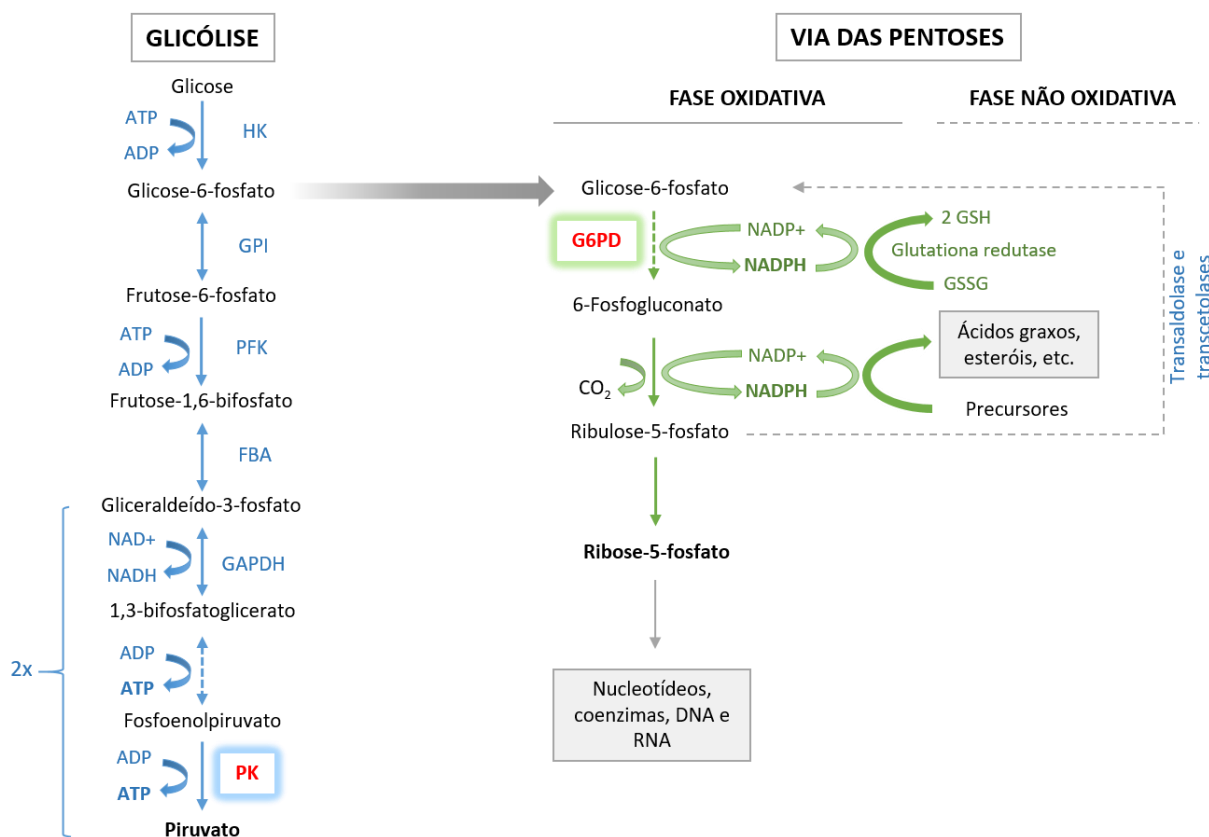


Figura 1.3 – Via glicolítica e via das pentoses. A primeira é a via principal e, a segunda, a via alternativa de oxidação da glicose presente no citoplasma de todas as células. Elas estão em equilíbrio e podem ocorrer simultaneamente, independentes uma das outras. O produto final da glicólise é o ATP. A via das pentoses

armazena a energia advinda da oxidação da glicose em NADPH. Em vias degradativas, como a glicólise, o substrato é oxidado gerando coenzimas reduzidas – NADH e FADH₂ – que serão oxidados para produzir ATP em vias subsequentes. Para a síntese de compostos biológicos, ocorre o reverso: há consumo de ATP e redução do substrato. Nesses casos, a coenzima utilizada é o NADPH, que é convertido em NADP⁺, que por sua vez é regenerado pela via das pentoses. A PK (piruvato cinase) atua na última etapa da via glicolítica e a G6PD (glicose-6-fosfato-desidrogenase) atua na primeira etapa da via das pentoses.

1.2.2 – Polimorfismos no gene *PKLR* e associação com a malária

As mutações no gene *PKLR* que levam à deficiência da PK ocasionam uma diminuição do aporte de ATP nos eritrócitos, diminuindo a meia-vida dessas células e provocando hemólise. A deficiência da PK é um traço autossômico recessivo bem caracterizado, com mais de 200 mutações definidas (60). Dentre os defeitos de enzimas glicolíticas, a deficiência da PK é a mais prevalente (1:20.000 na população caucasiana), e juntamente com a deficiência da G6PD, são a causa mais comum de anemia hemolítica não-esferocítica congênita (não-esferocítica se dá porque a hemácia não assume um formato de esfera, como em outras anemias) (61). Além da anemia hemolítica clássica, que varia de leve a totalmente compensada, onde são necessárias transfusões subsequentes, outros sintomas frequentes são reticulocitose, esplenomegalia e, mais raro, aumento de ferro (aumento de ferritina sérica), em que se faz necessário o uso de quelantes (55,62).

A associação da deficiência de PK com a malária foi descrita recentemente. Os primeiros estudos foram feitos em modelos de camundongo e culturas *in vitro*. Min-Oo e colaboradores, em 2003, observaram que camundongos recombinantes que carregam um locus de suscetibilidade a malária apresentaram resistência ao *Plasmodium chabaudi* devido a uma mutação (269T>A, éxon 3, Ile90Asn) de perda de função no gene *PKLR* (63). Já em 2008, dois trabalhos foram publicados comparando o crescimento do *Plasmodium* em eritrócitos humanos normais e com a deficiência em culturas *in vitro*, observando uma diminuição na infecção e replicação do parasita nos últimos. Interessante ressaltar que em 2009, pesquisadores do mesmo grupo constataram que pacientes que apresentaram resistência ao *Plasmodium* tiveram depleção de ATP nos eritrócitos (64,65).

A partir de 2010, os estudos genéticos em humanos começaram a endossar a relevância do gene *PKLR* com a malária. Nesse ano, *Alves e colaboradores* realizaram um estudo populacional em africanos de Cabo Verde, observando um desequilíbrio de ligação em extensão entre a região flanqueadora e o *loci* do gene *PKLR* nos indivíduos controles não infectados com malária (66). Os achados desse trabalho foram explorados pelo mesmo grupo. *Machado e colaboradores* realizaram a genotipagem de vários polimorfismos (SNPs e STRs) em diferentes grupos clínicos de malária de Angola e Moçambique e indivíduos PK-deficientes e sadios de Portugal, e obtiveram os seguintes resultados: i) as frequências alélicas dos marcadores do *PKLR* apresentaram-se altamente diferenciadas entre africanos e portugueses comparado aos marcadores neutros; ii) maior conservação da região do gene *PKLR* no grupo clínico de malária não complicada (desequilíbrio de ligação numa região mais extensa) e iii) associação positiva de um haplótipo com este grupo clínico (67). Ainda, um estudo subsequente realizado em 2012 mostrou que 4% dos indivíduos com malária de Moçambique, Angola, Guiné Equatorial e São Tomé e Príncipe apresentaram baixa atividade da enzima PK (68). Em 2015, *van Bruggen e colaboradores* caracterizaram a região genômica do *PKLR* na população da Tailândia e Senegal e identificaram variantes que podem alterar a estabilidade da enzima e proteger contra a malária (69).

Os últimos trabalhos correlacionando *PKLR* e malária foram publicados em 2017. *Laroque e colaboradores* indicaram que a suscetibilidade ao *P. chabaudi* em camundongos que apresentam a mutação protetora 90Ile>Asn é regulada por um locus denominado Char10, modulador da severidade da deficiência da PK e do nível de anemia (70). Por fim, o trabalho de *Faik e colaboradores*, publicado na *The Journal of Infectious Diseases*, levantou a hipótese de que a malária severa está relacionada com os polimorfismos do tipo CNV (do inglês, *copy number variations*) do gene *PKLR* na população do Gabão, África Central (71). Os trabalhos com populações humanas sustentaram os resultados anteriores obtidos em estudos *in vitro* e em camundongos. Nesse sentido, os estudos ao longo dos últimos 16 anos sugerem que a região do gene *PKLR* é alvo de pressão seletiva e mantém variantes genéticas em populações Africanas que impõem certa resistência à malária, justificando a sua frequência nessas populações mesmo frente ao prejuízo da deficiência enzimática. A questão que aparece a partir dessas evidências é: a frequência dos polimorfismos no gene *PKLR* é tão prevalente nas regiões endêmicas da malária

como as mutações de deficiência de *G6PD* e HbS? Para isso, é necessário expandir a análise genética em outras populações africanas e globais, além das estudadas até o momento.

Em indivíduos com deficiência na PK, o quadro de anemia é provocado pela redução da sobrevivência dos eritrócitos e pela eritropoiese irregular devido ao desequilíbrio nos níveis de ATP (72). Enquanto o eritrócito ainda está em fase de maturação, nas fases iniciais de reticulócito, ele é capaz de produzir ATP pela via fosforilativa. Porém, o eritrócito adulto perde o núcleo e as mitocôndrias e então o aporte de ATP depende exclusivamente da glicólise (73). Nesse sentido, a diminuição de ATP gera distúrbios na membrana da célula, provocando desidratação e lise das células. Além disso, a PK inibida aumenta o intermediário glicolítico 2,3-bifosfoglicerato, que inibe a G6PD, desativando o ciclo das pentoses e, com isso, aumentando o estresse oxidativo nas hemácias (74). Por esse motivo, o ciclo intracelular do *Plasmodium* fica prejudicado já que os eritrócitos não vivem o tempo suficiente para a completar a reprodução do parasito. Nesse aspecto, devido à lise prematura das células vermelhas com a consequente liberação de hemoglobina e heme, estes indivíduos podem desenvolver um aumento de ferro (28,65). Porém, para avaliar os níveis de ferro no paciente PK-deficiente, deve-se levar em consideração se o aumento não é proveniente de transfusões sanguíneas, já que nos casos mais graves de anemia, as transfusões são recorrentes (62).

1.2.3 – O gene *G6PD* e a via das pentoses

O gene *G6PD* é formado por 13 éxons, abrange cerca de 18.5kb e está localizado na região telomérica do braço longo do cromossomo X (Xq28). Este gene codifica a glicose-6-fosfato-desidrogenase (G6PD ou G6PDH), importante enzima metabólica do ciclo das pentoses. Como discutido anteriormente, as células animais degradam a glicose-6-fosfato até piruvato na via glicolítica. Porém, existem outros destinos catabólicos para a glicose-6-fosfato, entre eles o ciclo das pentoses. A via das pentoses-fosfato ou via das pentoses (PPP) é um braço da glicólise e funciona como uma via alternativa de obtenção de energia pela célula através do metabolismo da glicose. A G6PD é uma enzima citosólica expressa em todas as células humanas que tem importância na manutenção da integridade dos eritrócitos, evitando a oxidação da hemoglobina e de outras proteínas celulares (75). Essa enzima participa na primeira etapa do ciclo das

pentoses. A enzima G6PD oxida glicose-6-fosfato (G6P) a 6-fosfoglucoactona, reduzindo NADP a NADPH (Figura 1.3). Os produtos finais da via das pentoses são a pentose D-ribose-5-fosfato, utilizada na síntese de nucleotídeos (RNA e DNA), e a coenzima NADPH, principal doadora de hidrogênio em diversas reações enzimáticas e consumida na síntese de ácidos graxos (76). Eventualmente a ribose-5-fosfato por ser reciclada mediante a atividade de transcetolases e transaldolases regenerando glicose 6-fosfato, que podem seguir novamente a fase oxidativa, permitindo a formação contínua de NADPH. A G6PD funciona como a grande reguladora da via, é inibida quando as concentrações de NADPH/NADP⁺ estão altas e é ativa quando essa razão está baixa.

O NADPH (nicotinamida adenina dinucleotídeo fosfato) é responsável pela detoxificação celular, eliminando radicais livres e peróxidos de hidrogênio (H₂O₂). A remoção destes peróxidos é catalisada pela enzima glutathione peroxidase, que ao conjugar-se com a glutathione reduzida (GSH), forma a glutathione oxidada (GSSH). Essa, por sua vez, recebe um átomo de hidrogênio do NADPH, retornando a seu estado reduzido (GSH). O NADPH oxidado a NADP, na presença da G6PD, é reduzido novamente a NADPH (76). Esse ciclo é de extrema importância para os eritrócitos. As células vermelhas são ricas em radicais de oxigênio e peróxidos provenientes das reações de oxidação da hemoglobina, ou de fatores exógenos, como fármacos e infecções. Portanto, devido à importância na proteção contra o dano oxidativo, essas células apresentam elevadas concentrações de GSH (75). A relação da via glicolítica com a via das pentoses está esquematizada na Figura 1.3.

1.2.4 – Mutações no gene *G6PD* e associação com a malária

Evidências geográficas, epidemiológicas e *in vitro* fortalecem a hipótese de que a deficiência da G6PD protege contra o parasito *Plasmodium falciparum*. Embora a perda completa da função da enzima seja presumivelmente letal, como demonstrado em estudo *knockout* em camundongos, mutações que conferem uma deficiência parcial na atividade da enzima são comuns na África (77). Além disso, há um grau de conservação em certas regiões do gene que parecem ser essenciais para o desempenho da atividade enzimática e sobrevivência da célula (78). Apesar disso, mais de 180 variantes com importância clínica já foram descritas. Existe um

considerável número de variantes normais do *G6PD* na África, como a variante A e a variante B. Já as variantes relacionadas com a deficiência da enzima, particularmente frequentes (20-50%) na África e no Mediterrâneo, incluem a variante Med (C563T) comum no Sul da Europa, Oriente Médio e Índia e a variante A-, formada por duas mutações pontuais não-sinônimas, A376G e G202A, encontradas na África Subsaariana e em populações de origem Africana. Interessante ressaltar que as variantes A- e Med surgiram independentemente e aumentaram de frequência tão rápido que é difícil ser explicado pela deriva genética. Sugere-se que as mutações que levam ao alelo A- (ou variante A-) surgiram nos últimos 3.480 a 11.760 anos enquanto o alelo Med (ou variante Med) surgiu entre 1.600 e 6.640 anos atrás (79,80). Logo, essas datas indicam que as variantes foram fruto de pressão da malária, a qual teve um forte impacto no genoma humano há cerca de 10.000 anos (80). De fato, maiores frequências das variantes que conferem deficiência de G6PD são observadas em populações endêmicas para a malária (81), como na Figura 1.4.

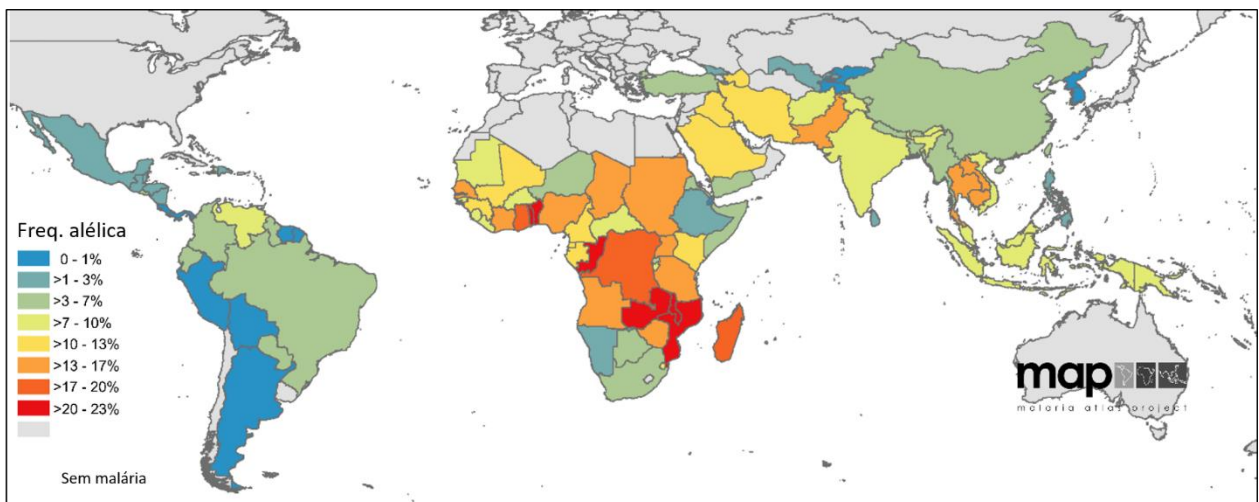


Figura 1.4 – Distribuição da frequência alélica das variantes de *G6PD* a nível global. (Adaptado de Howes, 2012).

A variante A- é uma das mais estudadas devido à sua relevância clínica e elevada frequência em africanos. A combinação da mutação A→G na posição 376 (éxon 5) e G→A na posição 202 (éxon 4) formam a variante A- (também chamada de Haplótipo A- ou Alelo A-) (82). Ambas as mutações são não-sinônimas do tipo *missense* (do inglês, “troca de sentido”), ou seja, a substituição de um único nucleotídeo altera o aminoácido na sequência do transcrito (existe também a mutação não-sinônima *nonsense* – “mutação sem sentido” – em que a alteração do

nucleotídeo leva à um códon de parada). A mutação A376G leva a alteração do aminoácido asparagina (Asn) para o aminoácido ácido aspártico (Asp) na posição 126 da proteína, enquanto a mutação G202A modifica o aminoácido metionina (Met) pelo aminoácido valina (Val) na posição 68 da proteína. A interação sinérgica das duas mutações próximas entre si no espaço da estrutura tridimensional da enzima provoca o dobramento inadequado da enzima, o que compromete a estabilidade e resulta em uma menor atividade (83). O haplótipo formado por essas mutações é encontrado em frequências superiores a 0.24 em populações africanas e diversos estudos já demonstraram o seu papel protetor frente a malária, como discutido no final deste tópico.

A maioria das mutações de deficiência da G6PD são substituições de nucleotídeo único que ocorrem nos éxons do gene e estão relacionadas a diferentes graus de atividade da enzima (84). A Figura 1.5 representa a distribuição dessas mutações nas regiões exônicas e a proporção por tipo de variação genética ao longo do gene *G6PD*. Em 1967, a OMS categorizou a deficiência de G6PD em cinco classes relacionadas ao grau de dano enzimático, como representado na Tabela 1.2. As variantes de classe I são as mais severas e ocorrem frequentemente nos éxons 6, 10 e 13, regiões de ligação ao substrato, interface dimérica e sítio estrutural do NADP+, respectivamente. Essas mutações levam à anemia hemolítica não esferocítica (CNSHA) e são mais facilmente detectadas a nível populacional devido à gravidade dos sintomas (84). As variantes de classe II possuem valores de atividade enzimática inferiores a 10%, acarretando sintomas e crises frequentes, tendo, como exemplo, a G6PD Med (C563T). A variante A- faz parte da classe III, que possui atividade enzimática entre 10 a 60% da atividade normal (83). Além dela, cita-se a variante G6PD A (A376G) e G6PD Asahi (G202A). Os indivíduos com as variantes da classe III geralmente são assintomáticos, mas a importância clínica está associada ao aparecimento de crises hemolíticas induzidas pela ingestão de medicamentos. As variantes de classe IV possuem de 60 a 150% da atividade normal, sem apresentar manifestações clínicas, tendo, como exemplo, a G6PD São Borja (G337A) (85). Por fim, as variantes de classe V possuem atividade enzimática aumentada (>150%) e são pouco descritas (86).

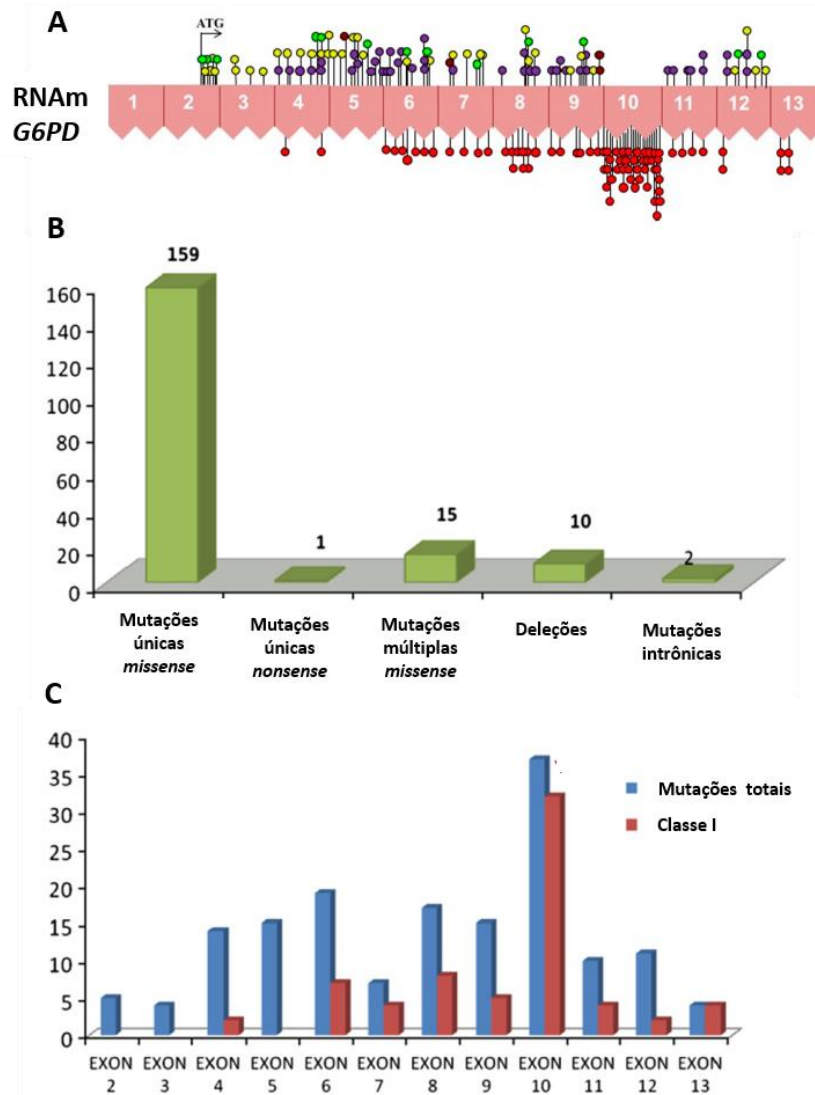


Figura 1.5 – Distribuição das mutações ao longo do gene *G6PD*. **A** – Disposição das mutações *missense* entre os 13 éxons do gene: mutações que conferem anemia hemolítica (vermelho), de classe II (roxo), de classe III (amarelo), de classe IV (marrom) e não reportadas (verde) (Adaptado de *Gómez-Manzo, 2016*). **B** – Número de mutações por tipo de variação e **C** – Número de mutações por éxons do gene. Os éxons com maior número de mutações são aquelas em que há maior quantidade de mutações de Classe I (Adaptado de *Minucci, 2012*).

Tabela 1.1 – Classificação das variantes de *G6PD* de acordo com os níveis de atividade enzimática (OMS).

| Classe | Porcentagem da atividade normal da G6PDH | Severidade da deficiência | Sintomas clínicos | Frequência em região geográfica | Variantes |
|--------|--|---------------------------|--|---------------------------------|---|
| I | <1% | Muito grave | Anemia hemolítica crônica | Rara, sem precisão geográfica | G6PD Bueno Aires G6PD Durban G6PD Volendam (C514T) |
| II | <10% | Grave | Anemia hemolítica aguda em resposta a drogas | Abundante no mundo | G6PD Med (C563T) G6PD Cassano G6PD Santamaria |
| III | 10-60% | Moderada | Anemia hemolítica crônica ou aguda | Abundante em regiões de malária | G6PD A- (A376G/ G202A) G6PD Seattle G6PD Canton G6PD A (A376G) G6PD Asahi (G202A) |
| IV | 60-150% | Atividade normal | Assintomático | Indefinido | G6PD São Borja (G337A) |
| V | >150% | Atividade aumentada | Assintomático | Indefinido | G6PD Hektoen |

(Adaptado de Silva, 2017)

Em geral, o mecanismo pelo qual as mutações em *G6PD* levam à deficiência da enzima correspondem à uma diminuição na estabilidade da enzima e uma redução nos níveis enzimáticos conforme o envelhecimento do eritrócito (87). Os indivíduos com deficiência da G6PDH são normalmente assintomáticos, mas podem desenvolver anemia grave no período neonatal ou anemia hemolítica aguda em resposta a potenciais indutores de estresse como algumas drogas (primaquina ou dapsona) ou o feijão fava (87). Apesar de a dapsona ter sido removida do mercado, o alarme está no uso da primaquina que é a única droga capaz de eliminar o *P. falciparum* na fase de gametócito com o uso em baixas doses (88). Portanto, a importância dessa deficiência tange o potencial risco de desenvolvimento dos sintomas frente o tratamento com primaquina em áreas endêmicas da malária (89). Considerando que o gene *G6PD* localiza-se no cromossomo X, a deficiência tem implicações importantes a nível populacional. Homens tendem a expressar mais frequentemente a deficiência da enzima. As mulheres heterozigotas, em contrapartida, apresentam duas populações de hemácias, criando um espectro de fenótipos, o que dificulta o diagnóstico (8,90).

Embora nos últimos anos algumas revisões e meta-análises tenham sido publicadas, estudos epidemiológicos ainda são controversos em relação à proteção se estender para ambos os sexos. Estudos caso-controle em diferentes populações Africanas (Kenya, Gambia, Maleses) comparando malária severa com malária assintomática demonstram que há associação de proteção entre mulheres heterozigotas e homens hemizigotos com a malária grave (91). Entretanto, Luzzato e colaboradores, que investigam a relação *G6PD versus* malária há mais de 40 anos, acreditam que a proteção se dá somente em mulheres heterozigotas, pois foi observado que as hemácias que apresentam o alelo selvagem em mulheres abrigam maior número de *P. falciparum* do que as células que expressam a mutação (92). Essa hipótese foi corroborada recentemente através do estudo de Uyoga e colaboradores (2015), difundido por um grupo de pesquisadores que tem grande *expertise* em malária, o qual mostrou que a proteção ocorre em mulheres heterozigotas que, além da resistência contra a malária grave, também tiveram menor mortalidade. Nesse mesmo estudo, homens não apresentaram proteção contra a malária, as suas hemácias ainda foram consideradas suscetíveis ao quadro de anemia gerado pela malária porque elas intrinsicamente sofrem de estresse oxidativo (91). O Consórcio *MalariaGen* também publicou um artigo na revista *Nature* em 2014 defendendo a hipótese de que o aumento da resistência se dá em heterozigose (93). Por fim, o papel de resistência das mutações em *G6PD* com a malária é indiscutível, confirmado em estudos genéticos e *in vitro*. Entretanto, mais trabalhos em diferentes populações são necessários para melhorar a conclusão acerca da proteção conferida por variantes nesse gene entre homens e mulheres.

As células vermelhas necessitam de ATP e de equivalentes redox para manter o formato bicôncavo da célula, as concentrações intracelulares de cátions, o estado reduzido da hemoglobina com um íon de ferro divalente, glutatona e os componentes da membrana. Caso haja deficiência em algum metabólito que prejudique a produção de ATP ou NADPH, a hemólise se instalará e as células lisadas serão removidas pelo sistema monócito-macrófago (28). A Figura 1.6 demonstra as diferenças entre os mecanismos que levam à hemólise entre a deficiência da PK e da G6PD. Ambas as deficiências levam à degradação da hemácia antes do período de 120 dias de vida celular, liberando os componentes da hemoglobina, como heme e ferro. Como esse

mecanismo pode influenciar na suscetibilidade a patógenos intracelulares, como as micobactérias, será o tópico discutido ao longo desse estudo.

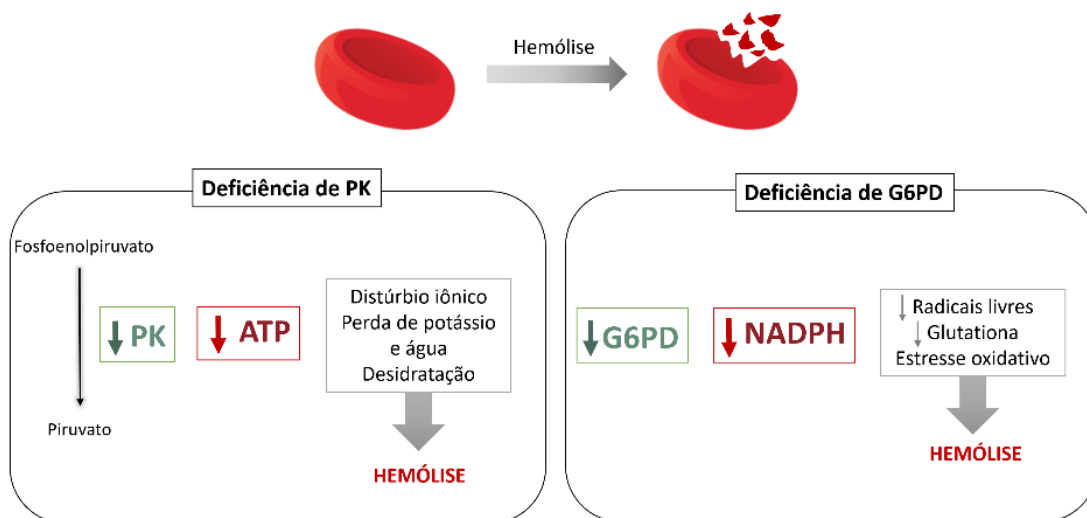


Figura 1.6 – Mecanismo de hemólise na deficiência da PK e da G6PD. A célula vermelha, na deficiência da PK, perde a integridade da membrana devido à diminuição de ATP. Na deficiência da G6PD, o aumento das espécies reativas de oxigênio (ROS) leva a degradação da hemácia frente à agentes indutores de ROS como a primaquina e a dapsona.

1.3 – Hanseníase

1.3.1 – Breve histórico e epidemiologia

A hanseníase é uma doença infecciosa milenar de caráter crônico que acomete a pele e os nervos periféricos podendo levar a danos físicos irreversíveis. É causada pelo patógeno intracelular obrigatório *Mycobacterium leprae*, que apresenta tropismo por macrófagos e células de Schwann de regiões superficiais do corpo, como a pele e os nervos periféricos (94). Por esse motivo, a manifestação se dá através do espessamento de nervos e de manchas eritematosas com limites imprecisos na pele, podendo ser acompanhada de perda de sensibilidade térmica, queda de pelos e perda de tônus muscular. As deformações de partes do corpo foram historicamente designadas como “lepra” (o nome foi mantido durante a tradução do Novo e Velho Testamentos), palavra utilizada negativamente como estigma social (95). A fim de evitar a segregação dos indivíduos afetados, a terminologia da doença foi substituída por “hanseníase”

em 1976 pelo Ministério da Saúde, em homenagem a Gerhard Armauer Hansen, que descobriu o bacilo da doença (96).

Apesar das controvérsias sobre a etiologia da doença, evidências arqueológicas confirmam a sua origem pré-histórica (97). O estudo de Monot e colaboradores a partir de genômica comparativa e análise filogenética de diferentes cepas, sugere que a doença surgiu no oeste africano e foi propagada para outras regiões do mundo pelo fluxo de migrações contínuas da humanidade (98). Em contrapartida, há indícios de que a hanseníase se originou na Ásia, a partir de análises em esqueletos encontrados na Índia e Tailândia, e se espalhou para a África e Europa no século IV a.C (99–101). Essa hipótese foi recentemente apoiada por Benjak e colaboradores que detectaram predominância de cepas ancestrais do bacilo no leste asiático (102). No Brasil, a doença foi introduzida por europeus e africanos durante o período colonial e o comércio de escravos no século XVIII (103).

Embora a hanseníase seja antiga, ainda é considerada um grave problema de saúde pública em países em desenvolvimento. Segundo o último levantamento da Organização Mundial da Saúde (OMS), 208.619 mil novos casos de hanseníase foram detectados no mundo em 2018. Esse número revela um decréscimo modesto, cerca de 15%, comparado aos últimos 10 anos (244.796 casos em 2009). O Brasil ainda não atingiu os objetivos globais de eliminação da hanseníase no mundo (definido pela prevalência menor do que 1 caso para 10.000 habitantes) da OMS, e juntamente com a Índia (120.334) e a Indonésia (17.017), apresentam os piores números no ranking de novos casos (28.660) (Figura 1.7). A detecção de casos novos apresentou um declínio no Brasil desde 2003 até 2016, entretanto, em 2017 esse número aumentou para 26.875 (em relação à 25.218 novos casos em 2016), e mais recentemente, para 28.660 novos casos, como demonstrado na Figura 1.8. Esse crescimento de 14% dos novos casos em dois anos aponta para a situação alarmante da doença no país. A incidência da hanseníase está relacionada com a falta de saneamento básico e as regiões Centro-oeste e Norte exibem as maiores taxas de detecção geral por 10.000 habitantes entre 2012-2016 no país, enquanto as menores taxas foram observadas no Sul e Sudeste. Em geral, os estados mais afetados são Mato Grosso, Tocantins, Rondônia, Pará e Maranhão. Nesse mesmo período, a taxa de detecção foi maior em homens para todas as faixa-etárias (104).

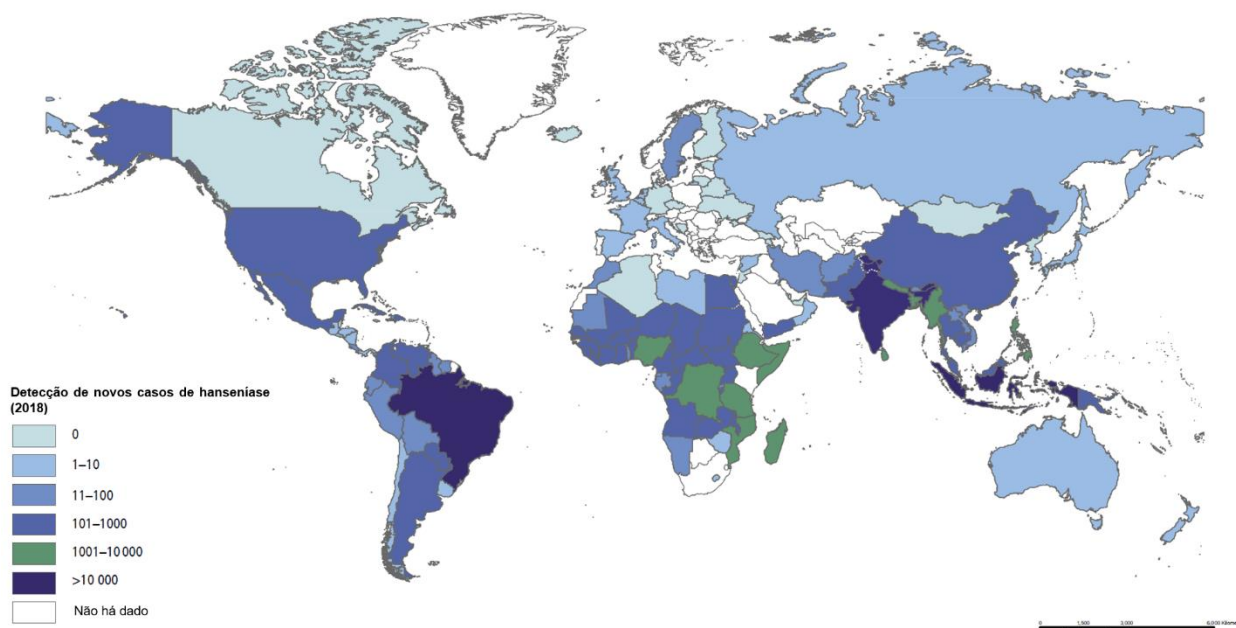


Figura 1.7 – Novos casos de hanseníase distribuídos no mundo em 2018 reportados pela OMS (*Weekly epidemiological record*, 2019). Fonte: <http://www.who.int/wer>

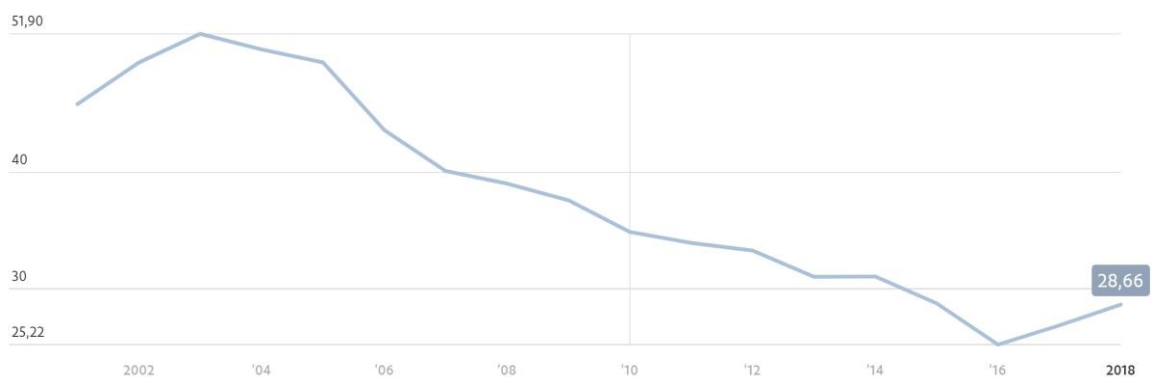


Figura 1.8 – Número total de novos casos no país de 2002 a 2018 (Ministério da Saúde).

1.3.2 – *Mycobacterium leprae*

O *Mycobacterium leprae* tem importância histórica pois foi o primeiro bacilo de infecção humana a ser identificado (1874), dando início a área da microbiologia. O segundo bacilo identificado foi o *Mycobacterium tuberculosis*, agente etiológico da tuberculose (105). Além disso, também foi uma das primeiras bactérias patogênicas a ter o genoma completamente

sequenciado. O *M. leprae* é uma bactéria gram positiva reta ou levemente encurvada com as bordas arredondadas. É um bacilo álcool-ácido resistente (BAAR), pois quando corado pelo método de Ziehl-Nilsen não descora pela ação da solução álcool-ácida. Desenvolve-se lentamente em temperaturas abaixo de 37 °C, e não cresce em meios de cultura axênicos *in vitro*, o que dificulta as investigações acerca da infecção (106). O modelo de infecção no coxim plantar de camundongos BALB/c ou atímicos (nude) constitui uma alternativa para o estudo da patogênese da hanseníase, permitindo o crescimento de bacilos viáveis para os estudos estruturais e bioquímicos (107). Outro modelo de infecção utilizado é o tatu de nove bandas (*Dasypus novemcinctus*), um hospedeiro animal com ocorrência no sul dos Estados Unidos e usado para o estudo do dano neural provocado pelo *M. leprae* (108–110).

O gênero *Mycobacterium* (família Mycobacteriaceae) contém mais de 180 espécies descritas atualmente dentre os quais o *M. leprae* e *M. tuberculosis* são as mais estudadas (111). Essas micobactérias têm como característica clássica a presença de lipídeos na forma de ácidos micólicos (ácidos graxos de elevado peso molecular) e lipoarabinomanana (LAM) em seu envelope celular. A estrutura capsular mais externa do *M. leprae* apresenta um glicolípido quimicamente distinto de outras bactérias, o glicolípido fenólico I (PGL-1) (112). Devido à especificidade do PGL-1 no *M. leprae*, a resposta humoral a esse bacilo é altamente específica (113). Essa estrutura tem relevância na interação patógeno-hospedeiro, pois está relacionada com o impedimento da passagem de drogas alvo de combate do bacilo, gerando resistência (107).

O sequenciamento do genoma do *M. leprae*, em 2000, trouxe importantes explicações acerca da dependência da célula hospedeira e da inabilidade do crescimento em cultura (114). O genoma do *M. leprae* é altamente conservado e reduzido em tamanho, apresentando 3.27 Mb (em comparação, o *M. tuberculosis* tem 4.41 Mb) (81,82). Devido à evolução redutiva ao qual o genoma desse bacilo foi submetido, genes importantes de funções catabólicas e respiratórias, de produção de sideróforos e de circuitos regulatórios foram perdidos (Figura 1.9) (115–117). Recentemente, o *Mycobacterium lepromatosis*, uma nova espécie de micobactéria que compartilha características microbiológicas, genéticas e clínicas similares com o *M. leprae* foi descrita. Como esse patógeno é capaz de levar a manifestações clínicas condizentes com a hanseníase, o *M. lepromatosis* é considerado um novo agente etiológico da doença (118). A similaridade entre esses bacilos levantou a hipótese da existência de um “complexo de *M.*

leprae”, análogo ao “complexo *M. tuberculosis*” existente (119). A Tabela 1.2 compara as principais diferenças genéticas entre *M. leprae*, *M. lepromatosis* e o *M. tuberculosis*. O genoma do *M. leprae* possui características únicas como o baixo número de genes codificantes (49,5% do genoma) e o elevado número de pseudogenes (1.116), enquanto 90,8% do genoma do *M. tuberculosis* são genes codificantes com pouquíssimos (6) pseudogenes (120,121). A sequência genômica do *M. lepromatosis* é tão reduzida quanto a do *M. leprae*, apresentando 3.21 Mb, e também é rica em pseudogenes (1.334) (115,122).

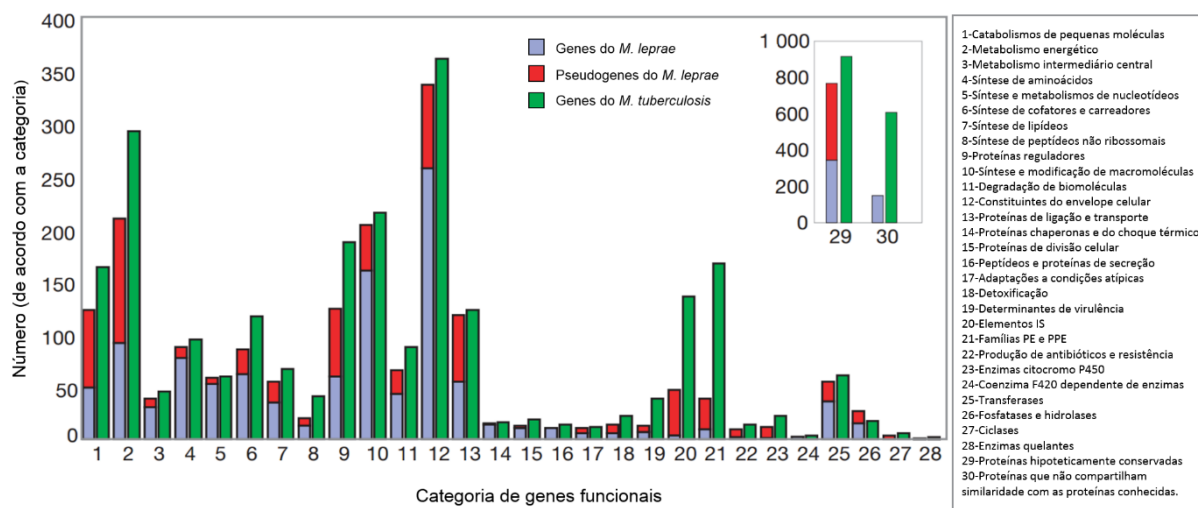


Figura 1.9 – Distribuição dos genes de *M. leprae* e *M. tuberculosis* de acordo com a categoria funcional.

O número de genes totais (cinza) e de pseudogenes (vermelho) do *M. leprae*, assim como o número dos genes totais (verde) do *M. tuberculosis* são mostrados nas barras em vertical. As categorias funcionais representadas pelos números 1-30 foram definidas no quadro (Adaptado de Cole, 2001).

Tabela 1.2 - Comparação entre os genomas do *M. leprae*, *M. lepromatosis* e *M. tuberculosis*.

| Característica | <i>M. leprae</i> | <i>M. lepromatosis</i> | <i>M. tuberculosis</i> |
|-------------------------|------------------|------------------------|------------------------|
| Tamanho do genoma (pb) | 3.268.203 | 3.206.741 | 4.411.532 |
| G+C (%) | 57,8 | 57,8 | 65,6 |
| Genes codificantes (n°) | 1.614 (45,9%) | 1.477 | 3.959 (90,8%) |
| Pseudogenes (n°) | 1.306 | 1.334 | 6 |

(Adaptado de Cole, 2001; Singh, 2015 e Benjak, 2018).

1.3.3 – Classificação clínica

A maioria dos indivíduos apresenta resistência ao *M. leprae*. Estima-se que cerca de 8-10% dos indivíduos expostos ao bacilo progridem para a doença ativa. A intensidade da resposta desenvolvida, influenciada pelo caráter genético do hospedeiro, direciona os fenótipos da doença, que podem desenvolver-se até 30 anos após a infecção (123). Uma vez que o mecanismo de resposta não seja suficiente, a hanseníase pode progredir para uma forma clínica inicial e transitória denominada intermediária (I) ou para uma das formas clínicas estáveis da doença (124).

A classificação das formas clínicas é de grande importância para o diagnóstico, pois é utilizada para determinar o tratamento apropriado. A mais utilizada é a de Ridley e Jopling (1966), que propõe a existência de um espectro de formas clínicas segundo o estágio clínico e imunológico dos pacientes. Nesse espectro, a hanseníase é subdividida em 5 formas clínicas: tuberculóide (TT), *borderline* tuberculóide (BT), *boderline boderline* (BB), *borderline* lepromatosa (BL) e lepromatosa (LL). As formas TT e LL representam os pólos que exibem as formas localizada e disseminada da doença, respectivamente. Os pacientes LL apresentam numerosas lesões com grande chance de comprometimento neural. Nesse polo a resposta imune humoral (Th2) é predominante, com perfil de citocinas ineficientes no controle do bacilo (IL-10 e IL-4) e, por isso, apresenta uma alta carga bacilar. Os indivíduos TT têm poucas lesões com bordas elevadas e bem delimitadas. Há maior atividade da imunidade celular (Th1), que induz a produção de citocinas tais como a IL-2 e IFN- γ , que vão ativar as respostas microbicidas. Com isso, o macrófago consegue conter a disseminação dos bacilos, e por conta disso, a baciloscopia é baixa ou negativa. Entretanto, a resposta imune celular intensifica o dano tecidual observado na lesão. Entre estes polos estão as formas intermediárias, que apresentam imunidade instável variando de acordo com a proximidade a cada forma polar (125). Ainda, existe a forma neural pura (NP), que se caracteriza pela ausência de lesões na pele e espessamento de nervos. Normalmente, o diagnóstico deste quadro é mais dependente de técnicas laboratoriais como a histopatologia ou a técnica de PCR em tempo real (do inglês, *polymerase chain reaction*), a partir da detecção de material genético bacteriano da biópsia do nervo (126).

Existe também a classificação operacional da OMS (1982) formulada para auxiliar no tratamento. Essa classificação é baseada no número de lesões e na carga bacilar, dividindo os pacientes em paucibacilares (PB) e multibacilares (MB). Pacientes PB apresentam baciloscopia negativa e até cinco lesões de pele, enquanto os MB têm baciloscopia positiva e seis ou mais lesões na pele. As formas clínicas BT e TT são normalmente classificadas como paucibacilares, enquanto as formas BB, BL e LL são os casos multibacilares.

Os episódios reacionais são processos inflamatórios intensos que acometem repentinamente os pacientes durante qualquer estágio da doença e até pós-cura. As perturbações imunológicas que levam aos episódios reacionais ainda não são entendidas, no entanto, postula-se que as formas polares são mais estáveis, enquanto as formas intermediárias são mais propensas a desenvolver esses estágios (127,128). A reação tipo 1, ou reação reversa (RR), é um quadro de inflamação aguda na pele e em nervos que ocorre principalmente em pacientes das formas *borderline*. Nesse episódio, há aumento das citocinas IL-2, IFN- γ e TNF e diminuição das citocinas IL-4, IL-5 e IL-10, evidenciando um súbito aumento da imunidade celular direcionada aos antígenos do *M. leprae* (129). A reação tipo 2, ou eritema nodoso hansênico (ENH), é um episódio crônico que ocorre com mais frequência em pacientes lepromatosos e *borderline*-lepromatosos. Manifesta-se por lesões eritematosas e nódulos inflamados acompanhados de sintomas sistêmicos como febre, prostração e dor. As reações tipo 1 e 2 ocorrem separadamente, mas podem progredir no mesmo paciente em momentos distintos (130). Acredita-se que os estágios reacionais compartilham os mesmos mecanismos de controle molecular, já que o perfil de citocinas liberados em ambos os episódios é semelhante (106,131). A Figura 1.10 resume a classificação das formas clínicas da hanseníase e seus episódios reacionais.

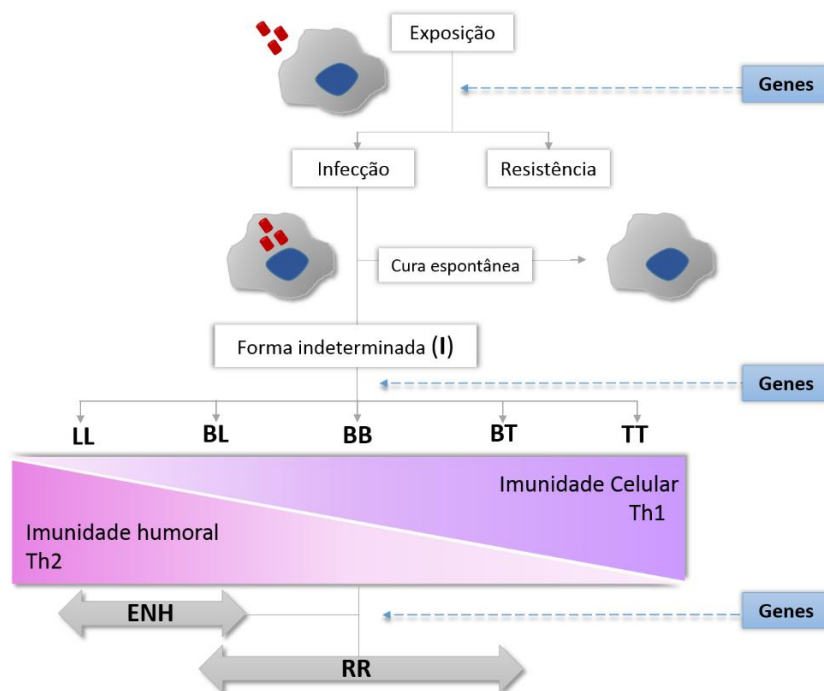


Figura 1.10 – Classificação das formas clínicas da hanseníase baseado em Ridley e Jopling (1966). Após a exposição, indivíduos suscetíveis progridem para uma das formas clínicas que dependem do tipo de resposta imune desenvolvida pelo hospedeiro. Os fatores genéticos influenciam os diferentes estágios desse modelo. LL – lepromatosa-lepromatosa; BL – *borderline*-lepromatosa; BB - *borderline-borderline*; BL – *borderline*-tuberculóide; TT – tuberculóide- tuberculóide. Os estágios reacionais compreendem o ENH (eritema nodoso hansênico) e a RR (reação reversa) (Adaptado de Mira, 2006 e Sauer, 2015).

1.3.4 – Transmissão, diagnóstico e tratamento

O mecanismo exato de transmissão da hanseníase ainda é incerto. Acredita-se que a propagação do *Mycobacterium leprae* se dê pela inalação do bacilo devido ao contato direto e prolongado com pacientes multibacilares sem tratamento (95). Diversas evidências sugerem que o trato respiratório superior é a principal rota de infecção. O DNA do bacilo já foi identificado na mucosa nasal e em *swab* bucal de pacientes e de indivíduos que residem no mesmo local dos pacientes, os chamados contatos domiciliares (132). Além disso, a taxa de infecção é maior quanto mais alto o contato com o *M. leprae*. Desse modo, pacientes com alta carga bacilar são considerados a principal fonte de transmissão da doença e o ambiente intradomiciliar é o principal foco de transmissão (94). Especula-se que quanto mais intenso o contato, maior a

exposição. No estado do Pará, região com alta incidência da doença e que possui uma densidade populacional média alta (4,1 pessoas por família), metade dos pacientes vivem em casas nas quais duas ou mais pessoas compartilham um quarto (133). Assim, uma atenção especial deve ser dada para os contatos domiciliares, já que eles apresentam mais chances de adoecer (134,135). Vale ressaltar que pacientes paucibacilares eventualmente podem apresentar uma carga bacilar alta o suficiente para viabilizar a transmissão (134).

Contudo, o contato com o bacilo não é suficiente para o desenvolvimento da doença. O status nutricional e as condições socioeconômicas, como a presença de saneamento básico e moradias pouco arejadas e populosas são alguns dos fatores ambientais que auxiliam no aparecimento da doença. Além desses, o caráter genético do indivíduo reflete a maior suscetibilidade ou resistência à doença, como discutiremos no tópico seguinte. Salienta-se que ainda não existe vacina para a hanseníase. Porém, estudos mostram que indivíduos que foram vacinados por BCG (do inglês, *Bacillus Calmette-Guérin*) na infância têm menos chances de desenvolver a doença e progredir para os polos mais severos (136). Por outro lado, um estudo de 2015 revelou que contatos domiciliares desenvolveram a hanseníase paucibacilar semanas após a administração da vacina BCG (137).

O controle da cadeia de transmissão é baseado na detecção precoce dos novos casos da doença e no tratamento adequado dos pacientes, principalmente os multibacilares que são disseminadores ativos do bacilo. Nesse cenário, a primeira complicação é visualizada: o diagnóstico é essencialmente clínico baseado na avaliação dermatoneurológica do paciente, por meio de testes de sensibilidade, palpação de nervos, avaliação da força motora e identificação de manchas hipopigmentadas ou avermelhadas na pele, e no histórico familiar. A hanseníase tem um longo período de incubação (média de 5, podendo chegar a mais de 20 anos), e as manifestações clínicas podem ocorrer anos após a infecção. Além disso, as lesões dermatológicas da doença são frequentemente confundidas com outras dermatoses, como alergias, doenças autoimunes e fungos (138). O diagnóstico laboratorial é realizado através da baciloscopia do raspado dérmico de lesões, lóbulos auriculares e cotovelos a fim de realizar a contagem de bacilos pela coloração de Ziehl-Nielsen. Porém, algumas formas clínicas apresentam baciloscopia negativa, e requerem a utilização de outros métodos de diagnóstico (139). A histopatologia é outro exame complementar que permite visualizar bacilos BAAR pela coloração de hematoxilina-

eosina e Fite-Faraco em biópsias da lesão na pele ou do nervo. A morfologia do tecido, a identificação de infiltrado mononuclear, epitélio granulomatoso e fibrose, características essenciais das lesões hansênicas também podem ser visualizados (140,141).

Testes sorológicos para a detecção de anticorpos anti-PGL1, um componente exclusivo da parede celular do *M. leprae*, também são utilizados em ensaios como ELISA, mas pouco empregados na clínica devido à baixa sensibilidade do método (106,142). Recentemente, testes baseados na detecção de uma proteína de fusão, denominada LID-1, mostram uma relação direta entre os títulos de anticorpos e a carga bacilar (143). Além disso, em casos difíceis (pacientes com lesão neural pura ou paucibacilares), podem ser utilizados ensaios moleculares baseados na técnica de reação em cadeia de polimerase (PCR) empregando amostras clínicas como linfa, sangue, secreção nasal e fragmentos de pele e nervos (144,145). Contudo, os métodos sorológicos e moleculares não são aplicados na rotina da atenção básica. O diagnóstico tardio tem sido um dos grandes obstáculos ao controle da hanseníase, pois inviabiliza o tratamento anterior ao acometimento neural e prejudica o bloqueio da transmissão (146).

A dapsona foi o primeiro medicamento eficaz no tratamento da hanseníase, capaz de inibir o crescimento bacteriano. Esse fármaco, descoberto em 1940, foi utilizado para hanseníase quanto para tuberculose. Porém, o surgimento de cepas resistentes levou a implementação da poliquimioterapia (PQT) pela OMS (1982), a qual é baseada na associação de três medicamentos: dapsona, rifampicina e clofazimina (138). Essa combinação evita a proliferação de cepas resistentes e, quando utilizada corretamente, apresenta uma alta chance de cura. A PQT inibe a capacidade de transmissão do bacilo já no início do tratamento. A administração é realizada gratuitamente seguindo a classificação operacional dos pacientes (146). Os pacientes PB utilizam uma combinação de rifampicina e dapsona com duração de 6 meses, já os pacientes MB fazem uso da combinação de rifampicina, dapsona e clofazimina durante o período de 12 meses (147). A talidomida é utilizada em pacientes que evoluem para episódios reacionais, como o ENH (148). Apesar do tratamento ser gratuito, a eficácia e a forma de administração são discutíveis pois são observadas altas taxas de recidiva da hanseníase entre os pacientes que finalizam o ciclo da PQT e que são, teoricamente, considerados curados (149,150).

1.3.5 – Papel do ferro na infecção pelo *M. leprae*

O ferro é um nutriente essencial tanto para o ser humano quanto para os microrganismos patogênicos, que tiveram que expandir sua capacidade de competir por esse elemento com o hospedeiro a fim de garantir a sua disseminação. Funções críticas estão associadas com o ferro em muitos processos celulares. Devido à capacidade de alternar entre dois estados de oxidação, ferroso (Fe^{2+}) ou férrico (Fe^{3+}), o ferro funciona como um catalisador redox, aceitando ou doando elétrons. Contudo, o potencial redox do ferro também gera toxicidade celular em condições de sobrecarga de ferro. O ferro pode reagir com as espécies reativas de oxigênio (ROS) geradas durante o curso da homeostase celular e danificar lipídios, DNA e proteínas. No organismo humano, o ferro é adquirido através da alimentação e da senescência das células vermelhas, onde há a maior concentração de ferro.

Além de mitigar a toxicidade, a regulação da distribuição do ferro pode ser considerada um mecanismo imunológico contra patógenos invasores. Mesmo na ausência de infecção, o metabolismo de ferro garante que ele seja dificilmente acessível a microrganismos patogênicos. Primeiro, a maioria do ferro encontra-se no ambiente intracelular, complexado dentro da hemoglobina nos eritrócitos. Segundo, a escassez de ferro extracelular livre também é um fator limitante para a disponibilidade de ferro a patógenos invasores (151,152). O ferro extracelular é ligado com alta afinidade pela transferrina (Tfr), proteína transportadora de ferro, que em indivíduos saudáveis é tipicamente menos de 50% saturada com ferro. Quando a ligação à Tfr excede a capacidade dessa proteína, moléculas do plasma como albumina, citrato, e aminoácidos podem funcionar, com menor afinidade, como quelantes de ferro (152). Frente à uma infecção, a retenção de ferro pelo organismo aumenta. A diminuição do ferro no plasma é visto em infecção por patógenos como o *Staphylococcus aureus*, e sugere-se que esse ferro seja armazenado em ferritina a fim de diminuir o ferro livre disponível para a bactéria (153). Logo, os patógenos desenvolveram formas de sequestrar esse elemento essencial à sobrevivência, como a capacidade de lisar a hemácia e liberar hemoglobina para extrair o ferro do heme.

A sobrevivência e virulência bacteriana durante o processo de infecção depende do sequestro de ferro, que também funciona como um cofator para diferentes enzimas do metabolismo celular (154). Em micobactérias, o ferro é cofator para hemoproteínas do

metabolismo do oxigênio e é utilizado por citocromos envolvidos no transporte de elétrons (155). Acredita-se que, para um crescimento ótimo, as micobactérias necessitam de cerca de 7 - 48 µg de ferro por grama de massa celular. Estudos *in vitro* mostram que concentrações de ferro abaixo dos valores de referência restringem o crescimento micobacteriano (156). Foi visto que a suplementação de ferro em indivíduos portadores de *M. tuberculosis* aumenta a multiplicação do bacilo e a chance de progressão para a tuberculose ativa (156,157). Indivíduos HIV (do inglês, *Human Immunodeficiency Virus*) positivos apresentam acúmulo de ferro nos macrófagos, o que pode explicar a grande predisposição às infecções possivelmente letais provocadas pelo complexo do *Mycobacterium avium* (158).

Um dos mecanismos mais especializados de captação de ferro por algumas micobactérias são os sideróforos, conhecidos como micobactinas (159). Os sideróforos são proteínas bem caracterizadas em *M. tuberculosis* que se ligam ao ferro com alta afinidade transportando-o para os compartimentos de interesse (160). O gênero *Mycobacterium* desenvolve-se no interior de fagossomos, vacúolos que se localizam no interior de macrófagos e são pobres em ferro. A micobactina, localizada na membrana da bactéria, forma sequestra o íon férrico, mesmo quando esse íon não é abundante, e o internaliza para o citoplasma bacteriano através dos receptores de regulação IrtAB (161,162). A aquisição do ferro citoplasmático do hospedeiro é, ainda, facilitado pela habilidade da micobactina de se difundir para fora do fagossomo, quelar ferro e se reassociar ao fagossomo através dos corpúsculos lipídicos (160). Existem ainda as carboximicobactinas (exoquelinas), que capturam e doam ferro para as micobactinas, além de roubar o ferro da transferrina para o fagossomo (163). O *M. tuberculosis* também adquire ferro através das proteínas transferrinas e lactoferrinas humanas (164,165). Além disso, essa espécie é capaz de captar ferro através do sistema heme/hemoglobina por mecanismos ainda não compreendidos (166). Por fim, a resposta do hospedeiro via ativação de IFN-γ em macrófagos diminui a expressão de receptores de Trf, diminuindo a concentração de ferro nos endossomos (167). Então, a ferroportina localizada nos fagolisossomos infectados com o *M. tuberculosis* fazem o efluxo do ferro desse compartimento celular. O bacilo, em contrapartida, consegue modular a expressão da ferroportina e circunvir esse mecanismo de defesa do hospedeiro (168). Estudos mostram que a inibição da síntese de micobactinas prejudica o crescimento do *M. tuberculosis* dentro dos macrófagos (161). Entretanto, a cepa nocaute para micobactina é capaz de manter a viabilidade

micobacteriana, ressaltando que apesar da importância dos sideróforos na sobrevivência micobacteriana, outros mecanismos para obtenção de ferro existem (169). A Figura 1.11 mostra os possíveis mecanismos de captação de ferro descritos até o momento no modelo de *M. tuberculosis*.

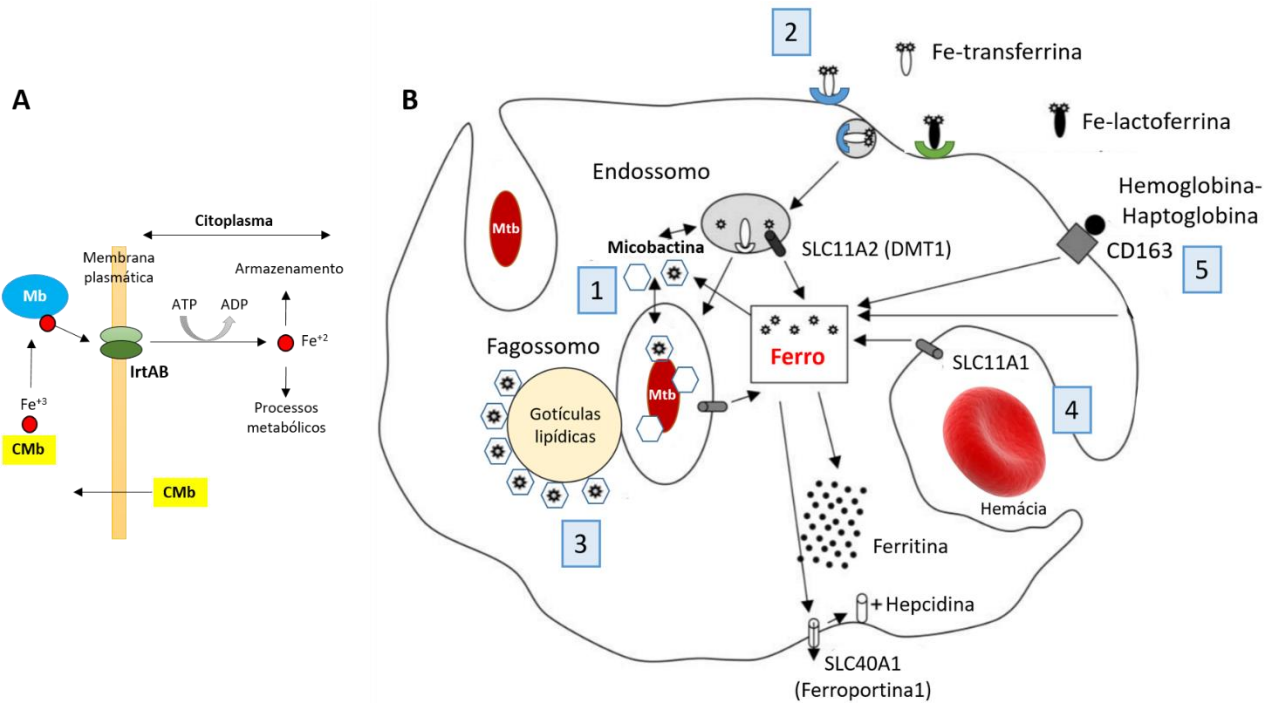


Figura 1.11 – Vias de aquisição de ferro pelo *M. tuberculosis* (Mtb) no nicho intracelular. A – O mecanismo mais especializado é a captura de Fe^{+3} do hospedeiro via micobactinas (Mb) e carboximicobactinas (CMb). **B –** Os seguintes mecanismos são propostos na literatura: **1)** internalização direta do ferro citoplasmático da célula hospedeira via micobactinas solúveis através do receptor IrAB (a carboximicobactina veicula ferro para micobactinas lipofílicas presentes na parede celular do Mtb (Figura 1.11 A); **2)** Captação via receptores de transferrina e lactoferrina através da interação com o endossomo (a transferrina se liga ao seu receptor e é internalizada para o endossomo, onde o baixo pH separa o ferro da transferrina e permite a reciclagem e a liberação do ferro para o citoplasma via receptor DMT-1, ou SLC11A2); **3)** Através da difusão da micobactina para fora do fagossomo e internalização ao fagossomo associado aos corpúsculos lipídicos; **4)** Diretamente do heme presente em hemácias via receptor SLC11A1/NRAMP1 presente dentro do vacúolo de endocitose da hemácia senscente (ainda especulativo) e, por fim, **5)** Através do complexo Hemoglobina-haptoglobina que é internalizado via receptor CD163. Quando liberado no citoplasma do macrófago, o ferro pode ser armazenado em ferritina, utilizado no metabolismo da célula ou exportado via Ferroportina 1 (receptor SLC40A1). A hepcidina é um regulador

do transporte de ferro, promovendo efluxo quando em baixas concentrações. Quando a hepcidina é aumentada (como ocorre na abundância de ferro ou na presença de inflamação), essa proteína se liga à Ferroportina 1, o complexo é internalizado e o efluxo de ferro é interrompido (Adaptado de Luo, 2005; Boelaert, 2007 e Cassat, 2013).

O *Mycobacterium leprae*, por sua vez, não apresenta o sistema especializado de sideróforos. Estudos da década de 80 sugeriam a presença de exoquelinas-*like* em experimentos *in vitro*, mas não havia evidências de que esse bacilo produzia essas moléculas. Acreditava-se que essas exoquelinas eram adquiridas através da produção de outras micobactérias na célula hospedeira (170–172). Após o sequenciamento do genoma do *M. leprae*, a hipótese estabelecida foi a de que essa espécie perdeu os genes responsáveis pela produção desses quelantes ao longo da evolução reductiva (173). Porém, esse patógeno provavelmente tem proteínas alternativas reguladoras de ferro, como as IREPs (do inglês, *iron regulated envelope proteins*), detectadas no envelope da bactéria, o que ajuda a explicar a sua sobrevivência no ambiente hostil do hospedeiro (174). Pessolani e colaboradores demonstraram a existência de um tipo de bacterioferritina (Brf) utilizado pelo *M. leprae*, a proteína de membrana MMP II, relacionado a homeostase de ferro nessa espécie (175). Contudo, ainda não existem trabalhos que mostrem o mecanismo exato pelo qual o *M. leprae* adquire ferro para sua sobrevivência.

Até o momento, estudos têm sugerido a contribuição de moléculas envolvidas na regulação das células vermelhas para a demanda de ferro do *M. leprae*. As hemoglobinas formam um complexo proteico que armazena a maior parte do ferro existente no organismo. A renovação de hemácias envolve a liberação de hemoglobina e o reaproveitamento do ferro contido no heme por um processo altamente regulado no interior de macrófagos (176). Frente à hemólise, a Hp se liga com alta afinidade à Hb e é internalizada via receptor CD163 (177). Eventualmente, a Hb livre no citoplasma pode também se ligar diretamente ao receptor CD163 (178). O heme internalizado é quebrado pela heme-oxigenase (HMOX-1) em monóxido de carbono, bilirrubina e ferro, que pode ser direcionado para armazenamento (ferritina), transporte (transferrina e ferroportina) ou outras demandas celulares (179).

Um estudo do nosso grupo demonstrou que macrófagos de pacientes lepromatosos apresentam uma maior expressão do receptor CD163 associado ao aumento dos estoques de

ferro e viabilidade do *M. leprae* (180). O receptor CD163 é um *scavenger* do complexo hemoglobina-haptoglobina e está relacionado ao ciclo de renovação das hemácias. A citocina IL-10 ativa a expressão de CD163 em macrófagos residentes e monócitos infiltrantes de tecidos em reações inflamatórias (181). Em 2017, demonstramos que biópsias de pacientes multibacilares (LL) apresentam aumento da expressão de hemoglobina, haptoglobina, heme-oxigenase 1 e receptor de transferrina 1 (Tfr1) comparado aos pacientes tuberculóides (BT). Observamos também um acúmulo de ferro nas lesões LL acompanhado da diminuição da Ferroportina 1 (182). Assim, o aumento de CD163 em macrófagos LL, além de induzir um fenótipo anti-inflamatório, pode indicar uma via de suplementação de ferro para o *M. leprae*.

Embora a captação e o metabolismo do ferro sejam indispensáveis para o processo de infecção micobacteriana, como esse elemento é utilizado e captado da célula hospedeira ainda não está claro. Os macrófagos são as células que mais armazenam ferro no organismo humano, porém, o processo pelo qual o ferro é trafegado até o fagossomo, onde se encontra o *M. leprae*, também é desconhecido (183). É possível que outras vias também regulem a disponibilidade de ferro no ambiente intracelular e o fluxo para o bacilo durante o curso da infecção.

1.3.7 – Suscetibilidade genética à hanseníase

O fato de a hanseníase ser contagiosa, mas apresentar baixa morbidade (já que a maioria das pessoas é naturalmente resistentes à doença) pode ter explicação genética. A heterogeneidade de formas clínicas, a capacidade de resistência da maioria dos indivíduos expostos ao patógeno e a baixa variabilidade do bacilo sugerem que o perfil genético do hospedeiro é crucial para o desenvolvimento da doença (114). A hipótese de que a hanseníase apresenta um componente hereditário vem sendo explorada desde a Idade Média e foi confirmada por estudos observacionais em gêmeos, análise de segregação complexa e análises de agregação familiar (pedigrees) (184,185). A partir da era genômica, foi possível identificar vários genes relacionados à hanseníase. Alguns dos achados são mais sólidos, como a associação com os genes *TLR1*, *NOD2*, *PARK2* e *TNF*, com replicações em populações independentes e conclusões acerca do papel funcional do polimorfismo na doença. Outros, todavia, ainda não foram bem caracterizados e apresentam direções de associação controversos em diferentes

populações. Atualmente, polimorfismos em genes da resposta imunológica são os mais bem caracterizados. Abordaremos inicialmente os genes envolvidos com a resposta imune inata, e, posteriormente, os envolvidos com a resposta imune adaptativa. Revisões acerca dos genes associados com a hanseníase podem ser acessadas nas referências: (123,128,186,187).

A resposta imune inata, a primeira linha de defesa contra o *M. leprae*, apresenta componentes cruciais para o combate inicial do bacilo além de direcionar para a resposta imune adaptativa (188). As células têm receptores de superfície que regulam as sinalizações da resposta inata, como os TLRs (*Toll-like receptors*), que reconhecem os microrganismos e ativam a célula; VDR (do inglês, *Vitamin D receptor*), receptor de vitamina D que ativa a via antimicrobiana mediada pela vitamina D juntamente com o TLR2 (*Toll-like receptor 2*), induzindo a produção de catelicidinas; MRC (*Manose receptor type C*), receptor de manose que juntamente com o TLR e NOD2 modulam a autofagia e as interações patógeno-hospedeiro e o receptor intracelular receptores NOD ou NLRs, (*Nod-like receptors*) que reconhece patógenos intracelulares (189). Os receptores TLR1/TLR2, presentes em monócitos/macrófagos, reconhecem lipoproteínas da parede das micobactérias e ativam essas células, normalmente induzindo a produção de vitamina D e peptídeos microbicidas contra o bacilo (190). Já os receptores TLR1/TLR6 estão relacionados à persistência do *M. leprae* nas células de Schwann (191).

O polimorfismo de base única, ou SNPs (do inglês, *single nucleotide polymorphisms*), são os marcadores genéticos mais estudados nos genes descritos a seguir. No gene *TLR1* o SNP N248S é muito estudado. O genótipo SS e NS desse polimorfismo foram associados à proteção contra o desenvolvimento de reação ENL à hanseníase *per se*, respectivamente, na população de Bangladesh (192). Já na população brasileira, um trabalho recente do nosso grupo revelou que a variante 248S está associada a suscetibilidade a hanseníase a partir de estudos caso-controle e em famílias conduzidos em populações de diferentes regiões do país. Demonstramos que essa variante provoca diminuição da razão *TNF/IL-10* em células polimorfonucleares estimuladas com *M. leprae*. Além disso, a meta-análise deste estudo junto com o de Bangladesh confirmou a associação de risco da variante 248S com a hanseníase (193). Além disso, dois polimorfismos (597C>T e um marcador microssatélite) no *TRL2* mostraram risco aumentado de desenvolvimento de reações reversas numa população da Etiópia (194). O gene *TLR4* também contém dois SNPs (896G>A, 1196C>T) com papel protetor ao desenvolvimento da hanseníase em africanos (195).

Recentemente, variantes nos genes *TLR1*, *TLR2* e *TLR4* foram associadas com a diferentes níveis de produção de citocinas em pacientes com hanseníase em Salvador (196).

Um estudo de rastreamento genômico em hanseníase foi conduzido em famílias Indianas e detectou a ligação de um *locus* de suscetibilidade na região 10p13 com pacientes paucibacilares (197). O gene *MRC1*, que codifica o receptor de manose, localiza-se nessa região e foi associado, posteriormente, ao risco de desenvolver a forma multibacilar da hanseníase no Malawi, Brasil e Vietnã (198,199). Em 2014, Grant e colaboradores verificaram a associação dos genes de *CUBN* (rs10904831) e *NEBL* (rs11012461) (ligado à região 10p13) ao desenvolvimento de hanseníase multibacilar em famílias Vietnamitas. Logo, a região 10p13 parece abrigar variantes genéticas que têm impacto sobre a polarização da hanseníase (200). O gene *MBL* codifica a proteína ligadora de manose que é solúvel e participa na resposta imune inata através da ativação do sistema complemento e na opsonização de patógenos. A diminuição dessa proteína no plasma predispõe à hanseníase lepromatosa (201). A variante 161G>A desse gene foi associada com a proteção ao grupo lepromatoso (LL) numa população do Nepal (202). Na população brasileira, a associação de polimorfismos no *MBL* apresenta resultados controversos em diferentes regiões (202–204).

O receptor NOD2 reconhece, auxilia na autofagia e modula a resposta ao *M. leprae*. Em 2009, o primeiro estudo do tipo GWAS (do inglês, *genome-wide association study*) com a hanseníase foi realizado por Zhang e colaboradores na população Chinesa, pontuando uma série de genes (*TNFSF15*, *NOD2*, *RIPK2*, *CCDC122*, *LACC1* e *HLA-DR-DQ*) associados à hanseníase. Nesse estudo, os SNPs rs9302752 e rs7194886 do gene *NOD2* foram associados com o risco ao desenvolvimento da hanseníase (205). Uma estudo posterior, também em Chineses, identificou a associação do rs7194886 com a tuberculose (206). Já numa população do Vietnã, a associação do rs9302752 foi replicada em famílias, sugerindo associação de risco para o desenvolvimento da forma multibacilar da hanseníase (207). Em seguida, o nosso grupo publicou uma replicação independente em um estudo caso-controle e em famílias utilizando diferentes populações brasileiras. Com isso, se confirmou a associação de risco do *NOD2* (rs8057431) e *CCDC122-LACC1* (rs4942254) com a hanseníase, sugerindo fortemente que a via de *NOD2* está associada com susceptibilidade à hanseníase (208). Ainda, em 2018, o *LACC1* (que gera uma proteína participante da via de *NOD2*) foi associado com a hanseníase em chineses (209). O *VDR* é um gene do receptor da vitamina D, que exerce uma variedade de efeitos na resposta imune inata. Apesar

de variações nesse gene terem sido associadas à hanseníase na população indiana, os resultados são ambíguos e estudos mais aprofundados são necessários (199,210). Em seguida, abordaremos a associação e o papel dos genes da resposta imune adaptativa e de algumas citocinas com a hanseníase.

Uma das primeiras regiões estudadas em hanseníase foi o *locus* HLA, na região 6p21 (211). A alternância entre os perfis de resposta tipo Th1 e Th2 devido às diferenças no processamento de antígenos e nas interações célula-célula é o que torna essa região ligada à predisposição a hanseníase (212). Um dos estudos mais consistentes entre a associação do complexo HLA com a hanseníase foi realizado pelo nosso grupo, em que foi observado associação dos alelos de *HLADR*015* e *HLADR*04* com o risco e a resistência à hanseníase, respectivamente, em populações do Vietnã e Rio de Janeiro (213). Mira e colaboradores, evidenciaram uma forte ligação da região *TNF/HLA* com a hanseníase em famílias vietnamitas (214).

O gene *TNF* também está localizado na região 6p21 e está envolvido com a modulação da resposta imune e a atividade pró-inflamatória. O SNP -308G>A, na região promotora do gene, é o mais estudado, porém os resultados em diferentes populações são discrepantes. Na população indiana, por exemplo, o alelo -308A está associado com a suscetibilidade à hanseníase (215). Nosso grupo realizou uma meta-análise utilizando também dados de estudo caso-controle e em famílias na população brasileira, desvendando que o alelo -308A está associado com a proteção à hanseníase em brasileiros, com a confirmação em diversas populações de regiões distintas (186). O gene *LTA*, que codifica a linfotóxina- α , também faz parte da superfamília de *TNF*. O SNP *LTA*+80 A>C foi associado com a susceptibilidade a hanseníase em jovens (<25 anos) Vietnamitas e Indianos. Já em brasileiros, a associação manteve a mesma direção, porém no limite de significância (216). O genes *TNF* e *LTA* atuam na manutenção da resposta granulomatosa ao *M. leprae*, o primeiro tem papel na formação, e o segundo na integridade do granuloma (217).

Em 1999, Moraes e colaboradores demonstraram que a expressão de IFN- γ e IL-12 é maior em pacientes reacionais, mostrando a importância dessas citocinas na infecção (131). O gene *IFNG*, que codifica o IFN- γ , contém o polimorfismo +874 T>A, que é o mais estudado. O alelo T deste SNP aumenta a produção de IFN- γ , promovendo resistência à tuberculose (218). Esse SNP também foi associado a proteção à hanseníase em um estudo caso-controle numa população brasileira, na qual foi observado maiores níveis de IFN- γ em células de indivíduos carreadores do

alelo protetor (219). Outro estudo na população do estado do Amazonas mostrou a associação do SNP +874 ligado a 12 repetições do microsatélite CA no íntron do gene *IFNG* com a Hanseníase (220). Ainda, foi visto que altas frequências desses microsatélites CA estão relacionadas com o desenvolvimento da Hanseníase em Goiás (221). Recentemente, Silva e colaboradores realizaram uma meta-análise recrutando os dados da literatura e adicionando novas populações, confirmando que o gene *IFNG* contribui para o curso da infecção (222).

SNPs do gene *IL10* também estão associados com a suscetibilidade à Hanseníase. A IL-10 tem papel supressor da resposta inflamatória, capaz de modular os níveis de TNF- α e IL-1 β (123). Um SNP muito estudado é o -819 C>T, localizado na região promotora do *IL10*. A associação de suscetibilidade a Hanseníase foi encontrada entre os carreadores do alelo T na população do Rio de Janeiro (223). Em populações do Sul do Brasil, tanto polimorfismos no gene *TNF* (-308G>A) quanto no gene *IL10* mostraram-se associados com a Hanseníase (224). Moraes e colaboradores, em 2004, evidenciaram a associação de proteção do haplótipo do *IL10* -3575A/-2849G/-2763C, enquanto o haplótipo -3575T/-2849A/-2763C foi associado com a suscetibilidade a Hanseníase (225). No ano seguinte, a replicação na população Indiana confirmou a associação de suscetibilidade desse haplótipo (226). Além do *IL10*, outros genes de citocinas também já foram associados com a Hanseníase. Polimorfismos no gene *IL12B*, *IL18* e *IL23R* mostraram associação em pacientes lepromatosos na Índia, China e México (227–229). Um outro gene, o *IL6*, apresenta polimorfismos que levam ao aumento da citocina IL-6 em pacientes com reação do tipo 2 em estudos caso-controle na população brasileira (230). Em 2017, nosso grupo confirmou a suscetibilidade de variantes tanto no *IL6* quanto no *NOD2* com o risco de desenvolver reações na Hanseníase (231).

Por fim, um dos genes fortemente associados são *PARK2* e *PACRG*, o primeiro codifica uma E3 ubiquitina ligase e o segundo é um gene co-regulador do *PARK2*, envolvido na proteólise mediada por ubiquitina. Ambos os genes estão relacionados com o direcionamento das bactérias intracelulares para a via de autofagia. Mira e colaboradores, em 2003, identificaram um pico de ligação na região do cromossomo 6q25-q27 relacionada com a suscetibilidade à Hanseníase a partir de um rastreamento genômico na população vietnamita (214). No ano seguinte, o mesmo grupo mostrou que o pico de ligação estava relacionado aos genes *PARK2* e *PACRG*, que não participavam das vias clássicas associadas a Hanseníase até então, e o estudo de replicação em

brasileiros e vietnamitas confirmaram que os SNPs *PARK2* -2599 e rs1040079 estão associados com a suscetibilidade à hanseníase (232). Um mapeamento de alta densidade e um estudo de desequilíbrio de ligação (LD) na região de *PARK2/PARCG* em vietnamitas e indianos identificou dois SNPs (rs1333955 e rs2023004) que capturam a associação nessas populações (233). Porém, a conclusão acerca da participação desse gene na hanseníase é complicada porque os SNPs estudados não são os mesmos nas diferentes populações e a estrutura genômica de LD em vietnamitas, brasileiros e indianos é complexa. Ainda, foi observado que há uma interação entre SNPs na região promotora (rs9365492 e rs9355403) e em genes de citocinas adjacentes que podem regular a predisposição à hanseníase em indianos (234). Polimorfismos no *PARK2* também estão associados com a febre tifoide e paratifoide, Parkinson e hepatite C (235–237). Isso mostra que algumas doenças podem compartilhar os mesmos genes de suscetibilidade.

O último gene abordado é o *LRRK2*, que codifica uma enzima que tem papel de cinase, atuando na regulação da autofagia, produção de ROS e citocinas inflamatórias. Esse gene foi inicialmente associado com a hanseníase no estudo GWAS em chineses (2009), e posteriormente replicado na população indiana (205,238). Recentemente, Fava e colaboradores (2016) demonstraram que o *LRRK2* está associado com o desenvolvimento da reação tipo 1 numa população vietnamita (239). O mesmo grupo, este ano, mostrou que o controle dos episódios reacionais da hanseníase está ligado à um efeito pleiotrópico entre os genes *LRRK2* e *PARK2*. Essa relação de pleiotropia foi vista também para o modelo de Parkinson, doença a qual o *LRRK2* foi inicialmente associado (240). Por último, neste trabalho, optamos por abordar os genes associados com a hanseníase mais bem descritos, mas outros genes envolvidos nas vias de resposta imune foram associados à hanseníase e seus fenótipos e podem ser encontrados nas revisões sugeridas (187). Ainda, ressalta-se que é possível que genes de outras vias regulatórias ou até de vias desconhecidas também participem no controle da predisposição à hanseníase. Apresentamos uma lista dos genes associados com a hanseníase ao longo dos últimos anos na Figura 1.12.

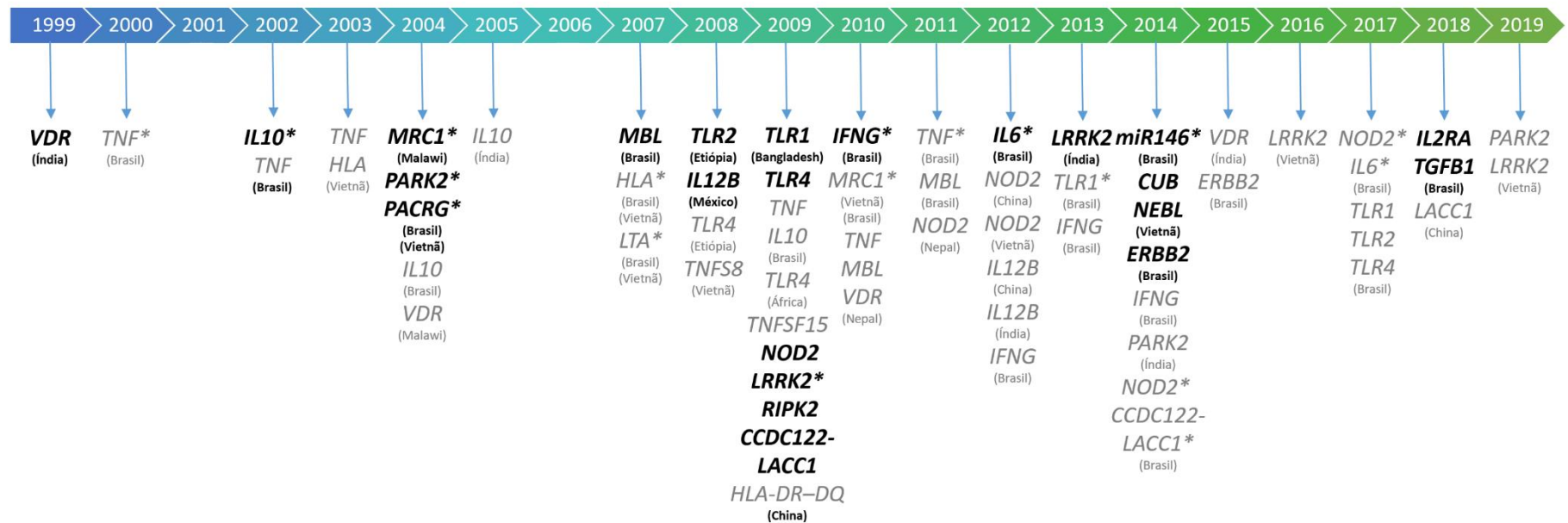


Figura 1.12 – Linha do tempo indicando os genes associados com a hanseníase nos últimos 20 anos. Abaixo do ano da publicação do estudo se encontra o gene associado com a hanseníase e a população estudada. A maioria dos genes aqui descritos foram baseados em estudo de gene candidato validados em estudos do tipo caso-controle. Para o completo apanhado dos genes associados com hanseníase, se faz necessário considerar aqueles desvendados em GWAS, e também os que ainda se encontram pouco caracterizados. Em negrito estão os genes associados pela primeira vez com a hanseníase na literatura e a população em que ele foi caracterizado. Em cinza estão os genes que foram replicados em populações distintas, validando ou não a associação. *Estudos em que o grupo do Laboratório de Hanseníase (LAHAN/Fiocruz) está envolvido.

1.4 – Heterogeneidade genética da população brasileira

Acredita-se que as populações ancestrais nativo-americanas são descendentes de asiáticos que migraram para a América através do estreito de Bering durante o período Pleistoceno (241). O povoamento das Américas ocorreu através de uma única onda migratória por volta de 20.000 anos atrás (242,243). O isolamento desse grupo ancestral durante séculos provocou o surgimento de populações culturalmente homogêneas e com baixa diversidade genética. A partir da chegada dos portugueses no Brasil, em 1500, o panorama de isolamento genético foi quebrado. A colonização refletiu na miscigenação entre europeus, africanos e nativo-americanos ao longo do território, dando origem à uma das populações mais heterogêneas do mundo. Interessante ressaltar que o homem moderno teve origem na África há cerca de 400.000 anos. As populações europeias e asiáticas surgiram do processo migratório “Out of Africa” há aproximadamente 60.000 anos, gerando populações independentes nos diferentes continentes (244). Nesse aspecto, como descrito por Pena e colaboradores (2009), a população brasileira representa o “ponto de encontro” dos três grandes grupos geográficos da humanidade (europeus, africanos e asiáticos, representados pelos nativo-americanos) (245).

A vinda dos europeus ao território brasileiro acarretou um declínio demográfico drástico das tribos nativo-americanas devido aos conflitos com os colonizadores e às doenças as quais os indígenas não eram adaptados. Somado a isso, a partir de 1550, cerca de 4 milhões de africanos, foram trazidos para o Brasil como mão-de-obra escrava para a agricultura e mineração (246). Os escravos provinham de diversas regiões africanas, com predominância de Angola, Congo e Moçambique (Figura 1.13) (247). Além disso, após a abertura dos portos no século 19, uma onda de imigrantes chegou ao território nacional, dentre italianos, espanhóis, alemães, e em menor proporção, sírios, libaneses, japoneses dentre outras nações. Segundo Callegari-Jacques e Salzano (1999), entre 1500 e 1972, 58% dos imigrantes que chegaram ao Brasil eram europeus, 40% africanos e 2% asiáticos (248). Nesse aspecto, incitada pelos “casamentos” entre homens portugueses e mulheres indígenas e escravas africanas, a contribuição sexual europeia foi majoritariamente masculina, enquanto as heranças maternas foram nativo-americanas e africanas, como demonstrado em estudos de marcadores em DNA mitocondrial e cromossomo Y (249,250).

Contudo, a distribuição dos grupos étnicos ao longo do território brasileiro não ocorreu de forma homogênea, e, portanto, as regiões geográficas brasileiras apresentam peculiaridades em relação às contribuições ancestrais africanas, europeias e nativo-americanas. Dados históricos e genéticos demonstram que a contribuição de nativo-americanos é maior na região Norte do país, e menor conforme se desloca para o Centro-Oeste e Sul, ao passo que a herança europeia é marcante no Sul e vai diminuindo conforme se avança para o Norte e Nordeste. O Sul do país teve considerável domínio compartilhado por lusitanos e espanhóis ao longo da colonização. Após a abolição da escravatura, a imigração de italianos e alemães às lavouras de café, preferencialmente no Sudeste e Sul, também fomentaram o contingente de europeus nessas regiões. Já a distribuição de africanos no território brasileiro foi bastante heterogênea, sendo que o Rio de Janeiro (38%) e a Bahia (25%) receberam os maiores contingentes, seguidos de Pernambuco (13%), São Paulo (12%), Maranhão (7%) e Pará (5%) (Figura 1.13) (251). Logo, a contribuição africana foi concentrada no Nordeste e Sudeste do Brasil, principalmente no litoral e nas regiões economicamente mais importantes. A distribuição de africanos nessas regiões é notória, já que é observado o “branqueamento” da população à medida que se distancia do litoral. A interiorização do africano no território se intensificou no século XVIII, em direção à exploração de minérios na Chapada da Diamantina. Com a exaustão das riquezas, os colonos enriquecidos retornaram ao litoral, porém os descendentes dos escravos permaneceram no interior e se misturaram com os povos nativos habitantes da região (251).

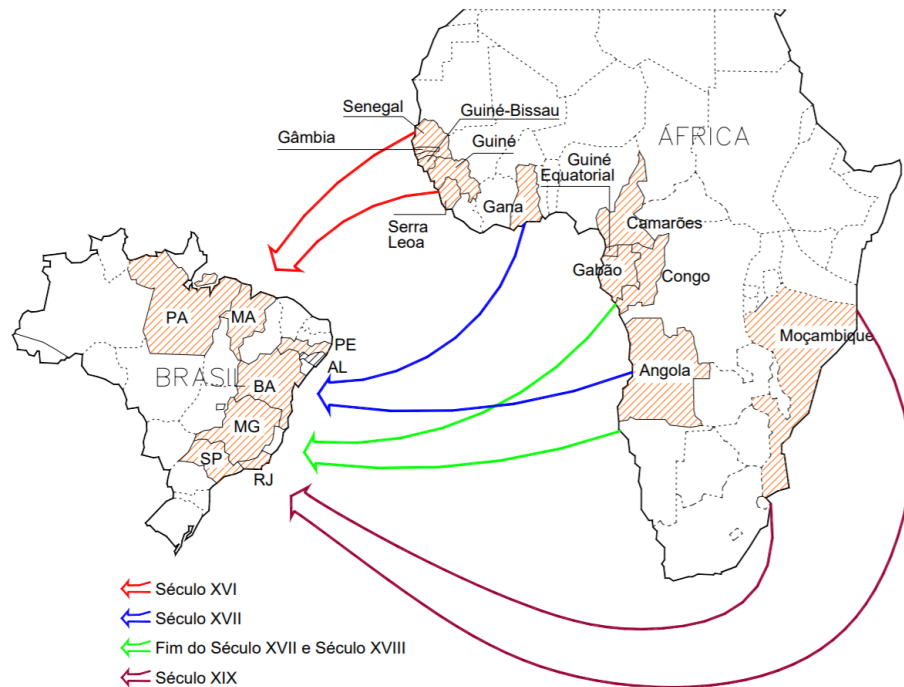


Figura 1.13 – Tráfego de africanos para o território brasileiro entre XVI e XIX. (Adaptado de Sousa, 2001).

Nesse contexto, a identidade genética de três populações ancestrais brasileiras, modeladas há milhares de anos de evolução, vem sendo modificada há cerca de 500 anos de fluxo gênico entre ameríndios, europeus e asiáticos. Essa característica pode ter implicações genéticas, históricas e evolutivas interessantes, como veremos.

1.4.1 – Estudos de associação em populações miscigenadas

A fim de identificar os genes candidatos ou regiões genômicas que contribuem para a suscetibilidade a uma doença complexa são usados os estudos de associação genética. Subentende-se por doença complexa aquela à qual o desenvolvimento está relacionado com a contribuição de diferentes genes, com pequeno efeito, e fatores ambientais. O estudo de associação genética testa a associação entre uma variação genética (o gene é escolhido *a priori*) e a doença através da comparação da frequência da variação em um grupo de indivíduos afetados com a doença de interesse (casos) e um grupo de indivíduos saudáveis (controles). Os estudos em larga escala, como os GWAS (do inglês, *Genome-Wide Association Studies*), permitem estudar

milhares de variantes genéticas ao longo do genoma. Esses estudos não precisam de *hipótese a priori* e permitem a identificação de novas regiões associadas com a doença (252,253).

A escolha do fenótipo deve estar alinhada com a hipótese a qual se quer testar, definindo critérios clínicos estritos para a definição dos casos a fim de obter um grupo homogêneo. Nesse sentido, os critérios de diagnóstico utilizados para a definição dos casos, principalmente quando a doença apresenta diferentes manifestações (fenótipos moderado e grave, por exemplo), deve ser bem padronizado e aplicado para todos os pacientes. Já a seleção dos controles pode ocorrer por dois métodos: I) seleção de indivíduos avaliados segundo a presença negativa da doença ou II) seleção de indivíduos randomicamente na população e que, nesse caso, o *status* da doença é desconhecido. Ressalta-se que para as doenças comuns (mais prevalentes), controles sabidamente não afetados pela doença permitem maior poder de detectar associação. Entretanto, como o *screening* da população para a presença ou ausência da doença apresenta dificuldades, o uso de controles sem o conhecimento histórico da doença se torna mais eficiente (254). Considerando a igualdade de exposição ao fator ambiental de risco, sugere-se que os controles ideais seriam os indivíduos que habitam a vizinhança dos casos, garantindo o mesmo tipo de exposição. Porém, como a seleção de um número representativo de indivíduos nessas condições é impraticável, muitos estudos fazem uso de controles provenientes de banco de doadores de sangue, assumindo que eles são representativos da população de casos (255). Deve-se considerar, ainda, que casos e controles não tenham grau de parentesco.

Os SNPs são os marcadores genéticos comumente testados em estudos de associação, mas os microssatélites, indels, VNTRs (do inglês, *variable-number tandem repeats*) e CNVs (do inglês, *copy-number variations*) também são utilizados. O controle de qualidade dos ensaios de genotipagem são importantes em estudos em larga escala como os GWAS, porém também se adequam aos estudos de gene candidatos. Muitos genótipos faltantes (*missings*) no grupo amostral pode enviesar os resultados, portanto o *call rate* (taxa de genotipagem) deve ser considerado. Além disso, deve-se avaliar se a frequência das variantes testadas está de acordo com o equilíbrio de Hardy-Weinberg. A não concordância com esse princípio pode indicar erros de genotipagem, acasalamento preferencial ou estratificação populacional (256).

Um dos grandes problemas na inferência causal de estudos epidemiológicos caso-controle são os fatores confundidores provenientes do viés de seleção. Nesse caso, a exposição não é

aleatoriamente distribuída entre os grupos de comparação, como em um ensaio clínico randomizado. A presença de estruturação, que é um tipo específico de confundimento, é provocada pela presença de subgrupos distintos dentro de um grupo populacional aparentemente homogêneo, ou seja, a estruturação ocorre quando as proporções de ancestralidade variam entre os grupos (257,258). Por exemplo, a presença de indivíduos com maior ancestralidade africana no grupo de casos em relação aos controles pode levar à falsa associação de um alelo com a doença simplesmente porque o alelo é mais frequente em africanos (Figura 1.14). Nesse caso, a super-representação do alelo nos casos se deve à predominância de mais indivíduos com elevada ancestralidade africana nesse grupo. A estruturação é decorrente da mistura étnica e é um complicador nos estudos de associação em populações miscigenadas, como a brasileira, pois pode mimetizar um sinal de associação falso-positivo ou perder efeitos de associações reais (259).

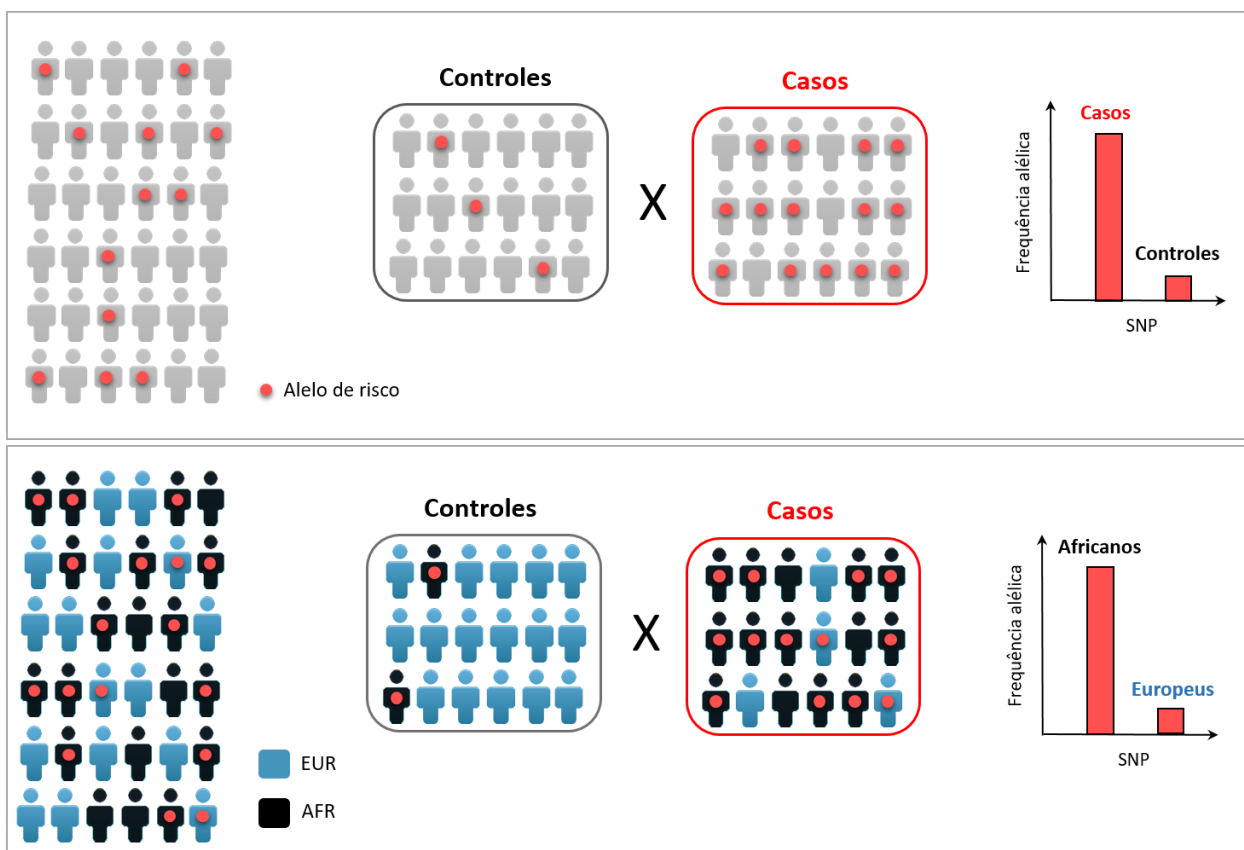


Figura 1.14 – Estruturação populacional e o viés de associação genética entre casos e controles. No quadro superior, o alelo de risco encontra-se aumentado em casos comparado aos controles, configurando associação com o desfecho. No segundo quadro, observa-se que há um desequilíbrio entre

os indivíduos que compõe os grupos. Logo, a associação ocorre porque o alelo de “risco” é mais frequente em indivíduos que carregam uma ancestralidade africana (em preto) do que os que apresentam maior ancestralidade europeia (em azul). Nesse exemplo, observa-se uma associação genética espúria.

Algumas estratégias para corrigir o efeito da estruturação são: I) Inclusão dos genótipos paternos dos indivíduos afetados (caso índice) como controle interno, e nesse caso, avalia-se o desequilíbrio na transmissão do alelo em um desenho baseado em famílias – TDT (do inglês, *transmission disequilibrium test*) – ou II) correção dos efeitos de estruturação populacional através do controle genômico de ancestralidade da população. O TDT não apresenta falso-positivo devido à estruturação populacional porque utiliza núcleos familiares. Nesse estudo, avalia-se o desequilíbrio na transmissão do alelo aos filhos afetados em relação ao esperado ao acaso (50%), utilizando o alelo não transmitido como controle (260). A dificuldade ocorre no recrutamento de famílias informativas, que devem ser organizadas em trios com pelo menos um filho afetado e pais heterozigotos, além de são ser prático para doenças de início tardio (261,262). A segunda abordagem é elaborada através da genotipagem de marcadores informativos de ancestralidade (AIMs), polimorfismos autossômicos que apresentam frequência alélicas extremas entre os diferentes grupos ancestrais que caracterizam a população de estudo.

Os AIMs tem sido estudados não somente no contexto dos estudos de associação, mas também de questões epidemiológicas, na genética forense, farmacogenética e genética de populações (263,264). Um maior número de marcadores para a estimativa da ancestralidade aumenta a confiança no resultado. Porém, atualmente existem diferentes painéis de AIMs bem caracterizados que utilizam números e tipos variados de marcadores e são eficientes na predição do componente ancestral a nível populacional e individual (265–267). Quanto maior a distância entre as populações parentais, menor o número de AIMs necessários para diferenciar duas populações teste (como os casos e controles, ou duas populações das diferentes Regiões Brasileiras) (268). Como a população brasileira é formada por grupos continentais, um número reduzido de marcadores é eficaz para caracterizar a população. Nos estudos GWAS, a estruturação populacional pode ser identificada através da Análise dos Componentes Principais (PCA), que modela as diferenças ancestrais genéticas entre casos e controles, e os Componentes Principais (PC) podem ser usados para a correção da análise (269).

Até hoje, poucos estudos abordaram possíveis “associações espúrias” devido à estratificação populacional. Isso porque até os anos 2000, não haviam formas de obter a ancestralidade genética e corrigir para o efeito da estratificação (270,271). Antigamente, com o objetivo de minimizar esses efeitos em populações miscigenadas, uma das maneiras amplamente utilizadas foi a classificação por etnia autodeclarada ou definida por um profissional da saúde. No Brasil, a classificação é baseada nos critérios do IBGE (brancos, pardos, negros, amarelos e indígenas) no momento da coleta dos voluntários do estudo, levando em consideração o histórico familiar do indivíduo (272). A princípio, essa estratégia é sensível no *matching* de casos e controles. Entretanto, é possível que haja diferentes proporções de ancestralidade dentro de um mesmo grupo étnico (258). Apesar dos estudos mais recentes mostrarem que há concordância entre a etnia e a ancestralidade genética na população brasileira, outros demonstram que há dissociação entre a cor de pele autodeclarada e ancestralidade (264,273–276). Além disso, a classificação por cor da pele está atrelada à condição socioeconômica, o que torna a definição muito subjetiva, e, portanto, não direta (275). Nesse aspecto, muitos pesquisadores da área são céticos quanto à utilização da etnia no ajuste das análises de associação.

No geral, para estimar a razão de chances (*Odds Ratio*) de desenvolver a doença, as análises de associação devem considerar o maior número de variáveis de confundimento possíveis que possam diferenciar os grupos de casos e controles. As variáveis confundidoras têm potencial em prever o desfecho e, por isso, confundem a identificação do real fator que leva ao desfecho. A utilização dessas variáveis no controle dos estudos de associação depende do tipo de estudo e do desfecho que se quer avaliar (277). No geral, para estudos de populações considera-se o sexo, idade, etnia ou ancestralidade genômica, classificação social, nível de escolaridade, etc. Por fim, a interpretação final do estudo de associação também deve ser cautelosa. Muitas vezes a variante causal com efeito na suscetibilidade não é a primeira a ser detectada pelo estudo devido ao desequilíbrio de ligação (LD), que é a associação não aleatória entre alelos de diferentes loci. Em LD, os marcadores moleculares vizinhos tendem a ser herdados juntos, e a informação de um deles prediz a do outro. Portanto, os estudos de associação podem desvendar o marcador não causal, mas que estão em LD com ele. Nesse aspecto, a avaliação final dos estudos de associação pode seguir três linhas: (1) associação direta, em que o SNP analisado é a variante causal que provoca a suscetibilidade; (2) associação indireta, em que a variante causal está em

LD com o SNP identificado na associação ou (3) resultado falso-positivo, provocado pela presença de potenciais variáveis confundidoras não mensuradas, como estruturação populacional, caso elas não sejam consideradas nas análises (254).

1. JUSTIFICATIVA

No decurso deste estudo, apresentamos como a seleção por patógenos exerce uma forte pressão seletiva sobre o genoma humano e molda a variabilidade genética entre diferentes populações. Mutações nos genes *PKLR* e *G6PD* que levam à resistência frente a infecção pelo *Plasmodium* são mantidas em populações africanas endêmicas para a malária. Entretanto, o dano funcional provocado por essas mutações, que tange à ocorrência de anemias hemolíticas e a consequente perda de ferro durante a hemólise, pode aumentar a suscetibilidade a patógenos intracelulares, como o *Mycobacterium leprae* e o *M. tuberculosis*. Sabendo-se que a população brasileira apresenta uma elevada contribuição africana, é possível que assinaturas de seleção da malária estejam presentes no genoma da população brasileira e alterem o perfil de suscetibilidade às doenças endêmicas no país, como a hanseníase. Portanto, exploramos a hipótese de *trade-off*, ou mecanismo evolutivo compensatório, em que polimorfismos de resistência à malária na África podem levar a suscetibilidade a patógenos intracelulares na população brasileira. Para desenvolver a nossa hipótese, utilizamos o modelo de hanseníase.

A hanseníase tem sido utilizada como modelo clássico de estudo de suscetibilidade genética em doenças infecciosas. A baixa patogenicidade do *M. leprae*, a diversidade de manifestações clínicas da doença e a elevada resistência humana (~90%) frente ao bacilo conduzem, em âmbito final, ao protagonismo dos genes do hospedeiro para o sucesso da infecção. No modelo de hanseníase, os genes de resposta imunológica são os mais bem caracterizados com a predisposição à doença, porém o papel de genes do metabolismo energético ainda não é definido. Uma das razões para esse estudo é que, na literatura, a associação de suscetibilidade dos polimorfismos do gene *PKLR* com patógenos intracelulares ainda não foi demonstrada em populações humanas. Em uma análise prévia a este trabalho (2015), foi identificada a associação genética do gene *PKLR* com a hanseníase em uma amostra do Rio de Janeiro. A fim de validar a relevância desse gene com a hanseníase, se fez necessário

replicar os achados iniciais em outras amostras da população brasileira. Além disso, acompanhando os estudos desenvolvidos em populações miscigenadas nos últimos 10 anos, o controle da associação por ancestralidade genômica se tornou imprescindível, permitindo conclusões mais robustas acerca da influência do gene no desenvolvimento da hanseníase, e de como a ancestralidade influencia na suscetibilidade a doença. Outra motivação a este estudo é que pouco foi investigado acerca das “pegadas de seleção” no gene *PKLR*, restringindo-se a avaliações nas frequências alélicas, haplotípicas e no padrão de LD da região do gene entre poucas populações africanas e portuguesas.

Atrelado a isso, objetivando compreender a interação entre o bacilo da hanseníase e sua célula hospedeira, estudos evidenciaram que a infecção pelo *Mycobacterium leprae* promove o aumento da captação de glicose e da atividade da enzima glicose-6-fosfato desidrogenase (G6PD). Paralelamente, um estudo de microarranjo identificou que as vias do metabolismo glicolítico e oxidativo são enriquecidas após a infecção. Portanto, genes relacionados à essas vias são potenciais candidatos ao estudo de associação genética com a hanseníase. O *G6PD*, assim como o *PKLR*, regula funções centrais das células vermelhas, e, por isso, está sob seleção por patógenos na África. Esse gene apresentou-se como forte candidato à associação em um estudo simultâneo do nosso grupo que demonstrou a relevância da via das pentoses durante a infecção pelo ml. Portanto, avaliamos a possível influência dos genes *PKLR* e *G6PD* com a predisposição a hanseníase, extrapolando a hipótese do papel desses genes para a suscetibilidade patógenos intracelulares. Finalmente, exploramos a vertente funcional na tentativa de explicar como as variantes genéticas modificam as interações celulares com o patógeno.

Ressalta-se que populações diferentes nem sempre compartilham os mesmos marcadores de suscetibilidade a doenças. Como o perfil genético da população brasileira é complexo e ainda pouco explorado no contexto da associação com as doenças infecciosas, é possível que novos marcadores de suscetibilidade sejam identificados nessa população. No contexto da hanseníase, a formulação de um painel de marcadores genéticos de suscetibilidade tem implicação clínica na predição de risco ao desenvolvimento da doença, principalmente nos indivíduos com alto risco de adoecimento, como os contatos domiciliares.

3 – OBJETIVOS

3.1 – Objetivo geral

Investigar a associação dos genes alvo de seleção natural, *PKLR* e *G6PD*, com as doenças micobacterianas, utilizando a hanseníase como um modelo de estudo, e avaliar a influência da ancestralidade na suscetibilidade à infecção.

3.2 – Objetivos específicos

- 1) Determinar a ancestralidade genômica de casos e controles entre os diferentes grupos populacionais utilizados no estudo de associação com a hanseníase a partir de 46 marcadores informativos de ancestralidade do tipo Indels;
- 2) Utilizar a ancestralidade genômica para ajuste da análise de associação do gene *PKLR* com a hanseníase na população do Rio de Janeiro e replicar o estudo de associação em diferentes populações brasileiras corrigindo para o efeito de estruturação populacional;
- 3) Investigar a presença de assinaturas de seleção natural na região do gene *PKLR* a partir de análises *in silico* em dados genômicos de europeus e africanos do 1000Genomes;
- 4) Realizar o estudo de associação do gene *G6PD* com a hanseníase em homens e mulheres da população do Rio de Janeiro, corrigindo para ancestralidade genômica;
- 5) Avaliar o papel funcional das variantes dos genes *PKLR* e *G6PD* com o desenvolvimento da hanseníase.

4 – METODOLOGIA

4.1 – Desenho de estudo

Para avaliar a influência dos genes *PKLR* e *G6PD* com as doenças micobacterianas, realizamos um mapeamento de variantes genéticas candidatas para o estudo retrospectivo de associação do tipo caso-controle com a hanseníase ou tuberculose (Figura 4.1). O primeiro passo foi caracterizar o perfil de ancestralidade das amostras utilizadas (N=4.498). Posteriormente, os estudos de associação foram realizados nas populações independentes seguido da replicação na população brasileira (N=5.293 para o estudo do *PKLR* e N=1.373 para o estudo do *G6PD*). A abordagem evolutiva foi explorada para o gene *PKLR*, e as análises funcionais foram realizadas para ambos os genes candidatos.

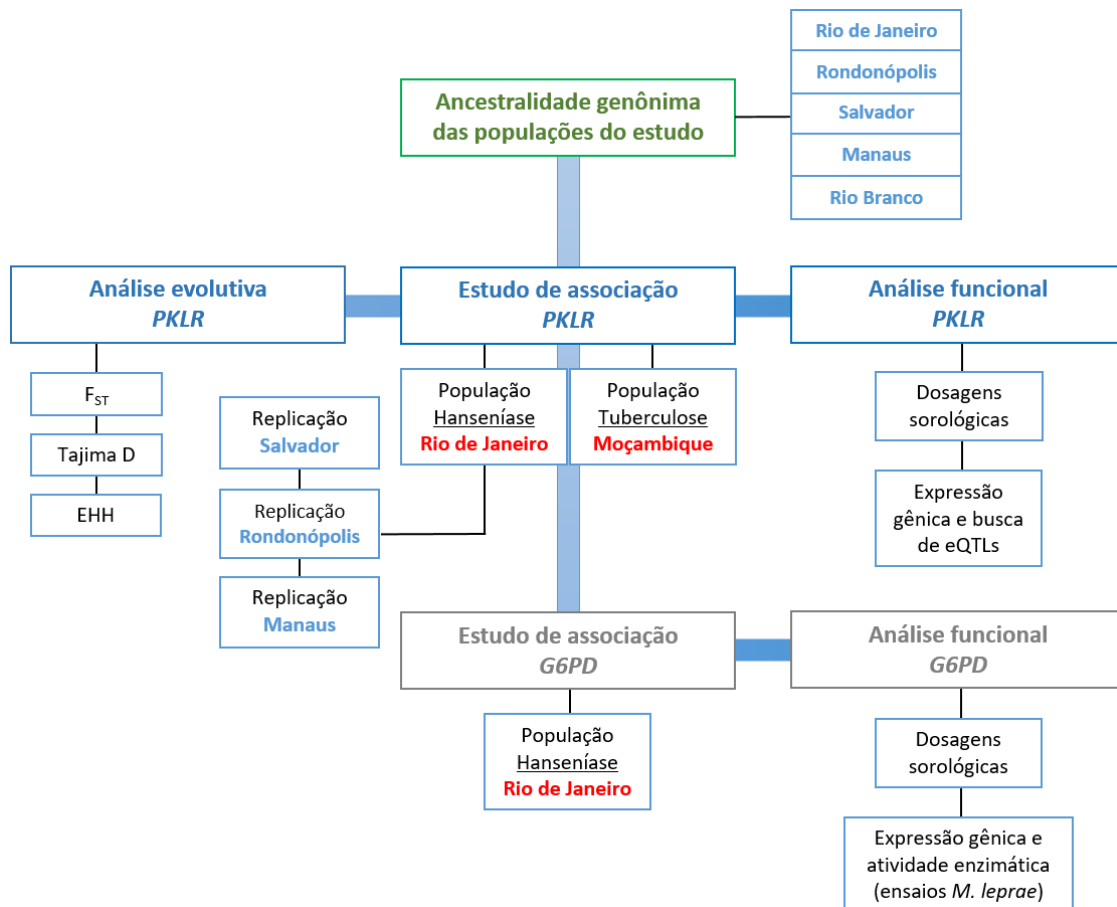


Figura 4.1 – Estratégia de estudo apresentada nesse trabalho.

4.2 – Análise de ancestralidade e estruturação populacional

4.2.1 – Populações do estudo de associação e ancestralidade genômica

O presente trabalho incluiu o estudo de 5.293 indivíduos de cinco populações de diferentes sub-regiões do Brasil e uma população africana.

O primeiro estudo de associação com a hanseníase ocorreu na população do Rio de Janeiro (RJ-região Sudeste; POP teste 1), principal do estudo. A replicação foi realizada em grupos amostrais representativos de outras regiões brasileiras em que a hanseníase é endêmica: Salvador (BA-região Nordeste; POP replicação 2), Rondonópolis (MT-região Centro-oeste; POP replicação 3); Manaus (AM-região Norte; POP replicação 4) e Rio Branco (AC-região Norte; POP replicação 5). A taxa de incidência da hanseníase em cada população está descrita na Tabela 4.1. Para todas as populações de casos com hanseníase, os pacientes foram classificados segundo os critérios de Ridley e Jopling e de acordo com a classificação operacional da OMS (PB – paucibacilares e MB – multibacilares). Para confirmar a hipótese de associação genética dos genes aqui avaliados com a suscetibilidade a doenças micobacterianas, foi utilizado um estudo caso-controle independente na população de Moçambique (Africanos; POP teste 6), em que o desfecho é a tuberculose. A descrição detalhada dessas populações pode ser acessada na Tabela 4.2.

- **População teste POP 1 (N = 1373): Rio de Janeiro (RJ)**

A população do Rio de Janeiro representa a população teste principal do estudo de associação e é composta por 597 indivíduos saudáveis doadores de medula óssea do Instituto Nacional de Câncer (INCA) e 776 pacientes diagnosticados com hanseníase pelo Ambulatório Souza Araújo da FIOCRUZ/RJ, centro de referência para diagnóstico clínico e tratamento de casos com a doença. Os pacientes são diagnosticados com base em exame clínico, baciloscópio e histopatológico, sendo que em alguns casos é necessário o diagnóstico molecular por PCR em tempo real. O grupo de pacientes contém aproximadamente 40% de casos diagnosticados com hanseníase no período de 1986 a 1999 e 60% de casos diagnosticados a partir de 2000 até 2016,

totalizando um tempo de 30 anos de recrutamento. Todos os casos utilizados no estudo são indivíduos não aparentados, sem histórico de HIV e que tiveram um período mínimo de 1 ano de acompanhamento no Ambulatório Souza Araújo. O grupo dos indivíduos saudáveis compreendem indivíduos sem relação histórica com a hanseníase e representantes da região demográfica do Rio de Janeiro.

A população de casos e controles do Rio de Janeiro tem sido utilizada para testar a associação de outros genes com a hanseníase em publicações recentes do laboratório (Cardoso et al., 2011; César-de-Mello, 2014 e de Sales-Marques et al., 2014). O DNA das amostras foi extraído a partir de sangue total pelo método de *salting out* no Laboratório de Hanseníase.

- **População de replicação POP 2 (N = 529): Salvador (BA)**

Os pacientes foram recrutados no período de 2011 a 2015 em dois centros localizados em Salvador: o Complexo Hospitalar Professor Edgar Santos da Universidade Federal da Bahia – referência para diagnóstico e tratamento da hanseníase no estado, e o Hospital Couto Maia, referência para o diagnóstico e tratamento de doenças infecciosas na Bahia. O número de pacientes foi 219, de ambos os sexos, com a idade entre 18 e 65 anos. O diagnóstico foi realizado clinicamente mediante avaliações dermatoneurológicas, teste de sensibilidade (estesiometria), baciloscopia da linfa e a biópsia de uma ou mais lesões para a definição da forma clínica da doença. Foram excluídos os pacientes com quadro infeccioso durante 30 dias prévios ao relato da entrevista e realização dos exames e os portadores de doenças autoimunes ou genéticas. Características adicionais estão descritas em Rego e colaboradores (2015) e Santana e colaboradores (2017).

Os controles consistiram em 310 doadores de sangue da Fundação de Hematologia e Hemoterapia da Bahia (HEMOBA), na cidade de Salvador (BA), recrutados em 2014 e 2015. Nenhum dos indivíduos declarou contato prévio ou grau de parentesco com pacientes de hanseníase em entrevista clínico-epidemiológica.

As amostras de DNA foram extraídas pelo método *salting out* e cedidas pela Universidade Federal da Bahia (UFBA/Salvador/BA).

- **População de replicação POP 3 (N=768): Rondonópolis (MT)**

As amostras de Rondonópolis (MT), uma região hiperendêmica para a hanseníase, como mostra a Tabela 4.1, foram arroladas de um estudo epidemiológico multidisciplinar conduzido no município. A população contém 411 indivíduos diagnosticados com hanseníase pelo atendimento ambulatorial do programa de saúde da família (PSFs) e 357 indivíduos saudáveis habitantes de Rondonópolis. Estes indivíduos foram selecionados em sistemas de mutirão para a detecção da hanseníase estabelecidos em locais públicos como faculdades, corporação de bombeiros, presídio, exército e em postos criados para participação voluntária. Um questionário foi previamente aplicado e a negativa para a história pessoal da doença, bem como para o exame clínico realizado nesta ocasião, foram critérios de inclusão no grupo dos controles. Estes foram selecionados ainda de forma a respeitar as características do grupo caso quanto às variáveis, etnia, sexo e idade. Dados da população podem ser acessados em Camargo e colaboradores, (2018).

As amostras de DNA desse grupo amostral foram extraídas pelo método de *salting out* e cedidas pelo Instituto Laura de Souza Lima (ILSL/Bauru/SP).

- **População de replicação POP 4 (N=1389): Manaus (AM)**

Os indivíduos da população de Manaus foram coletados na Fundação Alfredo da Matta (FUAM /Manaus) por demanda espontânea no período de março de 2014 a dezembro de 2017. Indivíduos acima de 10 anos foram incluídos no estudo. O grupo de controles é formado por 967 pessoas que buscavam atendimento dermatológico para fins empregatícios ou atividades recreativas e, após avaliações dermatoneurológicas, não apresentavam suspeita de hanseníase e declararam não ter histórico e não ser contatos de pacientes com hanseníase ou tuberculose. Os controles vivem na mesma área endêmica dos casos. O grupo de pacientes foi composto por 423 indivíduos diagnosticados pela clínica, baciloscopia e achados histológicos, de acordo com os sintomas apresentados. Informações sobre a população podem ser acessadas em Leituriundo e colaboradores, 2019.

As amostras de DNA dessa população foram extraídas por meio do kit comercial *Blood & Tissue* (Qiagen) e cedidas pela Fundação Alfredo da Mata (FUAM/Manaus/AM).

- **População de replicação POP 5 (N=768): Rio Branco (AC)**

As amostras de 157 pacientes de Rio Branco foram recrutadas para um estudo de levantamento de dados epidemiológicos e clínicos dos pacientes de hanseníase do município. A coleta foi realizada pelo Departamento de Dermatologia do Hospital das Clínicas (HC) do Acre, referência de hanseníase no estado, e que recebe encaminhamento de pacientes de Rio Branco e de outros municípios do estado. Foram incluídos pacientes maiores de 15 anos no estudo. Como controles, 611 doadores saudáveis do banco de sangue do Hemoacre e residentes do município de Rio Branco foram recrutados e as características de idade, gênero e histórico familiar com a hanseníase foram registradas.

As amostras de sangue foram cedidas pelo Departamento de Dermatologia do Hospital das Clínicas do Acre e o DNA foi extraído pelo método de *salting out* no Laboratório de Hanseníase da FIOCRUZ.

- **População teste POP 6 (N=466): Moçambique**

Amostras de sangue de 104 pacientes com tuberculose foram coletados de três Unidades Sanitárias da cidade e província de Maputo (Moçambique), Centro de Saúde Mavalane, Centro de Saúde de Polana Caniço e Hospital Geral da Machava. Os casos foram definidos como indivíduos com tuberculose pulmonar ativa e, como critério de elegibilidade, foram utilizados somente os indivíduos HIV negativo. Os critérios para diagnóstico basearam-se nos sintomas clínicos como dor no peito, perda de peso, fadiga, febre, mal-estar, sudorese noturna e tosse com expectoração, e laboratoriais, incluindo baciloscopia positiva e exame sugestivo em raio-X. Os controles são formados por 362 indivíduos doadores de sangue do Hospital Central de Maputo. Tais indivíduos eram HIV negativos e não apresentaram histórico prévio com tuberculose até o momento do estudo. Essa população também foi alvo de estudo em Mabunda e colaboradores, 2015.

As amostras de DNA de Moçambique foram extraídas a partir de sangue total pelo kit *QIAamp blood DNA extraction* (QIAGEN) e cedidas pelo Laboratório de Virologia Molecular do Instituto Nacional de Saúde de Maputo em Moçambique.

Tabela 4.1 – Taxa média de detecção de novos casos de hanseníase (por 100 mil habitantes) por região brasileira estudada, registrada pelo Ministério da Saúde entre 2012 e 2016.

| Região/Estado | Total | Masculino | Feminino |
|----------------------|--------------|------------------|-----------------|
| Norte | 34,26 | 40,52 | 27,81 |
| Acre | 18,23 | 23,97 | 12,37 |
| Amazonas | 14,83 | 17,83 | 11,77 |
| Nordeste | 23,42 | 25,53 | 21,39 |
| Bahia | 16,18 | 16,86 | 15,51 |
| Centro-Oeste | 37,27 | 41,29 | 33,28 |
| Mato Grosso | 88,9 | 93,2 | 84,38 |
| Sudeste | 5,31 | 6,05 | 4,59 |
| Rio de Janeiro | 7,18 | 7,98 | 6,44 |

Fonte: Sistema de Informação de Agravos de Notificação (Sinan). Boletim epidemiológico 2018.

Tabela 4.2 – Características gerais das populações de estudo.

| Característica | POP 1 Rio de Janeiro | | POP 2 Salvador | | POP 3 Rondonópolis | | POP 4 Manaus | | POP 5 Rio Branco | | POP 6 Moçambique | |
|------------------------------------|-------------------------|---------------|-------------------|---------------|-----------------------|---------------|-----------------|---------------|---------------------|---------------|---------------------|--------------|
| | CO (597) | CA (776) | CO (310) | CA (219) | CO (357) | CA (411) | CO (967) | CA (423) | CO (611) | CA (157) | CO (362) | CA (104) |
| Idade (média, DP) | 33±9 | 41±16 | 35±8 | 43±11 | 42±14 | 42±16 | 30±8 | 34±15 | 31±9 | 49±17 | 32±10 | 34±11 |
| Sexo (n, frequência) | | | | | | | | | | | | |
| M | 322 (0,53) | 504 (0,65) | 220 (0,71) | 131 (0,59) | 217 (0,61) | 250 (0,61) | 526 (0,54) | 288 (0,68) | 389 (0,63) | 98 (0,62) | 300 (0,83) | 59 (0,57) |
| F | 263 (0,44) | 270 (0,35) | 90 (0,29) | 88 (0,40) | 140 (0,39) | 161 (0,39) | 441 (0,45) | 135 (0,32) | 222 (0,33) | 59 (0,37) | 62 (0,17) | 45 (0,43) |
| Localização (n, frequência) | | | | | | | | | | | | |
| Sul | - | - | - | - | - | - | - | - | - | - | 295 (0,83) | 89 (0,91) |
| Centro | - | - | - | - | - | - | - | - | - | - | 30 (0,09) | 9 (0,09) |
| Norte | - | - | - | - | - | - | - | - | - | - | 29 (0,08) | 0 (0,00) |
| Pacientes (n, frequência) | | | | | | | | | | | | |
| MB | - | 499 (0,64) | - | 157 (0,71) | - | 310 (0,75) | - | 319 (0,75) | - | 120 (0,76) | - | - |
| PB | - | 274 (0,35) | - | 45 (0,21) | - | 96 (0,23) | - | 104 (0,25) | - | 37 (0,24) | - | - |
| Outros (NP, I) | - | 3 (0,10) | - | 17 (0,08) | - | 5 (0,02) | - | 0 (0,00) | - | 0 (0,00) | - | - |

Os dados são apresentados em frequências e podem não somar 1 devido aos dados faltantes (na informação de sexo). Legenda: CO=Controles; CA=Casos; DP=Desvio padrão; M=Masculino; F=Feminino; MB=Multibacilar; PB=Paucibacilar; NP=Neural pura e I=Indeterminada.

4.2.2 – Considerações éticas

Os procedimentos do presente trabalho na população do Rio de Janeiro foram aprovados pelo Comitê de Ética em Pesquisa da Fundação Oswaldo Cruz/RJ - parecer nº 16836 emitido em 27/03/2012, sob responsabilidade de Dra. Euzenir Nunes Sarno. O estudo nas populações subsequentes foram aprovados pelo comitê de ética da fundação de origem, como segue: Salvador – comitê de ética responsável 5577/UFBA/Faculdade de Medicina da Bahia (FMB), com data de Aprovação Ética do CEP/CONEP em 30/11/2014 sob registro nº 16206 (processo nº 25000.182612/2010-12); Rondonópolis – comitê de ética em pesquisa do Instituto Lauro de Souza Lima (CEP/ILSL, nº 172/09); Manaus – comitê de ética da Fundação Alfredo da Matta (CEP/FUAM, nº 555.62); Rio Branco – comitê de ética em pesquisa (CEP/UFAC, nº 750.553; CEP/Fiocruz, nº 775.694; CEP/HC/Acre, nº 910.309); e Moçambique – comitê de ética de Moçambique sobre o protocolo nº 399/CNBS/11. Todos os doadores de amostras biológicas participantes dos projetos assinaram o Termo de Consentimento Livre e Esclarecido (TCLE).

4.2.3 – Extração de DNA

As amostras de DNA foram extraídas pelo método *salting out*, com modificações (98). Inicialmente, 500 µL de sangue total foram adicionados à 1 mL de tampão de lise de hemácias (Tris-HCl 10 mM e EDTA 1mM) em *ependorf* de 1,5 mL e centrifugados a 2.800 rpm por 20 minutos. O sobrenadante foi descartado e, novamente, os tubos receberam 1 mL de tampão de lise de hemácias. Os tubos foram agitados no vórtex para dissolução do pellet e centrifugados por 20 minutos e a 2.500 rpm. Essa etapa correspondeu à lavagem das células para a retirada de hemácias e foi repetida de três a quatro vezes, até que se observasse um *pellet* claro. Em seguida, foram adicionados 200 µL do tampão de lise celular (Tris-HCl 10 mM, NaCl 400 mM, e EDTA 2 mM), 20 µL de SDS (10%) e 5 µL de proteinase K (20 mg/mL) para a degradação das membranas celulares e proteínas. Os tubos foram incubados em estufa a 56°C por no mínimo 8h (*overnight*). Após isso, adicionou-se 200 µL da solução de acetato de sódio saturado (6,83 M) e agitou-se os tubos no vórtex. Seguiu-se à centrifugação por 15 minutos a 3.000 rpm. Após isso, duas fases foram observadas: o sobrenadante, que contém o DNA, e o precipitado, que reserva os restos

orgânicos. O sobrenadante foi cuidadosamente coletado e armazenado em um *eppendorf* de 1,5 mL. Em seguida, adicionou-se etanol absoluto em volume 2 vezes o de sobrenadante recuperado, homogeneizando por inversão e observando o aglomerado de DNA formado. Após isso, os tubos foram centrifugados durante 15 minutos a 20.000 rpm e o sobrenadante descartado. Por fim, adicionou-se 200-400 μ L de etanol hidratado (70 %). Os tubos foram novamente centrifugados por 15 minutos a 20.000 rpm, o sobrenadante foi descartado e os tubos permaneceram na estufa até a secagem do *pellet* (30 minutos). A ressuspensão foi feita para 40-80 μ L de solução TE (Tris-HCl 5mM e EDTA 0,1mM) e os tubos permaneceram a 56°C por 30 minutos na estufa para a dissolução completa do *pellet* de DNA. A quantificação e avaliação da pureza do DNA extraído foi realizada por espectrofotometria utilizando o instrumento NanoDrop® ND-1000 (Thermo Fisher).

4.2.4 – Genotipagem dos marcadores informativos de ancestralidade (AIMs)

A fim de obter as estimativas de ancestralidade a nível individual e populacional em diferentes *subsets* da população brasileira, utilizamos o sistema multiplex composto por 46 marcadores informativos de ancestralidade do tipo Indels (46-AIMS-Indels) desenvolvido por Pereira e colaboradores (267). Os marcadores apresentam frequências alélicas distintas entre europeus, africanos e nativo-americanos e são capazes de discriminar eficientemente as proporções de ancestralidade de uma população heterogênea baseado nessas populações de referência, como demonstrado anteriormente (279,280). O painel de 46-AIM-Indels apresenta fragmentos curtos (menores de 250pb) e podem ser facilmente amplificados em uma única reação multiplex. A Tabela 4.3 descreve a relação dos marcadores utilizados, a localização genômica, o cromossomo relacionado e a caracterização dos alelos. Os marcadores foram agrupados em 4 painéis com cores representativas dos fluoróforos marcados em cada grupo de *primers*: 6-FAM (azul), VIC (verde), NED (cinza) e PET (vermelho). Em ANEXO I é possível encontrar a lista dos 46 pares de *primers* utilizados, bem com as características consideradas na definição da eficiência do painel: a frequência alélica nas três populações referência do painel HGDP-CEPH, o índice diferencial de frequência alélica (considerado como $\delta > 0.40$ entre duas populações) e o índice de diferenciação genética entre as populações (F_{ST}).

Para a amplificação das amostras no sistema de 46-AIM-Indels, utiliza-se o reagente QIAGEN Multiplex PCR kit, 0.1 mM de todos os *primers* (os 46 pares de *primers* são diluídos e armazenados juntos) e 0.5-10 ng/ μ L de DNA genômico em um volume final de 10 μ L de reação. As condições de reação foram: etapa inicial de 95°C por 15 minutos; 30 ciclos de 94°C por 30 segundos, 60°C por 90 segundos e 72°C por 45 segundos; e, por fim, 72°C por 60 minutos de extensão. As reações foram amplificadas no termociclador GeneAmp PCR System 9700 (Thermo Fisher). Os produtos de PCR foram preparados previamente para a eletroforese capilar pelo sistema de análise molecular ABI 3130 e ABI 3500 Genetic Analyzer (Thermo Fisher). Para isso, adicionamos 1 μ L do produto de PCR a 8.8 μ L de Hi-Di Formamide (Thermo Fisher) e 0.2 μ L de GeneScan 500 LIZ (Thermo Fisher) como padrão de peso molecular, totalizando um volume de 10 μ L. Os eletroferogramas foram analisados e os alelos genotipados no Genemapper 4.1 (Thermo Fisher). Para facilitar a genotipagem, os alelos curtos dos Indels (sequência de menor tamanho para o marcador) foram chamados de alelo 1 e os alelos longos (sequência de maior tamanho para o marcador) foram chamados de alelo 2 (Figura 4.2).

Tabela 4.3 – 46- AIM-Indels utilizados no estudo, incluindo a localização de cada marcador no genoma e os alelos relacionados.

| Código MID | Código rs | Cromossomo | Posição(bp) | Alelos descritos no dbSNP |
|-------------------|------------------|-------------------|--------------------|----------------------------------|
| MID-1470 | rs2307666 | 11 | 64729920 | -/GTTAC |
| MID-777 | rs1610863 | 16 | 6551830 | -/GAA |
| MID-196 | rs16635 | 6 | 99789775 | -/CAT |
| MID-881 | rs1610965 | 5 | 79746093 | -/ACTT |
| MID-3122 | rs35451359 | 18 | 45110983 | -/ATCT |
| MID-548 | rs140837 | 6 | 3708909 | -/CT |
| MID-659 | rs1160893 | 2 | 224794577 | -/CT |
| MID-2011 | rs2308203 | 2 | 109401291 | -/CTAGA |
| MID-2929 | rs33974167 | 8 | 87813725 | -/TA |
| MID-593 | rs1160852 | 6 | 137345857 | -/TT |
| MID-798 | rs1610884 | 5 | 56122323 | -/GGGAAA |
| MID-1193 | rs2067280 | 5 | 89818959 | -/AT |
| MID-1871 | rs2308067 | 7 | 127291541 | -/TT |
| MID-17 | rs4183 | 3 | 3192524 | -/TAAC |
| MID-2538 | rs3054057 | 15 | 86010538 | -/AACA |
| MID-1644 | rs2307840 | 1 | 36099090 | -/GT |
| MID-3854 | rs60612424 | 6 | 84017514 | -/TCTA |
| MID-2275 | rs3033053 | 14 | 42554496 | -/TCAGCAG |
| MID-94 | rs16384 | 22 | 42045009 | -/AAC |
| MID-3072 | rs34611875 | 18 | 67623917 | -/GCCCCCA |
| MID-772 | rs1610859 | 5 | 128317275 | -/TAG |
| MID-2313 | rs3045215 | 1 | 234740917 | -/ATTATAACT |
| MID-397 | rs25621 | 6 | 139858158 | -/TTCT |
| MID-1636 | rs2307832 | 1 | 55590789 | -/AA |
| MID-51 | rs16343 | 4 | 17635560 | -/TTTAT |
| MID-2431 | rs3031979 | 8 | 73501951 | -/ATTG |
| MID-2264 | rs34122827 | 13 | 63778778 | -/AAGT |
| MID-2256 | rs133052 | 22 | 41042364 | -/CAT |
| MID-128 | rs6490 | 12 | 108127168 | -/ATT |
| MID-15 | rs4181 | 2 | 42577803 | -/AAATACACAC |
| MID-2241 | rs3030826 | 6 | 67176774 | -/GTCCAATA |
| MID-419 | rs140708 | 6 | 170720016 | -/AATGGCA |
| MID-943 | rs1611026 | 5 | 82545545 | -/TGAT |
| MID-159 | rs16438 | 20 | 25278470 | -/CCCCA |
| MID-2005 | rs2308161 | 10 | 69800909 | -/AACAAT |
| MID-250 | rs16687 | 7 | 83887882 | -/CA |
| MID-1802 | rs2307998 | 5 | 7814345 | -/GGA |
| MID-1607 | rs2307803 | 3 | 108981031 | -/TG |
| MID-1734 | rs2307930 | 6 | 84476378 | -/CCAT |
| MID-406 | rs25630 | 6 | 14734341 | -/AG |
| MID-1386 | rs2307582 | 1 | 247768775 | -/AAACTATTCATTTTTACCCT |
| MID-1726 | rs2307922 | 1 | 39896964 | -/CAAGAACTATAAT/CACTATCTATTAT |
| MID-3626 | rs11267926 | 15 | 45526069 | -/AATATAATTTCTCCA |
| MID-360 | rs25584 | 12 | 112145217 | -/AA |
| MID-1603 | rs2307799 | 5 | 70828427 | -/TTGT |
| MID-2719 | rs34541393 | 20 | 30701405 | -/AACT |

Legenda- Código rs: Listagem dos marcadores de acordo com dbSNP (build 132); MID: Nomenclatura dos marcadores de acordo com o banco de dados *Marshfield Diallelic Insertion/Deletion Polymorphisms*.

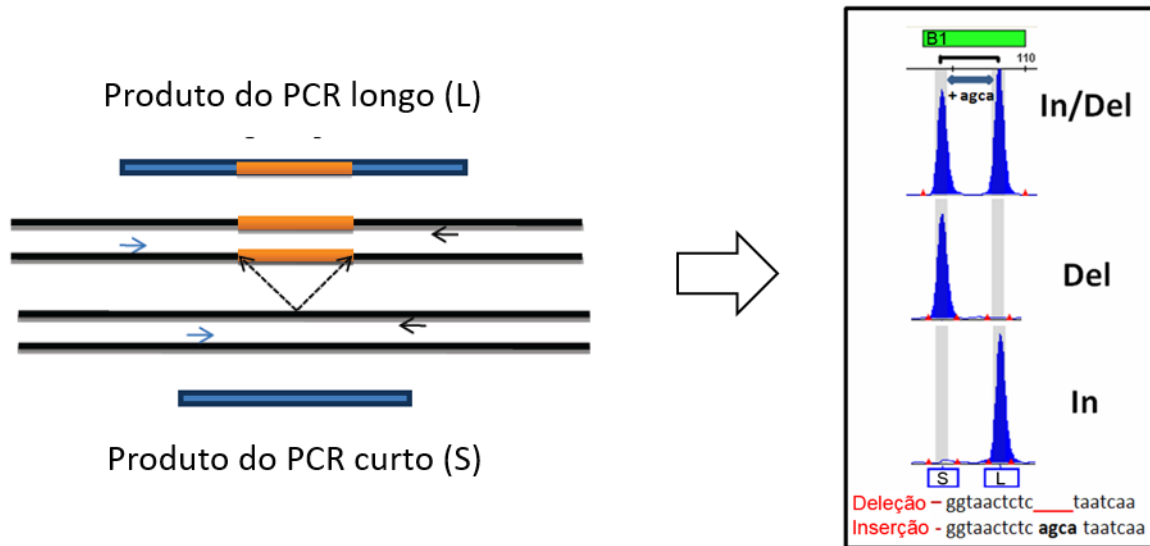


Figura 4.2 – Genotipagem dos marcadores do tipo Indels. Os *primers* da reação estão representados pelas setas e se encontram nas regiões flanqueadoras da sequência alvo. A presença do segmento de DNA (em laranja) irá resultar na amplificação do alelo longo (L). A ausência do segmento de DNA irá resultar na amplificação do alelo curto (S). O primeiro eletroferograma representa um indivíduo heterozigoto, enquanto os outros representam os indivíduos homozigotos para somente um dos alelos, curto (Del) ou longo (In) (Adaptado de Pereira & Gusmão, 2011).

4.2.5 – Análise de ancestralidade e estrutura populacional

As estimativas de ancestralidade individual e global foram calculadas através dos softwares Structure 2.3.3 (258) e Admixture 1.3.0 (281), utilizando como amostras de referência as populações africanas, europeias e nativo-americanas do painel de diversidade HGDP-CEPH (H952). As populações de referência do HGDP-CEPH, bem como o número de indivíduos de cada grupo populacional, estão relacionadas na Tabela 4.4. Em ambos os softwares o número de clusters ancestrais (K) utilizado foi igual a 3, suficientes para explicar a estrutura genética esperada nessas populações. O “barplot” foi realizado no software R 3.3.1, utilizando as estimativas individuais geradas no Admixture, já que essa ferramenta permite trabalhar com maior número de dados.

Tabela 4.4 – Caracterização das populações do painel HGDP-CEPH utilizadas como referência.

| Populações parentais HGDP-CEPH | Grupos populacionais | N | Ancestralidade (%) |
|---------------------------------------|---|----------|---------------------------|
| Africanos (N=105) | Bantu Sul da África – Botswana ou Namíbia, Angola, Lesoto e África do Sul | 8 | 96.54% |
| | Bantu do Quênia | 11 | |
| | Mandeka – Senegal | 22 | 98.84% |
| | Yoruba – Nigéria | 22 | 99.30% |
| | Biaka Pygmy – República da África Central | 23 | 99.67% |
| | Mbuti Pygmy – Congo | 13 | 99.95% |
| | San - Namíbia | 6 | 99.95% |
| Europeus (N=158) | Orcadianos – Ilhas Órcades, Escócia | 15 | 98.14% |
| | Adygei – Cáucaso Russo | 17 | 88.58% |
| | Rússia | 25 | 96.61% |
| | Basco – França | 24 | 97.97% |
| | França | 28 | 92.35% |
| | Bérgamo – Itália | 13 | 93.57% |
| | Sardenha – Itália | 28 | 97.55% |
| | Toscana – Itália | 8 | 99.61% |
| Nativo-americanos (N=64) | Colômbia | 7 | 95.50% |
| | Karitiana – Brasil | 14 | 99.99% |
| | Surui – Brasil | 8 | 99.95% |
| | Maya – México | 21 | 85.37% |
| | Pima – México | 14 | 99.66% |

Para facilitar a representação das ancestralidades, mantemos o vermelho para africanos, verde para europeus e azul para nativo-americanos nas figuras e tabelas desse estudo.

4.3 – Análise de associação genética

4.3.1 - Seleção de SNPs candidatos

A seleção dos “tag SNPs” no gene *PKLR* (chr1: 155.259.084-155.271.244 - GRCh37/hg19) e *G6PD* (chrX: 153.754.606-153.780.233 - GRCh37/hg19) foi realizada a partir da análise das variantes genéticas disponíveis publicamente em 1354 indivíduos da fase 3 do projeto 1000Genomes, sendo 504 africanos (excluindo os africanos miscigenados, ACB e ASW), 503 europeus e 347 nativo-americanos. De início, as frequências alélicas foram calculadas utilizando o PLINK (282). Em seguida, foi realizada a anotação funcional dos SNPs com o auxílio do software

Annovar (283) utilizando o genoma de referência hg19 (11 dez 2015). Em seguida, realizamos a Análise de Componente Principal (PCA) através do Eigensoft 4.2 (284) com o objetivo de selecionar possíveis SNPs funcionais dentre os principais clusters formados a partir dessa análise (Figura 4.3). Nesse sentido, o peso de cada SNP ou “SNP weight” referente ao Componente Principal 1 (PC1) e Componente Principal 2 (PC2) foram considerados. Assim, selecionamos os 30 SNPs com os maiores valores de “SNP weight” para o PC1 (denominados de “top SNPs”) e avaliamos a categoria funcional identificada pela anotação. O próximo passo foi realizar a análise de desequilíbrio de ligação (LD) entre os “top SNPs” no Haploview (285), considerando $r^2 > 0,8$ como elevado LD. Por fim, a partir da seleção dos SNPs representativos de cada bloco, foi feita a análise da frequência dos haplótipos em Europeus e Africanos com o auxílio do software Haploview. Ainda, como última análise, os “tag SNPs” disponíveis nas três coortes Brasileiras do projeto EPIGEN (90 indivíduos de Salvador, 88 de Bambuí e 87 de Pelotas) (<https://epigen.grude.ufmg.br>) também foram consultados para comparar a frequência dos SNPs/haplótipos no Brasil. Em resumo, os seguintes critérios foram utilizados para a seleção dos “tag SNPs”: I) Frequência alélica minoritária (MAF) $> 0,05$; II) Anotação funcional; III) Peso do SNP no PCA; IV) Análise de desequilíbrio de ligação ($r^2 < 0,8$); V) Frequência do haplótipo no 1000Genomes e Epigen e VI) Relatos na literatura. O fluxograma resumido da estratégia empregada para a seleção dos SNPs está representando na Figura 4.2. Os SNPs selecionados a partir dessa metodologia estão descritos na Tabela 4.5.

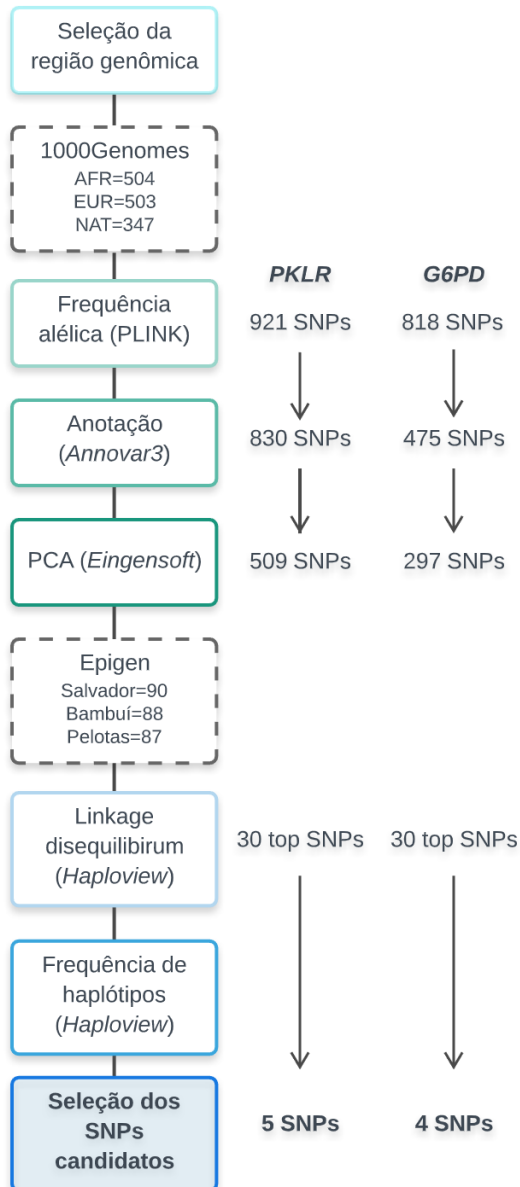


Figura 4.3 – Fluxograma de seleção dos SNPs funcionais na região do gene *PKLR* e *G6PD*. As populações utilizadas estão na caixa tracejada. A primeira etapa da metodologia descrita utilizou as populações do 1000Genomes. A partir da seleção dos “top SNP”, foi possível avaliar a distribuição das frequências alélicas, haplotípicas e o perfil de LD também nas populações brasileiras do Epigen, caso o SNP estivesse disponível no banco.

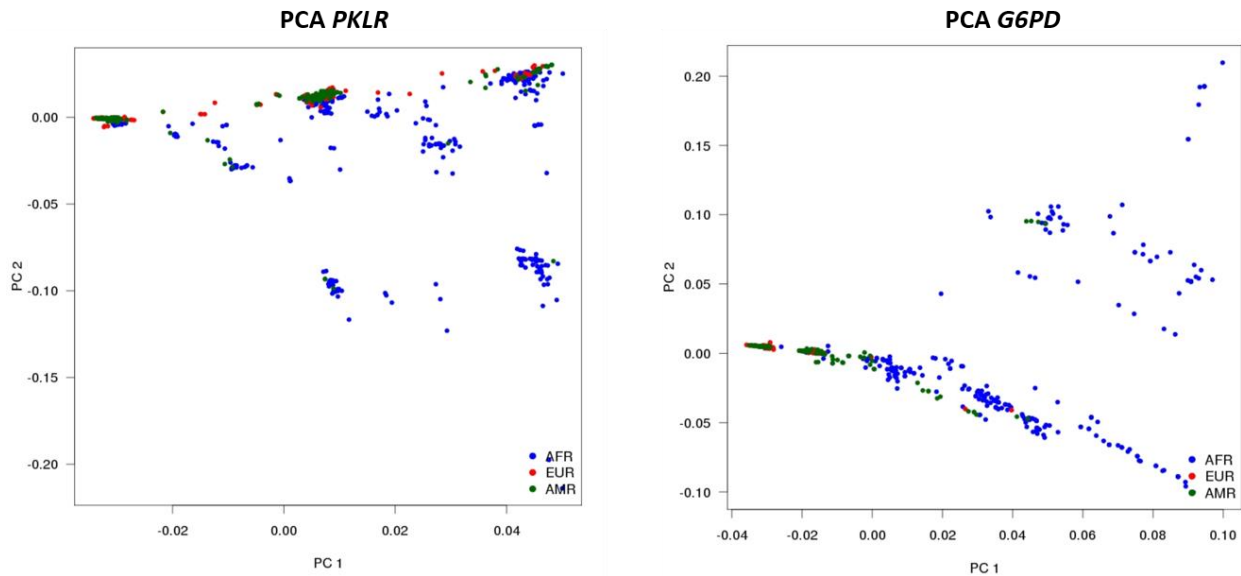


Figura 4.4 – Análises de Componente Principais (PCA) na população do 1000Genomes para a busca de SNPs funcionais. Nessa estratégia, utilizamos os SNPs anotados de uma região de 12.000pb *upstream* e *downstream* do loci do *PKLR* (chr1:155.259.084 - 155.271.244 – GRCh37/hg19) e de 5.000pb *upstream* e *downstream* do loci do *G6PD* (chrX:153.754.606 - 153.780.233 – GRCh37/hg19) para observar os cluster formados pelo Componente Principal 1 e 2 (PC1 e PC2). Posteriormente, selecionamos os 30 SNPs que apresentavam os maiores pesos (“eigenvalue”) dentro do PC1, o qual denominamos de “top SNPs”. Correlacionando os valores do PC com a informação da anotação, LD, frequência e citação na literatura, selecionamos os nossos SNPs candidatos. AFR-Africanos; EUR-Europeus e AME-Nativo-americanos.

Tabela 4.5 – Seleção dos SNPs nos genes *PKLR* e *G6PD* candidatos ao estudo de associação.

| SNP | A1 | A2 | Gene | Freq. A2 EUR | Freq. A2 AFR | Região | SNP weight ¹ (PC1) | Citação na literatura |
|--------------------|----|----|-------------|--------------|--------------|--------------------------|-------------------------------|------------------------------------|
| <i>PKLR</i> | | | | | | | | |
| rs11264355 | C | G | <i>HCN3</i> | 0.28 | 0.49 | Intron | 18º | Machado (2010)* |
| rs1052176 | G | T | <i>PKLR</i> | 0.28 | 0.47 | Éxon (<i>missense</i>) | 22º | Machado (2010)* Bruggen (2015)* |
| rs4620533 | C | G | <i>PKLR</i> | 0.27 | 0.56 | Intron | 14º | Machado (2010)* Bruggen (2015)* |

| | | | | | | | | |
|--------------------------------------|---|---|---------------------|------|------|--------------------------|-----|------------------|
| rs4971072 | A | G | <i>PKLR / FDPS</i> | 0.32 | 0.97 | Intergênica | 30º | - |
| rs11264359 | A | G | <i>FDPS</i> | 0.28 | 0.67 | Intron | 13º | Machado (2010) * |
| G6PD | | | | | | | | |
| rs1050829 (A376G)² | A | G | <i>G6PD</i> | 0.00 | 0.34 | Éxon (<i>missense</i>) | 13º | ** |
| rs1050828 (G202A)² | G | A | <i>G6PD</i> | 0.00 | 0.13 | Éxon (<i>missense</i>) | 28º | ** |
| rs2472393 | C | T | <i>G6PD / IKBKG</i> | 0.10 | 0.97 | Intron | 5º | - |

Frequência dos SNPs baseado no GRCh37. A1=Alelo 1; A2=Alelo 2; Freq.=Frequência; EUR=Europeus e AFR=Africanos
 Frequência alélica baseada na anotação dos SNPs e populações do 1000Genomes. As mutações. ¹Posição dos "top SNPs" ordenados por ordem crescente dos valores de "SNP weight" no PC1. ²Representação comumente encontrada na literatura. *Citação em trabalhos anteriores relacionando os SNPs do gene *PKLR* com a malária. **Citação ampla na literatura em estudos de associação com a malária.

4.3.2 – Genotipagem por PCR em tempo real

A genotipagem dos SNPs foi realizada ocorreu em placas de 96 poços através do sistema *TaqMan*[®] *SNP Genotyping Assay* de discriminação alélica por PCR em tempo real (Thermo Fisher). Neste sistema, um par de *primers* amplifica a região genômica onde se encontra o polimorfismo e cada alelo do polimorfismo é detectado por uma sonda marcada com os fluoróforos VIC ou FAM. A intensidade de fluorescência referente a cada fluoróforo determina os genótipos homocigotos (fluorescência aumentada para somente um dos fluoróforos) ou heterocigotos (fluorescência aumentada para os dois fluoróforos). As reações ocorreram na *StepOnePlus*[™] (Thermo Fisher) em volume final de 10 µL contendo *TaqMan*[®] *Genotyping Master Mix* (Thermo Fisher), 20 a 30 ng de DNA genômico e sonda *Taqman* (40X) (Thermo Fisher), cumprindo a ciclagem recomendada pelo fabricante. Os genótipos foram analisados no software *StepOne 2.3* (Thermo Fisher).

4.3.3 – Análises do estudo de associação

O controle de qualidade para a padronização das genotipagens em todas as populações foi realizado no PLINK 1.07 (282). Os critérios aplicados foram: I) Taxa de genotipagem > 90% em cada grupo populacional; II) SNPs em acordo com o equilíbrio de Hardy-Weinberg (HWE) nos grupos controles ($p > 0.05$) e III) Exclusão de SNPs com MAF < 0.05. O desvio do equilíbrio de HWE foi avaliado através do teste do Qui-quadrado. As frequências alélicas, genotípicas, carreadores do alelo minoritário e haplótipos entre casos e controles foram comparadas por regressão logística no software R 3.3.1, através dos pacotes “genetics” e “haplo.stats”. As frequências haplotípicas foram estimadas através de máxima verossimilhança, e comparadas entre casos e controles, conforme descrito. A análise de regressão logística foi utilizada para a obtenção da estimativa de associação, a OR (*Odds Ratio*, ou razão de chances), com 95% de intervalo de confiança (IC) para cada estudo. A correção para a estratificação populacional foi realizada incluindo as covariáveis – ancestralidade genética, sexo e idade (conforme o dado se encontrava disponível) – no modelo de associação. A indicação de maior susceptibilidade ocorre para $OR > 1$, enquanto $OR < 1$ sugere proteção para o nível de significância em p valor < 0.05. O desequilíbrio de ligação (LD) foi calculado para cada par de SNPs através do coeficiente r^2 pelo software Haploview 4.2 (285). Para as análises de associação, blocos com $r^2 > 0.70$ foram considerados em alto LD, blocos com $r^2 < 0.25$ considerados em baixo LD, e valores de r^2 entre esses dois extremos foram considerados variantes em LD moderado.

4.3.4 – Busca de eQTLs e correlação dos SNPs com a literatura

A pesquisa de SNPs eQTLs foi realizada através do banco eletrônico do “projeto GTEx” (*Genotype-Tissue Expression project*), que disponibiliza dados de expressão gênica e variantes genéticas globais em diferentes tecidos humanos, disponível em: <https://gtexportal.org/home/>.

Os dados do estudo GWAS na população de Chineses de Zhang e colaboradores (2009) está disponível para acesso no banco de dados genotípicos e fenotípicos dbGaP. Para verificar se havia relato da associação dos SNPs do gene PKLR com a hanseníase nesse estudo, utilizamos o número de acesso: phs000217.pha002872.

4.4 – Análises de seleção natural

4.4.1 – Dados genômicos

Para este estudo, foram obtidos os dados do genoma referente ao cromossomo 1 das populações do Projeto 1000 Genomes fase III, acessados através do link: <ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/>. O *dataset* inclui variantes em 504 Africanos e 503 Europeus. O pré-processamento dos dados incluiu a remoção dos Indels e *singletons* e a exclusão de SNPs com $MAF < 0.05$ através do VCFtools.

O *dataset* de Gouveia e colaboradores (2019), explorado nesse estudo, compreende duas populações africanas endêmicas para a malária: Ghana, composta 945 indivíduos de 35 tribos (principalmente das famílias linguísticas Kwa e Gur Niger-Congo), e Uganda que, com 568 indivíduos de 17 tribos (predominantemente da família linguística do Oeste do Nilo). Uganda foi dividida, para fins práticos, em Uganda Norte-Oeste (UNW) e Uganda Norte-Central (UNC) porque são separadas geograficamente por um desfiladeiro (Vale do Rift) que pode impedir o fluxo gênico entre elas. Apesar de ambas estarem em regiões endêmicas da malária, elas apresentam diferentes padrões de ancestralidade genética. Como controles foram utilizados 186 indivíduos do Sul africano, Sotho e Zulu, onde as clássicas variações da malária são reportadas em baixa frequência (280).

4.4.2 – Métodos de detecção de seleção estudados

A fim de investigar assinaturas de seleção na região do *PKLR*, inicialmente avaliamos as diferenças entre as frequências alélicas dos SNPs entre europeus e africanos pelo teste exato de Fisher. Em seguida, exploramos a abordagem “ F_{ST} outlier”, um teste baseado nas frequências alélicas de SNPs e comumente utilizado em genética de populações para detectar genes sob seleção. O cálculo do F_{ST} foi estimado para os pares de populações europeus e africanos no VCFtools. SNPs acima do quartil de 95% são considerados *outliers*. A análise de F_{ST} foi feita para os SNPs individuais e para janelas de 20 SNPs com sobreposição de 15 SNPs. Variantes com

MAF<0.05 foram excluídas para evitar subestimar o valor de F_{ST} , como descrito por Brandt e colegas (2017) (52).

O teste de Tajima D foi estimado no VCFtools para cada população em janelas de 1000 bases ao longo do genoma a fim de verificar desvios de neutralidade. O valor negativo do D de Tajima indica ocorrência de seleção positiva ou expansão populacional (maior número de variantes raras). Os valores positivos sugerem que a região sofreu seleção balanceadora ou contração populacional.

Os testes baseados em homozigidade do haplótipo estendido (EHH) foram o iHS (score de integração dos haplótipos) – baseado na padronização do log da razão das integrais do decaimento observado para os alelos ancestral e derivado em um SNP alvo – e o xpEHH, que compara o decaimento de homozigidade dado uma distância x do SNP “core” entre duas populações. Utilizamos europeus e africanos – excluindo os africanos miscigenados, africanos caribenhos de Barbados (ACB) e americanos descendentes de africanos (ASW) – do 1000Genomes. A região analisada compreendeu 37.000 pb (chr1: 155,247,000-155,284,000 bp) contendo o *loci* do *PKLR* e os genes flaqueadores. A análise do EHH, iHS e xpEHH foi realizada no VCFtools utilizando os haplótipos das populações do 1000Genomes em fase, realizado pelo SHAPEIT (286), e a interface gráfica foi gerada no pacote “rehh” em ambiente R (287) .

4.5 – Análises funcionais

4.5.1 – População do estudo funcional

A fim de fundamentar os resultados obtidos no estudo caso-controle, foi realizado um estudo funcional utilizando 148 voluntários sadios residentes na cidade do Rio de Janeiro e 141 pacientes de hanseníase do Ambulatório Souza Araújo (FIOCRUZ). Os participantes sadios incluem 43 homens (± 35 anos) e 105 mulheres (± 30 anos) com média de idade de 32 anos. Os indivíduos desse grupo declararam não apresentar nenhuma doença infecciosa, hepática, metabólica ou crônica e relação com alcoolismo. Já os pacientes do Rio de Janeiro apresentam 93 homens (± 43 anos) e 48 mulheres (± 49 anos) com idade média total de 45 anos. Além desses, um grupo de 59

pacientes de hanseníase (idade média de 43 anos) do município de Rondonópolis, apresentando 40 homens (± 47 anos) e 19 mulheres (± 36 anos), foi utilizado. Os critérios para classificação dos pacientes seguem os mesmos já descritos.

4.5.2 – Quantificações sorológicas

O sangue dos indivíduos foi coletado em tubos *vacutainer* (BD) específicos para soro durante o período da manhã na condição de jejum de 8 horas. Após 30 minutos, o soro foi centrifugado durante 15 minutos a 3.200 rpm e 16°C, e posteriormente armazenado a -20°C até o momento de leitura das proteínas. Os ensaios de quantificação do ferro sérico ($\mu\text{g/dL}$), ferritina (ng/mL) e haptoglobina (mg/dL) foram realizados a partir de protocolos padronizados em colaboração com o Laboratório Sérgio Franco (Rio de Janeiro). Para a análise das dosagens, estratificamos os grupos indivíduos por genótipos ou haplótipos, e utilizamos uma regressão linear pelo método dos mínimos quadrados considerando o ajuste para as covariáveis sexo, idade e forma clínica da hanseníase. O modelo linear estima a média de cada covariável e ajusta cada observação para esse valor. Após a remoção do efeito confundidor provocado pelas covariáveis, os dados ajustados das quantificações foram plotados em gráficos *boxplot* no ambiente R. Os *whiskers* (linha que representa a variabilidade fora do quartil superior e do quartil inferior) foram ordenados de acordo com o método de Tukey, com valores *outliers* mostrados caso sejam $1,5 \times$ intervalo interquartil (IQR) acima ou abaixo do 3º e 1º quartis, respectivamente.

4.5.3 – Extração de RNA de sangue total e síntese de cDNA

O RNA foi extraído de tubos de PAXgene (PreAnalytiX, Quiagen/BD), de acordo com o fabricante. O sangue foi centrifugado a 5000 x g por 10 minutos em temperatura ambiente, o sobrenadante removido e a ele foram adicionados 4mL de água (RNase *free*). Em seguida, o sedimento foi homogeneizado em vórtex e centrifugado a 5000 x g por 10 minutos. O sobrenadante foi removido e ao sedimento foram adicionados 350 μL de tampão BM1. A solução foi transferida para um tubo de 1,5mL e acrescentados 300 μL de BM2 e 40 μL de proteinase K. Em seguida, as amostras foram incubadas a 55°C durante 10 minutos sob agitação. Adicionou-se

a coluna *PAXgene Shredder Spin column*, fornecida pelo kit, e os tubos foram centrifugados a 20000 x g por 3 minutos. A esse sobrenadante coletado foram adicionados 700µL de isopropanol. Esta solução foi adicionada a uma nova coluna de membrana de sílica, também fornecida pelo kit (*Paxgene Spin column*) e centrifugada a 20000 x g por 1 minuto. Após, a membrana foi lavada com 350µL de tampão BM3 e centrifugado novamente a 20.000 x g por 1 minuto. Posteriormente, a membrana foi lavada por duas vezes com tampão BM4 e centrifugados a 20.000 x g por 2 minutos. A coluna foi então transferida para um tubo novo e centrifugado novamente a 20000xg por 1 minuto. A coluna foi transferida novamente para um novo tubo e foram adicionados 80µL de tampão de eluição BR5 seguido de centrifugação 20.000 x g por 1 minuto. O RNA foi desnaturado por 5 minutos e armazenados a -80°C.

O RNA foi quantificado pelo Nanodrop e tratado com DNase pelo kit DNA-free seguindo as instruções do fabricante (Thermo Fisher). A integridade do RNA foi verificada a partir de eletroforese em gel de agarose 1,2%. O total de 500ng de RNA foi transcrito em cDNA pelo kit SuperScript VILO Master Mix Synthesis. A reação ocorreu a 42°C durante 1 hora seguidos de 5 minutos a 85°C. O cDNA foi diluído para 5ng/µL e a quantificações confirmada pelo kit Qubit ssDNA Assay Kit (Thermo Fisher).

4.5.4 – Ensaios de infecção com *M. leprae* em monócitos primários

Os resultados gerados a partir dos ensaios de infecção com o *M. leprae* e as dosagens enzimáticas foram compartilhados com a tese de doutoramento de Medeiros (2018), onde se encontra a descrição mais detalhada dos ensaios aqui descritos (288). Monócitos primários foram isolados de indivíduos sadios (homens e mulheres) baseado na presença ou não das mutações de *G6PD* (variante A-), definidas pela genotipagem dos SNPs. Essas células foram obtidas de PBMCs (células mononucleares periféricas do sangue) a partir da sedimentação em Ficoll-Paque (GE Healthcare Life Science). PBMCs aderiram aos frascos de cultura por 2 horas em meio RPMI sem soro a 37°C. Células não aderentes foram removidas por lavagem com PBS (solução salina fosfatada). Monócitos aderentes foram cultivados em meio RPMI 1640 completo a 37°C e 5% de CO₂. Após 2 horas, os monócitos foram infectados com *M. leprae* vivo (MOI = 50:1 [bactéria/célula]) por 48 horas.

O RNA dos monócitos primários foi isolado a partir de TRIzol[®], de acordo com o fabricante (Thermo Fisher), o processamento e transcrição a cDNA foram realizados como descrito anteriormente.

4.5.5 – Ensaios de atividade enzimática

Os ensaios de atividade enzimática foram realizados em 1×10^6 células após 48 horas de infecção em cultura. As células foram lisadas em tampão hipotônico (25 mM KH_2PO_4 , 10 mM MgCl_2 , pH 7.4) e os extratos celulares foram usados em uma razão 1:2 com um meio de reação específico para a enzima contendo 50 mM Tris-HCl (pH 7.4) com 5 mM MgCl_2 e 1 mM glicose 6-fosfato e 0.2 mM NADP⁺. A presença do NADP reduzido foi determinada pela absorbância em 340 nm utilizando o espectrofotômetro de alta performance (BIOTEK).

4.5.6 – PCR em tempo real para análise da expressão gênica (qRT-PCR)

A fim de avaliar se os SNPs do gene *PKLR* regulam os níveis de expressão de genes da ferritina (FTL), haptoglobina (HP) e o próprio *PKLR*, realizamos ensaios de PCR em tempo real em amostras de sangue total de indivíduos carreadores e não carreadores dos genótipos de risco, utilizando o gene *RPL13a* como referência. O mesmo sistema foi utilizado para quantificar a expressão gênica do *G6PD* em sobrenadante de cultura de indivíduos carreadores e não carreadores da variante A-.

A reação de PCR em tempo real para avaliação da expressão gênica foi realizada utilizando-se o sistema Sybr Green (Thermo Fisher), seguindo as instruções do fabricante. As reações apresentaram volume final de 10 μL , onde foram adicionados 2 μL de cada cDNA, e 8 μL contendo os *primers* de interesse a 10 μM , Fast Sybr Green Master Mix 2X (Thermo Fisher) e água. Para cada reação foi amplificado o cDNA dos genes de interesse e o gene constitutivo RPL13a. As reações foram incubadas no sistema de PCR em tempo real StepOne Plus[®] (Thermo Fisher), seguindo as condições da reação: 95°C por 10 minutos, seguido de 40 ciclos de 95°C por 15 segundos e 60°C por 1 minuto. Ao final da reação de amplificação, as amostras foram submetidas a uma nova incubação para geração da curva de dissociação, onde se determina o ponto correspondente à temperatura de

dissociação dos oligonucleotídeos de suas sequências alvo. A sequência dos *primers* utilizados nesse estudo estão disponíveis na Tabela 4.6.

Os arquivos .RDML foram exportados para a utilização no LinRegPCR v.2012 para a estimativa da eficiência da PCR (289,290). Os valores de N_0 , que correspondem às unidades de fluorescência específicas do alvo no ciclo 0, foram exportadas do LinRegPCR e normalizadas utilizando o gene de referência (*RPL13a*) (291).

Tabela 4.6 – Sequência dos oligonucleotídeos utilizados nos ensaios de qRT-PCR.

| Gene | Fita | Sequência (5' -> 3') |
|---------------|-----------|----------------------------|
| <i>RPL13a</i> | Senso | GACAAGAAAAAGCGGATGGT |
| | Antisenso | GTACTTCCAGCCAACCTCGT |
| <i>PKLR</i> | Senso | AGCAAAGTCCATCCTGATTGGG |
| | Antisenso | CTCGGAGTCAATGTCCAGTAGG |
| <i>HP</i> | Senso | AGACCAGCAGATGCCCCACAGC |
| | Antisenso | GCGATATCCGTGACATCATTGC |
| <i>FTL</i> | Senso | AAAGTAATAGGGCTTCTGCCTAA |
| | Antisenso | AAGGCTTGTTAGGATAGTTAAGAAAG |
| <i>G6PD</i> | Senso | AGAGCTTTTCCAGGGCGAT |
| | Antisenso | CACCAGATGGTGGGGTAGAT |

Além disso, os ensaios de expressão gênica de um microarranjo em larga-escala desenvolvido pela Dra. Ana Carla Pereira Latini do Instituto Lauro Souza de Lima (ILSL) foram acessados para verificar a expressão de genes de interesse envolvidos com o metabolismo do ferro (*G6PD*, *PKLR*, *HMOX1*, *FTL*, *HFE*, *HP*, *CD163*, *HAMP*, *HEBP1*, *IDO1*, *IL10*, *TF* e *SLC40A1*). Esse estudo, em que os dados se encontram em fase de pré-processamento para a preparação do manuscrito, foi realizado em monócitos primários isolados de PBMC, como já descrito, de 25 indivíduos saudáveis de Bauru. Essas células foram estimuladas com *M. leprae* vivo (MOI 10:1) e a expressão quantificada após 48 horas de infecção.

Capítulo II

Ancestralidade genômica em grupos populacionais brasileiros utilizados em estudos de associação com a hanseníase

1 – Caracterização da ancestralidade global e individual das populações do estudo

A fim de avaliar a distribuição da ancestralidade genômica entre os casos e controles utilizados nos estudos de associação genética, utilizamos o painel de 46-AIMs-Indels para estimar as proporções de ancestralidade europeia, africana e nativo-americana entre os indivíduos. As estimativas de ancestralidade individual foram utilizadas (1º) para a correção da estratificação populacional nos estudos de associação com a hanseníase. Já a ancestralidade global foi usada (2º) para explorar o perfil de ancestralidade entre as populações das sub-regiões brasileiras adotadas nesse estudo, comparando com os dados prévios da literatura. Ressalta-se que o N amostral utilizado para a segunda abordagem é maior, pois nessa análise utilizamos todos os indivíduos em que foi possível estimar a ancestralidade, ao contrário do estudo de associação em que foram excluídas as amostras sem o dado da genotipagem ou com informação faltante.

Um trabalho prévio do laboratório, desenvolvido por Mota-Silva (2015), foi pioneiro na descrição das estimativas de ancestralidade da população de casos e controles do Rio de Janeiro utilizadas no estudo de associação com a hanseníase (292). Nesse estudo, a ancestralidade foi estimada para 539 pacientes e 580 controles. Porém, em vista do aumento do N amostral do grupo de casos para o estudo de associação apresentado no tópico seguinte, o primeiro passo foi calcular a ancestralidade dos novos pacientes do Rio de Janeiro (N=288) e comparar com o dado prévio. A Tabela 1.1 compara as proporções de ancestralidade em casos analisadas no trabalho de Mota-Silva e as descritas no presente estudo. Os 288 novos pacientes apresentam 32%, 53% e 15% de ancestralidade africana, europeia e nativo-americana, respectivamente, apresentando proporções similares às encontradas para os 539 pacientes de Mota-Silva.

Tabela 1.1 – Comparação das estimativas de ancestralidade entre os casos de hanseníase do Rio de Janeiro.

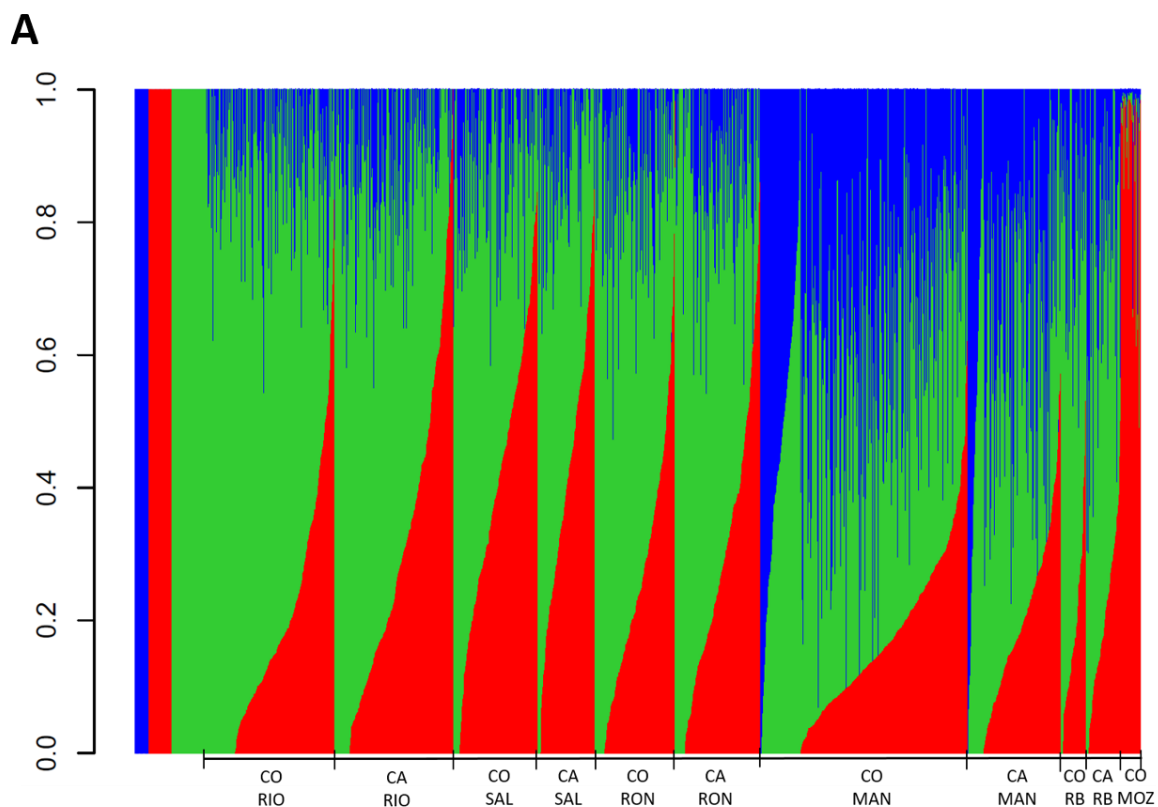
| Grupos | AFR | EUR | NAT | Total |
|-------------------|------------|------------|------------|--------------|
| Africanos | 99,6% | 0,2% | 0,2% | 105 |
| Europeus | 0,2% | 99,3% | 0,5% | 158 |
| Nativo-americanos | 0,1% | 2,2% | 97,7% | 64 |
| Casos (2015) | 30% | 57% | 13% | 539 |
| Casos (2019) | 32% | 53% | 15% | 288 |

As estimativas de ancestralidade iniciais foram calculadas pelo Structure.

A Figura 1.1 A engloba as proporções de ancestralidade de todas as populações do estudo, incluindo as populações parentais: nativo-americanos (azul), africanos (vermelho) e europeus (verde). Observa-se que, no geral, há uma maior contribuição europeia na população brasileira apresentada, seguido de africanos e, por fim, nativo-americanos. A comparação entre as populações brasileiras demonstra que Salvador apresenta o maior componente africano (maior área em vermelho), enquanto Manaus e Rio Branco refletem a elevada herança nativo americana (maior área em azul) presente na região Norte. Analisando as estimativas numéricas na Figura 1.1 B, é visto que Salvador carrega 47-48% de proporção africana, Manaus apresenta a maior proporção nativo-americana, 40-44%, seguido de Rio Branco, 24-26%. Rondonópolis, por sua vez, apresenta uma proporção aproximada com os casos do Rio de Janeiro, com um leve aumento do componente europeu e nativo-americano. A título de comparação, utilizamos também dados prévios da ancestralidade em controles de Moçambique, observando que essa população africana apresenta algum grau de miscigenação.

As proporções de ancestralidade numéricas podem ser visualizadas na Figura 1.1 B. A comparação entre casos e controles demonstra que não há grandes diferenças na composição da ancestralidade para os grupos populacionais brasileiros, exceto para o Rio de Janeiro. Apesar das proporções similares de nativo-americanos no Rio de Janeiro, uma maior ancestralidade europeia é vista em controles (70%) comparado aos casos (58%) e maior ancestralidade africana é exibida

nos casos (29%) comparado aos controles (18%). Essa diferença entre casos e controles do Rio de Janeiro foi a motivação para considerarmos a ancestralidade genômica para os ajustes do modelo de associação genética descritos a seguir. Em relação a Salvador, Rondonópolis, Manaus e Rio Branco, as diferenças entre as estimativas entre casos e controles são menores, tornando-os bons grupos para a replicação/estudo de associação do tipo caso-controle. Um resumo das estimativas de ancestralidade genômica entre as populações estudadas das diferentes regiões brasileiras pode ser acessado na Figura 1.2.



B

| População | Grupos | AFR | EUR | NAT | Total |
|----------------|-----------|-----|-----|-----|-------|
| Rio de Janeiro | Controles | 18% | 70% | 12% | 579 |
| | Casos | 29% | 58% | 13% | 824 |
| Salvador | Controles | 39% | 48% | 13% | 380 |
| | Casos | 40% | 47% | 13% | 258 |
| Rondonópolis | Controles | 25% | 59% | 16% | 351 |
| | Casos | 28% | 58% | 14% | 384 |
| Manaus | Controles | 15% | 41% | 44% | 941 |
| | Casos | 17% | 43% | 40% | 422 |
| Rio Branco | Controles | 19% | 57% | 24% | 114 |
| | Casos | 20% | 54% | 26% | 156 |
| Moçambique | Controles | 92% | 4% | 4% | 89 |

Figura 1.1 – Proporções de ancestralidade dos grupos populacionais estudados baseadas em 46 indels-AIMs. A. No *barplot*, os indivíduos estão representados em colunas verticais compostas pelas porcentagens (eixo y = 0-100%) de contribuição de cada população ancestral (Africanos (AFR) = vermelho; Europeus (EUR) = verde e Nativo-americanos (NAT) = azul). Os grupos de controles (CO) e casos (CA) para

cada população foram representados nessa ordem, exceto em Moçambique onde somente o dado em controles estava disponível (a ancestralidade em controles de Moçambique é um dado prévio do laboratório ainda não publicado). Legenda: RIO=Rio de Janeiro; SAL=Salvador; RON=Rondonópolis; MAN=Manaus, RB=Rio Branco e MOZ=Moçambique. **B.** Média das proporções de ancestralidade (%) em casos e controles de cada população calculada pelo Admixture (esse software foi utilizado para as análises englobando todas as populações).

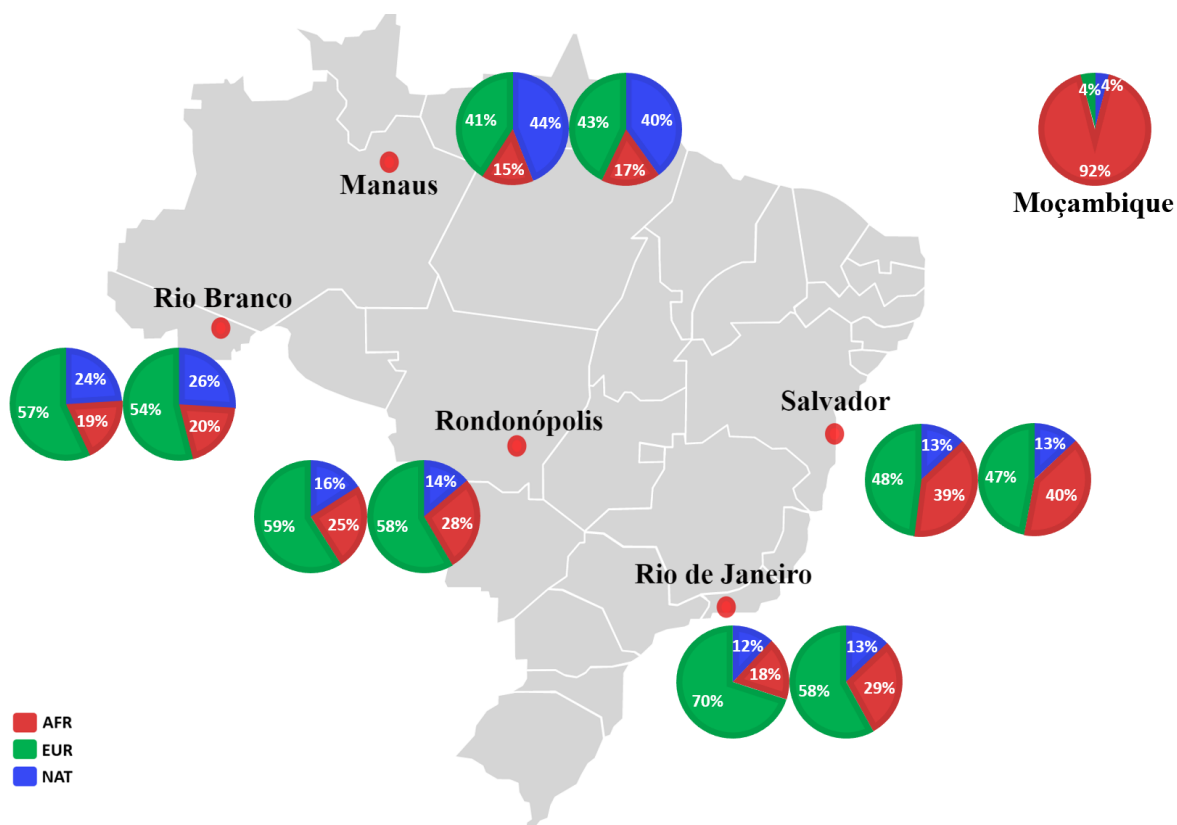


Figura 1.2 – Proporções (%) de ancestralidade genética entre casos e controles das populações endêmicas para a hanseníase entre as sub-regiões brasileiras. Controles e casos estão representados, nessa ordem (da esquerda para a direita), em cada população nos gráficos de pizza. Legenda – AFR: Africanos, EUR: Europeus e NAT: Nativo-americanos.

2 – Ancestralidade da população brasileira comparada aos estudos prévios

A fim de avaliar a correlação das estimativas de ancestralidade entre as diferentes populações brasileiras descritas na literatura, os dados de ancestralidade caracterizados nesse trabalho foram comparados com os estudos de Manta e colaboradores (2013), Pereira e colaboradores (2012), Santos e colaboradores (2009), Pena e colaboradores (2011) e Lima-Costa e colaboradores (2015). Para essa comparação, utilizamos somente os indivíduos sadios que são representativos das populações em questão. Na Tabela 2.1, observa-se que a maioria das populações desse estudo se correlacionam com os dados prévios da literatura quando avaliadas por região. As populações do Norte têm a maior proporção de nativo-americanos em relação às outras regiões brasileiras. À exceção de Santa Isabel do Rio Negro, que é uma comunidade indígena, Manaus é o município que apresenta a maior contribuição nativo-americana descrita em um município urbano no Brasil, vista por Manta e colaboradores (2013) e descritas nesse estudo. Rio Branco tem semelhança com a descrição da ancestralidade de Belém do estudo de Pereira (2012) e Manta (2013). Já a população de Salvador apresenta a maior herança africana dentre as populações Nordesteiras (Ilhéus – Pena, 2011 – e Alagoas e Pernambuco – Manta, 2013), descritas na literatura. Curiosamente, a população de Salvador do Epigen (50,5%), de Lima-Costa (2015), apresenta uma ancestralidade africana maior em relação aos indivíduos de Salvador deste estudo (39%). Rondonópolis apresentou proporções de ancestralidade esperadas para o Centro-Oeste, similar ao encontrado no estudo de Manta (2013) para o Mato Grosso do Sul. Em Manta e colaboradores (2013), a população do Rio de Janeiro apresentou 31% de ancestralidade africana, 55% europeia e 14% nativo-americana enquanto na amostra populacional utilizada nesse trabalho, observamos 18% de ancestralidade africana, 70% europeia e 13% nativo-americana. Considerando que os 46-AIMs-Indels foram os mesmos marcadores utilizadas por Manta, 2013, esperava-se que os valores de ancestralidade fossem similares entre as duas amostragens do Rio de Janeiro. Nesse caso, provavelmente a diferença está relacionada à origem geográfica das amostras no município do Rio de Janeiro, e não ao painel de AIMs utilizado. Com isso, concluímos que, diferentemente da população de indivíduos sadios do Rio de Janeiro, os demais grupos amostrais referentes às populações desse estudo estão de acordo com a descrição de ancestralidade esperada a partir de dados prévios da literatura.

Tabela 1.2 – Comparação da ancestralidade na população brasileira por diferentes estudos.

| Região | População | AFR | EUR | NAT | Referência |
|--------------|--|------------|------------|------------|--------------------|
| Norte | Santa Isabel do Rio Negro (AM) | 7% | 17% | 76% | Manta, 2013 |
| | Manaus (AM) | 16% | 46% | 38% | Manta, 2013 |
| | Belém (PA) | 17% | 54% | 29% | Pereira, 2012 |
| | Belém (PA) | 11,7% | 61,4% | 26,9% | Santos, 2009 |
| | Comunidades de afro-descendentes (MA, PA e AM) | 80,7% | 9,3% | 10% | Santos, 2009 |
| | Belém (PA) | 11% | 69% | 19% | Pena, 2011 |
| | Manaus (AM) | 15% | 41% | 44% | Este estudo |
| | Rio Branco (AC)¹ | 19% | 57% | 24% | Este estudo |
| Nordeste | Pernambuco | 28% | 57% | 15% | Manta, 2013 |
| | Alagoas | 27% | 55% | 19% | Manta, 2013 |
| | Salvador (BA) ² | 50,5% | 42,4% | 5,8% | Lima-Costa, 2015 |
| | Ilhéus (BA) | 29% | 60% | 9% | Pena, 2011 |
| | Salvador (BA) | 39% | 48% | 13% | Este estudo |
| Centro-Oeste | Mato Grosso do Sul | 26% | 59% | 15% | Manta, 2013 |
| | Terena (MS) | 9% | 13% | 78% | Manta, 2013 |
| | Rondonópolis (MT)¹ | 25% | 59% | 16% | Este estudo |
| Sudeste | Minas Gerais | 29% | 59% | 12% | Manta, 2013 |
| | Espírito Santo | 13% | 74% | 13% | Manta, 2013 |
| | Rio de Janeiro | 31% | 55% | 14% | Manta, 2013 |
| | São Paulo | 25% | 63% | 12% | Manta, 2013 |
| | Rio de Janeiro (RJ) | 17% | 74% | 7% | Pena, 2011 |
| | Bambuí (MG) ² | 9,6% | 83,8% | 5,4% | Lima-Costa, 2015 |
| | Rio de Janeiro (RJ) | 18% | 70% | 12% | Este estudo |
| Sul | Paraná | 18% | 71% | 11% | Manta, 2013 |
| | Santa Catarina | 11% | 80% | 9% | Manta, 2013 |
| | Rio Grande do Sul | 14% | 73% | 13% | Manta, 2013 |
| | Porto Alegre (RS) | 10% | 80% | 9% | Pena, 2011 |
| | Pelotas (RS) ² | 6,6% | 85,6% | 6,3% | Lima-Costa, 2015 |
| | Sul do Brasil | 1% | 95% | 4% | Santos, 2009 |

¹Populações em que ancestralidade está sendo descrita pela primeira vez. ²Populações do estudo EPINGEN. As estimativas de ancestralidade descritas para as populações deste estudo estão marcadas em negrito. Somente o grupo dos indivíduos saudáveis foram utilizados nessa comparação.

Capítulo III

***Associação genética, caracterização
evolutiva e análise funcional do gene PKLR
com a hanseníase***

1 – Análise de associação

1.1 – Associação dos SNPs do gene *PKLR* com a suscetibilidade à hanseníase após o ajuste por ancestralidade genômica na população do Rio de Janeiro

Anteriormente, identificamos a associação de suscetibilidade do gene *PKLR* com a hanseníase na população do Rio de Janeiro, ajustando o modelo de associação para sexo e etnia. Neste trabalho, utilizamos a ancestralidade genômica estimada no Capítulo II para controlar as análises, fornecendo um ajuste mais refinado. A Figura 1.1 representa a comparação entre os SNPs rs11264355, rs4620533, rs4971072 e rs11264359, inicialmente testados para a associação com a hanseníase no Rio de Janeiro, a partir de análises ajustadas e não ajustadas para as variáveis de confusão. Observa-se que o padrão de associação para esses SNPs é similar entre a análise não corrigida e a corrigida considerando a etnia (Figura 1.1 A e B). Já a correção por ancestralidade (Figura 1.1 C) altera o perfil de associação para alguns SNPs, como o rs4620533 e o rs4971072. A maior precisão desse ajuste é visualizada também pela diminuição dos intervalos de confiança referente aos valores de OR. Portanto, em populações heterogêneas, como a do Rio de Janeiro, a correção dos estudos de associação por ancestralidade pode minimizar a chance de associações espúrias provocadas pela estruturação populacional.

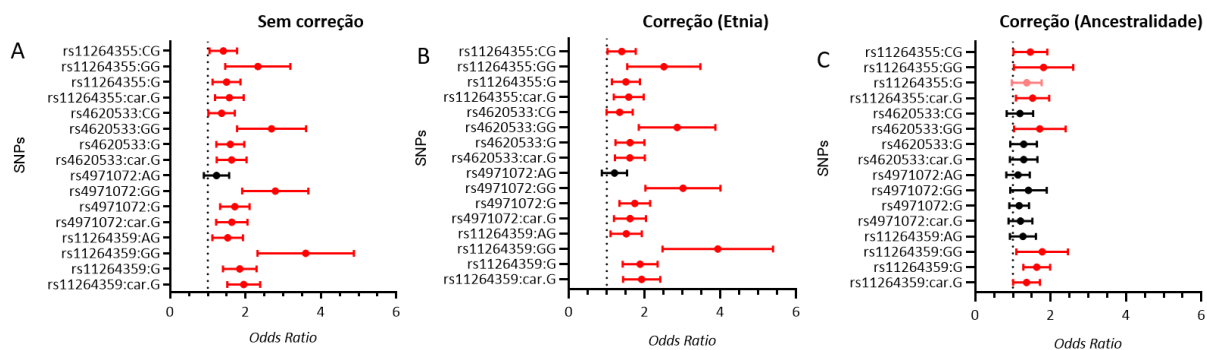


Figura 1.1 – Regressão logística dos SNPs do gene *PKLR* na população do Rio de Janeiro sem ajuste (A), ajustando para sexo e etnia (B) e ajustando para sexo e ancestralidade (C).

Em seguida, incluímos o rs1052176 no estudo, pois esse SNP apresentou-se como um forte candidato a SNP funcional e está localizado no éxon 12 do gene. A partir de então, utilizamos

somente a ancestralidade genômica e o sexo para o ajuste da regressão. Todos os SNPs apresentam frequências genotípicas em acordo com o equilíbrio de Hardy-Weimberg (HWE) em todas as populações de sadios estudadas. A qualidade da genotipagem geral dos SNPs está acima de 90% em todos os bancos de dados populacionais.

A Tabela 1.1 representa a associação dos SNPs testados na população do Rio de Janeiro, revelando associação de suscetibilidade à hanseníase para os genótipos CG (OR=1,42, $p=0,02$), GG (OR=1,70, $p=0,01$), o alelo G (OR=1,33, $p=0,05$) e os carreadores G (OR=1,48, $p=0,008$) do rs11264355; o genótipo TT (OR=1,23, $p=0,01$) e alelo T (OR=1,32, $p=0,05$) do rs1052176; o genótipo GG (OR=1,62, $p=0,01$) do rs4620533 e o genótipo GG (OR=1.69, $p=0.008$) e carreador G (OR=1.33, $p=0,03$) do rs11264359. O rs4971072, por sua vez, manteve a direção de risco para o genótipo GG, apresentando OR de 1,36, porém com p valor sugestivo de associação ($p=0.07$) após a correção por ancestralidade.

Tabela 1.1 – Frequência dos genótipos, alelos e carreadores dos SNPs do gene *PKLR* em casos e controles e regressão logística para associação com a hanseníase ou tuberculose.

| Desfecho | | Hanseníase | | | | | | | | | | | | | Tuberculose | | | |
|------------------|----------------|-------------------------------|---------------|--|------------------------------|---------------|---------------------------------|---------------|---------------------------------|----------------------------------|---------------|---------------------------------|----------------------------|---------------|---------------------------------|---------------------------|---------------|---|
| SNPs | | POP 1 teste Rio de Janeiro | | | POP 3 replicação Salvador | | | | | POP 4 replicação Rondonópolis | | | POP 5 replicação Manaus | | | POP 2 teste Moçambique | | |
| | | CO | CA | OR (IC) p^1 | CO | CA | OR (IC) p^2 | CO* | OR (IC) p^1 | CO | CA | OR (IC) p^1 | CO | CA | OR (IC) p^2 | CO | CA | OR (IC) p^3 |
| rs11264355 (C>G) | CC | 476 | 421 | | 281 | 195 | | | | 344 | 390 | | 260 | 127 | | 304 | 103 | |
| | | 210 (0.44) | 137 (0.33) | ref | 86 (0.31) | 61 (0.31) | ref | - | - | 145 (0.42) | 175 (0.45) | ref | 86 (0.33) | 46 (0.36) | ref | 70 (0.23) | 11 (0.11) | ref |
| | CG | 213 (0.45) | 204 (0.48) | 1.42 (1.04- 1.93) 0.02 | 146 (0.52) | 84 (0.43) | 0.88 (0.56- 1.40) 0.61 | - | - | 153 (0.44) | 157 (0.40) | 0.86 (0.62- 1.19) 0.39 | 121 (0.47) | 54 (0.43) | 1.30 (0.71- 2.38) 0.39 | 135 (0.44) | 50 (0.49) | 2.19 (1.04- 4.59) 0.03 |
| | | 53 (0.11) | 80 (0.19) | 1.70 (1.09- 2.65) 0.01 | 49 (0.17) | 50 (0.26) | 1.57 (0.90- 2.73) 0.10 | - | - | 46 (0.13) | 58 (0.15) | 0.98 (0.62- 1.55) 0.95 | 53 (0.20) | 27 (0.21) | 1.84 (0.89- 3.82) 0.09 | 99 (0.33) | 42 (0.41) | 2.27 (1.06- 4.84) 0.03 |
| | C | 633 (0.66) | 478 (0.57) | ref | 318 (0.57) | 206 (0.53) | ref | - | - | 443 (0.64) | 507 (0.65) | ref | 293 (0.56) | 146 (0.57) | ref | 275 (0.45) | 72 (0.35) | ref |
| | G | 319 (0.34) | 364 (0.43) | 1.33 (0.99- 1.78) 0.05 | 244 (0.43) | 184 (0.47) | 1.22 (0.82- 1.81) 0.31 | - | - | 245 (0.36) | 273 (0.35) | 0.95 (0.70- 1.30) 0.78 | 227 (0.44) | 108 (0.43) | 1.52 (0.65- 3.58) 0.32 | 333 (0.55) | 134 (0.65) | 1.39 (0.86- 2.27) 0.17 |
| | G carreador | 266 (0.56) | 284 (0.67) | 1.48 (1.10- 1.98) 0.008 | 195 (0.69) | 134 (0.69) | 1.05 (0.69- 1.61) 0.79 | - | - | 199 (0.57) | 215 (0.55) | 0.66 (0.66- 1.21) 0.47 | 174 (0.67) | 81 (0.64) | 1.44 (0.82- 2.55) 0.19 | 234 (0.77) | 92 (0.90) | 2.22 (1.10- 4.49) 0.02 |
| rs1052176 (G>T) | GG | 564 | 690 | | 262 | 193 | | 1246 | | 337 | 340 | | 506 | 307 | | 357 | 104 | |
| | | 253 (0.45) | 244 (0.35) | ref | 98 (0.37) | 65 (0.34) | ref | 375 (0.30) | ref | 140 (0.42) | 162 (0.48) | ref | 147 (0.29) | 102 (0.33) | ref | 73 (0.20) | 11 (0.11) | ref |
| | GT | 254 (0.45) | 314 (0.46) | 1.13 (0.87- 1.46) 0.33 | 118 (0.45) | 79 (0.41) | 1.04 (0.66- 1.63) 0.85 | 615 (0.49) | 1.20 (0.72- 2.01) 0.46 | 145 (0.43) | 128 (0.38) | 0.79 (0.57- 1.11) 0.18 | 243 (0.48) | 146 (0.48) | 1.08 (0.73- 1.60) 0.68 | 168 (0.47) | 42 (0.40) | 1.86 (0.89- 3.86) 0.09 |

| | | | | | | | | | | | | | | | | | | |
|-----------------|--------------------|---------------|---------------|---|---------------|---------------|---|----------------|---|---------------|---------------|---------------------------------|----------------|---------------|---------------------------------|---------------|---------------|---|
| | TT | 57 (0.10) | 132 (0.19) | 1.63 (1.11- 2.40) 0.01 | 46 (0.18) | 49 (0.25) | 1.70 (0.98- 2.95) 0.05 | 256 (0.21) | 2.00 (1.09- 3.65) 0.02 | 52 (0.15) | 50 (0.15) | 0.51 (0.51- 1.28) 0.38 | 116 (0.23) | 59 (0.19) | 1.09 (0.68- 1.76) 0.70 | 116 (0.32) | 51 (0.49) | 2.11 (1.00- 4.47) 0.04 |
| | G | 760 (0.67) | 802 (0.58) | ref | 314 (0.60) | 209 (0.54) | ref | 1365 (0.55) | ref | 425 (0.63) | 452 (0.66) | ref | 537 (0.53) | 350 (0.57) | ref | 314 (0.44) | 73 (0.35) | ref |
| | T | 368 (0.33) | 578 (0.42) | 1.32 (99- 1.77) 0.05 | 210 (0.40) | 177 (0.46) | 1.30 (0.87- 1.95) 0.19 | 1127 (0.45) | 1.41 (0.87- 2.27) 0.15 | 249 (0.37) | 228 (0.34) | 0.86 (0.62- 1.19) 0.37 | 475 (0.47) | 264 (0.43) | 0.95 (0.58- 1.55) 0.86 | 400 (0.56) | 135 (0.65) | 1.36 (0.85- 2.18) 0.19 |
| | T carread or | 311 (0.55) | 446 (0.65) | 1.23 (0.96- 1.57) 0.09 | 164 (0.63) | 128 (0.66) | 1.22 (0.80- 1.85) 0.34 | 871 (0.70) | 1.39 (0.86- 2.23) 0.17 | 197 (0.58) | 178 (0.53) | 0.80 (0.58- 1.09) 0.16 | 359 (0.71) | 205 (0.67) | 1.08 (0.75- 1.57) 0.65 | 284 (0.79) | 93 (0.89) | 1.96 (0.98- 3.95) 0.05 |
| rs4620533 (C>G) | CC | 565 | 430 | | 261 | 198 | | | | 343 | 398 | | 966 | 336 | | 325 | 102 | |
| | | 227 (0.40) | 132 (0.31) | ref | 69 (0.26) | 52 (0.26) | ref | - | - | 128 (0.37) | 163 (0.41) | ref | 270 (0.28) | 103 (0.31) | ref | 25 (0.08) | 3 (0.03) | ref |
| | CG | 268 (0.47) | 199 (0.46) | 1.15 (0.85- 1.55) 0.35 | 136 (0.52) | 87 (0.44) | 1.42 (0.95- 2.12) 0.08 | - | - | 163 (0.48) | 164 (0.41) | 0.80 (0.57- 1.11) 0.18 | 470 (0.49) | 170 (0.51) | 1.04 (0.73- 1.48) 0.79 | 130 (0.40) | 45 (0.44) | 2.41 (0.67- 8.70) 0.17 |
| | GG | 70 (0.12) | 99 (0.23) | 1.62 (1.08- 2.44) 0.01 | 56 (0.21) | 59 (0.30) | 1.31 (0.81- 2.12) 0.25 | - | - | 52 (0.15) | 71 (0.18) | 0.99 (0.63- 1.54) 0.97 | 226 (0.23) | 63 (0.19) | 0.98 (0.57- 1.23) 0.94 | 170 (0.52) | 54 (0.53) | 2.25 (0.63- 8.03) 0.21 |
| | C | 722 (0.64) | 463 (0.54) | ref | 274 (0.52) | 191 (0.48) | ref | - | - | 419 (0.61) | 490 (0.62) | ref | 1010 (0.52) | 376 (0.56) | ref | 180 (0.28) | 51 (0.25) | ref |
| | G | 408 (0.36) | 397 (0.46) | 1.25 (0.95- 1.65) 0.11 | 248 (0.48) | 205 (0.52) | 1.22 (0.82- 1.81) 0.31 | - | - | 267 (0.39) | 306 (0.38) | 0.95 (0.70- 1.28) 0.74 | 922 (0.48) | 296 (0.44) | 0.90 (0.53- 1.54) 0.71 | 470 (0.72) | 153 (0.75) | 1.11 (0.65- 1.89) 0.68 |
| | G carread or | 334 (0.59) | 293 (0.69) | 1.25 (0.94- 1.67) 0.11 | 192 (0.73) | 146 (0.74) | 1.04 (0.66- 1.64) 0.84 | - | - | 215 (0.63) | 205 (0.59) | 0.84 (0.62- 1.15) 0.29 | 696 (0.72) | 233 (0.70) | 1.02 (0.73- 1.43) 0.86 | 300 (0.92) | 99 (0.97) | 2.32 (0.66- 8.14) 0.18 |

| | | | | | | | | | | | | | | | | | | |
|------------------|--------------------|---------------|---------------|--|---------------|---------------|---------------------------------|---------------|---|---------------|---------------|---------------------------------|----------------|---------------|---------------------------------|---------------|---------------|----------------------------------|
| rs4971072 (A>G) | AA | 570 | 754 | | 267 | 195 | | | 338 | 380 | | 967 | 336 | | 347 | 103 | | |
| | | 181 (0.32) | 170 (0.23) | ref | 45 (0.17) | 30 (0.15) | ref | - | - | 99 (0.29) | 109 (0.29) | ref | 196 (0.20) | 78 (0.23) | ref | 3 (0.01) | 0 (0.00) | - |
| | AG | 283 (0.50) | 343 (0.45) | 1.11 (0.84- 1.46) 0.45 | 121 (0.45) | 80 (0.41) | 1.03 (0.57- 1.84) 0.90 | - | - | 166 (0.49) | 170 (0.45) | 0.92 (0.65- 1.32) 0.67 | 488 (0.50) | 180 (0.54) | 0.86 (0.60- 1.22) 0.40 | 6 (0.02) | 0 (0.00) | - |
| | | 106 (0.19) | 241 (0.32) | 1.36 (0.96- 1.92) 0.07 | 101 (0.38) | 85 (0.44) | 1.36 (0.76- 2.45) 0.29 | - | - | 73 (0.22) | 101 (0.27) | 1.15 (0.76- 1.76) 0.49 | 283 (0.29) | 78 (0.23) | 0.73 (0.48- 1.09) 0.12 | 338 (0.97) | 103 (1.00) | - |
| | A | 645 (0.57) | 683 (0.45) | ref | 211 (0.40) | 140 (0.36) | - | - | - | 364 (0.54) | 388 (0.51) | - | 880 (0.46) | 336 (0.50) | - | 12 (0.02) | 0 (0.00) | - |
| | G | 495 (0.43) | 825 (0.55) | 1.16 (0.91- 1.43) 0.22 | 323 (0.60) | 250 (0.64) | 1.21 (0.80- 1.83) 0.34 | - | - | 312 (0.46) | 372 (0.49) | 1.07 (0.79- 1.44) 0.65 | 1054 (0.54) | 336 (0.50) | 0.84 (0.49- 1.42) 0.51 | 682 (0.98) | 206 (1.00) | - |
| | G carread or | 389 (0.69) | 584 (0.77) | 1.17 (0.90- 1.53) 0.22 | 222 (0.83) | 165 (0.85) | 1.18 (0.68- 2.03) 0.54 | - | - | 239 (0.71) | 271 (0.72) | 0.99 (0.71- 1.38) 0.98 | 771 (0.79) | 258 (0.77) | 0.88 (0.61- 1.27) 0.50 | 344 (0.99) | 103 (1.00) | - |
| rs11264359 (A>G) | AA | 495 | 685 | | 213 | 186 | | 1246 | 344 | 388 | | 948 | 287 | | 326 | 104 | | |
| | | 205 (0.41) | 202 (0.29) | ref | 52 (0.24) | 39 (0.21) | ref | 272 (0.22) | ref | 131 (0.38) | 154 (0.37) | ref | 256 (0.27) | 91 (0.32) | ref | 21 (0.06) | 2 (0.02) | ref |
| | AG | 232 (0.47) | 323 (0.47) | 1.23 (0.94- 1.63) 0.12 | 99 (0.46) | 86 (0.46) | 1.26 (0.73- 2.19) 0.39 | 592 (0.48) | 1.39 (0.77- 2.51) 0.26 | 163 (0.47) | 170 (0.44) | 0.96 (0.69- 1.33) 0.82 | 462 (0.49) | 145 (0.51) | 1.02 (0.70- 1.47) 0.91 | 99 (0.30) | 26 (0.25) | 2.74 (0.57- 13.03) 0.20 |
| | | 58 (0.12) | 160 (0.23) | 1.69 (1.14- 2.50) 0.008 | 62 (0.29) | 61 (0.33) | 1.51 (0.83- 2.73) 0.17 | 382 (0.31) | 2.03 (1.07- 3.86) 0.02 | 50 (0.15) | 73 (0.19) | 1.24 (0.80- 1.94) 0.32 | 230 (0.24) | 51 (0.18) | 0.85 (0.54- 1.35) 0.51 | 206 (0.63) | 76 (0.73) | 4.08 (0.89- 18.61) 0.06 |
| A | 642 (0.65) | 727 (0.53) | ref | 203 (0.48) | 164 (0.44) | ref | 1136 (0.46) | ref | 425 (0.62) | 460 (0.59) | ref | 974 (0.51) | 327 (0.57) | ref | 141 (0.22) | 30 (0.14) | ref | |

| | | | | | | | | | | | | | | | | | | |
|---|----------------|---------------|---------------|---|---------------|---------------|---------------------------------|----------------|---------------------------------|---------------|---------------|---------------------------------|---------------|---------------|---------------------------------|---------------|---------------|----------------------------------|
| G | | 348 (0.35) | 643 (0.47) | 1.27 (1.65- 1.98) 0.06 | 223 (0.52) | 208 (0.56) | 1.24 (0.81- 1.90) 0.31 | 1356 (0.54) | 1.39 (0.89- 2.17) 0.13 | 263 (0.38) | 316 (0.41) | 1.08 (0.80- 1.47) 0.58 | 922 (0.49) | 247 (0.43) | 0.84 (0.48- 1.47) 0.55 | 511 (0.78) | 178 (0.86) | 1.70 (0.90- 3.22) 0.09 |
| | G carreador | 290 (0.59) | 483 (0.70) | 1.33 (1.02- 1.73) 0.03 | 161 (0.75) | 147 (0.79) | 1.35 (0.81- 2.26) 0.24 | 974 (0.79) | 1.59 (0.91- 2.78) 0.09 | 213 (0.62) | 243 (0.63) | 1.03 (0.75- 1.40) 0.84 | 692 (0.73) | 196 (0.69) | 0.96 (0.68- 1.37) 0.86 | 305 (0.93) | 102 (0.98) | 3.62 (0.80- 16.38) 0.09 |

O dado está apresentado em frequência. As populações em vermelho representam as populações alvo deste estudo. *Controles do Epigen (1246 indivíduos). Valores em negrito indicam os resultados significativos para $p < 0.05$. Valores não disponíveis (-). Os valores de OR foram calculados baseados nos genótipos ou alelos de referência (ref). Legenda: CO = Controles; CA = Casos; IC = Intervalo de confiança (95%).

¹ OR (*Odds Ratio*) e *P*-valor ajustado para as covariáveis sexo e ancestralidade.

² OR (*Odds Ratio*) e *P*-valor ajustado para as covariáveis sexo, idade e ancestralidade.

³ OR (*Odds Ratio*) e *P*-valor ajustado para sexo e idade.

1.2 – Replicação do estudo de associação do gene *PKLR* com a hanseníase em outras populações brasileiras

A força de associação de um gene com a doença é confirmada através da replicação do estudo em populações independentes. Com o intuito de verificar se a hipótese de associação do gene *PKLR* com a hanseníase no Rio de Janeiro é consistente, replicamos o estudo em outras populações brasileiras. A população de Manaus foi utilizada como representativa da região Norte brasileira, portanto não consideramos a associação genética do gene *PKLR* em Rio Branco no presente estudo. A Tabela 1.1 também representa a associação entre as populações de replicação (POP 3, POP 4 e POP 5). A partir desse ponto, o que denominados de alelos, genótipos e carreadores de risco referem-se à associação genética dos SNPs identificados na população do Rio de Janeiro. Em Rondonópolis e Manaus, populações do Centro-oeste e Norte, respectivamente, não foi visto uma diferença expressiva na frequência dos genótipos, alelos ou carreadores entre casos e controles para os SNPs estudados. Curiosamente, de forma inversa ao observado no Rio de Janeiro, em Manaus a frequência dos genótipos de risco TT-rs1052176 e GG-rs4620533 (0,23 em controles e 0,19 em casos), GG-rs4971072 (0,29 em controles e 0,23 em casos) e GG-rs11264359 (0,24 em controles e 0,18 em casos) apresentou-se aumentada no grupo de controles em relação aos pacientes. Em Rondonópolis, um leve aumento na frequência dos genótipos de risco GG-rs4620533 (0,15 em controles e 0,19 em casos), GG-rs4971072 (0,22 em controles e 0,27 em casos) e GG-rs11264359 (0,15 em controles e 0,19 em casos) foi observado nos casos, porém não houve associação significativa. Assim, nas populações de Rondonópolis e Manaus, o gene *PKLR* parece não ter relevância com a suscetibilidade à hanseníase.

Na população de Salvador, o cenário é parecido com o do Rio de Janeiro. O alelo G do rs11264355, rs4620533, rs4971072 e rs11264359 bem como o alelo T do rs1052176 são mais frequentes no grupo de casos comparado aos controles. O alelo T (rs1052176), por exemplo, que foi associado ao risco no Rio de Janeiro, ocorre com frequência de 0,46 em casos e de 0,40 em controles. Porém, após o ajuste por ancestralidade, sexo e idade, o único SNP que sustentou a associação de suscetibilidade com a hanseníase foi o rs1052176, apresentando OR de 1,70 e p valor *borderline* de 0,05. Assim, em Salvador, observamos uma associação com o risco de desenvolver a hanseníase para o rs1052176. Em seguida, uma outra estratégia foi utilizada para

melhorar a compreensão da relação do gene com a doença na população de Salvador. Os indivíduos sadios da coorte de Salvador do projeto Epigen, em que dois SNPs (rs1052176 e rs11264359) do *PKLR* foram genotipados previamente, foram utilizados como controles no lugar dos controles originais de Salvador. Com isso, a frequência do genótipo TT-rs1052176 permaneceu aumentada em casos (0,25 em casos e 0,21 em controles), e esse SNP foi associado significativamente com o risco à hanseníase (OR=2,00, $p=0,02$). O rs11264359 também apresentou um aumento do genótipo GG nos casos (0,33 em casos e 0,31 em controles), suficiente para levar à associação significativa com a hanseníase (OR=2,03, $p=0,02$). A partir desses dados, concluímos que a associação do gene *PKLR* com a suscetibilidade à hanseníase é mantida em Salvador, mas não em Rondonópolis e Manaus.

1.3 – Associação dos SNPs do gene *PKLR* com a suscetibilidade à tuberculose em Moçambique

A associação de suscetibilidade do gene *PKLR* com a tuberculose para os SNPs rs11264355, rs4620533, rs4971072 e rs11264359 foram gerados anteriormente (293). Dado que estamos avaliando a hipótese de suscetibilidade do gene *PKLR* em hanseníase, utilizamos a tuberculose para testar se a associação do gene é relevante também em outro modelo de doença micobacteriana. Para isso, utilizamos uma coorte de indivíduos africanos de Moçambique. A novidade apresentada por este estudo nessa população é a associação do rs1052176 com a tuberculose, em que a reanálise com os dados prévios pode ser vista na Tabela 1.1. Assim como para os genótipos CG (OR=2,19, $p=0,03$), genótipo GG (OR=2,27, $p=0,03$) e carreador de G (OR=2,22, $p=0,02$) do rs11264355 que foram significativamente associados com a suscetibilidade à hanseníase, o rs1052176 apresentou associação para o genótipo TT, com OR de 2,11 e p valor de 0,04. O rs4620533 não foi associado após a correção por sexo e idade em Moçambique, porém o rs11264359 apresentou uma associação sugestiva de risco com p valor 0,06 para o genótipo GG (OR=4,08, $p=0,06$) com a tuberculose. Portanto, o gene *PKLR* apresenta-se associado também com a tuberculose em Moçambique pelos SNPs rs11264355, rs1052176 e rs11264359.

As análises de associação descritas para cada população do estudo podem ser sumarizadas na Figura 1.2, que permite a visualização dos valores de OR referente aos genótipos de cada SNP associados entre as populações.

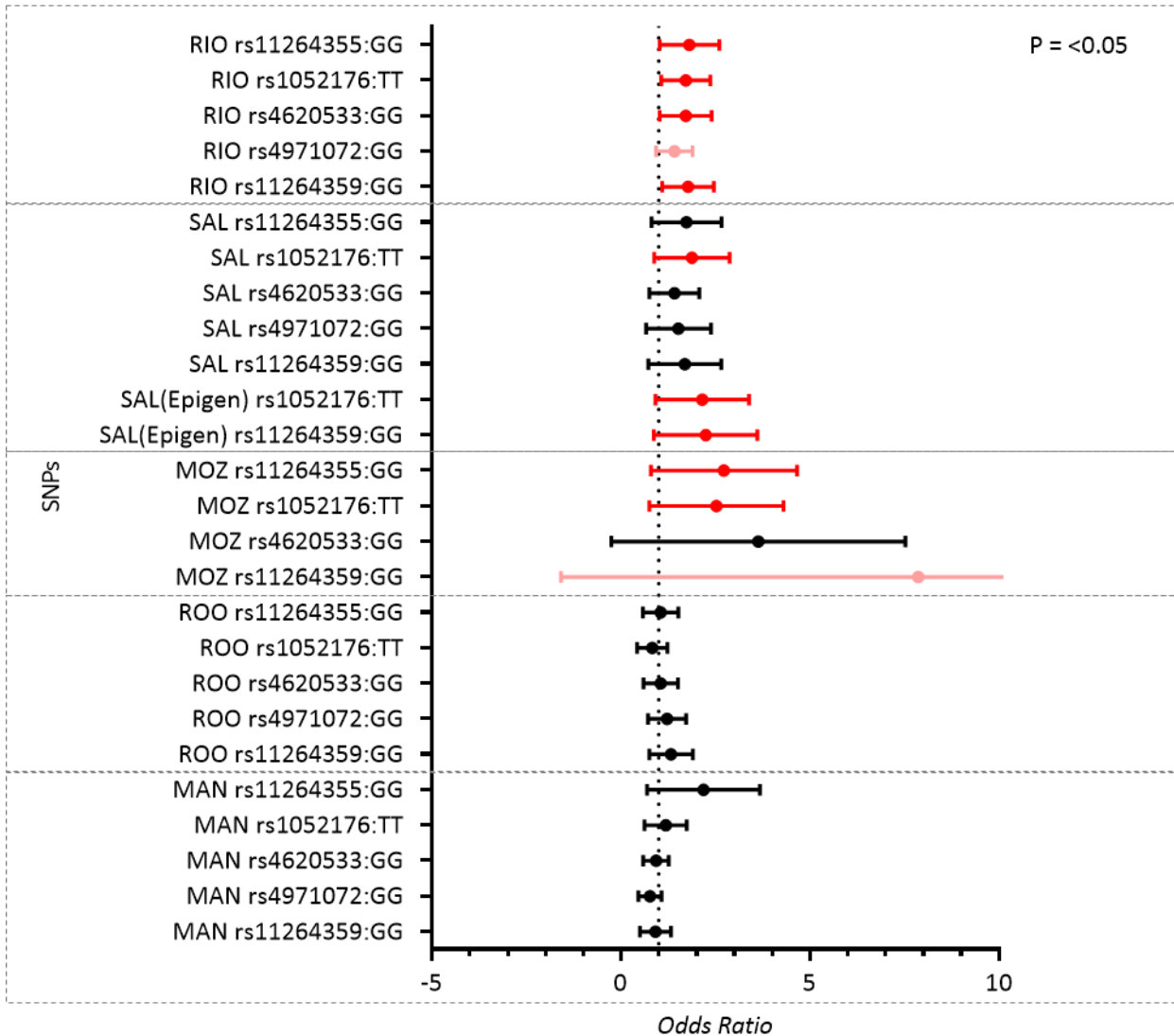


Figura 1.2 – Valores de *Odds ratios* (OR) e intervalos de confiança (95%) para a associação entre os polimorfismos do gene *PKLR* com a hanseníase e tuberculose. A associação com a hanseníase ocorreu nas populações brasileiras do Rio de Janeiro (RIO), Salvador (SAL), Rondonópolis (ROO) e Manaus (MAN) e associação com a tuberculose ocorreu em Moçambique (MOZ). Os valores de OR estão representados pelos pontos e as retas são os intervalos de confiança (95%). Em vermelho e rosa estão destacados os valores de OR com *p* valores significativos ($p < 0.05$) e *borderline*, respectivamente.

1.4 – Análise de desequilíbrio de ligação (LD) dos polimorfismos do gene *PKLR*

Para investigar o grau de segregação entre as variantes do *PKLR* associadas com a Hanseníase entre as diferentes populações brasileiras, avaliamos o perfil de LD entre elas. O esperado ao acaso é que as variantes genéticas tenham segregação independente, de forma aleatória (ou em equilíbrio de ligação). A análise de LD nos retorna se existem blocos haplotípicos formados por SNPs que estão segregando conjuntamente, ou seja, em desequilíbrio de ligação. Na Figura 1.3, é possível observar que a arquitetura genética entre as populações brasileiras é diferenciada. Em todas as populações, com exceção a Moçambique, é visto que os SNPs rs11264355, rs1052176 e rs4620533 formam um bloco em moderado a alto LD (considerando alto LD para $r^2 > 70$). De fato, comparando entre europeus e africanos do 1000Genomes, o bloco em LD existe em ambas as populações. Para as análises subsequentes, selecionamos como “tag SNP” desse bloco em LD o rs1052176 devido à sua posição em região codificante e à associação genética observada no Rio de Janeiro, Salvador e Moçambique.

Além disso, vimos que a população do Rio de Janeiro (A) tem maior semelhança com o perfil genético de Salvador (B), enquanto Rondonópolis (C) e Manaus (D) apresentam uma região genômica similar com um elevado grau de LD entre as variantes, próximo ao esperado em populações Europeias (F). Moçambique (E), por sua vez, apresenta maior número de variantes segregando de forma independente, como é esperado para populações africanas (F). Apenas os SNPs rs11264355 e rs1052176 estão em LD em casos e controles nessa última população. As populações de Bambuí, Pelotas e Salvador (G) do Epigen apresentam dois SNPs em LD completo com o rs1052176 (são eles o rs932972 e rs1052177, que se encontravam disponíveis no banco de dados). Na coorte de Salvador do Epigen os SNPs rs1052176, rs497107 e rs11264359 tem menores valores de r^2 (menor LD) do que a população de Salvador deste estudo, mais próximo do observado em africanos (1000Genomes).

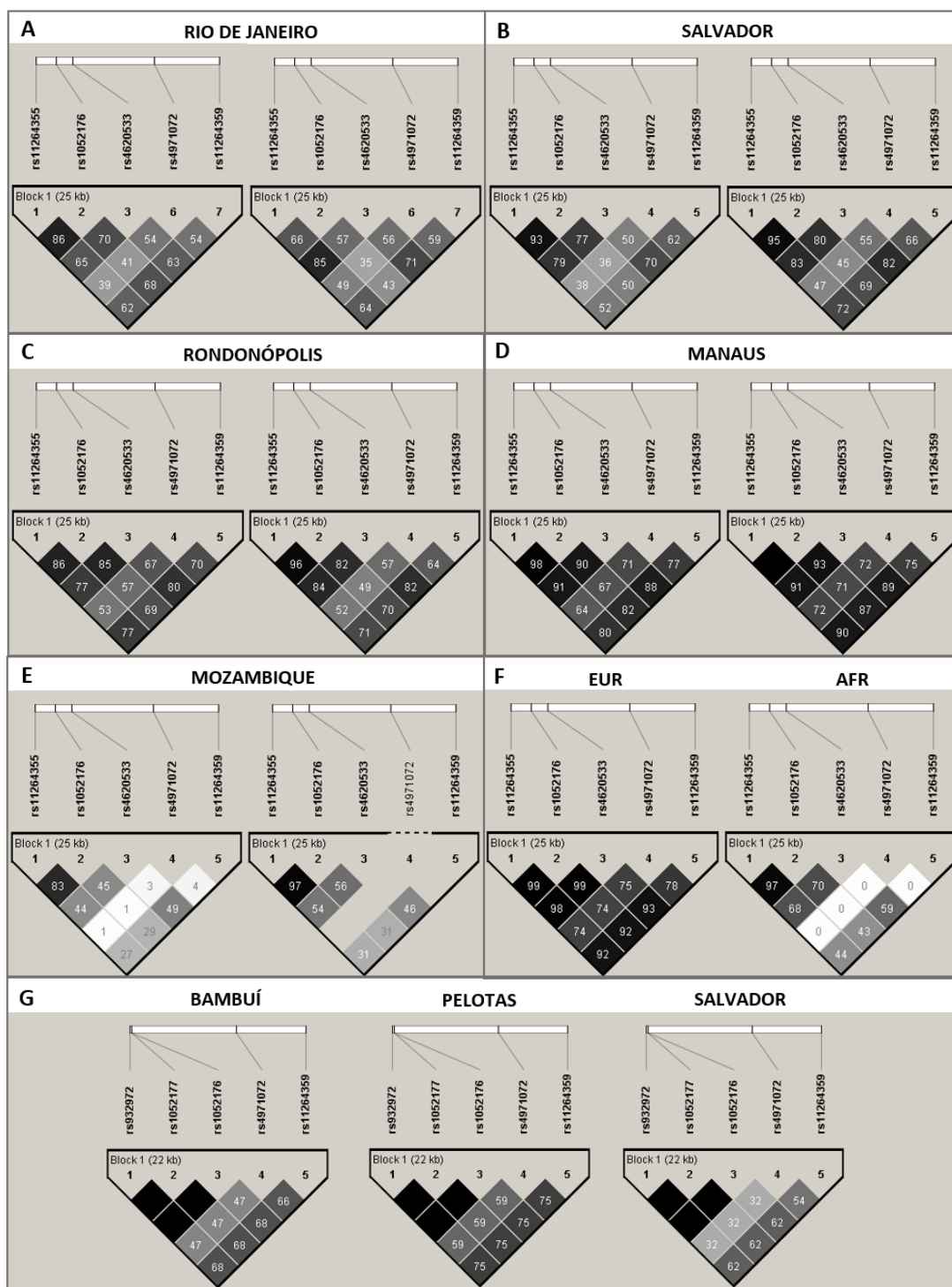


Figura 1.3 – Desequilíbrio de ligação entre os SNPs da região do *PKLR* entre as populações do estudo. O primeiro e o segundo gráfico em cada população representa o LD em controles e pacientes, respectivamente (exceto em F e G onde estão representadas as populações de indivíduos saudáveis do 1000Genomes e Epigen, respectivamente). Os SNPs estão representados de 1 a 5. Os losangos apresentam o valor de r^2 uma escala de branco a preto: branco – LD fraco ($r^2 < 0.5$); cinza – LD moderado ($0.5 < r^2 < 0.7$) e

preto – LD forte ($r^2 > 0,70$). Em G, foram plotados os SNPs do *PKLR* disponíveis nas populações do projeto Epigen (o rs11264355 e rs4620533 não foram genotipados nas coortes desse projeto).

1.4 – O haplótipo T/G/G e a associação com a suscetibilidade a hanseníase e tuberculose

Considerando o padrão de LD, os haplótipos estudados foram formados pelos SNPs rs1052176, rs4971072 e rs11264359. Na Tabela 1.2 observa-se que o haplótipo G/A/A tem frequência de 0,66 em europeus e praticamente não existe em africanos. Já o haplótipo T/G/G é mais frequente em africanos (0,47) do que em europeus (0,27) do 1000Genomes. Nas coortes do projeto Epigen, o haplótipo T/G/G também é mais frequente na população com maior ascendência africana, Salvador (0,45), em comparação com Pelotas e Bambuí (populações com elevada ancestralidade europeia), enquanto o haplótipo G/A/A tem frequência de 0,60 em Pelotas e Bambuí e 0,29 em Salvador.

No Rio de Janeiro, o haplótipo T/G/G tem frequência de 0,36 em casos e 0,28 em controles, e está significativamente associado com a suscetibilidade à hanseníase nessa população (OR=1,26 e $p=0,02$) após as correções por sexo e ancestralidade. A partir de então, denominamos o haplótipo T/G/G de “haplótipo de risco”. Observa-se que o haplótipo G/G/G (OR=1,88 e $p=0,0002$) também está associado significativamente ao risco nessa população, porém a frequência desse haplótipo é menor que 10% em casos (0,09) e controles (0,04), e, portanto, tem baixa ocorrência no nível populacional. A replicação em Salvador mostra que o haplótipo de risco também está aumentado em casos (0,45) em relação aos controles (0,37), e nessa população observamos que há uma sugestão de associação de suscetibilidade à hanseníase com p valor *borderline* (OR=1,30 e $p=0,07$). Em Moçambique também é visto associação de suscetibilidade para o haplótipo de risco com p valor *borderline* (OR=1,52 e $p=0,07$) com a tuberculose, sendo que a frequência desse haplótipo é de 0,64 em casos de TB e 0,54 em controles. Já em Rondonópolis e Manaus nenhuma associação significativa foi identificada para haplótipo e a hanseníase. Nessas populações, o haplótipo de risco não é predominante em casos. A partir desses dados, concluímos que o haplótipo T/G/G está associado com o desenvolvimento da hanseníase no Rio de Janeiro e Salvador, e com a TB em Moçambique. Destaca-se que o Rio de Janeiro e Salvador são as populações brasileiras que apresentam a maior ancestralidade africana,

e é provável que populações que não tenham grande contribuição africana não carreguem os “marcadores africanos”, e, portanto, não apresentem associação com as doenças micobacterianas.

Tabela 1.2 – Frequência de haplótipos do gene *PKLR* nas populações do estudo e associação com a hanseníase e TB.

| rs1052176 | rs4971072 | rs11264359 | AFR | EUR | BAM | PEL | SAL ¹ | RIO CA | RIO CO | OR <i>p</i> ^A | SAL CA | SAL CO | OR <i>p</i> ^B | MOZ CA | MOZ CO | OR <i>p</i> ^B | ROO CA | ROO CO | OR <i>p</i> ^A | MAN CA | MAN CO | OR <i>p</i> ^C |
|-----------|-----------|------------|------|------|------|------|------------------|-----------|-----------|-----------------------------|-----------|-----------|-----------------------------|-----------|-----------|-----------------------------|-----------|-----------|-----------------------------|-----------|-----------|-----------------------------|
| | | | | | | | | | | | | | | | | | | | | | | |
| T | G | G | 0.47 | 0.27 | 0.24 | 0.28 | 0.45 | 0.36 | 0.28 | 1.26 0.02 | 0.45 | 0.37 | 1.30 0.07 | 0.64 | 0.54 | 1.52 0.07 | 0.32 | 0.33 | 1.04 0.71 | 0.41 | 0.45 | 0.98 0.85 |
| G | G | A | 0.32 | 0.04 | 0.08 | 0.05 | 0.13 | 0.07 | 0.09 | 0.78 0.13 | 0.09 | 0.10 | 0.86 0.58 | 0.13 | 0.19 | ref | 0.09 | 0.06 | 1.31 0.20 | 0.05 | 0.05 | 0.77 0.30 |
| G | G | G | 0.19 | | 0.05 | 0.04 | 0.10 | 0.09 | 0.04 | 1.88 0.002 | 0.08 | 0.11 | 0.84 0.52 | 0.20 | 0.23 | 1.26 0.40 | 0.06 | 0.04 | 2.34 0.06 | 0.01 | 0.02 | 0.61 0.34 |
| G | A | A | | 0.66 | 0.60 | 0.60 | 0.29 | 0.37 | 0.52 | ref | 0.35 | 0.37 | ref | | 0.01 | | 0.49 | 0.51 | ref | 0.49 | 0.44 | ref |

O SNP rs1052176 é o “tag” representativo do bloco em LD com os SNPs rs11264355 e rs4620533. Haplótipos com frequência menor que 0.01 (<1%) não foram exibidos. Em negrito estão representados os valores significativos ($p < 0.05$) ou *borderline* ($0.05 < p < 0.07$) para a associação. Legenda: AFR = Africanos (1000Genomes); EUR = Europeus (1000Genomes); BAM = Bambuí (Epigen); PEL = Pelotas (Epigen); SAL¹ = Salvador (Epigen); RIO = Rio de Janeiro; SAL = Salvador; MOZ = Moçambique; ROO = Rondonópolis; MAN = Manaus; CA = Casos e CO = Controles. *Odds ratio* (OR) foi calculada baseado no haplótipo referência (ref).

^A Ajustado para sexo e ancestralidade.

^B Ajustado para sexo e idade.

^C Ajustado para sexo, idade e ancestralidade.

2 – Análises evolutivas

2.1 – As frequências alélicas das variantes do gene *PKLR* diferem entre europeus e africanos

Inicialmente, foi feita uma análise descritiva e exploratória das frequências alélicas dos SNPs do gene *PKLR* nas populações africanas e europeias do 1000Genomes. Para avaliar o quão a proporção entre as frequências é diferente entre as populações utilizamos o teste de Fisher. A Tabela 1.3 representa a frequência dos alelos entre europeus e africanos. Observa-se que os alelos apresentam frequência invertida entre as populações continentais e o teste de Fisher confirmar a diferença significativa dos SNPs entre essas populações ($p < 0.00001$).

Tabela 1.3 – Frequência alélica dos SNPs do gene *PKLR* em europeus e africanos do 1000Genomes.

| SNPs | Europeus | Africanos | Teste de Fisher (p) |
|--------------|----------|-----------|-------------------------|
| rs11264355-C | 0.72 | 0.51 | <0.00001 |
| rs11264355-G | 0.28 | 0.49 | |
| rs1052176-G | 0.72 | 0.51 | <0.00001 |
| rs1052176-T | 0.28 | 0.48 | |
| rs4620533-C | 0.72 | 0.43 | <0.00001 |
| rs4620533-G | 0.28 | 0.56 | |
| rs4971072-A | 0.67 | 0.03 | <0.00001 |
| rs4971072-G | 0.32 | 0.97 | |
| rs11264359-A | 0.72 | 0.67 | <0.00001 |
| rs11264359-G | 0.28 | 0.33 | |

Em vista dos dados de associação, simplificamos as análises subsequentes com foco nos 3 SNPs principais (rs1052176, rs4971072 e rs11264359). A análise do MAF (do inglês, *minor allele frequency*) ajuda a distinguir se as variantes genéticas estudadas são raras ou comuns na população, o que direciona as primeiras pistas acerca do tipo de seleção que está ocorrendo no gene alvo. Seleção positiva leva ao aumento súbito de variantes raras, enquanto a seleção balanceadora leva à manutenção de variantes com frequência intermediária na população. A Figura 1.4 explora o quão a frequência alélica das variantes rs1052176, rs4971072 e rs11264359 diferem das demais no cromossomo 1. Há maior densidade de variantes raras ($MAF < 0.05$) em europeus, enquanto em africanos há maior densidade de variantes comuns, com MAF entre 0,10

e 0,20. No geral, os SNPs de interesse apresentam uma distribuição que caracteriza variantes de frequência intermediária (0,28 a 0,56) em ambas as populações. Com exceção do rs4971072, o MAF dos SNP rs1052176 e rs11264359 é maior em africanos. No caso do rs4971072, o alelo minoritário em africanos corresponde ao alelo A (que tem frequência de 0,03 enquanto o alelo G é o majoritário – 0,97), enquanto em europeus corresponde ao alelo G (0,32).

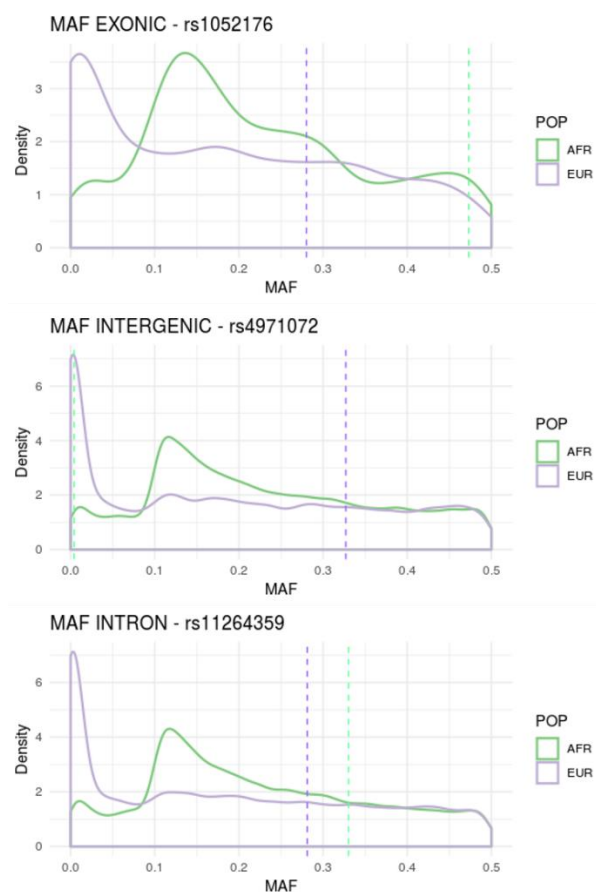


Figura 1.4 – Densidade das variantes genéticas do cromossomo 1 em europeus (EUR) e africanos (AFR). MAF=Frequência do alelo minoritário. A linha pontilhada representa o MAF do SNP de interesse na população europeia (em roxo) e na população africana (em verde).

2.2 – Análise de “ F_{ST} outlier” entre europeus e africanos

O F_{ST} entre pares de população foi calculado para as variantes do cromossomo 1 entre europeus e africanos. Os valores de F_{ST} para os SNPs individuais da região genômica do *PKLR* estão indicadas na Tabela 1.1 em ANEXO II. Na Figura 1.5, selecionamos 60 variantes da região do *PKLR*

que contém os SNPs rs1052176, rs4971072 e rs11264359 e comparamos os valores de F_{ST} individuais entre as variantes da região. O SNP rs1052176 (vermelho) apresentou $F_{ST} = 0,07$ (p valor=0,36) e o rs11264359 (verde) apresentou um alto $F_{ST} = 0,26$ com p valor = 0,08. O SNP rs4971072 (azul), por sua vez, possui o maior F_{ST} desse intervalo, $F_{ST} = 0,66$ e p valor = 0,02, indicando que esse SNP é um *outlier* em relação à distribuição esperada dos valores de F_{ST} referentes às variantes do cromossomo 1. Entretanto, para evitar resultados falso-positivos provocados por alelos que tenham um alto valor de F_{ST} devido à ocorrência da deriva, realizamos a análise de F_{ST} médio por janela de SNPs (20 SNPs com sobreposição de 15).

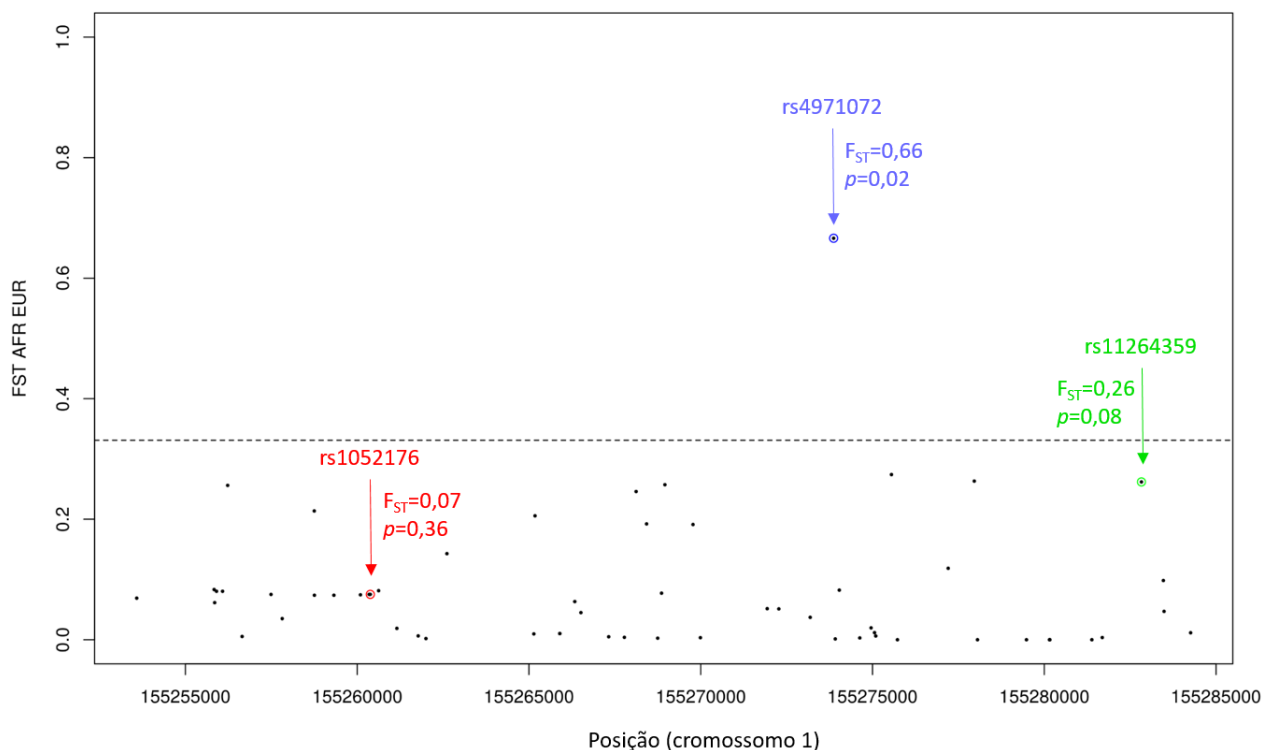


Figura 1.5 – Distribuição dos valores de F_{ST} para as variantes da região do gene *PKLR*. O eixo x representa as posições genômicas e o eixo y os valores de F_{ST} para as variantes entre europeus e africanos. Para a definição do ponto de corte em que consideraríamos o valor de F_{ST} como *outlier*, dividimos a distribuição dos valores de F_{ST} totais do cromossomo 1 por quartis (0,01; 0,05; 0,1; 0,25; 0,50; 0,75; 0,90; 0,95; 0,99) e a linha pontilhada (*threshold*) representa o penúltimo quartil dessa distribuição (95% - $F_{ST}>0,39$). Valores acima desse limiar indicam variantes *outliers*. Cada SNP foi representado por um ponto. O SNP rs1052176 foi destacado em vermelho, rs4971072 em azul e rs11264359 em verde.

Em seguida, para evitar resultados falso-positivos provocados por alelos que tenham um alto valor de F_{ST} devido a deriva genética (e não propriamente a seleção), combinamos as estimativas de F_{ST} para um intervalo de variantes através da análise de F_{ST} médio por janela de SNPs (20 SNPs com sobreposição de 15). A ideia dessa abordagem é que, devido ao desequilíbrio de ligação entre os sítios, o sinal de seleção em um determinado SNP será compartilhado pelos SNPs adjacentes. Desse modo, um sinal verdadeiro também será reproduzido na análise por janelas, enquanto um resultado espúrio não. Entretanto, não foi observada janela *outlier* para os intervalos que contém os SNPs rs1052176 ($F_{ST}=0,10$), rs11264359 ($F_{ST}=0,08$) e rs4971072 ($F_{ST}=0,21$). Essa última, apesar de não ser significativa, ainda é uma janela com valor elevado (Tabela 1.4).

Tabela 1.4 – Valores de F_{ST} para os SNPs individuais e por janelas em europeus e africanos.

| SNP | F_{ST} (SNP individual) | F_{ST} (Janelas) |
|------------|------------------------------|-----------------------|
| rs1052176 | 0,07 | 0,10 |
| rs4971072 | 0,66 | 0,21 |
| rs11264359 | 0,26 | 0,08 |

Janelas de 20 SNPs com sobreposição de 15 SNPs.

2.3 – Cálculo de Tajima D em europeus e africanos

O cálculo de Tajima D foi realizado para janelas de 1000pb no cromossomo 1 em africanos e europeus do 1000Genomes. A Figura 1.6 representa as 100 janelas *upstream* e *downstream* em relação à janela do SNP rs1052176. Observa-se que a maioria das janelas apresentam valores de Tajima D positivos ($D>0$) em ambas as populações nesse intervalo.

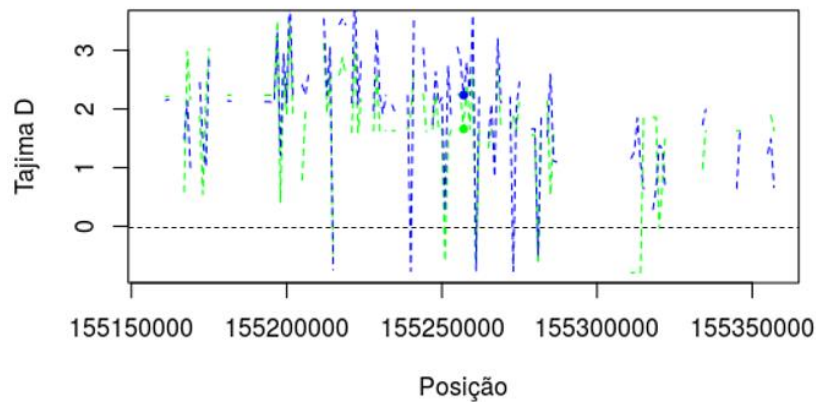


Figura 1.6 – Valores de Tajima D por janelas em africanos (linha azul) e europeus (linha verde). Cada janela em que o Tajima D foi calculado tem 1000pb. A janela do SNP rs1052176 (155260383 – Gh37) é representada pelo ponto azul e verde, nas respectivas populações, e 100 janelas foram representadas à direita e à esquerda dessa posição. Os espaços em branco representam as janelas em que não foi possível calcular o Tajima D devido à ausência de SNPs.

Focando no intervalo de janelas que contém os SNPs de interesse do gene *PKLR*, demonstrado na Tabela 1.5, calculamos o Tajima D considerando a ausência das variantes raras, pois elas podem inflar o valor de Tajima D, ou a presença das variantes raras, pois o valor de Tajima D depende do número de sítios segregantes no seguimento genômico, portanto o menor número de SNPs pode prejudicar o cálculo. É visto que, independentemente de considerar as variantes raras ou as variantes totais da região, os valores de Tajima D são similares entre as janelas em europeus e africanos. Assim, devido à escassez de SNPs por janela quando excluimos as variantes com menos de 5% de frequência ($MAF < 0,05$), consideramos os valores de Tajima D calculado para as variantes totais. A janela 1 (rs11264355), 17 (rs4971072) e 26 (rs11264359) apresentaram-se com sinais de Tajima D diferentes entre europeus e africanos, sugerindo diferentes regimes seletivos para esse seguimento genômico entre as populações. As janelas 4 (rs1052176) e 6 (rs4620533), entretanto, exibiram valores positivos entre europeus e africanos, concordantes com o regime de seleção balanceadora em ambas as populações. Como a maioria das janelas na região são positivas, há indicação de seleção balanceadora, porém não é distinguível a diferença dos regimes entre europeus e africanos.

Tabela 1.5 – Valores de Tajima D calculado por janelas de 1000pb na região do *PKLR*.

| Janela | Posição (Chr1) | MAF<0.05 | | | | Variantes totais | | | |
|-----------------|----------------|----------|--------------|---------|--------------|------------------|--------------|---------|--------------|
| | | Nº SNPs | Tajima D AFR | Nº SNPs | Tajima D EUR | Nº SNPs | Tajima D AFR | Nº SNPs | Tajima D EUR |
| 1 (rs11264355) | 155257000 | 1 | 2.23535 | 1 | 1.65796 | 3 | 0.42125 | 6 | -0.544685 |
| 2 | 155258000 | 2 | 2.79107 | 2 | 2.41415 | 6 | 0.96554 | 3 | 1.51136 |
| 3 | 155259000 | 1 | 2.23336 | 1 | 1.65263 | 3 | 1.11324 | 2 | 0.619624 |
| 4 (rs1052176) | 155260000 | 3 | 3.60451 | 3 | 2.66585 | 9 | 1.51539 | 4 | 1.88417 |
| 5 | 155261000 | 1 | -0.75543 | 1 | -0.601 | 5 | -1.2521 | 4 | -1.20741 |
| 6 (rs4620533) | 155262000 | 1 | 2.20966 | 1 | 1.64189 | 5 | 0.0664 | 3 | 0.0768556 |
| 7 | 155263000 | 0 | - | 0 | - | 6 | -0.9055 | 0 | - |
| 8 | 155264000 | 0 | - | 0 | - | 2 | -0.9483 | 0 | - |
| 9 | 155265000 | 3 | 1.50441 | 2 | 1.34088 | 6 | 0.23685 | 5 | -0.014184 |
| 10 | 155266000 | 2 | 2.0891 | 1 | 1.66327 | 5 | 0.63431 | 3 | 0.107406 |
| 11 | 155267000 | 1 | 0.8348 | 0 | - | 5 | 0.03412 | 2 | -0.997086 |
| 12 | 155268000 | 3 | 3.19904 | 3 | 2.68301 | 10 | 0.85124 | 6 | 1.0032 |
| 13 | 155269000 | 1 | 2.1056 | 1 | 1.66327 | 6 | 0.25775 | 2 | 0.59398 |
| 14 | 155270000 | 0 | - | 0 | - | 2 | -1.0134 | 0 | - |
| 15 | 155271000 | 0 | - | 0 | - | 6 | -1.1968 | 2 | -0.949456 |
| 16 | 155272000 | 1 | 2.19868 | 1 | 1.66327 | 1 | 2.19868 | 2 | 0.585714 |
| 17 (rs4971072) | 155273000 | 1 | -0.76747 | 1 | 1.8771 | 4 | -1.122 | 3 | 0.151368 |
| 18 | 155274000 | 1 | 2.04617 | 1 | 1.63649 | 4 | 0.69031 | 3 | 0.0867963 |
| 19 | 155275000 | 2 | 2.45746 | 2 | 2.19674 | 4 | 0.9823 | 4 | 0.836638 |
| 20 | 155276000 | 0 | - | 0 | - | 4 | -0.6886 | 0 | - |
| 21 | 155277000 | 1 | 1.90357 | 1 | 1.63649 | 9 | -0.9017 | 2 | 1.43531 |
| 22 | 155278000 | 0 | - | 0 | - | 4 | -1.3358 | 2 | -1.0627 |
| 23 | 155279000 | 1 | 1.66057 | 1 | 1.55828 | 4 | 0.01221 | 1 | 1.55828 |
| 24 | 155280000 | 1 | 1.66057 | 1 | 1.55828 | 3 | 0.06755 | 3 | -0.000576 |
| 25 | 155281000 | 1 | -0.48535 | 1 | -0.6481 | 6 | -1.1679 | 3 | -1.17395 |
| 26 (rs11264359) | 155282000 | 1 | 1.89132 | 1 | 1.65796 | 6 | -0.5306 | 1 | 1.65796 |

Representação das janelas que contém os SNPs de interesse referentes ao gene *PKLR*. As análises foram realizadas considerando as variantes totais ou excluindo as variantes raras.

2.4 – Análise da homozigosidade do haplótipo estendido (iHS e xpEHH)

O objetivo do estudo a partir das análises de xpEHH e iHS foi verificar a presença de varredura seletiva na região genômica indicando ocorrência de seleção positiva (valores de xpEHH e iHS acima de 2). Essas análises baseiam-se na hipótese de que a ocorrência de uma forte seleção em um determinado *locus* impulsiona novos alelos até uma frequência intermediária alta, carregando um haplótipo longo de baixa diversidade, sendo essa característica um sinal clássico de forte seleção positiva. Esses dados estão sumarizados na Tabela 1.1 em ANEXO II, onde observamos que nenhum dos valores de iHS tanto em europeus quanto em africanos demonstrou valores indicativos de seleção. Já a análise de xpEHH demonstra que existe um *sweep* seletivo em variantes (~25 SNPs) adjacentes nas posições que compreendem o gene *HCN3*, são elas: rs183293873, rs12044063, rs7367998, rs560609863, rs7520184, rs11264353, rs573791612, rs541165488, rs559902587, rs11264353, rs552421834, rs564241242, rs531296924, rs549806272, rs144733142, rs535535805, rs547542932, rs11264354, rs539661366, rs558388981, rs113065628, rs144035252, rs541106278, rs559445057, rs12724449. Na Figura 1.7 podemos observar o sinal de seleção dessas variantes ao longo da região composta por 1060 polimorfismos. Entretanto, o sinal de seleção parece ocorrer na população europeia (devido aos valores negativos do xpEHH), ao invés da africana, como o esperado.

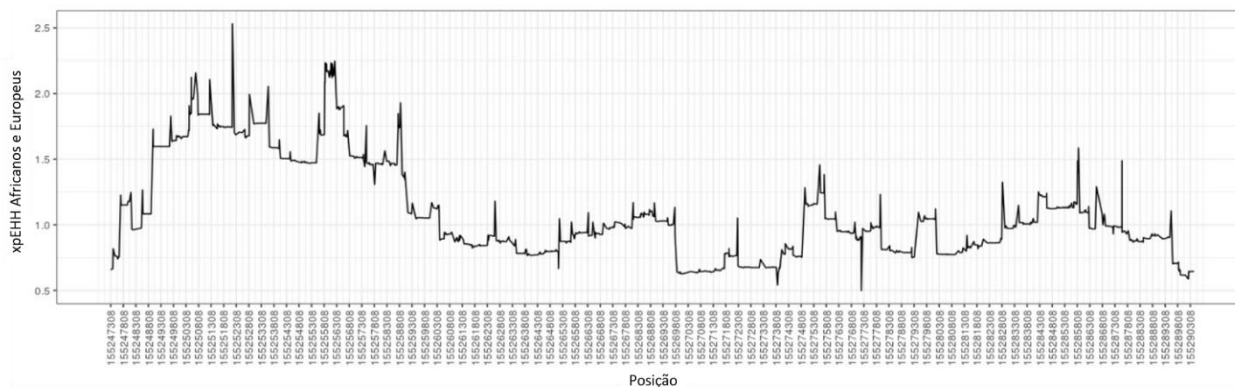


Figura 1.7 – Sweep de seleção no gene *HCN3* detectado pela análise de xpEHH (1:155247308-155290457) entre africanos e europeus. Os valores originais de xpEHH para toda a região são valores negativos. Adicionamos os dados do xpEHH (eixo y) em módulo e em intervalos de 500pb (eixo x) para facilitar a

visualização. O sinal ($x_{pEHH} > |2.00|$) foi visto para 25 SNPs entre 1060 variantes ao longo dos genes *HCN3*, *PKLR* e *FDPS*.

O próximo passo foi averiguar qual seria a relação dessas variantes sujeitas a um possível sinal de seleção com os polimorfismos do gene *PKLR*. Surpreendentemente, observamos que as variantes entre os genes *HCN3* e *PKLR* estão em alto LD em europeus e africanos, como visto na Figura 1.8. Em europeus o bloco é proeminente com muitas variantes em desequilíbrio completo ($r^2 > 95$). Em africanos, um menor padrão de LD é observado, mas o bloco entre as variantes de *sweep* no *HCN3* e o *PKLR* também é visto, principalmente para os rs11264355 e rs1052176 (apresentam os maiores valores de r^2 com as variantes adjacentes).

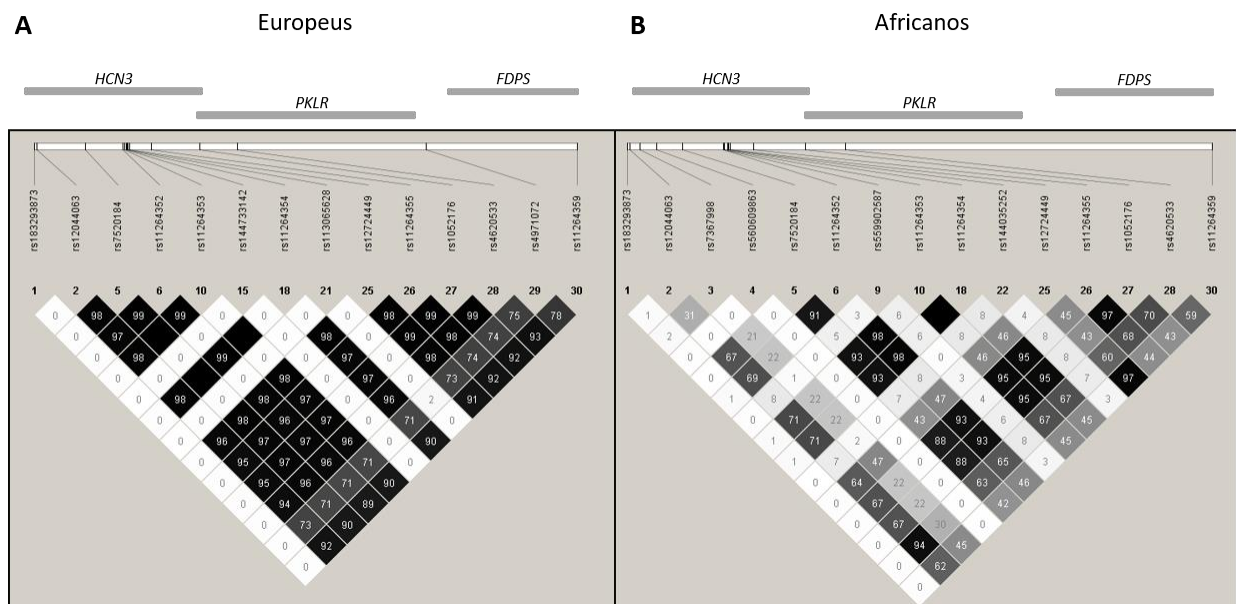


Figura 1.8 – LD entre as variantes dos genes *HCN3*, *PKLR* e *FDPS*. A e B representam o padrão de LD em europeus e africanos, respectivamente. Os blocos (em preto e cinza) configuram uma relação de alto LD entre essas variantes em ambas as populações, com destaque para os europeus. Os losangos apresentam uma escala de branco a preto: branco – LD fraco ($r^2 < 50$); cinza – LD moderado fraco ($50 < r^2 < 70$) e preto – LD forte ($r^2 > 70$).

A Tabela 1.6 mostra, em valores de r^2 , o desequilíbrio de ligação observado, confirmando que os valores de LD são muito elevados para o bloco formado pelos SNPs **rs7520184**, **rs11264352**, **rs11264353**, **rs11264354**, **rs11264355**, **rs1052176**, **rs4620533**, rs12044063, rs12724449 e rs11264359, com protagonismo ($r^2 > 0,88$) para os destacados em negrito.

Tabela 1.6 – LD entre os 25 SNPs com sinal de seleção pelo teste de xpEHH entre Europeus e Africanos e os SNPs associados à hanseniase.

| AFR / EUR (1000Genomes) | | | | | | |
|--------------------------|--------------------|--------------------|--------------------|--------------------|--------------------|------------|
| LD (r^2) | PKLR | | | | | eQTL |
| <i>HCN3</i> | rs11264355 | rs1052176 | rs4620533 | rs4971072 | rs11264359 | |
| rs183293873 ^a | 0.08 | 0.08 | 0.00 | 0.00 | 0.00 | Não |
| rs12044063 | 0.67 / 0.95 | 0.67 / 0.95 | 0.94 / 0.94 | 0.00 / 0.73 | 0.62 / 0.92 | Sim |
| rs7367998 | 0.22 | 0.22 | 0.30 | 0.01 | 0.45 | Não |
| rs560609863 ^a | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | - |
| rs7520184 | 0.88 / 0.97 | 0.88 / 0.97 | 0.63 / 0.96 | 0.00 / 0.71 | 0.42 / 0.90 | Sim |
| rs11264352 | 0.93 / 0.96 | 0.93 / 0.97 | 0.65 / 0.96 | 0.00 / 0.71 | 0.46 / 0.89 | Sim |
| rs573791612 ^a | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | - |
| rs541165488 ^a | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | - |
| rs559902587 ^a | 0.04 | 0.06 | 0.08 | 0.00 | 0.03 | Não |
| rs11264353 | 0.95 / 0.97 | 0.95 / 0.97 | 0.67 / 0.96 | 0.00 / 0.71 | 0.45 / 0.90 | Sim |
| rs552421834 ^a | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | - |
| rs564241242 ^a | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | - |
| rs531296924 ^a | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | - |
| rs549806272 ^a | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | - |
| rs144733142 ^a | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | Não |
| rs535535805 ^a | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | - |
| rs547542932 ^a | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | - |
| rs11264354 | 0.95 / 0.97 | 0.95 / 0.97 | 0.67 / 0.96 | 0.00 / 0.71 | 0.45 / 0.90 | Sim |
| rs539661366 ^a | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | - |
| rs558388981 ^a | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | - |
| rs113065628 ^a | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | Não |
| rs144035252 ^a | 0.08 | 0.08 | 0.07 | 0.00 | 0.03 | Não |
| rs541106278 ^a | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | Não |
| rs559445057 ^a | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | - |
| rs12724449 | 0.45 / 0.98 | 0.43 / 0.99 | 0.60 / 0.98 | 0.00 / 0.73 | 0.97 / 0.91 | Sim |

^aSNP fixado em Europeus (MAF=0.99-1.00). O LD está representado em africanos (AFR)/europeus (EUR). Valores em negrito correspondem aos SNPs que apresentam LD moderado a elevado em Africanos e Europeus do 1000Genomes, formando um bloco de SNPs em desequilíbrio com os SNPs associados à hanseniase. A coluna do eQTL representa os SNPs que regulam a expressão do gene *PKLR* descritos no GTEx.

Em seguida, exploramos o dado do decaimento de homozigosidade do haplótipo estendido (EHH), na Figura 1.9, utilizando como *core* os SNPs rs11264355 (A e B) e rs1052176 (C e D). Observa-se que, principalmente para o rs11264355, o decaimento entre o alelo ancestral e derivado em africanos (A) e europeus (B) é diferente. Em africanos, há uma maior área sob a curva para o alelo derivado (em vermelho) do *core* até a posição 155.28 Mb. Para o rs1052176 também é visto um maior decaimento em africanos (C) do *core* até a posição 155.24 Mb. Então, apesar do decaimento do alelo derivado em africanos ser suave em relação ao alelo ancestral, ocorrendo em aproximadamente 10.000 a 20.000 pb, esse grande haplótipo em homozigosidade pode estar se mantendo diferenciado dos europeus devido à seleção.

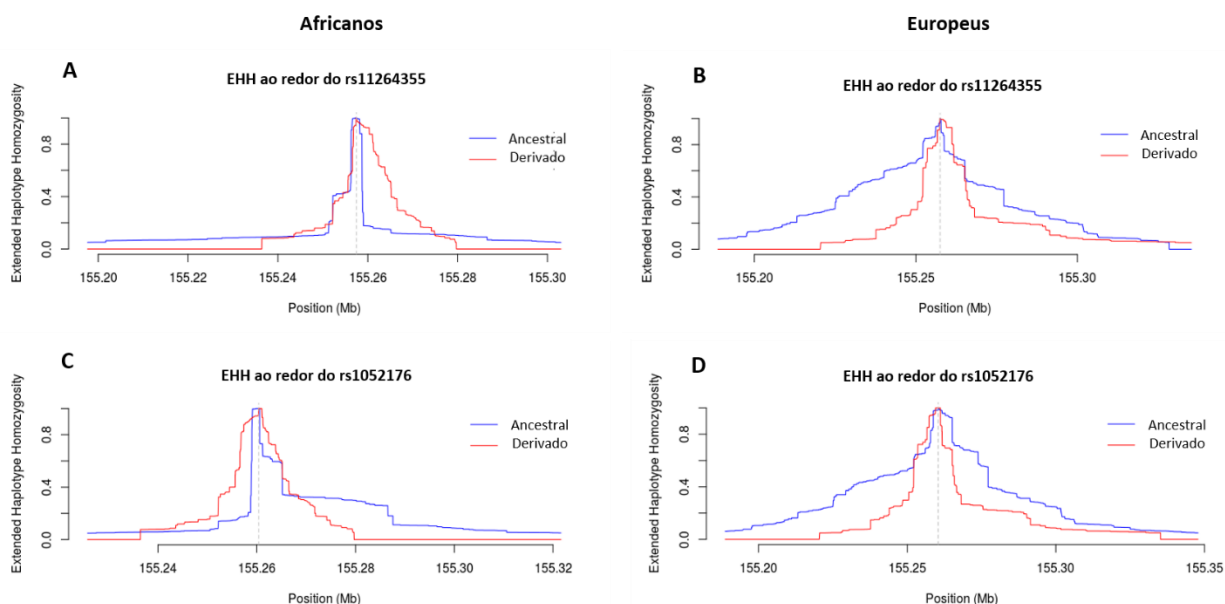


Figura 1.9 – Decaimento de homozigosidade do haplótipo a partir do SNP “core” em europeus e africanos. As linhas azuis e vermelhas representam o alelo ancestral e o derivado, respectivamente. A linha pontilhada representa o SNP “core”, sendo definido como o rs11264355 em A (Africanos) e B (Europeus) e rs1052176 em C (Africanos) e D (Europeus).

2.5 – F_{ST} e $xpEHH$ em uma base de dados de populações africanas

Ainda, na tentativa de verificar algum sinal de seleção em populações africanas endêmicas (Gana) e não endêmicas de malária proveniente do *dataset* de Gouveia e colaboradores publicado

em 2019 na *Plos Genetics*, exploramos a análise de F_{ST} , $xpEHH$ e a frequência alélica para os SNPs rs1052176 e rs11264359 desse estudo (Tabela 1.7). Os valores de F_{ST} para as variantes em africanos, no geral, não são elevados pois a maior diferenciação ocorre quanto maior é distância entre populações (como europeus e africanos). Embora o F_{ST} para o rs11264359 se destaque por ser considerado alto para a comparação entre populações africanas, a média por janela com 25 SNPs não é relevante (0,02) entre Gana e Sotho e Zulu. Entre essas populações, o $xpEHH$ apresentou sinal de *sweep* para o rs11264359. Porém, a frequência alélica para o alelo de risco G desse SNP é maior na população não endêmica para a malária (Alelo G = 0,89), Sotho e Zulu, do que em Gana (Alelo G = 0,71) ao contrário das expectativas (Tabela 1.8). Dessa forma, considerando todos os dados analisados, apesar da sugestão de assinaturas, a identificação de seleção na região do gene *PKLR* não é direta, e mais estudos com diferentes métodos são necessários.

Tabela 1.7 – Análise de F_{ST} , iHS e $xpEHH$ dos SNPs do gene *PKLR* nas populações africanas de Gouveia e colaboradores (2019).

| | Gana x Sotho e Zulu | Uganda Norte Central (UNC) x Sotho e Zulu | Uganda Norte Oeste (UNW) x Sotho e Zulu |
|----------------------------|---------------------------|---|---|
| F_{ST} | | | |
| rs1052176 | 0.038 | 0.0004 | 0.004 |
| rs11264359 | 0.079 | 0.031 | 0.025 |
| 25 SNPs window | 0.020* | 0.013 | 0.023 |
| iHS | | | |
| rs1052176 | 0.399 | -0.349 | -0.252 |
| rs11264359 | 0.194 | -0.031 | 0.345 |
| $xpEHH$ | | | |
| rs1052176 | 0.813 | -0.613 | -0.504 |
| rs11264359 | 2.253 | 0.813 | 1.000 |

A janela apresenta 25 SNPs (155234821-155312711), com exceção de * que apresenta 26 SNPs. Valores em negrito sugerem sinal de seleção.

Tabela 1.8 – Frequência alélica dos SNPs do *PKLR* na população de Gouveia e colaboradores (2019) e 1000Genomes.

| SNP | Gana | UNC | UNW | Sotho-Zulu | ESN | LWK | MAG | MSL | YRI | AFR | EUR |
|---------------------|-------------|------|------|-------------|------|------|------|------|------|------|------|
| rs11264355 C > G | - | - | - | - | 0.43 | 0.49 | 0.49 | 0.56 | 0.41 | 0.49 | 0.28 |
| rs1052176 G > T | 0.49 | 0.62 | 0.58 | 0.64 | 0.42 | 0.50 | 0.49 | 0.55 | 0.42 | 0.48 | 0.28 |
| rs4620533 C > G | - | - | - | - | 0.51 | 0.57 | 0.56 | 0.64 | 0.50 | 0.56 | 0.28 |
| rs4971072 G > A | - | - | - | - | 0.00 | 0.01 | 0.00 | 0.01 | 0.00 | 0.03 | 0.67 |
| rs11264359 G > A | 0.29 | 0.20 | 0.21 | 0.11 | 0.32 | 0.30 | 0.39 | 0.24 | 0.37 | 0.23 | 0.72 |

Frequência do segundo alelo. As populações africanas do 1000Genomes estão destacadas em cinza. Gana (945); UNC = Uganda Norte Central; UNW = Uganda Norte Oeste; Sotho-Zulu (186); **ESN** = Esan, Nigéria (173); **LWK** = Luhya, Quênia (116); **MAG** = Mandinka, Gambia (179); **MSL** = Mende, Serra Leoa (128); **YRI** = Yoruba, Nigéria (186); **AFR** = Africanos e **EUR** = Europeus.

3 – Análises funcionais

3.1 – Comparação entre os carreadores dos genótipos de risco e as dosagens séricas

Na tentativa de identificar uma correlação entre genótipo-fenótipo em relação aos indivíduos carreadores dos genótipos de risco associados à hanseníase, realizamos a comparação com ferro sérico, ferritina e haptoglobina em soro de indivíduos sadios e pacientes de hanseníase. Inicialmente, trabalhamos com o soro dos voluntários sadios e os pacientes do Ambulatório Souza Araújo. Em seguida, realizamos também a dosagem em um grupo de 59 pacientes de Rondonópolis, já que havia disponibilidade dessas amostras. As médias das quantificações nos indivíduos sadios e pacientes do Rio de Janeiro e Rondonópolis foram estratificadas por sexo na Tabela 1.9. Observa-se que tanto em sadios quanto nos pacientes as dosagens de ferro, ferritina e haptoglobina são maiores em homens, tornando necessário considerar a influência do sexo nas análises seguintes. Além disso, é visto que, na doença, os indivíduos do Rio de Janeiro apresentam menores valores de ferro sérico e aumento de ferritina e haptoglobina. Essas duas últimas proteínas, também consideradas proteínas de fase aguda, normalmente estão elevadas durante infecções.

Tabela 1.9 – Média das dosagens séricas entre homens e mulheres do estudo.

| Dosagens | Rio de Janeiro | | | | Rondonópolis | |
|--------------------------------------|-------------------|---------------------|----------------------|--------------------|---------------------|--------------------|
| | Sadios (N=148) | | Pacientes (N=124) | | Pacientes (N=59) | |
| | Homens (N=42) | Mulheres (N=104) | Homens (N=79) | Mulheres (N=45) | Homens (N=40) | Mulheres (N=19) |
| Ferro sérico ($\mu\text{g/dL}$) | 99,25 | 87,07 | 75,13 | 64,54 | 100,9 | 96,36 |
| Ferritina (ng/mL) | 148,36 | 39,49 | 218,80 | 83,60 | 327,84 | 153,60 |
| Haptoglobina (mg/dL) | 126,16 | 119,62 | 180,14 | 166,44 | 126,9 | 117,38 |

Valores de referência: 1) ferro – homens (59 a 160 $\mu\text{g/dL}$) e mulheres (37 a 150 mcg/dL); 2) ferritina – homens (24 a 336 ng/mL) e mulheres (11 a 307 ng/mL) e 3) haptoglobina (30 a 200 mg/dL) em adultos.

Na Figura 1.10, de maneira global, os níveis de ferro são maiores em controles e menores nos pacientes MB, e os níveis de ferritina e haptoglobina são levemente aumentados em pacientes MB na população do Rio de Janeiro (A). Já entre os pacientes de Rondonópolis (B), nenhuma diferença expressiva foi observada entre PB e MB.

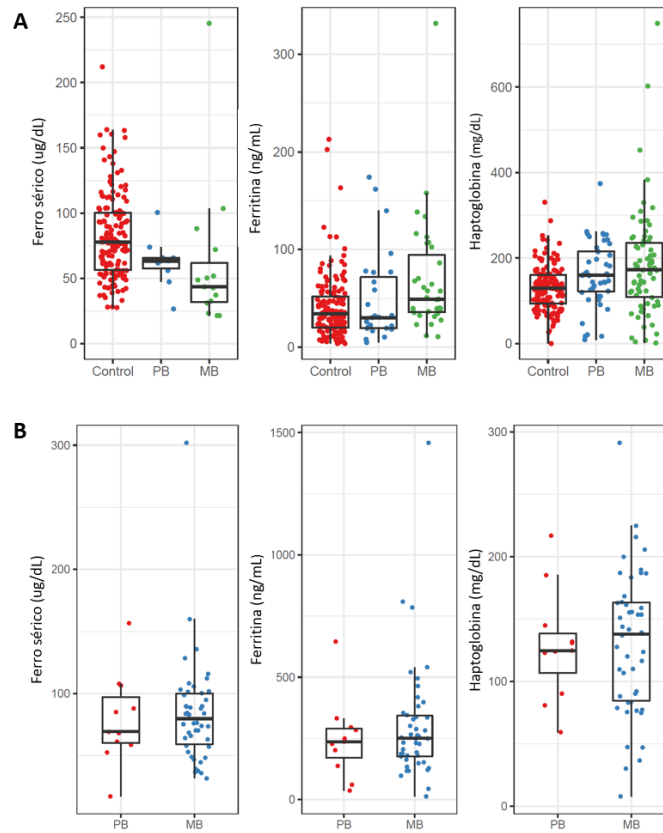


Figura 1.10 – Níveis de ferro sérico, ferritina e haptoglobina entre os indivíduos do estudo. As caixas dos *boxplots* representam o primeiro e o terceiro quartil e a linha central a mediana dos níveis séricos ajustados para idade e sexo por regressão linear. A – Quantificações entre indivíduos saudáveis e pacientes de hanseníase no Rio de Janeiro; B – Quantificações entre pacientes de hanseníase de Rondonópolis. Legenda: PB = pacientes paucibacilares; MB = pacientes multibacilares.

Os níveis de ferro e ferritina nos pacientes totais de Rondonópolis estão levemente aumentados em comparação com os pacientes do Rio de Janeiro. Já a haptoglobina apresenta-se em menores níveis em Rondonópolis comparado com os pacientes do Rio de Janeiro (Figura 1.11).

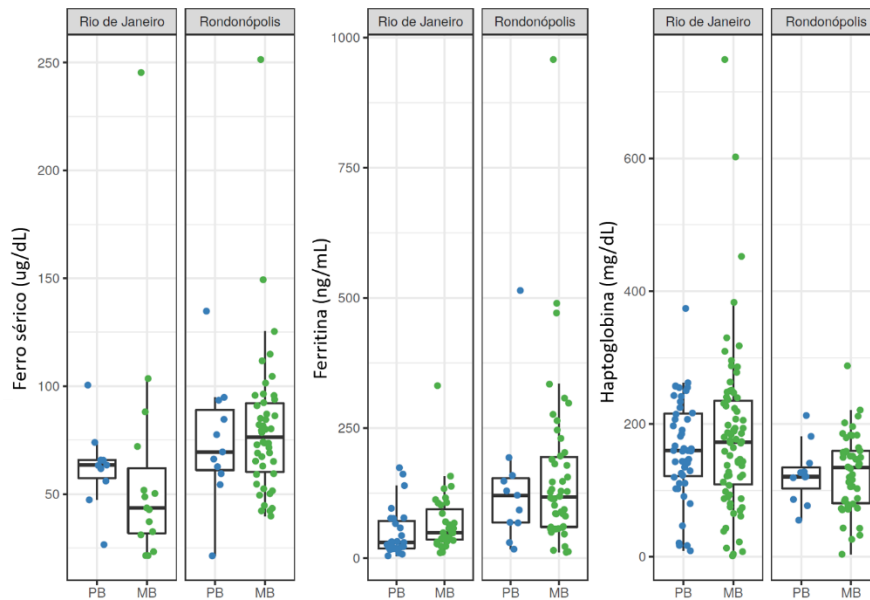


Figura 1.11 – Níveis de ferro sérico, ferritina e haptoglobina entre os pacientes do Rio de Janeiro e Rondonópolis. As caixas dos *boxplots* representam o primeiro e o terceiro quartis e a linha central a mediana dos níveis séricos ajustados para idade e sexo por regressão linear. PB = pacientes paucibacilares e MB = pacientes multibacilares.

A comparação entre os genótipos de risco, ajustando para a idade e sexo, demonstra que não há diferenças expressivas para os níveis de ferro, ferritina ou haptoglobina entre os genótipos dos SNPs rs1052176, rs4971072 e rs11264359 (Figura 1.12). Da mesma forma, comparando os níveis séricos entre os haplótipos (Figura 1.13), nenhuma diferença expressiva entre os grupos foi observada.

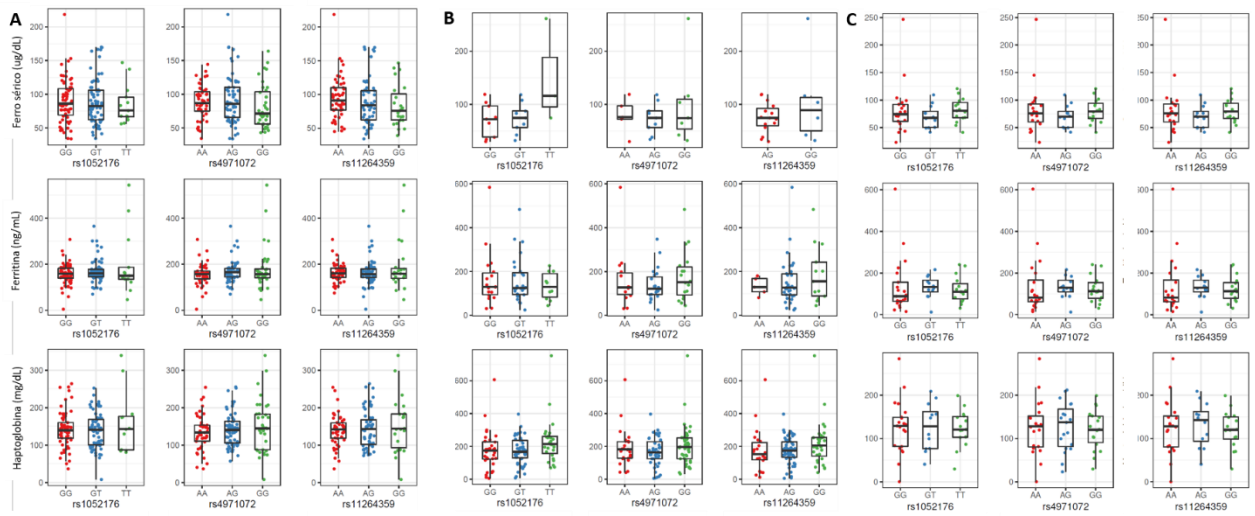


Figura 1.12 – Níveis de ferro sérico, ferritina e haptoglobina comparado com os genótipos de risco. As caixas dos *boxplots* representam o primeiro e o terceiro quartil e a linha central a mediana dos níveis séricos ajustados por regressão linear para idade e sexo em indivíduos saudáveis, e sexo, idade e forma clínica em pacientes. A – Quantificações em voluntários saudáveis (Rio de Janeiro); B – Quantificações em pacientes de hanseníase (Rio de Janeiro) e C – Quantificações em pacientes de hanseníase (Rondonópolis).

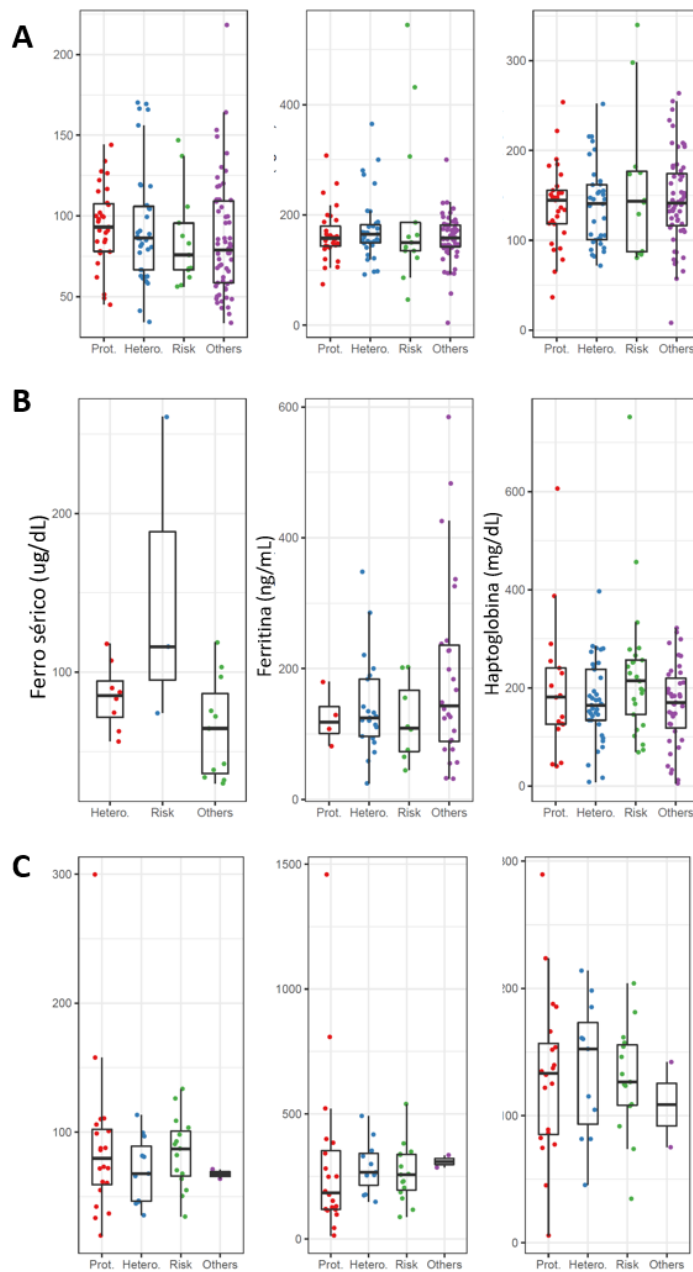


Figura 1.13 – Níveis de ferro sérico, ferritina e haptoglobina comparado com a combinação dos genótipos. As caixas dos *boxplots* representam o primeiro e o terceiro quartil e a linha central a mediana dos níveis séricos ajustados por regressão linear para idade e sexo em indivíduos saudáveis (A) e sexo, idade e forma clínica em pacientes (B e C). Legenda; Prot = haplótipo G/A/A; Risk = haplótipo T/G/G; Hetero. = haplótipo dos indivíduos heterozigotos e Others = outros haplótipos. A – Quantificações em voluntários saudáveis (Rio de Janeiro); B – Quantificações em pacientes de hanseníase (Rio de Janeiro) e C – Quantificações em pacientes de hanseníase (Rondonópolis).

Entretanto, considerando que a média de idade entre homens e mulheres dos indivíduos sadios e dos pacientes do Rio de Janeiro eram similares entre os grupos, em vez de utilizar a idade e o sexo como variáveis de ajuste, exploramos a comparação das dosagens séricas por sexo (Figura 1.14). Nesse caso, não ajustamos para idade e separamos os indivíduos pelo sexo feminino (F) e masculino (M). Assim, foi possível observar níveis sutilmente diferenciados, embora não significativos, para os carreadores dos genótipos de risco do SNP rs1052176. Um aumento na mediana de ferritina é observada nas mulheres do grupo TT em pacientes, enquanto nos homens esse aumento foi visto nos indivíduos sadios. Observa-se que a mediana para haptoglobina é maior entre as mulheres (sadias ou pacientes), porém esse leve aumento em homens somente foi visto em pacientes. Apesar da sugestão de modulação dos níveis de ferritina e haptoglobina pelo genótipo de risco TT, a estratificação por sexo diminui o N de comparação entre os grupos, sendo necessário aumentar o número de indivíduos para auxiliar as conclusões.

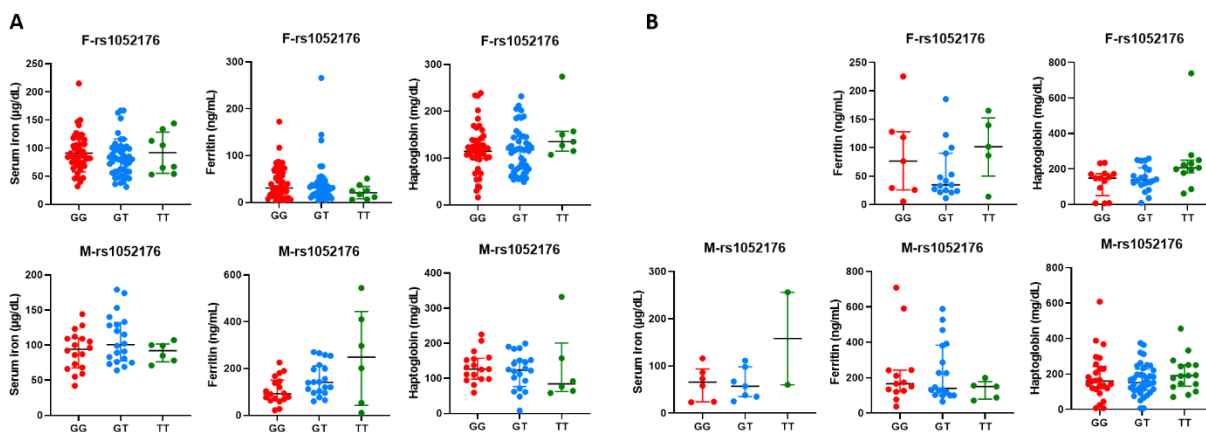


Figura 1.14 – Níveis de ferro sérico, ferritina e haptoglobina entre o sexo feminino (F) e masculino (M) comparado com os genótipos do SNP rs1052176. As barras horizontais representam as medianas com os intervalos interquartis (1^o e 3^o) da distribuição das amostras. Não foi possível comparar o ferro para o sexo feminino devido ao baixo N amostral. A – Quantificações em voluntários saudáveis (Rio de Janeiro) e B – Quantificações em pacientes de hanseníase (Rio de Janeiro).

3.2 – Busca de eQTLs do gene *PKLR* (GTEx)

Considerando que os polimorfismos estudados não apresentam uma correlação com os níveis séricos das dosagens avaliadas, nos questionamos se o papel funcional deles poderia estar relacionado com a regulação da expressão do gene *PKLR*, funcionando como eQTLs. A busca por eQTLs (SNPs que regulam a expressão gênica) tem sido muito utilizada para elucidar como os polimorfismos estão relacionadas às doenças complexas, uma vez que a maioria dos SNPs não altera a funcionalidade da proteína. Para isso, a partir da busca no portal GTEx, observamos que os SNPs rs1052176, rs4971072 e rs11264359 são eQTLs para o gene *PKLR*, porém a diferença de expressão significativa correlacionada aos genótipos desses SNPs somente foi identificada no tecido do nervo tibial. Nos outros tecidos disponíveis no GTEx, ou os SNPs não apresentaram diferença significativa para a expressão diferencial do gene, ou não foi possível quantificar a expressão do gene *PKLR* no tecido (devido à baixa expressão desse gene). Na Figura 1.15, é visto que o genótipo TT para o rs1052176, bem como os genótipos GG para o rs4971072 e rs11264359 modulam negativamente a expressão do gene *PKLR* em biópsias de nervo.

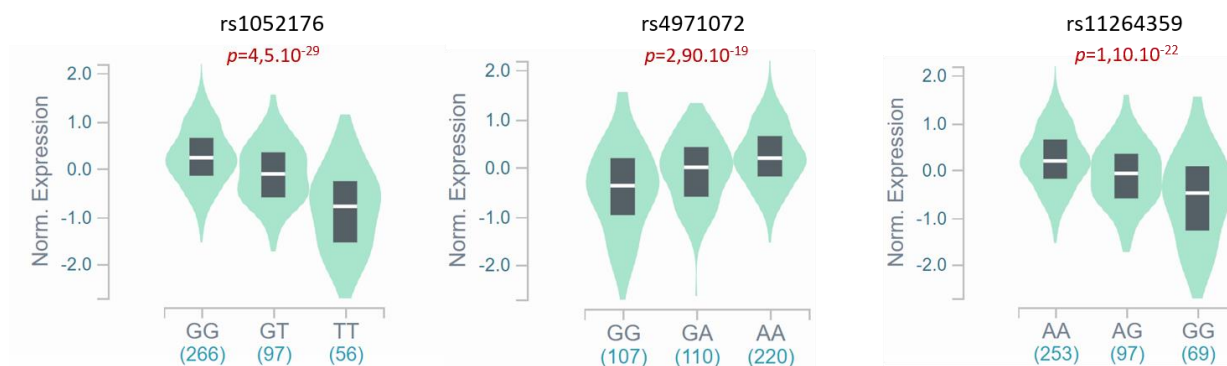


Figura 1.15 – Expressão do gene *PKLR* comparado aos genótipos dos SNPs eQTLs. O gráfico de violino (similar ao *boxplot*), baseado no GTEx, representa a distribuição da expressão do gene *PKLR* em biópsias de nervo tibial por grupos de genótipos: homocigotos para o alelo de referência, heterocigoto e homocigotos para o alelo alternativo. O número de indivíduos de cada grupo está representado em parênteses.

Capítulo IV

***Associação de polimorfismos do gene G6PD
com a hanseníase e o impacto funcional
dessas variantes na infecção***

1 – Análise de associação

1.1 – Associação dos polimorfismos do gene *G6PD* com a hanseníase na população do Rio de Janeiro

Dado que o gene *G6PD* está no cromossomo X, as mulheres apresentam dois alelos para a mutação (heterozigotas ou homozigotas), enquanto os homens carregam somente uma cópia do alelo, ou seja, são hemizigotos. Portanto, as análises de associação foram estratificadas por homens e mulheres, como demonstrado na Tabela 1.1. Inicialmente, avaliando a frequência das variantes nas populações do 1000Genomes, observamos que o alelo T referente ao polimorfismo rs2472393 (C>T) apresenta frequência alélica de aproximadamente 90% tanto em homens quanto em mulheres na população africana (AFR). Em europeus (EUR) a frequência desse SNP está invertida, ou seja, o alelo C é o mais frequente. Da mesma forma, para as mutações clássicas A376G e G202A de troca de aminoácido e com efeitos conhecidamente deletérios para o transcrito, a frequência do alelo G (A376G) é de 33% e 34% e do alelo A (G202A) é 13% e 14% em homens e mulheres africanos, respectivamente. Notavelmente não se observa frequência de nenhuma dessas mutações em europeus (0%), o que sugere que uma pressão seletiva na África é responsável por manter essas mutações elevadas, mesmo que danosas, devido ao seu caráter vantajoso contra a malária.

Em relação à frequência dos SNPs no Rio de Janeiro, primeiramente destacamos que não houve desvio de HWE em mulheres nos grupos casos e controles (rs1050829 – Controles $p=0,16$ e Casos $p=0,39$; rs1050828 – Controles $p=1,00$ e Casos $p=1,00$ e rs2472393 – Controles $p=0,10$ e Casos $p=0,11$). A mutação rs1050829 (A376G) apresentou frequência aumentada de hemizigotos G- (14%), heterozigotas AG (24%) e homozigotas GG (4%) em casos comparado aos controles, porém a associação não foi mantida após a correção. No mesmo sentido, a mutação rs1050828 (G202A), que é a mais rara na população geral, apresentou um aumento de hemizigotos A- (6%) e heterozigotas GA (15%) nos pacientes em comparação aos controles, sem associação genética com a hanseníase após a atribuição dos confundidores sexo e ancestralidade na regressão. Em relação ao SNP intrônico rs2472393, observa-se que há maior frequência de homens hemizigotos T- (44%) e mulheres homozigotas TT (25%) comparado aos controles. Entretanto, essa diferença de frequência entre casos e controles não foi suficiente para levar à associação desse SNP após a

correção para sexo e ancestralidade. Os resultados de associação para os polimorfismos do gene *G6PD* estão resumidamente descritos na Figura 1.1.

Tabela 1.1 – Frequência dos polimorfismos do gene *G6PD* nas populações do 1000Genomes e no Rio de Janeiro e associação com a hanseníase.

| 1000Genomes | | | Rio de Janeiro | | | |
|-----------------------|----------------|----------------|----------------|---------------|-------------------------------------|-----------------------------|
| rs1050829 (A376G) A>G | | | | | | |
| Homens | AFR N = 319 | EUR N = 240 | CA N = 308 | CO N = 327 | OR (IC 95%) p valor | OR (IC 95%) p valor* |
| A- | 215 (0.67) | 239 (0.99) | 264 (0.86) | 306 (0.94) | ref | ref |
| G- | 104 (0.33) | 1 (0.00) | 44 (0.14) | 21 (0.06) | 2.42 (1.40-4.18) 0.001 | 1.62 (0.89-2.92) 0.10 |
| Mulheres | AFR N = 342 | EUR N = 263 | CA N = 163 | CO N = 256 | OR (IC 95%) p valor | OR (IC 95%) p valor* |
| AA | 147 (0.43) | 261 (0.99) | 118 (0.72) | 217 (0.85) | ref | ref |
| AG | 155 (0.45) | 2 (0.004) | 39 (0.24) | 35 (0.14) | 2.04 (1.23-3.40) 0.005 | 1.19 (0.66-2.15) 0.54 |
| GG | 40 (0.12) | 0 (0.00) | 6 (0.04) | 4 (0.02) | 2.75 (0.76-9.96) 0.12 | 0.82 (0.19-3.51) 0.79 |
| A | 449 (0.66) | 524 (0.992) | 275 (0.84) | 469 (0.92) | ref | ref |
| G | 235 (0.34) | 2 (0.002) | 51 (0.16) | 43 (0.08) | 2.02 (1.09-3.72) 0.02 | 1.08 (0.53-2.18) 0.82 |
| rs1050828 (G202A) G>A | | | | | | |
| Homens | AFR N = 319 | EUR N = 240 | CA N = 309 | CO N = 321 | OR (IC 95%) p valor | OR (IC 95%) p valor* |
| G- | 278 (0.87) | 240 (1.00) | 292 (0.94) | 308 (0.97) | - | - |
| A- | 41 (0.13) | 0 (0.00) | 17 (0.06) | 10 (0.03) | 1.79 (0.80-3.98) 0.15 | 1.27 (0.55-2.95) 0.56 |
| Mulheres | AFR N = 342 | EUR N = 263 | CA N = 168 | CO N = 258 | OR (IC 95%) p valor | OR (IC 95%) p valor* |
| GG | 256 (0.75) | 263 (1.00) | 141 (0.84) | 236 (0.91) | ref | ref |
| GA | 78 (0.23) | 0 (0.00) | 26 (0.15) | 21 (0.08) | 2.07 (1.12-3.82) 0.01 | 1.21 (0.60-2.43) 0.58 |
| AA | 8 (0.02) | 0 (0.00) | 1 (0.01) | 1 (0.00) | 1.67 | 1.15 |

| | | | | | | |
|-------------------------|------------------------|------------------------|-----------------------|-----------------------|--------------------------------------|-------------------------------------|
| | | | | | (0.10-26.96) 0.71 | (0.06-19.60) 0.91 |
| G | 590 (0.86) | 526 (1.00) | 308 (0.92) | 493 (0.96) | ref | ref |
| A | 94 (0.14) | 0 (0.00) | 28 (0.08) | 23 (0.04) | 1.94 (0.87-4.36) 0.10 | 1.18 (0.48-2.93) 0.70 |
| rs2472393 C>T | | | | | | |
| Homens | AFR N = 319 | EUR N = 240 | CA N = 290 | CO N = 123 | OR (IC 95%) p valor | OR (IC 95%) p valor* |
| C- | 20 (0.06) | 204 (0.85) | 162 (0.56) | 93 (0.76) | ref | ref |
| T- | 299 (0.94) | 36 (0.15) | 128 (0.44) | 30 (0.24) | 2.44 (1.52-3.92) 0.0002 | 1.64 (0.97-2.79) 0.06 |
| Muheres | AFR N = 342 | EUR N = 263 | CA N = 191 | CO N = 230 | OR (IC 95%) p valor | OR (IC 95%) p valor* |
| CC | 4 (0.01) | 210 (0.80) | 68 (0.36) | 119 (0.52) | ref | ref |
| CT | 36 (0.11) | 51 (0.19) | 76 (0.40) | 85 (0.37) | 1.56 (1.01-2.40) 0.04 | 1.11 (0.69-1.80) 0.65 |
| TT | 302 (0.88) | 2 (0.007) | 47 (0.25) | 26 (0.11) | 3.1 (1.79-5.56) 0.0001 | 1.37 (0.70-2.68) 0.34 |
| C | 44 (0.06) | 471 (0.90) | 212 (0.55) | 323 (0.70) | ref | ref |
| T | 640 (0.94) | 55 (0.10) | 170 (0.45) | 137 (0.30) | 1.89 (1.26-2.82) 0.001 | 1.17 (0.73-1.86) 0.49 |

O dado está apresentado em frequência. Valores em negrito indicam os resultados significativos para $p < 0.05$. Valores não disponíveis (-). Os valores de OR foram calculados baseados nos genótipos ou alelos de referência (ref). Legenda: AFR = Africanos; EUR = Europeus; CO = Controles; CA = Casos e IC = Intervalo de confiança (95%).

* OR (*Odds Ratio*) e *P*-valor ajustado para as covariáveis sexo e ancestralidade.

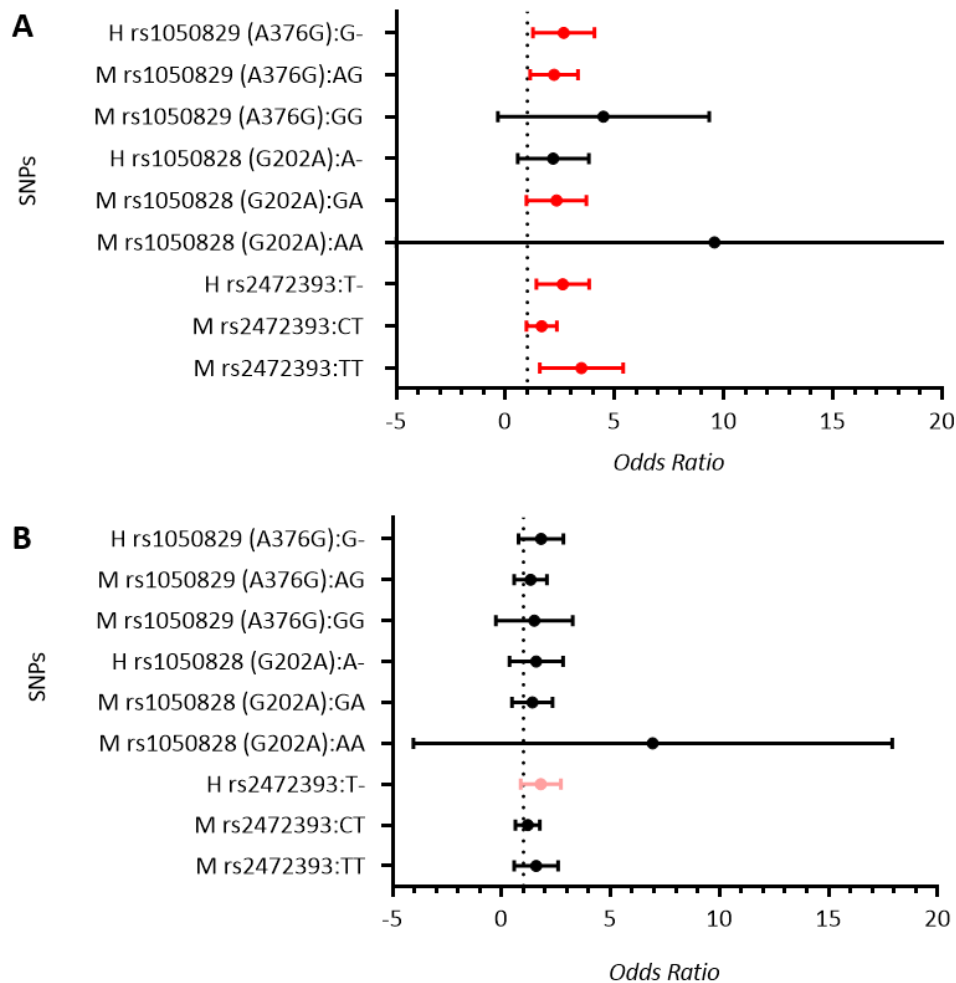


Figura 1.1 – Valores de *Odds ratios* (OR) e intervalos de confiança (95%) para a associação entre os polimorfismos do gene *G6PD* e a hanseníase. A – Valores de OR não ajustados e B – Valores de OR ajustados para sexo e ancestralidade. H = Homens e M = Mulheres. Em vermelho e rosa estão destacados os valores de OR com *p* valores significativos ($p < 0.05$) e *borderline*, respectivamente. Os maiores intervalos de confiança ocorreram para comparações em que há frequência muito baixa ou zerada entre um dos grupos.

1.2 – Análise de LD entre os polimorfismos do gene *G6PD*

A fim de observar o padrão de segregação das variantes de *G6PD* selecionadas nesse estudo, comparamos o perfil de LD entre as populações europeus e africanas do 1000Genomes, os controles e casos do Rio de Janeiro e as coortes do Epigen (Figura 1.2). Em geral, as variantes

estudadas não estão em um bloco em LD nas populações avaliadas. Em A, a frequência das variantes rs1050829 e rs1050828 em europeus é muito baixa e o LD é inexistente entre elas e o rs2472393. Em africanos, existe um baixo LD entre rs1050829 e rs1050828, e em geral as variantes segregam independentemente. Já em B, casos e controles apresentam um LD moderado entre os SNPs rs1050829 e rs1050828, perfil que também é observado nas populações do Epigen, em C. Assim, concluímos que o rs2472393 tem segregação independente em relação aos SNPs exônicos rs1050829 e rs1050828, que por sua vez apresentam um desequilíbrio moderado na população brasileira, diferente do observado em europeus e africanos.

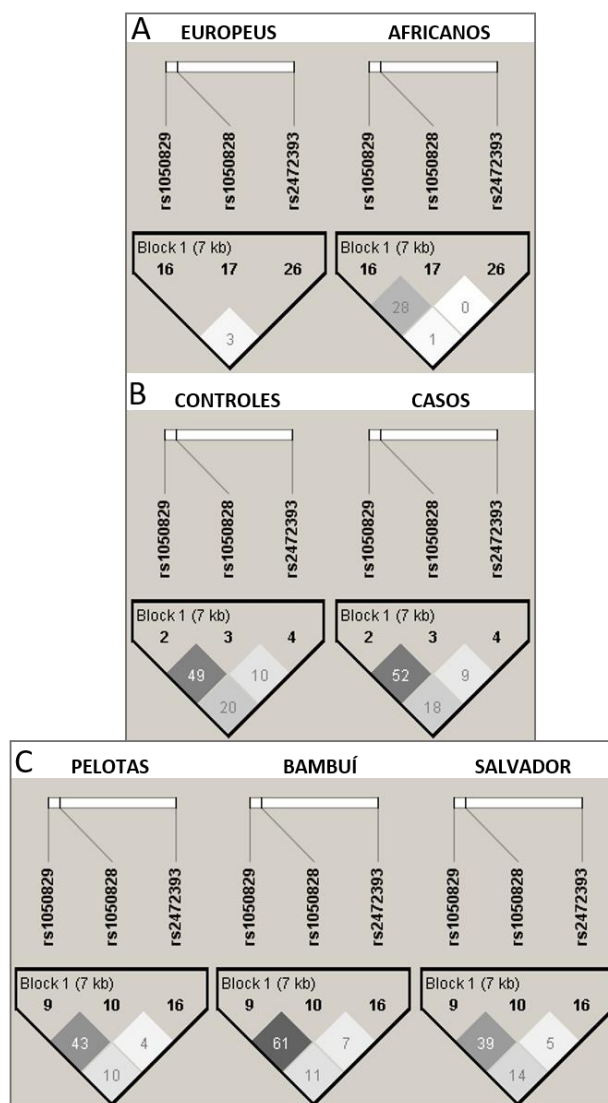


Figura 1.2 – Desequilíbrio de ligação entre os SNPs da região do *G6PD* nas populações do 1000Genomes (A), Rio de Janeiro (B) e Epigen (C). Os losangos apresentam o valor de r^2 uma escala de branco a preto: branco – LD fraco ($r^2 < 0.5$); cinza – LD moderado ($0.5 < r^2 < 0.7$) e preto – LD forte ($r^2 > 0.7$).

2 – Análise funcional

2.1 – Comparação entre os genótipos das mutações em *G6PD* e as dosagens séricas

Independente da ausência de associação do *G6PD* com a suscetibilidade a hanseníase, comparamos as variantes genéticas com os ensaios funcionais objetivando correlacionar o efeito das mutações A376G e G202, que sabidamente alteram a atividade da enzima, no nosso modelo de estudo. Inicialmente, realizamos a comparação das dosagens de ferro, ferritina e haptoglobina com os genótipos das variantes rs1050829 (A376G) e rs1050828 (G202) a fim de avaliar uma correlação funcional genótipo-fenótipo. A Figura 1.3 mostra que, para a variante rs1050828 (G202) (A), não observamos diferenças expressivas para as dosagens entre os genótipos homocigotos GG e heterocigotos GA em mulheres e homens hemizigotos G- e A-. Para o rs1050829 (A376G) (B), também não foi visto diferença para o ferro sérico e ferritina entre os genótipos homocigotos AA e heterocigotos AG no grupo das mulheres e em homens hemizigotos A- e G-. Já para a haptoglobina, apesar de a diferença não ser expressiva, pontuamos um aumento dos níveis dessa proteína em mulheres heterocigotas para a mutação em comparação com as homocigotas selvagens AA. Em homens, entretanto, os carreadores do alelo mutante (hemizigotos G-) apresentaram uma leve diminuição da haptoglobina em vista dos selvagens (hemizigotos A-). Essa diferença pode ter ocorrido devido ao N amostral. Por fim, concluímos que as mutações de *G6PD* não afetam significativamente os níveis de ferro, ferritina e haptoglobina em indivíduos saudáveis.

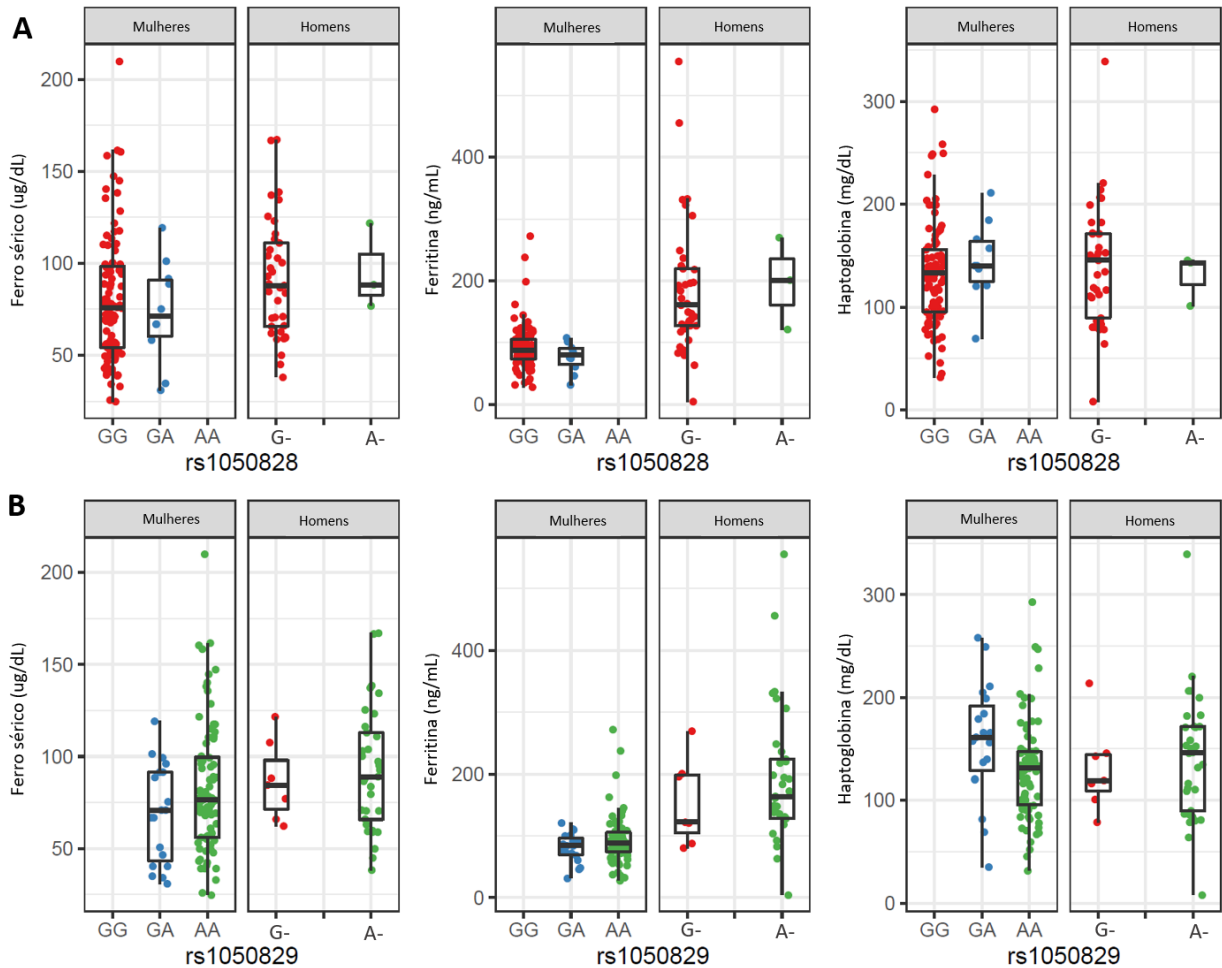


Figura 1.3 – Níveis de ferro sérico, ferritina e haptoglobina por SNP do gene *G6PD* em indivíduos saudáveis. As caixas dos *boxplots* representam o primeiro e o terceiro quartil e a linha central a mediana dos níveis séricos ajustados para idade e sexo por regressão linear. A – Quantificações entre indivíduos saudáveis para o rs1050828 (G202A); B – Quantificações entre indivíduos saudáveis para o rs1050829 (A376G).

2.2 – Impacto funcional das variantes de *G6PD* na infecção pelo *M. leprae*

Dado que resultados anteriores do grupo demonstraram que a via da *G6PD* tem importância no processo de infecção pelo *M. leprae* uma vez que células de Schwann infectadas com o bacilo aumentam a expressão de *G6PD* e de enzimas reguladores dessa via. Logo, mesmo que nenhuma associação significativa tenha sido observada no estudo caso-controle do Rio de Janeiro para as variantes desse gene e a hanseníase, o próximo passo foi comparar as variantes

de “risco” com as características funcionais. Entende-se como variantes de risco aquelas em que a mutação está atrelada a alteração na funcionalidade da enzima. Assim, em vista da importância funcional bem caracterizada das variantes A376G e G202 na atividade da enzima glicose-6-fostato-desidrogenase, exploramos a expressão gênica e a atividade enzimática entre homens e mulheres sadios carreadores e não carreadores dessas mutações.

Inicialmente, uma análise de microarranjo em larga-escala desenvolvida pelo Instituto Lauro Souza de Lima (ILSL) avaliou o perfil da expressão gênica em monócitos de indivíduos sadios do município de Bauru 48 horas após o estímulo com *M. leprae* (MOI 10:1). O resultado desse microarranjo foi acessado para avaliar a modulação da expressão de genes relacionados ao metabolismo de ferro após a infecção, como *PKLR*, *G6PD*, *HFE*, *HMOX*, *FTL*, *HP*, entre outros, e se encontram em ANEXO I (comunicação pessoal com a Dra. Ana Carla Pereira – dados não publicados). Na Figura 1.4 A, demonstramos que há um aumento na expressão de *G6PD* nas células estimuladas na população de Bauru. Em B, observamos o mesmo padrão de expressão do gene *G6PD*, de forma significativa, após a infecção em monócitos primários derivados de indivíduos sadios do Rio de Janeiro, corroborando o achado anterior. Comparando entre as mulheres homozigotas e heterozigotas para as mutações da variante A-, em C, observamos que há um aumento da expressão de *G6PD*, embora não significativo, e não detectamos diferença da expressão baseado no genótipo. Já em D, é possível observar que o aumento da expressão de *G6PD* ocorre na presença da variante A- em homens, independentemente da infecção. Ou seja, no caso dos homens, a própria célula estimula a expressão do gene *G6PD* como um mecanismo compensatório da presença da mutação. Dado que os homens possuem um alelo único para a mutação, esse gênero tem maior chance de expressar o fenótipo, justificando o resultado significativo nesse grupo.

Além da expressão de *G6PD* aumentada, monócitos estimulados com *M. leprae* após 48h de infecção também induzem o aumento da atividade enzimática da G6PD de forma significativa, como é observado em F. Por último, em G, tentamos identificar se a alteração na atividade também poderia estar atrelada à presença das mutações (variante A-). Entretanto, não é observado diferença significativa da atividade da enzima a nível de genótipo. Ressaltamos que a tendência de diminuição da atividade enzimática nos homens que carregam a variante A- (G-/A-) vai de acordo com a literatura, já que essas mutações são descritas como potentes redutoras da

atividade da enzima. Porém, um estudo com um maior N amostral é necessário. Com isso, concluímos que a infecção pelo *M. leprae* estimula tanto a atividade da G6PD quanto a expressão de *G6PD*, e essa também é maior nos homens carreadores da variante A-.

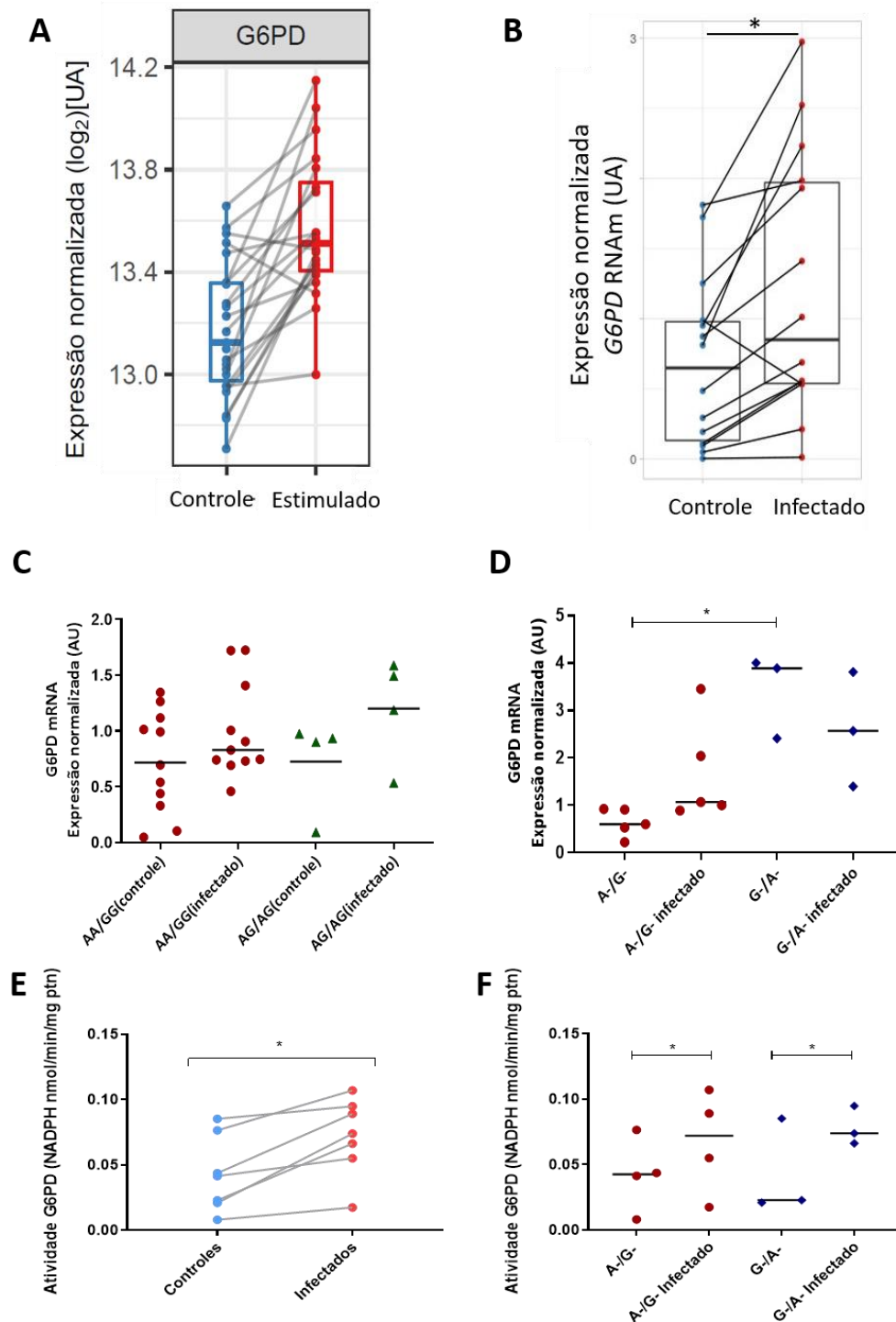


Figura 1.4 – Expressão do gene *G6PD* e atividade enzimática em monócitos de infectados com *M. leprae*. **A** – Monócitos primários de indivíduos sadios de Bauru (SP). A a expressão do *G6PD* foi quantificada 48 horas após a infecção (MOI 10:1) (dados em preparação para o manuscrito). **B** – As células foram coletadas de indivíduos do Rio de Janeiro (RJ) e a expressão analisada 48 horas após a infecção (MOI 50:1). **C** – Expressão de *G6PD* em mulheres homozigotas selvagens (AA/GG) e heterozigotas (AG/GA) para as mutações G202A/A376G; **D** – Expressão de *G6PD* em homens hemizigotos (A-/G-) selvagens e hemizigotos (G-/A-) para as mutações G202A/A376G. **E** – Atividade da *G6PD* em monócitos de doadores homens sadios após 48h de infecção com *M. leprae*. **F** – Atividade da *G6PD* em homens hemizigotos (A-/G-) selvagens e hemizigotos (G-/A-) para as mutações G202A/A376G. A comparação entre os controles e os infectados foram realizadas utilizando o teste de Wilcoxon ($*p<0.05$) e entre os grupos de genótipos a comparação utilizou o teste de Mann-Whitney ($*p<0.05$). Os dados dessa figura foram partilhados com o estudo de Medeiros, 2018 (288) (o manuscrito com primeira autoria compartilhada se encontra em ANEXO III).

Capítulo V

Discussão e Conclusões

1. DISCUSSÃO

Ao longo desse trabalho, discutimos como as relações evolutivas entre patógeno-hospedeiro moldam o padrão de variabilidade genética humana e influenciam na diversidade de fenótipos observados na natureza e na suscetibilidade a doenças complexas. Levantamos a hipótese de que eventos evolutivos passados, como a seleção de variantes genéticas que promovem resistência à infecção pelo *Plasmodium sp.*, deixam “assinaturas” no genoma que são propagadas ao longo das gerações. Tais variantes genéticas, contudo, manifestam um fenótipo não muito atraente do ponto de vista de um indivíduo sadio, já que o preço da resistência muitas vezes é o desequilíbrio de alguma via metabólica ou a disfunção de estruturas celulares. Esse é um aspecto interessante da seleção natural, que atua moldando os fenótipos que são possíveis, mas que nem sempre são os mais desejáveis. No caso da malária, o efeito protetor tange à imperfeição das células vermelhas e a deficiência de enzimas metabólicas cruciais à sobrevivência, mas que, em contrapartida, garante a sobrevivência do indivíduo no ambiente hostil. Essa via evolutiva de perda e ganho é o que denominamos de *trade-off*, e na ausência de uma tradução adequada, o termo será expresso como “mecanismo evolutivo compensatório”.

Evidências acerca da provável presença dessas assinaturas de seleção em genes que regulam enzimas eritrocitárias, como o gene *PKLR* e o *G6PD*, estão demonstradas nas populações africanas endêmicas para a malária. O déficit funcional provocado em indivíduos com esses genes “assinados” é a chance de ocorrência de anemias hemolíticas. Alguns patógenos intracelulares capturam o ferro proveniente do heme, que por sua vez é liberado em eventos de hemólise. Portanto, sugerimos que variantes genéticas selecionadas para a resistência a patógenos na África podem levar a suscetibilidade a microrganismos intracelulares, como as micobactérias. Essa hipótese é endossada pela estrutura genética da população brasileira, que carrega seguimentos genômicos provenientes de três populações ancestrais, incluindo os africanos. Sendo possível detectar os alelos de resistência à malária em brasileiros, os genes *PKLR* e *G6PD* seriam bons candidatos ao estudo de doenças provocadas por micobactérias e que sejam prevalentes no Brasil, como a hanseníase e tuberculose.

As variantes genéticas dos genes *PKLR* e *G6PD* estudadas nesse trabalho são descritas como protetoras para a malária em africanos, população que é endêmica para a malária há

milênios (66–68,77,78,294,295). Essas variações são encontradas em regiões brasileiras não acometidas pela malária. Ademais, como a população brasileira é muito recente, apresentando cerca de 500 anos, ainda não ocorreu tempo suficiente para gerar alteração no padrão genético da população devido à seleção natural. Nesse aspecto, considerando o histórico de miscigenação brasileira, essas variantes genéticas assinaladas no genoma da nossa população é fruto da migração dos africanos ao território brasileiro. Assim, considerando que essas variantes são “africanas”, quanto maior a porcentagem de africanos na população brasileira de estudo (como Rio de Janeiro e Salvador), maior a chance de detecção desses polimorfismos. Vale ainda ressaltar que devido à elevada ancestralidade africana nos casos de hanseníase do Ambulatório Souza Araújo em comparação aos controles do Rio de Janeiro, provenientes de banco de doadores de medula óssea, um aumento considerável na frequência desses polimorfismos é esperado nos pacientes. Por conta disso, afim de evitar associações genéticas espúrias com a doença, a ancestralidade genética foi considerada como variável de ajuste, juntamente com o sexo e a idade (quando o dado se encontrava disponível), na regressão logística das populações do Rio de Janeiro e de replicação.

1.1 – Suscetibilidade do gene *PKLR* às doenças micobacterianas

O cenário ideal para os estudos em genética humana seria conhecer em nível molecular cada indivíduo e os determinantes genéticos de cada traço fenotípico. Em vista de inviabilidade dessa abordagem, especialmente em doenças complexas, busca-se caracterizar e agrupar indivíduos semelhantes e desenvolver métodos para comparar esses grupos. A relevância dessa estratégia é o aumento da acurácia dos genes identificados como possíveis preditores de risco da doença em estudo (259,271). A associação genética do gene *PKLR* com as doenças micobacterianas foi desvendada na população do Rio de Janeiro e Moçambique anteriormente (293). Porém, os ajustes aplicados na correção do modelo de associação da população do Rio de Janeiro basearam-se no sexo e etnia. Em vista de um estudo paralelo do nosso grupo que identificou diferenças a nível de ancestralidade global entre os pacientes de hanseníase acolhidos no Ambulatório Souza Araújo (FIOCRUZ) e os controles do banco de sangue do INCA (292), fez-se necessário revisitar a análise de associação desse gene. O grupo dos pacientes apresenta uma

porcentagem de ancestralidade similar ao encontrado para a população de controles do Rio de Janeiro, como em Manta e colaboradores, 2013 (279). Assim, o diferencial entre os grupos se deve à seleção dos controles provenientes de banco de doadores de medula óssea, que provavelmente são moradores de outras regiões geográficas da grande cidade do Rio de Janeiro. Os controles ideais, nesse caso, poderiam ser os indivíduos habitantes das vizinhanças dos pacientes, como os contatos domiciliares que não adoeceram. Os controles do banco de sangue de Salvador (Hemoba) não apresentaram essa diferença em relação aos casos de hanseníase. Dessa forma, na tentativa de lapidar a associação do gene *PKLR* com a hanseníase no Rio de Janeiro, sabendo-se da diferença entre casos e controles, aplicamos o ajuste por ancestralidade genômica no modelo de associação na regressão logística. Apesar da potencial perda de poder da análise, a associação foi mantida, confirmando a relação do gene *PKLR* com a suscetibilidade a hanseníase no Rio de Janeiro.

Antes de abordar as peculiaridades da associação do gene *PKLR* nas populações de replicação, considerações sobre o histórico de cada população são ponderadas. Rio de Janeiro e Salvador são as populações brasileiras que apresentam a maior ancestralidade africana, condizente com o histórico demográfico ocorrido entre elas (279,296,297). Salvador foi capital do Brasil por 214 anos até a sede ser alterada para o Rio de Janeiro. Ambas as cidades possuíam os principais portos por onde chegavam os navios negreiros, e, juntas, enviaram mais viagens de escravos do que qualquer outro porto da Europa (246,298). Manaus e Rio Branco, em contrapartida, apresentam uma maior contribuição europeia e nativo-americana. Curiosamente, considerando que Manaus é uma capital de importância econômica central no Norte e que apresenta fluxo de imigrantes de outras regiões do Norte e do Nordeste, a ancestralidade nativo-americana foi maior do que a esperada e observada em outros estudos (279). Rondonópolis teve a ancestralidade examinada pela primeira vez nesse trabalho, e, em geral, se aproxima da encontrada no Rio de Janeiro, com a exceção de que há maior contribuição nativo-americana, como descrito em outros estudos para a região Centro-Oeste. No geral, o perfil de ancestralidade identificado nesse estudo está de acordo com os estudos prévios em relação às diferentes regiões brasileiras descritas na literatura (299).

A replicação da associação do gene *PKLR* com as doenças micobacterianas, observada inicialmente no Rio de Janeiro e em Moçambique, foi confirmada em Salvador, mas não em

Manaus e Rondonópolis. Rio de Janeiro e Salvador são as populações brasileiras que apresentam a maior ancestralidade africana, condizente com o histórico demográfico ocorrido entre elas (279,296,297). Portanto, observamos que a associação genética do *PKLR* é detectada em populações com maior ancestralidade africana, o que corrobora a nossa hipótese já que os polimorfismos estudados são prevalentes em populações africanas, sugerimos que a associação de suscetibilidade do *PKLR* é dependente da ancestralidade. Nesse caso, descrevemos a ancestralidade local, no nível do *loci*, ou seja, a associação será detectada quanto mais “africano” é o seguimento genômico do *PKLR*. E, nesse caso, há maior chance de encontrar ancestralidade local africana em indivíduos que apresentam maior ancestralidade global africana. Isto é, quanto maior o componente africano na população, maior a chance de identificar a associação. Por outra perspectiva, discute-se que a correção por ancestralidade pode minimizar a associação de genes dependentes de ancestralidade, exaltando ainda mais os estudos de replicação em populações independentes na intenção de compreender a associação.

O estudo em populações miscigenadas tem muito valor em genética de populações. Em primeiro lugar, porque a maior porcentagem de trabalhos na literatura prioriza investigações em populações com ancestralidade europeia (300). Europeus, contudo, representam um quinto da população humana. Sabe-se que as variantes de risco variam profundamente entre populações devido às diferenças na deriva genética, ao padrão de LD entre os marcadores causais e os alelos neutros, e ainda à relação com o ambiente ou a outros modificadores genéticos que ampliam o efeito de uma variante causal em uma população e não em outra (257). Portanto, o estudo em europeus restringe o apuramento das variantes causais em doenças complexas em outras populações do mundo. Em segundo lugar, a distribuição das doenças difere ao redor do mundo. Países em desenvolvimento com características tropicais são endêmicos para doenças não frequentes em europeus. Portanto, estudos em europeus nem sempre refletem a descoberta de variantes causais que possam auxiliar para formulação de vacinas, métodos diagnóstico, melhora de tratamento para os agravos mais recorrentes no país. O Brasil ocupa um lugar privilegiado no tocante a diversidade, refletindo a concentração de populações milenares que se misturaram ao longo dos últimos 500 anos em um único território, criando um padrão genético que envolve variantes presentes em quase todo o mundo. Assim, o estudo no cenário brasileiro pode trazer novidades ainda não exploradas na área da genética humana.

A relevância do *PKLR* no desenvolvimento da hanseníase ainda foi corroborada pela associação dos SNPs rs1052176, rs4971072 e rs11264359 na população Chinesa Han do estudo GWAS de Zhang e colaboradores (2009), como visto na Tabela 1.1 dessa seção. Os dados desse estudo pan-genômico estão disponíveis para acesso no banco dbGAP do NCBI (número de acesso: phs000217.pha002872). Nessa população, o alelo T do rs1052176 foi associado com a hanseníase na direção de risco (Alelo G – OR=0,78 e $p=0,01$), como observado no Rio de Janeiro. Já os SNPs rs4971072 e rs11264359 sugerem uma associação de proteção para o alelo G com a hanseníase (Alelo G – OR=0,81; $p=0,04$ e OR=0,82; $p=0,06$, respectivamente), ao contrário do visto no Rio de Janeiro. A associação de alelos opostos de *locus* bialélicos com a mesma doença, o que é denominado associação “flip-flop”, pode parecer confusa quando avaliada em populações similares e normalmente são consideradas espúrias. Porém, esse tipo de associação pode ocorrer em populações diferentes, como é o caso das populações brasileira e asiática, e indicar efeitos heterogêneos da mesma variante devido às diferenças na arquitetura genética da população, ao perfil de LD e ao ambiente ao qual essas populações são expostas (301). Sugerimos que o rs1052176 apresenta-se como o principal SNP associado com a hanseníase, pois além de estar em região codificante, foi o único em que o alelo T foi associado ao risco tanto na população do Rio de Janeiro, quanto em Salvador, Moçambique e China. Assim, concluímos que a replicação da associação genética com valores de OR moderados entre as populações indica que o *PKLR* exibe um papel modesto, mas relevante na suscetibilidade à hanseníase.

Tabela 1.1 – Associação do gene *PKLR* com a hanseníase na população de Chineses Han de Zhang e colaboradores (2009).

| Chineses Han (GWAS – Zhang, 2009) | | | | |
|-----------------------------------|---------|---------|--------------|----------------|
| SNP | Alelo 1 | Alelo 2 | OR (Alelo 2) | <i>p</i> valor |
| rs1052176 | T | G | 0.78 | 0.01 |
| rs4971072 | A | G | 0.81 | 0.04 |
| rs11264359 | A | G | 0.82 | 0.06 |

Em negrito se encontram os valores significativos ($p<0.05$).

O paradigma do estudo genético em doenças infecciosas vem ganhando novas vertentes conforme os estudos em diferentes modelos são desvendados. Até os anos 90, postulava-se que imunodeficiências monogênicas raras levavam a suscetibilidade a múltiplas infecções (hipótese

“um gene, múltiplas infecções”) enquanto as doenças infecciosas comuns estavam relacionadas com modelos de herança poligênica de vários genes (hipótese “uma infecção, múltiplos genes”). Atualmente, também é aceita a hipótese “um gene, uma infecção”, que descreve a existência de um grupo primário de imunodeficiências que predispõe a uma única infecção e que, paralelamente, as doenças comuns refletem a herança de um ou poucos genes principais de suscetibilidade (do inglês, *major susceptibility genes*) (302). Essa última hipótese ganhou força em estudos recentes em que mutações raras em genes do eixo IL-23/IFN- γ que provocam imunodeficiências relacionadas a produção de citocinas aumentam a suscetibilidade a micobactérias ambientais avirulentas, ou MDMS (do inglês, *Mendelian susceptibility to mycobacterial disease*), e a tuberculose. Nesse contexto, é provável que os fenótipos complexos sejam regulados tanto por variantes comuns, com frequência acima de 5% (hipótese “doença comum, variantes comuns”), quanto por variantes raras (hipótese “doença comum, variante rara”) que podem ser consideradas principais no desencadeamento das doenças observadas nas populações (303,304).

Dentro do modelo da hanseníase, acredita-se que poucos genes têm papel central para a incidência da doença, enquanto outros exercem influência sutil (305). O gene *PKLR* apresenta-se como um candidato em que polimorfismos com frequências intermediárias estão relacionados com um valor de OR modesto (1,2 a 1,7) para a associação com a hanseníase. Essas variantes de efeito sutil podem funcionar como preditores de risco juntamente com os marcadores prévios identificados. Porém, o desdobramento funcional dos polimorfismos no *PKLR* não foi totalmente desvendado nesse estudo, o que prejudica extrapolar a importância central ou não desse gene na doença. Embora a comparação das quantificações séricas entre homens e mulheres tenha sugerido que o genótipo de risco TT pareça modular os níveis de ferritina e haptoglobina, não observamos alterações expressivas. Entretanto, ressalta-se que os níveis de ferro são modulados por diversos fatores como exercício físico, alimentação, tabagismo, além de outros genes que não foram considerados nas análises (306). Além disso, o aumento na contagem de reticulócitos (onde a atividade da enzima é aumentada em relação a hemácias maduras) pode compensar uma possível alteração enzimática nas hemácias (28). Outro ponto é que como os polimorfismos associados com a hanseníase têm frequência elevada na população, a probabilidade de um efeito funcional ser expressivo em soro é baixa. Portanto, a sugestão de alterações sutis provocados

pela presença do genótipo de risco corrobora a hipótese genética identificada, apesar de mais estudos com um N amostral maior seja necessário.

A partir de um estudo do nosso grupo publicado por Barbosa e colaboradores em 2017, ficou claro a dependência do ferro para o *M. leprae* na forma mais grave da hanseníase. Nesse estudo, identificamos um maior depósito de ferro e menor expressão de ferroportina (relacionada com o efluxo de ferro) em pacientes lepromatosos (LL), além do aumento da expressão de proteínas envolvidas com o metabolismo de ferro (hemoglobina, haptoglobina, heme-oxigenase 1, receptor de transferrina 1 e do receptor CD163). Além disso, foi visto que a IL-10 pode modular a expressão de CD163, aumentando a captação de ferro e favorecendo a persistência do *M. leprae* em pacientes LL em comparação aos pacientes BT (180,182). É provável que algumas dessas alterações fisiológicas possam ser acentuadas devido à presença do haplótipo de risco T/G/G e melhor “visualizadas” a nível celular em ensaios *in vitro*. A hipótese a qual associa a relação das proteínas reguladoras de ferro com o haplótipo de risco associado com a suscetibilidade à hanseníase está representada na Figura 1.1.

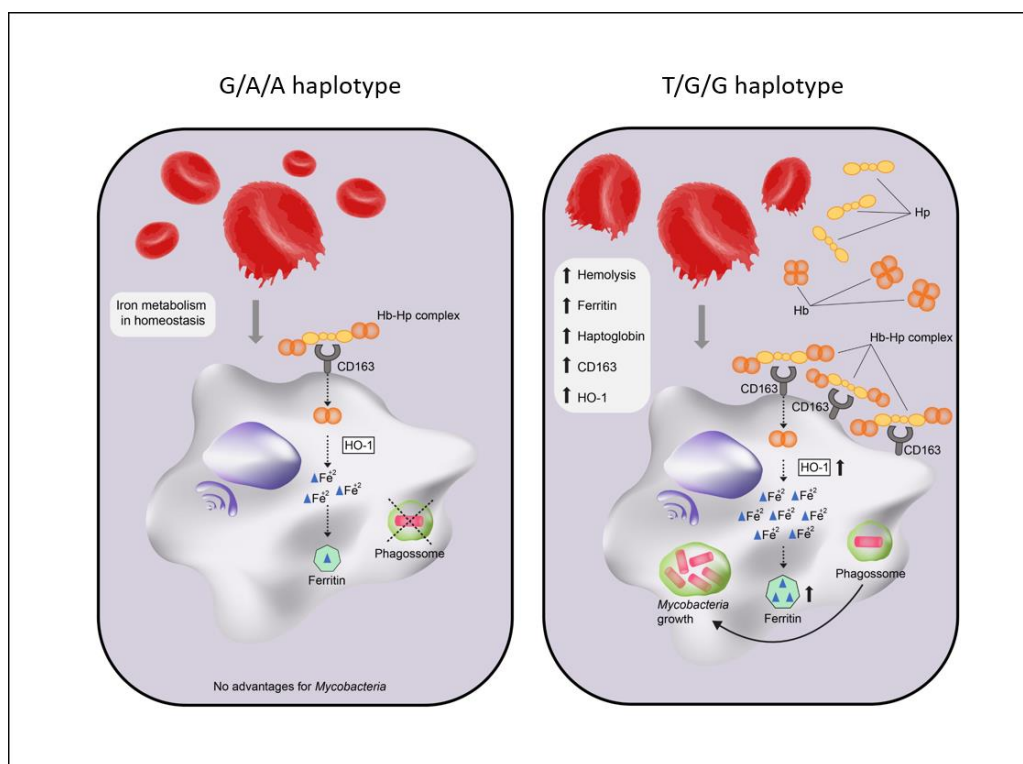


Figura 1.1 – Modelo proposto para o mecanismo de sobrevivência micobacteriana em indivíduos portadores dos diferentes haplótipos do gene *PKLR*. O esquema ilustra a possível via que promove o

aumento dos níveis de ferro e, conseqüentemente, a persistência de *M. leprae* em monócitos e macrófagos. Supomos que carreadores do haplótipo T/G/G possam apresentar algumas hemácias com menor meia-vida e que, conseqüentemente, tenham maior chance de hemólise. Devido à liberação de hemoglobina (Hb), a haptoglobina (Hp) pode se encontrar levemente aumentada. O complexo Hp-Hb internaliza o heme nos macrófagos via CD163. A enzima heme-oxigenase 1 (HO-1) cliva o heme em bilirrubina, CO e ferro, que é armazenado em ferritina. O mecanismo descrito ocorre em homeostase para evitar que o ferro livre provoque estresse oxidativo no interior das células. No entanto, sugerimos que esse mecanismo é aprimorado nos carreadores das variantes de risco para “sequestrar” ferro, levando a uma vantagem para o crescimento micobacteriano.

Em geral, a maioria dos estudos de associação, e principalmente os GWAS, revelam que o controle da suscetibilidade à hanseníase está intimamente ligado a variantes de pequeno efeito (OR<1,5). Contudo, a maioria dos efeitos funcionais dessas variantes identificadas permanecem sem explicação. Aproximadamente 90% das variantes associadas ocorrem em sítios fora das regiões codificantes do gene (45% em regiões intrônicas, 43% em regiões intergênicas), o que sugere a provável chance de muitas delas atuarem como agentes reguladores (307). Alinhado ao fato de que a maioria das variantes em estudos de associação não são relacionadas com mudanças na proteína, outras alterações funcionais são possíveis, principalmente para polimorfismos em regiões regulatórias que podem modificar a expressão gênica. Além da expressão, as variantes podem também alterar o *splicing*, alguns RNAs não codificantes (microRNAs) e a estabilidade do RNAm (308). Nesse estudo, a hipótese de que os SNPs estudados funcionam como cis-eQTLs, diminuindo a expressão do gene *PKLR*, traz uma explicação para o fenômeno genético observado. Entretanto, esse dado baseou-se na expressão em nervo tibial somente (no painel do GTEx), pois o índice de detecção da expressão do gene *PKLR* é baixo em outros tecidos, e então o estudo das variantes eQTLs nesse gene apresentam dificuldades. Contudo, os dados em nervo são interessantes uma vez que o *M. leprae* tem predileção por nervos periféricos. Na tentativa de replicar o achado observado em nervo, utilizamos amostras de sangue total provenientes de pacientes com hanseníase e indivíduos saudáveis a fim de correlacionar os SNPs com a expressão dos genes *PKLR*, *FTL* (ferritina) e *HP* (haptoglobina), mas encontramos dificuldade na detecção da expressão do gene *PKLR* em sangue total e essa abordagem do estudo precisa ser melhor avaliada (ANEXO II). Ainda, outras modificações

fisiológicas são passíveis de ocorrer. Análises computacionais para avaliar o efeito de SNPs não-sinônimos, como o rs1052176, nos transtornos de células vermelhas mostraram que esse SNP pode alterar a transcrição devido a modificações no sítio de ligação ao ribossomo (309). Por fim, recentemente, os SNPs rs8177970 e rs8177963 (excluído nos filtros de MAF para seleção de SNPs desse estudo) foram correlacionados com a deficiência da PK devido a alterações no *splicing* (310).

Apesar do papel funcional dos SNPs associados com a hanseníase ainda não serem totalmente decifrados, a relevância genética do gene *PKLR* com a hanseníase e tuberculose, apresentados neste estudo, podem contribuir para uma expansão do modelo atual de suscetibilidade genética as doenças micobacterianas, ao permitir a inclusão de um novo gene que participa do mecanismo de suscetibilidade à doença. A implicação futura desse estudo é a definição de um painel de marcadores de suscetibilidade útil no valor preditivo para o desenvolvimento da hanseníase, que poderá ser implementado no acompanhamento clínico de contatos domiciliares.

1.2 – Pegadas de seleção nos genes *PKLR* e *G6PD*

A identificação de marcadores alvo de seleção natural pode ser informativo para identificar alelos de risco associados a doenças (2,311). A frequência elevada do haplótipo de risco do *PKLR* em casos de Salvador, Rio de Janeiro, Moçambique e africanos do 1000Genomes em relação aos grupos controles e europeus, sugere a presença de “haplótipos africanos” associados com a doença. Essa motivação nos levou a investigar assinaturas evolutivas na região do gene *PKLR*. A partir de 2010 as primeiras descrições da prevalência de mutações de deficiência da PK em populações africanas como Moçambique, Angola, São Tomé e Príncipe, Guiné Equatorial, e Cabo Verde começaram a surgir. Machado e colaboradores suspeitaram de possíveis marcas de seleção na região do gene *PKLR* baseado na maior diferenciação de SNPs e STRs (microsatélites) entre africanos de regiões endêmicas (Moçambique e Angola) e portugueses comparado a marcadores neutros. Além disso, uma região genômica ampla em alto LD foi vista no grupo de malária não grave, mas não no grupo de malária grave, reforçando que a malária pode estar exercendo pressão seletiva sobre a região do gene *PKLR* (66–68). Em 2015, o grupo do Dr. Philippe

Gros, pioneiro na demonstração de que as mutações no gene *PKLR* protegem contra a malária em modelo murino e *in vitro* (63,64), avaliou um longo haplótipo de variantes prevalentes em populações endêmicas da malária em Senegal e na Tailândia, porém nenhuma abordagem evolutiva foi aplicada (69). Além da seleção natural, os eventos demográficos (deriva, fluxo migratório e dinâmica populacional) também regulam as frequências alélicas do genoma, e, portanto, o fenômeno de seleção natural somente é corroborado quando evidenciado por diferentes testes (30). Assim, nesse trabalho, exploramos algumas metodologias para detecção de seleção na região do *PKLR*, porém, apesar dos dados serem consistentes com a hipótese formulada, as análises não foram conclusivas para afirmar a presença de pegadas de seleção no gene.

Além das suposições prévias da literatura, o primeiro indício de que a região poderia ser alvo de seleção foi a grande diferença alélica dos polimorfismos estudados entre europeus e africanos. A frequência desses SNPs se encontra invertida entre essas populações. Então, objetivando detectar desvios de neutralidade na região do gene, inicialmente realizamos a análise de Tajima D. No modelo neutro de seleção, as frequências alélicas flutuam de forma aleatória e não influenciam no *fitness* do indivíduo. Sob seleção positiva, há diminuição da variabilidade genômica ligada ao alelo selecionado, o que é chamado de *sweep* (ou varredura) seletivo. Uma menor diversidade nucleotídica é observada assim como um excesso de variantes raras, provocadas pelo “efeito carona” ao redor do alelo que tende à fixação (34,51). Assim, nossos dados não apontam para a ocorrência de seleção positiva na região do *PKLR*, pois apesar de observarmos um grande bloco em LD com o gene adjacente, não foi visto excesso de alelos raros. A análise de Tajima D também não aponta para a ocorrência de seleção positiva na região. A maioria das janelas de 1000pb apresentam valores positivos, sugerindo a ocorrência de seleção balanceadora. Outro ponto que corrobora a possível ocorrência desse último regime seletivo é a grande densidade de alelos com frequência intermediária (%), em vez da predominância de alelos raros, como esperado inicialmente para seleção positiva.

Em seguida, realizamos a análise de F_{ST} *outlier*, uma abordagem muito utilizada para averiguar diferenciação elevada de SNPs entre duas populações, que pode indicar seleção. Entre europeus e africanos, há SNPs que são extremamente diferenciados, como é o caso do SNP rs4971072 do gene *PKLR* (frequência de 0,9 numa população e praticamente ausente na outra).

O fato de alguns SNPs serem mais diferenciados do que o resto dos SNPs da região genômica é indicativo de que eles estiveram sob seleção. Entretanto, como a diferenciação de variantes pode estar atrelada à deriva genética, é recomendado avaliar o índice de F_{ST} em janelas de SNPs, pois espera-se que a seleção altere a frequência de um conjunto de variantes ao invés de uma somente (51). Nos dados, apesar do SNP rs4971072 ser um *outlier* para a análise e o rs11264359 apresentar um valor próximo ao quartil de 90%, os valores por janelas não ultrapassam o limite para ser considerada uma janela *outlier*. Portanto, embora a análise apresente uma direção, não foi capaz de identificar intervalos de SNPs com diferenciação extrema que justifiquem a ocorrência de seleção.

Já a análise de homozigidade haplotípica entre populações, o xpEHH, trouxe resultados interessantes. Apesar da varredura de seleção não ocorrer sobre o gene *PKLR*, um grande bloco em LD foi observado entre os SNPs dos genes *HCN3* e *PKLR*. Nesse aspecto, caso alguns dos SNPs no gene *HCN3* desempenhem papel regulatório no gene *PKLR*, o sinal de seleção poderia ser uma explicação inicial. Segundo o projeto GTEx (312), as variantes sob seleção no gene *HCN3* são eQTLs relacionado ao *PKLR*, confirmando a correlação entre esses dois genes. Além disso, o gene *HCN3* codifica um canal voltagem-dependente de transporte de íons sódio e potássio que é muito expresso em células precursoras de eritrócitos. Os polimorfismos nesse gene são candidatos à seleção pela malária no trabalho de Machado e colaboradores, onde foi visto uma alta taxa de heterozigidade nesse gene relacionado com um fenótipo clínico de malária na população de Angola (67). O questionamento levantado a partir desses resultados é: por que o sinal de seleção vem atuando em um gene adjacente? A resposta não é fácil de se delinear. Especulamos que devido à importância do gene *PKLR* na fisiologia humana, alterações no *loci* genético podem ser danosas, enquanto as variantes em um gene adjacente podem exercer efeitos mais suaves no gene a favor da seleção.

O tamanho da região afetada pelo *sweep* depende da força da seleção e da velocidade em que o alelo selecionado alcança a frequência elevada. Geralmente, *sweeps* rápidos atingem grandes regiões (estima-se que uma vantagem seletiva de 1% leva à varredura de uma região de 600 Kb). Nesse caso, aumenta-se a chance de detecção de seleção, porém, dificulta a identificação do alelo causal (31). No nosso caso, observamos um *sweep* parcial, pois a região afetada é limitada a menos de 30 SNPs. Outro aspecto curioso é que o *sweep* parcial identificada no

HCN3/PKLR direciona o sinal para a seleção na Europa, e não na África, como inicialmente pensado. Isso torna a compreensão das assinaturas de seleção no gene ainda mais complexa. Ainda, no *dataset* de Gouveia e colaboradores, apesar do sinal de seleção para o rs11264359 ter sido identificado pelo xpEHH entre a população de Gana, endêmica para malária, e Sotho e Zulu, o alelo G (alelo de risco à hanseníase) apresentou-se aumentado em Sotho e Zulu, e não em Gana, como o esperado. Esses resultados combinados sugerem a ocorrência de seleção na região baseado no teste xpEHH, porém a direção da associação e a população onde ela está ocorrendo não está clara.

Detectar fortes evidências de seleção no genoma é raro, principalmente porque depende do método que está sendo empregado. A região do HLA, por exemplo, é classicamente estudada como região alvo de seleção por diferentes métodos (36,50,52). Contudo, até por envolver um *locus* extenso, altamente polimórfico e regulada por diferentes genes, ainda não foi possível detectar seleção para várias regiões relacionadas ao HLA. Um outro exemplo clássico é o gene da lactase, o *LCT*, que apresenta um dos sinais de seleção positivas mais fortes já estudados (47,311). Apesar disso, identificar alelos sob seleção até mesmo nesse gene é desafiador. Haplótipos longos em uma região de 98 kb ao redor do gene *LCT* são distribuídos em elevada frequência a nível global. A única diferença entre o haplótipo associado à persistência da lactase e o haplótipo ancestral é uma substituição G→C na posição 14.010 do gene (311). Outro obstáculo é que a maioria dos *scans* genômicos foram formulados para identificar assinaturas de seleção positiva, e poucos buscaram verificar seleção balanceadora. Um dos problemas é a detecção desse tipo de seleção em escala temporal (35). Sugere-se que a seleção balanceadora não é muito comum no genoma humano, porém deve ser mais frequente do que o estimado até hoje. Isso provavelmente se deve aos métodos e dados empregados que não permitem obter uma estimativa menos conservadora. Nesse aspecto, a fim de melhorar as técnicas para verificar seleção no genoma, muitos pesquisadores atuam modificando ou criando novos métodos baseado nas características genômicas e nas modificações esperadas pelo tipo de seleção a fim de melhorar os testes de seleção (36,41). Outro ponto importante de salientar é que os testes de seleção têm janela de tempo. Isso significa que se a mutação ocorreu há milhares de anos após a fixação da mutação vantajosa, alguns testes não irão evidenciar seleção, como o teste D de Tajima, por exemplo. Isso ocorre porque, com o passar do tempo, novas mutações podem surgir na região e a recombinação

gênica pode diluir a frequência de um alelo que fora fixado. Esse não é o caso da suposta seleção por malária, que ocorreu recentemente (5-7.000 anos) (31,49,80).

Neste trabalho, não houve aprofundamento em análises para detecção de seleção natural no gene *G6PD* devido à variedade de achados que confirmam tal hipótese. Tishkoff e colaboradores (2001), através da análise de microssatélites/RFLP, verificaram que há menor diversidade haplotípica em populações Africanas submetidas à malária, bem como altos níveis de LD nos cromossomos que contêm as variantes A- e Med, confirmando menor diversidade genética nesses cromossomos. Além disso, através de análises de coalescência, o mesmo estudo inferiu que a idade do alelo A- é de aproximadamente 6.350 anos (variando entre 3.840 – 11.760), consistente com a discussão do período temporal em que malária atuou como força seletiva na África, levando ao aumento dessas mutações (294). Sabeti e colaboradores (2002) também demonstraram, a partir de análise de EHH, que existe um longo braço de haplótipo em homozigidade relacionado ao alelo A da mutação G202A (que forma o alelo A-), e a idade estimada desse alelo foi de 2.325 anos (1.200 – 3.862) (1). Nesse estudo, métodos tradicionais como o D de Tajima e o D de Fu e Li não captaram a seleção, consistente com o baixo poder de detectar seleção recente desses testes. Salienta-se que o teste EHH foi inicialmente introduzido para detectar seleção positiva recente, e mesmo que esse não seja o regime principal na região do *G6PD* (estudos sugerem que esse gene é alvo de seleção balanceadora), o EHH foi capaz de identificar seleção. Apesar de Verrelli e colaboradores reportarem que o alelo A- não teve a sua origem recente (acredita-se que foi por volta de 100.000 anos), eles sugerem que o aumento na frequência desse alelo na África ocorreu devido a eventos recentes, como a malária, o que justifica a ausência dessa variante em populações não-Africanas (313).

Interessante que a seleção no gene *G6PD* afetou uma região >1.6 Mb do cromossomo X, demonstrando que a seleção pode ter profundo efeito na variabilidade nucleotídica ao redor de longas distâncias genômicas, mesmo em populações africanas. Além disso, esse padrão de LD em uma extensa região do gene ou é mais jovem do que as idades estimadas em outros estudos ou ocorreu devido à uma forte pressão seletiva (314). Ainda, apesar da vantagem evolutiva frente a malária das variantes em *G6PD*, elas não atingem fixação. Os alelos da deficiência são encontrados em frequências acima de 25% em determinadas populações, o que está aquém das frequências em direção a fixação, o que sugere que as mulheres homozigotas para as variantes

estariam em desvantagem (315). Portanto, atualmente, é bem aceita a hipótese de que a região do *G6PD* é alvo de seleção balanceadora.

1.3 – Aspectos funcionais das variantes do gene *G6PD* e a patogênese da hanseníase

Dados anteriores do nosso laboratório demonstraram que há uma intensa alteração no metabolismo de células de Schwann provocados pela infecção com o *M. leprae* (316). Foi visto que há um aumento do transportador de glicose, GLUT-1, aumento de enzimas produtoras de NADPH e de combate de ROS como a enzima málica citosólica (ME-1) e a glutatona redutase (GSR), e aumento de expressão de *G6PD* em células de Schwann da linhagem ST8814. Ainda, a infecção provocou a diminuição da liberação de lactato nessas células, da atividade da enzima de conversão a lactato (LDH-M) e do potencial elétrico das mitocôndrias. Esse trabalho concluiu que o *M. leprae* pode modular o metabolismo da glicose para aumentar a geração de poder redutor (NADPH), facilitando a regeneração de glutatona e conseqüentemente o controle dos radicais livres, favorecendo a infecção (316). Esses resultados foram compilados na Figura 1.2. Os dados em monócitos primários, fomentado na tese de doutoramento de Medeiros (2018), corroboram os achados anteriores, já que foi observado o aumento da expressão de *G6PD* e da atividade da enzima após a infecção em indivíduos sadios de forma independente da presença das mutações. A expressão de *G6PD* também está aumentada em lesões de pele e de nervos de pacientes com hanseníase, como demonstrado pelo nosso grupo. Além disso, a utilização do inibidor de *G6PD* (6-AN) foi capaz de reverter os efeitos vantajosos ao *M. leprae*, anteriormente encontrados, e diminuir a viabilidade do bacilo de forma significativa (288). Logo, dada a importância do *G6PD* na infecção, esse gene mereceu ser melhor explorado. Para isso, avaliamos a contribuição de variantes genéticas no gene *G6PD* para a predisposição à hanseníase.

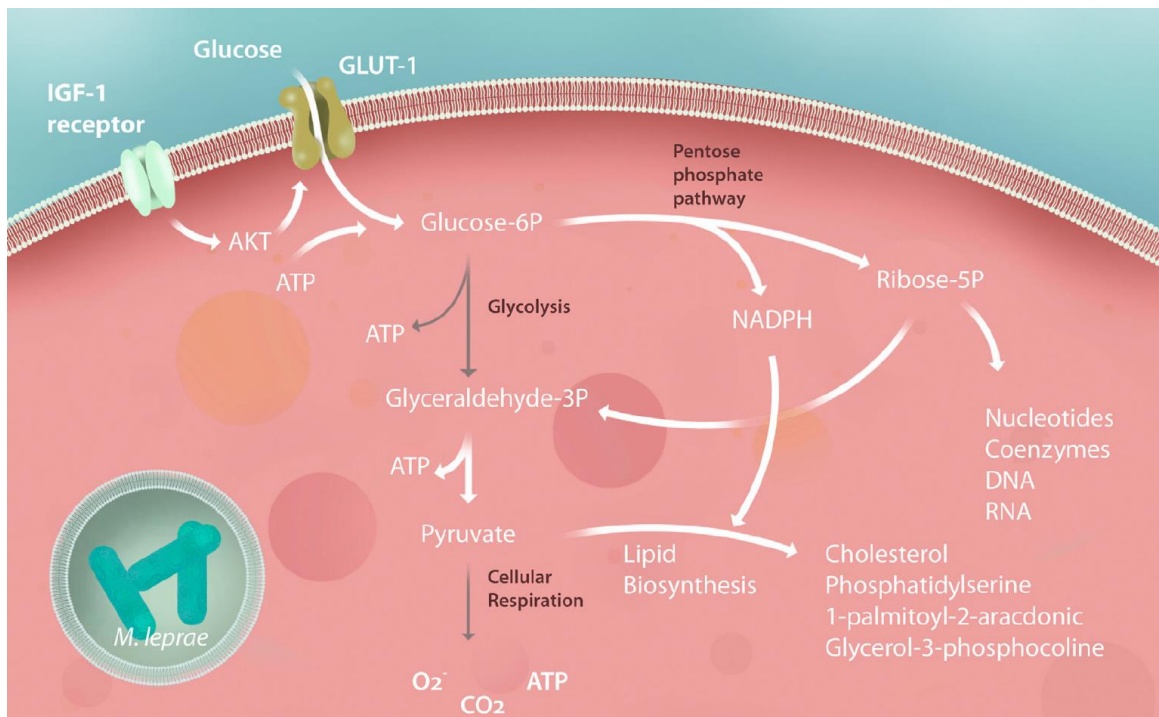


Figura 1.2 – Modulação do metabolismo de células de Schwann (CS) para a sobrevivência micobacteriana. Após a infecção, os transportadores do tipo GLUT-1 são superexpressados e ocorre aumento da captação de glicose via sinalização Akt. A glicose é direcionada para a via das pentoses (PPP) ao invés da glicólise (essa via é regulada negativamente nesse cenário). O piruvato gerado pelo PPP (via Ribose-5P) é rapidamente convertido para citrato e subseqüentemente convertido em lipídios, interrompendo o ciclo do ácido tricarbóxico, a respiração e o potencial de energia mitocondrial das células de Schwann. O NADPH gerado pela PPP mantém a produção de lipídeos necessário ao *M. leprae*, a produção de glutatona e a produção de nucleotídeos para a replicação micobacteriana (Adaptado de Toledo-Pinto, 2018 e Medeiros, 2016).

Além da variante intrônica (rs2472393) identificada pela metodologia de seleção de SNPs funcionais, as mutações G202A e A376G no gene *G6PD* (também apontadas como candidatas pela metodologia sugerida), que conferem deficiência pela variante A-, foram selecionadas como objeto deste estudo. Isso porque representam as mutações mais frequentes nas regiões africanas endêmicas para a malária, além de estarem relacionadas com o maior dano enzimático, apesar de existirem outras variantes genéticas associadas com a deficiência da enzima. Verrelli e colaboradores (2002) realizaram uma análise extensiva da variabilidade nucleotídica em 5.2 Kb do gene *G6PD* em populações africanas diversas e identificaram somente duas mutações de

deficiência grave comum em todas as populações, G202A e A376G (variante A-) (313). Essas variantes, e outras, se correlacionam com a distribuição geográfica do *Plasmodium falciparum* na África, corroborando a bem estabelecida hipótese de que a deficiência da G6PD reduz o risco de infecção pelo *P. falciparum* (315).

Considerando que o gene *G6PD* é alvo de pressão seletiva na África e a deficiência da glicose-6-fosfato-desidrogenase protege contra a malária, inicialmente levantou-se a hipótese (I) de que as variantes que levam à deficiência da G6PD poderiam estar relacionadas com a suscetibilidade à hanseníase, assim como as do gene *PKLR*. Isso porque a enzima G6PD deficiente poderia levar ao estresse oxidativo nas hemácias e interferir na meia-vida dessas células, promovendo a liberação de ferro e favorecendo o crescimento de bactérias intracelulares. Por outro lado, como os dados de Medeiros e colaboradores demonstram que a G6PD é crucial para a infecção pelo *M. leprae*, mutações que levam ao dano funcional da enzima podem prejudicar a sobrevivência do bacilo e aumentar a resistência à hanseníase, e, portanto, uma segunda hipótese (II) foi formulada. Entretanto, nenhuma das duas hipóteses construídas foi corroborada já que não houve associação dos polimorfismos estudados com a hanseníase, os resultados funcionais ou a viabilidade do bacilo (dado apresentado em Medeiros, 2018). Por fim, uma terceira hipótese (III) baseada nos dados globais foi proposta: embora as variantes genéticas de proteção contra a malária não estejam correlacionadas com a predisposição à hanseníase, o gene *G6PD* não deixa de ser relevante já que a bactéria modula significativamente a expressão e a atividade da enzima no processo da infecção. Todavia, no presente trabalho avaliamos a associação de apenas quatro variantes do *G6PD* com a hanseníase, e é possível que variantes na região promotora do gene que alterem o perfil de expressão e quantidade de enzima produzida alterem a predisposição à doença.

Em acordo com os resultados, a ausência de associação genética do gene *G6PD* com a hanseníase após o ajuste para sexo e ancestralidade indica que, a nível genético, as variantes estudadas nesse gene não apresentam relevância na suscetibilidade à doença. Cabe ressaltar que os indivíduos carreadores da variante A- do sexo masculino apresentam uma expressão basal aumentada do *G6PD* em relação aos não carreadores. Isso pode ser apontado como um possível mecanismo de compensação do efeito deletério na atividade da enzima, contrabalanceado pelo aumento da expressão. Tal característica foi observada de forma significativa somente em

homens porque nesses, devido a presença de somente um alelo, os efeitos fisiológicos são mais acentuados (76). Em contrapartida, como as mulheres heterozigotas apresentam o alelo selvagem e o alelo alternativo, o efeito deletério não é tão proeminente já que há um balanço entre os transcritos. Nesse caso, devido à inativação aleatória do cromossomo X, as mulheres são consideradas um mosaico celular, onde algumas células irão expressar a enzima com a mutação, porém outras células irão expressar a enzima normal (317). Essa diferença entre homens e mulheres é importante porque pode alterar o perfil de suscetibilidade à doença baseado na diferença de expressão entre os sexos. Como exemplo, foi visto que o haplótipo A- protege os homens hemizigotos contra a malária em Mali, no oeste africano, mas não as mulheres heterozigotas (318). Outros estudos já apontam que a proteção ocorre tanto para hemizigotos quanto para mulheres heterozigotas. E, ainda, estudos apontam que o maior *fitness* se dá para as mulheres heterozigotas (76,315).

O aumento da expressão de *G6PD* independente da variante A- após o estímulo com *M. leprae* em monócitos foi de acordo com os dados prévios encontrados em células de Schwann e em monócitos de voluntários sadios de Bauru estimulados com *M. leprae* vivo (dados não publicados). Portanto, é indiscutível a relevância da via das pentoses no processo de infecção. Porém, o mecanismo pelo qual o patógeno é capaz de utilizar estas modulações em prol da sua persistência na célula ou como estes podem interferir na patogenia da hanseníase ainda é obscuro, e, portanto, apresentamos algumas suposições. Sabe-se que o *M. leprae* induz a formação de corpúsculos lipídicos no macrófago, levando à famosa característica de *foamy macrophages* (do inglês, “macrófagos espumosos”) e que essas vesículas são indispensáveis para a sobrevivência intracelular desse patógeno (319). Esses corpúsculos são formados por colesterol da célula hospedeira sequestrados por receptores induzidos pelo bacilo (320). A produção de NADPH na via das pentoses é utilizado na biossíntese redutora de ácidos graxos e compostos esteróides. Assim, (I) acredita-se que o aumento da via das pentoses pode estar redirecionando NADPH para a produção de lipídeos, favorecendo o crescimento da bactéria. Essa hipótese é ainda favorecida por dados prévios que demonstraram que a diminuição dos níveis de lactato no sobrenadante das células infectadas também pode estar sendo direcionada para a produção de lipídeos. Resultados preliminares do Laboratório de Hanseníase verificaram, através de carbono radioativo, que o carbono proveniente do lactato fora detectado em glicerídeos. Nesse caso, é

possível que o carbono a ser exportado na forma de lactato esteja sendo reutilizado na exacerbada síntese de lipídeos sabidamente modulada pelo *M. leprae*. Ademais, (II) o NADPH atua na regeneração do sistema glutationa, diminuindo as espécies reativas de oxigênio (ROS) promovidas pela resposta do hospedeiro, criando um ambiente ainda mais favorável para a sobrevivência bacteriana. O *M. leprae* induz a produção de citocinas como IFN- γ e TNF, que por sua vez induzem o aumento na produção de ROS, potentes agentes microbicidas que atuam na morte celular de patógenos intracelulares (321,322). Logo, tanto a indução de lipídeos via NADPH e o controle de ROS é favorável à sobrevivência do bacilo. Além do disposto, (III) um outro subproduto da via das pentoses, a ribulose-5-fostato, é constituinte de nucleotídeos e participa na formação de D-arabinose, um componente da parede celular da micobactéria (arabinogalactana) (323). Então, o *M. leprae* também pode estar direcionando a via das pentoses a favor da produção de parede celular a partir de metabólitos do hospedeiro durante a sua replicação.

Vale ressaltar que o nosso estudo avaliou as alterações provocadas pelas mutações em monócitos, e não em células vermelhas como a literatura usualmente apresenta. As mutações em leucócitos não são muito estudadas e pouco se sabe sobre o nível da proteína nessas células mutantes, defendendo a posterior relevância desses ensaios. As mutações da variante A- provocam um mau dobramento da proteína e gera instabilidade celular, podendo levar à oxidação ou à inativação da enzima frente a agentes indutores de ROS (295,317,324). Não identificamos alteração no nível de expressão do gene *G6PD* e da atividade da enzima em comparação com os indivíduos carreadores das mutações, entretanto, como os monócitos, diferentemente das hemácias apresentam síntese proteica, eles podem estar fornecendo níveis ativos da enzima. Assim, conforme o observado, é provável que para manter os níveis ideais de enzima ativa, os indivíduos portadores do haplótipo A⁻ aumentem a expressão basal do *G6PD*.

2 – CONSIDERAÇÕES FINAIS

A proposta do atual estudo foi investigar a associação genética dos polimorfismos nos genes *PKLR* e *G6PD*, descritos como alvos de seleção pela malária, com a predisposição a doenças micobacterianas, tendo a hanseníase como modelo de doença. A tuberculose, assim como a hanseníase, é uma doença que acomete aproximadamente 10% dos indivíduos expostos ao bacilo e depende do *background* genético do hospedeiro. Vários genes de suscetibilidade (*IFNG*, *IL10* e *VDR*) em comum entre essas doenças já foram descritos (222,325,326). Os mecanismos de exposição, infecção e progressão do estado de latência à doença ativa são processos semelhantes entre as duas micobactérias. Portanto, é possível extrapolar que a associação genética identificada na hanseníase possa ter um efeito semelhante na tuberculose, apesar da necessidade de estudos mais profundos entre ambas as doenças.

Um dos critérios para validação de um gene candidato é que haja consistência entre a replicação observada em diferentes populações. O gene *PKLR* foi consistentemente associado com a hanseníase após os ajustes, através de replicações em brasileiros e de um estudo caso-controle independente em africanos avaliando o desfecho com a tuberculose. Ainda, a relevância do gene no desenvolvimento da hanseníase foi retratada em um estudo independente em uma população asiática. A vertente evolutiva do gene *PKLR* também foi explorada, sugerindo a ocorrência de seleção na região genômica, mas que provavelmente não foi tão extrema a ponto de ser detectada pelas metodologias e banco de dados utilizados. Embora a relação fenótipo-genótipo não tenha sido completamente desvendada da maneira empregada nesse trabalho, faz-se necessário continuar as análises em busca do entendimento do papel do *PKLR* com a hanseníase e tuberculose. Por fim, esse estudo confirma a associação do gene *PKLR* com a hanseníase através de uma hipótese que permite o estudo de mecanismos histórico-evolutivos, abrangendo o conhecimento do fluxo gênico entre populações, da variabilidade genética humana e suscetibilidade à patógenos, bem como a discussão de como a estrutura genética de diferentes populações pode influenciar no desenvolvimento de doenças infecciosas.

Por outro lado, o gene *G6PD* perdeu a associação genética após as correções por ancestralidade, demonstrando que esse gene tem pouco ou nenhum efeito no controle genético da hanseníase. Entretanto, a relevância da via das pentoses é inegável em vista dos resultados

encontrados. O destaque em relação ao *G6PD* está ligado à patogênese da hanseníase, uma vez que o *M. leprae* é capaz de modular os mecanismos da maquinaria celular a favor da sua sobrevivência na célula independente do componente genético (em relação ao gene *G6PD*). Trazemos algumas explicações no sentido de como o patógeno é capaz de utilizar-se de modulações celulares em prol da sua persistência no hospedeiro. A relevância desse gene em tuberculose, entretanto, ainda está para ser esclarecido. A discussão da via das pentoses pode, ainda, tanger a descoberta de alvos farmacológicos interferentes dessa via, sendo uma alternativa para o tratamento da hanseníase. Com isso, nossos dados exibem a complexidade dos componentes genéticos que influenciam no desenvolvimento das doenças micobacterianas em que existem genes que têm papel maior ou menor no caráter da predisposição à infecção, mas que também existem patógenos que subvertem mecanismos celulares e são capazes de burlar os efeitos genéticos e fisiológicos para manter a sua multiplicação.

O nosso estudo, além de fortalecer a hipótese de que a via das pentoses é crucial para o desenvolvimento do bacilo, contribuiu para o desvendamento de um novo gene (o *PKLR*) de suscetibilidade à hanseníase e tuberculose relacionado com a via do metabolismo energético, ainda não associada com a hanseníase no âmbito da predisposição genética.

3 – CONCLUSÕES

Baseado nos resultados demonstrados no presente estudo, podemos concluir que:

- A associação do gene *PKLR* com a suscetibilidade à hanseníase foi confirmada na população do Rio de Janeiro e validada em Salvador, mas não em Rondonópolis e Manaus, após o controle por ancestralidade genômica;
- A associação de suscetibilidade do gene *PKLR* com a tuberculose também foi confirmada em Moçambique;
- O haplótipo T/G/G (rs1052176/rs4971072/rs11264359) está associado com o risco de desenvolver doenças micobacterianas nas populações brasileiras que apresentam maior ancestralidade africana, Rio de Janeiro e Salvador, e em Moçambique;
- O estudo evolutivo sugere a presença de assinaturas de seleção na região do gene *PKLR* em variantes que parecem modular a expressão do gene, porém o regime ocorrido é complexo e merece ser estudado a partir de outras metodologias e bancos de dados populacionais;
- As variantes de risco do gene *PKLR* não alteram os níveis de ferro, ferritina e haptoglobina sérica de forma significativa em indivíduos sadios e pacientes com hanseníase, mas parecem modular sutilmente os níveis de ferritina e haptoglobina entre homens e mulheres. Além disso, as variantes modulam a expressão do gene *PKLR*.
- O gene *G6PD* não está associado com a suscetibilidade à hanseníase na população do Rio de Janeiro após o controle por ancestralidade genômica;
- A via da glicose-6-fosfato desidrogenase está aumentada na infecção pelo *M. leprae* e é crucial para a sobrevivência do bacilo.

4 – REFERÊNCIAS BIBLIOGRÁFICAS

1. Sabeti PCC, Reich DEE, Higgins JMM, Levine HZPZP, Richter DJJ, Schaffner SFF, et al. Detecting recent positive selection in the human genome from haplotype structure. *Nature* [Internet]. 2002;419(6909):832–837. Available from: <http://www.nature.com/nature/journal/v419/n6909/full/nature01140.html?free=2>
2. Quintana-murci L. Genetic and epigenetic variation of human populations: An adaptive tale. *Comptes rendus - Biol* [Internet]. 2016;339(7–8):278–83. Available from: <http://dx.doi.org/10.1016/j.crv.2016.04.005>
3. Barreiro LB, Quintana-Murci L. From evolutionary genetics to human immunology: How selection shapes host defence genes. *Nat Rev Genet* [Internet]. 2010;11(1):17–30. Available from: <http://dx.doi.org/10.1038/nrg2698>
4. Nédélec 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;167(3):657-669.e21.
5. Hedrick PW. Resistance to malaria in humans: the impact of strong, recent selection. *Malar J* [Internet]. 2012;11(1):349. Available from: *Malaria Journal*
6. Krishna R. Dronamraju. *Infectious Disease and Host-pathogen Evolution*. Cambridge University Press; 2004. 1–370 p.
7. Akide-Ndunge OB, Ayi K, Arese P. The Haldane malaria hypothesis: facts, artifacts, and a prophecy. *Redox Rep*. 2003;8(5):311–6.
8. Hedrick PW. Population genetics of malaria resistance in humans. *Heredity (Edinb)* [Internet]. 2011;107(4):283–304. Available from: <http://dx.doi.org/10.1038/hdy.2011.16>
9. Karlsson EK, Kwiatkowski DP, Sabeti PC. Natural selection and infectious disease in human populations. *Nat Rev Genet* [Internet]. 2014;15(6):379–93. Available from: <http://dx.doi.org/10.1038/nrg3734>
10. Mohandas N, An X. Malaria and human red blood cells. *Med Microbiol Immunol* [Internet]. 2012 Nov 11;201(4):593–8. Available from: <http://link.springer.com/10.1007/s00430-012-0272-z>
11. Snow RW, Omumbo JA, Lowe B, Molyneux CS, Obiero JO, Palmer A, et al. Relation between severe malaria morbidity in children and level of *Plasmodium falciparum* transmission in Africa. 1997;349:1650–4.
12. WHO. *World Malaria Report*. 2018. 210 p.
13. Munthali A. *Managing Malaria in Under-Five Children in a Rural Malawian Village*. *Nord J African*

- Stud. 2005;14(2):127–46.
14. Kwiatkowski DP. How malaria has affected the human genome and what human genetics can teach us about malaria. *Am J Hum Genet.* 2005;77(2):171–92.
 15. Allison AC. PROTECTION AFFORDED BY SICKLE- CELL TRAIT AGAINST SUBTERTIAN. 1954;290–4.
 16. Machado P, Mendes C, do Rosário VE, Arez AP. A contribuição dos polimorfismos humanos do eritrócito na proteção contra a malária. *Rev Pan-Amaz Saude.* 2010;1(4):85–96.
 17. Stuart MJ, Nagel RL, Jefferson T. Sickle-cell disease. *Lancet.* 2004;364(9442):1343–60.
 18. Weatherall DJ, Clegg JB. Public Health Reviews Inherited haemoglobin disorders : an increasing global health problem. 2001;79(8).
 19. Friedman MJ, Roth EF, Nagel RL, Trager W. Plasmodium falciparum: Physiological interactions with the human sickle cell. *Exp Parasitol.* 1979;47(1):73–80.
 20. Weatherall DJ. PHENOTYPE–GENOTYPE RELATIONSHIPS IN MONOGENIC DISEASE: LESSONS FROM THE THALASSAEMIAS. *Nat Rev Genet.* 2001;4:245–55.
 21. Modiano G, Morpurgo G, Terrenato L, Novelletto A, Rienzo A Di, Colombo B, et al. Protection against Malaria Morbidity : Near-Fixation of the α -Thalassemia Gene in a Nepalese Population. 1991;1925(1887):390–7.
 22. Horuk R. The Duffy antigen receptor for chemokines DARC / ACKR1. 2015;6(June):5–7.
 23. Chitnis CE, Chaudhuri A, Horuk R, Pogo AO, Miller LH. The Domain on the Duffy Blood Group Antigen for Binding Plasmodium vivax and P. knowlesi Malarial Parasites to Erythrocytes. *J Exp Med.* 1996;184:1531–6.
 24. Barnwell JW, Nichols ME, Rubinstein P. In vitro evaluation of the role of the Duffy blood group in erythrocyte invasion by Plasmodium vivax. *J Exp Med.* 1989;169(1795–1802).
 25. Ryan JR, Stoute JA, Amon J, Dunton RF, Mtalib R, Koros J, et al. Evidence for transmission of Plasmodium vivax among a Duffy antigen negative population in western Kenya. 2006;75(4):575–81.
 26. Cavasini CE, Mattos LC De, Ar Á, Couto DA, Sc V, Couto DA, et al. Duffy blood group gene polymorphisms among malaria vivax patients in four areas of the Brazilian Amazon region. *Malar J.* 2007;6(167):1–8.
 27. Langhi Jr DM, Bordin JO. Duffy blood group and malaria. *Hematology.* 2006;11(5):389–98.
 28. Jacobasch G, Rapoport SM. Hemolytic anemias due to erythrocyte enzyme deficiencies. *Mol Aspects Med* [Internet]. 1996 Apr;17(2):143–70. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/8813716>
 29. Wright S. Evolution in Mendelian Populations. *Genetics* [Internet]. 1931 Mar;16(2):97–159.

Available from: <http://www.ncbi.nlm.nih.gov/pubmed/17246615>

30. Akey JM, Eberle MA, Rieder MJ, Carlson CS, Shriver MD, Nickerson DA, et al. Population history and natural selection shape patterns of genetic variation in 132 genes. *PLoS Biol.* 2004;2(10).
31. Sabeti PC, Schaffner SF, Fry B, Lohmueller J, Varilly P, Shamovsky O, et al. Positive natural selection in the human lineage. *Science (80-)*. 2006;312(5780):1614–20.
32. Hernandez RD, Kelley JL, Elyashiv E, Melton SC, Auton A, McVean G, et al. Classic selective sweeps were rare in recent human evolution. *Science (80-)*. 2011;331(6019):920–4.
33. McCoy RC, Akey JM. Selection plays the hand it was dealt: Evidence that human adaptation commonly targets standing genetic variation. *Genome Biol.* 2017;18(1):1–4.
34. Nielsen R. Molecular Signatures of Natural Selection. *Annu Rev Genet.* 2005;39(1):197–218.
35. Andrés AM, Hubisz MJ, Indap A, Torgerson DG, Degenhardt JD, Boyko AR, et al. Targets of balancing selection in the human genome. *Mol Biol Evol.* 2009;26(12):2755–64.
36. Bitarello BD, De Filippo C, Teixeira JC, Schmidt JM, Kleinert P, Meyer D, et al. Signatures of long-term balancing selection in human genomes. *Genome Biol Evol.* 2018;10(3):939–55.
37. De Filippo C, Key FM, Ghirotto S, Benazzo A, Meneu JR, Weihmann A, et al. Recent Selection Changes in Human Genes under Long-Term Balancing Selection. *Mol Biol Evol.* 2016;33(6):1435–47.
38. Asthana S, Schmidt S, Sunyaev S. A limited role for balancing selection. *Trends Genet [Internet]*. 2005 Jan;21(1):30–2. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0168952504003038>
39. Bubb KL, Bovee D, Buckley D, Haugen E, Kibukawa M, Paddock M, et al. Scan of human genome reveals no new loci under ancient balancing selection. *Genetics.* 2006;173(4):2165–77.
40. Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, et al. Genome-wide detection and characterization of positive selection in human populations. *Nature.* 2007;449(7164):913–8.
41. Bitarello BD. Seleção balanceadora no genoma humano: relevância biológica e consequências deletérias. Universidade de São Paulo; 2016.
42. Nielsen R, Hellmann I, Hubisz M, Bustamante C, Clark AG. Recent and ongoing selection in the human genome. *Nat Rev Genet [Internet]*. 2007 Nov;8(11):857–68. Available from: <http://www.nature.com/articles/nrg2187>
43. Holsinger KE, Weir BS. Genetics in geographically structured populations: Defining, estimating and interpreting F_{ST} . *Nat Rev Genet.* 2009;10(9):639–50.
44. Voight BF, Kudaravalli S, Wen X, Pritchard JK. A map of recent positive selection in the human genome. *PLoS Biol.* 2006;4(3):0446–58.

45. Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZXP, Pool JE, et al. Sequencing of 50 human exomes reveals adaptation to high altitude. *Science* (80-). 2010;329(5987):75–8.
46. Carvajal-Rodriguez A. Simulation of Genomes: A Review. *Curr Genomics*. 2008;9(3):155–9.
47. Bersaglieri T, Sabeti PC, Patterson N, Vanderploeg T, Schaffner SF, Drake JA, et al. Genetic signatures of strong recent positive selection at the lactase gene. *Am J Hum Genet*. 2004;74(6):1111–20.
48. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* [Internet]. 1989 Nov;123(3):585–95. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/2513255>
49. Kreitman M. Methods to Detect Selection in Populations with Applications to the Human. *Annu Rev Genomics Hum Genet*. 2000;
50. Bitarello BD, Francisco R dos S, Meyer D. Heterogeneity of dN/dS Ratios at the Classical HLA Class I Genes over Divergence Time and Across the Allelic Phylogeny. *J Mol Evol*. 2016;82(1):38–50.
51. Vitti JJ, Grossman SR, Sabeti PC. Detecting Natural Selection in Genomic Data. *Annu Rev Genet*. 2013;47(1):97–120.
52. Brandt D, Cesar J, Goudet J, Meyer D. Population Differentiation at the HLA Genes. 2017;214668.
53. Booker TR, Jackson BC, Keightley PD. Detecting positive selection in the genome. *BMC Biol*. 2017;15(1):1–10.
54. Nelson DL, Cox MM. *Lehninger Principles of Biochemistry*. 3^a. Worth Publishers; 2000.
55. Zanella A, Fermo E, Bianchi P, Valentini G. Red cell pyruvate kinase deficiency: Molecular and clinical aspects. *Br J Haematol*. 2005;130(1):11–25.
56. Berghout J, Higgins S, Loucoubar C, Sakuntabhai A, Kain KC, Gros P. Genetic diversity in human erythrocyte pyruvate kinase. *Genes Immun* [Internet]. 2012;13(1):98–102. Available from: <http://dx.doi.org/10.1038/gene.2011.54>
57. Brown KA. Erythrocyte metabolism and enzyme defects. *Lab Med*. 1996;27(5):329–33.
58. Zanella A, Fermo E, Bianchi P, Roberto L, Valentini G. Pyruvate kinase deficiency : The genotype-phenotype association. 2007;217–31.
59. Etiemble J, Picat C, Boivin P, Sanguines C, Inserm U, Era C, et al. A Red Cell Pyruvate Kinase Mutant with Normal L-Type PK in the Liver. 1982;256–8.
60. Grace RF, Zanella A, Neufeld EJ, Morton DH, Eber S, Yaish H, et al. Erythrocyte pyruvate kinase deficiency: 2015 status report. *Am J Hematol*. 2015;90(9):825–30.
61. Beutler E, Gelbart T. Estimating the prevalence of pyruvate kinase deficiency from the gene frequency in the general white population. *Blood*. 2000;95(11):3585–8.

62. Zanella A, Berzuini A, Colombo MB, Guffanti A, Lecchi L, Poli F, et al. Iron status in red cell pyruvate kinase deficiency: study of Italian cases. *Br J Haematol*. 1993;83(3):485–90.
63. Min-Oo G, Fortin A, Tam M-F, Nantel A, Stevenson MM, Gros P. Pyruvate kinase deficiency in mice protects against malaria. *Nat Genet* [Internet]. 2003;35(4):357–62. Available from: <http://www.nature.com/doi/10.1038/ng1260>
64. Ayi K, Liles WC, Gros P, Kain KC. Adenosine Triphosphate Depletion of Erythrocytes Simulates the Phenotype Associated with Pyruvate Kinase Deficiency and Confers Protection against *Plasmodium falciparum* In Vitro. *J Infect Dis* [Internet]. 2009;200(8):1289–99. Available from: <https://academic.oup.com/jid/article-lookup/doi/10.1086/605843>
65. Ayi K, Min-Oo G, Serghides L, Crockett M, Kirby-Allen M, Quirt I, et al. Pyruvate Kinase Deficiency and Malaria. *N Engl J Med* [Internet]. 2008;358(17):1805–10. Available from: <http://www.nejm.org/doi/abs/10.1056/NEJMoa072464>
66. Alves J, Machado P, Silva J, Gonçalves N, Ribeiro L, Faustino P, et al. Analysis of malaria associated genetic traits in Cabo Verde, a melting pot of European and sub Saharan settlers. *Blood Cells, Mol Dis* [Internet]. 2010;44(1):62–8. Available from: <http://dx.doi.org/10.1016/j.bcmed.2009.09.008>
67. Machado P, Pereira R, Rocha AM, Manco L, Fernandes N, Miranda J, et al. Malaria: Looking for selection signatures in the human PKLR gene region. *Br J Haematol*. 2010;149(5):775–84.
68. Machado P, Manco L, Gomes C, Mendes C, Fernandes N, Salomé G, et al. Pyruvate Kinase Deficiency in Sub-Saharan Africa: Identification of a Highly Frequent Missense Mutation (G829A;Glu277Lys) and Association with Malaria. *PLoS One*. 2012;7(10).
69. Van Bruggen R, Gualtieri C, Iliescu A, Cheepsunthorn CL, Mungkalasut P, Trape JF, et al. Modulation of malaria phenotypes by pyruvate kinase (pklr) variants in a Thai population. *PLoS One* [Internet]. 2015;10(12):1–18. Available from: <http://dx.doi.org/10.1371/journal.pone.0144555>
70. Min-oo G, Tam M, Ponka P, Stevenson MM, Gros P. The mouse Char10 locus regulates severity of pyruvate kinase deficiency and susceptibility to malaria. 2017;1–18.
71. Faik I, Tong H Van, Lell B, Meyer CG, Kremsner PG, Velavan TP. Pyruvate Kinase and Fcy Receptor Gene Copy Numbers Associated With Malaria Phenotypes. 2017;216.
72. Zanella A, Bianchi P. Red cell pyruvate kinase deficiency: From genetics to clinical manifestations. *Best Pract Res Clin Haematol*. 2000;13(1):57–81.
73. Williams TN. Human red blood cell polymorphisms and malaria. *Curr Opin Microbiol*. 2006;9(4):388–94.
74. Min-Oo G, Gros P. Erythrocyte variants and the nature of their malaria protective effect. *Cell Microbiol*. 2005;7(6):753–63.

75. Beutler E. G6PD deficiency. *Blood*. 1994;84(11):3613–36.
76. Luzzato L. Glucose 6-phosphate dehydrogenase deficiency: from genotype to phenotype. *Haematologica*. 2006;91(10):1303–6.
77. Longo L, Vanegas OC, Patel M, Rosti V, Li H, Waka J, et al. Maternally transmitted severe glucose 6-phosphate dehydrogenase deficiency is an embryonic lethal. *EMBO J*. 2002;21(16):4229–39.
78. Notaro R, Afolayan A, Luzzatto L. Human mutations in glucose 6-phosphate dehydrogenase reflect evolutionary history. *FASEB J*. 2000;14(3):485–94.
79. Reich DE, Reich DE, Cargill M, Cargill M, Bolk S, Bolk S, et al. Linkage disequilibrium in the human genome. *Nature* [Internet]. 2001;411(6834):199–204. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/11346797>
80. Tishkoff SA, Williams SM. Genetic analysis of African populations: Human evolution and complex disease. *Nat Rev Genet*. 2002;3(8):611–21.
81. Howes RE, Dewi M, Hogg MM, Battle KE, Padilla CD, Baird JK, et al. G6PD Deficiency Prevalence and Estimates of Affected Populations in Malaria Endemic Countries : A Geostatistical Model-Based Map. 2012;9(11).
82. May J, Meyer CG, Großterlinden L, Ademowo OG, Mockenhaupt FP, Olumese PE, et al. Red cell glucose-6-phosphate dehydrogenase status and pyruvate kinase activity in a Nigerian population. *Trop Med Int Heal*. 2000;5(2):119–23.
83. Gómez-Gallego F, Garrido-Pertierra A, Bautista JM. Structural Defects Underlying Protein Dysfunction in Human Glucose-6-phosphate Dehydrogenase A⁻ Deficiency. *J Biol Chem*. 2000 Mar;275(13):9256–62.
84. Minucci A, Moradkhani K, Hwang MJ, Zuppi C, Giardina B, Capoluongo E. Glucose-6-phosphate dehydrogenase (G6PD) mutations database: Review of the “old” and update of the new mutations. *Blood Cells, Mol Dis*. 2012;48(3):154–65.
85. Calabro V, Mason PJ, Filosa S, Civitelli D, Cittadella R, Tagarelli A, et al. Genetic heterogeneity of glucose-6-phosphate dehydrogenase deficiency revealed by single-strand conformation and sequence analysis. *Am J Hum Genet*. 1993;52(3):527–36.
86. Pereira LLMD, Bravin CA, Cintra TS, Cassa WSP, Santos TA, Fonseca A, et al. Prevalence of G6PD deficiency and molecular characterization of G202A, A376G and C563T polymorphisms in newborns in Southeastern Brazil. *Einstein (São Paulo)* [Internet]. 2019 Jan 17;17(1):1–10. Available from: <https://journal.einstein.br/article/prevalence-of-g6pd-deficiency-and-molecular-characterization-of-g202a-a376g-and-c563t-polymorphisms-in-newborns-in-southeastern-brazil/>
87. Luzzatto L, Nannelli C, Notaro R. Glucose-6-Phosphate Dehydrogenase Deficiency. *Hematol Oncol*

- Clin North Am [Internet]. 2016 Apr;30(2):373–93. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0889858815001938>
88. Chernof CD. Dapsone-Induced Hemolysis in G6PD Deficiency. 2015;5–7.
 89. Avalos S, Mejia RE, Banegas E, Salinas C, Gutierrez L, Fajardo M, et al. G6PD deficiency , primaquine treatment , and risk of haemolysis in malaria - infected patients. 2018;1–11.
 90. BEUTLEK E, YEH M, FAIRBANKS VF. The Normal Human Female As a Mosaic of X-Chromosome Activity: : studies using the gene FOR G-6-PD-deficiency as a marker. *Obstet Gynecol Surv.* 1962;17(5):763.
 91. Uyoga S, Ndila CM, Macharia AW, Nyutu G, Shah S, Peshu N, et al. Glucose-6-phosphate dehydrogenase deficiency and the risk of malaria and other diseases in children in Kenya: A case-control and a cohort study. *Lancet Haematol.* 2015;2(10):e437–44.
 92. Luzzatto L, Usanga EA, Reddy S. Glucose-6-Phosphate Dehydrogenase Deficient Red Cells: Resistance to Infection by Malarial Parasites. *Science (80-)* [Internet]. 1969 May 16;164(3881):839–42. Available from: <http://www.sciencemag.org/cgi/doi/10.1126/science.164.3881.839>
 93. Malaria Genomic Epidemiology Network. Reappraisal of known malaria resistance loci in a large multicenter study. *Nat Genet* [Internet]. 2014 Nov 28;46(11):1197–204. Available from: <http://www.nature.com/articles/ng.3107>
 94. Lastória JC, de Abreu MAMM. Leprosy: A review of laboratory and therapeutic aspects - Part 2. *An Bras Dermatol.* 2014;89(3):389–401.
 95. White C, Franco-Paredes C. Leprosy in the 21st century. *Clin Microbiol Rev.* 2015;28(1):80–94.
 96. Grzybowski A, Sak J, Pawlikowski J, Iwanowicz-Palus G. Gerhard Henrik Armauer Hansen (1841-1912)-The 100th anniversary of the death of the discoverer of *Mycobacterium leprae*. *Clin Dermatol.* 2013;31(5):653–5.
 97. Hulse E. Leprosy and Ancient Egypt. *Lancet* [Internet]. 1972 Nov;300(7785):1024–5. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0140673672924336>
 98. Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* 1987;16(3):1215.
 99. Molto JE. Humerus varus deformity in Roman period burials from Kellis 2, Dakhleh, Egypt. *Am J Phys Anthropol* [Internet]. 2000 Sep;113(1):103–9. Available from: <http://doi.wiley.com/10.1002/1096-8644%28200009%29113%3A1%3C103%3A%3AAID-AJPA9%3E3.0.CO%3B2-A>
 100. Robbins G, Mushrif Tripathy V, 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):1–8.

101. Tayles N, Buckley HR. Leprosy and tuberculosis in iron age Southeast Asia? *Am J Phys Anthropol.* 2004;125(3):239–56.
102. Bouth RC, Baptista IMFD, Loiseau C, Miyamoto Y, Girma S, Kodio M, et al. Phylogenomics and antimicrobial resistance of the leprosy bacillus *Mycobacterium leprae*. *Nat Commun.* 2018;9(1).
103. Monot M. On the Origin of Leprosy. *Science* (80-) [Internet]. 2005 May 13;308(5724):1040–2. Available from: <http://www.sciencemag.org/cgi/doi/10.1126/science/1109759>
104. Ministério da Saúde B. Caracterização da situação epidemiológica da hanseníase e diferenças por sexo, Brasil, 2012-2016. *Bol epidemiológico* [Internet]. 2018;49(4):1–12. Available from: <http://portalarquivos2.saude.gov.br/images/pdf/2018/janeiro/31/2018-004-Hanseniasse-publicacao.pdf>
105. Fogel N. Tuberculosis: A disease without boundaries. *Tuberculosis* [Internet]. 2015 Sep;95(5):527–31. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S1472979214206950>
106. Scollard DM, Adams LB, Gillis TP, Krahenbuhl JL, Truman RW, Williams DL. The continuing challenges of leprosy. *Clin Microbiol Rev.* 2006;19(2):338–81.
107. Shepard CC, Chang YT. Effect of Several Anti-Leprosy Drugs on Multiplication of Human Leprosy Bacilli in Foot-Pads of Mice. *Exp Biol Med* [Internet]. 1962 Mar 1;109(3):636–8. Available from: <http://ebm.sagepub.com/lookup/doi/10.3181/00379727-109-27293>
108. Kirchheimer WF. Experimental leprosy in the nine banded armadillo. *Public Health Rep.* 1975;90(6):483–5.
109. Truman R. Leprosy in wild armadillos. *Lepr Rev.* 2005;76(3):198–208.
110. Sharma R, Lahiri R, Scollard DM, Pena M, Williams DL, Adams LB, et al. The armadillo: A model for the neuropathy of leprosy and potentially other neurodegenerative diseases. *DMM Dis Model Mech.* 2013;6(1):19–24.
111. Tortoli E. Microbiological features and clinical relevance of new species of the genus *Mycobacterium*. *Clin Microbiol Rev.* 2014;27(4):727–52.
112. Hunter SW, Brennan PJ. A novel phenolic glycolipid from *Mycobacterium leprae* possibly involved in immunogenicity and pathogenicity. *J Bacteriol.* 1981;147(3):728–35.
113. Spencer JS, Brennan PJ. The role of *Mycobacterium leprae* phenolic glycolipid I (PGL-I) in serodiagnosis and in the pathogenesis of leprosy. *Lepr Rev.* 2011;82(4):344–57.
114. Alcaïs A, Mira M, Casanova JL, Schurr E, Abel L. Genetic dissection of immunity in leprosy. *Curr Opin Immunol.* 2005;17(1):44–8.
115. Singh P, Cole ST. *Mycobacterium leprae*: genes, pseudogenes and genetic diversity. *Future Microbiol* [Internet]. 2011 Jan;6(1):57–71. Available from:

<https://www.futuremedicine.com/doi/10.2217/fmb.10.153>

116. Cole ST, Eiglmeier K, Parkhill J, James KD, Thomson NR, Wheeler PR, et al. Massive gene decay in the leprosy bacillus. *Nature*. 2001;409(6823):1007–11.
117. Monot M, Honoré N, Garnier T, Zidane N, Sherafi D, Paniz-Mondolfi A, et al. Comparative genomic and phylogeographic analysis of *Mycobacterium leprae*. *Nat Genet*. 2009;41(12):1282–9.
118. Spencer JS, May GS, Seo Y-H, Schoberle T, Han XY, Li W, et al. A New *Mycobacterium* Species Causing Diffuse Lepromatous Leprosy. *Am J Clin Pathol*. 2008;130(6):856–64.
119. Scollard DM. Infection with *Mycobacterium lepromatosis*. *Am J Trop Med Hyg* [Internet]. 2016;95(3):500–1. Available from: <https://www.futuremedicine.com/doi/10.2217/fmb.10.153>
120. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, et al. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*. 1998;393(6685):537–44.
121. Cole ST. Comparative mycobacterial genomics. *Curr Opin Microbiol*. 1998;1(5):567–71.
122. Singh P, Benjak A, Schuenemann VJ, Herbig A, Avanzi C, Busso P, et al. Insight into the evolution and origin of leprosy bacilli from the genome sequence of *Mycobacterium lepromatosis*. *Proc Natl Acad Sci U S A*. 2015;112(14):4459–64.
123. Moraes MO, Cardoso CC, Vanderborght PR, Pacheco AG. Genetics of host response in leprosy. *Lepr Rev*. 2006;77(3):189–202.
124. Eichelmann K, González SEG. Leprosy. An Update: Definition, Pathogenesis, Classification, Diagnosis, and Treatment. *PLoS One*. 2013;104(7):554–63.
125. Ridley DS, Jopling WH. Classification of leprosy according to immunity. A five-group system. *Int J Lepr Other Mycobact Dis Off Organ Int Lepr Assoc*. 1966 Sep;34(3):255–73.
126. Jardim MR, Antunes SLG, Santos AR, Nascimento OJM, Nery JAC, Sales AM, et al. Criteria for diagnosis of pure neural leprosy. *J Neurol*. 2003;250(7):806–9.
127. Sauer MED, Salomão H, Ramos GB, D'Espindula HRS, Rodrigues RS a., Macedo WC, et al. Genetics of leprosy: Expected and unexpected developments and perspectives. *Clin Dermatol* [Internet]. 2015;33(1):99–107. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0738081X14002296>
128. Fava VM, Dallmann-Sauer M, Schurr E. Genetics of leprosy: today and beyond. *Hum Genet* [Internet]. 2019;(0123456789):1–12. Available from: <http://link.springer.com/10.1007/s00439-019-02087-5>
129. Yamamura M, Uyemura K, Deans RJ, Weinberg K, Rea TH, Bloom BR, et al. Defining protective responses to pathogens: Cytokine profiles in leprosy lesions. *Science* (80-). 1991;254(5029):277–

- 9.
130. Hungria EM, Bühner-Sékula S, de Oliveira RM, Aderaldo LC, Pontes A de A, Cruz R, et al. Leprosy reactions: The predictive value of *Mycobacterium leprae*-specific serology evaluated in a Brazilian cohort of leprosy patients (U-MDT/CT-BR). *PLoS Negl Trop Dis*. 2017;11(2):1–17.
 131. Moraes MO, Sarno EN, Almeida a. S, Saraiva BCC, Nery J a C, Martins RCL, et al. Cytokine mRNA expression in leprosy: A possible role for interferon- γ and interleukin-12 in reactions (RR and ENL). *Scand J Immunol*. 1999;50(5):541–9.
 132. 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. *Am J Trop Med Hyg*. 2008;78(3):518–21.
 133. Barreto JG, Guimarães L de S, Leão MRN, Ferreira DVG, de Araújo Lima RA, Salgado CG. Anti-PGL-I seroepidemiology in leprosy cases: Household contacts and school children from a hyperendemic municipality of the Brazilian Amazon. *Lepr Rev*. 2011;82(4):358–70.
 134. da Costa Martins AC, Miranda A, de Oliveira MLW del R, Bühner-Sékula S, Martinez A. Nasal mucosa study of leprosy contacts with positive serology for the phenolic glycolipid 1 antigen. *Braz J Otorhinolaryngol*. 2010;76(5):579–87.
 135. Fine PEM, Sterne JAC, Pönnighaus JM, Bliss L, Saul J, Chihana A, et al. Household and dwelling contact as risk factors for leprosy in northern Malawi. *Am J Epidemiol*. 1997;146(1):91–102.
 136. Coppola M, van den Eeden SJF, Robbins N, Wilson L, Franken KLMC, Adams LB, et al. Vaccines for leprosy and tuberculosis: Opportunities for shared research, development, and application. *Front Immunol*. 2018;9(FEB):1–12.
 137. Richardus RA, Butlin CR, Alam K, Kundu K, Geluk A, Richardus JH. Clinical manifestations of leprosy after BCG vaccination: An observational study in Bangladesh. *Vaccine* [Internet]. 2015;33(13):1562–7. Available from: <http://dx.doi.org/10.1016/j.vaccine.2015.02.017>
 138. White C. Sociocultural considerations in the treatment of leprosy in Rio de Janeiro, Brazil. *Lepr Rev*. 2002;73(4):356–65.
 139. Cocito C, Delville J. Biological, chemical, immunological and staining properties of bacteria isolated from tissues of leprosy patients. *Eur J Epidemiol*. 1985;1(3):202–31.
 140. Massone C, Nunzi E, Cerroni L. Histopathologic diagnosis of leprosy in a nonendemic area. *Am J Dermatopathol*. 2010;32(4):417–9.
 141. Antunes SLG, Chimelli L, Jardim MR, Vital RT, Nery JA da C, Corte-Real S, et al. Histopathological examination of nerve samples from pure neural leprosy patients: obtaining maximum information to improve diagnostic efficiency. *Mem Inst Oswaldo Cruz* [Internet]. 2012 Mar;107(2):246–53.

Available from: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0074-02762012000200015&lng=en&tlng=en

142. De Moura RS, Calado KL, Oliveira MLW, Bühner-Sékula S. Sorologia da hanseníase utilizando PGL-I: Revisão sistemática. *Rev Soc Bras Med Trop*. 2008;41(SUPPL. 2):11–8.
143. de Souza MM, Netto EM, Nakatani M, Duthie MS. Utility of recombinant proteins LID-1 and PADL in screening for *Mycobacterium leprae* infection and leprosy. *Trans R Soc Trop Med Hyg*. 2014;108(8):495–501.
144. Martinez AN, Ribeiro-Alves M, Sarno EN, Moraes MO. Evaluation of qPCR-Based assays for leprosy diagnosis directly in clinical specimens. *PLoS Negl Trop Dis*. 2011;5(10):1–8.
145. Santos AR, Nery JC, Duppre NC, Gallo MEN, Filho JTG, Suffys PN, et al. Use of the polymerase chain reaction in the diagnosis of leprosy. *J Med Microbiol*. 1997;46(2):170–2.
146. Moschella SL. An update on the diagnosis and treatment of leprosy. *J Am Acad Dermatol*. 2004;51(3):417–26.
147. Cruz RC da S, Bühner-Sékula S, Penna MLF, Penna GO, Talhari S. Leprosy: current situation, clinical and laboratory aspects, treatment history and perspective of the uniform multidrug therapy for all patients. *An Bras Dermatol* [Internet]. 2017 Dec;92(6):761–73. Available from: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0365-05962017000600761&lng=en&tlng=en
148. Haslett PAJ, Roche P, Butlin CR, Macdonald M, Shrestha N, Manandhar R, et al. Effective Treatment of Erythema Nodosum Leprosum with Thalidomide Is Associated with Immune Stimulation. *J Infect Dis*. 2005;192(12):2045–53.
149. Walker SL. New therapeutics for leprosy. *Br J Dermatol* [Internet]. 2019 Oct 16;bjd.18519. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/bjd.18519>
150. Linder K, Zia M, Kern W V., Pfau RKM, Wagner D. Relapses vs. reactions in multibacillary leprosy: Proposal of new relapse criteria. *Trop Med Int Heal*. 2008;13(3):295–309.
151. Pandey M, Talwar S, Bose S, Pandey AK. Iron homeostasis in *Mycobacterium tuberculosis* is essential for persistence. *Sci Rep*. 2018;8(1):1–9.
152. Cassat JE, Skaar EP. Iron in infection and immunity. *Cell Host Microbe*. 2013;13(5):509–19.
153. Ganz T, Nemeth E. Iron homeostasis in host defence and inflammation. *Nat Rev Immunol* [Internet]. 2015 Aug 10;15(8):500–10. Available from: <http://www.nature.com/articles/nri3863>
154. Dussurget O, Smith I. Interdependence of mycobacterial iron regulation, oxidative-stress response and isoniazid resistance. *Trends Microbiol*. 1998;6(9):354–8.
155. Voss JJ De, Rutter K, Schroeder BG, Iii CEB, Voss JJDE. Iron Acquisition and Metabolism by Iron

- Acquisition and Metabolism by Mycobacteria. 1999;181(15):4443–51.
156. Boelaert JR, Vandecasteele SJ, Appelberg R, Gordeuk VR. The Effect of the Host's Iron Status on Tuberculosis. *J Infect Dis.* 2007;195(12):1745–53.
 157. Murray MJ, Murray AB. Adverse effect of iron repletion on infection. *Clin Res.* 1978;26(3):1113–5.
 158. Douvas GS, May MH, Crowle AJ. Transferrin, Iron, and Serum Lipids Enhance or Inhibit *Mycobacterium avium* Replication in Human Macrophages. *J Infect Dis.* 1993;167(4):857–764.
 159. Madigan CA, Cheng TY, Layre E, Young DC, McConnell MJ, Debono CA, et al. Lipidomic discovery of deoxysiderophores reveals a revised mycobactin biosynthesis pathway in *Mycobacterium tuberculosis*. *Proc Natl Acad Sci U S A.* 2012;109(4):1257–62.
 160. Luo M, Fadeev EA, Groves JT. Mycobactin-mediated iron acquisition within macrophages. *Nat Chem Biol.* 2005;1(3):149–53.
 161. De Voss JJ, Rutter K, Schroeder BG, Su H, Zhu Y, Barry CE. The salicylate-derived mycobactin siderophores of *Mycobacterium tuberculosis* are essential for growth in macrophages. *Proc Natl Acad Sci U S A.* 2000;97(3):1252–7.
 162. Rodriguez GM. Control of iron metabolism in *Mycobacterium tuberculosis*. *Trends Microbiol.* 2006;14(7):320–7.
 163. Gobin J. Exochelins of *Mycobacterium tuberculosis* remove iron from human iron-binding proteins and donate iron to mycobactins in the *M. tuberculosis* cell wall. *J Exp Med* [Internet]. 1996 Apr 1;183(4):1527–32. Available from: <http://www.jem.org/cgi/doi/10.1084/jem.183.4.1527>
 164. Olakanmi O, Schlesinger LS, Ahmed A, Britigan BE. The Nature of Extracellular Iron Influences Iron Acquisition by *Mycobacterium tuberculosis* Residing within Human Macrophages. *Infect Immun.* 2004;72(4):2022–8.
 165. Boradia VM, Malhotra H, Thakkar JS, Tillu VA, Vuppala B, Patil P, et al. *Mycobacterium tuberculosis* acquires iron by cell-surface sequestration and internalization of human holo-transferrin. *Nat Commun.* 2014;5.
 166. Tullius M V., Harmston CA, Owens CP, Chim N, Morse RP, McMath LM, et al. Discovery and characterization of a unique mycobacterial heme acquisition system. *Proc Natl Acad Sci U S A.* 2011;108(12):5051–6.
 167. Olakanmi O, Britigan BE, Schlesinger LS. Gallium disrupts iron metabolism of mycobacteria residing within human macrophages. *Infect Immun.* 2000;68(10):5619–27.
 168. Van Zandt KE, Sow FB, Florence WC, Zwilling BS, Satoskar AR, Schlesinger LS, et al. The iron export protein ferroportin 1 is differentially expressed in mouse macrophage populations and is present in the mycobacterial-containing phagosome. *J Leukoc Biol.* 2008;84(3):689–700.

169. Lambrecht, 1993 Inability to detect mycobactin in Mycobacteria infected tissues suggests an alternative iron acquisition mechanism by Mycobacteria in vivo.pdf.
170. Momotani E, Wuscher N, Ravisse P, Rastogi N. Immunohistochemical identification of ferritin, lactoferrin and transferrin in leprosy lesions of human skin biopsies. *J Comp Pathol.* 1992;106(3):213–20.
171. Hall RM, Wheeler PR, Ratledge C. Exochelin-mediated iron uptake into *Mycobacterium leprae*. *Int J Lepr.* 1983;51(4):490.
172. Sharman GJ, Williams DH, Ewing DF, Ratledge C. Determination of the structure of exochelin MN, the extracellular siderophore from *Mycobacterium neoaurum*. *Chem Biol.* 1995;2(8):553–61.
173. Grosset JH, Cole ST. Genomics and the chemotherapy of leprosy. *Lepr Rev.* 2001;72(4):429–40.
174. Wheeler PR. Recent Research into the Physiology of *Mycobacterium leprae*. Vol. 31, *Advances in Microbial Physiology.* 1990. 71–124 p.
175. Pessolani MC. Purification, characterization, gene sequence, and significance of a bacterioferritin from *Mycobacterium leprae*. *J Exp Med* [Internet]. 1994 Jul 1;180(1):319–27. Available from: <http://www.jem.org/cgi/doi/10.1084/jem.180.1.319>
176. Thom CS, Dickson CF, Gell DA, Weiss MJ. Hemoglobin Variants: Biochemical Properties and Clinical Correlates. *Cold Spring Harb Perspect Med.* 2013;3(3):1–22.
177. Mendonça VRR, Luz NF, Santos NJG, Borges VM, Gonçalves MS, Andrade BB, et al. Association between the Haptoglobin and Heme Oxygenase 1 Genetic Profiles and Soluble CD163 in Susceptibility to and Severity of Human Malaria. Adams JH, editor. *Infect Immun* [Internet]. 2012 Apr;80(4):1445–54. Available from: <http://iai.asm.org/lookup/doi/10.1128/IAI.05933-11>
178. Schaer CA, Schoedon G, Imhof A, Kurrer MO, Schaer DJ. Constitutive endocytosis of CD163 mediates hemoglobin-heme uptake and determines the noninflammatory and protective transcriptional response of macrophages to hemoglobin. *Circ Res.* 2006;99(9):943–50.
179. Schaer DJ, Schleiffenbaum B, Kurrer M, Imhof A, Bächli E, Fehr J, et al. Soluble hemoglobin-haptoglobin scavenger receptor CD163 as a lineage-specific marker in the reactive hemophagocytic syndrome. *Eur J Haematol.* 2005;74(1):6–10.
180. Moura DF, Mattos KA De, Nery AC, Pinheiro RO, Sarno EN. CD163 favors *Mycobacterium leprae* survival and persistence by promoting anti-inflammatory pathways in lepromatous macrophages. 2012;2925–36.
181. Sulahian TH, Högger P, Wahner AE, Wardwell K, Goulding NJ, Sorg C, et al. Human monocytes express CD163, which is upregulated by IL-10 and identical to p155. *Cytokine.* 2000;12(9):1312–21.
182. de Mattos Barbosa MG, da Silva Prata RB, Andrade PR, Ferreira H, de Andrade Silva BJ, da Paixão

- de Oliveira JA, et al. Indoleamine 2,3-dioxygenase and iron are required for *Mycobacterium leprae* survival. *Microbes Infect.* 2017;19(11):505–14.
183. Finch CA, Huebers H. Perspectives in Iron Metabolism. *N Engl J Med* [Internet]. 1982 Jun 24;306(25):1520–8. Available from: <http://www.nejm.org/doi/abs/10.1056/NEJM198206243062504>
184. Abel L, Demenais F. Detection of major genes for susceptibility to leprosy and its subtypes in a Caribbean island: Desirade island. *Am J Hum Genet.* 1988;42(2):256–66.
185. Feitosa MF, Borecki I, Krieger H, Beiguelman B, Rao DC. The genetic epidemiology of leprosy in a Brazilian population. *Am J Hum Genet.* 1995;56(5):1179–85.
186. Cardoso CC, Pereira AC, Sales-Marques C, Ozório Moraes M. Leprosy susceptibility: genetic variations regulate innate and adaptive immunity, and disease outcome. *Future Microbiol.* 2011;6(5):533–549.
187. Cambri G, Mira MT. Genetic susceptibility to leprosy—from classic immune-related candidate genes to hypothesis-free, whole genome approaches. *Front Immunol.* 2018;9(JUL):1–9.
188. Prevedello FC, Mira MT. Hanseníase: uma doença genética? *An Bras Dermatol.* 2007;82(5):451–9.
189. Mazini PS, Alves HV, Reis PG, Lopes AP, Sell AM, Santos-Rosa M, et al. Gene association with leprosy: A review of published data. *Front Immunol.* 2016;6(JAN):1–17.
190. Krutzik SR, Ochoa MT, Sieling PA, Uematsu S, Ng YW, Legaspi A, et al. Activation and regulation of Toll-like receptors 2 and 1 in human leprosy. *Nat Med.* 2003;9(5):525–32.
191. Mattos KA, Oliveira VGC, D’Avila H, Rodrigues LS, Pinheiro RO, Sarno EN, et al. TLR6-Driven Lipid Droplets in *Mycobacterium leprae*- Infected Schwann Cells: Immunoinflammatory Platforms Associated with Bacterial Persistence . *J Immunol.* 2011;187(5):2548–58.
192. Schuring RP, Hamann L, Faber WR, Pahan D, Richardus JH, Schumann RR, et al. Polymorphism N248S in the Human Toll-Like Receptor 1 Gene Is Related to Leprosy and Leprosy Reactions. *J Infect Dis.* 2009;199(12):1816–9.
193. De Sales Marques C, Brito-De-Souza VN, Guerreiro LTA, 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. *J Infect Dis.* 2013;208(1):120–9.
194. Bochud P, Hawn TR, Siddiqui MR, Saunderson P, Britton S, Abraham I, et al. Toll-Like Receptor 2 (TLR2) Polymorphisms Are Associated with Reversal Reaction in Leprosy. *J Infect Dis* [Internet]. 2008 Jan 15;197(2):253–61. Available from: <https://academic.oup.com/jid/article-lookup/doi/10.1086/524688>
195. 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. *Eur J Clin Microbiol Infect Dis.* 2009;28(9):1055–65.
196. Santana N de L, Rêgo JL, Oliveira JM, de Almeida LF, Braz M, Machado LMM, et al. Polymorphisms in genes TLR1, 2 and 4 are associated with differential cytokine and chemokine serum production in patients with leprosy. *Mem Inst Oswaldo Cruz.* 2017;112(4):260–8.
 197. 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. *Nat Genet.* 2001;27(4):439–41.
 198. Alter A, De Léséleuc L, Van Thuc N, Thai VH, Huong NT, Ba NN, et al. Genetic and functional analysis of common MRC1 exon 7 polymorphisms in leprosy susceptibility. *Hum Genet.* 2010;127(3):337–48.
 199. Fitness J, Floyd S, Warndorff DK, Sichali L, Malema S, Crampin AC, et al. Large-scale candidate gene study of leprosy susceptibility in the Karonga district of northern Malawi. *Am J Trop Med Hyg.* 2004;71(3):341–9.
 200. Grant A V., 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. *Hum Genet.* 2014;133(7):883–93.
 201. Dornelles LN, Pereira-Ferrari L, Messias-Reason I. Mannan-binding lectin plasma levels in leprosy: Deficiency confers protection against the lepromatous but not the tuberculoid forms. *Clin Exp Immunol.* 2006;145(3):463–8.
 202. de Messias-Reason IJ, Boldt ABW, 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. *J Infect Dis.* 2007;196(9):1379–85.
 203. Vasconcelos LRS, Fonseca JPL, Carmo RF do, Mendonça TF de, Pereira VRA, Lucena-Silva N, et al. Mannose-binding lectin serum levels in patients with leprosy are influenced by age and MBL2 genotypes. *Int J Infect Dis [Internet].* 2011 Aug;15(8):e551–7. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S120197121100097X>
 204. Sapkota BR, Macdonald M, Berrington WR, Misch EA, Ranjit C, Siddiqui MR, et al. Association of TNF, MBL, and VDR polymorphisms with leprosy phenotypes. *Hum Immunol.* 2010;71(10):992–8.
 205. Zhang FR, Huang W, Chen SM, Sun LD, Liu HX, Li YP, et al. Genomewide association study of leprosy. *N Engl J Med.* 2009;361:2609–18.
 206. Pan H, Dai Y, Tang S, Wang J. Polymorphisms of NOD2 and the risk of tuberculosis: A validation study in the Chinese population. *Int J Immunogenet.* 2012;39(3):233–40.
 207. Grant A V, 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. *J Infect Dis* [Internet]. 2012;206(11):1763–7. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22984114>
208. Sales-Marques C, Salomão H, Fava VM, Alvarado-Arnez LE, Amaral EP, Cardoso CC, et al. NOD2 and CCDC122-LACC1 genes are associated with leprosy susceptibility in Brazilians. *Hum Genet* [Internet]. 2014;133(12):1525–32. Available from: <http://link.springer.com/10.1007/s00439-014-1502-9>
209. 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. *Am J Hum Genet*. 2018;102(5):794–805.
210. Roy S, Frodsham A, Saha B, Hazra SK, Mascie-Taylor CGN, Hill AVS. Association of Vitamin D Receptor Genotype with Leprosy Type. *J Infect Dis*. 1999;179(1):187–91.
211. de Vries RRP, Mehra NK, Vaidya MC, Gupte MD, Khan PM, Van Rood JJ. HLA-linked Control of Susceptibility to Tuberculoid Leprosy and Association with HLA–DR types. *Tissue Antigens*. 1980;16(4):294–304.
212. Mira MT. Genetic host resistance and susceptibility to leprosy. *Microbes Infect*. 2006;8(4):1124–31.
213. Vanderborcht PR, Pacheco AG, Moraes ME, Antoni G, Romero M, Verville A, et al. HLA-DRB1* 04 and DRB1* 10 are associated with resistance and susceptibility, respectively, in Brazilian and Vietnamese leprosy patients. *Genes Immun*. 2007;8(4):320–4.
214. Mira MT, Alcaïs A, Van Thuc N, Thai VH, Huong NT, Ba NN, et al. Chromosome 6q25 is linked to susceptibility to leprosy in a Vietnamese population. *Nat Genet*. 2003;33(3):412–5.
215. Roy S, McGuire W, Mascie-Taylor CGN, Saha B, Hazra SK, Hill AVS, et al. Tumor Necrosis Factor Promoter Polymorphism and Susceptibility to Lepromatous Leprosy. *J Infect Dis*. 1997;176(2):530–2.
216. Alcaïs A, Alter A, Antoni G, Orlova M, Van Thuc N, Singh M, et al. Stepwise replication identifies a low-producing lymphotoxin- α allele as a major risk factor for early-onset leprosy. *Nat Genet*. 2007;39(4):517–22.
217. Hagge DA, Saunders BM, Ebenezer GJ, Ray NA, Marks VT, Britton WJ, et al. Lymphotoxin-a and TNF have essential but independent roles in the evolution of the granulomatous response in experimental leprosy. *Am J Pathol* [Internet]. 2009;174(4):1379–89. Available from: <http://dx.doi.org/10.2353/ajpath.2009.080550>
218. Sallakci N, Coskun M, Berber Z, Gürkan F, Kocamaz H, Uysal G, et al. Interferon- γ gene+874T-A polymorphism is associated with tuberculosis and gamma interferon response. *Tuberculosis*. 2007;87(3):225–30.

219. 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. *Hum Genet.* 2010;128(5):481–90.
220. Silva GAV, Santos MP, Mota-Passos I, Boechat AL, Malheiro A, Naveca FG, et al. IFN- γ +875 microsatellite polymorphism as a potential protection marker for leprosy patients from Amazonas state, Brazil. *Cytokine* [Internet]. 2012;60(2):493–7. Available from: <http://dx.doi.org/10.1016/j.cyto.2012.04.043>
221. Reynard MP, Turner D, Junqueira-Kipnis AP, De Souza MR, Moreno C, Navarrete C V. Allele frequencies for an interferon- γ microsatellite in a population of Brazilian leprosy patients. *Eur J Immunogenet.* 2003;30(2):149–51.
222. Silva GA V, Naveca FG, Ramasawmy R, Boechat AL. Association between the IFNG +874A/T gene polymorphism and leprosy resistance: A meta-analysis. *Cytokine* [Internet]. 2014;65(2):130–3. Available from: <http://dx.doi.org/10.1016/j.cyto.2013.12.002>
223. Santos AR, Suffys PN, Vanderborght PR, Moraes MO, Vieira LMM, Cabello PH, et al. Role of tumor necrosis factor-alpha and interleukin-10 promoter gene polymorphisms in leprosy. *J Infect Dis* [Internet]. 2002;186(11):1687–91. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/12447749>
224. Franceschi DSA, Mazini PS, Rudnick CCC, Sell AM, Tsuneto LT, Ribas ML, et al. Influence of TNF and IL10 gene polymorphisms in the immunopathogenesis of leprosy in the south of Brazil. *Int J Infect Dis.* 2009;13(4):493–8.
225. Moraes MO, Pacheco a G, Schonkeren JJM, Vanderborght PR, Nery J a C, Santos a R, et al. Interleukin-10 promoter single-nucleotide polymorphisms as markers for disease susceptibility and disease severity in leprosy. *Genes Immun.* 2004;5(7):592–5.
226. Malhotra D, Darvishi K, Sood S, Sharma S, Grover C, Relhan V, et al. IL-10 promoter single nucleotide polymorphisms are significantly associated with resistance to leprosy. *Hum Genet.* 2005;118(2):295–300.
227. Alvarado-Navarro A, Montoya-Buelna M, Muñoz-Valle JF, López-Roa RI, Guillén-Vargas C, Fafutis-Morris M. The 3'UTR 1188 A/C polymorphism in the interleukin-12p40 gene (IL-12B) is associated with lepromatous leprosy in the west of Mexico. *Immunol Lett.* 2008;118(2):148–51.
228. 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. *Am J Hum Genet.* 2012;91(5):935–41.
229. Ali S, Srivastava AK, Chopra R, Aggarwal S, Garg VK, Bhattacharya SN, et al. IL12B SNPs and copy

- number variation in IL23R gene associated with susceptibility to leprosy. *J Med Genet.* 2013;50(1):34–42.
230. Sousa ALM, Fava VM, Sampaio LH, Martelli CMT, Costa MB, Mira MT, et al. Genetic and immunological evidence implicates interleukin 6 as a susceptibility gene for leprosy type 2 reaction. *J Infect Dis.* 2012;205(9):1417–24.
 231. Sales-Marques C, Cardoso CC, Alvarado-Arnez LE, Illaramendi X, Sales AM, Hacker M de A, et al. Genetic polymorphisms of the IL6 and NOD2 genes are risk factors for inflammatory reactions in leprosy. *PLoS Negl Trop Dis.* 2017;11(7):1–16.
 232. Mira MT, Alcaïs A, Van Thuc H, Moraes MO, Di Flumeri C, Thai VI, et al. Susceptibility to leprosy is associated with PARK2 and PACRG. *Nature.* 2004;427(6975):636–40.
 233. 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. *Hum Genet.* 2013;132(1):107–16.
 234. Chopra R, Kalaiarasan P, Ali S, Srivastava AK, Aggarwal S, Garg VK, et al. PARK2 and proinflammatory/anti-inflammatory cytokine gene interactions contribute to the susceptibility to leprosy: A case-control study of North Indian population. *BMJ Open.* 2014;4(2):1–7.
 235. Al-Qahtani AA, Al-Anazi MR, Al-Zoghaibi FA, Abdo AA, Sanai FM, Al-Hamoudi WK, et al. PARK2 polymorphisms predict disease progression in patients infected with hepatitis C virus. *Ann Hepatol.* 2016;15(6):824–33.
 236. Ali S, Vollaard a. M, Widjaja S, Surjadi C, Van De Vosse E, Van Dissel JT. PARK2/PACRG polymorphisms and susceptibility to typhoid and paratyphoid fever. *Clin Exp Immunol.* 2006;144(3):425–31.
 237. Konovalova E V, Lopacheva OM, Grivennikov IA, Lebedeva OS, Dashinimaev EB, Khaspekov LG, et al. Mutations in the Parkinson's Disease-Associated PARK2 Gene Are Accompanied by Imbalance in Programmed Cell Death Systems. *Acta Naturae [Internet].* 2015;7(4):146–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26798503>
 238. 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).
 239. 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. Johnson C, editor. *PLoS Negl Trop Dis [Internet].* 2016 Feb 4;10(2):e0004412. Available from: <https://dx.plos.org/10.1371/journal.pntd.0004412>

240. Fava VM, Xu YZ, Lettre G, Van Thuc N, Orlova M, Thai VH, et al. Pleiotropic effects for Parkin and LRRK2 in leprosy type-1 reactions and Parkinson's disease. *Proc Natl Acad Sci*. 2019;116(31):15616–24.
241. Moreno-Mayar JV, Potter BA, Vinner L, Steinrücken M, Rasmussen S, Terhorst J, et al. Terminal Pleistocene Alaskan genome reveals first founding population of Native Americans. *Nature*. 2018;553(7687):203–7.
242. Silva WA, Bonatto SL, Holanda AJ, Ribeiro-dos-Santos AK, Paixão BM, Goldman GH, et al. Mitochondrial genome diversity of native americans supports a single early entry of founder populations into America. *Am J Hum Genet*. 2002;71(1):187–92.
243. Amorim CEG, Nunes K, Meyer D, Comas D, Bortolini MC, Salzano FM, et al. Genetic signature of natural selection in first Americans. *Proc Natl Acad Sci U S A*. 2017;114(9):2195–9.
244. Gunz P, Bookstein FL, Mitteroecker P, Stadlmayr A, Seidler H, Weber GW. Early modern human diversity suggests subdivided population structure and a complex out-of-Africa scenario. *Proc Natl Acad Sci* [Internet]. 2009 Apr 14;106(15):6094–8. Available from: <http://www.pnas.org/cgi/doi/10.1073/pnas.0808160106>
245. Pena SDJ, Bastos-Rodrigues L, Pimenta JR, Bydlowski SP. DNA tests probe the genomic ancestry of Brazilians. *Brazilian J Med Biol Res*. 2009;42(10):870–6.
246. Ribeiro D. O povo brasileiro. Companhia das Letras. 1995. 1–447 p.
247. Curtin PD. The Atlantic slave trade : a census. Madison: University of Wisconsin Press; 1969.
248. Callerigari-Jacques S, Salzano F. Brazilian Indian/non-Indian interactions and their effects. 51st ed. *Ciência Cultural*; 1999. 166-174. p.
249. Alves-Silva J, da Silva Santos M, Guimarães PEM, Ferreira ACS, Bandelt H-J, Pena SDJ, et al. The Ancestry of Brazilian mtDNA Lineages. *Am J Hum Genet* [Internet]. 2000 Aug;67(2):444–61. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0002929707626548>
250. Abe-Sandes K, Silva WA, Zago MA. Heterogeneity of the Y chromosome in Afro-Brazilian populations. *Hum Biol* [Internet]. 2004 Feb;76(1):77–86. Available from: <http://www.jstor.org/stable/41466214>
251. IBGE. Brasil: 500 anos de povoamento. Rio de Janeiro; 2000. 1–237 p.
252. Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. *Am J Hum Genet* [Internet]. 2012;90(1):7–24. Available from: <http://dx.doi.org/10.1016/j.ajhg.2011.11.029>
253. Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet*. 2005;6(2):95–108.
254. Lewis CM, Knight J. Introduction to genetic association studies. Cold Spring Harb Protoc.

- 2012;7(3):297–306.
255. Golding J, Northstone K, Miller LL, Smith GD, Pembrey M. Differences between blood donors and a population sample: Implications for case-control studies. *Int J Epidemiol.* 2013;42(4):1145–56.
 256. Clayton DG, Walker NM, Smyth DJ, Pask R, Cooper JD, Maier LM, et al. Population structure, differential bias and genomic control in a large-scale, case-control association study. *Nat Genet.* 2005;37(11):1243–6.
 257. Marchini J, Cardon LR, Phillips MS, Donnelly P. The effects of human population structure on large genetic association studies. *Nat Genet.* 2004;36(5):512–7.
 258. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics.* 2000;155(2):945–59.
 259. Pritchard JK, Donnelly P. Case-control studies of association in structured or admixed populations. *Theor Popul Biol.* 2001;60(3):227–37.
 260. Schaid DJ. Transmission disequilibrium, family controls, and great expectations. *Am J Hum Genet.* 1998;63(4):935–41.
 261. Spielman RS, Ewens WJ. The TDT and other family-based tests for linkage disequilibrium and association. *Am J Hum Genet.* 1996;59(5):983–9.
 262. Spielman RS, McGinnis RE, Ewens WJ. Transmission test for linkage disequilibrium: The insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet.* 1993;52(3):506–16.
 263. Pereira F dos SCF, Guimarães RM, Lucidi AR, Brum DG, Paiva CLA, Alvarenga RMP. A systematic literature review on the European, African and Amerindian genetic ancestry components on Brazilian health outcomes. *Sci Rep.* 2019;9(1):1–11.
 264. Parra FC, Amado RC, Lambertucci JR, Rocha J, Antunes CM, Pena SDJ. Color and genomic ancestry in Brazilians. *Proc Natl Acad Sci U S A.* 2003;100(1):177–82.
 265. Halder I, Shriver M, Thomas M, Fernandez JR, Frudakis T. A panel of ancestry informative markers for estimating individual biogeographical ancestry and admixture from four continents: Utility and applications. *Hum Mutat.* 2008;29(5):648–58.
 266. Nassir R, Kosoy R, Tian C, White PA, Butler LM, Silva G, et al. An ancestry informative marker set for determining continental origin: Validation and extension using human genome diversity panels. *BMC Genet.* 2009;10:1–13.
 267. Pereira R, Phillips C, Pinto N, Santos C, dos Santos SEB, Amorim A, et al. Straightforward inference of ancestry and admixture proportions through ancestry-informative insertion deletion multiplexing. *PLoS One.* 2012;7(1).
 268. Tian C, Kosoy R, Nassir R, Lee A, Villoslada P, Klareskog L, et al. European population genetic

- substructure: Further definition of ancestry informative markers for distinguishing among diverse European ethnic groups. *Mol Med*. 2009;15(11–12):371–83.
269. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006;38(8):904–9.
270. Knowler WC, Williams RC, Pettitt DJ, Steinberg AG. Gm(3;5,13,14) and type 2 diabetes mellitus: An association in American Indians with genetic admixture. *Am J Hum Genet*. 1988;43(4):520–6.
271. Thomas DC, Witte JS. Point: Population stratification: A problem for case-control studies of candidate-gene associations? *Cancer Epidemiol Biomarkers Prev*. 2002;11(6):505–12.
272. IBGE. Características étnico-raciais da população brasileira. Vol. 39, Produção da Pecuária Municipal. Rio de Janeiro; 2013. 1–208 p.
273. Lins TC, Vieira RG, Abreu BS, Grattapaglia D, Pereira RW. Genetic composition of Brazilian population samples based on a set of twenty eight ancestry informative SNPs. *Am J Hum Biol*. 2010;22(2):187–92.
274. Lins TC, Vieira RG, Abreu BS, Gentil P, Moreno-Lima R, Oliveira RJ, et al. Genetic heterogeneity of self-reported ancestry groups in an admixed Brazilian population. *J Epidemiol*. 2011;21(4):240–5.
275. Fernanda Lima-Costa M, Rodrigues LC, Barreto ML, Gouveia M, Horta BL, Mambrini J, et al. Genomic ancestry and ethnorracial self-classification based on 5,871 community-dwelling Brazilians (The Epigen Initiative). *Sci Rep*. 2015;5.
276. Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, et al. Genetic Structure of Human Populations. 2002;298(December).
277. Sedgwick P. Analysing case-control studies: Adjusting for confounding. *BMJ*. 2013;346(7890):4–5.
278. Camargo RM de, Silva WL da, Medeiros P, Belone A de FF, Latini ACP. Polymorphisms in the TGFB1 and IL2RA genes are associated with clinical forms of leprosy in Brazilian population. *Mem Inst Oswaldo Cruz*. 2018;113(12):e180274.
279. Saloum de Neves Manta F, Pereira R, Vianna R, Rodolfo Beuttenmüller de Araújo A, Leite Góes Gitaí D, Aparecida da Silva D, et al. Revisiting the Genetic Ancestry of Brazilians Using Autosomal AIM-Indels. *PLoS One*. 2013;8(9):1–11.
280. Manta FSN, Pereira R, Caiafa A, Silva DA, Gusmão L, Carvalho EF. Analysis of genetic ancestry in the admixed Brazilian population from Rio de Janeiro using 46 autosomal ancestry-informative indel markers. *Ann Hum Biol*. 2013;40(1):94–8.
281. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*. 2009;19(9):1655–64.

282. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81(3):559–75.
283. Wang K, Li M, Hakonarson H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 2010;38(16):1–7.
284. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet.* 2006;2(12):2074–93.
285. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: Analysis and visualization of LD and haplotype maps. *Bioinformatics.* 2005;21(2):263–5.
286. Delaneau O, Coulonges C, Zagury J. Shape-IT : new rapid and accurate algorithm for haplotype inference. 2008;14:1–14.
287. Gautier M, Vitalis R. rehh : an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. *Bioinformatics.* 2012;28(8):1176–7.
288. Medeiros RCA. Modulação metabólica induzida pelo *Mycobacterium leprae* e implicações para a suscetibilidade à hanseníase. Instituto Oswaldo Cruz; 2018.
289. Ramakers C, Ruijter JM, Lekanne Deprez RH, Moorman AFM. Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neurosci Lett.* 2003;339(1):62–6.
290. Ruijter JM, Ramakers C, Hoogaars WMH, Karlen Y, Bakker O, van den hoff MJB, et al. Amplification efficiency: Linking baseline and bias in the analysis of quantitative PCR data. *Nucleic Acids Res.* 2009;37(6).
291. Peirson SN. Experimental validation of novel and conventional approaches to quantitative real-time PCR data analysis. *Nucleic Acids Res.* 2003;31(14):73e – 73.
292. Rafaela Mota da Silva. Utilização de marcadores informativos de ancestralidade em indivíduos casos e controles da hanseníase. Centro Universitário IBMR; 2015.
293. Lima-Bezerra OC. Análise de polimorfismos no gene PKLR e associação com a hanseníase. 2015.
294. Tishkoff SA, Varkonyi R, Cahinhinan N, Abbes S, Argyropoulos G, Destro-Bisol G, et al. Haplotype diversity and linkage disequilibrium at human G6PD: Recent origin of alleles that confer malarial resistance. *Science (80-).* 2001;293(5529):455–62.
295. Howes RE, Battle KE, Satyagraha AW, Baird JK, Hay SI. G6PD Deficiency. Global Distribution, Genetic Variants and Primaquine Therapy [Internet]. Vol. 81, *Advances in Parasitology*. Elsevier; 2013. 133–201 p. Available from: <http://dx.doi.org/10.1016/B978-0-12-407826-0.00004-7>
296. Magalhães Da Silva T, Sandhya Rani MR, De Oliveira Costa GN, Figueiredo MA, Melo PS, Nascimento JF, et al. The correlation between ancestry and color in two cities of Northeast Brazil with

- contrasting ethnic compositions. *Eur J Hum Genet*. 2015;23(7):984–9.
297. Pena SDJ, di Pietro G, Fuchshuber-Moraes M, Genro JP, Hutz MH, Kehdy F de SG, et al. The genomic ancestry of individuals from different geographical regions of Brazil is more uniform than expected. *PLoS One*. 2011;6(2).
298. Mariuzzo P. Atlas do comércio transatlântico de escravos. *Cienc Cult* [Internet]. 2011 Jan;63(1):59–61. Available from: http://cienciaecultura.bvs.br/scielo.php?script=sci_arttext&pid=S0009-67252011000100021&lng=pt&tlng=pt
299. Souza AM de, Resende SS, Sousa TN de, Brito CFA de. A systematic scoping review of the genetic ancestry of the Brazilian population. *Genet Mol Biol* [Internet]. 2019 Sep;42(3):495–508. Available from: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1415-47572019000400495&tlng=en
300. Sirugo G, Williams SM, Tishkoff SA. The Missing Diversity in Human Genetic Studies. *Cell* [Internet]. 2019 Mar;177(1):26–31. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0092867419302314>
301. Lin PI, Vance JM, Pericak-Vance MA, Martin ER. No gene is an island: The flip-flop phenomenon. *Am J Hum Genet*. 2007;80(3):531–8.
302. Casanova JL, Abel L. Human genetics of infectious diseases: A unified theory. *EMBO J*. 2007;26(4):915–22.
303. Gibson G. Rare and common variants: Twenty arguments. *Nat Rev Genet*. 2012;13(2):135–45.
304. Pritchard JK. The allelic architecture of human disease genes: common disease-common variant... or not? *Hum Mol Genet*. 2002;11(20):2417–23.
305. Alcaïs A, Abel L, Casanova J-L. Human genetics of infectious diseases: between proof of principle and paradigm. *J Clin Invest* [Internet]. 2013;5(9):21–31. Available from: <http://www.dovepress.com/getfile.php?fileID=16748%5Chttp://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed11&NEWS=N&AN=2013449942>
306. Cook JD, Baynes RD, Skikne BS. Iron deficiency and the measurement of iron status. *Nutr Res Rev*. 1992;5(1):198–202.
307. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A*. 2009;106(23):9362–7.
308. Borel C, Deutsch S, Letourneau A, Migliavacca E, Montgomery SB, Dimas AS, et al. Identification of cis- and trans-regulatory variation modulating microRNA expression levels in human fibroblasts. *Genome Res*. 2011;21(1):68–73.

309. Rajith B, Doss CGP. Path to facilitate the prediction of functional amino acid substitutions in red blood cell disorders - A computational approach. *PLoS One*. 2011;6(9).
310. Bagla S, Bhambhani K, Gadgeel M, Buck S, Jin J-P, Ravindranath Y. Compound heterozygosity in PKLR gene for a previously unrecognized intronic polymorphism and a rare missense mutation as a novel cause of severe pyruvate kinase deficiency. *Haematologica* [Internet]. 2019 Sep;104(9):e428–31. Available from: <http://www.haematologica.org/lookup/doi/10.3324/haematol.2018.214692>
311. Tishkoff SA, Reed FA, Ranciaro A, Voight BF, Babbitt CC, Silverman JS, et al. Convergent adaptation of human lactase persistence in Africa and Europe. *Nat Genet* [Internet]. 2007 Jan 10;39(1):31–40. Available from: <http://www.nature.com/articles/ng1946>
312. Marrow B, Secreted S, Protect C. The Genotype-Tissue Expression (GTEx) project. *Nat Genet*. 2014;71(11):3831–40.
313. Verrelli BC, McDonald JH, Argyropoulos G, Destro-Bisol G, Froment A, Drousiotou A, et al. Evidence for balancing selection from nucleotide sequence analyses of human G6PD. *Am J Hum Genet*. 2002;71(5):1112–28.
314. Saunders MA, Slatkin M, Garner C, Hammer MF, Nachman MW. The extent of linkage disequilibrium caused by selection on G6PD in humans. *Genetics*. 2005;171(3):1219–29.
315. Clarke GM, Rockett K, Kivinen K, Hubbart C, Jeffreys AE, Rowlands K, et al. Characterisation of the opposing effects of G6PD deficiency on cerebral malaria and severe malarial anaemia. *Elife*. 2017;6:1–30.
316. Medeiros RCA, Girardi K do C de V, Cardoso FKL, Mietto B de S, Pinto TG de T, Gomez LS, et al. Subversion of Schwann Cell Glucose Metabolism by *Mycobacterium leprae*. *J Biol Chem*. 2016 Oct;291(41):21375–87.
317. M D Cappellini GF. Glucose-6-phosphate dehydrogenase deficiency. *Lancet*. 2008;371(9606):64–74.
318. Guindo A, Fairhurst RM, Doumbo OK, Wellems TE, Diallo DA. X-linked G6PD deficiency protects hemizygous males but not heterozygous females against severe malaria. *PLoS Med*. 2007;4(3):516–22.
319. Mattos KA, Lara FA, Oliveira VGC, Rodrigues LS, D’Avila H, Melo RCN, et al. Modulation of lipid droplets by *Mycobacterium leprae* in Schwann cells: A putative mechanism for host lipid acquisition and bacterial survival in phagosomes. *Cell Microbiol*. 2011;13(2):259–73.
320. Mattos KA, Oliveira VCG, Berrêdo-Pinho M, Amaral JJ, Antunes LCM, Melo RCN, et al. *Mycobacterium leprae* intracellular survival relies on cholesterol accumulation in infected macrophages: A potential target for new drugs for leprosy treatment. *Cell Microbiol*.

2014;16(6):797–815.

321. Teles RM, Moraes MO, Geraldo NT, Salles AM, Sarno EN, Sampaio EP. Differential TNF α mRNA regulation detected in the epidermis of leprosy patients. *Arch Dermatol Res.* 2002;294(8):355–62.
322. Dupré-Crochet S, Erard M, Nüße O. ROS production in phagocytes: why, when, and where? *J Leukoc Biol.* 2013;94(4):657–70.
323. Weston A, Stern RJ, Lee RE, Nassau PM, Monsey D, Martin SL, et al. Biosynthetic origin of mycobacterial cell wall galactofuranosyl residues. *Tuber Lung Dis.* 1998;78(2):123–31.
324. Babalola MO, Imaga NA, Samuel TA, Diriwari IP, Kolade O, Ezeamalu I, et al. Genetic Polymorphisms of Glucose-6-Phosphate Dehydrogenase in Lagos, Nigeria. *Hemoglobin.* 2018 Jan;42(1):47–50.
325. Cao Y, Wang X, Cao Z, Cheng X. Association of vitamin D receptor gene taqi polymorphisms with tuberculosis susceptibility: A meta-analysis. *Int J Clin Exp Med.* 2015;8(6):10187–203.
326. Ke Z, Yuan L, Ma J, Zhang X, Guo Y, Xiong H. IL-10 polymorphisms and tuberculosis susceptibility: An updated meta-analysis. *Yonsei Med J.* 2015;56(5):1274–87.

ANEXO I

1 – Lista dos pares de *primers* referente aos 46-AIMs-Indels e as características desse painel de marcadores

Tabela 1.1 – Sequência dos primers utilizados no multiplex dos 46-AIMs-Indels.

| Número MID | Número rs | Primer marcado com a fluorescência | Primer não marcado |
|------------|------------|------------------------------------|----------------------------|
| MID-1470 | rs2307666 | 6FAM-GAGTCTGACCCTTCATAAGC | gCCATGGTGATATTACGTCCC |
| MID-777 | rs1610863 | 6FAM-TGGAAGACACGTCCCTAAGAG | gTATTCCTCCAGGCTCTTTGC |
| MID-196 | rs16635 | 6FAM-CCAAGTTCTAGCCATATGGA | gTTTCTTACTATCTTCTCTGACCAT |
| MID-881 | rs1610965 | 6FAM-TTGGCTCCCTATGATAATCC | gTTTCTTGTGTTCCCAAAGTTCTCC |
| MID-3122 | rs35451359 | 6FAM-TCACAAGTCCGGAATACCAG | gAGTTATGGGATGGGAAGGAG |
| MID-548 | rs140837 | 6FAM-AGTCAGGACTGAAGAAACCC | gTTTCTCAGTAAAACAAGAGCCCGTG |
| MID-659 | rs1160893 | 6FAM-CACTGCATCAGACTGACTTC | gTTTCTTGGCTGCTTTGCTTTGAATT |
| MID-2011 | rs2308203 | 6FAM-TGAGAACTAGGAGCTCTGG | gTTTCTTAAAGAGCCACTGACAT |
| MID-2929 | rs33974167 | 6FAM-TGTGATGTGGATAGGCAAGG | gAGGCTCCATTTGTTAAGAGG |
| MID-593 | rs1160852 | 6FAM-TGCTCACTTTAGTGGAGACC | gTTTCTGCTTTAGGTCCCTTCTG |
| MID-798 | rs1610884 | 6FAM-ACGACAGTGTTCACAAAGAG | gCTGTTGTCTGACCTGTGAAG |
| MID-1193 | rs2067280 | 6FAM-GCTGGGTAGTTTTTCTCTCC | gTTCCACCATCTACCTTCTATG |
| MID-1871 | rs2308067 | 6FAM-TTGTAGTCAGAGAGTGTGCC | gAGCCTTTTCCCTAACGTCAC |
| MID-17 | rs4183 | 6FAM-AGAACTGCAACCTCCAAG | GATCCAGACACTGAAGATG |
| MID-2538 | rs3054057 | 6FAM-CTCGCAAAGTAGGCAAGTTC | gACACCAACAATCTTGGCACC |
| MID-1644 | rs2307840 | 6FAM-ACACCACTGAAGATCTGACC | GGTCTAAAGTCAGTGCACAG |
| MID-3854 | rs60612424 | VIC-TCACCTTATTAGGGTTGC | GCCAGGGATTTAGTGTAGAG |
| MID-2275 | rs3033053 | VIC-CTACCTGACTACCACCTATG | gACCCAGCCTATCTGACTTTG |
| MID-94 | rs16384 | VIC-TGGTGGCTCATGCACTTTTG | gTTTACAGGGTCTCGCTATGATGC |
| MID-3072 | rs34611875 | VIC-AGCTTTTTCCGGCAACTCTC | gTTTGGATGTGTCTGAGCTCAAC |
| MID-772 | rs1610859 | VIC-GTCTCRTTTTCTGCACTAG | gTTTCTTATCCTTCTGCTCACTCTAC |
| MID-2313 | rs3045215 | VIC-GCACACATGCAGAAATGCAG | GTTGTAACATCTGTGAGGTC |
| MID-397 | rs25621 | VIC-TGGGCTTCTTCTGGGAAAAC | gCCACATTAGGCGTTTTGTCT |
| MID-1636 | rs2307832 | VIC-TTAGGAAGAGGTGCTATGGG | gCCTCCTTTGAAGACACACAG |
| MID-51 | rs16343 | VIC-AAGATTGGAGGGAAAAGTGC | gCGTCTCCACCTTCTTTTTTC |
| MID-2431 | rs3031979 | VIC-AGGAGGAGCTGATAGACTTC | GCAGTGTGCAACTGATACG |
| MID-2264 | rs34122827 | VIC-CTTTGGCTATCCTGTCTCAC | GTAGGAGACCACTCACATTC |
| MID-2256 | rs133052 | NED-ATCGAACCTTCTAAGGAC | GCAAGAAAAGGAATCCAGGC |
| MID-128 | rs6490 | NED-ATCAGGAGACAATCCAGCAG | gTCCAGCCATTAGACAAAGG |
| MID-15 | rs4181 | NED-GGGTATTTGCCTCATCTCC | gTTTCTAGGTATTCTCTGTTCCAC |
| MID-2241 | rs3030826 | NED-ACATACACGTGGAAGACTGC | gTTACTGTGACTGATCCAATAG |
| MID-419 | rs140708 | NED-CAGGAAAGTATGGCCATTC | GTCCATGTTTTCTTTGAGCATC |
| MID-943 | rs1611026 | NED-TCTTCTACCCCTGTTAGTG | GACAAGATCACTAGCTTGAC |
| MID-159 | rs16438 | NED-ACCAGAGCACTACAGCCTTT | gCAAGGYAGTAAACAATGAGGG |
| MID-2005 | rs2308161 | NED-TGTAGCGGCAATATAGGCAG | GAAAGTTGTGGCTTAACTGG |
| MID-250 | rs16687 | NED-ATGGAGCAGTAAAGCAGCAC | GTCACCTTTGGTTTTGCAGG |
| MID-1802 | rs2307998 | NED-ACGGTCACTTTGTAGCTCC | gCCAGTTGAGAATCACTGCAC |
| MID-1607 | rs2307803 | NED-TGTTGCAGAAGAACTCAACC | GATAAGCACCTAACTCCAG |
| MID-1734 | rs2307930 | PET-TTCGTGTTCTCACTGTCC | GTGCATCCCATACAACCTGAC |
| MID-406 | rs25630 | PET-TGGCTGCTGTAGATTGTAGG | gACAAATGGACAACGGCCAAG |

| | | | |
|-----------------|------------|--------------------------|--------------------------|
| MID-1386 | rs2307582 | PET-AGAGGATCATGGAGACCAAC | gtttATGTTCCAAGTCAGCAGCAC |
| MID-1726 | rs2307922 | PET-GGTCCAAATGCACCACAATC | GCTCTGCTATTTTGGTTTGC |
| MID-3626 | rs11267926 | PET-TGTTGGTTCTCTCCTTTTCC | GGTGACCCCTTCTTTATCTC |
| MID-360 | rs25584 | PET-AGATCAACTGCCAATCTGGG | gCTCAAGTGACCAACCCACCT |
| MID-1603 | rs2307799 | PET-TTACAATTTCAAGCCTCCGC | GGAGCTGTTAGTCTGAGTAG |
| MID-2719 | rs34541393 | PET-GTCAGGAGTCTAGAACTTC | GGGTGATGAAATGTTCCGAA |

Um dos pares de *primers* de cada AIM foi marcado com os fluoróforos 6FAM, VIC, NED e PET, diferenciados didaticamente por diferentes cores.

Tabela 1.2 – Características dos 46 marcadores informativos de ancestralidade. Frequências alélicas, valores de δ e F_{ST} para os 46 AIM-INDELS nas populações Africanas (AFR), Europeias (EUR) e Nativo-Americanas (NAT) do painel de diversidade HGDP-CEPH.

| MID | Frequência do alelo 1 | | | Frequência do alelo 2 | | | Diferencial de frequência alélica (δ) | | | F_{ST} | | |
|-----------|-----------------------|------|------|-----------------------|------|------|--|---------|---------|----------|---------|---------|
| | AFR | EUR | NAM | AFR | EUR | NAM | AFR/EUR | AFR/NAM | EUR/NAM | AFR/EUR | AFR/NAM | EUR/NAM |
| MID-1470 | 0.16 | 0.56 | 0.03 | 0.84 | 0.44 | 0.97 | 0.40 | 0.13 | 0.53 | 0.28 | 0.08 | 0.43 |
| MID-777 | 0.39 | 0.29 | 0.84 | 0.61 | 0.71 | 0.16 | 0.10 | 0.45 | 0.56 | 0.02 | 0.34 | 0.45 |
| MID-196 | 0.54 | 0.47 | 0.06 | 0.46 | 0.53 | 0.94 | 0.07 | 0.48 | 0.41 | 0.00 | 0.39 | 0.30 |
| MID-881 | 0.32 | 0.88 | 0.99 | 0.68 | 0.12 | 0.01 | 0.56 | 0.67 | 0.11 | 0.50 | 0.62 | 0.07 |
| MID-3122 | 0.55 | 0.99 | 1.00 | 0.45 | 0.01 | 0.00 | 0.44 | 0.45 | 0.01 | 0.48 | 0.39 | 0.00 |
| MID-548 | 0.34 | 0.20 | 0.17 | 0.66 | 0.80 | 0.83 | 0.14 | 0.17 | 0.02 | 0.05 | 0.06 | 0.00 |
| MID-659 | 0.28 | 0.10 | 0.26 | 0.72 | 0.90 | 0.74 | 0.17 | 0.02 | 0.15 | 0.09 | -0.01 | 0.08 |
| MID-2011 | 0.16 | 0.79 | 0.91 | 0.84 | 0.21 | 0.09 | 0.63 | 0.74 | 0.12 | 0.56 | 0.70 | 0.04 |
| MID-2929 | 0.79 | 0.66 | 0.98 | 0.21 | 0.35 | 0.02 | 0.14 | 0.19 | 0.32 | 0.04 | 0.13 | 0.23 |
| MID-593 | 0.70 | 0.01 | 0.01 | 0.31 | 0.99 | 0.99 | 0.69 | 0.69 | 0.00 | 0.73 | 0.64 | -0.01 |
| MID-798 | 0.30 | 0.64 | 0.12 | 0.71 | 0.36 | 0.88 | 0.35 | 0.18 | 0.53 | 0.21 | 0.08 | 0.41 |
| MID-1193 | 0.18 | 0.10 | 0.14 | 0.82 | 0.90 | 0.86 | 0.08 | 0.04 | 0.04 | 0.02 | 0.00 | 0.00 |
| MID-1871 | 0.08 | 0.33 | 0.29 | 0.92 | 0.67 | 0.71 | 0.25 | 0.21 | 0.04 | 0.15 | 0.14 | 0.00 |
| MID-17 | 0.80 | 0.34 | 0.90 | 0.21 | 0.66 | 0.10 | 0.46 | 0.10 | 0.56 | 0.34 | 0.03 | 0.46 |
| MID-2538 | 0.00 | 0.49 | 0.06 | 1.00 | 0.51 | 0.94 | 0.49 | 0.06 | 0.43 | 0.45 | 0.08 | 0.32 |
| MID-1644 | 0.23 | 0.95 | 0.20 | 0.77 | 0.05 | 0.81 | 0.72 | 0.03 | 0.76 | 0.73 | 0.00 | 0.79 |
| MID-3854 | 0.75 | 0.01 | 0.01 | 0.25 | 0.99 | 0.99 | 0.74 | 0.74 | 0.01 | 0.77 | 0.69 | 0.00 |
| MID-2275 | 0.49 | 0.10 | 0.52 | 0.51 | 0.91 | 0.48 | 0.39 | 0.04 | 0.43 | 0.33 | -0.01 | 0.41 |
| MID-94 | 0.07 | 0.16 | 0.65 | 0.93 | 0.85 | 0.35 | 0.08 | 0.58 | 0.49 | 0.03 | 0.56 | 0.43 |
| MID-3072 | 0.10 | 0.94 | 0.98 | 0.90 | 0.06 | 0.02 | 0.84 | 0.88 | 0.04 | 0.83 | 0.86 | 0.01 |
| MID-772 | 0.94 | 0.97 | 0.78 | 0.06 | 0.03 | 0.22 | 0.03 | 0.16 | 0.19 | 0.00 | 0.11 | 0.19 |
| MID-2313 | 0.11 | 0.25 | 0.73 | 0.89 | 0.75 | 0.27 | 0.14 | 0.62 | 0.48 | 0.06 | 0.59 | 0.38 |
| MID-397 | 0.77 | 0.77 | 0.34 | 0.23 | 0.23 | 0.66 | 0.01 | 0.43 | 0.42 | 0.00 | 0.32 | 0.31 |
| MID-1636 | 0.15 | 0.79 | 0.99 | 0.85 | 0.22 | 0.01 | 0.64 | 0.85 | 0.21 | 0.57 | 0.81 | 0.15 |
| MID-51 | 0.06 | 0.64 | 0.87 | 0.94 | 0.36 | 0.13 | 0.59 | 0.81 | 0.23 | 0.52 | 0.81 | 0.11 |
| MID-2431 | 0.05 | 0.11 | 0.23 | 0.95 | 0.89 | 0.77 | 0.06 | 0.18 | 0.12 | 0.02 | 0.14 | 0.05 |
| MID-2264* | 0.16 | 0.47 | 0.13 | 0.84 | 0.45 | 0.87 | 0.30 | 0.03 | 0.33 | 0.21 | 0.00 | 0.23 |
| MID-2256 | 0.00 | 0.22 | 0.84 | 1.00 | 0.78 | 0.16 | 0.22 | 0.84 | 0.61 | 0.19 | 0.87 | 0.54 |
| MID-128 | 0.02 | 0.48 | 0.05 | 0.98 | 0.52 | 0.95 | 0.46 | 0.02 | 0.43 | 0.39 | 0.00 | 0.33 |
| MID-15 | 0.20 | 0.45 | 0.84 | 0.81 | 0.55 | 0.16 | 0.26 | 0.64 | 0.38 | 0.13 | 0.58 | 0.25 |
| MID-2241 | 0.81 | 0.29 | 0.07 | 0.19 | 0.71 | 0.93 | 0.52 | 0.74 | 0.22 | 0.42 | 0.70 | 0.13 |
| MID-419 | 0.92 | 0.81 | 0.56 | 0.08 | 0.19 | 0.44 | 0.11 | 0.36 | 0.25 | 0.04 | 0.31 | 0.14 |
| MID-943 | 0.18 | 0.79 | 0.40 | 0.82 | 0.21 | 0.60 | 0.62 | 0.22 | 0.39 | 0.54 | 0.11 | 0.29 |

| | | | | | | | | | | | | |
|-----------------|------|------|------|------|------|------|------|------|------|------|-------|------|
| MID-159* | 0.75 | 0.57 | 0.73 | 0.25 | 0.43 | 0.27 | 0.18 | 0.02 | 0.17 | 0.07 | -0.01 | 0.05 |
| MID-2005 | 0.03 | 0.67 | 0.33 | 0.97 | 0.33 | 0.68 | 0.64 | 0.30 | 0.35 | 0.59 | 0.30 | 0.21 |
| MID-250 | 0.82 | 0.72 | 0.33 | 0.18 | 0.29 | 0.67 | 0.11 | 0.50 | 0.39 | 0.03 | 0.41 | 0.26 |
| MID-1802 | 0.54 | 0.00 | 0.00 | 0.46 | 1.00 | 1.00 | 0.54 | 0.54 | 0.00 | 0.59 | 0.49 | 0.00 |
| MID-1607 | 0.42 | 0.16 | 0.56 | 0.58 | 0.84 | 0.45 | 0.26 | 0.13 | 0.39 | 0.16 | 0.03 | 0.31 |
| MID-1734 | 0.85 | 0.84 | 0.38 | 0.15 | 0.17 | 0.63 | 0.02 | 0.48 | 0.46 | 0.00 | 0.40 | 0.39 |
| MID-406 | 0.05 | 0.82 | 0.70 | 0.95 | 0.18 | 0.30 | 0.78 | 0.66 | 0.12 | 0.74 | 0.67 | 0.04 |
| MID-1386 | 0.14 | 0.22 | 0.77 | 0.86 | 0.78 | 0.23 | 0.08 | 0.63 | 0.55 | 0.02 | 0.58 | 0.46 |
| MID-1726 | 0.18 | 0.70 | 0.76 | 0.82 | 0.30 | 0.24 | 0.52 | 0.58 | 0.06 | 0.43 | 0.51 | 0.00 |
| MID-3626 | 0.12 | 0.70 | 0.04 | 0.88 | 0.30 | 0.96 | 0.58 | 0.09 | 0.66 | 0.50 | 0.04 | 0.58 |
| MID-360* | 0.59 | 0.86 | 0.52 | 0.32 | 0.14 | 0.48 | 0.27 | 0.07 | 0.34 | 0.14 | 0.03 | 0.27 |
| MID-1603 | 0.08 | 0.39 | 0.82 | 0.92 | 0.61 | 0.18 | 0.31 | 0.74 | 0.43 | 0.21 | 0.73 | 0.30 |
| MID-2719 | 0.45 | 0.35 | 0.47 | 0.55 | 0.66 | 0.53 | 0.10 | 0.02 | 0.12 | 0.02 | -0.01 | 0.03 |

* Para estes marcadores foi observado a presença de um alelo 3 com frequência = 1- (frequência do alelo 1 + frequência do alelo2)

ANEXO II

1 – Resultados complementares

1.1 – Resultado dos testes de seleção (F_{ST} , iHS e $xpEHH$) das variantes do gene *PKLR*

Tabela 1.5 – Valores de F_{ST} , iHS e $xpEHH$ dos SNPs da região do gene *PKLR*.

| Gene | Localização | SNP | Posição (Chr1) | F_{ST} EUR x AFR | iHS AFR | iHS EUR | $xpEHH$ AFRxEUR |
|--------------------|----------------------|-------------------|------------------|--------------------|---------------|---------------|-----------------|
| <i>HCN3</i> | Intron | rs183293873 | 155250521 | - | - | - | -2.1227 |
| <i>HCN3</i> | Intron | rs12044063 | 155250691 | - | - | - | -2.1581 |
| <i>HCN3</i> | Intron | rs7367998 | 155251250 | - | - | - | -2.1062 |
| <i>HCN3</i> | Intron | rs560609863 | 155252157 | - | - | - | -2.5306 |
| <i>HCN3</i> | Intron | rs7520184 | 155253583 | 0.069 | - | - | -2.0537 |
| <i>HCN3</i> | Intron | rs11264352 | 155255837 | 0.083 | - | - | -2.0909 |
| <i>HCN3</i> | Intron | rs573791612 | 155255843 | - | - | - | -2.08524 |
| <i>HCN3</i> | Intron | rs541165488 | 155255844 | - | - | - | -2.0781 |
| <i>HCN3</i> | Intron | rs559902587 | 155255852 | 0.061 | - | - | -2.2337 |
| <i>HCN3</i> | Intron | rs11264353 | 155255906 | 0.080 | - | - | -2.2232 |
| <i>HCN3</i> | Intron | rs552421834 | 155255907 | - | - | - | -2.1670 |
| <i>HCN3</i> | Intron | rs564241242 | 155255929 | - | - | - | -2.1715 |
| <i>HCN3</i> | Intron | rs531296924 | 155255944 | - | - | - | -2.1711 |
| <i>HCN3</i> | Intron | rs549806272 | 155255968 | - | - | - | -2.1738 |
| <i>HCN3</i> | Intron | rs144733142 | 155256019 | - | - | - | -2.1237 |
| <i>HCN3</i> | Intron | rs535535805 | 155256050 | - | - | - | -2.1767 |
| <i>HCN3</i> | Intron | rs547542932 | 155256067 | - | - | - | -2.1767 |
| <i>HCN3</i> | Intron | rs11264354 | 155256082 | 0.080 | - | - | -2.2332 |
| <i>HCN3</i> | Intron | rs539661366 | 155256093 | - | - | - | -2.1396 |
| <i>HCN3</i> | Intron | rs558388981 | 155256117 | - | - | - | -2.1396 |
| <i>HCN3</i> | Intron | rs113065628 | 155256126 | - | - | - | -2.1207 |
| <i>HCN3</i> | Intron | rs144035252 | 155256131 | - | - | - | -2.2257 |
| <i>HCN3</i> | Intron | rs541106278 | 155256178 | - | - | - | -2.1441 |
| <i>HCN3</i> | Intron | rs559445057 | 155256189 | - | - | - | -2.1332 |
| <i>HCN3</i> | Intron | rs12724449 | 155256231 | 0.256 | - | - | -2.2468 |
| <i>HCN3</i> | Intron | rs141141634 | 155256651 | 0.005 | - | - | -1.6869 |
| <i>HCN3</i> | Intron | rs11264355 | 155257492 | 0.075 | 0.5883 | 1.5265 | -1.7555 |
| <i>HCN3</i> | <i>Missense</i> | rs35001694 | 155257818 | 0.035 | - | - | -1.3076 |
| <i>HCN3/PKLR</i> | Downstream/ 3'UTR | rs3814319 | 155258752 | 0.213 | -1.1374 | 0.0000 | -1.7345 |
| <i>HCN3/PKLR</i> | Downstream/ 3'UTR | rs3814318 | 155258756 | 0.073 | 0.4441 | 1.5357 | -1.8465 |
| <i>HCN3/PKLR</i> | Downstream/ 3'UTR | rs8847 | 155259323 | 0.073 | 0.8985 | -0.2687 | -1.1649 |
| <i>HCN3/PKLR</i> | Downstream/ 3'UTR | rs932972 | 155260096 | 0.074 | 0.9138 | 1.5049 | -1.1687 |
| <i>PKLR</i> | 3'UTR | rs1052177 | 155260350 | 0.075 | 0.9080 | 1.5026 | -1.1447 |
| <i>PKLR</i> | Sinônimo | rs1052176 | 155260383 | 0.075 | 0.9080 | 1.5061 | -1.1491 |

| | | | | | | | |
|------------------|--------------------|-------------------|------------------|--------------|----------------|----------------|----------------|
| PKLR | Intron | rs143970593 | 155260625 | 0.081 | 2.7306 | 0.0000 | -0.9411 |
| PKLR | Intron | rs8177992 | 155261156 | 0.018 | - | - | -0.8716 |
| PKLR | Intron | rs3762272 | 155261777 | 0.006 | 0.0000 | 0.0000 | -0.8383 |
| PKLR | Intron | rs554407398 | 155262002 | 0.001 | - | - | -0.8514 |
| PKLR | Intron | rs4620533 | 155262613 | 0.142 | 1.4367 | 1.4260 | -1.1801 |
| PKLR | Intron | rs575392819 | 155265143 | 0.009 | 0.2984 | 0.4605 | -0.6666 |
| PKLR | Intron | rs2071053 | 155265177 | 0.205 | -0.1638 | -0.1274 | -1.0474 |
| PKLR | Intron | rs138344553 | 155265896 | 0.010 | - | - | -0.9321 |
| PKLR | Intron | rs142339464 | 155265900 | 0.010 | - | - | -0.9321 |
| PKLR | Intron | rs8177968 | 155266335 | 0.063 | 0.8993 | 0.0000 | -1.0929 |
| PKLR | Intron | rs8177967 | 155266513 | 0.045 | 0.1932 | 0.0000 | -1.0212 |
| PKLR | Intron | rs115499949 | 155267322 | 0.004 | - | - | -0.9873 |
| PKLR | Intron | rs142640960 | 155267779 | 0.003 | - | - | -0.9858 |
| PKLR | Intron | rs12067675 | 155268120 | 0.245 | -0.5732 | 0.0458 | -1.1695 |
| PKLR | Intron | rs12741350 | 155268425 | 0.192 | - | - | - |
| PKLR | Intron | rs181993389 | 155268748 | 0.002 | - | - | -1.0715 |
| PKLR | Intron | rs186226317 | 155268860 | 0.077 | -0.8432 | 0.0000 | -1.1063 |
| PKLR | Intron | rs11264357 | 155268958 | 0.257 | 2.4584 | 1.8552 | -1.1670 |
| PKLR | Intron | rs3020781 | 155269776 | 0.191 | -0.1962 | 0.0571 | -1.1327 |
| PKLR | Sinônimo | rs8177962 | 155269991 | 0.003 | - | - | -0.6415 |
| PKLR | Upstream | rs116269004 | 155271936 | 0.051 | -0.2554 | 0.0000 | -0.8194 |
| PKLR | Upstream | rs7524950 | 155272274 | 0.051 | 1.0932 | 1.1857 | -1.0514 |
| PKLR | Upstream | rs74118437 | 155273188 | 0.037 | - | - | -0.7356 |
| PKLR/FDPS | Intergênico | rs4971072 | 155273869 | 0.666 | 0.0000 | -0.3807 | -0.5415 |
| PKLR/FDPS | Intergênico | rs141353488 | 155273917 | 0.001 | - | - | -0.6497 |
| PKLR/FDPS | Intergênico | rs139040128 | 155274035 | 0.082 | -0.1736 | 0.0000 | -0.8103 |
| PKLR/FDPS | Intergênico | rs182120474 | 155274631 | 0.002 | - | - | -0.7542 |
| PKLR/FDPS | Intergênico | rs12032720 | 155274960 | 0.019 | 0.7018 | 1.1398 | -1.2832 |
| PKLR/FDPS | Intergênico | rs142839664 | 155275060 | 0.011 | - | - | -1.1618 |
| PKLR/FDPS | Intergênico | rs74422316 | 155275095 | 0.006 | - | - | -1.1445 |
| PKLR/FDPS | Intergênico | rs7534795 | 155275553 | 0.274 | -0.9151 | 0.3029 | -1.4561 |
| PKLR/FDPS | Intergênico | rs10908461 | 155275727 | 0.000 | 0.1181 | 1.0392 | -1.3825 |
| FDPS | Upstream | rs78003115 | 155277204 | 0.118 | 0.0000 | -0.3978 | -0.4985 |
| FDPS | Upstream | rs6672284 | 155277963 | 0.263 | -0.6813 | 0.1957 | -1.2314 |
| FDPS | Upstream | rs557717725 | 155278057 | 0.000 | - | - | -0.8129 |
| FDPS | Intron | rs2297480 | 155279482 | 0.000 | 0.4141 | 0.9353 | -1.0953 |
| FDPS | Intron | rs566939675 | 155280152 | 0.000 | - | - | -1.0528 |
| FDPS | Intron | rs558224690 | 155280157 | 0.000 | - | - | -1.0457 |
| FDPS | Intron | rs11264358 | 155280164 | 0.000 | 1.3357 | 0.3310 | -1.1215 |
| FDPS | Intron | rs540631227 | 155281386 | 0.000 | - | - | -0.7996 |
| FDPS | Intron | rs10458626 | 155281689 | 0.003 | 0.0000 | 0.0000 | -0.8729 |
| FDPS | Intron | rs11264359 | 155282829 | 0.261 | -0.7450 | 0.0555 | -1.3239 |
| FDPS | Intron | rs59650797 | 155283467 | 0.098 | - | - | -1.1482 |
| FDPS | Intron | rs57041449 | 155283488 | 0.047 | - | - | -1.0751 |
| FDPS | Intron | rs10796941 | 155284261 | 0.011 | - | - | -1.2512 |

Valores de F_{ST} para 60 SNPs individuais, iHS referente à cada população e xpEHH entre europeus e africanos do 1000Genomes. Os valores faltantes (-) podem não ter sido calculados devido aos filtros utilizados em cada análise. Valores em negrito indicam seleção (>2). Em vermelho estão representados os SNPs associados à hanseníase. A variante *missense* (mutação de troca de sentido) ou sinônima estão em região codificante.

1.1 – Análise de expressão gênica global de genes relacionados ao metabolismo de ferro em monócitos de indivíduos saudáveis de Rondonópolis

Avaliamos o perfil de expressão global dos genes relacionados com o metabolismo de ferro, e incluímos os genes *G6PD* e *PKLR*, em amostras de monócitos primários estimulados com *M. leprae* após 48h de infecção (comunicação pessoal com a Dra. Ana Carla – dados não publicados). Observa-se que o *PKLR* está diminuído, enquanto o *G6PD* é superexpressado após o estímulo. Cabe ressaltar que os genes *CD163*, *HP* e *HMOX1* tem a expressão diminuída frente a infecção (Figura 1.1).

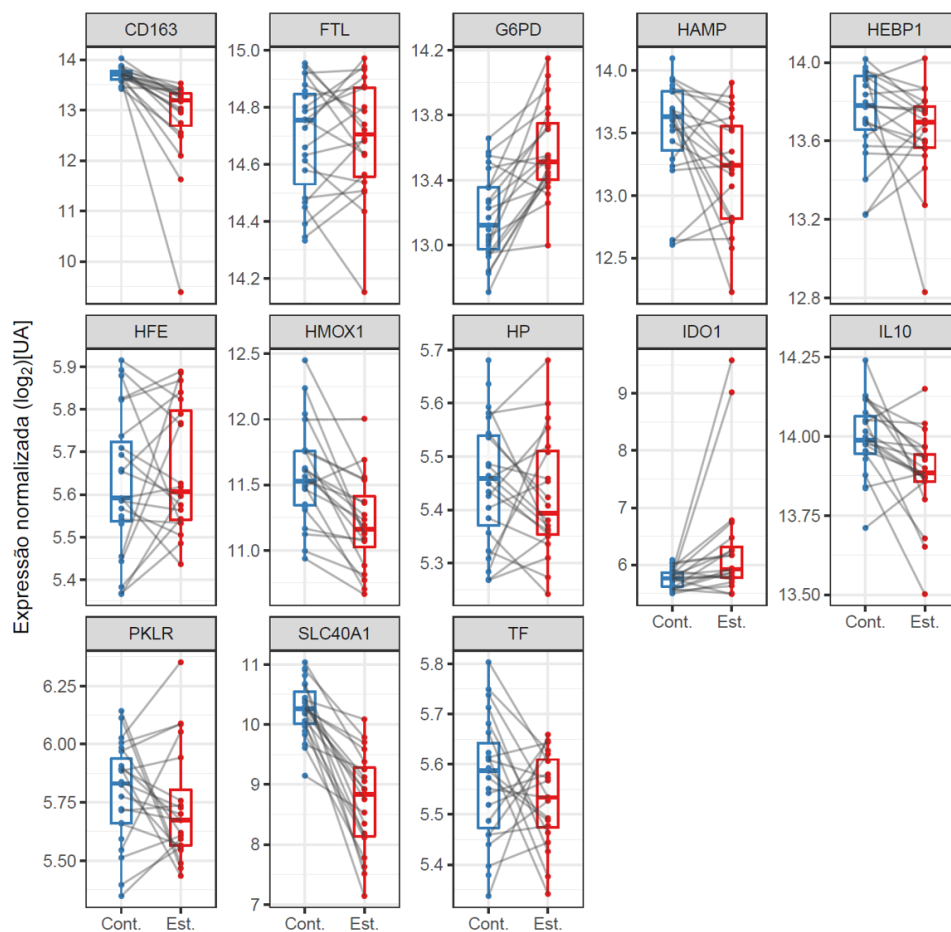


Figura 1.1 – Expressão global dos genes relacionados com o metabolismo do ferro em monócitos primários estimulados com *M. leprae*. Os sobrenadantes foram coletados após 48h de infecção com MOI (1:10).

1.2 – Análise de expressão gênica dos genes *PKLR*, *FTL* e *HP* em sangue total

A fim de verificar possíveis diferenças de expressão dos genes da ferritina (*FTL*) e haptoglobina (*HP*) devido à presença do genótipo de risco do rs1052176, além de confirmar se o esse SNP se apresenta como um eQTL do gene *PKLR* em sangue total, avaliamos a expressão global desses genes entre os grupos estudados, e, posteriormente, entre os genótipos (Figura 1.2). Em acordo com os resultados em soro, observa-se que a expressão de *FTL* (A) e *HP* (B) está levemente acentuada em MB comparado a PB e aos indivíduos saudáveis. A busca por eQTLs em pacientes e indivíduos saudáveis, entretanto, não foi eficaz pois o N amostral ficou prejudicado para a expressão do gene *PKLR*, uma vez que esse gene é pouquíssimo expresso em sangue total, como mostra a Figura 1.3. A pesquisa em um outro banco de eQTLs denominado DICE (do inglês, *Database of Immune Cell eQTLs Expression. Epigenomics*) também reporta o *PKLR* como um gene pouco expresso (<1 transcrito por milhão – TPM – em todas as células) (Figura 1.4).

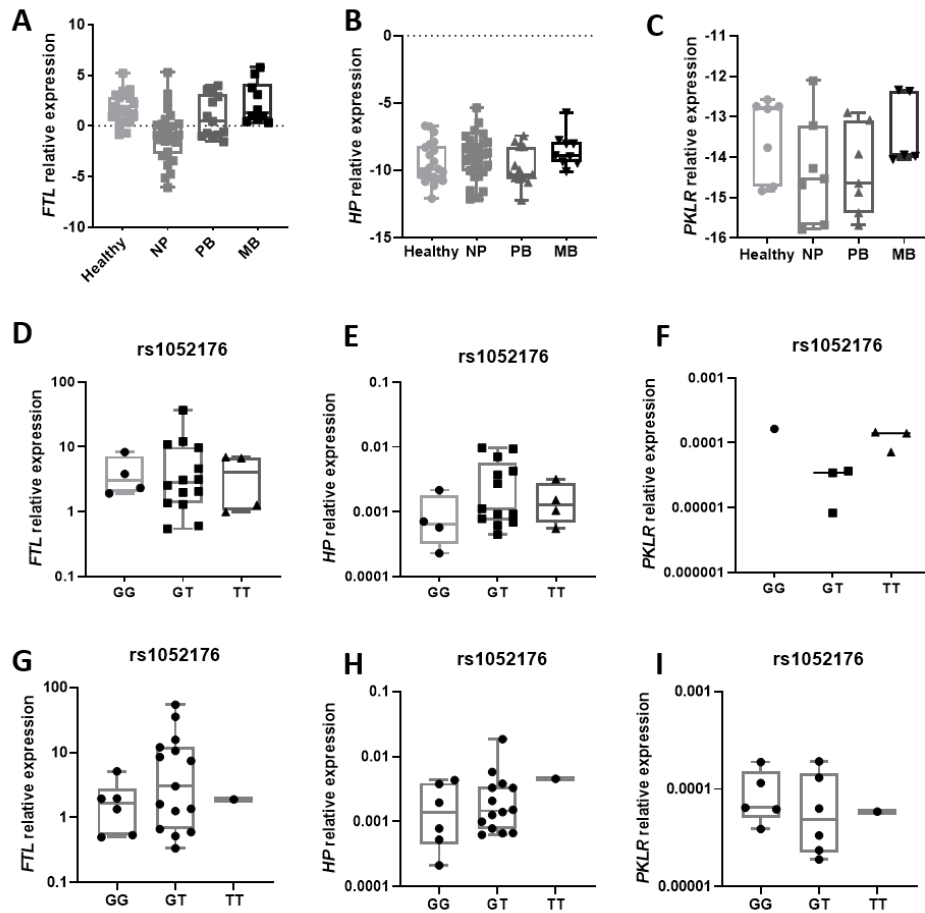


Figura 1.2 – Expressão dos genes *FTL*, *HP* e *PKLR* em sangue total. A, B e C – Expressão gênica entre os grupos estudados (saudáveis, pacientes NP, pacientes PB e pacientes MB). Os dados se encontram transformados para log2 para melhor visualização. D, E e F – Expressão gênica *versus* genótipos do SNP rs1052176 em indivíduos saudáveis. O eixo y está em log10. G, H e I – Expressão gênica *versus* genótipos do SNP rs1052176 em pacientes MB e PB.

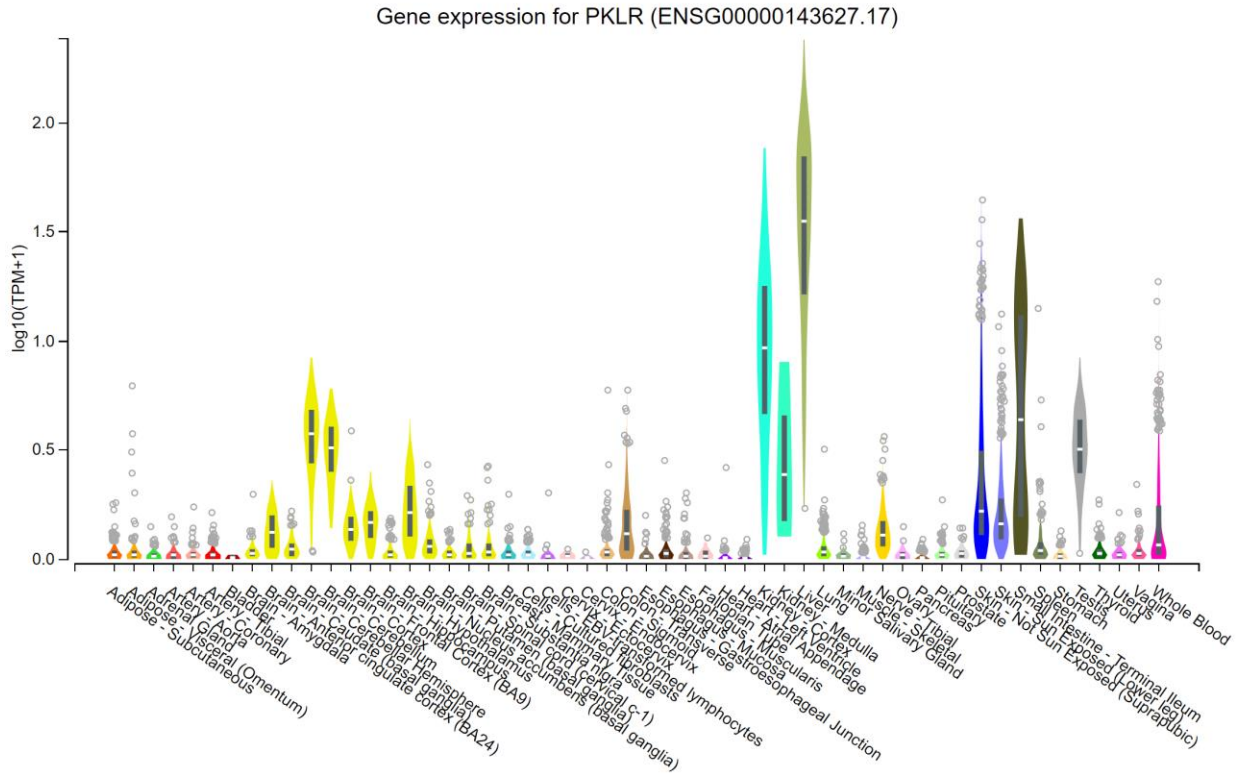


Figura 1.3 – Níveis de expressão do gene *PKLR* em diferentes tecidos segundo o projeto GTEx. O gene é muito expresso no fígado, rins e intestino. Porém, pouco expresso em outros tecidos, bem como em sangue total (destacado em rosa), o que prejudica a detecção pela metodologia empregada nesse estudo.

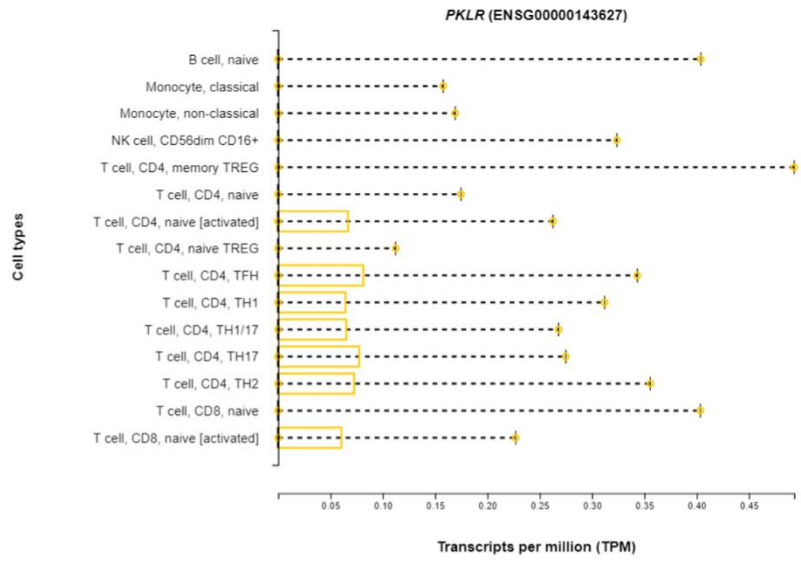


Figura 1.4 – Níveis de expressão do gene *PKLR* em linfócitos e monócitos do banco DICE. O gene *PKLR* é pouco expresso nessas células, apresentando menos de 1 TPM (transcritos por milhão) por célula.

ANEXO III

1 – Artigos em preparação para submissão:

1.1 – Artigo em vias de submissão para a revista *Plos Pathogens* (dezembro de 2019)

Título: Pathogen-selected polymorphisms in the *PKLR* gene are associated with mycobacterial susceptibility in Brazilian and African populations.

Autores: Ohanna Cavalcanti de Lima-Bezerra, Nédio Mabunda, Lucia Elena Alvarado Arnez, Graça Salomé, Amina de Sousa, Fernanda Souza Gomes Kehdy, Caroline de Sales-Marques, Fernanda Saloum Neves Manta, Rafaela Mota Dias, Lais Ferreira Pereira, Kelly Nunes, Mateus Gouveia, Thyago Leal Calvo, Ana Carla Pereira Latini, André Luiz Leturiondo, Fabíola da Costa Rodrigues, Ariani Batista Noronha, Cynthia de Oliveira Ferreira, Carolina Talhari, Jamile Leão Rêgo, Lea Cristina de Carvalho Castellucci, Eduardo Tarazona-Santos, Elizeu Fagundes de Carvalho, Diogo Meyer, Roberta Olmo Pinheiro, Antonio Guilherme Pacheco, Ilesh V Jani e Milton Ozório Moraes.

Pathogen-selected polymorphisms in the *PKLR* gene are associated with mycobacterial susceptibility in Brazilian and African populations

1 Ohanna Cavalcanti de Lima-Bezerra¹, Nédio Mabunda², Lucia Elena Alvarado Arnez¹, Graça Salomé², Amina de
2 Sousa², Fernanda Souza Gomes Kehdy¹, Caroline de Sales-Marques¹, Fernanda Saloum Neves Manta¹, Rafaela
3 Mota Dias¹, Lais Ferreira Pereira¹, Kelly Nunes⁸, Mateus Gouveia, Thyago Leal Calvo¹, Ana Carla Pereira Latini³,
4 André Luiz Leturiondo⁴, Fabíola da Costa Rodrigues⁴, Ariani Batista Noronha⁴, Cynthia de Oliveira Ferreira⁴,
5 Carolina Talhari⁴, Jamile Leão Rêgo⁵, Lea Cristina de Carvalho Castellucci⁵, Eduardo Tarazona-Santos⁶, Elizeu
6 Fagundes de Carvalho⁷, Diogo Meyer⁸, Roberta Olmo Pinheiro¹, Antonio Guilherme Pacheco⁹, Ilesh V Jani² and
7 Milton Ozório Moraes¹

¹Laboratory of Leprosy, Oswaldo Cruz Institute, FIOCRUZ, Rio de Janeiro/RJ, Brazil.

²Laboratory of Molecular Virology, National Institute of Health, Maputo, Mozambique.

³Instituto Lauro de Souza Lima (ILSL), Bauru/SP, Brazil.

⁴Alfredo da Matta Foundation, Manaus, Amazonas, Brazil.

⁵Instituto Nacional de Ciência e Tecnologia em Doenças Tropicais, Salvador /BA, Brazil.

⁶Departament of Biology, Institute of Biological Sciences, Federal University of Minas Gerais (UFMG), Belo Horizonte/MG, Brazil.

⁷DNA Diagnostic Laboratory (LDD), State University of Rio de Janeiro (UERJ), Rio de Janeiro/RJ, Brazil.

8 ⁸Laboratory of Evolutionary Genetics and Biology, Instituto de Biociências, University of São Paulo, São Paulo/SP, Brazil,

⁹Science Computational Program (PROCC), FIOCRUZ, Rio de Janeiro/RJ, Brazil.

9 Corresponding Author

10 Dr. Milton Ozório Moraes

11 Laboratório de Hanseníase, Instituto Oswaldo Cruz,

12 Fundação Oswaldo Cruz, FIOCRUZ-RJ,

13 Rio de Janeiro, Brazil

14 E-mail: milton.moraes@fiocruz.br

15 Keywords: *PKLR*, genetic association, mycobacteria, ancestry, pathogen selection

16 **Abstract (300 words)**

1.2 – Artigo em vias de submissão para a revista *Plos One* (janeiro de 2020)

Título: Characterization of human glucose-6-phosphate dehydrogenase A⁻ mutation in *Mycobacterium leprae* infection and leprosy predisposition.

Autores: Rychelle Clayde Affonso Medeiros^{1a}, Ohanna Cavalcanti de Lima-Bezerra^{1a}, Leonardo Ribeiro Batista-Silva, Fernanda de Souza Gomes Kehdy, Robertha Mariana Rodrigues Lemes, Karina Girardi de Vasconcelos, Thyago Leal Calvo, Flávio Alves Lara e Milton Ozório Moraes.

Characterization of human glucose-6-phosphate dehydrogenase A-mutation in *Mycobacterium leprae* infection and leprosy predisposition

Rychelle Clayde Affonso Medeiros^{1a}, Ohanna Cavalcanti de Lima-Bezerra^{1a}, Leonardo Ribeiro Batista-Silva¹, Fernanda de Souza Gomes Kehdy¹, Robertha Mariana Rodrigues Lemes², Karina Girardi de Vasconcelos², Thyago Leal Calvo¹, Flávio Alves Lara² and Milton Ozório Moraes¹

¹Laboratório de hanseníase, Oswaldo Cruz Institute, Rio de Janeiro, 21040-900, RJ, Brazil. ²Laboratório de Microbiologia celular, Oswaldo Cruz Institute, Rio de Janeiro, 21040-900, RJ; ³Theses authors contributed equally to this work.

Corresponding Author

Dr. Milton Ozório Moraes

Laboratório de Hanseníase, Instituto Oswaldo Cruz,
Fundação Oswaldo Cruz, FIOCRUZ-RJ,
Rio de Janeiro, Brazil
Email: milton.moraes@fiocruz.br

2 – Artigos submetidos como colaboradora ao longo do período de doutoramento:

2.1 – Artigo submetido na revista *Cells Reports* (outubro de 2019) e disponível em pré-print

Título: Congenital Zika Syndrome is associated with maternal genetic background

Autores: Tamiris Azamor, Daniela Prado Cunha, Andréa Marques Vieira da Silva, Ohanna Cavalcanti de Lima Bezerra, Marcelo Ribeiro-Alves, Thyago Leal Calvo, Fernanda de Souza Gomes Kehdy, Fernanda Saloum de Neves Manta, Thyago Gomes Pinto, Laís Pereira Ferreira, Elyzabeth Avvad Portari, Letícia da Cunha Guida, Leonardo Gomes, Maria Elisabeth Lopes Moreira, Elizeu de Carvalho, Cynthia Chester Cardoso, Marcelo Muller, Patrícia Cristina da Costa Neves, Zilton Vasconcelos, Milton Ozorio Moraes.

Congenital Zika Syndrome is associated with maternal genetic background

Authors

Tamiris Azamor^{1,2}, Daniela Prado Cunha³, Andréa Marques Vieira da Silva², Ohanna Cavalcanti de Lima Bezerra¹, Marcelo Ribeiro-Alves⁴, Thyago Leal Calvo¹, Fernanda de Souza Gomes Kehdy¹, Fernanda Saloum de Neves Manta¹, Thyago Gomes Pinto¹, Lais Pereira Ferreira¹, Elyzabeth Avvad Portari³, Leticia da Cunha Guida³, Leonardo Gomes³, Maria Elisabeth Lopes Moreira³, Elizeu de Carvalho⁵, Cynthia Chester Cardoso⁶, Marcelo Muller², Patricia Cristina da Costa Neves², Zilton Vasconcelos³, Milton Ozorio Moraes^{1*}

1 Laboratório de Hanseníase. Instituto Oswaldo Cruz. Fiocruz, Brazil.

2 Vice-Diretoria de Desenvolvimento Tecnológico. Instituto de Tecnologia em Imunobiológicos, Fiocruz, Brazil.

3 Unidade de Pesquisa Clínica, Instituto Nacional de Saúde da Mulher, da Criança e do Adolescente Fernandes Figueira. Fiocruz, Brazil.

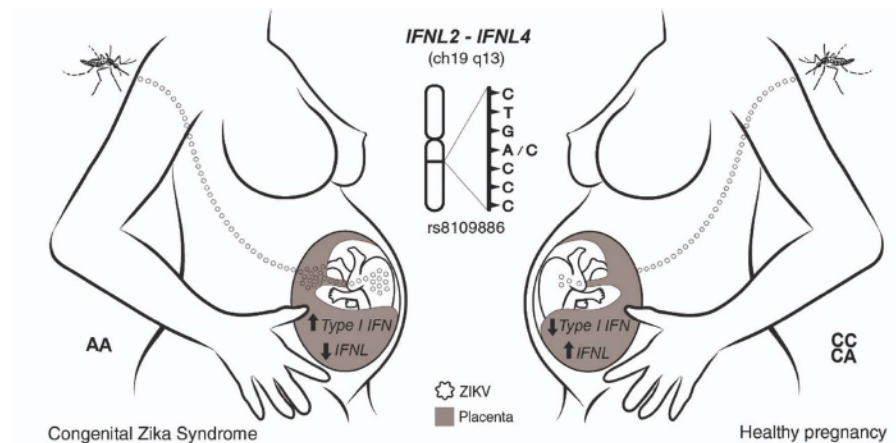
4 Laboratório de Pesquisa Clínica em DST/AIDS. Instituto Nacional de Infectologia. Fiocruz, Brazil.

5 Laboratório de Diagnóstico por DNA. Universidade do Estado do Rio de Janeiro, Brazil.

6 Laboratório de Virologia Molecular. Universidade Federal do Rio de Janeiro, Brazil.

*Correspondence: milton.moraes@fiocruz.br

Graphical abstract



2.2 – Artigo submetido na revista *Plos Neglected Diseases* (5 de setembro de 2019)

Título: Association of *NOD2* and *IFNG* single nucleotide polymorphisms with leprosy in the Amazon ethnic admixed population.

Autores: André Luiz Leturiondo, Ariani Batista Noronha, Carla Yael Ribeiro Mendonça, Cynthia de Oliveira Ferreira, Lucia Elena Alvarado-Arnez, Fernanda Saloum de Neves Manta, Ohanna Cavalcanti de Lima Bezerra, Elizeu Fagundes de Carvalho, Milton Ozório Moraes, Fabíola da Costa Rodrigues e Carolina Talhari.

PLOS Neglected Tropical Diseases
Association of NOD2 and IFNG single nucleotide polymorphisms with leprosy in the Amazon ethnic admixed population
 --Manuscript Draft--

| | |
|--|--|
| Manuscript Number: | PNTD-D-19-01538 |
| Full Title: | Association of NOD2 and IFNG single nucleotide polymorphisms with leprosy in the Amazon ethnic admixed population |
| Short Title: | Association of NOD2 and IFNG in the Amazon ethnic admixed population |
| Article Type: | Research Article |
| Keywords: | leprosy; genetic ancestry; SNPs; admixed population |
| Corresponding Author: | André Luiz Leturiondo Alfredo da Matta Foundation Manaus, Amazonas BRAZIL |
| Corresponding Author Secondary Information: | |
| Corresponding Author's Institution: | Alfredo da Matta Foundation |
| Corresponding Author's Secondary Institution: | |
| First Author: | André Luiz Leturiondo |
| First Author Secondary Information: | |
| Order of Authors: | André Luiz Leturiondo Ariani Batista Noronha Carla Yael Ribeiro Mendonça Cynthia de Oliveira Ferreira Lucia Elena Alvarado-Arnez Fernanda Saloum de Neves Manta Ohanna Cavalcanti de Lima Bezerra Elizeu Fagundes de Carvalho Milton Ozório Moraes Fabiola da Costa Rodrigues Carolina Talhari |
| Order of Authors Secondary Information: | |
| Abstract: | <p>Background Leprosy is a chronic infectious disease, caused by <i>Mycobacterium leprae</i>, which affects skin and peripheral nerves. Polymorphisms in genes associated with autophagy, metabolism, innate and adaptive immunity confer susceptibility to leprosy. However, these associations need to be confirmed through independent replication studies in different ethnicities. The population from Amazon state (northern Brazil) is admixed and it contains the highest proportion of Native American genetic ancestry in Brazil.</p> <p>Methods We conducted a case-control study for leprosy in which we tested thirteen previously associated SNPs in key immune response regulating genes: TLR1 (rs4833095), NOD2 (rs751271, rs8057341), TNF (rs1800629), IL10 (rs1800871), CCDC122/LACC1 (rs4942254), PACRG/PRKN (rs9356058), IFNG (rs2430561), IL6 (rs2069845), LRRK2 (rs7298930, rs3761863), IL23R (rs76418789) and TYK2 (rs55882956). Genotyping was carried out by allelic discrimination in 967 controls and 412 leprosy patients. Association with susceptibility was assessed by logistic regression analyses adjusted for the following covariates: gender, age and ancestry.</p> |

3 – Artigos publicados como colaborador durante o período do doutoramento:

3.1 – Artigo publicado na revista *Microbes and Infection* (24 de junho de 2017)

Título: Indoleamine 2,3-dioxygenase and iron are required for *Mycobacterium leprae* survival.

Autores: Mayara Garcia de Mattos Barbosa¹, Rhana Berto da Silva Prata¹, Priscila Ribeiro Andrade, Helen Ferreira, Bruno Jorge de Andrade Silva, Jessica Araujo da Paixao de Oliveira, Tayna Quintella Assis, Thiago Gomes de Toledo-Pinto, Ohanna Cavalcanti de Lima Bezerra, Jose Augusto da Costa Nery, Patricia Sammarco Rosa, Marcelo Torres Bozza, Flavio Alves Lara, Milton Ozorio Moraes, Veronica Schmitz, Euzenir Nunes Sarno, Roberta Olmo Pinheiro.



Original article

Indoleamine 2,3-dioxygenase and iron are required for *Mycobacterium leprae* survival

Mayara Garcia de Mattos Barbosa ^{a,1}, Rhana Berto da Silva Prata ^{a,b,1}, Priscila Ribeiro Andrade ^a, Helen Ferreira ^a, Bruno Jorge de Andrade Silva ^a, Jéssica Araújo da Paixão de Oliveira ^a, Tainá Quintella Assis ^a, Thiago Gomes de Toledo-Pinto ^a, Ohanna Cavalcanti de Lima Bezerra ^a, José Augusto da Costa Nery ^a, Patricia Sammarco Rosa ^d, Marcelo Torres Bozza ^b, Flávio Alves Lara ^c, Milton Ozório Moraes ^a, Veronica Schmitz ^a, Euzenir Nunes Sarno ^a, Roberta Olmo Pinheiro ^{a,*}

^a Leprosy Laboratory, Oswaldo Cruz Institute, Oswaldo Cruz Foundation, Rio de Janeiro, RJ, Brazil

^b Inflammation and Immunity Laboratory, Federal University of Rio de Janeiro, Rio de Janeiro, RJ, Brazil

^c Cellular Microbiology Laboratory, Oswaldo Cruz Foundation, Rio de Janeiro, RJ, Brazil

^d Laboratory Animal House, Lauro de Souza Lima Institute, Bauru, SP, Brazil

Received 28 November 2016; accepted 24 June 2017

Available online ■■■

Abstract

Our previous study has demonstrated that IL-10 may modulate both indoleamine 2,3-dioxygenase (IDO) and CD163 expression in lepromatous leprosy (LL) cells, favoring *Mycobacterium leprae* persistence through induction of regulatory pathways and iron storage. Here, we observed that in LL lesion cells there is an increase in the expression of proteins involved in iron metabolism such as hemoglobin (Hb), haptoglobin, heme oxygenase 1 and transferrin receptor 1 (TfR1) when compared to tuberculoid leprosy (BT) cells. We also found increased iron deposits and diminished expression of the iron exporter ferroportin 1 in LL lesion cells. Hemin, but not FeSO₄ stimulation, was able to enhance *M. leprae* viability by a mechanism that involves IDO. Analysis of cell phenotype in lesions demonstrated a predominance of M2 markers in LL when compared with BT lesion cells. A positive correlation between CD163 and PPARG with the bacillary index (BI) was observed. In contrast, TNF, STAT1 and CSF2 presented a negative correlation with the BI. In summary, this study demonstrates that iron may regulate IDO expression by a mechanism that involves IL-10, which may contribute for the predominance of M2-like phenotype in LL lesions that favors the phagocytosis and maintenance of *M. leprae* in host cells.

© 2017 Institut Pasteur. Published by Elsevier Masson SAS. All rights reserved.

Keywords: Leprosy; Macrophages; Iron; Indoleamine 2,3-dioxygenase; *Mycobacterium leprae*

1. Introduction

Macrophages are pivotal cells in the pathogenesis of leprosy, caused by *Mycobacterium leprae*. Each of the

polarized clinical presentations in leprosy is associated with different macrophagic populations in the host tissue. Whereas tuberculoid leprosy (BT) macrophages present a pro-inflammatory profile, in lepromatous leprosy (LL) these cells form an ideal niche for parasitic entry and replication, with increased scavenger receptors and anti-inflammatory cytokines expression [1–4]. It has been recently demonstrated that the exposure to *M. leprae* has an inhibitory effect on a pro-inflammatory phenotype through a mechanism phenolic

* Corresponding author. Leprosy Laboratory, Oswaldo Cruz Foundation, Avenida Brasil, 4365, Mangueiras, Rio de Janeiro, RJ, ZIP code 21040-360, Brazil.

E-mail address: roberta.olmo@pq.cnpq.br (R.O. Pinheiro).

¹ These authors contribute equally for this manuscript.

3.2 – Artigo publicado na revista *The FASEB Journal* (22 de novembro de 2019)

Título: Liquid-liquid phase separation and fibrillation of the prion protein modulated by a high-affinity DNA aptamer.

Autores: Carolina O. Matos, Yulli M. Passos, Mariana J. do Amaral, Bruno Macedo, Matheus H. Tempone, Ohanna C. L. Bezerra, Milton O. Moraes, Marcius S. Almeida, Gerald Weber, Sotiris Missailidis, Jerson L. Silva, Vladimir N. Uversky, Anderson S. Pinheiro e Yraima Cordeiro.


RESEARCH ARTICLE

Liquid-liquid phase separation and fibrillation of the prion protein modulated by a high-affinity DNA aptamer

Carolina O. Matos, Yulli M. Passos, Mariana J. do Amaral, Bruno Macedo, Matheus H. Tempone

... See all authors 

First published: 22 November 2019 | <https://doi.org/10.1096/fj.201901897R>

Read the full text 



PDF



TOOLS



SHARE

Abstract

Structural conversion of cellular prion protein (PrP^C) into scrapie PrP (PrP^{Sc}) and subsequent aggregation are key events associated with the onset of transmissible spongiform encephalopathies (TSEs). Experimental evidence supports the role of nucleic acids (NAs) in assisting this conversion. Here, we asked whether PrP undergoes liquid-liquid phase separation (LLPS) and if this process is modulated by NAs. To this end, two 25-mer DNA aptamers, A1 and A2, were selected against the globular domain of recombinant murine PrP (rPrP⁹⁰⁻²³¹) using SELEX methodology. Multiparametric structural analysis of these aptamers revealed that A1 adopts a hairpin conformation. Aptamer binding caused partial unfolding of rPrP⁹⁰⁻²³¹ and modulated its ability to undergo LLPS and fibrillate. In fact, although free rPrP⁹⁰⁻²³¹ phase separated into large droplets, aptamer binding increased the number of droplets but noticeably reduced their size. Strikingly, a modified A1 aptamer that does not adopt a hairpin structure induced formation of amyloid fibrils on the surface of the droplets. We show here that PrP undergoes LLPS, and that the PrP interaction with NAs modulates phase separation and promotes PrP fibrillation in a NA structure and concentration-dependent manner. These results shed new light on the roles of NAs in PrP misfolding and TSEs.

Supporting Information 